DISS. ETH NO. 24361

CLIMATE EXTREMES AND THEIR IMPACT ON ECOSYSTEM–ATMOSPHERE INTERACTIONS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

SEBASTIAN DOMINIK SIPPEL

M. Sc. Geoökologie, Universität Bayreuth

M. Sc. Environmental Change and Management, University of Oxford

born on 20.12.1987

citizen of Germany

accepted on the recommendation of

Prof. Dr. Sonia I. Seneviratne, examiner

Dr. Miguel D. Mahecha, co–examiner

Prof. Dr. Martin Heimann, co–examiner

Prof. Dr. Nicolas Gruber, co–examiner

2017

Contents

List of Figures

List of Tables

Summary

Extreme weather and climate events (summarised as 'climate extremes' from here onwards) are a crucial aspect of Earth's climatic variability. However, climate extremes are frequently associated with adverse impacts on socio-economic and ecological systems. For example, heat in combination with drought may severely affect the functioning of terrestrial ecosystems, and in some cases these events have the potential to undo several years of ecosystem carbon sequestration. Moreover, the intensity and frequency of several types of climate extremes, such as heat, cold, and heavy rainfall, have been changing in recent years. These changes are projected to continue in the 21st century, thus raising concerns about the capacity of ecological and socio-economic systems to cope with these events in the future.

Nonetheless, our scientific understanding of climate extremes and the mechanistic pathways through which these events propagate into ecological or socio-economic systems, remains limited. The impact of climate extremes varies widely depending on their type and spatio-temporal structure, and these impacts are mediated by the vulnerability and exposure of the system under scrutiny. Therefore, the quantification of these phenomena, and the attribution to their respective drivers across space and time is often ambiguous. Accordingly, closing scientific knowledge gaps and improving methodologies to scrutinise climate extremes and their impacts constitutes a research priority of high societal relevance.

The overarching objective of the present PhD thesis is to improve the quantification of, and contribute to the understanding of climate extremes and their impact on ecosystem-atmosphere interactions. To address these objectives, the thesis relies on joint analyses and integration of observation-based datasets and model ensemble simulations. Specifically, the thesis explores (1) a wide range of generic statistical-methodological considerations, (2) approaches to enable sound process-oriented model ensemble simulations using observation-based constraints, towards (3) a comprehensive attribution of ecosystem impacts arising from climate extremes.

1. **Statistical quantification of extremes in observed or simulated spatio-temporal gridded datasets** (Part I).
   An investigation and quantification of extremes in spatio-temporal datasets requires robust statistical methodologies and diagnostics. Therefore, the thesis scrutinises statistical methods, both empirically and analytically, to explore recent changes in temperature and precipitation extremes in gridded observations. These analyses reveal that conventional statistical methods that are based on a reference period standardisation might induce substantial biases in spatially aggregated estimates of extremes. For example, the occurrence of extremes that exceed two standard deviations in standardised data could be overestimated by 48.2% outside a given reference period of 30 years in independent and identically distributed Gaussian data. Analytical corrections for these kinds of statistical errors are derived in the thesis.
   Because climate extremes are inevitably rare in temporally and spatially limited observational records, ensemble simulations constitute an indispensable and complementary tool to scrutinise climate extremes from a statistical perspective, circumventing small sample issues in observations. Hence, the thesis also illustrates how model ensembles can be used as surrogate observations to benchmark statistical methods and metrics for an accurate assessment of climate extremes in observations.

2. **Observation-based constraints improve model ensemble simulations of climate extremes and ecosystem impacts** (Part II).
   Climate model ensemble simulations generated for the purpose of quantifying and attributing climate extremes typically exhibit biases in their output that hinder any straightforward simulation or assessment of impacts. Therefore, I develop, apply, and evaluate tools to constrain climate model ensembles based on observational diagnostics related to land-atmosphere interactions. The application of these constraints simultaneously reduces multivariate biases in model ensembles and thus might offer a novel route

to bias correction for climate impact simulations and analyses of climate extremes.

3. **Extremes events in the terrestrial biosphere: drivers and attribution** (Part III).
Linking or attributing extreme responses in the terrestrial biosphere to climatic drivers is not straightforward because respective analyses often rely on small sample sizes or even singular events in observations. Therefore, I construct an ensemble of climate-ecosystem impact simulations, constrained by observational diagnostics developed in Part II, that is designed (a) to systematically investigate and attribute changes in the intensity and frequency of simulated ecosystem productivity extremes ('EPEs') to the respective drivers, and (b) to assess the effect of timing and seasonal interaction of EPEs in the terrestrial biosphere. Thus, a perspective centred on ecosystem impacts is adopted.
An analysis of these simulations reveals that (a) recent trends in the intensity of EPEs in Europe are contrasting seasonally, i.e. spring EPEs show consistent trends towards increased carbon uptake, while trends in summer EPEs are predominantly negative (higher net carbon release under drought and heat in summer) or close to neutral. Furthermore, the analyses reveal that (b) spring-summer interacting carbon cycle effects due to climate extremes and thus their timing plays an important role in shaping EPEs in Europe. These interacting effects include both partial compensation of drought or heat wave induced carbon losses in summer due to increased carbon uptake in the preceding spring (driven by higher temperatures), and conversely, spring 'carry-over' effects into summer arising from depleted soil moisture that exacerbates summer carbon losses.

In conclusion, the thesis lays out a comprehensive framework for systematically quantifying and attributing the impacts of climate extremes in the terrestrial biosphere using joint analyses of observations and model ensembles. The thesis shows that firstly, scrutinising statistical methods and diagnostics, and evaluating observation-based constraints on model ensembles, are key to an improved understanding as well as quantification of climate extremes and their impacts. Secondly, a consequent probabilistic interpretation of climate-ecosystem model

ensemble simulations offers novel perspectives on the mechanistic pathways and interacting effects of terrestrial ecosystem responses to climate extremes.

Zusammenfassung

Extreme Wetter- und Klimaereignisse (hier zusammengefasst als 'Klimaextreme') sind ein zentraler Aspekt klimatischer Variabilität des Erdsystems. Allerdings sind diese Ereignisse häufig mit negativen Auswirkungen auf sozioökonomische und ökologische Systeme verbunden. Als Beispiel können Hitzewellen genannt werden, die in Verbindung mit Trockenheit die Funktionsweise terrestrischer Ökosysteme nachhaltig beeinträchtigen, und in einigen Fällen sogar die Netto-Kohlenstoffaufnahme einiger Jahre zunichte machen können. Ferner wurden in den letzten Jahren und Jahrzehnten Veränderungen in der Intensität wie auch Häufigkeit von Klimaextremen, wie beispielsweise Hitze- oder Kältewellen und Starkniederschlägen festgestellt. Diese Veränderungen werden sich voraussichtlich im 21. Jahrhundert fortsetzen, und infolgedessen geben diese Prognosen Anlass zu Bedenken, ob und inwiefern ökologische und sozioökonomische Systeme diese Ereignisse in Zukunft bewältigen können.

Dennoch ist das wissenschaftliche Verständnis von Klimaextremen und den Prozessen, die Auswirkungen in ökologischen und sozioökonomischen Systemen verursachen, derzeit begrenzt. Die Auswirkungen von Klimaextremen variieren stark je nach Art und räumlich-zeitlicher Struktur des jeweiligen Ereignisses, und Auswirkungen werden außerdem durch Vulnerabilität und Exposition des jeweiligen Systems beeinflusst. Deshalb ist die Quantifizierung von Klimaextremen und deren Auswirkungen, wie auch die Zuordnung zu deren jeweiligen Ursachen in Raum und Zeit oft nicht eindeutig, und im Allgemeinen unsicher. Dementsprechend stellen der wissenschaftliche Erkenntnisgewinn und methodische Verbesserungen zur Analyse von Klimaextremen und deren Auswirkungen ein wichtiges Forschungziel mit hoher gesellschaftlicher Relevanz dar.

Das vorrangige Ziel dieser Dissertation ist es, die Quantifizierung und das Verständnis von Klimaextremen und deren Auswirkungen auf die Funktionsweise

terrestrischer Ökosysteme, insbesondere Ökosystem-Atmosphäre-Interaktionen, zu verbessern. Zentraler methodischer Ansatzpunkt dieser Arbeit ist dabei die Analyse und Integration von Beobachtungs-basierten Datensätzen und modell-basierten Ensemble-Simulationen. Im Einzelnen untersucht die Dissertation (1) statistisch-methodische Fragestellungen zur Quantifizierung von Klimaextremen, (2) Ansätze, die verbesserte prozess-orientierte Ensemble-Simulationen mit Hilfe beobachtungs-basierter Eigenschaften des Klimasystems ermöglichen, um (3) eine umfassende Zuordnung der Ökosystem-Auswirkungen von Klimaextremen zu deren Ursachen vorzunehmen.

1. **Statistische Quantifizierung von Extremen in beobachteten und simulierten räumlich-zeitlichen Datensätzen** (Thema I).

   Die Analyse und Quantifizierung von Extremen in räumlich-zeitlichen Datensätzen erfordert robuste statistische Methoden. Daher untersucht diese Dissertation sowohl empirisch als auch analytisch statistische Methodik, die zur Diagnostizierung von Veränderungen in Temperatur- und Niederschlagsextremen in gitter-basierten Beobachtungsdatensätzen verwendet werden. Diese Analysen zeigen, dass konventionelle statistische Methoden, denen eine Standardisierung auf Basis einer Referenzperiode zugrunde liegt, erhebliche Fehler in räumlich aggregierten Schätzungen von Extremereignissen hervorrufen können. Zum Beispiel würde die Auftretenswahrscheinlichkeit von Extremen, die in standardisierten Daten zwei Standardabweichungen überschreiten, um 48,2% außerhalb eines gegebenen 30-jährigen Referenzzeitraums in unabhängigen und identisch verteilten Gaußschen Daten überschätzt werden. Eine analytische Korrektur dieses statistischen Artefakts wird in der Dissertation hergeleitet.
   Klimaextreme treten definitionsgemäß in zeitlich und räumlich begrenzten Beobachtungsdatensätzen relativ selten auf. Deshalb stellen modell-basierte Ensemble-Simulationen ein wichtiges komplementäres Instrument dar, um Klimaextreme aus statistischer Perspektive zu untersuchen und das Problem kleiner Stichproben in Beobachtungen zu umgehen. Diese Dissertation zeigt daher ebenso auf, wie Ensemble-Simulationen als Surrogat-Beobachtungen verwendet werden können, um statistische Methoden und

Diagnostiken zur exakten Bewertung von Klimaextremen in Beobachtungen zu evaluieren.

2. **Verbesserung von Ensemble-Simulationen zur Analyse von Klimaextremen und Ökosystem-Auswirkungen durch beobachtungs-basierte Constraints**[1] (Thema II)
   Modell-basierte Ensemble-Simulationen zeigen häufig systematische Fehler in simulierten Klimavariablen, die eine direkte Anwendung zur Quantifizierung und ursächlichen Zuordnung von Klimaextremen und deren Auswirkungen erschweren. Deshalb entwickle und evaluiere ich in dieser Dissertation Methoden, die eine Filterung von Ensemble-Simulationen mit Hilfe beobachtungs-basierter Diagnostiken (z.B. Diagnostiken von Ökosystem-Atmosphäre-Interaktionen), ermöglichen. Die Anwendung dieser Filter reduziert systematische Fehler in mehreren Variablen und der multivariaten Korrelationsstruktur in Ensemble-Simulationen und eröffnen so eine neue Möglichkeit zur systematischen Fehlerkorrektur für Simulationen von Klimafolgen oder Analysen von Klimaextremen.

3. **Extremereignisse in der terrestrischen Biosphäre: Ursachen und Zuordnung** (Thema III)

   Die Verknüpfung oder Zuordnung von extremen Ökosystemreaktionen zu klimatischen Ursachen ist oft nicht direkt möglich, da sich solche Analysen häufig auf kleine Stichprobengrößen oder sogar einzelne Ereignisse in Beobachtungen stützen. Deshalb generiere ich Ensemble-Simulationen des Klima-Ökosystem Wirkungsgefüges mit Hilfe eines Ökosystemmodells, die (a) zur systematischen Untersuchung von Veränderungen in Intensität und Häufigkeit von simulierten Extremen in Ökosystemproduktivität (‘EÖP’), und zur Zuordnung der jeweiligen Ursachen verwendet werden können, und (b) die Rückschlüsse über das zeitliche und saisonale Zusammenwirkens von EÖPs in der terrestrischen Biosphäre zulassen. Somit wird eine auf die Ökosystem-Auswirkungen von Klimaextremen fokussierte Perspektive eingenommen.

[1] hier: Constraints sinngemäß als ‘Filter’ übersetzt

Eine Analyse dieser Simulationen zeigt (a) saisonal gegenläufige Trends in der Intensität von simulierten EÖPs in Europa, d.h. EÖPs im Frühjahr zeigen robuste Trends hin zu erhöhter Ökosystem-Kohlenstoffaufnahme, während EÖPs im Sommer überwiegend negative (d.h. höhere Netto-Kohlenstofffreisetzung unter Trockenheit und Hitze im Sommer) oder neutrale Trends aufweisen. Diese Analysen zeigen außerdem, dass (b) Ökosystem-Interaktionen zwischen Frühling und Sommer, und somit der Zeitpunkt des Auftretens von Klimaextremen, eine wichtige Rolle für EÖPs in Europa einnehmen. Diese Wechselwirkungen beinhalten sowohl die Teilkompensation von Dürre- oder Hitze-induzierten Kohlenstoffverlusten im Sommer aufgrund einer erhöhten Kohlenstoffaufnahme im vorangegangenen Frühling (aufgrund höherer Temperaturen); wie auch den gegensätzlichen Effekt, nämlich eine negative Nachwirkung von Frühlingseffekten im Sommer durch reduzierte Bodenfeuchtigkeit, die Ökosystem-Kohlenstoffverluste im Sommer verschärfen kann.

Insgesamt legt die Dissertation einen umfassenden methodischen Ansatz für die systematische Quantifizierung und ursächliche Zuordnung von Klimaextremen und deren Auswirkungen auf Ökosystem-Atmosphäre-Interaktionen vor, der auf einer Analyse von Beobachtungen und Ensemble-Simulationen basiert. Im Hauptergebnis zeigt die Dissertation, dass eine umfangreiche Untersuchung von statistischen Methoden zur Quantifizierung von Klimaextremen, und die Anwendung von beobachtungs-basierten Diagnostiken als Filter für Ensemble-Simulationen entscheidend zu einem besseren Verständnis sowie Quantifizierung von Klimaextremen und deren Auswirkungen beitragen kann. Weiterhin eröffnet eine probabilistische Interpretation von Klima-Ökosystem Ensemblesimulationen neue Perspektiven auf Prozesse und wechselwirkende Effekte in der Funktionsweise terrestrischer Ökosysteme unter Klimaextremen.

Acknowledgements

This thesis would not have been possible without the support of many people whom I wish to thank wholeheartedly:
First and foremost, I would like to thank Miguel Mahecha for his continuous support and guidance throughout all stages of this thesis, his enthusiasm, and inspiring ideas. I am grateful for in-depth scientific discussions, many opportunities to participate in workshops, conferences, group meetings, may it be at the institute or elsewhere, and for a positive and collaborative spirit in the 'Empirical Inference of the Earth System' group. I am especially grateful to Sonia Seneviratne for supervising me as an external PhD student, for invaluable scientific discussions, feedback, and her interest in this thesis. She allowed me to join the Land-Climate Dynamics group at ETH Zurich for a research stay, which was a memorable experience and a very productive time. I am indebted to Friederike Otto for motivating me to pursue this thesis, her research ideas and scientific discussions, for allowing me to continue working with weather@home ensembles, and hosting me during a short but memorable research stay at the Environmental Change Institute. Special thanks to Markus Reichstein for giving me the chance to join a vibrant research environment in the Department for Biogeochemical Integration, his approach to science that is admirable, scientific insights and interest in my work, and for the opportunity to join innovative scientific activities in the realm of extreme events and their societal impact; special thanks also to Martin Heimann for giving me the chance to pursue an unconventional research idea, for his interest in ensemble-based approaches and scientific ideas, for supervising and evaluating the thesis. Further I would like to thank Nicolas Gruber for co-examining the thesis.

Furthermore, I would like to thank Jakob Zscheischler, Matthias Forkel, Milan Flach, Fabian Gans, and Holger Lange for help, scientific discussions, a critical view, sharing ideas, and crucial feedback at various stages of the thesis. I acknowledge scientific comments and feedback, and discussions with Myles Allen, Nuno Carvalhais, Martin Jung, Mirco Migliavacca, René Orth, Jonas Peters, Anja Rammig, Kirsten Thonicke, and Sönke Zaehle.

With regard to specific chapters, warm thanks to Lisa Alexander, Myles Allen, Jannis von Buttlar, Fabian Gans, Markus Reichstein, Jens Schumacher, Sonia Seneviratne, and Hans von Storch for comments on Chapter 2 and Silvana Schott for graphical support; Markus Donat, Fabian Gans, and Holger Metzler for discussions on scientific content and the analytical approximation in Chapter 3. I appreciate productive discussions and the exchange of ideas with Richard Jones, Antje Weisheimer, Lukas Gudmundsson, Jose Manuel Gutierrez Llorente on model biases and bias correction in general and specifically on the resampling approach (presented in Chapter 5), and Nuno Carvalhais, Mirco Migliavacca, and Sönke Zaehle for discussions on ecosystem model simulations. Chapter 8 improved thanks to helpful comments from Nuno Carvalhais, Kazuhito Ichii, Martin Jung, Sonia Seneviratne, and Sebastian Wolf. The comments and scientific insights of numerous peer reviewers helped to improve Chapters 2 through 8. Ulrich Weber, Urs Beyerle and Jan Sedlacek are acknowledged for help with data preparation and maintenance. Birgitta Wiehl and Barbara Wittneben provided crucial administrative support.

I thank numerous friends and colleagues at the MPI for Biogeochemistry, especially in the Department for Biogeochemical Integration for making it a truly special and inspiring work environment. Also, thanks to members of the Land-Climate Dynamics group at ETH Zurich and at the ECI Global Climate Science Programme at the University of Oxford for their hospitality.

Finally, and most importantly, I thank my family and Julia for their support.

The thesis was supported by the German National Academic Foundation (Studienstiftung des Deutschen Volkes), the European Union's Horizon 2020 research and innovation programme under grant agreement No. 640176 (BACI, 'Detecting changes in essential ecosystem and biodiversity properties towards a Biosphere Atmosphere Change Index'), the European Space Agency (STSE project CAB-LAB), and the International Max Planck Research School for Global Biogeochemical Cycles (IMPRS-gBGC). I acknowledge the World Climate Research Programme (WCRP) and the International Centre for Theoretical Physics (ICTP), which have supported the research presented in Chapter 4 through the WCRP-ICTP Summer School on Extremes (2014). This thesis uses a variety of observation-based data sources and reanalyses as well as model output and all data providers and maintainers are gratefully acknowledged and referenced in the respective chapters[2].

[2] Some specific acknowledgements:
Observation-based and reanalysis datasets: In Chapter 2, the Twentieth Century Reanalysis Project (`http://www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2.html`) is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office. I acknowledge the WMO ETCCDI Linkage Project LP100200690 for HadEX2 and GHCNDEX datasets aquired through `www.climdex.org` (Chapter 3). The E-OBS dataset from the EU-FP6 project ENSEMBLES (`http://ensembles-eu.metoffice.com`) and the data providers in the ECA&D project (`http://www.ecad.eu`) are gratefully acknowledged (Chapters 4, 5, and 6). ERA-Interim data was provided courtesy ECMWF (`http://www.ecmwf.int/en/research/climate-reanalysis/era-interim`; Chapters 5, 6, 7 and 8). CRU data were obtained from the University of East Anglia Climate Research Unit (CRU), British Atmospheric Data Centre (`http://badc.nerc.ac.uk/data/cru`; Chapters 5 and 7). I also acknowledge the creators and providers of Berkeley Earth Observations (Rohde et al., 2013), CRUNCEP dataset (`http://dods.extra.cea.fr/data/p529viov/cruncep/readme.htm`), Global Precipitation Climatology Centre (GPCC) Full Data Reanalysis Version 6.0 at Deutscher Wetterdienst (Schneider et al., 2011, 2014), and the WATCH-Forcing-Data-ERA-Interim (WFDEI, a product of the EU-EMBRACE project, Weedon et al., 2014) (all used in Chapter 5). Furthermore, I acknowledge the Global Modeling and Assimilation Office and the Goddard Earth Sciences (GES) Data and Information Service Center (DISC) for the dissemination of MERRA, MERRA-LAND, and GLDAS-2 products (GL-NOAH-PF), the latter of which were acquired as part of the mission of NASA's Earth Science Division. The CFSR data are from the Research Data Archive which is maintained by the Computational and Information Systems Laboratory at the National Center for Atmospheric Research (NCAR). NCAR is sponsored by the National Science Foundation. The original data are available from the RDA (`http://dss.ucar.edu`) in data set number ds093.0. I further acknowledge the Japanese 25yr Re-Analysis and JMA Climate Data Assimilation System (JCDAS) for the dissemination of JRA-25 data (all used in Chapter 7). The JRC-TIP (Joint Research Centre Two-Stream Inversion Package) FAPAR dataset based on satellite broadband albedo observations (Pinty et al., 2011) is provided by The Inversion Lab and the Joint Research Centre of the European Commission, and I thank Michael Voßbeck, Bernhard Pinty, Nadine Gobron and Thomas Kaminksi for making the data available (Appendix A).
Models: Model ensemble simulations constitute the main research tool in this thesis, and I would like to thank the Oxford eResearch Centre for their technical expertise, the Met Office Hadley Centre PRECIS team for their technical and scientific support for the development and application of weather@home and all of the volunteers who have donated their computing time to climateprediction.net (used in Chapters 4, 5, 6, and 8). Moreover, I acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and thank the climate modelling groups for producing and making available their model output (Chapter 7). For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

1. Introduction

1.1. Motivation for the present thesis

‘"Die hohen Fluthen haben fast ein Menschenalter uns in Ruh’ gelassen; wenn aber eine von den schlimmen wiederkommt (...), so kann mit einem Mal die ganze Herrlichkeit zu Ende sein;" (...) *"Vor dreißig Jahren ist der alte Deich gebrochen; dann rückwärts vor fünfunddreißig, und wiederum vor fünfundvierzig Jahren; seitdem aber, (...) haben die höchsten Fluthen uns verschont. Der neue Deich aber soll trotz solcher hundert und aber hundert Jahre stehen;"’* Hauke Haien, The Dykemaster in Theodor Storm’s novella ‘The Rider on the White Horse’[1] from 1888, contemplates about the risk of severe storm surges and floods that are recurring regularly on time scales from decades to centuries, and that have the potential to impose disastrous impacts on coastal communities.

Extreme weather and climate events, such as storms, floods, cold spells, heatwaves or droughts have long affected human societies - with often adverse and sometimes catastrophic impacts, and as a source of great concern within communities and the society at large. For example, tree-ring based proxy records (Büntgen et al., 2011b; Cook et al., 2015; Luterbacher et al., 2016) and documentary evidence (Büntgen et al., 2011a; Wetter et al., 2014) reveal pronounced hydro-climatic variability and extremes over past centuries in Europe–with individual extreme events such as drought and heat in 1540 (labelled a ‘worst-case’ scenario, Wetter et al., 2014), or severe rain and cold periods 1315–1317 that are associated with food shortages and famine (e.g. the ‘Great European Famine’, 1315–1317, Lucas, 1930) that under some conditions could even overturn social

[1] Theodor Storm. 1888/2011. Der Schimmelreiter (‘The Rider on the White Horse’). ISBN 3458362169.

and political power relationships (Bauch, 2016).
But, conversely, extreme weather and climate events (summarised as ‘climate extremes’ in this introduction) also serve as a source of admiration for the strong forces of nature, inspiring writers, poets, and scientists throughout centuries. The mere possibility of occurrence of climate extremes can trigger coherent societal planning or even innovation (a simple and straightforward example is the Dykemaster’s invention to build flatter dykes based on mathematical considerations in the tale cited above), and -under some circumstances- might lead to societal cohesion and adaptation in the aftermath of climate extremes (Luterbacher and Pfister, 2015).

In more recent years, various types of climate extremes have occurred, on all continents, and on a range of spatial and temporal scales (AghaKouchak et al., 2012; Coumou and Rahmstorf, 2012). These events, modulated by vulnerability and exposure of any system under consideration, regularly impose substantial impacts on human societies and ecosystems (Easterling et al., 2000; IPCC, 2012). The impacts range from the loss of human lives (Le Tertre et al., 2006; Gasparrini et al., 2015), economic losses either through direct or indirect effects (Smith and Katz, 2013; Zander et al., 2015; Burke et al., 2015), effects on terrestrial ecosystems and ecological communities (Parmesan et al., 2000; Thibault and Brown, 2008; Jentsch et al., 2009) and the global carbon cycle (Reichstein et al., 2013; Zscheischler et al., 2014b). Moreover, the role of climate extremes in triggering or shaping human conflict continues to be investigated and debated (Scheffran et al., 2012; Schleussner et al., 2016a).

The occurrence of a few ‘high-impact’ climate extremes in recent years raised attention and awareness among the public, policy-makers and institutions to prepare for these events under changing climatic conditions (WMO, 2011; IPCC, 2012). For example, a heat summer in Central Europe that was unprecedented in centuries occurred in 2003 (Luterbacher et al., 2004) - a statistically very unlikely event with estimated return periods of at least several thousand years even if recent warming would be taken into account (Schär et al., 2004). The event had been interpreted as a first sign of increasing variability of summer temperatures (Schär et al., 2004) and was associated with strong land-atmosphere feedbacks that might be expected in future European climate (Seneviratne et al., 2006; Fischer et al., 2007). The death toll of this event was large, estimates range from 40.000-70.000 excess deaths due to heat (WMO, 2011), and socio-economic costs due to crop failure and forest fires were in the range of around 13 billion Euros (García-Herrera et al., 2010). Furthermore, drought and heat undid several years of net carbon sequestration in European ecosystems (Ciais et al., 2005), which thus constitutes a positive feedback in the climate system exacerbating climate change.

The role of climate extremes as a crucial feature of the Earth's climate, and the impacts of climate extremes on human societies and ecosystems (IPCC, 2012) provides the main motivation for the present thesis. In this Chapter, I first provide an introduction to the notion of climate extremes in the context of climate variability and change. This includes a short overview of observed changes in climate extremes that are most relevant for terrestrial ecosystems and the carbon cycle (temperature extremes, heavy precipitation, and drought). Uncertainties around the quantification of these phenomena are shortly discussed, and model ensemble simulations are introduced (Section 1.2). Second, I review key land-atmosphere and ecosystem-atmosphere interactions that arise through basic (bio-)physical and biogeochemical principles and mechanistic links between the energy, water, and carbon (Section 1.3). Third, I focus on the impacts of climate extreme events on the terrestrial carbon cycle (Section 1.4). Lastly, the structure of the thesis and key findings of each chapter are outlined (Section 1.5).

1.2. Climate extremes - background, definitions & examples

1.2.1. Climatic variation and extremes

The Earth's climate varies over a wide range of time scales - from seconds to millions of years (Gettelman and Rood, 2016). This continuum of variability can be understood as a response to deterministic insolation forcing with daily, annual and longer-term (e.g. 'Milankovich' cycles, see Hays et al. (1976)) periodicities, and a transfer of spectral energy across frequencies through non-linear dynamics in the Earth system associated with atmosphere-ocean, atmosphere-land (including atmosphere-ecosystem), and atmosphere-cryosphere interactions (Huybers and Curry, 2006). Hence, variation in the Earth system can be conceptualised as the super-position of characteristic periodicities, including trends, and stochastic components (Ghil et al., 2011). In this context, the occurrence of extreme values, or climate extremes, constitutes a crucial aspect and manifestation of climatic variability, across space and time, and against the backdrop of long-term climatic changes (e.g. Tingley and Huybers, 2013).

1.2.2. On the definition of climate extremes

Extremes are commonly understood as very large, and unusual deviations from a normal state of any system under consideration. Hence, a simple and straightforward definition is that climate extremes can be described as the '*occurrence of a value of a weather or climate variable above (or below) a threshold value near the upper (or lower) ends of the range of observed values of the variable*' (IPCC, 2012). Accordingly, extreme values typically constitute the (upper or lower) tail of a univariate probability distribution (Ghil et al., 2011).

However, it should be emphasised that no universally accepted definition of climate extremes exists (Stephenson et al., 2008). For instance, the Oxford English Dictionary defines *extreme* as 'the outermost', 'farthest from the centre', or 'very advanced in any direction; utmost; uttermost', but also as 'Going to great lengths', and 'opposed to moderate'[2]. While all of these terms might correspond to the intuitive understanding of extremes referred to above, it highlights that any quantitative definition is inherently relative and thus somewhat subjective, and depends on what is considered extreme by the observer or in any given context (Stephenson et al., 2008). Likewise, the definition in IPCC (2012) cited above does not account for unusual sequences of events, or unusual bi- or multivariate constellations of individual variables (so-called 'compound events', e.g. IPCC (2012); Leonard et al. (2014)), and is not *per se* relevant for impacts. Hence, an obvious alternative starting point to define or diagnose extremes would be to start from the distribution of impact variables in the system under consideration and assess climate variables that led to these extreme impacts in a 'backwards' manner (Smith, 2011; Zscheischler et al., 2013). However, the IPCC (2012) definition has been applied, implicitly or explicitly, in a large number of studies and proved useful for practical applications (see IPCC, 2012, and references therein),

[2]Oxford English Dictionary Online, `http://www.oed.com`.

because it allows a straightforward quantification of climate extremes (Sillmann et al., 2013b,a) and detailed mathematical description of the tails of the probability distributions (Coles et al., 2001). Therefore, it serves as a useful conceptual model to approach climate extremes, and is adopted as such in the present thesis. However, as the different chapters in this thesis have different objectives, the definition of climate extremes, and the choice of variables, is specified separately in each chapter.

1.2.3. Quantifying climate extremes in a changing climate

Conceptual framework In a changing climate, the frequency, intensity and spatio-temporal characteristics (e.g. affected area, duration, or time-area integral) of climate extremes are expected to change (Mearns et al., 1984; Meehl et al., 2000; Easterling et al., 2000). A simple conceptual framework is shown in Figure 1.1, following IPCC (2012) but originating earlier (presumably Meehl et al. (2000)) that illustrates how changes in the (a) mean, (b) variance, and (c) shape of a univariate probability distribution could result in changes in the tails. Although this schematic is highly simplified and not based on physical considerations, it can yield useful insights: First, even relatively small changes in the mean of a (climate) variable can lead to disproportionate changes in the number or frequency of climate extremes in the tails, including record-breaking events (Rahmstorf and Coumou, 2011) (whereas the intensity of the events would scale with the mean shift) – with potentially profound implications if any particular impact would be triggered by the exceedance of a fixed threshold (Mearns et al., 1984). Second, changes in variance can have large effects on the frequency and intensity of climate extremes. Based on theoretical work, it has been shown that for a comparable change in mean and variance of a univariate probability distribution, the frequency of these events is more sensitive to changes in variability than in averages (Katz and Brown, 1992). Third, it is also conceivable that non-linear mechanisms in the Earth system might lead to changes in the symmetry of a climate variable's probability distribution, i.e. changes only in one tail of the distribution. Such non-linear changes in only one tail of the distribution might be well-expected based on physical reasoning for some variables (e.g. precipitation, see Section 1.3), but the detection of this phenomenon still poses challenges (see e.g. Chapter 2).

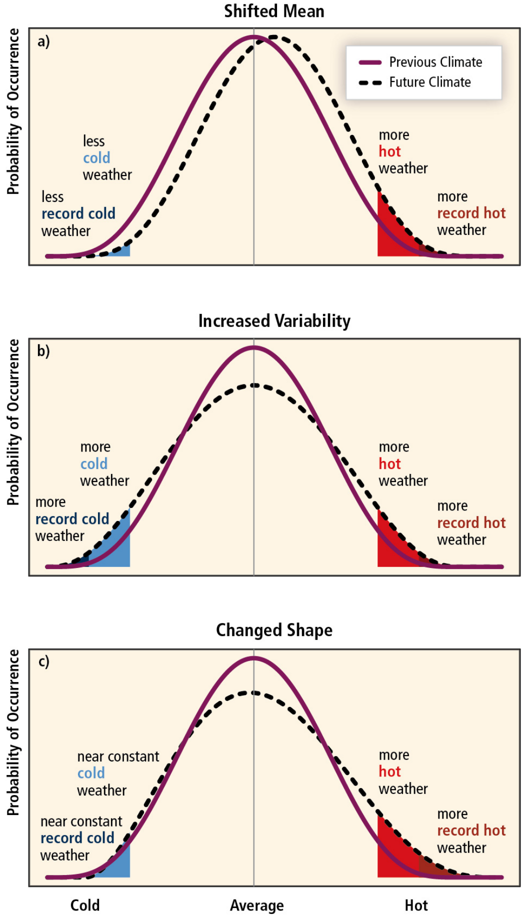


FIGURE 1.1.: **Hypothesised changes in the distribution of temperature**: a) simple mean shift, b) increased temperature variability (no shift in the mean), and c) effects of an altered shape of the distribution. From IPCC (2012).

Quantifying change in climate and weather extremes An obvious starting point to detect, quantify, and understand potential changes in the frequency or intensity of climate extremes is to analyse long-term meteorological observations.

Over past decades, many studies have shown that the occurrence probabilities and intensity of several types of climate extremes have been changing in observations. Here, I provide a few examples and brief overview of observed changes and first-order expectations of changes in temperature extremes, heavy precipitation extremes, and drought at large spatial scales. These hydro-meteorological extremes are highly relevant for terrestrial ecosystems and the carbon cycle (Reichstein et al., 2013; Zscheischler et al., 2014b, see also Section 1.4), and are thus considered as the main hydro-meteorological hazards investigated in this thesis. A more detailed examination and review of climate extremes is given elsewhere (e.g. IPCC, 2012).

Temperature extremes On a global scale, observations point towards a widespread increase in positive temperature extremes consistent with expectations in a generally warming climate. These trends have been shown in maximum and minimum daily temperatures (Alexander et al., 2006; Donat et al., 2013b), in the area affected by temperature extremes (Hansen et al., 2012; Dittus et al., 2015), the duration of heat waves (Perkins et al., 2012), and an increase in record-breaking monthly temperatures (Coumou and Robinson, 2013). Accordingly, reductions in the occurrence of cold conditions are observed globally (Alexander et al., 2006), although this does not mean that cold events are not occurring any more (e.g. Cattiaux et al., 2010; Kodra et al., 2011). These trends also hold qualitatively on continental scales, where individual studies have shown that increases in observed temperature extremes have been widespread, for instance in the Mediterranean (Kuglitsch et al., 2010), North America (Peterson et al., 2008; Grotjahn et al., 2016), Australia (Alexander and Arblaster, 2009), China (Zhou and Ren, 2011), and many other regions (see, for instance, IPCC, 2012). In summary, IPCC (2012) conclude that it is *very likely*[3] that an increase of warm days and nights, and a decrease of their cold counterparts, has occurred at the global scale. Furthermore, projections point towards continued increases in temperature extremes (Kharin and Zwiers, 2000; Kharin et al., 2007; Orlowsky and Seneviratne, 2012; Sillmann et al., 2013a) and scale approximately linearly with global temperatures, but with larger slopes (Seneviratne et al., 2016). How-

[3]In IPCC-terminology this denotes a confidence of 90–100% (Mastrandrea et al., 2010)

ever, it is important to recognise that temperature extreme events are not merely random events, but are typically associated with distinct atmospheric conditions such as quasi-stationary anticyclonic circulation anomalies or atmospheric blocking in mid-latitudes (Xoplaki et al., 2003; Meehl and Tebaldi, 2004). Heat waves can be further enhanced through interactions with the land surface (Seneviratne et al. (2006), see more detailed description and basic physical principles of land-atmosphere interactions introduced in Section 1.3).

Precipitation extremes In a warmer world, substantial changes to the hydrological cycle are expected (Stocker et al., 2013). Radiation-induced energy budget changes in the troposphere and near the surface lead to an acceleration of the hydrological cycle, and hence increased evapotranspiration from the ground that enhances atmospheric moisture content (Trenberth, 1999; Allen and Ingram, 2002). The water-holding capacity of the atmosphere increases by approximately 7% per °C of warming at the surface following the Clausius-Clapeyron equation. These considerations have led to the expectation that changes in heavy rainfall extremes are physically constrained by that rate (Trenberth, 2011), albeit only in regions without major changes in atmospheric circulation (Pall et al., 2007). However, uncertainty remains as scaling rates of extreme precipitation and atmospheric water vapour with temperature are not identical, and the former affected by several factors (O'Gorman and Schneider, 2009) with large model spread in heavy precipitation scaling rates in the tropics (ibid.), and some observational records and model simulations that even indicate higher scaling rates for extremes at sub-daily time scales (Lenderink and Van Meijgaard, 2008; Bao et al., 2017). Hence, in combination, a certain redistribution of precipitation, i.e. enhanced heavy precipitation events, and somewhat smaller increases in mean precipitation (Allen and Ingram, 2002) in tandem with potentially longer dry spells (but which is still uncertain, Fischer and Knutti, 2013), might be expected in the near future and at the global scale (Trenberth, 2011). Indeed, increases in heavy precipitation extremes are projected in many land regions of the globe (Fischer and Knutti, 2015), although local and regional variation is prevalent over land (see for instance a regional simulation over the United States, Prein et al., 2016). In observational time series, widespread significant trends in precipitation extremes and totals are found in observations at the global scale, with a clear majority of stations showing upward trends (Westra et al., 2013), and in an averaged signal across wet and –to a lesser extent– dry regions (Donat et al. (2016), see also Chapter 3). At the global scale the number of record-breaking events has increased faster than expected in a stationary climate, and consistent with Clausius-Clapeyron expectations (Lehmann et al., 2015). However, on regional scales, changes are often relatively noisy and spatially somewhat heterogeneous (Alexander et al., 2006; Donat et al., 2013b). A markedly high sensitivity of precipitation extremes to temperature increases has been found in the tropics and at high latitudes (Westra et al., 2013)IPCC (2012) conclude that it is *likely*[4] that heavy precipitation extremes have increased both in frequency and as a proportion of heavy rainfall over many areas of the globe.

Drought Drought constitutes a major hazard in many regions of the world, because it can severely affect agricultural systems and thus food production but also terrestrial biota and ecosystems. Therefore, the question to which extent drought is changing in a warming climate is crucial. However, drought is a complex phenomenon that can be defined in various ways based on a deficit in precipitation, soil moisture, or runoff, that is commonly referred to as meteorological drought, agricultural drought (or 'soil-moisture drought'), or hydrological drought, respectively (IPCC, 2012, leaving aside water scarcity here, which is at least partly a socio-economic phenomenon).

As soil-moisture droughts and hydrological droughts ultimately arise from a lack of water on land, these events depend on multiple variables that determine the water balance at the land surface. In particular, a critical precipitation deficit (or meteorological drought) is typically a necessary but not sufficient condition for the development of soil-moisture or hydrological drought, because evapotranspiration (the flux of water from the land surface to the atmosphere, see Section 1.3 for a more detailed description), and initial water storage conditions equally shape drought events. Evapotranspiration is jointly controlled by 1) the supply of water from the land surface, and 2) the evaporative demand of the atmosphere (also called 'potential evapotranspiration', PET), which is related to the ability of the

[4] 66–100% probability (Mastrandrea et al., 2010).

atmosphere to evaporate, absorb, and transport water. Hence, PET depends on a variety of variables, including radiation, wind speed, and vapour-pressure deficit (VPD); whereas temperature affects PET indirectly through its effect on VPD (Seneviratne et al., 2012).

Besides the complexities in the meteorological drivers from which drought conditions arise, observational estimates of evapotranspiration and precipitation as the main drivers of drought are often notoriously uncertain (Seneviratne et al., 2010a; Wang and Dickinson, 2012; Trenberth et al., 2014) and observations are (sometimes severely) limited in space and time.

As a result, assessments of historical trends in drought at the regional or global scale are often uncertain (Seneviratne et al., 2012; Trenberth et al., 2014). For example, different formulations of drought indicators show widely differing sensitivities to temperature change (Milly and Dunne, 2011; Sheffield et al., 2012), depending on whether the computation of PET is based on simplified and typically empirical approaches that depend to a large extent on temperature (e.g. Thornthwaite, 1948), or whether it is derived from physics-based formulations such as the Penman-Monteith equation (see e.g. Wang and Dickinson, 2012, for an introduction). Hence, the Intergovernmental Panel on Climate Change concluded with *medium confidence* that ‘some regions of the world have experienced more intense and longer drought [...], but opposite trends also exist’ (Seneviratne et al., 2012), thus reflecting large uncertainties in historical drought assessments.

Moreover, developing theory-based expectations, or evaluating model projections of future changes in drought occurrence, intensity or duration is even more complex. This is because, in addition to the uncertainties outlined above, several feedback mechanisms and interactions between meteorological variables are likely to shape future drought events (Seneviratne, 2012, see more details on land-atmosphere interactions in Section 1.3), notwithstanding indirect effects through variations or possible changes in large-scale oceanic or atmospheric circulation regimes that affect droughts (Trenberth et al., 2014).

One might anticipate that in a warmer world, increased heating at the land surface might enhance evapotranspiration and thus could exacerbate droughts (Dai, 2011; Trenberth et al., 2014). However, under dry conditions land-atmosphere interactions can enhance temperatures close to the surface (see Section 1.3), and thus high temperatures might be (partly) a consequence, rather than the cause, of drought (Sheffield et al., 2012). Moreover, several negative feedbacks between a warmer climate and droughts might also be expected: First, increased evapotranspiration could lead to a corresponding increase in relative humidity (and thus decrease in VPD), thus partially counteracting drought (Seneviratne, 2012). Second, evapotranspiration from a dry land surface becomes limited by soil moisture, which then naturally prevents further drying (Seneviratne et al., 2012). Third, it is expected that under higher CO_2 concentrations, terrestrial plants' water use efficiency increases (Drake et al., 1997), leading to reduced evapotranspiration via reduced stomatal conductance, thus reducing the occurrence of soil moisture droughts in process-oriented models (Burke, 2011).

In summary, drought trends at the global scale, and future projections of drought in a warmer climate remain uncertain, depending crucially on the drought indicators used, and the considered processes and underlying assumptions. For example, Burke and Brown (2008) showed that drought indicators that incorporate the atmospheric demand for moisture show an increase of 5%–45% of the land surface in drought, which is significant albeit its large spread (whereas indicators based on precipitation alone show little change). Nonetheless, the dominant source of uncertainty in projections of soil-moisture drought across an ensemble of state-of-the-art models relates to the formulation of the underlying model (Orlowsky and Seneviratne, 2013). This is in contrast to projections of (for instance) temperature extremes (as outlined above), which are primarily related to the specified scenario of greenhouse gas forcing, and thus highlights uncertainties in model projections of future drought. Nonetheless, several regions have indeed experienced trends towards a higher frequency of drought conditions in recent years (*medium confidence* according to IPCC, 2012), such as for example South-Central Europe (Stahl et al., 2010).

Uncertainties in the quantification of climate extremes at large spatio-temporal scales Analyses of extreme events in observational data are challenging, because observational records are typically limited in length, often contain heterogeneities, and extremes are rare by definition (Nicholls, 1996), all of which makes analyses sensitive to outliers. Therefore, on-going monitoring, data availability, data exchange, and data quality control are all crucial prerequisites for reliable analyses (Nicholls, 1996; Alexander et al., 2006; Donat et al., 2013b; Alexander, 2016). The detection and quantification of changes in climate extremes depends on the variable, definition of thresholds or metrics, and spatial and temporal scales of analysis. Moreover, the detection and quantification of climate extremes are contingent on the statistical methods applied (AghaKouchak et al., 2012). Since assessments of observed trends in climate extremes under climate change routinely enter, inform and influence public discourse[5], an accurate quantification of these phenomena in tandem with explicit definitions and metrics is vital to enable informed discussions about climate change impacts. Therefore, a key issue in this context is to ascertain statistical accuracy and robustness of the metrics that are used to quantify climate extremes. Part I of the present thesis revisits state-of-the-art methodologies that are widely used to detect globally or regionally aggregated signals of climate extremes (see conceptual Figure 1.2a) in spatio-temporal datasets of temperature (Chapter 2) and precipitation (Chapter 3). However, these methodological results are generic in that they apply to any other spatio-temporal data as well.

1.2.4. Attribution of climate extremes and their impacts

A question that arises frequently in the public discussion in the aftermath of specific climate extremes is whether anthropogenic climate change could have played a role, or could even be blamed, for the occurrence of a particular climate extreme event. However, there is no straightforward answer to this question (Allen, 2003), because any climate extreme event could occur in an unperturbed climate as well and the observational record is inevitably limited. Nonetheless, insights into frequency and magnitude of specific classes of climate extremes can be derived from a probabilistic perspective - hence addressing the question how the odds of occurrence of these events have changed in response to specific forcings (Allen, 2003; Stone and Allen, 2005; Stott et al., 2016). Because observed weather constitutes only one of many possible trajectories of the system, a probabilistic approach requires a collection of possible trajectories from repeated model simulations

[5]e.g. `http://www.nytimes.com/2012/08/07/science/earth/extreme-heat-is-covering-more-of-the-earth-a-study-says.html`

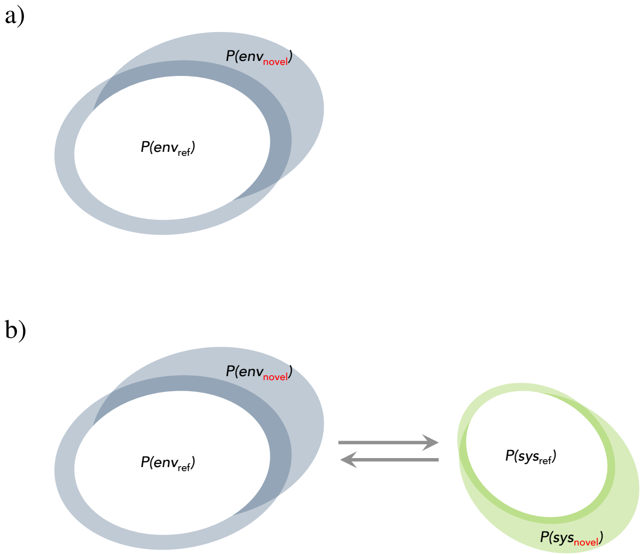


FIGURE 1.2.: **Conceptualised quantification of changes in climate extremes and associated impacts.** a) State-of-the-art methodologies to quantify changes in climate extremes (depicted as changes in the probability distribution $P(env_{ref})$ to $P(env_{novel})$) are revisited in Part I of this thesis. b) Conceptualised relationship between a multivariate distribution of climate variables and the distribution of an impact variable. Part II of this thesis deals with suitable constraints on $P(env)$ to improve the simulation of climate extremes, and Part III assesses the impacts of climate extremes in the terrestrial biosphere ($P(sys)$) using ensemble simulations of a climate and ecosystem model. Please note that climate impacts are also modulated by vulnerability and exposure (IPCC, 2012), which is omitted in this figure for simplicity. Figure courtesy M. Mahecha.

(Gneiting and Raftery, 2005), typically called ensemble simulations, to be able to address rare events in the tails of the distribution.

Probabilistic attribution of climate extremes therefore usually aims to get a notion of how the odds of a specific type of climate extreme might have changed due to a change in external forcing (Stone and Allen, 2005; Stott et al., 2016). For example, the Russian heat wave in 2010 was a meteorologically highly unusual event that broke long-term records (Barriopedro et al., 2011; Tingley and Huybers, 2013) causing over 55.000 casualties (Guha-Sapir et al., 2011) and severe reductions in vegetation productivity (Bastos et al., 2014). The event was associated with a long-lived blocking situation that arises due to natural climate variability (Dole et al., 2011). Conversely, however, Rahmstorf and Coumou (2011) showed that the frequency of events similar in magnitude to the Russian heat wave had increased by a factor of roughly five - implying that with an 80% chance this disastrous event would not have occurred without climate warming. Using large ensembles of climate model simulations, Otto et al. (2012) showed that a probabilistic attribution of the odds of events does not contradict that the event mainly originated from natural climate variability. Furthermore, ensemble-based attribution studies investigated and attributed extremes in more impact-related variables such as floods (Pall et al., 2011; Schaller et al., 2016) and heat-health related metrics (Mitchell et al., 2016a), but detailed understanding or tools for relating attribution research to impacts in various fields is still widely lacking (Hansen and Stone, 2016; Otto, 2016), and event attribution results often depend on the specific framing of the attribution question and the event definition (Stott et al., 2016). Moreover, as models are inevitably imperfect, several uncertainties remain in attempts to attribute climate extremes that are due to model biases and poor reliability (Massey et al., 2015; Bellprat and Doblas-Reyes, 2016), and the imperfect representation of long-term trends (Min et al., 2013), amongst others.

Nonetheless, model ensemble simulations constitute a powerful tool to characterise the probability distribution of possible weather states in response to various climate forcings - and hence to assess weather-related risks in forecast (e.g. Gneiting and Katzfuss, 2014) and hindcast (Massey et al., 2015). This includes assessments of return times of climate extremes, potential interactions of driving variables, and associated (simulated) impacts. Therefore, climate model ensemble simulations are used extensively in this thesis: This includes approaches to scrutinise the probability distribution of climate variables (and changes therein, but not focusing on impacts explicitly, depicted conceptually in Figure 1.2a), by using model ensembles to benchmark statistical inferences about rare climate extremes (Part I, Chapter 4). Further, model ensembles are used to develop and apply bias correction methods based on observation-based constraints (Part II); and finally to assess climate-impact relationships in model ensembles explicitly

(Part III, Chapter 8), which includes scrutinising extremes and recent changes in ecosystem impact variables ‘backwards’ (conceptually illustrated in Figure 1.2b).

1.3. Energy, water, and carbon: Processes and land-atmosphere interactions

The Earth’s atmosphere and the land surface are intimately linked. This coupling includes both the direct atmospheric influence on the land surface, including ecological and hydrological systems, but also associated feedbacks induced by land surface processes that in turn shape the atmosphere and climate (Bonan, 2015). A conceptual scheme that illustrates the mutual links between the atmosphere and the land surface, and associated processes is shown in Fig. 1.3. Land-atmosphere interactions occur primarily via biogeophysical (i.e., physical transfer of energy and moisture) and biogeochemical (i.e., cycling of elements) processes, and on almost instantaneous time scales up to centuries (Bonan, 2015) and perhaps even on evolutionary time scales (Lovelock and Watson, 1982). For example, slow interactions would include changes in structure or species composition changes in ecosystems that would affect the energy balance via albedo feedbacks (Ganopolski et al., 1998), or carbon cycle feedbacks to climate change in the 21st century (Friedlingstein et al., 2014). Conversely, fast interactions occur for instance as a direct response of plants’ stomatal conductance to variation in weather variables such as light, temperature or moisture availability. In this section, I illustrate the main mechanisms of land-atmosphere interactions that link the energy and water balances at the land surface, and the terrestrial carbon cycle.

1.3.1. The surface energy balance

The Earth’s climate and weather is driven primarily by solar insolation forcing. The solar constant, i.e. the energy transferred by photon flux (or the integral of the sun’s electromagnetic spectrum) at the top of the atmosphere amounts to 1361 W m^{-2} (scientific measurements of this fundamental quantity date back at least to the early 19th century, Abbot, 1914), which equals 341 W m^{-2} in incoming shortwave radiation averaged over diurnal and seasonal variation and globally at the top of the atmosphere (Trenberth et al., 2009). The net radiation (R_{net}) at

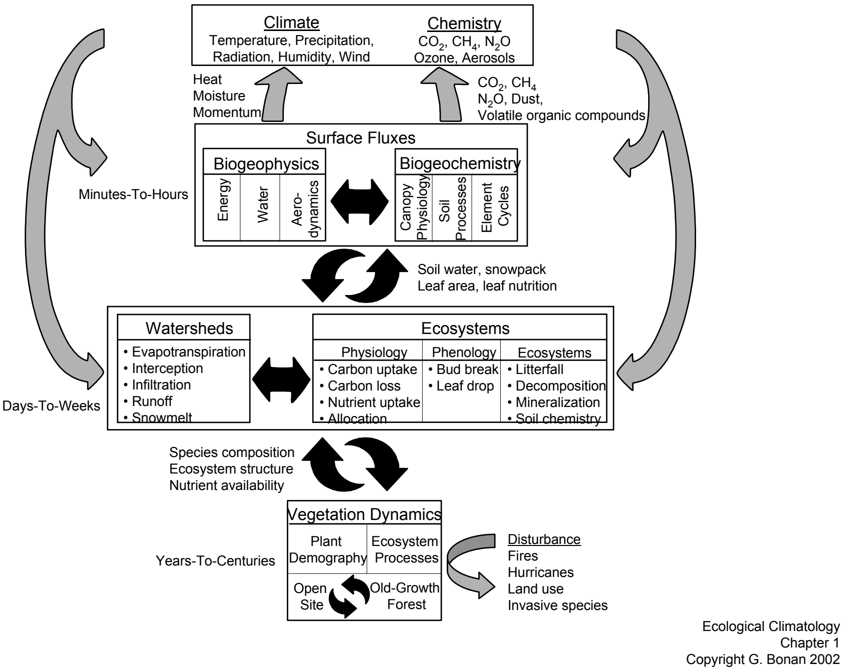


FIGURE 1.3.: Conceptual depiction of ecosystem-atmosphere interactions. Terrestrial ecosystems affect weather, climate, and atmospheric composition through biogeophysical and biogeochemical processes (detailed in Section 1.3), mediated by watershed and ecosystem dynamics on longer time scales. From Bonan (2015).

the land surface is given by (see e.g. Bonan, 2015)

$$R_{net} = (S\downarrow - S\uparrow) + (L\downarrow - L\uparrow). \tag{1.1}$$

The incoming shortwave radiation[6] ($S\downarrow = 161.2$ W m^{-2}, $S_{land}\downarrow = 145.1$ W m^{-2}) and longwave radiation ($L\downarrow = 333$ W m^{-2}, $L_{land}\downarrow = 303.6$ W m^{-2}) components are balanced by their outgoing counterparts ($S\uparrow = 23.1$ W m^{-2}, $S_{land}\uparrow = 39.6$ W m^{-2} and $L\uparrow = 396$ W m^{-2}, $L_{land}\uparrow = 383.2$ W m^{-2}). At the land surface, net radiation is partitioned into the turbulent fluxes of sensible

[6]all values given as long-term averages for the globe (no subscript) and global average over land (subscript $_{land}$) according to Trenberth et al. (2009)

($H = 17$ W m^{-2}, $H_{land} = 27$ W m^{-2}) and latent heat ($\lambda ET = 80.0$ W m^{-2}, $\lambda ET_{land} = 38.5$ W m^{-2}, λ denotes the energy required for vaporisation of water, and ET is evapotranspiration), and the ground heat flux ($|G| < 1$ W m^{-2}, $|G_{land}| < 1$ W m^{-2}), i.e.

$$R_{net} = H + \lambda ET + G. \tag{1.2}$$

All terms of the energy balance vary in space and time (Fasullo and Trenberth, 2008), and differ between land and ocean (Trenberth et al., 2009). The energy used by plants to drive photosynthesis does not show up directly in long-term averages of the surface energy balance, because chemical energy produced by photosynthesis (driven by $S\downarrow$) and stored in organic compounds is later released by decomposition that leads to surface heating, which would yield a long-term net zero energy balance under an equilibrium assumption. Earth's vegetation converts on average only about 0.27% of the incident photosynthetically active radiation (which is about half of $S\downarrow$) into biomass (Hall and Rao, 1999). Nonetheless, the relationship between the absorption of photosynthetically active radiation by vegetation, based on incident shortwave radiation, and biomass produced by the plants' photosynthetic machinery forms the basis for an important class of photosynthesis models, namely light-use efficiency models (Field et al., 1998).

Land-atmosphere interactions emerge directly from the energy balance: For instance, the albedo of the land surface ($\alpha = \frac{S\uparrow}{S\downarrow}$) directly links the structure of the land surface (i.e. plant form and structure, ultimately: 'life') to surface climate through its control on the surface energy balance. These insights led to the notion that climate and living organisms could evolve in tandem, albeit under simplified assumptions (Lovelock and Watson, 1982), but also that active albedo management can be used to mitigate hot temperature extremes (Davin et al., 2014).

1.3.2. The surface water balance

The water balance at the land surface is given by (see e.g. Bonan (2015), but these fundamental ideas go back at least to Thornthwaite and Mather (1955))

$$\frac{dS}{dt} = P - ET - R, \tag{1.3}$$

where changes in water storage ($\frac{dS}{dt}$) are balanced by precipitation inputs (P) and losses through evapotranspiration (ET) and runoff (R), with some underlying soil water storage S. This long-term water balance subsumes surface runoff and drainage commonly as runoff, and neglects fluxes such as capillary water rise, plant-induced hydraulic redistribution of water (Horton and Hart, 1998), lateral transport, and anthropogenic irrigation.

The surface water and energy balances are linked through evapotranspiration (ET) from the land surface. ET is the sum of (1) physical evaporation from the soil or water surfaces (ET_{evap}), (2) interception of water from leaf surfaces (ET_{interc}), and (3) transpiration from plants (ET_{transp}), i.e. water loss through plant stomata. More than half of the global evapotranspiration flux is made up by transpiration ($\frac{ET_{transp}}{ET} \approx 60\% \pm 15\%$), with typically higher values in moist forest ecosystems, and almost half of global land precipitation is lost through transpiration ($\frac{ET_{transp}}{P} \approx 45\%$, according to Schlesinger and Jasechko (2014)). Hence, the importance of transpiration in the global water cycle highlights the role of terrestrial ecosystems, and in particular plants' stomata, in mediating water fluxes, with direct links to carbon assimilation and the carbon cycle.

A direct insight that follows from a joint consideration of the land energy and water balances is that the ET flux is limited either by the available energy, or the availability of water on the ground (Budyko, 1974). Land-atmosphere interactions such as soil moisture-temperature and soil moisture-precipitation coupling can arise through these controls and feedbacks with the energy and water balance typically on daily to seasonal time scales (see e.g. Seneviratne et al. (2010a) for a detailed overview). These interactions arise because the evaporative fraction (defined as the ratio between latent heat and the sum of the turbulent energy fluxes, i.e. $EF = \frac{\lambda ET}{\lambda ET + H} = \frac{\lambda ET}{R_{net} - G}$) is highly variable in space and time (Pitman, 2003) and mediated by soil moisture availability. This variation in water and energy controls on evapotranspiration has important implications for climate variability and extremes: For example, a positive radiation anomaly under wet land surface conditions would enhance evapotranspiration, which is insensitive to soil moisture in this regime, thus causing no direct feedbacks with the atmosphere, or even a weak dampening effect due to latent cooling (Seneviratne et al., 2010a). Conversely, under soil moisture limitation, i.e. under dry-transitional land surface conditions, the climate anomaly would be amplified due to reduced evapotranspiration and enhanced sensible heating, which warms the boundary layer and thus leads to even higher temperatures (ibid.). This mechanism has been shown in observations and models to be at work in many high-impact heat waves in recent years (Fischer et al., 2007; Hirschi et al., 2011; Whan et al., 2015), and the expansion of dry-transitional climate regimes might even enhance climate variability and climate extremes in a future climate (Seneviratne et al., 2006; Fischer and Schär, 2009). However, climate models disagree considerably on the representation of land-atmosphere coupling (Chapter 7), and hence these processes constitute a key weakness in present-day Earth system models leading to considerable biases in atmospheric variables (Pitman, 2003) but also in simulated carbon cycle impacts (Chapter 5). Nonetheless, the recent availability of observations-based benchmarking datasets of evapotranspiration allows to constrain model ensembles to a certain degree by observational data (see e.g. Fischer et al., 2012; Stegehuis et al., 2013). In Part II of this thesis, I explore whether climate model ensemble simulations can be constrained by observation-based metrics to yield an improved simulation of climate extremes and ecosystem impacts. These approaches include the development (Chapter 5) and application (Chapter 6) of a constraint-based bias correction methodology that uses the distribution of summer temperature as a constraint; and a land-atmosphere coupling metric is used to improve the representation of temperature extremes in a multi-model ensemble (Chapter 7).

1.3.3. The terrestrial carbon cycle

Carbon dioxide (CO_2) constitutes an important greenhouse gas in the atmosphere that exerts major control on global climate (Lacis et al., 2010). In its present elevated concentration it contributes an additional radiative forcing of $+1.82$ W m^{-2} relative to 1750 (Myhre et al., 2013, the total additional radiative forcing due to anthropogenic activities is estimated at $+2.83$ W m^{-2}). Therefore, an understanding of the global carbon cycle, including fluxes between different reservoirs, is imperative.

The carbon cycle describes the cycling of carbon in organic and inorganic form between the ocean, the atmosphere, the land, and the lithosphere. Carbon release from fossil fuel combustion, cement production and land use changes have led to an increase in the atmospheric CO_2 concentration from around 280 ppm in pre-industrial times to above 400 ppm in 2016 (Betts et al., 2016), and an associated major perturbation and redistribution of carbon between its reserves (Ciais et al., 2014). Here, I focus on the terrestrial component of the carbon cycle, i.e. atmosphere-ecosystem exchanges of carbon. A detailed introduction to the global carbon cycle, including natural and perturbed reservoir sizes and fluxes is given elsewhere (e.g. Ciais et al., 2014).

The ecophysiological basis of ecosystem-atmosphere carbon exchange

At site level, net ecosystem carbon uptake (NEP) is given by the difference between the ecosystem's gross primary productivity (GPP) and respiration (R_{eco}) losses (Chapin III et al., 2006; Schulze, 2006) that occur either through plant autotrophic respiration (R_A) or decomposition in soils (heterotrophic respiration, R_H),

$$NEP = GPP - R_{eco} = GPP - R_A - R_H = NPP - R_H. \qquad (1.4)$$

NPP is the net uptake of carbon after accounting for autotrophic respiration losses. Often, the ecosystem carbon balance is simplified by using NEP interchangeably with net ecosystem exchange (NEE) but with opposite sign (i.e. $NEE = -NEP$). However, similarly to the water balance discussed above, this approach ignores several fluxes that would have to be considered at larger spatial scales (Körner, 2003), for instance lateral or vertical fluxes of dissolved organic carbon, carbon emissions due to land use change or fire, and transport of carbon by humans (timber or crop harvest, etc.) or animals.

On the ecosystem level, GPP subsumes plant photosynthetic uptake of CO_2, that is the chemical reaction by which CO_2 and water are converted into plant organic compounds using photochemical energy absorption in the chloroplasts (Hall and Rao, 1999) that are small organelles in leaves of green plants. The release of organic carbon into the atmosphere via ecosystem respiration, that is the flux inverse to carbon uptake, consists of plant or microbial metabolic processes (R_A and R_H, respectively) and depends on temperature (Lloyd and Taylor, 1994; Mahecha et al., 2010b), but also on the availability of water (Lloyd and Taylor, 1994) and carbon supply (Högberg et al., 2001; Ilie et al., 2016).

Plants regulate diffusion of CO_2 into the leaf's interior through opening their stomata. Stomatal conductance regulation can be thought of as a process by which plants maximise CO_2 assimilation (which diffuses from the atmosphere into inner-cellular air spaces), and minimise transpiration water losses (Medlyn et al., 2011). This key concept constitutes a fundamental physiological link between the carbon and water cycle (Fatichi et al., 2015). These basic principles underlie ecosystem models that simulate water-carbon processes at the land surface (see Sellers et al. (1997); Bonan (2015) for a general introduction; and Sitch et al. (2003) for the LPJ model that is used in Chapter 5 and 8). A variety of ecosystem-atmosphere interactions and longer-term carbon-climate feedbacks arise from this connection: For example, under increased CO_2 concentrations it is expected that stomatal conductance reduces, with an associated reduction in transpiration, causing higher runoff rates (Betts et al., 2007) but also higher temperatures (Cao et al., 2010), including under heat waves (Kala et al., 2016), and these plant physiological adjustments imply reduced drought stress projections (Swann et al., 2016).

In the context of ecosystem respiration, the explicit temperature dependence of photosynthesis kinetics and soil microbial activity, and thus soil respiration, raises concerns of pronounced climate-carbon cycle feedbacks in a warmer climate (Luo, 2007; Heimann and Reichstein, 2008; Bond-Lamberty and Thomson, 2010; Crowther et al., 2016). Nonetheless, ecosystem-atmosphere feedback chains are complex (Heimann and Reichstein, 2008), particularly on multi-decadal time scales that typically exceed time horizons of ecosystem experiments, and involve interactions with other biogeochemical cycles such as nitrogen and phosphorous (Falkowski et al., 2000; Luo, 2007; Gruber and Galloway, 2008; Zaehle et al., 2010).

Measuring and modelling ecosystem-atmosphere exchange of carbon
However, despite ecological theory and evidence for ecosystem-atmosphere interactions from ecological experiments and long-term monitoring, the quantification of carbon fluxes, and hence ecosystem-atmosphere interactions on large spatial scales remains a difficult task. This is because direct carbon flux measurements are only available at site scale. The most popular measurement approaches are forest or ecosystem inventories (Pan et al., 2011) and the Eddy covariance (EC) technique (e.g. Foken, 2008b). While forest inventories have typically a relatively poor temporal resolution, the eddy covariance technique measures turbulent exchange of water and carbon as an integrated signal between an ecosystem footprint and the atmosphere's boundary layer in high temporal resolution (typically aggregated to half-hourly fluxes). Therefore, the EC technique has been used extensively over the past two decades to investigate carbon dynamics, seasonality, and inter-annual variability, extremes, and relationships to other ecological variables in a large variety of ecosystems (e.g. Wofsy et al., 1993; Baldocchi, 2008). However, unresolved methodological issues such as the energy balance closure problem (Foken, 2008a) remain, and the partitioning of the measured net flux (NEE) into GPP and respiration R_{eco} is challenging (Reichstein et al., 2005). Regional and global networks of EC towers are now in routine use to describe carbon cycle dynamics (Baldocchi et al., 2001) and the joint exploitation of multiple site measurements often improves confidence in results, e.g. for the assessment of drought impacts on ecosystem carbon uptake (Wolf et al., 2016).

Nonetheless, despite the availability of point-based measurements, top-down constraints from atmospheric measurements (Graven et al., 2013), satellite and air-borne observations that can act as proxies for carbon cycle dynamics, terrestrial carbon cycle dynamics remain uncertain (Le Quéré et al., 2009). Because of these uncertainties in observations of the land component of the carbon cycle, residuals in the global carbon budget are often attributed to land processes (ibid.).

Empirical and process-oriented biogeochemical models constitute indispensable and complementary tools to scrutinise and conceptualise carbon cycle dynamics. On the one hand, process-oriented models encapsulate ecological understanding about relevant ecological processes (Sellers et al., 1997). These tools are widely used to project carbon cycle processes and dynamics in the 21st century and to study potential climate-carbon cycle feedbacks (e.g. Cox et al., 2000; Bonan, 2015) - although uncertainty on the relevant processes, parameterisation schemes, dynamics and feedbacks makes their projections uncertain (Heimann and Reichstein, 2008), and leads to often considerable deviations from observations (Mahecha et al., 2010a). On the other hand, empirical models typically employ a statistical relationship or machine learning algorithm to extrapolate (or: 'upscale') an ensemble of point measurements to global fields of carbon or water fluxes, and often also employ satellite measurements for extrapolation (Jung et al., 2011; Tramontana et al., 2016). Alternatively, considerations around ecosystem water use efficiency, i.e. the link between carbon uptake and water lost via transpiration at the ecosystem level, can be used to estimate carbon fluxes (Beer et al., 2007, 2009). Empirical models have proven useful for instance to explain interannual variability in the terrestrial carbon cycle (Jung et al., 2017). Also, empirical estimates of carbon fluxes are widely used to benchmark process-oriented models (Luo et al., 2012; Anav et al., 2013; Sippel et al., 2016b).

Uncertainties in the terrestrial carbon sink Based on combined evidence from atmosphere and ocean observations it can be inferred that the biosphere acts as a sink of carbon at present (Le Quéré et al., 2009). Terrestrial vegetation absorbs about 30% of present-day anthropogenic carbon emissions, while another 27% are taken up by the oceans and 43% remain in the atmosphere (Le Quéré et al., 2009). Hence, carbon sequestration by terrestrial ecosystems provides an important ecosystem service in mitigating the increase of CO_2 in the atmosphere. Although ultimately terrestrial reservoirs and the ocean can only slow down the increase in atmospheric CO_2 (Field et al., 1998), it remains a crucial question whether the presently observed sink is only a temporary slowdown - potentially induced by faster tree growth but eventually halted due to unchanged stocks (Körner, 2017), or whether structural changes in ecosystems can absorb and store substantial amounts of carbon for longer time periods, as might be indicated by an observed increasing trend in the ampltiude of the seasonal cycle of

CO_2 in high latitudes (Graven et al., 2013). Climate-carbon cycle simulations reflect these key uncertainties, as process model projections do not agree in whether the biosphere will act as a carbon sink or source under climate change in the 21st century (Friedlingstein et al., 2014). In summary, the fate of the terrestrial biosphere in the 21st century is a key uncertainty in global climate projections (Heimann and Reichstein, 2008).

1.4. Climate extremes and their impact on the terrestrial carbon cycle

A striking feature that emerges from global-scale carbon cycle observations is that the biosphere's ability to absorb carbon from the atmosphere is highly variable between years - including years in which the 'land carbon sink' turns into a net source of carbon (Le Quéré et al., 2009). Variability in land-atmosphere carbon exchange is inextricably linked to variability in weather and climate factors (Jung et al., 2017), and climate extremes contribute significantly to inter-annual variability in ecosystem carbon uptake (Zscheischler et al., 2014b). Hence, climate extremes are indeed key features that affect terrestrial ecosystem dynamics via various ecophysiological pathways, and can affect ecosystem structure and function (Smith, 2011; Reyer et al., 2013; Frank et al., 2015). Evidence for the effects of climate extremes on terrestrial ecosystems is available from multiple ecological archives and exploration tools such as tree rings analyses (Babst et al., 2012; Williams et al., 2013; Rammig et al., 2015), in-situ observation networks (Ciais et al., 2005; Reichstein et al., 2007; Wolf et al., 2016), ecosystem manipulation experiments (Knapp et al., 2002; Jentsch et al., 2007), satellite observations (Chambers et al., 2007; Zscheischler et al., 2013), and biogeochemical model simulations (Van Oijen et al., 2014; Rolinski et al., 2015).

Ecosystems sequester carbon slowly, but carbon release is thought to occur fast (Körner, 2003; Frank et al., 2015), for instance due to fire, windthrow, harvest, or drought. This expectation is consistent with a highly skewed distribution of carbon cycle anomalies globally, where in most regions losses induced by negative carbon uptake anomalies strongly exceed positive uptake anomalies (Zscheischler et al., 2014c). For example, the European heatwave and drought 2003 has been shown to undo four years of carbon sequestration (Ciais et al., 2005), thereby raising concerns that more frequent or intense climate extremes might turn ecosystems into carbon sources in the future. In fact, losses in ecosystem carbon have been reported for a number of large climate extremes in recent years such as droughts in Europe (Ciais et al., 2005; Reichstein et al., 2007), North America (Schwalm et al., 2012; Wolf et al., 2016), Australia (Ma et al., 2016), and the Amazon (Phillips et al., 2009; Lewis et al., 2011). Further, it has been shown that seasonal interactions of water and carbon dynamics, and soil moisture interactions, can shape the evolution of climate extremes (Seneviratne et al., 2010a; Teuling et al., 2010; Wolf et al., 2016). On the global scale a deficit in water availability is the main driver of carbon losses due to climate extremes (Zscheischler et al., 2014b).

However, a conceptual generalisation of the ecosystem impacts induced by climate extremes, or even upscaling to larger regions is a very difficult task for several reasons. First, ecosystem impacts of climate extremes are inherently nonlinear, including abrupt climatic thresholds that induce damages (Reichstein et al., 2013). Second, impacts can occur through direct and indirect pathways, and might occur concurrent to the climate extreme or lagged (see e.g. Anderegg et al. (2015) for an example of drought legacy on forest ecosystems, and Frank et al. (2015) for a conceptual overview and detailed review). Not all events that are extreme from a climatological perspective induce extreme impacts (Smith, 2011). Third, not all data streams that are used to investigate ecosystem extremes are equally suitable to investigate different types of ecosystem impacts. For example, satellite proxies of vegetation productivity might overlook effects in evergreen vegetation, because leaf or canopy properties do not show strong changes despite changing physiology (Frank et al., 2015). Fourth, productivity and respiration are partly sensitive to different driving variables and can thus be affected differently, which might therefore lead to differential net responses (Schwalm et al., 2010). Fifth, different ecosystem types and different plant species respond differently to climate extremes (Teuling et al., 2010; Babst et al., 2012; Yang et al., 2016), responses might change seasonally (Wolf et al., 2016), and physiological and phenological processes might interact (Reyer et al., 2013). Finally, the observational record is limited in time (continuous satellite records of more than a decade or two are only now becoming available (Schimel et al., 2015)) and space (in-situ observation networks and tree-ring archives are typically skewed towards temperate and boreal biomes (Baldocchi et al., 2016; Babst et al., 2017)). Further, extremes are rare by definition such that the number of ecosystem extremes that are available for comprehensive analysis is simply small. Therefore, crucial gaps remain in the understanding and quantification of ecosystem responses to climate extremes (Beier et al., 2012), and in particular regarding how different events are interacting in space and time.

Hence, long time series that allow to investigate ecosystem impacts of climate extremes, their interaction, and driving variables would constitute a crucial starting point to improve and generalise the impacts of climate extremes on terrestrial ecosystems. In Part III of this thesis, I introduce bias-corrected climate-ecosystem ensemble simulations that are designed for a comprehensive analysis of ecosystem extreme responses, and interacting carbon cycle effects due to climate extremes on a sub-continental scale in Europe.

1.5. Structure of the thesis

The central aim of this PhD thesis is to improve the quantification of, and contribute to the understanding of climate extremes and their impact on ecosystem-atmosphere interactions by a joint analysis and integration of observational datasets with model ensemble simulations. To arrive at these objectives, I first revisit methodological choices that allow a statistically robust quantification of climate extremes in either observational or simulated spatio-temporal datasets (Part I, Chapters 2–2). Second, I develop and apply tools to constrain and bias-correct climate model ensemble simulations with observational data in order to derive physically plausible and realistic datasets for an assessment of climate extremes and simulation of impacts (Part II, Chapters 5–7). Third, I assess and attribute the impacts of climate extremes in the terrestrial biosphere at regional scales in mid-latitude regions with a focus on spring-summer interactions of extreme ecosystem responses (Part III, 8, and Appendix A). Figure 1.4 illustrates the structure of the thesis in a conceptual sketch.

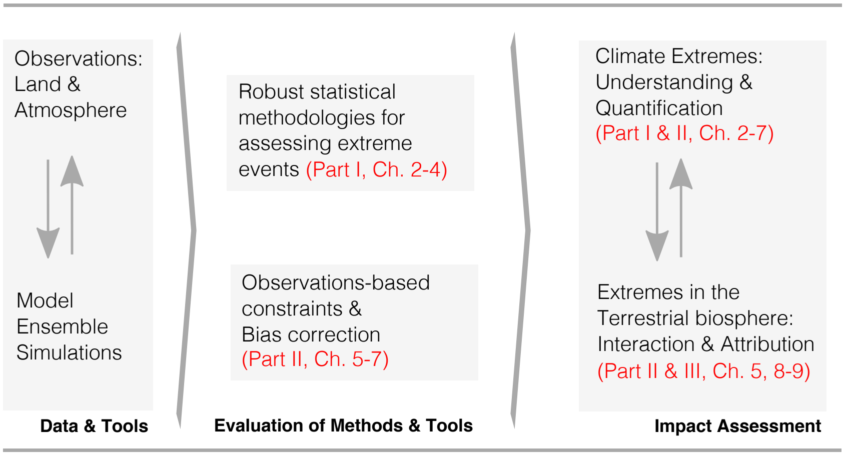


FIGURE 1.4.: **Structure of the thesis.** The central theme of this thesis is an integration and use of both model ensemble simulations and observational datasets to improve the understanding, quantification and attribution of climate extremes and their impact in terrestrial ecosystems. This includes 1) an evaluation of methodological choices, datasets, and models: This is achieved by revisiting statistical methodologies to quantify climate extremes in observed or simulated spatio-temporal datasets (Part I), and by identifying and testing useful observation-based constraints for assessments of climate extremes and their ecosystem impact (Part II). The evaluation of methods directly feeds into 2) an assessment of climate extremes (Part I & II) and their impacts in the terrestrial biosphere (Part II & III), including an attribution scheme for extreme ecosystem responses and interactions between different events.

Chapter 2 revisits a conventional statistical methodology that is used to quantify changes in climate variability and the occurrence of climate extremes (e.g. in temperature) in spatio-temporal observational datasets. It is shown that conventional standardisation of gridded data relative to the local mean and standard deviation of a reference period leads to an artificial increase in climate extremes and variability in the time steps that lie outside of a given reference period. In time-invariant Gaussian data with a reference period length of 30 years, the overestimation of '2-sigma extremes' would amount to 48.2%. It is also shown that the statistical artefacts can be corrected analytically assuming a Gaussian distribution, and earlier studies are revised to correct for normalisation-induced biases in estimating the occurrence of climate extremes.

Chapter 3 builds directly on Chapter 2 and elaborates further on a related phenomenon - namely statistical artefacts that are induced by standardisation of non-negative climate variables such as e.g. precipitation, i.e. dividing a random variable by a sample mean derived from a fixed reference period. The chapter reinvestigates the question whether observed precipitation extremes and annual totals have been increasing in the world's dry regions over the last 60 years. Despite recently postulated increasing trends, it is demonstrated that large uncertainties prevail that still preclude a definite answer to this question due to (1) statistical artefacts induced by data processing as noted above, and (2) the choice of dryness definition. Furthermore, an analytical description of the artefact induced by standardisation is induced that allows to estimate and correct for these biases.

Chapter 4 illustrates how climate model ensemble simulations can serve as a useful test bed for assessing the statistical robustness of methodological approaches–even if only small sample sizes are available. An empirical analysis of a large ensemble simulation is compared to inferences about rare climate extremes based on extreme value theory–in a case study of cold extremes and heavy precipitation at the regional scale in Europe. It is found that the parameter choices in extreme value statistical analysis are indeed crucial (e.g. the choice of 'block size' for selecting climate extremes), and biases could result if chosen inappropriately. Hence, model ensemble simulations can inform parameter choices for inferences about climate extremes in observations that are inherently limited in spatial and temporal extent.

In Chapter 5, a novel bias correction methodology is developed that is designed to minimise biases in regional climate model ensemble simulations while preserving multivariate correlations between variables and physical consistency on a seasonal time scale. The method uses an observed temperature distribution as a constraint to resample ensemble members, and is shown to considerably reduce biases in non-constrained variables such as precipitation or radiation. The representation of climate extremes is improved, and it is shown, using a biogeochemical model, that an accurate representation of climate forcing is a prerequisite for a plausible simulation of extreme impacts in the terrestrial biosphere.

In Chapter 6, the role of human-induced warming in Central European heatwaves in summer 2015 is assessed using a regional climate model ensemble, following the bias correction in Chapter 5, in tandem with an extreme value analysis based on observations. The attribution analysis shows that human-induced warming plays a role in occurrences of heatwaves in Europe, but quantitative estimates of risk ratios differ between observations and models.

Chapter 7 highlights the intimate links between the state of the land surface during heat events and observed biases in the representation of temperature extremes (cf. Chapter 5) in multi-model ensembles. Best estimates of land-atmosphere coupling and its uncertainties are inferred from a set of 54 different combinations of observations-based benchmark datasets of temperature and evapotranspiration, which are used to constrain a multi-model ensemble of climate simulations. This procedure is shown to reduce the magnitude of temperature extremes at present and in future predictions, which is highly relevant for predicting climate impacts for instance in the terrestrial biosphere.

In Chapter 8, an explicit assessment of extreme ecosystem impacts due to climate extremes at regional scale is presented using an ensemble of climate-ecosystem model simulations. It is shown that spring and summer trends over the last 25 years in ecosystem productivity extremes contrast each other (higher carbon uptake in spring extremes, carbon losses in summer extremes), which are driven by a spring vs. summer reversal in the response of ecosystem productivity to recent climatic changes (mainly to warming). Furthermore, evidence for interactions between spring and summer climate extremes affecting terrestrial ecosystems (cf. Appendix A) is presented.

Appendix A is a short commentary that highlights the role of the seasonal timing of climate extremes in triggering impacts in terrestrial ecosystems, and land-atmosphere feedbacks that amplify summer drought and result from early vegetation activity in spring. The commentary summarises a recent study that illustrated this phenomenon in the United States in 2012 (Wolf et al., 2016), where positive carbon cycle impacts due to a warm spring could compensate for drought-induced losses in ecosystem carbon uptake in summer. The study by Wolf et al. (2016) and Appendix A thus briefly present a motivation and conceptual basis of Chapter 8 in its current form.

In Chapter 9, I conclude on the research presented in this thesis, and potential directions of future research that could build upon the methods, tools and hypotheses that were developed in this thesis.

Part I.

Statistical quantification of extremes in observations and model ensembles

2. Quantifying changes in climate variability and extremes: pitfalls and their overcoming[1,2]

Abstract

Hot temperature extremes have increased substantially in frequency and magnitude over past decades. A widely used approach to quantify this phenomenon is standardizing temperature data relative to the local mean and variability of a reference period. Here we demonstrate that this conventional procedure leads to exaggerated estimates of increasing temperature variability and extremes. For example, the occurrence of '2-sigma extremes' would be overestimated by 48.2% compared to a given reference period of 30 years with time-invariant simulated Gaussian data. This corresponds to an increase from a 2.0% to 2.9% probability of such events. We derive an analytical correction revealing that these artifacts prevail in recent studies. Our analyses lead to a revision of earlier reports (e.g. Huntingford et al., 2013): For instance we show that there is no evidence for a recent increase in normalised temperature variability. In conclusion, we provide an analytical pathway to describe changes in variability and extremes in climate observations and model simulations.

2.1. Introduction

Quantifying to what extent the magnitude and frequency of extreme events are changing is a priority in climate change research (IPCC, 2012; Seneviratne et al.,

[1]This chapter is published as Sippel, S., J. Zscheischler, M. Heimann, F. E. L. Otto, J. Peters, and M. D. Mahecha. 2015. *Geophysical Research Letters* **42**(22), 9990–9998. doi:10.1002/2015GL066307.

[2]Supplementary material that complements this Chapter with more detailed explanations is available in Appendix B.

2014). In recent years, unusually hot temperature extremes have occurred and these events are increasingly exceeding the range of historical variability (Rahmstorf and Coumou, 2011; Mora et al., 2013). Considerable scientific debate has sparked around whether present-day changes in extreme events are due to the shifting mean climatology, or whether we are also confronted with changing variability (Hansen et al., 2012; Huntingford et al., 2013; Alexander and Perkins, 2013; Mora et al., 2013; Seneviratne et al., 2014). Of particular focus in this context are changes in temperature extremes, which have direct impacts upon human wellbeing and likewise affect ecosystem services and global biogeochemical cycles (IPCC, 2012; Reichstein et al., 2013).

A widely used approach to address this question relies on normalizing climate data relative to a reference period (Hansen et al., 2012; Coumou and Robinson, 2013; Huntingford et al., 2013; Kamae et al., 2014; Curry et al., 2014) aiming to objectively compare temperature variability and extremes across space and time. This approach conventionally derives standardised anomalies by locally subtracting the mean (μ_{ref}) from and dividing the observations by the standard deviation (σ_{ref}) estimated from some reference period:

$$z = \frac{X - \mu_{ref}}{\sigma_{ref}} \tag{2.1}$$

The idea is to rank or count events based on departures from the local climatology (as defined by the reference period) in units of standard deviation (σ). Transformations of this kind underpin studies of changes in the occurrence of monthly or seasonal temperature extremes (Hansen et al., 2012; Coumou and Robinson, 2013; Kamae et al., 2014; Curry et al., 2014) and variability (Huntingford et al., 2013). Further, so-derived standardised anomalies have been used to determine continental-scale rankings of the most significant meteorological or geophysical extreme events (Grumm and Hart, 2001; Hart and Grumm, 2001; Root et al., 2007; Graham and Grumm, 2010), and Kodra and Ganguly (2014) study asymmetry in the distributions of temperature extremes using a variant of this methodology.

In this paper, we demonstrate that this conventional normalisation procedure inevitably leads to erroneous and exaggerated estimates of temperature extremes and variability outside a specified 'reference period'. Furthermore, we derive an analytical correction that accounts for these statistical artifacts and allows for an accurate quantification of large-scale climate variability and extremes.

2.2. Methodology and results

2.2.1. Normalisation-induced artefacts and an analytical correction for quantifying extremes

To test the suitability of the reference-period normalisation, we conduct Monte-Carlo simulations with independent and identically distributed random variables drawn from a standard Gaussian distribution ($\mathcal{N}(\mu = 0, \sigma^2 = 1)$). This numerical experiment is set-up in analogy to investigations of monthly or seasonally standardised extremes (see Hansen et al., 2012, for an example) in gridded temperature data with $k = 10^4$ time series ('grid cells') and $n = 60$ data points per time series ('years of data'), but consisting of purely random Gaussian variables (i.i.d.). For each time series we generate anomalies and subsequently standardise these based on the conventional procedure (Eq. 2.1). Both mean ($\hat{\mu}_{ref}$) and standard deviation ($\hat{\sigma}_{ref}$) are estimated from each time series' first 30 values (i.e. $n_{ref} = 30$). The number of values exceeding σ extremes are counted at each time step in the original and normalised dataset (Figure 2.1, grey and red line, respectively).

Given that the statistical properties of the artificial data are time-invariant, there should be no change in the number of extremes across the dataset. However, in fact we find substantial increases in the number of extreme events outside the reference period along with a reduction in extremes within the reference period (Figure 2.1a, R code to reproduce these results in Section B.1). A quantification of 2σ extremes across all grid cells in the artificial dataset leads to a considerable increase (red line in Figure 2.1a) in the out-of-base period relative to the reference period of about 48.2%. Considering only the out-of-base period the number of 2σ (3σ) events would be overestimated by 29.1% (131.0%) relative to the original Gaussian data (black line in Figure 2.1a), which corresponds to an increase

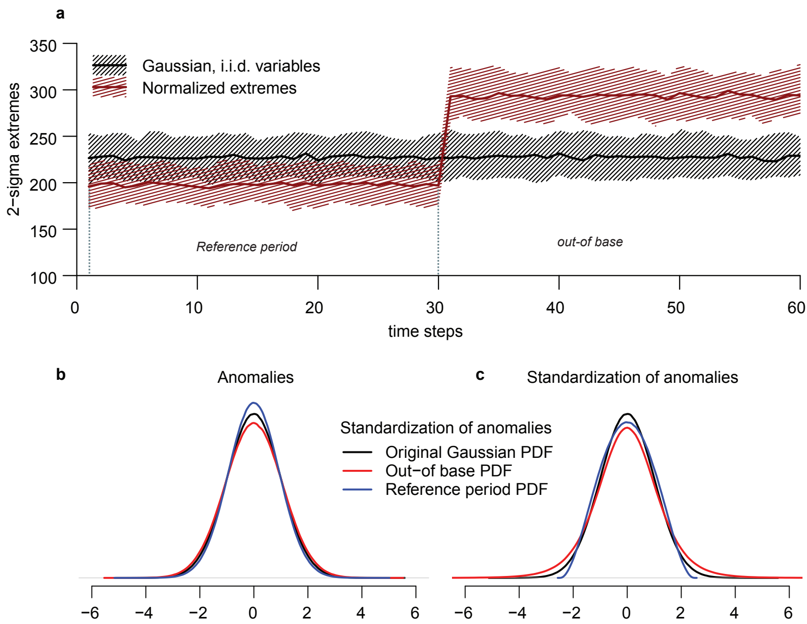


FIGURE 2.1.: Biases in the detection of extreme events in stationary and independent Gaussian data induced by normalisation. a) Occurrences of positive 2-sigma extremes in artificial Gaussian time series based on 10,000 replicates over 60 time-points before normalizing the data (black line), and after normalizing each replicate using the first 30 samples as reference period. b) Illustration of variance inflation and reduction through the generation of anomalies in the out-of-base (blue) vs. reference period (red) PDF ($n_{ref} = 8$ for illustration). c) Changing tails in normalised (i.e., divided by the SD estimate) Gaussian variables ($n_{ref} = 8$ for illustration). Coloured shading in (a) indicates the 5th to 95th percentile in repeated simulations.

from a 2.3% (1.3‰) chance to 2.9% (3.1‰). For illustration purposes, the distributions at a random time step inside and outside the reference period across all time series is shown in Figure 2.1b and 2.1c for anomalies and standardised variables, respectively. Overall, the artificial experiment reveals potentially severe artefacts in the widely applied reference period normalisation. In the following paragraphs, we reveal the consequences of this conventional normalisation and derive an analytical solution for the induced artefacts.

To understand the origin of the apparent increase in extremes we have to consider that the ‘true’ values for mean and variability are inherently unknown, which changes Eq. 2.1 to:

$$z = \frac{X - \hat{\mu}_{ref}}{\hat{\sigma}_{ref}}. \tag{2.2}$$

The estimates of the mean ($\hat{\mu}_{ref}$) and standard deviation ($\hat{\sigma}_{ref}$) are random variables with well-known statistical properties (Von Storch and Zwiers, 2001), drawn from an independent sample in case of analyzing the out-of-base period (Zhang et al., 2005) (see Section B.2 for a detailed statistical description), and subsequently pooled in space. Consequently, the biases between both periods are induced by a combination of two effects, firstly the generation of anomalies ($X_{anom} = X - \hat{\mu}_{ref}$), and secondly the standardisation ($z = \frac{X_{anom}}{\hat{\sigma}_{ref}}$) (Figure 2.1b,c): The generation of anomalies systematically increases (decreases) the variance across grid cells in the out-of-base (reference) period (Tingley, 2012), but does not affect the underlying distribution (Section B.2). However, the local standardisation of each time series induces qualitative changes to the (spatial) distribution (for an analytical derivation see Section B.2) such that heavier tails outside the reference period are induced (Figure 2.1c). This qualitative difference stems from the fact that any time point in the out-of-base period follows a *t*-distribution with $n_{ref} - 1$ degrees of freedom (Section B.2). Hence, the heavier tails generated by the conventional standardisation lead to a consistent and potentially severe overestimation of extreme events in the out-of-base period (Figure 2.1a) for relatively short, but in practice often used, sometimes unavoidable, reference periods. However, the distribution after normalisation can be derived analytically (Section B.2), and hence the biases can be rectified separately both for the reference and the out-of-base periods. Specifically, instead of counting 2σ (3σ) extremes in the out-of-base period, a search for the corresponding percentile threshold in the variance-adjusted *t*-distribution (2.12σ (3.32σ), respectively, if $n = 30$) would allow for the detection of the correct number of events (Figure 2.2a, Figure B1 for an illustration of the correction procedure). Further, it is worth noting that even with an increasing number of samples in the reference period, the convergence to small biases is slow. For autocorrelated data the artefacts are even more pronounced owing to a smaller effective sample size (Figure B2a and Figure B2b, respectively).

Before applying the proposed analytical correction we have to consider that temperatures at monthly or seasonal time scales are typically non-stationary (Ji et al., 2014), i.e. simulated or observed time series might contain spatially and temporarily diverse trends. Using Monte-Carlo type simulations of normalised Gaussian time series with changing trends and variability we find that both exerts strong influence on the magnitude of the biases (Section B.3). Increasing (decreasing) trends or variability in the out-of-base period severely deflates (inflates) the biases for the upper tail (Figure B2a,b). These insights are equally applicable to the lower tail of the distribution if the sign of the trend is reversed. To assess the issue of non-stationarity in more detail, we consider trends and changes in variability in the artificial dataset introduced in Figure 2.1. First, random linear trends are added in the out-of-base period to each random Gaussian time series, where the magnitudes of the trends at the last time step are drawn randomly for each grid cell from a uniform distribution in the interval $[-1 \leq \delta \leq 1]$ in units of σ (Figure 2.2b). Second, we investigate a trend in the out-of-base period coinciding with randomly assigned changes in variability ($0.8 \leq \sigma \leq 1.2$, Figure 2.2c).

Following the solution for stationary time series outlined above, we offer an analytical correction that allows handling of the additional artefacts induced by non-stationarities (Section B.4). In essence, normalizing non-stationary data induces a non-central version of Student's *t*-distribution. This analytical distribution can be used to avoid normalisation-induced biases entirely if changes in the trend or variability are known (Figure 2.2b,c). Likewise, estimating the trend and/or changes in variability largely allowes for removing the biases (Figure 2.2b,c). As above, σ-extremes are counted based on the biased estimate of the conventional procedure (red line), and based on the application of the suggested correction procedure using known (blue) and estimated (green) trends and changes in variability. Throughout this paper, Singular Spectrum Analysis (SSA), a non-linear spectral decomposition methodology (Golyandina and Zhigljavsky, 2013; von Buttlar et al., 2014) is used to estimate trend components, before the analytical correction procedure based on the noncentral *t*-distribution is applied. Trends are extracted as 31-year and larger components using a 45-year SSA window length ($L = 45$).

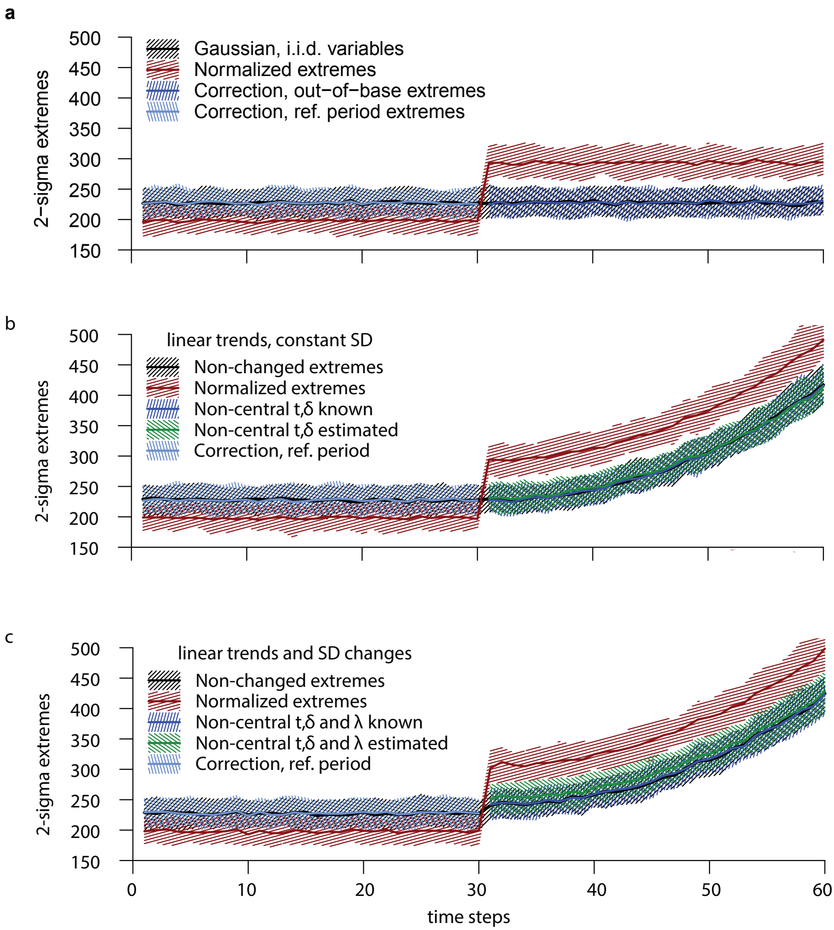


FIGURE 2.2.: Correction of normalisation-induced biases in stationary and non-stationary time series consisting of independent random variables. Detecting 2-sigma extreme events in a) Stationary Gaussian time series, b) Gaussian time series with random linear trends added in the out-of-base period ($-1 < \delta_{t=60} < 1$, in units of σ), c) Gaussian time series with random linear trends ($-1 < \delta_{t=60} < 1$, in units of σ) and changing variance ($0.8\sigma_{ref} < \lambda\sigma_{ref} < 1.2\sigma_{ref}$) in the out-of-base period. In each panel, coloured shading indicates the 5th to 95th percentile in repeated simulations ($k = 10^4$ simulated time series in all panels).

2.2.2. Quantifying extremes in Earth observation data

In this subsection, we assess how monthly temperature extremes on land have changed over the second half of the 20th century in the Northern hemisphere up to present by applying the statistical approach outlined above. In order to avoid potential inhomogeneities related to gridded observations, we analyse the state-of-the-art Twentieth Century Reanalysis dataset (Compo et al., 2011) (Version 2). The reanalysis dataset assimilates only surface pressure measurements and monthly sea surface temperatures into an atmosphere and land general circulation model (Compo et al., 2011) and is hence independent from station temperature measurements. The dataset has been specifically designed to assess climate variability and extremes statistics 'spanning the instrumental record', and has been demonstrated to reproduce the observed temperature trends and variability to a very large extent (Compo et al., 2011).

In our analysis, we first interpolate the dataset to a 2° x 2° regular latitude-longitude grid, and mask ocean pixels. Second, we estimate separately for each month and grid cell the trend component, local mean and (non-detrended and detrended) standard deviation in two different reference periods (1921-1950 and 1951-1980). Thirdly, each pixel time series is normalised using both reference periods and the detrended and non-detrended σ_{ref} estimates. For each month we calculate the area affected by 2σ and 3σ extremes, using the conventional normalisation approach and our correction. We use the trend estimates for our correction, but assume an approximately unchanged variance over the past decades (Huntingford et al., 2013). Lastly, we derive seasonal averages of the 'area affected by extremes' for Northern hemisphere summer (JJA, Figure 2.3).

Our analysis reveals that the exceedance of monthly 2σ and 3σ temperature extremes in summer has indeed increased substantially over the Northern hemisphere (Figure 2.3a,b for land areas in the NH outer tropics). However, the bias-adjusted time series show a consistently slower and smoother increase as compared to the conventionally applied uncorrected normalisation procedure. A break point analysis using piecewise linear regression (Toms and Lesperance, 2003) based on our revised figures indicates that the recent rapid increase in hot summer months in the Northern hemisphere (2σ and 3σ events) started to emerge around the late 1980s or early 1990s (Figure 2.3b).

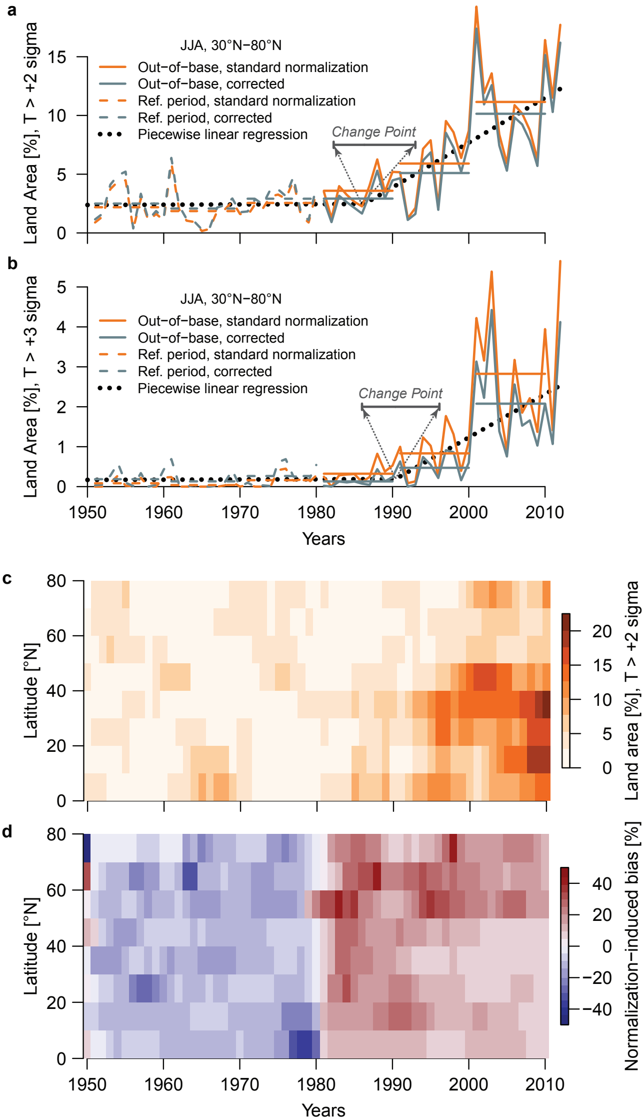


FIGURE 2.3.: Full caption is displayed on the next page.

The magnitude of the biases and the discontinuities at the reference and out-of-base period are robust across different reference periods, and also hold if trends

FIGURE 2.3.: (continued) Increase in normalised hot temperature extremes in a spatio-temporal dataset (20th Century Reanalysis (Compo et al., 2011)). a,b) Time series of fraction of extratropical Northern hemisphere land area covered by positive monthly 2σ (a) and 3σ (b) events in summer (reference period: 1951-1980). Horizontal lines indicate decadal averages for the conventional normalisation procedure (light blue) and our proposed correction (orange). c) Zonal evolution of fraction of land area covered by monthly positive 2σ extremes in Northern hemisphere summer. d) Zonal evolution of relative biases induced by the conventional normalisation approach.

are subtracted before estimating local variability (Coumou and Robinson, 2013) (Figure B3 and Figure B4). Increases in extremes relative to local variability show a clear zonal pattern (Figure 2.3c) with the largest increases in the tropics and subtropics. Therefore, biases induced by the normalisation are largest in areas where the trend is relatively small compared to local variability (Figure 2.3d). However, it is worth noting that peculiarities of the station-based observational record such as urban heat islands or local land-use changes are not accounted for in the 20th Century Reanalysis (Parker, 2011). In addition, the availability of pressure observations varies through time (Compo et al., 2011). As such, the main purpose of the present analysis is to illustrate the potential biases induced by reference period standardisation in spatio-temporal datasets.

2.2.3. Implications for large-scale assessments of variability and asymmetry

Normalisation-induced biases are not only relevant for assessments of extremes, but a careful consideration of such statistical pre-processing techniques is equally important for analysis of variability and asymmetry in spatio-temporal datasets. An example is provided by a recent study that investigated whether temperature variability has changed over the second half of the 20th century on global and continental scales (Huntingford et al., 2013). The authors argue that annual temperatures in low-variance regions have become more variable over the past decades, whilst global temperature variability has remained near constant. This explanation stems from the authors' observation that normalised variability has increased more than absolute (spatial) variability (16% vs. 2% increases between 1963-

1980 and 1981-1996). Using the 20th Century reanalysis dataset we reproduce the increases in the annual, global, area-weighted standard deviation (12.9% vs. 1.8% increases, when using the conventional data processing scheme (Huntingford et al., 2013), Figure 2.4).

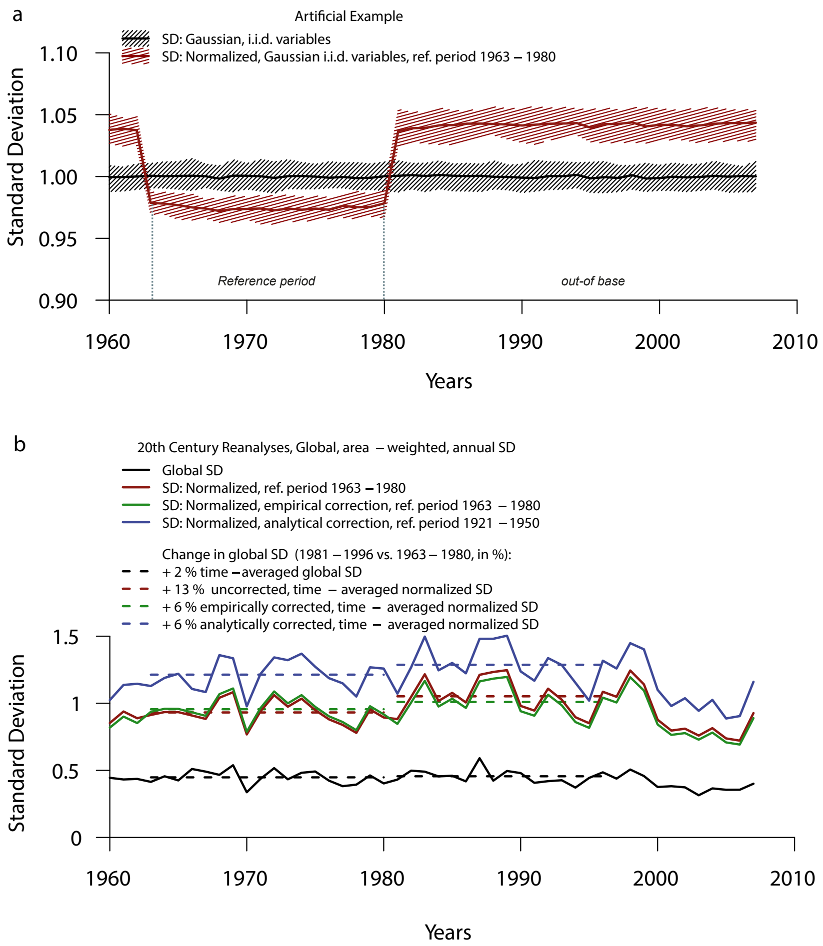


FIGURE 2.4.: Normalisation-induced changes in variability. a,b) Time series of normalised variability following the data processing scheme of Huntingford et al. (2013) in an artificial example ($k = 10^4$ time series) with i.i.d. Gaussian variables (a) and in the 20th Century Reanalysis dataset (b).

However, an artificial experiment in analogy to the previous subsection shows that the conventional normalisation procedure changes the standard deviation of the data (Figure 2.4a), and in particular yields an increase in standard deviation between the reference and the out-of-base period. Therefore, we correct the conventionally normalised standard deviation of annual temperatures in the 20th Century Reanalysis dataset empirically and analytically. The former is achieved by simulating the reduction in standard deviation in artificial Gaussian data (Fig. 2.4a), whereas the latter is achieved by using an earlier reference period (1921-1950) and the application of our analytical correction. The empirical and analytical corrections reduce the increase in normalised variability from 12.9% to 5.6% and 6.0%, respectively (see Fig. 2.4b). A permutation-based significance test (Fay and Shaw, 2010) shows that the increases in mean corrected normalised standard deviation between both periods are not significant ($p_{empirical} = 0.147$ and $p_{analytical} = 0.110$), whereas conventional normalisation yields a highly significant increase ($p_{conventional} = 0.004$). Hence, the relatively small and non-significant difference between the increases in standardised and absolute variability might indeed be due to the explanation offered previously (Huntingford et al., 2013), and potentially related to major El-Niño events in the latter period (Fedorov and Philander, 2000). If the periods before and after 1980 are extended to derive a larger sample, this reduces the increase in normalised variability to only 2% (1981-2006 vs. 1955-1980). Thus, based on our proposed normalisation we cannot confirm that changes across low-variance regions have occurred over the past decades. Nonetheless, our results underpin that global temperature variability has not changed (Huntingford et al., 2013), and additionally show that this finding holds both in absolute and normalised terms.

Finally, another recent study (Kodra and Ganguly, 2014) reports that asymmetry in temperature distributions of seasonal extreme values at daily time scale (both minima and maxima, i.e. the hottest and coldest day per season) is strongly increasing towards both the cold and hot tails in model projections of future climate conditions relative to a recent period. As a pre-processing step, the authors derive ‘anomalies’ of seasonal extremes by subtracting the mean of the recent (historical) climatology of seasonal extremes from both periods. This procedure leads to narrower distributions in the reference period and a broader distribution in the future (independent) period (see Section B.2). This variance inflation in skewed extreme value distributions leads to the observed effect even in stationary time series, and should hence be interpreted with caution (Figure B5, and Section B.6).

2.3. Outlook and conclusion

The observation that a commonly used normalisation of temperature data is inappropriate for assessing changes in variability, extremes, and asymmetry is of general validity and should also be considered in investigations of other climatological and Earth observations. The steadily growing archives of Earth observations derived from both ground based as well as satellite remote sensing data requires reconsidering conventional data analytic approaches such as standardisation. For instance, extremes in gridded standardised anomalies of rainfall and storms (Grumm and Hart, 2001; Hart and Grumm, 2001; Root et al., 2007; Graham and Grumm, 2010; Curry et al., 2014) have been studied using varieties of the conventional standardisation procedure and are potentially distorted by the artefacts discussed in this paper. Further, our results might facilitate the interpretation of single climatic extreme events or trends that are frequently characterised in terms of standardised departure from climatology, both inside and/or outside the climatological reference period (Schär et al., 2004; Barriopedro et al., 2011; Xu et al., 2012; Ramos et al., 2014; Cook et al., 2015). Although our analytical treatment using the t-distribution is confined to distributions that can be approximated as Gaussian, we emphasise that the induction of biases in the tails due to dependent/independent estimators of location and scale are fundamental and hold indeed across a wide range of distributions. Furthermore, because temperature extremes are bounded (Nogaj et al., 2006), approximations of temperature values by distributions with infinite tails (such as Gaussian and the t-distribution) might poorly estimate the most extreme temperatures. Here we offer a correction which adjusts biases in variability and extremes induced by a widely used data preprocessing approach. Alternatively, statistically more advanced but readily available tools, such as the theory of extreme values (Katz et al., 2013; Nogaj et al., 2006) offer complementary approaches to quantify extreme events under non-stationary conditions that are not affected by the statistical issues reported in this paper.

In conclusion, data normalisation for the detection of changes in extremes or variability has to be applied with caution: otherwise there is a risk to arbitrarily inflate both extremes and variability in the time periods under scrutiny. Our study demostrates how to avoid biases of this kind. However, our analyses do not call into question the major qualitative results that were outlined in previous studies (Hansen et al., 2012; Seneviratne et al., 2014): hot temperature extremes have increased considerably on the global scale, a trend which is most likely to continue throughout the 21st century (Coumou and Robinson, 2013; Sillmann et al., 2013a).

3. Have precipitation extremes and annual totals been increasing in the world's dry regions over the last 60 years?[1,2]

Abstract

Daily precipitation extremes and annual totals have increased in large parts of the global land area over the past decades. These observations are consistent with theoretical considerations of a warming climate. However, until recently these trends have not been shown to consistently affect dry regions over land. A recent study, published by Donat et al. (2016), now identified significant increases in annual-maximum daily extreme precipitation (Rx1d) and annual precipitation totals (PRCPTOT) in dry regions. Here, we revisit the applied methods and explore the sensitivity of changes in precipitation extremes and annual totals to alternative choices of defining a dry region (i.e. in terms of aridity as opposed to precipitation characteristics alone). We find that (a) statistical artifacts introduced by data pre-processing based on a time-invariant reference period lead to an overestimation of the reported trends by up to 40 %, and that (b) the reported trends of globally aggregated extremes and annual totals are highly sensitive to the definition of a 'dry region of the globe'. For example, using the same observational dataset, accounting for the statistical artifacts, and based on different aridity-based dryness definitions, we find a reduction in the positive trend of

[1]This chapter is published as Sippel, S., J. Zscheischler, M. Heimann, H. Lange, M. D. Mahecha, G. J. van Oldenborgh, F. E. L. Otto, and M. Reichstein. 2017. *Hydrology and Earth System Sciences* **21**, 441–458. doi:10.5194/hess-21-441-2017. The statistical results presented in this chapter underlie an Addendum published as Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A., and Maher, N. 2017. *Nature Climate Change* **7**, 154–158. doi:10.1038/nclimate3160.

[2]Supplementary material that complements this Chapter with more detailed explanations is available in Appendix C.

Rx1d from the originally reported $+1.6\,\%\,decade^{-1}$ to $+0.2$ to $+0.9\,\%\,decade^{-1}$ (period changes for 1981–2010 averages relative to 1951–1980 are reduced to -1.32 to $+0.97\,\%$ as opposed to $+4.85\,\%$ in the original study). If we include additional but less homogenised data to cover larger regions, the global trend increases slightly (Rx1d: $+0.4$ to $+1.1\,\%\,decade^{-1}$), and in this case we can indeed confirm (partly) significant increases in Rx1d. However, these globally aggregated estimates remain uncertain as considerable gaps in long-term observations in the Earth's arid and semi-arid regions remain. In summary, adequate data pre-processing and accounting for uncertainties regarding the definition of dryness are crucial to the quantification of spatially aggregated trends in precipitation extremes in the world's dry regions. In view of the high relevance of the question to many potentially affected stakeholders, we call for a well-reflected choice of specific data processing methods and the inclusion of alternative dryness definitions to guarantee that communicated results related to climate change be robust.

3.1. Introduction

Daily precipitation extremes are expected to increase over large parts of the global land area roughly by 6–7 % per °C of warming due to a higher atmospheric water-holding capacity as specified by the Clausius–Clapeyron equation (Allen and Ingram, 2002; Trenberth et al., 2003). Quantifying and predicting changes in precipitation characteristics due to climate change is crucial for water availability assessments and adaptation to climate change (IPCC, 2012; Greve et al., 2014). On a global scale, daily precipitation extremes have been observed to intensify (Donat et al., 2013b; Westra et al., 2013; O'Gorman, 2015), consistent with global model simulations (Fischer and Knutti, 2015), and coincide with a global-scale increase in observed annual precipitation totals (Donat et al., 2013b). However, there is little information to date on how precipitation characteristics have changed in the past over dry land areas and how they will change in the future. Donat et al. (2016) investigated whether and to what extent daily precipitation extremes (Rx1d) and annual precipitation totals (PRCPTOT) have increased over the last 60 years using observational data. The authors identified rapid increases in Rx1d over dry regions, which strongly outpace the corresponding increases over wet areas, and found a similar pattern for PRCPTOT.

The question whether precipitation extremes increase in dry regions is highly relevant in the context of climate change adaptation, as generally dry areas may be less prepared to deal with precipitation extremes (Ingram, 2016). Consequently, the recent report on increasing Rx1d in dry areas was highlighted in major Science journals (including *Nature* News (Tollefson, 2016), and *Nature Climate Change* (Ingram, 2016)) and received a lot of media coverage[3], which indicates the importance of this topic for the scientific community, the public and decision makers.

However, scrutinizing the findings by Donat et al. (2016) reveals two major issues of concern: first, the applied statistical approach introduces two systematic biases that lead to a substantial overestimation of the increase in PRCPTOT and Rx1d of up to 40 % in dry regions. Wet regions, by contrast, are only affected to a limited degree due to an approximate cancellation of errors in trend estimates. Second, the definition of a dry region used in Donat et al. (2016) based on PRCPTOT and Rx1d alone does only partly reflect the water balance and thus water availability (for instance, it ignores losses through evapotranspiration). Furthermore, defining dryness based on low Rx1d (Donat et al., 2016) takes a decision on whether a region is dry or not based on only 1 day in the year. The chosen definitions thus induce considerable uncertainty in the reported results. If we test alternative but well-established definitions of a 'dry region' (based on water supply and demand, either implicitly or explicitly; see Köppen, 1900; Greve et al., 2014) and apply the appropriate statistical tools, we find strongly reduced trends and period changes (1981–2010 averages relative to the 1951–1980 reference period) in PRCPTOT and Rx1d in the world's dry regions. An accurate quantification of trends and changes in precipitation characteristics is of high relevance and a

[3] http://www.huffingtonpost.com/entry/global-warming-will-bring-extreme-rain-and-flooding-study-finds_us_56e081c7e4b0860f99d796ab, https://www.theguardian.com/environment/2016/mar/08/hotter-planet-spells-harder-rains-to-come-study, https://www.sciencedaily.com/releases/2016/03/160308105625.htm, http://phys.org/news/2016-03-global-world-driest-areas.html, http://www.abc.net.au/news/2016-03-08/climate-change-could-bring-more-rain-to-deserts-study/7229236, http://www.asce.org/magazine/20160412-climate-change-to-cause-more-precipitation-in-dry-regions,-researchers-say/

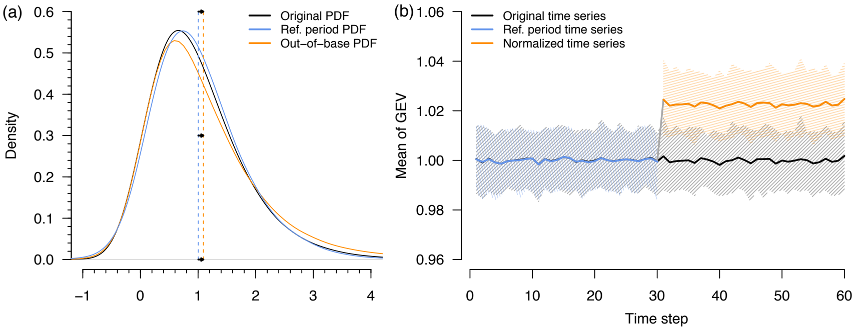


FIGURE 3.1.: Conceptual example of biases in the mean induced by normalisation based on a fixed reference period. a) Probability distributions and their respective means for an artificial dataset of 10^4 grid cells each comprised of random variables sampled from a generalised extreme value distribution (GEV; $\mu = 1$, $\sigma = 1$, $\xi = 0$, sample size $n_{\mathrm{ref}} = 8$ for illustration) distribution, and normalised following Donat et al. (2016) with different reference periods. b) Shift in the mean of spatially aggregated variables due to reference period normalisation ($n_{\mathrm{ref}} = 30$ following Donat et al., 2016, confidence intervals denote the 5th to 95th percentile). Code to reproduce this example is provided in the Supplement.

crucial prerequisite in the context of making climate change adaptation decisions (e.g. IPCC, 2014).

3.2. On data pre-processing based on a time-invariant reference period

As a first step in the analysis of Donat et al. (2016), the authors normalise the 60-year time series in the gridded HadEX2 dataset (Donat et al., 2013b) for each grid point with the sample mean of a 30-year reference period (1951–1980), which is a widespread procedure in climate science. However, this procedure artificially increases the mean of the spatial distribution in the out-of-base period (1981–2010) in all investigated time series, simply because variability in the sample means inflates the signal in the latter period (Sippel et al., 2015b). To illustrate this point, consider two hypothetical climate regions of the same size; in region one, the mean of a precipitation quantity increases between two periods (from 100 to 200 mm, say), for example due to a few large extremes, whereas it decreases by exactly the same amount in region two (i.e. from 200 to 100 mm). Consequently, over the combined period the spatial average and the spread of the two regions would be statistically indistinguishable. However, normalizing by the mean of the first time period would imply that the spatial average across both regions for the second period is 1.25 (the average of 0.5 and 2), i.e. a spurious increase of 25 % between both periods. This issue is illustrated in Figure 3.1 for an artificial dataset that consists of $n = 10^4$ time series (e.g. 'grid cells') that are drawn randomly and independently from a generalised extreme-value (GEV, Coles et al., 2001) distribution. The GEV distribution provides an asymptotical limit model for maxima derived from a sequence of random variables with a fixed block size (Coles et al., 2001, e.g. Rx1d,), and is therefore appropriate to illustrate this issue. Normalizing each time series in the artificial dataset by its mean in the first period yields a spatial 'reference period distribution' that is different from the spatial 'out-of-base period distribution' (and from the original GEV distribution; Figure 3.1a). In particular, this normalisation leads to increased spatial averages in the out-of-base period (Figure 3.1b). Furthermore, the normalisation procedure induces a considerable increase in the variance, skewness, and higher statistical moments in the spatial distribution in the out-of-base period (see e.g. Figure 3.1a), which would be of relevance if higher statistical moments (e.g. changes in spatial variance) were studied. The reason for this difference lies in the fact that the estimated sample means (of the reference period) are statistically dependent to reference period time series, but (virtually) independent to the time period that lies outside of the reference period (Zhang et al., 2005; Sippel et al., 2015b). It is worth noting that these biases can be understood analytically (Section C.1). The expected value Δ_{bias}, defined as the relative bias in the out-of-base period, can be well approximated for each grid cell with

$$\Delta_{\text{bias}} \approx \frac{\sigma^2}{\mu^2 n_{\text{ref}}}, \tag{3.1}$$

where μ, σ, and n_{ref} denote the time series' mean, standard deviation, and reference period length, respectively (Section C.1). Thereby, it can immediately be seen that the introduced bias is systematically positive outside of the reference

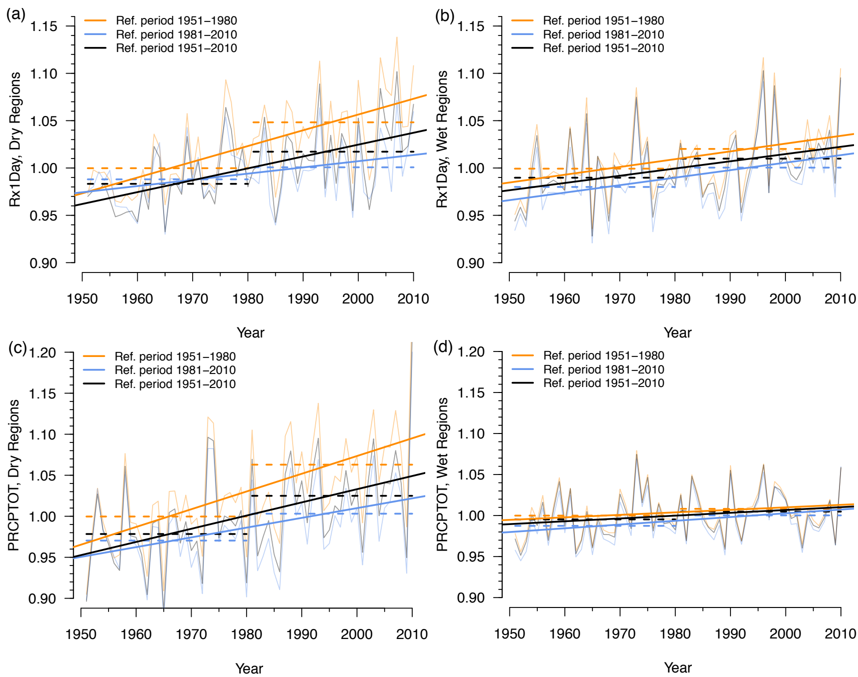


FIGURE 3.2.: Normalisation-induced biases on time series and trend estimates. a, b) Time series, trends, and 30-year means of spatially aggregated heavy precipitation (Rx1d) in (a) dry and (b) wet regions. c, d) Time series, trends, and 30-year means of spatially aggregated total precipitation (PRCPTOT) in (a) dry and (b) wet regions. Orange lines are taken from Donat et al. (2016) (ref. period: 1951–1980), black lines are corrected for biases (ref. period: 1951–2010), and blue lines indicate a hypothetical 1981–2010 reference period.

period, and it is proportional to the ratio of $\frac{\sigma^2}{\mu^2}$ for any fixed reference period length.

An additional statistical bias stems from the choice of the world's 30 % wettest and 30 % driest regions based on the climatology of PRCPTOT and Rx1d in the reference period (1951–1980). Because 30 years is fairly short to derive a robust climatology of the tails of the precipitation distribution, the computed changes in wet and dry regions are distorted by the 'regression to the mean' phenomenon (Galton, 1886; Barnett et al., 2005). To illustrate this issue, recall the conceptual two-region example quoted above, where variation between the two available

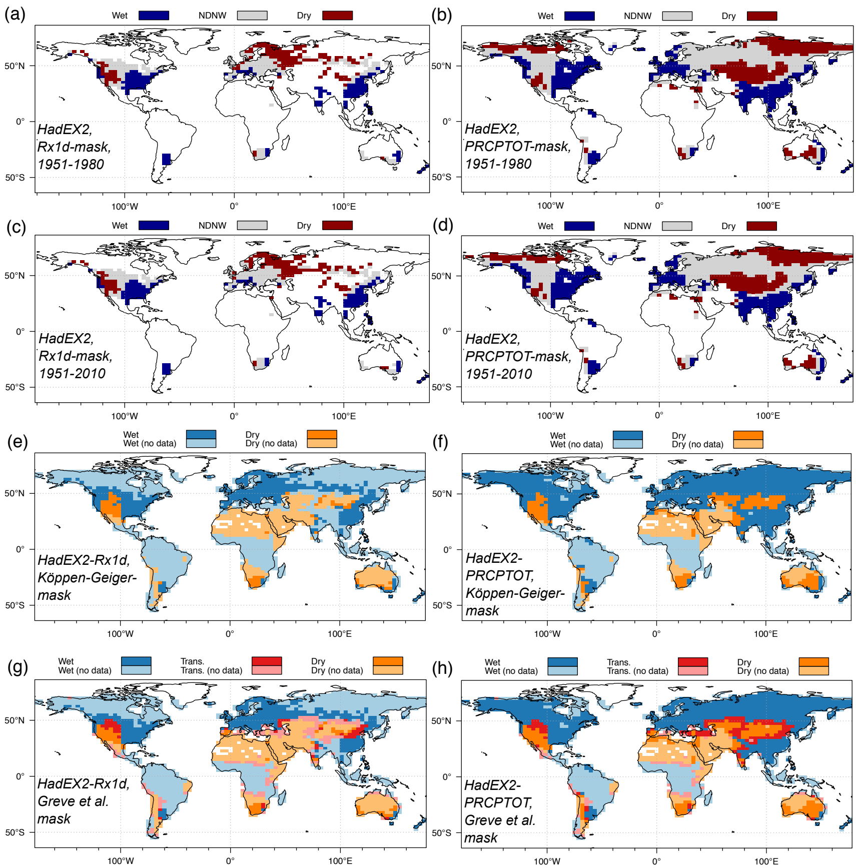


FIGURE 3.3.: Different mask of the world's dry and wet regions. a)–d) Dryness/wetness masks based on 1951–1980 and HadEX2 (a, b; see Donat et al. (2016)) and 1951–2010 (c, d; to avoid 'regression to the mean' selection bias, see text) for Rx1d (left panels) and PRCPTOT (right panels). 'NDNW' indicates neither dry nor wet areas, white inland areas indicate less than 90 % data availability in the HadEX2 dataset and were not considered. e, f) Dry regions based on the Köppen–Geiger classification as updated by Kottek et al. (2006) and data availability in HadEX2. g, h) Dry and transitional regions following Greve et al. (2014) and data availability in HadEX2.

TABLE 3.1.: Statistical pre-processing uncertainties and biases in period increments and trend slopes

World Region	Precipitation characteristic	Ref. Period (Normalisation)	Ref. Period (region selection)	Period Increment[1] [%]	Bias [%]	Sen slope [decade^{-1}]	Bias [%]	Type of bias
Dry (HadEX2, 30% lowest Rx1day)	Rx1d	1951-1980	1951-1980	4.85	40.4	0.016	33.3	[2]
	Rx1d	1981-2010	1981-2010	1.29	-62.7	0.006	-50.0	[3]
	Rx1d	1951-2010	1951-2010	3.45	0.0	0.012	0.0	[4]
	Rx1d	1951-1980	1951-2010	3.97	15.1	0.014	16.7	[5]
	Rx1d	1951-2010	1951-1980	4.33	25.3	0.014	16.7	[6]
Wet (HadEX2, 70% highest Rx1day)	Rx1d	1951-1980	1951-1980	2.09	2.2	0.007	8.7	[2]
	Rx1d	1981-2010	1981-2010	2.09	2.2	0.007	-1.5	[3]
	Rx1d	1951-2010	1951-2010	2.04	0.0	0.007	0.0	[4]
	Rx1d	1951-1980	1951-2010	2.41	18.1	0.008	16.0	[5]
	Rx1d	1951-2010	1951-1980	1.73	-15.3	0.006	-4.8	[6]
Dry (HadEX2, 30% lowest PRCPTOT)	PRCPTOT	1951-1980	1951-1980	6.32	32.9	0.020	40.4	[2]
	PRCPTOT	1981-2010	1981-2010	3.38	-29.0	0.010	-29.5	[3]
	PRCPTOT	1951-2010	1951-2010	4.76	0.0	0.015	0.0	[4]
	PRCPTOT	1951-1980	1951-2010	5.74	20.8	0.019	27.5	[5]
	PRCPTOT	1951-2010	1951-1980	5.34	12.2	0.017	14.9	[6]
Wet (HadEX2, 70% highest PRCPTOT)	PRCPTOT	1951-1980	1951-1980	0.83	-13.7	0.003	-13.6	[2]
	PRCPTOT	1981-2010	1981-2010	1.30	35.5	0.005	28.9	[3]
	PRCPTOT	1951-2010	1951-2010	0.96	0.0	0.004	0.0	[4]
	PRCPTOT	1951-1980	1951-2010	1.32	38.5	0.005	38.2	[5]
	PRCPTOT	1951-2010	1951-1980	0.40	-58.6	0.002	-52.4	[6]

[1] Period increment denotes the change in period means between 1981-2010 vs. 1951-1980.
[2] Combination of ‘Normalisation’ and ‘Regression to mean’ (RTM) bias, ‘early’ ref. period (i.e. following Donat et al. (2016))
[3] Combination of ‘Normalisation’ and ‘RTM’ bias, ‘late’ ref. period
[4] Ref. Period covering the entire temporal domain (no bias)
[5] ‘Normalisation’ bias only
[6] ‘RTM’ bias only. Red indicates period increments and trend estimates based on the 1951–1980 reference period; blue indicates period increments and trend estimates based on the 1981–2010 reference period.

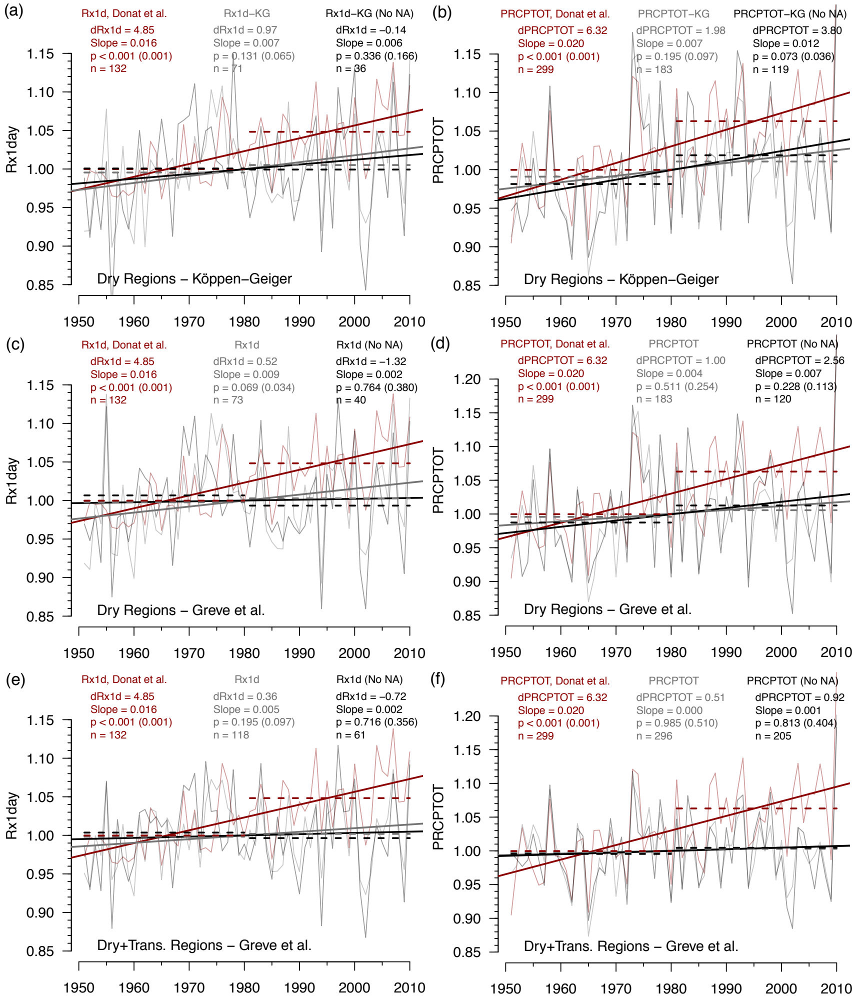


FIGURE 3.4.: a)–f) Time series, trends, and 30-year means of spatially aggregated heavy precipitation (Rx1d, a, c, e) and annual rainfall totals (PRCPTOT, b, d, f) in dry regions following (a, b) the Köppen–Geiger classification (Kottek et al., 2006), (c, d) Greve et al. (2014), and (e, f) dry and transitional regions combined (Greve et al., 2014). Red lines are drawn as reported in Donat et al. (2016) for comparison, i.e. based on the 1951–1980 reference period and dryness defined as 'moderate extreme precipitation' (Rx1d) and annual precipitation totals (PRCPTOT). Grey and black lines are corrected for statistical artefacts (1951–2010 reference period), and dry regions are defined based on aridity. Grey lines report 90 % complete time series, black lines report only data with 100 % complete temporal coverage. All p values are given for two-sided (one-sided) Mann–Kendall trend tests.

time periods would be entirely due to random causes. If any of the two periods would be chosen to stratify the dataset in one dry and one wet region, this would result in opposing changes (i.e. dry gets wetter, wet gets drier) in the independent period. In other words, selecting from the dry (wet) end of the spatial distribution in one subset of the dataset, or ‘reference period’, will result in a higher probability for wetter (drier) conditions in the remaining years if any type of random variation plays a role (Table 3.1, and Figure 3.2 for changes due to both statistical effects). Although random variations in 30-year averages are not very large (cf. Figure 3.3a and 3.3b and Figure 3.3c and 3.3d), it is important to consider this effect as it is indeed noticeable in the reported results (Table 3.1).

The chosen normalisation approach combined with the spatial point selection method results in a bias toward PRCPTOT and Rx1d increasing at a faster rate in dry regions compared to wet regions. Over dry regions, both effects lead to an overestimation of the trends in precipitation totals and extremes by +40.3 and +33.2 % (+32.9 and +40.4 % overestimation in the reported period changes from 1951–1980 to 1981–2010), respectively (Figure 3.2, Table 3.1). In contrast, in wet regions both errors roughly cancel each other out in the case of extremes (increase by only +8.7 %) and lead to a small underestimation of the increase in total precipitation (−13.7 %). In summary, we find that the applied pre-processing steps are crucial to accurately quantify changes in precipitation extremes and annual totals. In the study under scrutiny, if the dryness definition is kept, trends and period increments are corrected to much lower values, but the trends and period increments remain positive and significant (see Figure 3.2).

3.3. On the definition of a dry region

Climatological dryness is typically not determined by water supply alone but also depends on atmospheric water demand, i.e. the ability to evaporate water from the land surface (Köppen, 1900). This means that *‘we cannot tell whether a climate is moist or dry by knowing precipitation alone; we must know whether precipitation is greater or less than potential evapotranspiration’*, as Charles Warren Thornthwaite put it in a landmark paper (Thornthwaite, 1948); a statement that is indeed mirrored in present-day literature (e.g. Hulme, 1996; Cook et al., 2004; Feng and

Fu, 2013; Greve et al., 2014; Sherwood and Fu, 2014; Huang et al., 2015), and international reports (Middleton and Thomas, 1992; Millennium Ecosystem Assessment, 2005; Adeel et al., 2005). Metrics and indicators that are typically used to determine climatological dryness and changes therein are derived from this concept, e.g. the aridity index as the ratio of precipitation to potential evapotranspiration (e.g. Hulme, 1996; Greve et al., 2014; Milly and Dunne, 2016). However, in other studies dry regions are defined based on monthly or annual precipitation totals (Allan et al., 2010; Sun et al., 2012; Liu and Allan, 2013). Donat et al. (2016) defined dry regions for the PRCPTOT analysis based on low annual precipitation totals, and dry regions for the Rx1d analysis are based on moderate annual-maximum daily precipitation. Consequently, this latter definition takes a decision whether a region is dry or not based on the precipitation amount of a single day per year. Regions in northern Europe, such as parts of Scandinavia or the Netherlands, fall in the ‘dry’ class because of relatively small annual-maximum daily precipitation (Figure 3.3). Hence, different notions of what constitutes a dry region can contrast each other, resulting in regions being dry in one definition and wet in another (e.g. parts of north-eastern Europe; Figure 3.3). These variations in dryness definitions consequently induce uncertainties in the interpretation of changes in precipitation extremes and totals in the ‘world’s dry regions’. These definition-related differences can be substantial – for example, as much as 50.8 % (PRCPTOT) and 71.8 % (Rx1d) of the ‘dry grid cells’, following the respective definitions in Donat et al. (2016), are neither arid nor semi-arid (Section C.2, Figure C3), and would thus not be considered dry if a definition based on both water supply and atmospheric demand were to be used.

To clarify this issue, we test the sensitivity of the reported increases in Rx1d and PRCPTOT to the choice of dryness definition by using a variety of different dryness definitions (Figure 3.3). Hence, we evaluate trends and period increments in Rx1d and PRCPTOT in

1. regions that fall below the global 30 % quantile in HadEX2 in the respective diagnostic (Rx1d or PRCPTOT), following Donat et al. (2016);

2. dry regions (‘B-climates’) from a traditional climate classification based on temperature and precipitation (Köppen, 1900; Kottek et al., 2006);

3. dry regions as identified from an aridity-based definition of dryness (Greve et al., 2014);

4. dry and transitional regions combined from the latter definition (Greve et al., 2014).

In addition, we test uncertainties related to the temporal coverage of the dataset by relying on time series with at least 90 % coverage (cf. Donat et al., 2016) and furthermore also analyse only time series without missing values (100 % coverage).

Our results show that, if dry regions are defined based on water availability (i.e. dry regions following either Greve et al. (2014) or Köppen (1900)) and statistical artefacts are accounted for, in dry or dry and transitional regions combined, the trends reduce from the originally reported 1.6 % decade^{-1} (2.0 % decade^{-1}) to +0.2 to +0.9 % decade^{-1} (+0.0 to +1.2 % decade^{-1}) for Rx1d (PRCPTOT), respectively (see Figure 3.4). The uncertainty range reflects the choice of the aridity mask used and the temporal coverage of the time series considered (see Tables 3.2 and 3.3). Similarly, period changes between 1951–1980 and 1981–2010 would be reduced to −1.32 to +0.97 % (+0.5 to +3.8 %) as opposed to +4.85 % (+6.3 %) for Rx1d (PRCPTOT) in the original study. Although the trends remain positive, based on a two-sided Mann–Kendall test, no significant trends in Rx1d and PRCPTOT can be detected in the world's dry regions (Figure 3.4). However, the coverage of the world's arid regions with long-term observational monitoring data is rather sparse and largely confined to arid and semi-arid regions in North America and Eurasia (Figure 3.3), and thus large uncertainties remain. A few of the data gaps in HadEX2 in arid and semi-arid regions can be filled with available data from the less homogenised GHCNDEX dataset (Donat et al., 2013a). In the dry (Köppen, 1900; Greve et al., 2014) and dry-transitional regions (Greve et al., 2014) of this merged dataset, the magnitude of the trends and period changes remains largely the same for Rx1d (trends: +0.4 to +1.1 % decade^{-1}; period changes: −0.16 to +1.41 %), but with now more significant p values due to a higher data coverage (Table 3.2). For PRCPTOT, the HadEX2–GHCNDEX-merged dataset reveals on average increased and significant trends (+0.6 % to +1.9 % decade^{-1}) and period changes (+1.7 to +5.1 %). The reported results are consistent with earlier studies that report modest increases in Rx1d and PRCP-

TOT in predominantly arid and semi-arid subsidence regions based on model simulations (Kharin et al., 2007; Fischer and Knutti, 2015), and in observations for individual subtropical regions such as Australia or the Mediterranean (Westra et al., 2013; Lehmann et al., 2015). If 'the world's dry regions' are defined based on falling below a global 30 % threshold in Rx1d or PRCPTOT in the HadEX2 dataset (Donat et al., 2016), we indeed confirm robust increases in both Rx1d and PRCPTOT. Thus, the originally reported robust increases in both diagnostics are highly sensitive to the definition of a 'dry region', and appear to stem from regions with relatively moderate extreme (Rx1d) or average (PRCPTOT) precipitation, such as regions in northern Europe (Rx1d, Figure 3.3) or north-eastern Siberia (PRCPTOT, Figure 3.3).

3.4. Conclusions

Monitoring and an accurate quantification of trends in meteorological risks in a rapidly changing Earth system is a prerequisite to well-informed decision-making in the context of climate change adaptation (IPCC, 2014). In this context, short reference periods that are defined on a subset of the available dataset for normalisation or data pre-processing purposes should be avoided, as this procedure inevitably introduces biases (Zhang et al., 2005; Sippel et al., 2015b). In the present study under scrutiny, these statistical effects reduce the reported trends and period changes by up to 40 %, but the direction of the overall signal remains unchanged (i.e. increasing trends in Rx1d and PRCPTOT in regions of moderate extreme precipitation and low annual totals, respectively).

Furthermore, the definition of a 'dry region' induces considerable uncertainty in quantifying changes in Rx1d and PRCPTOT in such areas. If dryness is defined based on water supply and demand (i.e. aridity), we find much smaller trends and period increments in Rx1d and PRCPTOT, which are almost exclusively positive but in many cases insignificant (Tables 3.2 and 3.3). Hence, overall we can confirm an indication towards increases in both metrics in the world's dry regions. However, it is important to stress that many of the world's dry regions, such as large arid and semi-arid regions in Africa, the Arabian Peninsula, and partly South America, are not covered by monitoring datasets that are available at present. This fact highlights the importance of consistent, long-term monitoring efforts, data quality control, development and maintenance of long-term datasets (Alexander et al., 2006; Donat et al., 2013b,a), and also emphasises that the results reported here should be regarded as indicative only for those arid regions where data are available.

In summary, understanding and disentangling ongoing changes in precipitation characteristics in the world's dry regions remains a research priority of high relevance. In this context, our paper demonstrates that (1) data pre-processing can introduce substantial bias, and (2) trends and period changes can be sensitive to the specific choice of dryness definition that is used; therefore, we urge authors to be considerate and specific regarding both choices and to consider associated uncertainties.

TABLE 3.2.: Uncertainties regarding the definition of a 'dry region', Rx1d.

Dry Region Definition	Dataset	Ref. Period	Temporal Coverage (%)	Period Increment[1] [%]	Trend Slope [decade^{-1}]	two-sided p-value (one-sided)	Sample size
Donat et al. (2016), global 30% quantile in Rx1d	HadEX2[2]	1951-1980	90%	4.85	0.016	< 0.001 (< 0.001)	132
Donat et al. (2016), global 30% quantile in Rx1d	HadEX2[2]	1951-1980	100%	6.07	0.020	< 0.001 (< 0.001)	57
Donat et al. (2016), global 30% quantile in Rx1d	HadEX2[2]	1951-2010	90%	3.45	0.012	< 0.001 (< 0.001)	132
Donat et al. (2016), global 30% quantile in Rx1d	HadEX2[2]	1951-2010	100%	4.24	0.017	< 0.001 (< 0.001)	57
Köppen (1900), dry climates ('B-climates')	HadEX2[2]	1951-2010	90%	0.97	0.007	0.131 (0.064)	71
Köppen (1900), dry climates ('B-climates')	HadEX2[2]	1951-2010	100%	−0.14	0.006	0.336 (0.167)	36
Greve et al. (2014), dry regions	HadEX2[2]	1951-2010	90%	0.52	0.009	0.069 (0.034)	73
Greve et al. (2014), dry regions	HadEX2[2]	1951-2010	100%	−1.32	0.002	0.764 (0.380)	40
Greve et al. (2014), dry+transitional regions	HadEX2[2]	1951-2010	90%	0.36	0.005	0.195 (0.097)	118
Greve et al. (2014), dry+transitional regions	HadEX2[2]	1951-2010	100%	−0.72	0.002	0.716 (0.356)	61
Köppen (1900), dry climates ('B-climates')	HadEX2-GHCNDEX[3]	1951-2010	90%	1.41	0.011	0.058 (0.029)	127
Köppen (1900), dry climates ('B-climates')	HadEX2-GHCNDEX[3]	1951-2010	100%	0.68	0.008	0.101 (0.050)	78
Greve et al. (2014), dry regions	HadEX2-GHCNDEX[3]	1951-2010	90%	1.16	0.010	0.049 (0.024)	124
Greve et al. (2014), dry regions	HadEX2-GHCNDEX[3]	1951-2010	100%	0.00	0.005	0.243 (0.121)	80
Greve et al. (2014), dry+transitional regions	HadEX2-GHCNDEX[3]	1951-2010	90%	0.81	0.007	0.099 (0.049)	191
Greve et al. (2014), dry+transitional regions	HadEX2-GHCNDEX[3]	1951-2010	100%	−0.16	0.004	0.270 (0.134)	120

[1] Period increment denotes the change in period means between 1981-2010 vs. 1951-1980.
[2] HadEX2 is the same dataset used in the original study (Donat et al., 2016).
[3] HadEX2-GHCNDEX is a merged version, where GHCNDEX data (Donat et al., 2013a) has been added to HadEX2 data in arid regions.

TABLE 3.3.: Uncertainties regarding the definition of a ‘dry region’, PRCPTOT.

Dry Region Definition	Dataset	Ref. Period	Temporal Coverage (%)	Period Increment[1] (%)	Trend Slope ($decade^{-1}$)	two-sided p-value (one-sided)	Sample size
Donat et al. (2016), global 30% quantile in PRCPTOT	HadEX2[2]	1951-1980	90%	6.32	0.020	< 0.001 (< 0.001)	299
Donat et al. (2016), global 30% quantile in PRCPTOT	HadEX2[2]	1951-1980	100%	5.93	0.015	0.002 (0.001)	108
Donat et al. (2016), global 30% quantile in PRCPTOT	HadEX2[2]	1951-2010	90%	4.76	0.015	< 0.001 (< 0.001)	299
Donat et al. (2016), global 30% quantile in PRCPTOT	HadEX2[2]	1951-2010	100%	4.37	0.010	0.157 (0.077)	108
Köppen (1900), dry climates (‘B-climates’)	HadEX2[2]	1951-2010	90%	1.98	0.007	0.195 (0.100)	183
Köppen (1900), dry climates (‘B-climates’)	HadEX2[2]	1951-2010	100%	3.80	0.012	0.073 (0.036)	119
Greve et al. (2014), dry regions	HadEX2[2]	1951-2010	90%	1.00	0.004	0.511 (0.254)	183
Greve et al. (2014), dry regions	HadEX2[2]	1951-2010	100%	2.56	0.007	0.228 (0.113)	120
Greve et al. (2014), dry+transitional regions	HadEX2[2]	1951-2010	90%	0.51	0.000	0.985 (0.510)	296
Greve et al. (2014), dry+transitional regions	HadEX2[2]	1951-2010	100%	0.92	0.001	0.813 (0.404)	205
Köppen (1900), dry climates (‘B-climates’)	HadEX2-GHCNDEX[3]	1951-2010	90%	3.47	0.013	0.030 (0.015)	234
Köppen (1900), dry climates (‘B-climates’)	HadEX2-GHCNDEX[3]	1951-2010	100%	5.14	0.019	0.009 (0.004)	175
Greve et al. (2014), dry regions	HadEX2-GHCNDEX[3]	1951-2010	90%	2.63	0.011	0.077 (0.038)	231
Greve et al. (2014), dry regions	HadEX2-GHCNDEX[3]	1951-2010	100%	4.20	0.017	0.024 (0.012)	170
Greve et al. (2014), dry+transitional regions	HadEX2-GHCNDEX[3]	1951-2010	90%	1.67	0.006	0.200 (0.099)	356
Greve et al. (2014), dry+transitional regions	HadEX2-GHCNDEX[3]	1951-2010	100%	2.47	0.009	0.084 (0.041)	275

[1] Period increment denotes the change in period means between 1981-2010 vs. 1951-1980.
[2] HadEX2 is the same dataset used in the original study (Donat et al., 2016).
[3] HadEX2-GHCNDEX is a merged version, where GHCNDEX data (Donat et al., 2013a) has been added to HadEX2 data in arid regions.

4. Combining large model ensembles with extreme value statistics to improve attribution statements of rare events[1,2]

Abstract

Gaining a better understanding of rare weather events is a major research challenge and of crucial relevance for societal preparedness in the face of a changing climate. The main focus of previous studies has been to apply a range of relatively distinct methodologies to constrain changes in the odds of those events, including both parametric statistics (extreme value theory, EVT) and empirical approaches based on large numbers of dynamical model simulations.

In this study, the applicability of EVT in the context of probabilistic event attribution is explored and potential combinations of both methodological frameworks are investigated. In particular, this study compares empirical return time estimates derived from a large model ensemble with parametric inferences from the same data set in order to assess whether statements made about events in the tails are similar. Our analysis is illustrated using a case study of cold extremes and heavy rainfall in winter 2013/14 in Europe (focussing on two regions: North-West Russia and the Iberian Peninsula) for a present-day (including 'anthropogenic' influences) and an alternative 'non-industrial' climate scenario. We show that parametric inferences made about rare 'extremes' can differ considerably from estimates based on large ensembles. This highlights the importance of an ap-

[1]This chapter is published as Sippel, S., D. Mitchell, M. T. Black, A. J. Dittus, L. Harrington, N. Schaller, and F. E. L. Otto. 2015. *Weather and Climate Extremes* **9**, 25–35. doi:10.1016/j.wace.2015.06.004.

[2]Supplementary Online Material (SOM) that provides additional information but that is not necessary for understanding the scientific content of this Chapter is available under `http://www.sciencedirect.com/science/article/pii/S2212094715300050`

propriate choice of block and sample sizes for parametric inferences of the tails of climatological variables. For example, inferences based on annual extremes of daily variables are often insufficient to characterise rare events due to small sample sizes (i.e. with return periods > 100 *years). Hence, we illustrate how a combination of large numerical simulations with EVT might enable a more objective assessment of EVT parameters, such as block and sample size, for any given variable, region and return period of interest. By combining both methodologies, our case study reveals that a distinct warming of cold extremes in winter has occured throughout Europe in the 'anthropogenic' relative to the non-industrial climates for given sea surface temperatures in winter 2013/14. Moreover, heavy rainfall events have become significantly more frequent and more pronounced in North and North-East Europe, while other regions demonstrate no discernible changes. In conclusion, our study shows that EVT and empirical estimates based on numerical simulations can indeed be used to productively inform each other, for instance to derive appropriate EVT parameters for short observational time series. Further, the combination of ensemble simulations with EVT allows to significantly reduce the number of simulations needed for statements about the tails.*

4.1. Introduction

It is a major scientific challenge to better understand extreme meteorological events and potential changes in the odds of their occurence in a warming climate (IPCC, 2012; Zhang et al., 2014). This is due to a number of reasons, including limitations of the observational record to capture rare extreme events, and issues of data availability and quality. Moreover, structural and parametric model uncertainties, as well as the proverbial chaotic nature of weather (Lorenz, 1963) hinder any straightforward attribution of causality between climatic drivers and any particular extreme weather event.

To overcome these difficulties, many scientific studies use either one of the following approaches:

First, *extreme value theory* (EVT) has been developed to provide a means to model the tails of statistical distributions based on mathematical theory (Coles et al., 2001). Such an analysis allows statistical statements to be made based on parametric extreme value distributions (see Wigley, 2009, for illustrative examples). For example, scientific assessments have been made to investigate trends in temperature and precipitation extremes in the 21st century in atmosphere-ocean coupled models (Kharin and Zwiers, 2000; Kharin et al., 2007, 2013), allowing the estimation of return levels and their associated statistical uncertainties. Further illustrative applications of the (univariate) EVT framework elucidate causes for geophysical extremes, such as the connection between atmospheric modes of variability and cold extremes (Sillmann et al., 2011). However although EVT is increasingly used in climatological studies to constrain the odds of rare events (Katz, 2010), including extensions to account for non-stationarity, multivariate and spatial extremes (see Ghil et al., 2011, for a review), Katz et al. (2013) argues that its full potential has not yet been tapped for many geophysical applications.

Second, an alternative approach to improve the understanding of extremes and their changing odds in a non-stationary climate has been to deploy very large ensembles of dynamical models, namely *probabilistic event attribution* (PEA, Stone and Allen, 2005; Allen, 2003). This methodology is used extensively to sample rare events and subsequently estimate their probabilities under different climate forcing scenarios (Stott et al., 2004; Otto et al., 2012; Massey et al., 2015). The latter often serves to estimate the anthropogenic contribution ('fraction of attributable risk') to changes in the meteorological risk of present-day weather and climate extremes (Allen, 2003; Stott et al., 2013; Bindoff et al., 2013; Christidis et al., 2013). Importantly, an assessment of this type addresses the odds of specific extreme weather events - often those that had happened in a particular year such as droughts, heat waves or cold spells (Herring et al., 2014). Notable extensions to the PEA methodology include the attempt to account for more impact-related variables, for instance through a coupling with hydrological models to assess floods (Pall et al., 2011). Nonetheless, PEA assessments are typically based on rather data-intensive empirical estimates of return times, and rely to a large extent on dynamical model simulations.

Our study addresses the following research questions:

1. Is the statistical framework of EVT applicable in the context of a probabilistic assessment of extreme events? Accordingly, can both methodological frameworks be productively combined to inform each other?

2. Using a *combined* methodology, how have meteorological extremes at daily time scales in the European winter of 2013/14 changed relative to a pre-industrial climate?

Based on our first research question, we envision an application in which both methodological frameworks could inform each other in order to a) derive insights about appropriate parameter choices (i.e. required sample and block sizes) for the application of statistical models based on EVT for the meteorological characteristics of any variable or region of interest; and b) given informed parameter choices, how many numerical simulations are actually needed to estimate a given 'target return period' to a satisfactory degree of accuracy?

Hence, our study details a joint assessment of both methodologies and evaluates whether statements made about the tails of meteorological variables such as temperature and precipitation are comparable. This methodological comparison might serve as a starting point to reconcile the two statistical frameworks for climatological applications, i.e. to inform each other about relevant parameter choices (EVT) or the number of samples needed to estimate a specific return level. To illustrate this comparison and to address the second research question, a large ensemble of atmosphere-only regional climate simulations for the European 2013/14 winter season is investigated as a case study along with a 'non-industrial' climate scenario of winter 2013/14 (i.e. with anthropogenic forcings removed (Schaller et al., 2014), see Section 4.2).

The particular season of interest, winter 2013/14 in Europe, provides an interesting case study, because it came along with exceptionally mild temperatures, severe storm depressions, both winter dryness and heavy precipitation on regional to sub-continental scales. Significant but diverse societal impacts were associated with those events, for instance exceptionally early vegetation greening and a reduction of fossil fuel consumption for heating due to the absence of severe frosts in some regions[3]. Seasonal temperatures ranked among the highest ever recorded in a range of countries according to national weather services (e.g. Austria, Denmark, France, Germany, the Netherlands, Norway, Poland, Slovakia, Switzerland, and the UK, e.g. Figure 4.1, Deutscher Wetterdienst (2014)). When it comes to seasonal rainfall anomalies, a remarkable east-west divide persisted over most of

[3] http://www.pecad.fas.usda.gov/highlights/2014/03/EU_12march2014/

the winter, where central and south-eastern parts of the continent received exceptionally low rainfall, whereas its most western stretches, such as Ireland and the UK, experienced a record wet season (Huntingford et al. (2014); Figure 4.1). These remarkable patterns resulted from a synoptic situation with many storm depressions that moved along the English Channel, over the British Isles and into the North Sea, hence advecting warm air into Central and East European regions, and causing rainfall and severe winds in Britain and along the Atlantic coast. This synoptic situation is also reflected by seasonal geopotential height anomalies (Figure 4.1), which were strongly negative over the North Atlantic and the British Isles, whereas positive anomalies prevailed over Eastern Europe (see Huntingford et al. (2014) for a more detailed discussion).

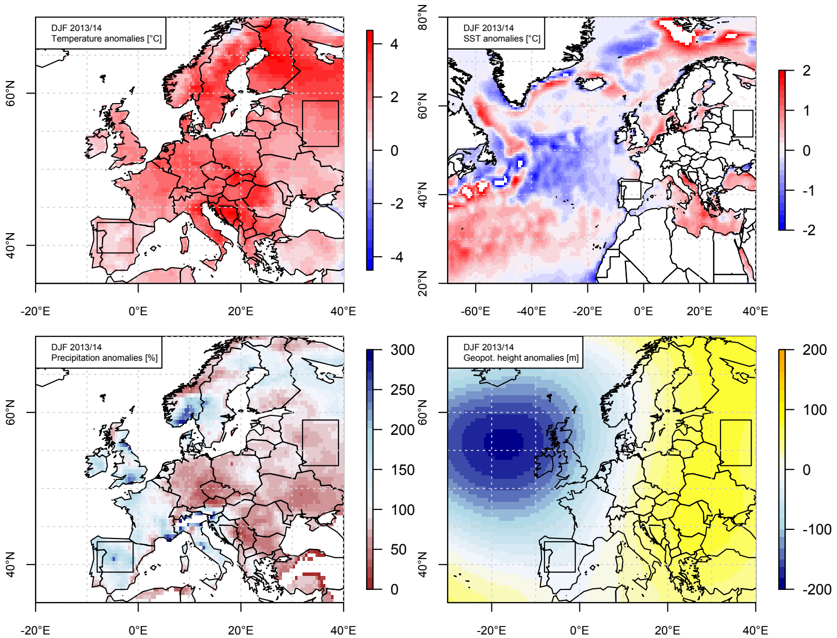


FIGURE 4.1.: Synoptic analysis of winter 2013/14 in Europe: Seasonal temperature anomalies (top left), SST anomalies (top right), anomalies in cumulative rainfall (bottom left), and geopotential height anomalies (bottom right). Temperature and precipitation data were taken from E-OBS, SSTs and geopotential height anomalies were calculated from ERA-Interim (reference period: 1981-2010). The study regions over Spain and Russia are drawn as rectangular boxes.

This study's analysis focuses on cold temperature and heavy rainfall extremes, which allows to state how the odds of occurence of extremes in these two variables have changed between a 'non-industrial' climate and the contemporary winter climate in 2013/14. These two variables provide a good case study, because we expect temperature to be relatively spatially coherent, and precipitation to be somewhat noisier both in space and time. We illustrate our methodological approach as well as the attribution analysis for two spatially averaged regions, North-West Russia and the Iberian Peninsula, as well as for the entire European model domain.

In Section 4.2, we describe the experimental setup, evaluate and bias-adjust the regional climate model and outline the statistical methodology to estimate return times. In Section 4.3, we first outline the results of the methodological comparison (EVT vs. empirical return time estimates), and discuss related issues such as parameter choices for a potential combination of both methodologies. Second, the illustrative attribution case study of winter minimum temperatures and precipitation is presented. Lastly, we draw some conclusions about the applicability of EVT based return time estimates in the context of probabilistic event attribution (Section 4.4).

4.2. Material and methods

Model structure and experiment setup In this study, we analyse large ensemble simulations of the HadAM3P atmosphere-only, global circulation model with an embedded, identically formulated regional model for Europe (HadRM3P), which has been used extensively elsewhere (Jones, 2004; Massey et al., 2015). The global (nested regional) models are run with a spatial resolution of 1.875°x1.25° (0.44°x0.44°) on a rotated grid identical to the EURO-CORDEX region[4], with 19 vertical levels and a temporal resolution of 15 (5) minutes (Massey et al., 2015). The model is based on the atmospheric component of the HadCM3 general circulation model (see Pope et al. (2000) for a full description) with improvements with respect to the calculation of clouds and convection, and a more realistic coupling of vegetated surfaces with the soil (Massey et al., 2015).

[4] http://www.euro-cordex.net/About-Euro-Cordex.1864.0.html

Since atmosphere-only simulations were conducted, observed sea surface temperatures (SSTs) and sea ice fractions for the observed period (DJF 2013/2014) are provided to the model from the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) dataset (Stark et al., 2007; Donlon et al., 2012). Further model drivers include the observed atmospheric composition (CO_2, CH_4, N_2O, halocarbons and ozone), natural and anthropogenic emissions of different sulfur species, and solar anomalies (see Massey et al. (2015) for a more detailed description and evaluation of the modelling framework). Initial conditions are perturbed in the global circulation model on 1st December for each ensemble member (ibid.).

As observed SST patterns from 'the world that might have been' (i.e. the 'non-industrial' scenario) in the absence of anthropogenic emissions are not known precisely, estimates are made using some of the state-of-the-art coupled ocean-atmosphere models taken from the Coupled Model Intercomparison Project, phase 5 (CMIP5, Taylor et al. (2012)). Eleven of these models have run 'natural forcings only' simulations of the historical climate, and these are subtracted from the 'all forcings' simulations to obtain an estimate for the change in the SSTs (hereafter, delta SSTs). The differencing is performed on climatological monthly means over the last decade available, i.e. 1996-2005. The delta SSTs are then used to change the observed SSTs accordingly. To sample uncertainty, we use these different CMIP5 models which cover the main modelling groups from around the world (see Schaller et al., 2014, for details). All non-anthropogenic forcings such as aerosols, volcanoes and the solar cycle are kept constant in both scenarios.

The large model ensemble investigated in this paper is derived through the weather@home framework, in which citizen scientists donate idle computer time in order to perform computationally intensive calculations in a distributed manner. This approach provided model ensemble simulations for 13260 DJF periods in an industrial world and 22129 for the non-industrial case, using the variety of different SST reconstructions. Data preprocessing consists of regridding the regional model ensembles to a regular 0.5° grid over Europe, using a second order conservative remapping scheme (Jones, 1999). Subsequently, a meteorological sanity check is conducted, in which all ensemble members with meteorologically implausible values are removed, before we validate our model and analyse the ensemble's statistics of extremes (see below). We derive both 1-day spatial averages over selected regions and 5-day grid-based averages for minimum temperatures and heavy precipitation, where both aggregation steps are computed on the original gridded time series. Two regions were chosen to represent different European climates with predominantly maritime/Mediterranean (Spain) and continental (North-West Russia) influences (Table 4.1). The grid-based European-scale analysis is noisier due to a lower level of aggregation, but nevertheless provides valuable spatially explicit details. Due to model spinup time, the first two weeks of December are disregarded, after which it was checked that no remainder spinup effects are detectable.

Region	Eastern boundary (°E)	Western boundary (°E)	Southern boundary (°N)	Northern boundary (°N)
Spain	-8	-1	39	43
NW Russia	32	39	53	59

TABLE 4.1.: Regions used in this study and their geographical boundaries.

Model validation and bias adjustment A high-quality grid-based European land-only observational data set in 0.5° resolution (E-OBS, version 10.0, Haylock et al., 2008) is used in order to quantify biases in simulated meteorological variables and to conduct a simple synoptic assessment for winter 2013/14 (Figure 4.1). Since our ensemble is based on the SST patterns of one winter season, an assessment of model performance would not be representative based on the ensemble alone. Hence, we use 50 randomly chosen ensemble members per year (i.e. 1300 model years in daily resolution) for the winter seasons from 1986-2010 from an identical model setup (Massey et al., 2015) for the purpose of validation. Differences in statistical distributions are assessed graphically by quantile-quantile plots. The spread of the ensemble is illustrated similarly to Massey et al. (2015) by appending randomly chosen ensemble members without replacement in order to derive 50 winter time series for each of the years 1986-2010 (Figure 4.1).

The model's winter simulations of daily temperature show relatively good agreement with the distribution of daily minimum temperatures in E-OBS in both regions, although with a slight cold bias over NW Russia. For the whole European domain, larger biases are observed in Scandinavia, and towards the southern margins of the regional model domain (Supplementary Online Figure 2). Nonetheless, it can be noted that the regional model performs better in simulating temperatures in winter as compared to summer (Massey et al., 2015). Hence, we conclude that our model simulates temperatures to a reasonable degree, and this also holds for percentiles relatively far away from the mean (Figure. 4.1, Supplementary Online Figure 1).

Precipitation simulations do not always agree favourably with observations. Considerable wet biases towards the upper tails of the distributions of daily rainfall over the two regions remain, as well as for most grid cells throughout Europe (Supplementary Online Figure 1). Here, we use a very simple bias adjustment methodology to account for this bias. Due to the obvious positivity constraint, an additive correction of biases, which is often applied to climatic variables such as temperature (Hempel et al., 2013; Sippel and Otto, 2014), is not feasible for precipitation. Hence, we determine a multiplicative correction factor similar to Hempel et al. (2013), which quantifies biases in the 97.5th percentile:

$$c = \frac{OBS_{97.5th}}{MOD_{97.5th}} \tag{4.1}$$

Subsequently, daily rainfall values are scaled by c, which removes some of the biases in the high percentiles. Although using a single percentile is a somewhat subjective choice, we argue that it is relatively robust with respect to the observations, since in the period used for model validation (DJF 1986-2010), the 97.5th percentile corresponds approximately to the 50th largest value, hence a relatively robust sample. This simple multiplicative adjustment yields a better match of simulated and observed rainfall amounts also in higher quantiles, without any invasive changes to the distribution. Importantly however, scaling the absolute values with an adjustment factor does not affect any relative changes between fitted extreme value distributions.

Further, it is important to note that an acceptable simulation of daily precipitation statistics does not warrant satisfying simulation at monthly or seasonal time scales. For an evaluation and discussion for model performance at monthly time scales, we refer the interested reader to Massey et al. (2015). Moreover, while we acknowledge that the resolution of the regional model is too coarse to resolve local convection or thunderstorm-related activity, Supplementary Online Figure 1 demonstrates that the distribution of daily rainfall events in the model agrees broadly with the observations for both regions, including its tails.

Statistical estimates of return periods The primary objective of this study is to compare statistical inferences for the tails of meteorological variables based on EVT with empirical return time estimates. The ensemble simulations are conducted for one season only (DJF 2013/14 in a 'natural' and 'anthropogenic' scenario), hence stationarity for the EVT based estimates of the tails is assumed. Further, we fit generalised extreme value (GEV) distributions of the form (Coles et al., 2001)

$$G(z) = exp(-[1 + \xi\frac{z-\mu}{\sigma}]^{-(\frac{1}{\xi})}), \tag{4.2}$$

to a sample of 1-day (5-day) minimum temperatures and maximum cumulative rainfall events for each simulated winter season for each area-averaged region (grid cell). Here, μ, σ and ξ denote the location, scale and shape parameter of the GEV distribution. Unless otherwise stated, confidence intervals representing 5-95$\%$ parametric uncertainty are given based on the normality of the GEV parameter estimates (Coles et al., 2001). To address the influence of GEV parameter choice (block and sample size) on the return time estimates (Section 4.3), we resample the large ensemble to derive different block and sample sizes for various return time estimates. This procedure is iteratively repeated for each parameter combination in order to derive resampling based 5$\%$-95$\%$ confidence intervals for return time estimates that are comparable to the empirical estimates.

For the analysis of rare winter extremes, a resampling strategy is used in order to avoid biases associated with an extrapolation from 1-yr extremes to several hundred year return level extremes (see Section 4.3), which might also entail a very different dynamical structure of the atmospheric circulation in the real world. Therefore, 10-yr block extremes are drawn from the large sample by a random selection of ten ensemble members, from which only the most extreme value is retained. This procedure is repeated 200 times (for both regions and for each CMIP5 model's SST reconstruction) to derive a statistical distribution of 10-year block extremes, which is subsequently used to fit a GEV distribution as specified above.

Throughout our analysis, a Generalised Maximum Likelihood Estimation (GMLE) approach is used for fitting the parametric model to the data (Martins and Stedinger, 2000), which is conducted using the extRemes software package (Gilleland and Katz, 2011). We also tested the GEV parameter estimation using the L-moment and MLE methods: these were found to yield estimates very similar to the GMLE method that we employ here. All statistical analysis is performed in the R statistical environment (R Development Core Team, 2013) using the add-on packages 'boot' and 'ADGofTest'.

Empirical return time estimates are constructed by plotting the sorted values of the ensemble against its rank. To assess uncertainty of this empirical estimate, we derive bootstrapped uncertainty intervals (5%-95%) by resampling ($n = 5000$ ensemble members, $R = 1000$ times).

Evaluation of fitted extreme value distributions The parametric fits are evaluated in a three-fold approach:

First, we use adjusted mean residual life plots (Coles et al., 2001) in order to test whether the exceedance of any threshold u yields an approximately linear scaling of the 'residual means' (i.e. the average of the values exceeding the threshold u). This concept is frequently used to determine an appropriate threshold for peak over threshold models with a prior declustering of extremes. It can be shown that the residual means follow a linear function of the threshold, if the peak over threshold model is appropriate (Coles et al. (2001), p. 79). Here, this idea is slightly modified, and we plot the 'mean residual life' of the seasonal block maxima, thus it could be seen similarly to a seasonal declustering approach (i.e. assuming that any two extreme events in one season are not independent). Present non-linearities in these plots might indicate that extreme events are subject to different physical/dynamical climatic regimes, and will be further discussed/evaluated below.

Second, each fitted GEV (both regional and grid cell based) is tested for its goodness of fit using a parametric Anderson-Darling (AD) test based on a significance level of $\alpha = 5\%$. We chose the AD test over a Kolmogorov-Smirnov (KS) test used in earlier studies (Kharin et al., 2007), because it is more sensitive to the tails of the distribution by implementing a weight function instead of a maximum distance approach such as the KS test.

Lastly, in order to evaluate deviations of the fitted GEV distributions from the empirical large ensemble for rare events in the tails (Section 4.3), we adopt a somewhat ad-hoc but practically useful definition of 'biases' (see Figure 4.2 and associated discussion): Since our focus is on 'rare events', we determine the maximum absolute difference in return levels in the interval of 100 to 1000 years (i.e. 99th to 99.9th percentile) between the fitted GEV's and the empirical return levels of the large ensemble, using monotonic Hermite spline interpolation to derive a continuous curve for the latter. To compare the biases in GEV fits from the empirical ensemble with the 'expected biases' inherent in any GEV model for a given block size (Section 4.3), we simulate a large number of random values from the fitted extreme value distributions for each region. Subsequently, we determine the distance ('bias', as defined above) between GEV fits from this data using each block size of interest from a 'large empirical GEV sample' ($n = 15000$). Hence, these artificial simulations mimic the comparison between the empirical ensemble and GEV fits with different block sizes. The uncertainty of an empirical estimate of the tail is tested by resampling from a known GEV model (Supplementary Online Figure 4), the variance of which becomes large for very high return periods.

4.3. Results and discussion

In this section, we first test a combination of stationary EVT analysis with a large ensemble of numerical simulations and present a systematic evaluation of the parameter choices in EVT-based assessments regarding its effects on return time estimates for meteorological variables (Section 4.3.1). Subsequently, we analyse changes in cold temperature and heavy precipitation extremes in winter 2013/14 relative to a pre-industrial scenario in the large ensemble simulation using extreme value theory (Section 4.3.2).

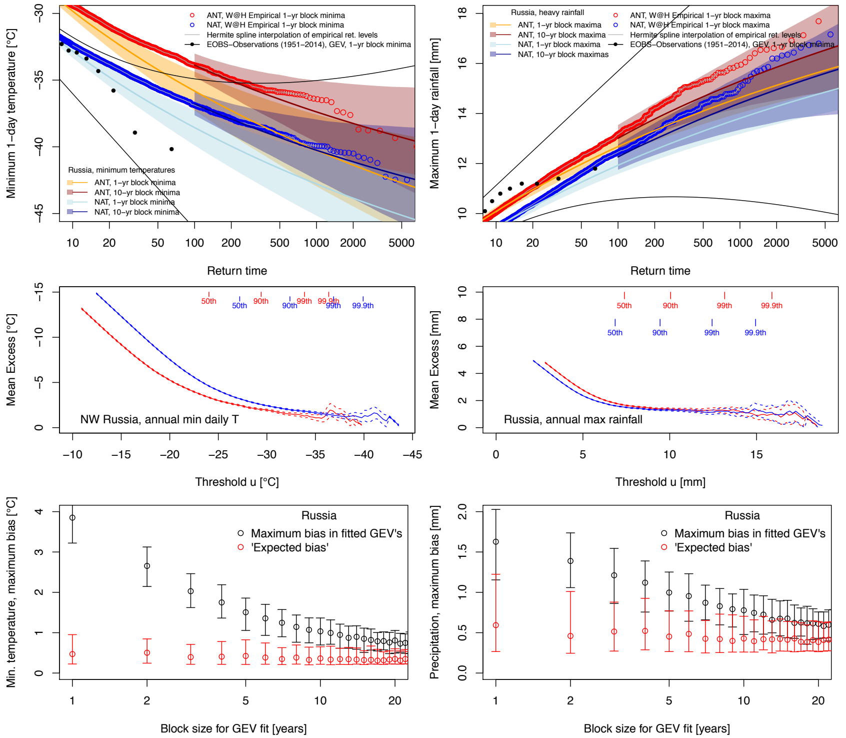


FIGURE 4.2.: Return level plots of GEV distributions fitted to 1-yr (n=1000) and 10-yr (n=100) block extremes of daily minimum temperatures (top left) and heavy precipitation (top right). Coloured dots reflect the 'empirical' large ensemble, observations are denoted in black (EOBS, 1951-2014). Shading represents parametric uncertainty as taken from the fitted generalised extreme value distributions. (Middle panels) Mean residual life plot of annual block minima of daily temperatures (left) and block maxima of daily rainfall (right) for NW Russia. (Bottom) Biases in rare events (100 to 1000 year return periods, as defined in Section 4.2) estimated from GEV distributions as a function of block size for the model ensemble and as would be expected by sampling from 'ideal' GEV distributions.

4.3.1. Combining extreme value analysis with large ensemble simulations

A comparison of the fitted GEV distributions based on resampled sub-ensembles with the empirical estimate of the tail for the NW Russia region is presented in

Figure 4.2 (top), including a stationary GEV fit to the E-OBS observations for illustrative purposes only (1951-2014, black dots and line)[5].

The methodological comparison reveals that GEV-based inferences with large block sizes (e.g. 10-yr return periods, Figure 4.2 (top), dark-blue / dark-red line and shading) agree well with the empirical estimate (circles). However, inferences made for shorter return periods (e.g. 1-yr events: orange / light blue) overestimate (minimum temperatures) or underestimate (maximum rainfall) return levels of rare events (e.g. 100+ year return levels). This analysis is presented in for the NW Russia region, and occurs similarly over Spain (Supplementary Online Figure 3), although less pronounced. These differences are important to consider, because a relatively large proportion of the GEV's fitted to resampled sub-ensembles (n = 1000, annual block extremes) is not rejected by a statistical Goodness-of-Fit test[6], and could thus be misinterpreted if only a small ensemble were available. However, these differences in the inferences about the tails can be readily detected in the mean residual life plots, for example in the NW Russia region (Figure 4.2, middle) with a non-linear breakpoint approximately around the median (marked as 50th percentile, corresponding to 2-yr return events). Hence, extreme value statistics of seasonal minimum daily temperatures or precipitation might not be rare enough in order to satisfactorily constrain events that are located far in the tails. This could potentially lead to notorious biases in statistical models, which are most pronounced for large return periods (i.e. 99th to 99.9th percentile in this case) if the chosen block size is too small. Comparing these biases with biases in the tails for independent and identically GEV-distributed artificial data (see Section 4.2) for a detailed description of the resampling strategy to obtain these 'expected biases in the tails') for any given block size shows that for large enough block sizes these biases are reduced (Figure 4.2). Hence, although the ultimate reason for non-adequate statistical model fits for rare events are limited sample and block sizes (Fisher and Tippett, 1928; Coles et al., 2001), characteristics of climatological variables such as serial correlation, climatic vari-

[5]However, it should be kept in mind that the observations are based on a 63-year period, including potential non-stationarities and cover a variety of synoptic conditions, whereas the model ensemble is run conditional on 2013/14 SST's.

[6]Based on the AD-test, the proportion of the null hypothesis *not rejected* is for the anthropogenic (natural) ensemble: 41% (30%) NW Russia, 100% (100%) Spain (minimum temperatures), and 99% (98%) NW Russia, 99% (100%) Spain (heavy rainfall).

ability and noise, or potential dynamic regime changes under extreme conditions might considerably amplify these deviations.

The illustrative example highlights that the choice of block size is critical in climatological applications of EVT. To further and more systematically investigate this issue, we conducted a range of resampling experiments to assess the influence of parameter choice on an EVT-based estimation of climatological events in the extreme tails. Those parameter choices are inherently a trade-off between bias (short block size) and variance (due to small sample sizes for large blocks Coles et al., 2001), which is illustrated in Figure 4.3 for two different return times (20 and 1000-years) in NW Russia. We note that from a practical perspective, for example for the analysis of relatively short observational time series, the choice of block size depends not only on the available sample size and climatological variable of interest, but also on the ‘target return time’ upon which a statement should be made (Figure 4.3, see also tabulated values in Supplementary Online Table 1 and 2). To this end, it is interesting to note that these biases require careful consideration if, for example, statistical models are derived based on annual extremes of daily variables, which is widely being done (see for example: Coles et al., 2001). On the other hand however, GEV-based inferences with larger block sizes allow to derive very consistent statements for high return intervals, for which a reduced number of ensemble simulations are already sufficient (e.g. compare GEV-based inference with an ensemble of size $n = 1000$ in Figure 4.3 with the empirical estimate, $n = 13260$).

At this point, a couple of cautionary remarks might be appropriate. First, it should be noted that in this paper we investigate the simplest case of an application of EVT: Daily extremes determined from seasonal blocks under stationary conditions (i.e. Winter 2013/14 under anthropogenic *or* natural forcing conditions). Hence, it should be stressed that EVT can also be applied under non-stationary conditions (Kharin and Zwiers, 2000; Kharin et al., 2007, 2013) and with covariates accounting for additional information (see for example Sillmann et al., 2011). Furthermore, peak-over-threshold models constitute an important alternative to modelling block maxima with GEV’s (Coles et al., 2001); a detailed investigation of this in terms of informed parameter choices based on ensemble simulations could be a topic for future study.

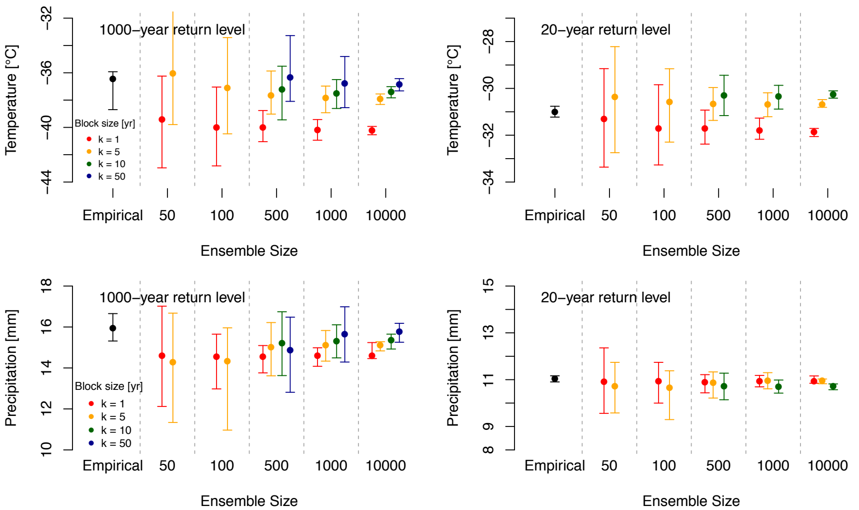


FIGURE 4.3.: Illustration of bias-variance trade-offs for estimates of 1000-year (left) and 20-year (right) return time events of daily minimum temperature (top) and maximum precipitation (bottom) in the Russia region for the bootstrapped empirical ensemble and GEV's fitted to resampled ensembles using various sample and block sizes.

Second, as we are concerned here about the statistics of rare events, and thus the dynamical structure of such events is not investigated. In Europe, such rare events might be related to relevant modes of atmospheric variability, such as for example the North Atlantic Oscillation (NAO) (Sillmann et al., 2011). Therefore, our analysis and estimation of return periods of extremes is conditional on sea surface temperature patterns that were present in winter 2013/14 in the Euro-Atlantic region with the NAO being in its positive phase (Huntingford et al., 2014).

Third, it should be pointed out that an analysis of rare events is inherently uncertain. In this paper, we are addressing statistical (Section 4.3.1) and scenario reconstruction uncertainties (Section 4.3.2). Hence, possibly large uncertainties that might stem from the models' (imperfect) structure or parametrisation are not examined here, although an attempt was made to implicitly account for such issues using the empirical bias correction for precipitation (Section 4.2).

To summarise, our analysis reveals that seasonal block extremes in an ensemble of regional model simulations of daily meteorological variables might not be robust enough to infer statistical statements on the odds of particularly rare events of both temperature and precipitation. A resampling scheme is shown to improve the fits to the tails based on larger than annual block sizes. Therefore, the combination of a large number of dynamical model simulations with statistical extreme value models might enable a more informed selection of parameter choices for EVT-based inferences. In return, EVT-based estimates might point at the number of numerical simulations needed to adequately constrain a given return period of interest (Figure 4.3). Hence, we conclude that for climatological applications both methodologies might benefit from a statistical setup in which EVT and large numerical simulations inform each other, for example to choose EVT parameters for the analysis of relatively short observational time series.

4.3.2. The anthropogenic influence on European minimum temperatures and precipitation in winter 2013/14

In this subsection, we present and discuss how climatic changes between the counterfactual scenario and the present might have altered the odds of cold extremes and heavy precipitation as an application of the extreme value analysis outlined above. We also illustrate for two regions how uncertainties in the reconstruction of a counterfactual past might induce uncertainties in attribution statements. Finally, we discuss our results in the context of changes in extremes throughout Europe.

Temperatures In a winter season such as DJF2013/14 in Europe, minimum temperatures have warmed significantly and unambiguously in both study regions (Figure 4.4, Supplementary Online Figure 6) and throughout Europe (Supplementary Online Figure 7). For example, the location parameter of GEV distributions fitted to 10-yr resampled minimum temperatures in NW Russia has shifted significantly under all scenarios (Figure 4.4). However, the reconstruction of a 'non-industrial world' scenario induces considerable uncertainties, with a warming of roughly two and four degrees at the lower and upper end, respectively, of the CMIP5 models used for reconstruction (Figure 4.4). Hence, scenario uncertainties are larger in magnitude than statistical uncertainties resulting from fitting statistical models in this type of study. Furthermore, the decreasing odds of extremely cold temperatures in the two regions studied in this paper seem to a very large proportion caused by a shift in the location parameter of the GEV, rather than by changes in the scale or shape of the distribution. In fact, none of the different SST reconstructions shows a significant change in scale or shape in any region under study (not shown), and computing GEV parameters over each grid cell of the European model domain yields only minor and largely non-significant changes in the shape and scale parameters of seasonal cold extremes. This finding indicates that the year-to-year variability of seasonal cold extremes (around the shifting mean) has not changed markedly in our model, though the interpretation of individual GEV parameters is to be made with caution (Gilleland, pers. comm.).

Nonetheless, testing different assumptions about potential chages in the scale and shape of the tails is a highly topical issue in climatology - not least because recent findings point at a decreasing temperature variability at the sub-seasonal scale in northern latitudes (Screen, 2014). Consequently, we further investigate this issue in our model ensemble with a focus on the tails. To do so, we compare the present-day warming relative to the pre-industrial scenario in winter maximum temperatures with the warming in the coldest winter temperatures (i.e. a 'differential warming of winter temperature extremes' is defined as the difference between the warming in the warmest and coldest winter temperatures expressed through 100-year return levels). To this end, we find a clear, spatially coherent and widely significant pattern (Figure 4.5): In large areas of North and Central Europe, cold extremes have warmed considerably stronger than warm extremes. Only in the Mediterranean region and towards the eastern edges of our model domain this pattern is not as clearly pronounced. This finding is qualitatively consistent with previous studies that have shown that daily minimum temperatures (nights) are warming faster than maximum temperatures (days) in observations (Alexander et al., 2006; Donat et al., 2013b) and that sub-seasonal temperature variability in northern latitudes is decreasing (Screen, 2014), both of which might contribute to the differential warming seen here. Mechanisms behind the day-night assymetry might indeed include stronger night-time effects of increased greenhouse gas forcing, whereas changes in sub-seasonal variability in northern latitudes might be driven by the Arctic amplification, i.e. temperatures of northerly winds might have warmed faster than southerly winds over the last decades (Screen, 2014). Disentangling these effects is not the focus of this paper, but would provide an interesting topic for further study.

In brief, our analysis suggests that cold and warm temperatures extremes have warmed considerably since pre-industrial times, but the upper and lower (extreme) tail might indeed warm at different rates. However, our present analysis does not show any evidence that the year-to-year variability of seasonal cold temperature extremes has changed.

Precipitation When it comes to wintertime heavy rainfall events, changes between the 'non-industrial' (NAT) and anthropogenic (ANT) scenarios are less pronounced and vary among regions and the CMIP5 models used for reconstructing the SST patterns. We find a significant shift towards stronger heavy precipitation events in NW Russia (Figure 4.4, bottom), whereas in Spain no significant overall changes are shown by the model (Supplementary Online Figure 6). Moreover, the counterfactual world reconstructions clearly show that scenario uncertainty is large when it comes to heavy rainfall:
In NW Russia all except one SST reconstructions for the 'non-industrial' scenario lead to a significant increase in the location parameter, however results for Spain show the sign of the location parameter to differ between SST estimates leading to an overall small but non-significant increase in the location parameter. Like the results for temperature, we also observe that the scale and shape parameters are not changing significantly across the studied regions.

To understand further, we derived GEV fits for heavy precipitation for each grid cell of the European model domain (Supplementary Online Figure 8). Most attribution studies conducted to date have been looking at regional averages, mainly because spatial (or temporal) aggregation reduces the level of noise. Although we acknowledge that this type of spatially explicit analysis presented here might involve considerable uncertainties, particularly as local features such as processes on a sub-grid cell scale might not be well-represented in the model, we argue that Supplementary Online Figure 8 allows to identify European regions that show

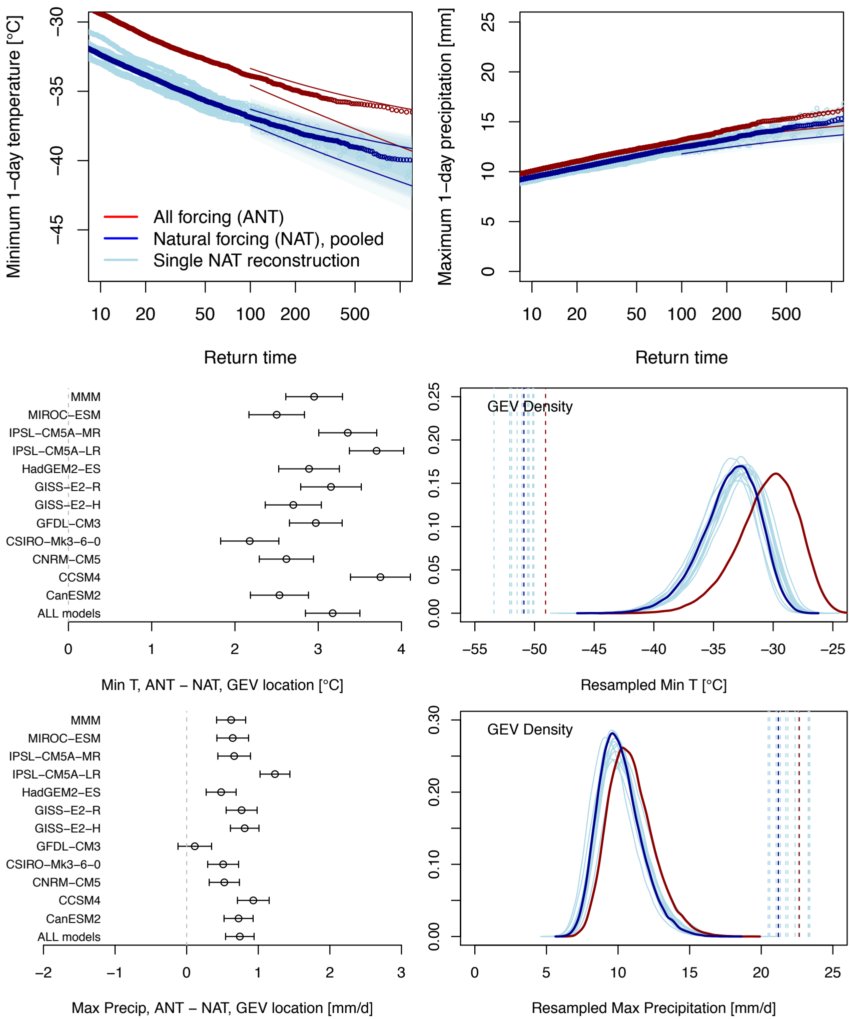


FIGURE 4.4.: (Top panel) Return periods of seasonal block minimum temperatures (left) and heavy rainfall (right) in NW Russia. (Middle panel) Warming in the GEV's location parameter (left) and densities of the fitted GEV distributions for 10-yr resampled block minimum 1-d temperatures. (Bottom) Changes in the location parameter (left) and GEV density (right) of 10-yr resampled heavy precipitation events. Changes in the GEV's location parameter (middle and bottom left panel) and single SST reconstructions (turquoise lines) are given for the 11 different natural SST estimates, and the CMIP5 models from which the estimates were obtained are listed in Schaller et al. (2014).

a spatially coherent signal of human-induced changes in 100-year return levels of daily rainfall. Whilst the overall pixel-based signal is much noisier and does not point at strongly pronounced changes in extreme winter rainfall in Central or Southern Europe, we are able to identify regions in North and North-East Europe that exhibit a spatially coherent signal of increasing 100-yr return levels (Figure 4.5). Here again, those changes can be attributed to a shift in the location parameter of the GEV, rather than changes in scale or shape (Supplementary Online Figure 8). In conclusion, we find clear indications that winter rainfall extremes are changing in parts of North Europe, whilst in southern regions, particularly in the Mediterranean no clear statement can be made at present. Although the mechanisms behind intensified extreme rainfall are still debated (O'Gorman and Schneider, 2009), they can be conceptualised as a subtle interplay between thermodynamical effects (i.e. the amount of moisture held within a fixed volume of air, described by the well-known Clausius-Clapeyron relationship, e.g. Held and Soden (2006)) and large-scale atmospheric dynamics in a warming climate (Emori and Brown, 2005). Nonetheless, our results in European regions agree qualitatively well with previous findings of intensified daily rainfall in model simulations for the mid-high latitudes, and relatively minor changes in the extreme percentiles over the Mediterranean (Pall et al., 2007). Likewise, Westra et al. (2013) show that maximum precipitation events at daily time scales are becoming more intense in the observational record for most stations globally, with least pronounced changes occuring in drier sub-tropical regions, such as the Mediterranean.

4.4. Conclusion

In this study we examined two commonly used techniques for assessing the odds of extreme weather events in a changing climate. The purpose of which was to test if using statistical inferences on relatively small sample sizes (as is common in observational studies) would give quantitatively similar results to using large sample sizes of the simulated climate (as is used in complementary experiments, e.g. Stott et al. (2004); Otto et al. (2012)). When it comes to attribution statements it is important to account for such potential differences, because statements are

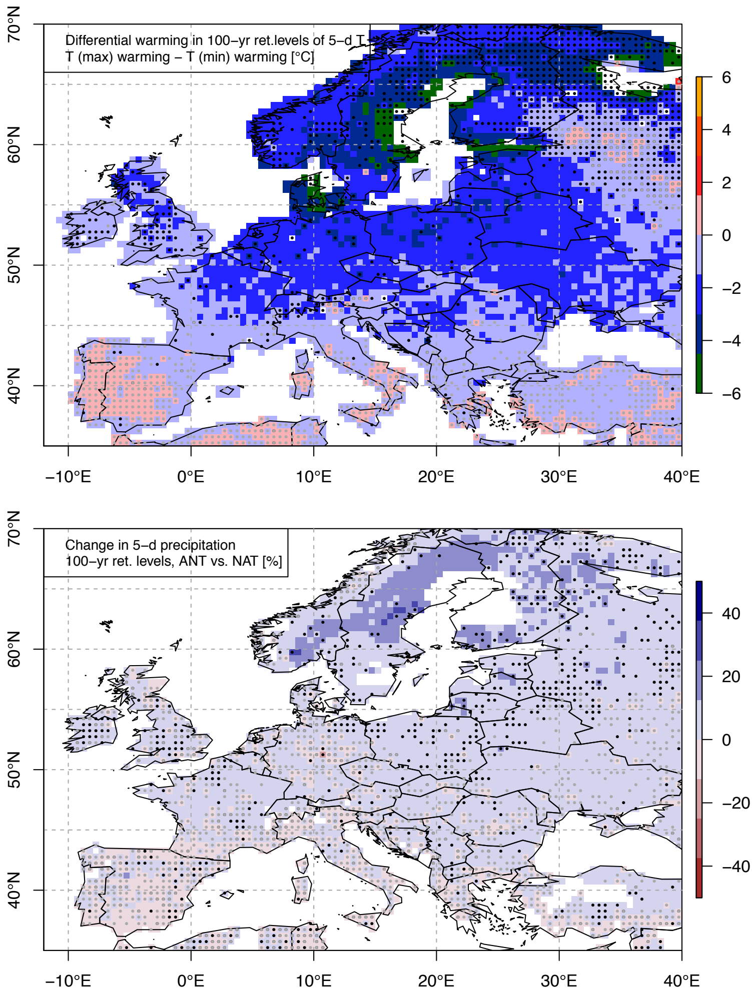


FIGURE 4.5.: (Top) Difference in warming of the warm and the cold tail of 100-year return periods of daily winter extremes. (Bottom) Percent changes in 100 year return levels of 5-day rainfall sums in Europe between DJF2013/14 and a counterfactual 'non-industrial' winter season with a similar sea surface temperature pattern. Black dots indicate poor goodness-of-fit as indicated by the AD-test, while grey stippling indicates non-significant changes in return levels.

often based on the evaluation of relatively subtle changes in the tails. The analysis is then used to understand how the lower and (upper) tails of temperature (rainfall) distributions change under anthropogenic climate change in a European winter season.

We show that for some regions of Europe, the definition of what counts as extreme data can drastically change how the extreme value distribution (in this case the GEV) is fitted. In some cases, for instance winter temperatures over NW Russia, the GEV model based on an annual block size does not fit well to empirical estimates of the tails from thousands of ensemble members of a climate simulation. The reason for the observed disparity may well be because of different dynamic regimes under very extreme meteorological conditions that do not occur in every seasonal simulation. As such, a careful choice of parameters is crucial when using EVT for understanding extreme events, especially if small sample sizes akin to observational data are used. We argue that large ensemble simulations might offer a route to test the robustness of such parameter choices for any particular variable or region of interest. Further, a combination of GEV-based inference with ensemble simulations allows to reduce the number of required simulations substantially for estimating high return periods. For example, we show that when analysing extreme temperatures over Russia, a statement regarding the 1000-year return period can be made by fitting an extreme value distribution to a sample size of 1000 years, whereas empirical estimates would require an order of magnitude larger sample size. Similar conclusions can be drawn for Spain.

Using the refined resampling technique for understanding rare extremes and with respect to the case study of the unusual winter 2013/14, we find a widespread warming pattern throughout Europe, which led to a reduction of return periods of very cold winter days (as derived from seasonal minima). This is accompanied by an increase in warm winter anomalies both in frequency and magnitude throughout the model domain. Crucially, the observed warming of daily winter temperature minima is larger than the maxima, showing an asymmetry in the changes in extremes. Finally, predominantly northern parts of Europe show significant increases in unusually extreme daily rainfall events, emphasizing the importance of considering extreme events on a regional basis.

Part II.

Observations-based constraints to improve the simulation of climate extremes and ecosystem impacts

5. A novel bias correction methodology for climate impact simulations[1,2]

Abstract

Understanding, quantifying and attributing the impacts of extreme weather and climate events in the terrestrial biosphere is crucial for societal adaptation in a changing climate. However, climate model simulations generated for this purpose typically exhibit biases in their output that hinders any straightforward assessment of impacts. To overcome this issue, various bias correction strategies are routinely used to alleviate climate model deficiencies most of which have been criticised for physical inconsistency and the non-preservation of the multivariate correlation structure. In this study, we introduce a novel, resampling-based bias correction scheme that fully preserves the physical consistency and multivariate correlation structure of the model output. This procedure strongly improves the representation of climatic extremes and variability in a large regional climate model ensemble (HadRM3P, `http://www.climateprediction.net/weatherathome`*), which is illustrated for summer extremes in temperature and rainfall over Central Europe. Moreover, we simulate biosphere-atmosphere fluxes of carbon and water using a terrestrial ecosystem model (LPJmL) driven by the bias corrected climate forcing. The resampling-based bias correction yields strongly improved statistical distributions of carbon and water fluxes, including the extremes. Our results*

[1] This chapter is published as Sippel, S., F. E. L. Otto, M. Forkel, M. R. Allen, B. P. Guillod, M. Heimann, M. Reichstein, S. I. Seneviratne, K. Thonicke, and M. D. Mahecha. 2016. *Earth System Dynamics* **7**, 71–88. doi:10.5194/esd-7-71-2016.

[2] Supplementary Online Material (SOM) that provides additional information but that is not necessary for understanding the scientific content of this Chapter is available under `http://www.earth-syst-dynam.net/7/71/2016/esd-7-71-2016-supplement.pdf`

thus highlight the importance to carefully consider statistical moments beyond the mean for climate impact simulations. In conclusion, the present study introduces an approach to alleviate climate model biases in a physically consistent way and demonstrates that this yields strongly improved simulations of climate extremes and associated impacts in the terrestrial biosphere. A wider uptake of our methodology by the climate and impact modelling community therefore seems desirable for accurately quantifying changes in past, current and future extremes.

5.1. Introduction

Weather and climate extreme events such as heat waves, droughts or storms cause major impacts upon human societies and ecosystems (IPCC, 2012). In recent years, these climatic events have changed in intensity and frequency in many parts of the world (Barriopedro et al., 2011; Donat et al., 2013b; Seneviratne et al., 2014) and changes are likely to continue throughout the 21st century (Sillmann et al., 2013a). Therefore, improving the scientific understanding of these events, including the link to impacts, constitutes an important research challenge (IPCC, 2012; Zhang et al., 2014).

The impacts of climate extremes and potential changes therein are strongly felt in the terrestial biosphere. For example, heat and drought events trigger ecological responses (Reyer et al., 2013; Frank et al., 2015), which in turn induces changes to the cycling of water and carbon through such systems with potential feedback to the atmosphere and climate system (Reichstein et al., 2013; Frank et al., 2015). Indeed, on continental to global scales, it has been shown that large-scale reductions in photosynthetic uptake of carbon by plants are mainly driven by water limitations (Zscheischler et al., 2014b,c). Furthermore, it has been demonstrated that a single large event such as the European heat and drought summer 2003 alone might undo several years of ecosystem carbon sequestration (Ciais et al., 2005), thus potentially jeopardizing the terrestrial carbon sink potential (Lewis et al., 2011).

A widely debated question in this realm is whether the observed changes in the occurrence of climatic extremes and associated impacts can be attributed to specific changes in climate forcing, both anthropogenic or natural (Allen, 2003; Stone and Allen, 2005; Stone et al., 2009). To this end, large climate model ensembles are needed in order to derive robust probabilistic conclusions about changes in the odds of these events (Bindoff et al., 2013; Massey et al., 2015), because direct assessments of rare extremes are often prohibited by the lack of long and good quality observational time series. Hence, climate models are indispensable tools to study present and future climate extremes on various spatial and temporal scales, and the availability of such simulations is often a prerequisite for studying climate impacts.

However, despite considerable progress in recent years, global and regional climate models typically exhibit biases in various statistical moments of their simulated variables (Ehret et al., 2012; Wang et al., 2014), which often impedes direct assessments of climate extremes (Otto et al., 2012; Sippel and Otto, 2014) or simulating impacts (Maraun et al., 2010; Hempel et al., 2013). These biases are often due to an imperfect representation of physical processes in the models, parametrisations of sub-grid scale processes, and an over- or underestimation of feedbacks with the land-atmosphere or ocean-atmosphere feedbacks (Ehret et al., 2012; Mueller and Seneviratne, 2014). Due to the various origins of model biases, these biases are frequently varying depending on weather patterns both spatially and temporally, for instance in the distributed weather@home ensemble-based modelling framework (Massey et al., 2015) or in an ensemble of regional climate models (Vautard et al., 2013).

To alleviate this issue, various bias correction schemes have been developed in recent years that generally aim to statistically transform biased model output in order to derive more realistic simulations (see e.g. Maraun et al., 2010; Teutschbein and Seibert, 2012). To do so, a statistical relationship ('transfer function') is built between the statistical distribution of an observed and simulated variable (Piani et al., 2010). Such methods span a wide range from very simple parametric transformations adjusting simulated means to observations (i.e. also called the 'delta method' (additive) or 'linear scaling' (multiplicative), (Teutschbein and Seibert, 2012)) to sophisticated, nonparametric approaches that aim to correct various statistical moments of the simulated distributions such as quantile mapping approaches (Wood et al., 2004; Gudmundsson et al., 2012).

However, the application of bias correction implicitly requires that a range of assumptions are met, which might be questionable in many cases and are discussed in detail in Ehret et al. (2012). Most importantly, the application of bias correction implicitly assumes that the statistical transformation improves the simulated output time series ('effectiveness'), whilst the signal of interest, e.g. the climate change signal or properties of the extremes, remains accurately detectable ('reliability'). Those assumptions are not always fulfilled since statistical bias correction methods are not based on physical principles, but operate rather heuristically on an observed model-data mismatch. To this end, even relatively simple methods that are designed to adjust ‘only’ simulated long-term monthly means to observations (e.g. Hempel et al., 2013) lack a sound physical rationale to whether these adjustments are to be made additively or multiplicatively. Further, the assumption of time invariant biases that currently underlies state-of-the-art bias correction procedures (Christensen et al., 2008; Ehret et al., 2012) might be especially critical for century-long climate simulations spanning several degrees of warming (Christensen et al., 2008; Buser et al., 2009) including changing land-atmosphere feedback processes (Seneviratne et al., 2006). Recent studies have shown that this assumption is questionable for future climate simulations (Maraun, 2012; Bellprat et al., 2013), and have made attempts to address time-dependent biases.

Furthermore, an adjustment of daily variability does not necessarily improve monthly statistics, thus emphasizing the role of time scales at which bias correction is conducted (Haerter et al., 2011). Lastly, if impact simulations are to be conducted with bias-corrected output of numerical climate models, the multivariate correlation structure between climate variables deserves attention: Most bias-correction schemes that are currently in use to simulate impacts have been suggested to correct each variable separately (Hempel et al., 2013) and hence dependencies between variables are often not retained. This is especially critical for assessments of extreme events and ‘compund events’ (Leonard et al., 2014), where inter-variable interactions, such as soil moisture-temperature feedbacks might play an important role (Seneviratne et al., 2006). Although recent progress has been made to derive bivariate bias correction schemes (Hoffmann and Rath, 2012; Piani and Haerter, 2012; Li et al., 2014), to the best of our knowledge currently no bias correction scheme retains a multivariate correlation structure of a larger set of input variables for impact simulations.

In conclusion, accounting for biases in climate model output is crucial in order to produce credible climate model simulations. Nonethless, statistical transformations are to be applied with caution and the changes induced to the simulated statistical moments, multivariate dependencies and spatio-temporal patterns deserve considerable attention. Since the tails of statistical distributions are especially sensitive to changes in statistical moments such as the mean and variance (Katz and Brown, 1992), the latter holds in particular for assessments of extreme events and highlights the need for physically consistent ways to alleviate climate model biases.

In this paper, we demonstrate how a physically consistent bias correction of a regional climate model ensemble might aid to better simulate climatic extreme events and impacts in the terrestrial biosphere (see Figure 5.1 for the methodological workflow of the paper).

First, we introduce a novel methodology to alleviate biases in the output of climate model ensembles that successfully circumvents major deficiencies of statistical bias correction (Section 5.3): an ensemble-based probabilistic resampling approach retains the physical consistency of the regional climate model output. This includes the preservation of the multivariate correlation structure, and the procedure is shown to considerably improve the simulation of various statistical moments of the simulated variables. Secondly, we assess contemporary temperature and precipitation extreme events in Central Europe on monthly to seasonal time scales by comparing a widely used 'standard' statistical bias correction methodology (Hempel et al., 2013) with the original model simulations and the probabilistic resampling (Section 5.4.1 and 5.4.2). This evaluation also focuses on the uncertainty induced by different observational datasets used as a basis for any bias correction approach. Thirdly, we explicitly test how differently corrected climatic data propagates into the simulation of impacts on major component fluxes of terrestrial carbon (net ecosystem exchange (NEE), gross primary production (GPP) and ecosystem respiration (Reco)) and water cycling (actual evapotranspiration, AET) in the terrestrial biosphere using a dynamic vegetation model (LPJmL, Section 5.4.3). To this end, we demonstrate that different ways to deal with biases in climate simulations yield both qualitatively and quantitatively different results regarding simulated impacts, which affect both central moments of the distribution as well as extremes and variability.

5.2. Data

5.2.1. Climate model simulations

In this study, regional climate model ensemble simulations spanning 26 years (1986-2011) with approx. 800 ensemble members per year from the weather@home

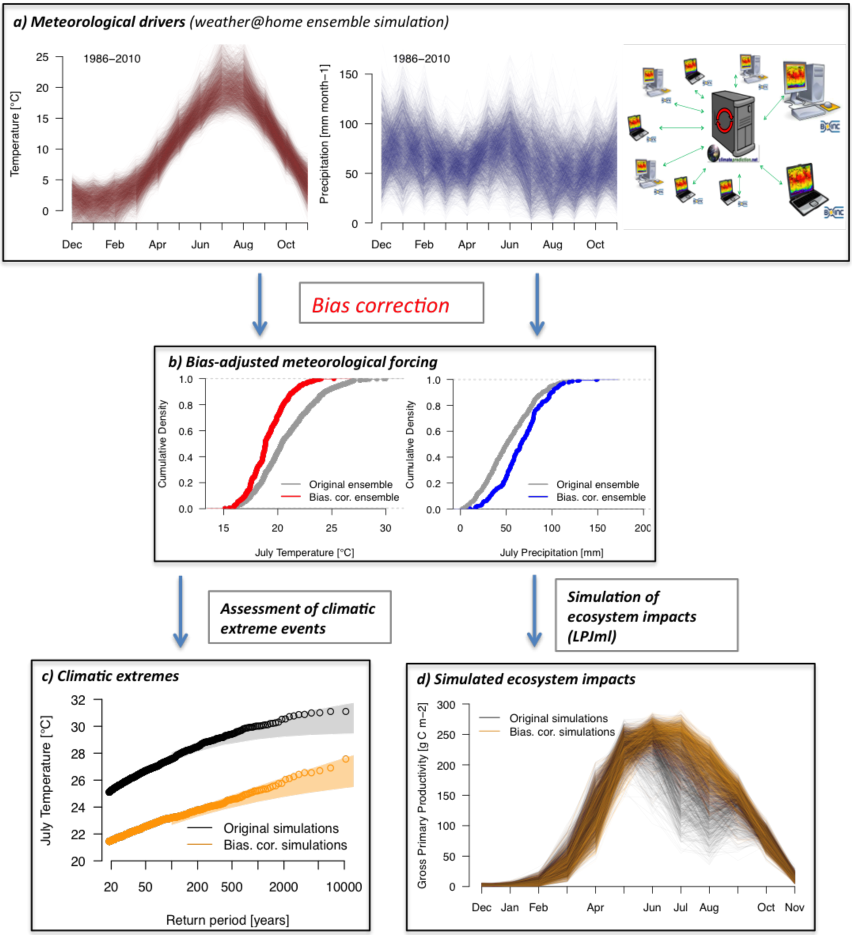


FIGURE 5.1.: Methodological workflow of the study. a) Generation of regional climate model simulations using a large ensemble modelling framework (*climateprediction.net/weatherathome*). b) Adjustment of biases in the regional climate model's output. c) Assessment of weather and climate extreme events. d) Ensemble simulation of ecosystem-atmosphere fluxes of carbon and water using the LPJmL model.

distributed computing platform are investigated (Massey et al., 2015). The 'atmosphere-only' simulations were conducted over the European region (identical to the EURO-CORDEX region Giorgi et al., 2009) using a regional model

(HadRM3P) on a rotated grid nested into the global HadAM3P model. Both models share the same model formulation and are described in Pope et al. (2000). The regional (global) simulations are run with a spatial resolution of 0.44°x0.44° (1.875°x1.25°) with 19 vertical levels, and the temporal resolution is set to 5 (15) minutes (Massey et al., 2015). The models are driven by observed sea surface temperatures and sea ice fractions, the observed composition of the atmosphere (greenhouse gases, aerosols) and anomalies in the solar cycle (Massey et al., 2015). To derive different ensemble members, the initial conditions of the driving GCM are perturbed on 1st December of each 1-year simulation (ibid.). For further analysis and bias correction, the ensemble simulations were remapped to 0.5 ° spatial resolution using a conservative remapping scheme (Jones, 1999).

Massey et al. (2015) demonstrate that the ensemble setup described above produces a realistic representation and statistics of European weather events, including the extremes for most seasons and regions. However, despite these encouraging results, a relatively large mismatch remains between the statistical distribution of the ensemble simulations and the observations in Northern hemisphere summer, which holds for the means of simulated seasonal temperature and precipitation (Massey et al., 2015) as well as for higher statistical moments, shown in the Supplement against the ERA-Interim reanalysis dataset (Dee et al., 2011). Especially in more continental parts of the European model region, HadRM3P shows a pronounced hot and dry bias in simulated summer weather (Supplementary Online Figures S1–S3b). However, note that the ensemble setup still captures the entire range of the observed distribution (Supplementary Online Figure S1). In HadRM3P, these biases are likely related to an imperfect parametrisation of cloud processes in the model, leading to an overestimation of incoming solar radiation, which in turn triggers warm and dry summer conditions (R. Jones, 2015, pers. comm.) that are further amplified by strong soil moisture-temperature coupling in the model (Supplementary Online Figure S4). In this context, it is worthwhile to note that these biases are not a peculiarity of the regional climate model employed in this study, but indeed hold for many dynamically downscaled regional climate model simulations over Europe (Buser et al., 2009; Boberg and Christensen, 2012).

5.2.2. Simulation of atmosphere-biosphere carbon and water fluxes

To assess terrestrial biosphere impacts of bias correcting regional climate simulations, we simulate ensembles of atmosphere-biosphere fluxes of carbon (NEE, GPP, Reco) and water (AET) using the Lund-Potsdam-Jena managed land scheme (LPJmL, Version 3.5, Sitch et al., 2003; Bondeau et al., 2007), a state-of-the-art process-based dynamic global vegetation model that accounts for human land use. We follow Schulze (2006) and Chapin III et al. (2006) in their definition of major components of carbon cycling in terrestrial ecosystems: Gross primary productivity (GPP) denotes the vegetation's gross photosynthetic uptake of carbon from the atmosphere, whereas ecosystem respiration (Reco) is defined as the respiratory release of carbon by plants and microbes in the ecosystem, i.e. including both (autotrophic) plant respiration and (heterotrophic) soil organic matter decomposition. Net ecosystem exchange (NEE) constitutes the net carbon flux from the ecosystem to the atmosphere, i.e. the difference between Reco and GPP.

LPJmL simulates vegetation dynamics (growth, competition and mortality) and fully coupled cycling of carbon (photosynthesis, autotrophic & heterotrophic respiration) and water (transpiration, evaporation, interception, runoff) in terrestrial ecosystems and managed systems (Sitch et al., 2003; Gerten et al., 2004; Bondeau et al., 2007). The model is driven with monthly or daily climatic input data (temperature, precipitation, incoming shortwave radiation & net longwave radiation), atmospheric carbon dioxide concentrations and soil texture. Vegetation structure in LPJmL is characterised by the fractional coverage of 11 plant functional types that differ in their bioclimatic limits and ecophysiological parameters. Vegetation dynamics and competition are explicitly represented using a set of allometric and empirical equations and updated annually (Sitch et al., 2003).

GPP in LPJmL follows the process-oriented coupled photosynthesis and water balance scheme of the BIOME3 model (Haxeltine and Prentice, 1996). Subsequently, autotrophic (growth and maintenance) respiration is subtracted from GPP, and the net carbon uptake is allocated to plant compartments based on a set of allometric constraints (Sitch et al., 2003). Ecosystem heterotrophic respiration depends on temperature and moisture in each litter and soil carbon pool; carbon decomposition dynamics are simulated as first-order kinetics with specified decomposition rate in each pool (Sitch et al., 2003). Water cycling in LPJmL has been improved by Gerten et al. (2004) and Schaphoff et al. (2013), where actual evapotranspiration (the sum of evaporation, transpiration and interception) is computed as a function from atmospheric demand and soil moisture supply. Phenology and photosynthesis-related parameters in the LPJmL version used in this paper have been optimised against remote sensing observations for an improved simulation of natural vegetation greenness dynamics (Forkel et al., 2014), including the introduction of a novel phenology scheme.

LPJmL has been applied in a range of studies assessing ecosystem responses to anomalous climatic conditions (Rammig et al., 2015; Van Oijen et al., 2014; Zscheischler et al., 2014c; Rolinski et al., 2015). Rolinski et al. (2015) argued that the model might be able to capture various ecosystem physiological responses to climatic extreme events such as heat or drought through various pathways. These include a water stress response through reduced stomatal conductance, which in turn decreases both photosynthetic carbon uptake and transpiration. Further, the model responds to very high temperatures by a photosynthesis inhibition and increased respiration (Rammig et al., 2015).

In this paper, we use the weather@home climate data to derive ensemble-based simulations of the functioning of terrestrial ecosystems. LPJmL simulations are conducted in natural vegetation mode (i.e. no human land use, fire or permafrost) in 0.5° spatial resolution and monthly time steps over Central Europe. For each bias-corrected ensemble dataset, 2000 years of spinup to equilibrate soil carbon pools were conducted, using randomly chosen years from the first 10 years of the HadRM3P ensemble. Subsequently, atmosphere-biosphere fluxes were simulated at the monthly time scale for 1986-2010 over Central Europe (see Figure 4.1 for methodological workflow). This procedure was repeated five times to check that no carry-over effects from the randomised spinup affect simulated biosphere-atmosphere carbon fluxes in the transient period. Since this was not the case, differences in carbon and water fluxes and their extremes can be directly attributed to the bias correction of the climatic forcing in the transient period, and are analysed in Section 5.4.3.

5.2.3. Observations

Any statistical assessment or correction method of biases requires reference datasets, and the quality of bias adjustment is thus restricted by the quality of observations or reanalysis data available (Ehret et al., 2012; Hempel et al., 2013). Consequently, the sensitivity of 'bias corrected' model output to any given set of observations needs to be tested. In this study, a range of observational datasets is used in order to characterise uncertainty induced by using different observations for bias correction. In total, seven different temperature and precipitation datasets consisting of gridded observations/reanalysis were used (one at a time) for the univariate bias correction (Section 5.4.2) and are detailed in Table 5.1. The simultaneous correction of multiple variables for the impact simulations in the terrestrial biosphere presented in Section 5.4.2 are conducted using ERA-Interim as reference dataset (Dee et al., 2011, see Table 5.1).

To conduct the sensitivity analysis of climatic extremes and associated biosphere impacts to the type of bias correction applied, we select one focus region in Central Europe. This region roughly encompasses Germany ($47.5 - 55.0°$N, $6.0 - 15.0°$E) and consists of temperate mid-latitude climate with maritime influence to the North-West and more continental characteristics to the East. In addition, to sample local (i.e. grid cell scale) variability we test different bias correction scheme on one single grid cell located in Central Germany ('Jena pixel', $50.75°$N, $11.75°$E).

5.3. Methods

In this section, we describe the different bias correction methods deployed in this study. First, a bias correction methodology for impact simulations that has been adopted widely is summarised (Hempel et al., 2013). Second, we introduce the novel resampling-based bias correction scheme and lastly the methodologies for evaluation are described.

5.3.1. Statistical bias correction

Hempel et al. (2013) presented a bias-correction that is designed to preserve long-term trends in simulated impacts and that has been used widely in simulating

TABLE 5.1.: Datasets used for bias correction and evaluation.

Name of dataset	Climate variables	Domain & Orig. Resolution	Provider & Reference
Berkeley Earth Observations (gridded experimental)	Tair	Europe, 0.25°, monthly, 1850-2012	`http://www.berkeleyearth.org`, Rohde et al. (2013)
Climate Research Unit (CRU), High-resolution gridded datasets	Tair, Precip.	Global, 0.5°, monthly, 1901-2012	Climate Research Unit, `http://www.cru.uea.ac.uk/cru/data/hrg/`, Harris et al. (2014)
CRUNCEP	Tair, Precip., SWdown, LWdown	Global, 0.5°, daily, 1948-2012	`http://dods.extra.cea.fr/data/p529viov/cruncep/readme.htm`
Global Precipitation Climatology Centre monthly precipitation (GPCC)	Precip.	Global, 0.5°, monthly, 1901-2012	Global Precipitation Climatology Center (GPCC), `http://gpcc.dwd.de/`, Schneider et al. (2014)
E-OBS gridded dataset	Tair, Precip.	Europe, 0.5°, daily, 1951-2014	European Climate Assessment & Dataset (ECA&D), `http://www.ecad.eu`, Haylock et al. (2008)
ERA-Interim, Version 2 (ERAI)	Tair, Precip., SWdown, LWdown, LE	Global, $\approx$ 0.7°, 6-hourly, 1979-2014	European Centre for Medium Range Weather Forecasts (ECMWF), `http://apps.ecmwf.int/datasets/data/interim-full-daily/`, Dee et al. (2011)
Model Tree Ensembles	LE	Global 0.5°, monthly, 1982-2011	MPI Biogeochemistry Jena, Jung et al. (2011)
WATCH-harmonised (WFDharmonised)	Tair, Precip., SWdown, LWdown	Europe, 0.5°, daily, 1901-2012	MPI Biogeochemistry Jena, Weedon et al. (2011); Beer et al. (2014)
WATCH ERA-Interim (WFDEI)	Tair, Precip., SWdown, LWdown	Global, 0.5°, daily, 1979-2012	Weedon et al. (2014)

effects of climatic changes in different sectors such as water, agriculture, ecosystems, health, coastal infrastructure, and agro-economy (see Warszawski et al., 2014, for an overview).

The approach builds on earlier, conventional statistical bias correction schemes (Piani et al., 2010; Haerter et al., 2011) and is based on linear transfer functions of the form

$$x_{cor} = a + bx. \tag{5.1}$$

Here, x and x_{cor} represent the simulated and corrected climatic variable, a and b are coefficients to be calibrated.

In Hempel et al. (2013) the transfer function is applied additively (for temperature, i.e. $b = 1$), such that

$$a = \overline{T_{obs}} - \overline{T_{mod}}; \tag{5.2}$$

where $\overline{T_{mod}}$ and $\overline{T_{obs}}$ represent the means of simulated and observed monthly temperatures, respectively.

To account for positivity constraints for precipitation and radiation components, Hempel et al. (2013) suggested a multiplicative adjustment of those variables (i.e. $a = 0$), such that

$$b = \frac{\overline{x_{obs}}}{\overline{x_{mod}}}. \tag{5.3}$$

These parametric transformations are applied on each grid cell and for each month separately to account for potential temporal and spatial structure in the biases. By applying this transfer function, long-term monthly means of the simulated distributions are matched with those in observations for each grid cell (Hempel et al., 2013). In addition to adjusting monthly means, Hempel et al. (2013) also adjust daily variability about the monthly means, but (importantly) the year-to-year variability at monthly time scales remains unchanged. In our present analysis, we follow this conventional bias correction scheme for comparison and denote it by ‘ISIMIP’.

Furthermore, to isolate the effects of bias-correcting the full suite of impact variables (temperature, precipitation and radiation) vs. correcting simulated precipitation only, we conduct impact simulations with a ‘precipitation only’ bias-corrected scenario (‘PRECIPCOR’).

5.3.2. A novel resampling-based ensemble bias correction scheme

In this subsection, we introduce a novel ‘bias correction scheme’ suitable for ensemble simulations that retains the physical consistency and multivariate correlation structure of the model output. The idea is to resample plausible ensemble members from a large ensemble simulation given the statistical distribution of an observable meteorological metric (‘constraint’). The procedure is illustrated using the weather@home ensemble described above.

The largest biases in the HadRM3P simulation occur in the summer season (JJA) over the European model domain, where the model ensemble produces too frequent and too pronounced hot and dry conditions (Massey et al., 2015). Importantly however, the ensemble spans the entire distribution of observed summer conditions in most parts of Europe, i.e. some (but too few) ensemble members produce relatively wet and cold summers. Therefore, our resampling-based correction approach is designed to alleviate the representation of summer conditions in the model ensemble.

The bias correction procedure consists of the following steps and is illustrated in Figure 5.2:

1. Define an observable meteorological metric that is poorly represented (‘biased’) in the model ensemble. In this paper, we use summer mean temperatures over Central Europe, which are relatively well-constrained in observational datasets.

2. Estimate the probability distribution function of the meteorological constraint from observational datasets using e.g. a kernel density fit ($\hat{f}_{obs}(x)$, see e.g. Figure 5.2a, blue line for an illustration), where x denotes the constraint. Here, we use a Gaussian kernel with reliable data-based bandwidth selection (Sheather and Jones, 1991) fitted over the observed meteorologi-

cal constraint for the period 1986-2011 using various observational datasets (one at a time).

3. Estimate the probability distribution of the meteorological constraint in the model ensemble using the same estimation procedure as above over all ensemble members and all years ($\hat{f}_{mod}(x)$, see Figure 5.2a, red line). The deviation between the red and blue line in Figure 5.2a illustrates the temperature bias in the weather@home ensemble.

4. Derive a transfer function that maps any given quantile in the observations ($q_{obs,X}$) to the respective quantile in the model ensemble ($q_{mod,X}$, see Figure 5.2b), such that $q_{mod,X} = TF(q_{obs,X})$ using the fitted kernels $\hat{f}_{obs}(x)$ and $\hat{f}_{mod}(x)$ to determine empirical quantile functions. For example, a 'median temperature' summer over Central Europe (approx. 17.2°C) would correspond to the 50th percentile in the observations-based kernel (by definition). The transfer function would then map the 50th percentile in $\hat{f}_{obs}$ to the corresponding 20.4th percentile in $\hat{f}_{mod}$ (i.e. average summer temperatures of 17.2° would correspond to the 20.4th percentile in the model ensemble, see Figure 5.2b). In this study, we use Cubic Hermite splines (Fritsch and Carlson, 1980) to determine the transfer function shown in Figure 5.2b.

5. Derive a new 'bias-corrected' ensemble (of sample size n) by randomly resampling n times from observed percentiles ($q_{obs,X}$) and retaining the ensemble member that corresponds to $q_{mod,X}$ as given by the transfer function ($n = 800$ per year in our study).

Hence, the outlined procedure does not adjust any output variable in the model ensemble thus preserving physical consistency, but rather selects plausible ensemble members. This procedure invariably leads to a reduction in the effective ensemble size: For example in the HadRM3P ensemble, roughly the hottest 20% of simulations are effectively not chosen for the resampled ensemble since they are implausibly hot (Figure 5.2a). However, an evaluation of the sample size in the bias corrected ensemble shows that at least 4% of the ensemble simulations match any decile of observations (Figure 5.2d, in an unbiased ensemble exactly

10% of ensemble simulations would match each decile of observations), corresponding to an effective sample size of at least approx. 1000 model years (= ensemble members) per decile of observations (Figure 5.2d).

In conclusion, the outlined approach to bias correction is conceptually similar to earlier ideas of assigning weights to different regional climate model projections based on each model's performance in order to derive probabilistic multi-model projections (Piani et al., 2005; Collins, 2007; Knutti, 2010; Christensen et al., 2010). However, instead of a weight assignment specific ensemble members are selected and combined into a new ensemble using the statistical distribution of observed meteorological constraints.

5.3.3. Analysis methodology

In Section 5.4.1 the outlined bias correction method is evaluated for the simulation of temperature, rainfall and radiation using standard evaluation metrics such as seasonal mean values and interannual variability. Further, we evaluate soil moisture coupling in the original and bias corrected ensemble against reanalysis data and upscaled observations by computing the correlation between summer mean temperatures and the mean latent heat flux following Seneviratne et al. (2006).

Moreover, we analyse empirical return times of the original and bias-corrected ensembles that are derived by plotting each ensemble value against its rank both for climatological extremes (Section 5.4.2: monthly summer temperatures and cumulative summer rainfall) and simulated ecosystem-atmosphere annual fluxes of water and carbon (Section 5.4.3).

To further understand discrepancies between the bias-corrected ensemble simulations and observed climate extremes (Section 5.4.2), we characterise the tails of simulated and observed variables by extreme value theory (Coles et al., 2001). Hence, generalised extreme value distributions (GEV) are derived from monthly temperature and precipitation in a procedure similar to Sippel et al. (2015a), i.e. by resampling block-maxima in randomly concatenated 10-year sequences of ensemble data and fitted to a GEV model using generalised maximum likelihood estimation. In observational data, only a relatively small sample size is available (mostly 1901-2014 only) that is additionally plagued by non-stationarity and

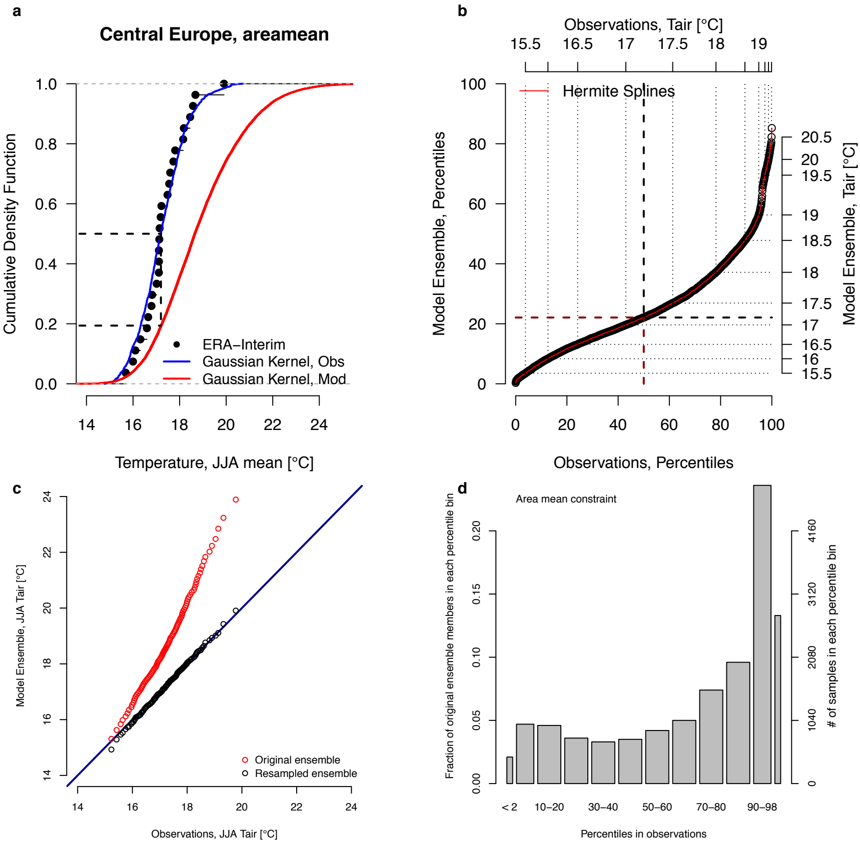


FIGURE 5.2.: Illustration of ensemble-based resampling methodology. a) Empirical cumulative density function of JJA mean temperatures over Central Europe in ERA-Interim. The non-parametric fit to the cumulative density using a Gaussian kernel for observations and the model ensemble are shown by the blue and red lines, respectively. b) A transfer function between the observed and modelled distribution is derived using Cubic Hermite splines. c) Quantile-quantile plot for the original and resampled ensemble for the JJA temperature constraint. d) Fraction of original ensemble members in percentile bins of the observed distribution (blue line in (a)), i.e. 'effective ensemble size' after resampling.

does not match the period in which ensemble simulations are available (1986-2011). Hence, for monthly temperatures we subtract the trend and seasonal cycle from the original time series using Singular Spectrum Analysis (von Buttlar et al., 2014), and subsequently resample (monthly) summer temperature anomalies (for the whole time series) by adding a trend and seasonal cycle component drawn randomly from the period of available ensemble simulations (1986-2011, each observational dataset is analysed separately). Approximate stationarity was assumed for seasonal precipitation, and hence no further adjustments were made. Lastly, GEV models were fitted to the observations following the procedure as described above.

5.4. Results

This section is structured as follows: First, we evaluate the bias correction procedure both for resampling based on an area mean and grid cell based constraint. Second, climate extreme statistics and their sensitivity to bias correction schemes are investigated (Section 5.4.2). More specifically, the probabilistic resampling scheme introduced in sect. 3.2 is evaluated against a conventional bias correction scheme (Hempel et al., 2013, Section 5.3.1) and compared against the uncorrected simulations and different observational datasets. Third, we illustrate how biases and their ‘correction’ propagate into climatic impacts exemplified by simulations ecosystem water and carbon fluxes in Central European natural vegetation.

5.4.1. Evaluation of resampling bias correction

An evaluation of the distribution of variables in the resampled ensemble in Central Europe shows that it not only improves the simulation of seasonal mean temperatures (which it does by construction), but also yields considerable improvements to the simulation of rainfall and radiation components (Figure 5.3). This suggests that these biases are related to specific synoptic situations in summer, justifying to apply the bias correction approach to summer months. Hence, the multivariate covariance structure between temperature, precipitation and radiation as simulated by HadRM3P appears to be well represented in the model simulations posterior to the updating procedure given the reanalysis/observational data. Moreover, while this procedure also improves the simulation of summer temperatures and precipitation on a monthly time scale, virtually no changes in the ensemble statistics are induced to non-summer months (Supplementary Online Figure S1), indicating that the time scales of temporal decorrelation are short enough for a successful application of the resampling procedure. However, while conventional statistical bias correction following Hempel et al. (2013) adjusts monthly means of the distributions of precipitation and radiation (by construction), changes are induced by the multiplicative adjustment to the width and shape of the distribution, including its tails (Figure 5.3, see also Section 5.4.2).

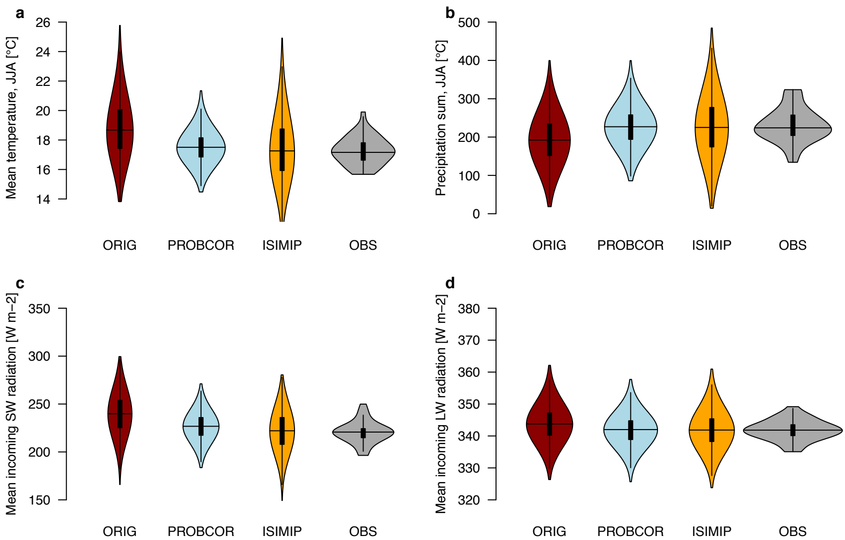


FIGURE 5.3.: Evaluation of the resampling bias correction methodology for the study area in Central Europe for (a) temperature, (b) precipitation, (c) incoming short-wave radiation, and (d) incoming long-wave radiation. Both sides of each violin are constructed as rotated, equal-area kernel density estimates, and a standard boxplot is drawn inside each violin.

An evaluation of the resulting spatial patterns of the resampling bias correction shows that the representation of the simulated statistical distributions of temperature and precipitation are considerably improved in Central Europe (area mean constraint) and across the entire European model region (single grid cell constraints, Supplementary Online Figures S2a–S3b). Remarkably, this holds not only for seasonal averages, but also for higher statistical moments such as the inter-decile range.

Furthermore, we test the representation of land-atmosphere coupling in the original and resampled model ensemble by investigating the correlation strength between summer mean temperatures (T) with latent heat (LE) fluxes following Seneviratne et al. (2006). The original HadRM3P ensemble shows strong water limitation of evapotranspiration in summer (negative correlation between LE and T) for most temperate and Mediterranean European regions, thus overestimating soil moisture control compared to reanalysis data and upscaled observations (Supplementary Online Figure S4). In the resampled ensemble, land-atmosphere coupling remains strongly soil moisture controlled in the Mediterranean regions, but reduces in temperate European regions, resulting in spatial patterns that resemble those of land-atmosphere coupling in ERA-Interim (Supplementary Online Figure S4). The latter finding indicates that the procedure of eliminating implausible ensemble members also yields an improved representation of physical processes such as land-atmosphere coupling in the resampled ensemble.

5.4.2. Sensitivity of climatic extremes to bias correction

Summertime temperature extremes

Summertime monthly extreme temperatures are shown in Figure 5.4 as a spatial average for the study region located over Central Europe and for an illustrative and randomly chosen grid cell ('Jena grid cell').

The location, slope and shape of the lines in the return time plots shown in Figure 5.4 reveal that the tails of simulated monthly temperature extremes are highly sensitive to the type of bias correction applied, both for a regional average and a single grid cell: Uncorrected simulations overestimate both location and scale (i.e. slope of the line in the return time plot) of positive temperature anomalies in summer, while this is not the case for anomalously cold summer months (Figure 5.4). An additive adjustment of monthly means (orange lines in Figure 5.4, Hempel et al., 2013) preserves slope and shape of the tail, i.e. preserves the year-to-year variability of simulated monthly temperatures (and biases therein) in the ensemble. Note that this procedure cannot account for the asymmetry between the upper and lower tail of simulated monthly temperatures - i.e. the offset correction leads to an overcorrection of cold months, whereas the statistics of the hot tails improve only marginally. This is confirmed by a statistical extreme value analysis (Supplementary Online Figures S5a–S6b): The temperature offset approach adjusts only the location of the GEV yielding spurious artefacts in the (originally well simulated) cold tail, whilst not accounting fully for the upper tail due to the aforementioned asymmetries. This is a fundamental drawback of using linear parametric transfer functions, i.e. even if the variability of the simulated distributions would have been adjusted along with the means (see e.g. Sippel and Otto, 2014), the outlined 'asymmetry' issue would not necessarily improve. On the other hand, the probabilistic resampling procedure alters both the location and slope of the lines in the return time plot, where resampling based on a spatial average as well as on a grid cell constraint yield relatively similar representations of the tails. An evaluation of the extreme value statistics shows that the probabilistic procedure indeed considerably improves the statistical characteristics of the simulated tails in the ensemble compared to (long-term) observations (Supplementary Online Figure S5a–S6b). To this end, resampling the original ensemble changes location and scale of the extreme value distributions, but the shape parameter of the tails remain effectively unchanged. Some caution is required due to the relatively scarce availability of observed monthly mean temperatures (i.e. 1901-2014), which induces considerable uncertainties to the parameters of the fitted GEV distributions. Moreover, the different time periods of observations and ensemble simulations (1986-2011) impede a direct 'evaluation' of the bias correction. Nonetheless, this indicative comparison yields very promising results of bias-correcting without invasive changes to the simulated statistical distribution.

Lastly, our analysis shows that any bias correction based on a single grid-cell level induces some uncertainty due to the choice of observational dataset. This is an important issue to consider if impact model simulations on a grid cell scale are to be conducted, whereas regional averages are not as strongly affected. Figure 5.4 shows that resampling the ensemble based on a spatial average constraint reduces this uncertainty as compared to adjusting monthly means or resampling on a grid cell scale.

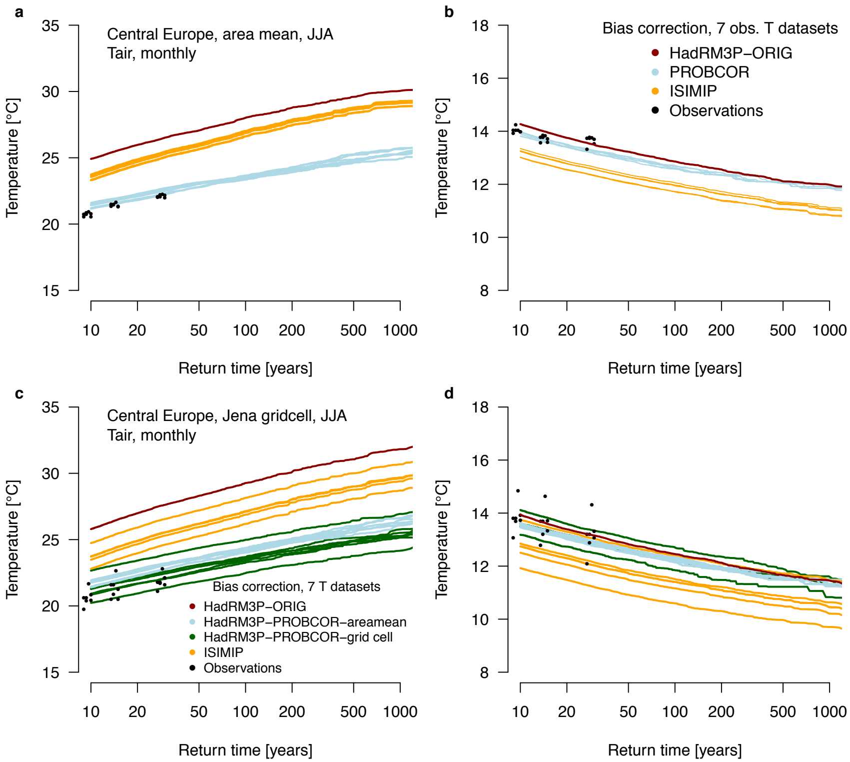


FIGURE 5.4.: Return times of hot (a,c) and cold (b,d) temperature extremes in summer (JJA) in the original regional model simulations ('ORIG'), in the resampled ensemble ('PROBCOR') and the mean-adjusted ensemble ('ISIMIP'). Plots are shown as spatial averages over Central Europe (top panels) and for an illustrative grid cell (Jena, bottom panels). Black dots in each plot indicate empirical return times estimated from observations taken from 7 different datasets that were used for bias correction.

Summertime rainfall extremes

We extend the analysis of the previous paragraph to investigate how resampling based on a temperature constraint alters the representation of summer precipitation in a large ensemble simulation. The original HadRM3P simulated summer seasons are too dry in average over Central Europe, which is largely due to a much too dry lower tail (Figure 5.5), whilst simulated heavy monthly precipitation matches relatively well the available observational data (Figure 5.5).

The tails of simulated (cumulative) seasonal precipitation are sensitive to bias correction. As above, the plots in Figure 5.5 illustrate that a statistical adjustment of the means can be detrimental to statistics of extremes and variability. For instance, scaling monthly means to match observations (Hempel et al., 2013) leads to an inflation of very wet seasons that are physically implausible given the observations (Figure 5.5, orange lines). Likewise, the (biased) dry tail in HadRM3P improves only to a very limited extent if the scaling approach is used. The extreme value analysis (Supplementary Online Figures S6a and S6b) shows that the multiplicative adjustment changes both location and scale of the tail distribution - and that both parameters are not necessarily improving (indeed often deteriorating, see e.g. scale parameters in Supplementary Online Figure S6a and S6b) by applying a simple statistical bias correction. However, resampling based on a *temperature* constraint yields a new ensemble, in which the simulation of both tails has improved (Figure 5.5, Supplementary Online Figure S6b). Only minor changes have been induced to the (well-simulated) wet tail, whilst the previously strongly biased dry tail has considerably improved (Figure 5.5), indicating that temperature-based resampling as deployed here successfully separates 'plausible' ensemble members from the (unrealistic) hot and dry members. The extreme value analysis shows that resampling largely alters the location of the simulated distribution of seasonal rainfall extremes, whilst the scale and shape of the tails remain largely unchanged.

To conclude, it was shown that resampling based on a univariate observations-based temperature constraint improves the simulation of rainfall variability and extremes by teasing out ensemble members that are implausibly hot and dry in our case study region.

5.4.3. The impact of bias correction on simulated ecosystem water and carbon fluxes

In this subsection, we present HadRM3P-LPJmL ensembles of simulated fluxes of carbon and water and discuss bias correction methods with a focus on the extreme tails of the simulated distributions. Further, we investigate the sensitivity of the simulated carbon fluxes to an accurate representation of rainfall in the climatic input data.

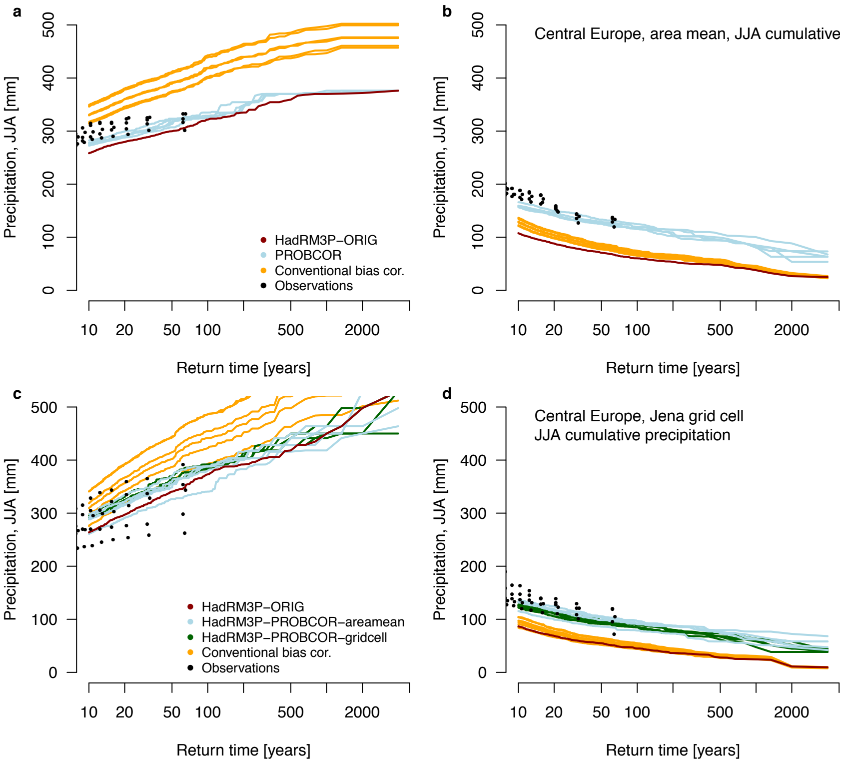


FIGURE 5.5.: Return times of wet (a,c) and dry (c,d) rainfall extremes in summer (JJA) in the original regional model simulations ('ORIG'), in the resampled ensemble ('PROBCOR') and the mean-adjusted ensemble ('ISIMIP'). Plots are shown as spatial averages over Central Europe (top panels) and for an illustrative grid cell ('Jena pixel', bottom panels). Black dots in each plot indicate empirical return times estimated from observations taken from 7 different datasets that were used for bias correction.

Annual mean fluxes across the large ensemble of NEE, GPP, Reco, and AET are shown in Table 5.2 for the 1986-2010 period for each bias correction and the control simulation. Conventional statistical bias correction that matches monthly means of the HadRM3P ensemble *exactly* to those of the ERA-Interim control climate simulation yields differences in fluxes of -6.6%, -7.5% and -4.7% for GPP, Reco and AET, respectively. Note that differences in the resampled HadRM3P ensemble are less pronounced (-4.2%, -4.5%, and -2.0%, respectively), although no attempt has been made to adjust the statistical properties of the model output. Those differences in simlated annual mean fluxes are related to higher statistical moments of the statistical distributions and shown in Figure 5.6.

To this end, simulated GPP, NEE, and AET show strong asymmetry in their simulated distributions (Figure 5.6): Negative anomalies in GPP and AET are much more pronounced than positive ones; this holds also for NEE but with an inverted sign (ecosystem carbon release corresponds to positive fluxes). However, the simulation of these extremes is highly sensitive to bias correction, where the lower tails of GPP and AET in the original and statistically bias corrected ensemble strongly overestimate reductions in carbon and water flux. In contrast, negative GPP and AET anomalies in the resampled ensemble (corresponding to positive ones in NEE) exhibit a much less pronounced lower tail and asymmetry and agree well with the control simulations.

For example, a positive anomaly in NEE corresponding to a 30-year return period exceeds $+200$ g C m^{-2} year^{-1} in the conventionally bias corrected simulations and the original ensemble, whereas such an anomaly in the resampled ensemble hardly reaches $+150$ g C m^{-2} year^{-1} (Figure 5.6b) roughly corresponding to an empirical 30-year return event in the ERA-Interim control simulations. Similar arguments can be made for negative anomalies in annual GPP and annual AET (Figure 5.6). The different tails of the simulations occur because the original meteorological ensemble implies large hot and dry biases in summer, inducing negative anomalies in ecosystem-atmosphere carbon and water cycling. These biases are not accounted for by conventional statistical bias correction but they are alleviated if an ensemble resampling scheme is used (see previous subsection). However, this is remarkable because monthly means of precipitaton in PRECIPCOR and ISIMIP are identical to the control climate simulation, which highlights the importance to consider statistical moments beyond the mean for impact simulations.

However, note that the positive tails of GPP and AET are not as strongly affected. Furthermore, ecosystem respiratory fluxes show a relatively lower sensitivity to bias correction (i.e. to hot and dry summer conditions).

Further, we investigate whether different bias correction schemes induce different sensitivities of LPJmL simulated carbon fluxes to rainfall. Here, the rela-

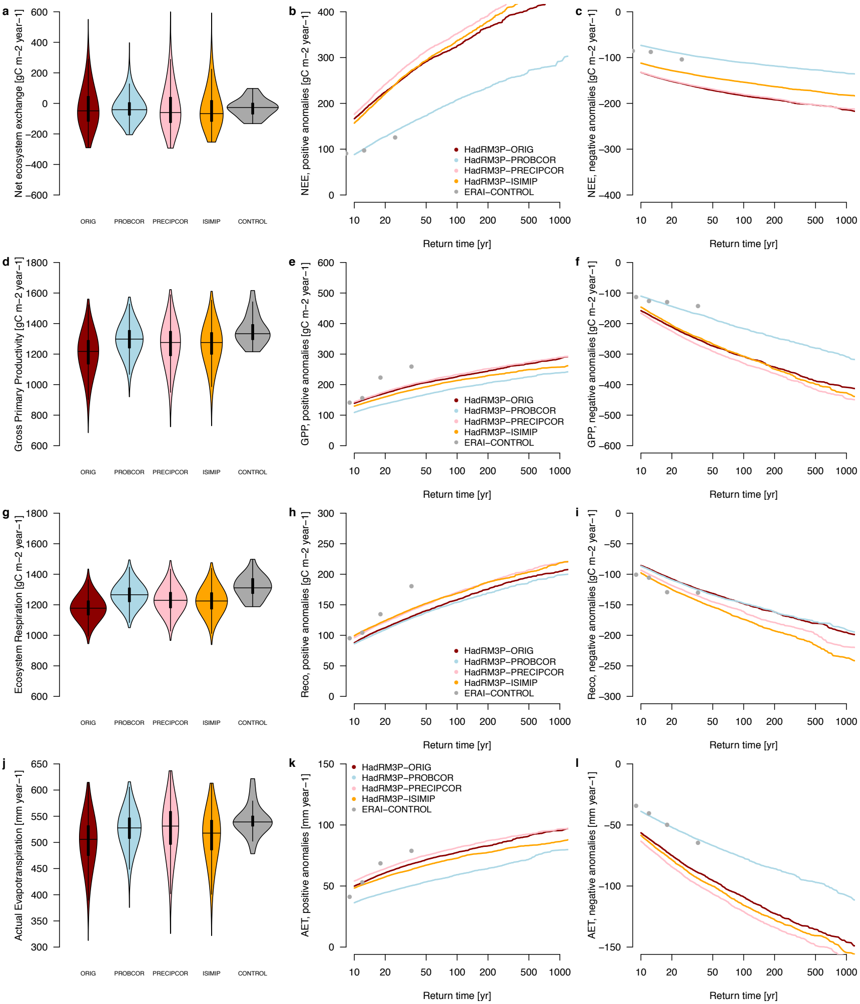


FIGURE 5.6.: LPJmL simulated distributions of ecosystem-atmosphere carbon and water fluxes for Central European natural vegetation for each bias correction scheme. Each row shows the simulated distribution and the upper and lower tail of NEE (a,b,c), GPP (d,e,f), Reco (g,h,i) and AET (j,k,l), respectively. a,d,g,j) Both sides of each violin are constructed as rotated, equal-area kernel density estimates, and a standard boxplot is drawn inside each violin.

tionship between a growing season rainfall proxy (April-September rainfall sums) and annual NEE is characterised using piecewise linear regression (Figure 5.7a-

TABLE 5.2.: Annual mean ecosystem-atmosphere water and carbon fluxes simulated by LPJml.

Bias Correction Method	NEE (g C m^{-2} a^{-1})	GPP (g C m^{-2} a^{-1})	Reco (g C m^{-2} a^{-1})	ET (mm a^{-1})
HadRM3P-ORIG	-26.5	1206.4	1179.9	501.5
HadRM3P-PROBCOR	-30.3	1295.8	1265.5	525.9
HadRM3P-ISIMIP	-31.6	1262.3	1230.7	525.2
HadRM3P-PRECIPCOR	-38.2	1263.2	1225.0	511.2
ERAI-CONTROL	-28.4	1353.3	1324.8	536.7

d). Figure 5.7e shows that LPJmL simulated annual NEE responds to rainfall in a roughly similar way across different bias correction schemes, which highlights the need of an accurate representation of precipitation in climate impact simulations in the terrestrial biosphere. However, characterizing the annual NEE response for each quantile of the rainfall distribution shows that the resampled rainfall distribution (PROBCOR) leads to a less negative NEE response to rainfall (larger slopes in Figure 5.7f), whereas a dry summer tail (in the ORIG, ISIMIP, and PRECIPCOR simulations) yields a generally stronger NEE response (more negative sloped in Figure 5.7f).

In conclusion, different bias correction methods induce different statistical properties of simulated ecosystem-atmosphere fluxes of carbon and water. This affects the variability and skewness of NEE, GPP and AET simulations (as shown in Figure 5.6), where hot and dry biases in summer imply a disproportional reduction in carbon and water fluxes in climatically ‘unfavourable’ years. Conventional statistical bias correction cannot account for this issue, whereas the novel probabilistic bias correction schemes alleviates those biases to a very large extent.

5.5. Discussion

In this paper, we have introduced a novel ensemble-based resampling bias correction approach that retains the physical consistency and multivariate correlation structure of regional climate model output. The approach thus relies on a physically consistent set of climate model simulations (i.e. closure of water and energy balances). The methodology is conceptually similar to earlier approaches designed to constrain future probabilistic climate predictions based on observa-

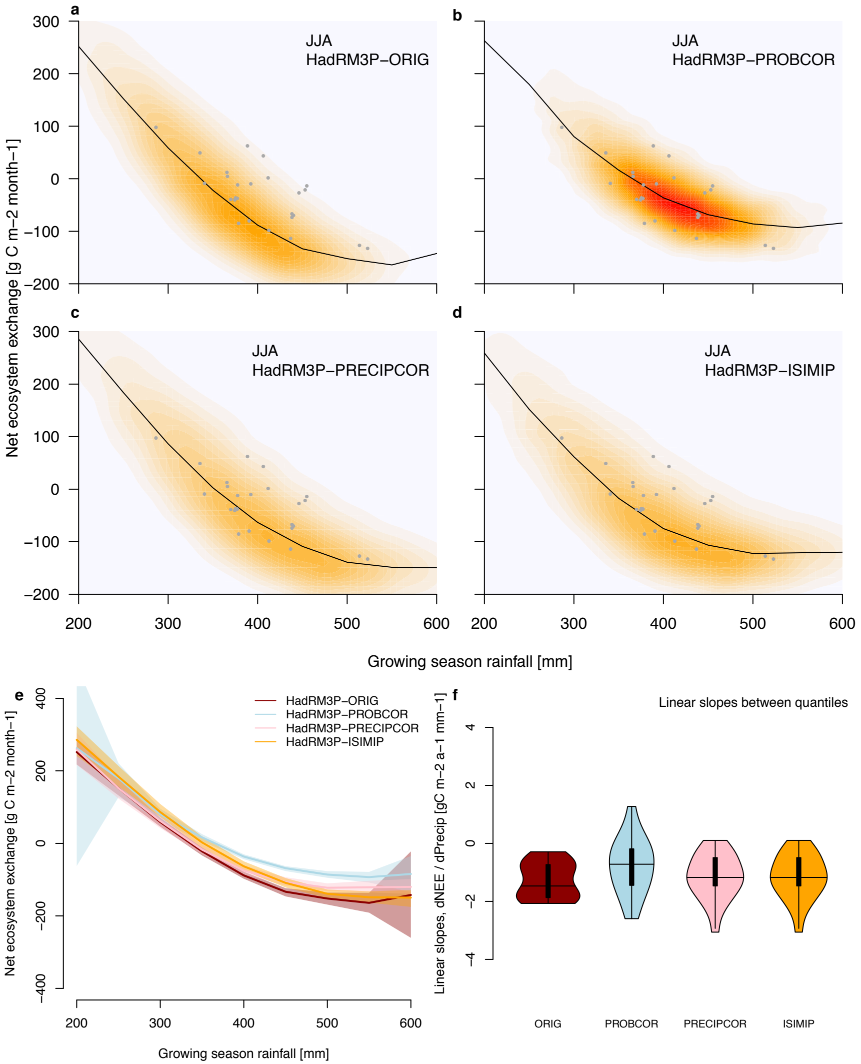


FIGURE 5.7.: a–d) Kernel density plots of the sensitivity of simulated annual NEE to growing season rainfall in LPJmL under four different bias correction schemes. Grey dots denote ERA-Interim control simulations in each plot, black lines indicate piecewise linear regressions. e) Piecewise linear regression relations for each bias correction scheme. Shaded colours indicate confidence intervals (5-95th percentile of piecewise linear regression derived by bootstrapping). f) Distribution of linear regression slopes (dNEE / dRainfall) between regularly spaced quantiles of the rainfall distribution for each bias correction scheme, shown as violin plots.

tional constraints (Piani et al., 2005; Collins, 2007). Its application has been shown in this paper to yield considerably improved simulations of weather and climate extremes. Remarkably, the improvement holds for variables that have not been constrained upon (i.e. constraining on seasonal mean temperatures improves the representation of mean and extreme precipitation), which indeed emphasises the importance to bias correct in a physically meaningful way.

Furthermore, simple but widely used statistical bias correction methodologies (e.g. Hempel et al., 2013) have been evaluated with respect to the effect on the representation of weather and climate extremes on monthly to seasonal time scales. These methods cannot account for biases associated with e.g. specific synoptic situations that result in biases in higher statistical moments of the simulated distributions, which indeed emphasises the importance to bias correct in a physically meaningful way. We demonstrated that this shortcoming of conventional methodologies can be detrimental to statistics of weather and climate extremes and their variability. More sophisticated statistical bias-correction schemes (see Gudmundsson et al., 2012, for an overview) that might have an improved skill in rectifying biases in higher statistical moments (such as e.g. asymmetries in simulated distributions) have not been explicitly tested in this study. However, the fundamental question of how physical consistency can be preserved after bias correction (Ehret et al., 2012), including multivariate dependencies between variables, remains elusive. Therefore non-linear and nonparametric bias correction techniques (Gudmundsson et al., 2012) might potentially improve statistics of extreme events if a large enough sample of observations is available, but cannot retain physical consistency (Sippel and Otto, 2014) and may ultimately fall short for correcting a set of input variables.

To this end, we have explicitly simulated an ensemble of ecosystem-atmosphere fluxes of carbon and water using a state-of-the-art biosphere model (LPJmL) in order to test the sensitivity to bias correction. Similarly to above, we find that bias correction induces strong effects on the representation of extremes and variability in carbon and water fluxes (Section 5.4.3). Mechanistically, the stark contrast between the bias correction schemes can be traced back to the sensitivity of the LPJmL model to dry conditions (see e.g. Rammig et al., 2015; Rolinski et al., 2015): NEE, GPP and AET in Central Europe are to a large extent driven by the availability of rainfall in the growing season, except for wet conditions, under which the relationship levels off (Figure 5.7). Bias correction strongly affects the variability and extremes of rainfall (as shown above), thus inducing pronounced asymmetries in simulated water and carbon fluxes (Figure 5.7f, Figure 5.6). Therefore, our results highlight the importance to account not only for biases in the mean but also for higher moments in the climatic input in order to generate robust insights into the past, present and future climate impacts. Our results demonstrate that physically consistent bias correction schemes might be preferable for this task. Moreover, it has been shown recently that climatic drivers exert multivariate controls on ecosystem responses such as phenology and vegetation greenness dynamics (Forkel et al., 2015), therefore accurate ecosystem impact simulations requires bias correction schemes that preserve the correlation structure of climatic data.

However, several limitations of the present methodology should be discussed: First, probabilistic resampling based on a regional observational constraint cannot account for biases on very large regional or continental scales if the biases show a spatially or temporally heterogenous structure or gradients. In the latter case, resampling-based bias correction could lead to spurious artefacts in the spatio-temporal structure of the bias-corrected model domain. Hence, a careful evaluation of the ensemble resampling approach has to be made - particularly with a focus on the spatial and temporal extent of the constraint and the resampled ensemble: A trade-off exists between resampling on small domains (e.g. grid-cell based) that is sensitive to the choice of observational dataset, and very large domains that might be prone to a spatio-temporal bias structure. Secondly, the resampling approach requires relatively large ensemble sizes to be effective: in order to plausibly cover the climate space in any particular location, the simulated ensemble should cover the entire observed distribution. However, this condition does not necessarily restrict resambling-based bias correction methods to large ensemble simulations: For example, under the assumption of ergodicity for a given time period, resampling shorter time periods (e.g. single years) from smaller ensembles such as CORDEX regional simulations (Giorgi et al., 2009) would provide a promising topic for further study. In this context, the applicability of the resampling methodology would depend on the remaining effective sample size after the resampling step. The latter is a function of the biases in the model and the number of ensemble members available, and could be tested in an evaluation step similarly to Figure 5.2d. Thirdly, the applicability of bias correction methods for future projections is currently unclear, since previous studies have shown that biases in climate projections (e.g. for the 21st century) might not be stationary (Ehret et al., 2012; Maraun, 2012). However, an application of the resampling approach to future projections similarly to the current practice of statistical bias correction (Hempel et al., 2013, e.g.) would be straightforward, i.e. based on a calibration using present or past conditions. Lastly, a clear distinction between bias correction and statistical downscaling is crucial (Maraun, 2013): While the resampling bias correction is designed to account for the former, no attempt of statistical downscaling or bridging any scale mismatches is made (see, e.g. Maraun, 2013, for a detailled discussion).

Notwithstanding these limitations however, we show the usefulness of the novel bias correction scheme that might be a useful and physically consistent alternative to conventional statistical bias correction as long as global and regional dynamical climate models suffer from pertinent biases.

5.6. Conclusions

In this paper, we introduced a novel bias correction method that retains physical consistency and the multivariate correlation structure of the climate model output based on an ensemble resampling approach. We showed that such an approach strongly improves

a) statistics of weather and climate extreme events, and

b) the simulation of climate impacts such as ecosystem-atmosphere fluxes of carbon and water, including extremes and variability therein.

The methodology could be readily taken up in probabilistic event attribution studies that deploy large ensembles simulations (see Stott et al., 2013, for an overview) in order to more realistically describe the statistics of (changing) extreme events.

Furthermore, detecting and attributing the impacts of climatic variability and extremes on hydrological and socio-ecological systems has emerged as a highly topical research area (Stone et al., 2009, 2013), including demonstrated interest by stakeholders across various sectors (Schiermeier, 2011; Stott and Walton, 2013; Sippel et al., 2015c). To this end, our study showed that it is crucial to account for higher statistical moments in biased climatic input data, and to correct climatic biases in a physically consistent way. Therefore, our methodology could be taken up by the climate impact modelling community to reduce climate forcing biases to a very large extent without requiring any modifications to the climate model output.

6. The role of anthropogenic warming in 2015 Central European heat waves[1,2,3]

Abstract

Station-based observations and bias-corrected model simulations show that the frequency of short-term heat waves in Central Europe has increased, albeit quantitative estimates of risk ratios differ considerably between methods.

6.1. Summer 2015 in Europe

The summer 2015 in Europe was highly unusual, as persistent heat and dryness prevailed in large parts of the continent. In Central and Eastern Europe, a combination of record-low seasonal rainfall (Orth et al., 2016) and record-high monthly July/August temperatures were observed over an area stretching from France to Western Russia (Figure D1). The anomalous temperatures were caused by a sequence of four intense heat waves that struck the region from the end of June to early September (e.g. Figure 6.1a). It is precisely the few-day heat that causes problems with human health, especially when combined with high humidity (McGregor et al., 2010). Here we analyse seasonal maxima of 3-day mean temperature ($\mathrm{Tair}_{\mathrm{3d,\ max}}$) and seasonal maxima of 3-day daily maximum wet bulb temperature ($\mathrm{WBTX}_{\mathrm{3d,\ max}}$), a measure of human thermal discomfort that combines

[1]This chapter is published as Sippel, S., F. E. L. Otto, M. Flach, and G. J. van Oldenborgh. 2016. In Herring, S. C., Hoell, A., Hoerling, M. P., Kossin, J. P., Schreck III, C. J., and Stott, P. A. (Eds.), Explaining Extremes of 2015 from a Climate Perspective. *Bulletin of the American Meteorological Society*, **97**(12), S51–S56. doi:10.1175/BAMS-D-16-0149.

[3]Supplementary material that complements this Chapter with more detailed explanations is available in Appendix D.

temperature and humidity and is a proxy for heat stress on the human body (Fischer and Knutti, 2013; Sherwood and Huber, 2010).

The series of heat waves began with a strongly meandering jetstream, i.e. summertime 'omega-blocking' (Dole et al., 2011), and the advection of very warm subtropical air into Central and Western Europe (Figure D1). Later in the season, the jetstream was displaced to the north, so that stable high-pressure systems could prevail over Central and East Europe bringing heat there. The first heat wave in early July was hence most pronounced in western parts of the continent, while South-Central and East-Central Europe experienced the highest temperatures in the subsequent heat waves later in the season (Figure 6.1b).

Anomalies in the hottest 3-day mean temperature reached up to +6°C relative to climatology (Figure 6.1c,d) and temperature records tumbled: This included nation-wide records[4] (Kitzingen, Germany: 40.3°C), various station-records stretching from France to the Balkan countries and Southern Sweden[5], night-time temperatures (Vienna, Austria: 26.9°C), record 3-day mean temperatures across Central Europe (Figure 6.11e), and inland water temperatures (e.g. Lake Constance). Europe experienced the hottest August ever recorded (NOAA National Centers for Environmental Information, 2016), and the entire summer season ranked 3rd after the unusual summers of persistent heat in 2003 and 2010 with their hotspots in France and Western Russia, respectively (Barriopedro et al., 2011; Stott et al., 2004). This extraordinary sequence of events raises the question to what extent human-induced climate change played a role in short-term heat waves beyond natural climate variability.

A potential anthropogenic contribution to the summer 2015 heat events had already been investigated in near real-time[6], and in the present paper we build upon and substantiate the previous analysis: We investigate two diagnostics ($\mathrm{Tair}_{3d,\,max}$ and $\mathrm{WBTX}_{3d,\,max}$) at four locations in long-term station-based observational records and in a large ensemble of consistently bias-corrected regional climate model simulations.

[4] https://weather.com/news/climate/news/europe-heat-wave-poland-germany-czech-august-2015
[5] http://www.meteofrance.fr/actualites/26913226-episode-de-tres-fortes-chaleurs-en-france
[6] http://www.climatecentral.org/europe-2015-heatwave-climate-change

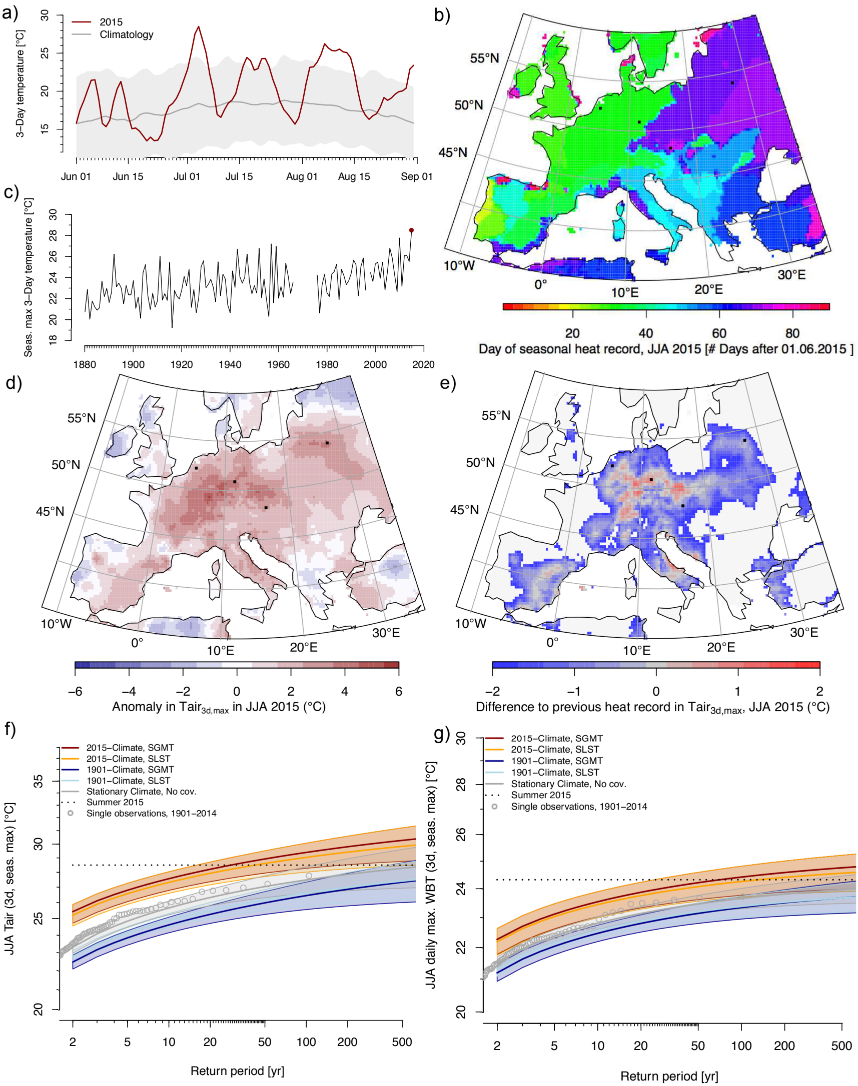


FIGURE 6.1.: a) Time series of 3-daily mean temperatures in summer 2015 at the Jena site (grey shading denotes $\pm 2\sigma$ deviations relative to long-term inter-annual variability). b) Day of seasonal temperature record in summer 2015. Full caption is continued on the following page.

FIGURE 6.1.: (continued) c) Annual time series of seasonal maximum of 3-day mean temperatures ($Tair_{3d,\,max}$) at the Jena site (summer 2015 is marked by a red dot). d) Anomalies in Tair3d,max over Europe in summer 2015 relative to 1981-2010. e) Difference to previous heat records (1950-2014) in $Tair_{3d,\,max}$ in the EOBS dataset. Positive differences indicate a new heat record in JJA 2015. f, g) Return time plots of GEV fits for $Tair_{3d,\,max}$ and $WBTX_{3d,\,max}$, respectively, at the Jena site. Red (orange) lines indicate the fit for 2015 climate, darkblue (lightblue) lines indicate the fit for 1901 climate for a smoothed global mean temperature covariate (smoothed local summer temperature covariate).

6.2. Methods and data

First, we analyse long-term observational data (115 years of data for each station) from the ECA&D dataset (Klein Tank et al., 2002) of four Central and East European stations that were affected by the heat waves in summer 2015 (Table 6.1), using data from 1901 onwards. For each station, annual time series of $Tair_{3d,\,max}$ and $WBTX_{3d,\,max}$ are calculated for July-August. $WBTX_{3d,\,max}$ is derived from daily maximum air temperature and vapour pressure (computed from relative humidity and daily mean temperature) using an iterative procedure based on the psychrometric equation[7] (Sullivan and Sanders, 1974). Subsequently, generalised extreme value (GEV) statistical models are fitted to the data (Coles et al., 2001) excluding the year 2015, using two different assumptions about changes in climate:

i. A 'local' station-based covariate to the location parameter of the GEV (21-year smoothed local summer temperatures, SLST) as a proxy for any changes to local climate;

ii. A 'global' covariate to the location parameter (21-year smoothed global mean temperatures, SGMT) as a proxy for anthropogenic influence on climate (Van Oldenborgh et al., 2012).

To avoid over-fitting the relatively low number of data points, no dependence in the scale or shape parameter is assumed. Probability ratios based on the GEV as a metric to quantify human-induced change in the odds of extreme events ($PR =$

[7] `http://www.srh.noaa.gov/epz/?n=wxcalc_rh`

$\frac{p_{ANT}}{p_{NAT}}$, Fischer and Knutti (2015)) were obtained by calculating the probability of an event as warm or warmer than the observed 2015-event in a 2015-climate (p_{ANT}), and in 1901 as a proxy for pre-industrial climate.

Second, a model ensemble-based assessment using the global general circulation model HadAM3P (1.875° x 1.25° x 15min resolution) and a dynamically downscaled regional variant (HadRM3P, 0.44° x 0.44° x 5min resolution) is conducted to complement the empirical analysis (see Massey et al., 2015, for all details regarding the model setup). Initial condition ensembles are generated for an 'anthropogenic scenario' (ANT, n = 2286), in which the model is driven by observed (2015) sea surface temperatures (SST) and anthropogenic forcings in atmosphere-only mode for one year at a time (starting December 1st, Massey et al. (2015)); and a 'natural scenario' (NAT, n = 4414) with all anthropogenic forcings (i.e., greenhouse gases, aerosols, halocarbons and ozone) set to pre-industrial levels and 11 different estimates of 'natural' SSTs (Schaller et al., 2014). For each of the four locations (centred over a 1°x1° grid cell), a resampling bias correction strategy based on an observational constraint is applied to the model ensemble (Sippel et al., 2016a), because the raw model output is notoriously too hot and dry (Black et al., 2015; Massey et al., 2015) severely compromising attribution statements (Figure D2). The seasonal maximum 21-day average temperature from the E-OBS dataset (Haylock et al., 2008) is used as resampling constraint and a percentile-based transfer function is calibrated for each station separately on the 1986-2010 climatology using an identical model setup (Massey et al., 2015). Subsequently, both 'natural' and 'anthropogenic' simulations are resampled using the derived relationship (Sippel et al., 2016a). In contrast to widely used methods like quantile-quantile mapping, resampling retains the full multivariate structure and physical consistency of the model output, but reduces the available ensemble size and chooses colder and wetter ensemble members therefore alleviating the hot and dry bias (Sippel et al., 2016a). In the context of event attribution it is applied for the first time in this paper (Figure 6.2a-d, see next section). To avoid potential mean biases due to station location, the mean of the resampled ensemble is adjusted to the station mean (Figure D2c-d). Results are demonstrated exemplarily for one station (Jena), and probability ratios are reported for all stations.

6.3. Results and discussion

The statistical analysis of estimated return times of $\text{Tair}_{\text{3d, max}}$ reveals that 2015-like heat events occur in present day climate approximately every 27 years in Jena with the one-sided 5% lower confidence bound at 16 years (Figure 6.1). Including both the local and global climate change covariates into the GEV fit demonstrates a profound increase in return times of those types of events relative to earlier years for both $\text{Tair}_{\text{3d, max}}$ and $\text{WBTX}_{\text{3d, max}}$ in Jena (Figure 6.1f,g) and all other locations with probability ratios typically exceeding a value of ten (Table 6.1). The intensity of heat waves increases by about 3 degrees in $\text{Tair}_{\text{3d, max}}$ but only 1.1K in $\text{WBTX}_{\text{3d, max}}$ (Figure 6.1f,g). In spite of this difference, the increase in the probability ratio is similar.

A similar analysis is conducted in a very large ensemble of model simulations. The 21-day resampling constraint considerably improves the representation of short-term heat waves by avoiding physically implausible simulations (Figure 6.2a-d) and improving the simulated variability of heat waves (Figure D2c,d). The correlation structure between the temperature constraint and short-term heat stress ($\text{WBTX}_{\text{3d, max}}$) in the observations is reproduced in the resampled model ensemble, but not in the original model ensemble (Figure 6.2a,c). This indicates that robust attribution statements for impact-related, and thus multivariate quantities (such as $\text{WBTX}_{\text{3d, max}}$) require a physically consistent bias correction of model output.

Consistent with the observations, the model-based assessment shows a shift in the return periods towards more frequent and more pronounced summer heat stress (Figure 6.2b) in all locations (Table 6.1) and both bias-corrected and original simulations. The probability ratios derived from the bias-corrected model ensembles range from 1.1 to 2.9 ($\text{Tair}_{\text{3d, max}}$) for the four locations (PR=1.3-3.1 for $\text{WBTX}_{\text{3d, max}}$ in Jena and De Bilt), depending on the magnitude of the 2015-event, the model-simulated warming and inter-annual variability. These estimates are thus lower than those estimated from the observations, but can be largely explained by method- and data-related differences: For instance, the statistical method assumes that the trend is caused fully by anthropogenic factors, while the model analysis is based on a ‘real counterfactual’ scenario but tends to underestimate warming trends in temperature extremes in Europe (Min et al.,

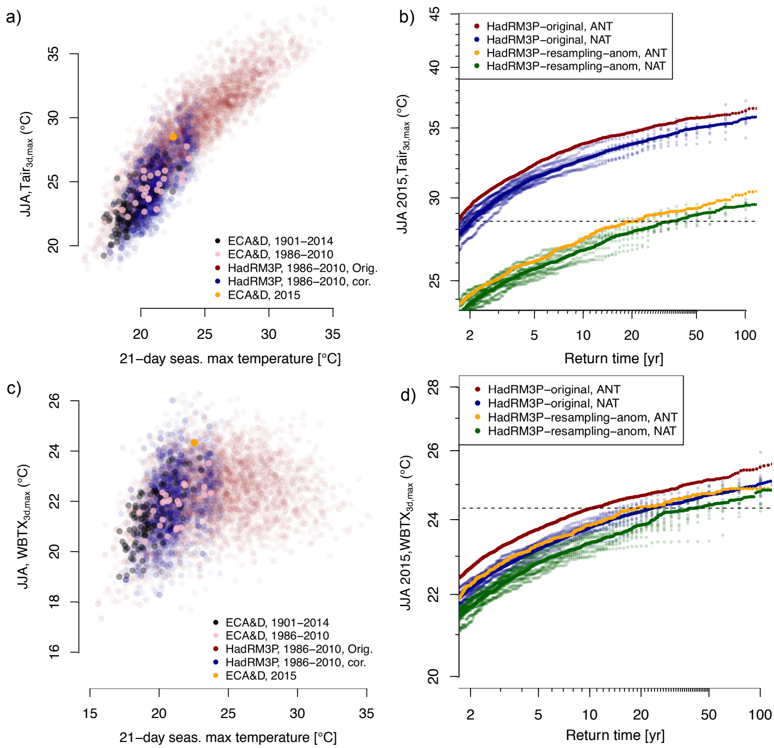


FIGURE 6.2.: a,c) Correlation between 21-day seasonal maximum temperature (observational constraint for resampling bias correction) and impact-related quantities ($\mathrm{Tair}_{\mathrm{3d,\,max}}$ and $\mathrm{WBTX}_{\mathrm{3d,\,max}}$, respectively). Pink dots correspond to 1986-2010, the period used for calibration of the bias correction resampling function. b,d) Return time plots for original and bias-corrected model output for $\mathrm{Tair}_{\mathrm{3d,\,max}}$ and $\mathrm{WBTX}_{\mathrm{3d,\,max}}$, respectively.

2013). The mean observed change across all locations between 2015 and 1901 of 3.1K ($\mathrm{Tair}_{\mathrm{3d,\,max}}$) and 2.2K ($\mathrm{WBTX}_{\mathrm{3d,\,max}}$) is much larger than in the original (+1.1K in $\mathrm{Tair}_{\mathrm{3d,\,max}}$ and +0.5K in $\mathrm{WBTX}_{\mathrm{3d,\,max}}$) and bias corrected (+0.9K in $\mathrm{Tair}_{\mathrm{3d,\,max}}$ and +0.5K in $\mathrm{WBTX}_{\mathrm{3d,\,max}}$) model simulations. Hence, replacing the model-simulated warming by the observed change between 1901 and 2015 causes roughly a tripling of probability ratios for the bias-corrected simulations at all locations (e.g. 3.4-8.7 for $\mathrm{Tair}_{\mathrm{3d,\,max}}$, and 2.7 to exceeding 10 for $\mathrm{WBTX}_{\mathrm{3d,\,max}}$, cf. Table 6.1). Furthermore, uncertainties due to event selection (Christiansen, 2015), dependence on the spatial and temporal scale (Angélil et al., 2014), high non-linearity in attribution metrics such as the probability ratio (Figure D2), and a slightly higher variability on sub-monthly time scales in the model simulations than in the observations despite bias correction further contribute to model-data discrepancies and variability in the presented estimates of the probability ratios.

6.4. Conclusion

In conclusion, the multi-method analysis applied in this paper provides consistent evidence that human-induced climate change has contributed to the increase in the frequency and intensity of short-term heat waves and heat stress such as the Central and East Europe 2015 event. However, quantitative estimates of the risk ratio at local scales can differ widely depending on the exact methodologies applied, thus highlighting large method- and data-related uncertainties. In this study, due to the large discrepancy between observed and modelled trends in temperature extremes the model-estimated probability ratios are lower than those estimated from the observations.

TABLE 6.1.: Location of meteorological stations and probability ratios estimated from observed and simulated data. Very large PR with a lower bound (5% confidence interval) exceeding 10 are reported as > 10. PR from the model output are given as 5th to 95th percentile of 100 bootstrapped replicates (n = 1000). A PR range exceeding one would be significant at 95% confidence under a one-sided test. PR for the original model simulations (i.e., non-bias corrected) are indicated for comparison only.

Station	De Bilt[a]	Jena	Minsk	Vienna
Country	Netherlands	Germany	Belarus	Austria
Location	52°06'N, 5°11'E	50°55.5', 11°35'E	53°52'N, 27°32'E	48°14'N, 16°21'E
$Tair_{3d,\,max}$, 2015 (°C)	25.2	28.5	27.3	29.1
$PR_{HadRM3P,\,BC\text{-}anom}$	1.2-1.4	1.1-2.5	1.7-2.5	1.8-2.9
$PR_{HadRM3P,\,BC\text{-}anom}$, obs. trend	4.7-7.5	4.1-8.7	3.4-5.2	>10
PR_{EOBS}, GEV-GMT	>10	>10	>10	>10
$WBTX_{3d,\,max}$ (2015, °C)	22.9	24.3	n.a.[b]	n.a.[b]
$PR_{HadRM3P,\,BC\text{-}anom}$	1.3-1.8	1.5-3.1	n.a.[b]	n.a.[b]
$PR_{HadRM3P,\,BC\text{-}anom}$, obs. trend	>10	2.7-7.7	n.a.[b]	n.a.[b]
PR_{EOBS}, GEV-GMT	>10	>8.6	n.a.[b]	n.a.[b]

a. The observed De Bilt series contains a well-known inhomogeneity in 1950, so the homogenised series from KNMI was used instead.
b. Humidity data was not available for Vienna and Minsk in the ECA&D dataset for the year 2015.

7. Refining multi-model projections of temperature extremes by evaluation against land-atmosphere coupling diagnostics[1,2]

Abstract

The Earth's land surface and the atmosphere are strongly interlinked through the exchange of energy and matter. This coupled behaviour causes various land-atmosphere feedbacks, and an insufficient understanding of these feedbacks contributes to uncertain global climate model projections. For example, a crucial role of the land surface in exacerbating summer heat waves in mid-latitude regions has been identified empirically for high-impact heat waves, but individual climate models differ widely in their respective representation of land-atmosphere coupling. Here, we compile an ensemble of 54 combinations of observations-based temperature (T) and evapotranspiration (ET) benchmarking datasets and investigate coincidences of T anomalies with ET anomalies as a proxy for land-atmosphere interactions during periods of anomalously warm temperatures. First, we demontrate that a large fraction of state-of-the-art climate models from the Coupled Model Intercomparison Project (CMIP5) archive produces systematically too frequent coincidences of high T anomalies with negative ET anomalies in mid-latitude regions during the warm season and in several tropical regions year-round. These coincidences (high T, low ET) are closely related to the representation of temperature variability and extremes across the

[1]This chapter is published as Sippel, S., J. Zscheischler, M. D. Mahecha, R. Orth, M. Reichstein, M. Vogel, and S. I. Seneviratne. 2016. *Earth System Dynamics* **8**, 387–403. doi:10.5194/esd-8-387-2017.

[2]Supplementary Online Material (SOM) that provides additional information but that is not necessary for understanding the scientific content of this Chapter is available under `http://www.earth-syst-dynam.net/8/387/2017/esd-8-387-2017-supplement.pdf`

multi-model ensemble. Second, we derive a land-coupling constraint based on the spread of the T-ET datasets and consequently retain only a subset of CMIP5 models that produce a land-coupling behaviour that is compatible with these benchmark estimates. The constrained multi-model simulations exhibit more realistic temperature extremes of reduced magnitude in present climate in regions where models show substantial spread in T-ET coupling, i.e. biases in the model ensemble are consistently reduced. Also the multi-model simulations for the coming decades display decreased absolute temperature extremes in the constrained ensemble. On the other hand, the differences between projected and present-day climate extremes are affected to a lesser extent by the applied constraint, i.e. projected changes are reduced locally by around 0.5°C to 1°C - but this remains a local effect in regions that are highly sensitive to land-atmosphere coupling. In summary, our approach offers a physically consistent, diagnostic-based avenue to evaluate multi-model ensembles, and subsequently reduce model biases in simulated and projected extreme temperatures.

7.1. Introduction

The exchange of matter and energy between the land surface and the atmosphere is a crucial feature of the Earth's climate (Seneviratne et al., 2010a; Bonan, 2015; van den Hurk et al., 2016). On one hand, the atmosphere exerts a key influence on land surface processes such as vegetation growth by supplying light, water and carbon dioxide (Köppen, 1900). On the other hand, the land surface feeds back to the atmosphere, for example through the partitioning of energy into latent and sensible heat fluxes, or by modifying land surface properties, thus implying a direct link to near-surface climate (Koster et al., 2004; Seneviratne et al., 2010a). Conceptually, coupling between the atmosphere and the land surface is often classified into two qualitatively different regimes, a so-called 'energy-limited' and 'water-limited' regime (Seneviratne et al., 2010a): In the wet (energy-limited) regime, the land surface is largely controlled by the atmosphere through radiation (see conceptual Figure 7.1a,b), implying a positive association between near-surface temperature (T) and evapotranspiration (ET). In contrast, in a dry, water-limited state, the land controls near-surface climate through a lack of soil moisture, and a corresponding reduction in evapotranspiration and latent cooling (see conceptual Figure 7.1a,b) with a negative association between T and ET. Therefore, the state of the land surface and land-atmosphere feedbacks modulate and amplify climatic extreme events such as heat waves in mid-latitude regions (Seneviratne et al., 2006; Fischer et al., 2007; Hirschi et al., 2011; Whan et al., 2015; Hauser et al., 2016). An understanding of these feedbacks might yield improved seasonal predictability of extremes (Quesada et al., 2012), and could help to constrain and better predict model-simulated present and future climate variability in these regions (Seneviratne et al., 2006; Lorenz et al., 2012; Dirmeyer et al., 2013; Seneviratne et al., 2013; van den Hurk et al., 2016; Davin et al., 2016).

However, at present large uncertainties and methodological inconsistencies prevail in both understanding and quantification of land-atmosphere coupling at various spatial and temporal scales, which relate to

i. scarcity of accurate observational products of soil moisture or evapotranspiration at large spatiotemporal scales and relatively short observational periods (Seneviratne et al., 2010a),

ii. the metrics and variables used to quantify land-atmosphere coupling differ widely in the variables they address (Seneviratne et al., 2010a), and in emphasizing either the whole distribution (Dirmeyer, 2011; Lorenz et al., 2012; Miralles et al., 2012), or the tails of relevant variables (Zscheischler et al., 2015).

As a consequence, uncertainties and methodological inconsistencies contribute to a greatly diverging representation of land-atmosphere coupling in state-of-the art climate models (Koster et al., 2004; Boé and Terray, 2008, see also Figure 7.1a,b for a simple conceptual example), and further contribute to uncertainties related to projected increases in summer temperature variability in the 21st century in mid-latitude regions (Seneviratne et al., 2006; Dirmeyer et al., 2013). In this context, it has been noted that accurate simulations of temperature variability and extremes require a realistic representation of land-atmosphere interactions (Seneviratne et al., 2006; Fischer et al., 2012; Bellprat et al., 2013). In other words, biases in temperature variability and extremes might in part stem from an unrealistic representation of land-atmosphere interactions (Fischer et al., 2012; Lorenz et al., 2012; Davin et al., 2016), likely leading to temperature-dependent biases in multi-model ensembles (Boberg and Christensen, 2012; Bellprat et al., 2013).

A model evaluation focus on interpretable land-atmosphere coupling diagnostics might serve as a complementary strategy to traditional model validation and testing (Seneviratne et al., 2010b; Santanello et al., 2010; Mueller et al., 2011b; Mueller and Seneviratne, 2014). Hence, this approach is intended towards testing and understanding the spread and physical consistency in simulated relationships in state-of-the-art multi-model ensembles (e.g. the Coupled Model Intercomparison Project, CMIP5 Taylor et al., 2012) against available observations-based datasets. For example, in the context of land-atmosphere coupling, earlier studies used bivariate correlation- or regression-based metrics to test and evaluate coupling behaviour (Hirschi et al., 2011; Lorenz et al., 2012). Conceptually, the notion of 'diagnostic-based model evaluation' as discussed here is consistent with so-called 'pattern-oriented model evaluation' (Grimm and Railsback, 2012; Reichstein et al., 2011) - the latter being applied in the context of evaluating simulated and observed patterns at multiple scales in a data-driven way (e.g. in the context of ecosystem carbon turnover times, Carvalhais et al., 2014).

In the context of extracting credible and relevant information from large (multi-)model ensembles, weighting or selecting models based on observations-based constraints has become increasingly popular recently (Tebaldi and Knutti, 2007; Knutti, 2010), as a priori model ensembles might be seen as a somewhat arbitrary collection of model runs (or 'ensembles of opportunity'). For example, empirical and/or physics-based criteria have been used to constrain snow-albedo feedbacks (Hall and Qu, 2006), constrain carbon cycle projections (Cox et al., 2013; Wenzel et al., 2014; Mystakidis et al., 2016), or in the context of refining precipitation projections (Orth et al., 2016). Moreover, empirical diagnostics are applied to select models for event attribution analyses (Perkins et al., 2007; King et al., 2016; Otto et al., 2015) and analyses of drought projections based on model performance (Van Huijgevoort et al., 2014), or to resample large initial-condition ensembles to alleviate biases without distorting the multivariate structure of climate model output (Sippel et al., 2016a). In the context of land-atmosphere coupling, Fischer

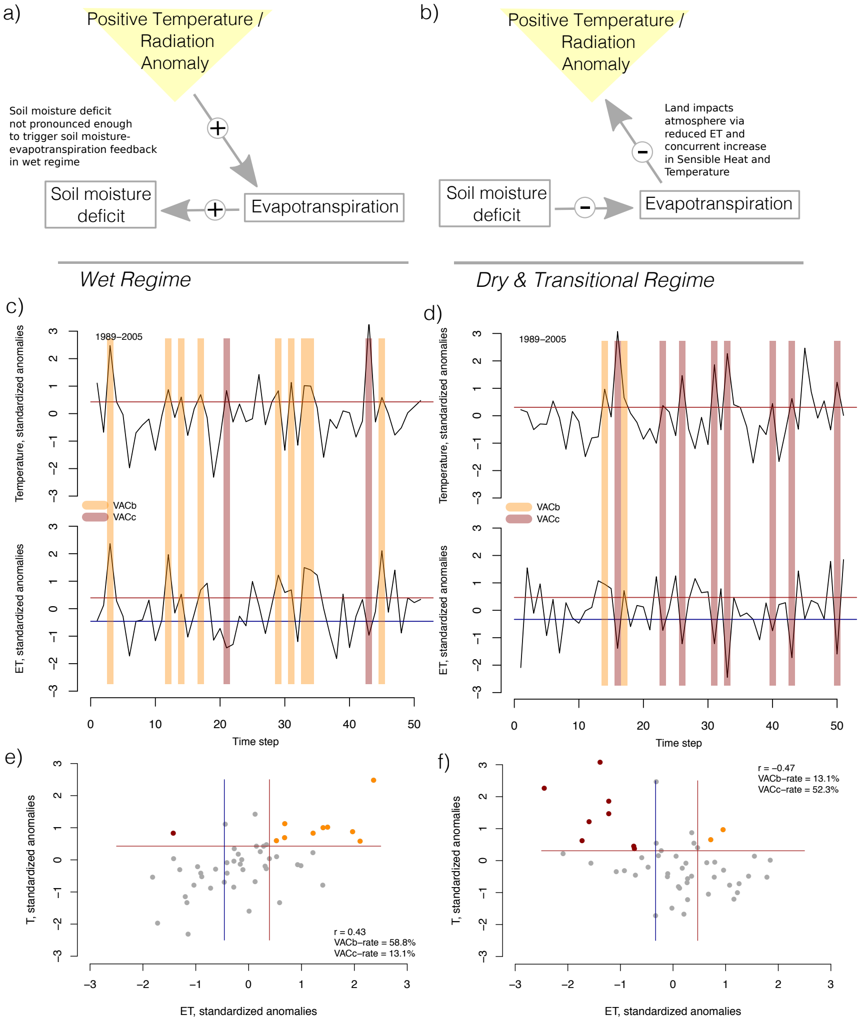


FIGURE 7.1.: Illustration of qualitatively contrasting warm season temperature-evapotranspiration (T-ET) coupling in global climate models. a, b) Conceptual illustration of T-ET coupling in (a) wet, and (b) dry & transitional regimes. In wet regimes T and ET are positively associated (atmosphere impacts land), while in dry & transitional regimes T and ET are negatively associated due to soil moisture feedbacks (i.e., land impacts atmosphere via reduced ET amd concurrent increases in sensible heat and T). Full caption is continued on the following page.

FIGURE 7.1.: (continued) c–f) Different CMIP5 models show contrasting T-ET coupling behaviour in a mid-latitude region in summer (Central Europe, spatial average, JJA, 1989-2005): (c,e) NorESM1-M produces predominantly wet regimes, i.e. a positive T-ET coupling, while (d,f) ACCESS1-3 produces predominantly dry regimes (negative T-ET coupling), illustrated as time series (c-d) and in the T-ET plane (e-f). Red lines in (c-f) indicate th_{upper} for T and ET, blue lines indicate th^{T}_{lower} (70th and 30th percentile in each individual time series, respectively).

et al. (2012) and Stegehuis et al. (2013) have constrained a regional model ensemble over Europe using present-day interannual variability of summer temperature, and observations-based estimates of summer sensible heat fluxes. However, these studies came to somewhat conflicting results with respect to the obtained change in warming projections, which probably was due to the underlying choices of datasets to obtain the constraints (Stegehuis et al., 2013). Hence, care is needed in that these practices might not necessarily translate into improved future climate projections or reduced uncertainties. That is because the selection of relevant metrics is clearly not trivial but subjective, and because good model performance w.r.t. any given metric does not translate directly into (more) reliable projections (Knutti, 2008).

Therefore, the starting point for the present analysis -in the sense of being necessary, but not sufficient to assure reliability of future climate projections- is that physically motivated, observations-based diagnostics might offer

1. a link to identify and interpret relevant processes across multiple models (i.e., model evaluation), and

2. to reduce biases by focusing the interpretation of multi-model ensembles on models that are ‘right for the right reasons’. Most notably climate impacts, including extremes, typically depend on the multivariate structure of climate variables, where simple univariate statistical bias correction methods are prone to failure (Ehret et al., 2012; Cannon, 2016).

In this study, we first evaluate land-atmosphere coupling in state-of-the-art global climate models from the CMIP5 archive and a large ensemble of observations-based ET datasets (Mueller et al., 2013) that has been compiled to address the aforementioned uncertainties in land-atmosphere coupling. In our analyses a land-atmosphere coupling metric that is based on coincidences of temperature and evapotranspiration anomalies is applied. The idea behind a coincidence metric as opposed to a traditional univariate evaluation of model simulated ET fluxes or temperature is that it is insensitive to biases in the simulated means or variances, and thus focusses only on an abstract property of the data, namely the bivariate dependence structure of T and ET. Secondly, we derive a model constraint based on the physically motivated land-coupling diagnostic and the ensemble of benchmarking datasets in order to explore the implications of a reduced ensemble but with land-atmosphere coupling that is *within the range* of the benchmarking datasets.

7.2. Data and methods

7.2.1. Datasets for T-ET coupling analysis and model evaluation

Global temperature and evapotranspiration datasets In order to evaluate T-ET coupling in global climate models, an ensemble of 18 gridded ET estimates, taken from the LandFlux-EVAL multi-data set synthesis project (Mueller et al., 2013), are combined with three different observations-based and reanalysis-driven temperature datasets, yielding in total 54 T-ET combinations (see Table 7.1). T-ET coincidence rates are calculated from each of those 54 combinations to evaluate and constrain the multi-model ensemble of global climate models (Section 3). The ensemble of ET reference datasets has been generated by combining a wide range of different ET estimates, consisting of five diagnostic (based on remote sensing or in-situ observations) products, five land surface models driven by observed climate forcing and four reanalysis products (Mueller et al., 2013). The three temperature datasets are based on one observational product (Climate Research Unit dataset, Harris et al., 2014) and two reanalysis products (ERA-Interim reanalysis (ERAI, Dee et al., 2011), and Climate Forecast System Reanalysis (CFSR, Saha et al., 2010), see Table 7.1 for details). The large number of T-ET dataset combinations is used in order to take uncertainties in both T- and ET datasets into account. We have tested that the spread between individual ET datasets is substantially larger than the spread between individual

T datasets (not shown). This indicates that the largest source of uncertainty stems from the choice of ET dataset, and therefore we consider only three different T datasets. Each of the 54 T-ET dataset combinations (denoted as ‘T-ET coupling benchmarks’ in the remainder of the paper) is consistently derived from observations, and thus can be expected to represent relevant features in T-ET coupling under different assumptions that underlie diagnostic datasets, reanalyses and land surface models. Therefore, these datasets represent a very large spread of plausible T-ET coupling estimates, and the spread can be considered as a conservative benchmark for model evaluation (including observational noise, i.e. allowing a wide range of T-ET coupling in models). However, it should be emphasised that the datasets are not independent realisations. Thus, we only use the spread of the T-ET coupling benchmarks, but we do not interpret the probability distribution of dataset combinations.

For the analysis of historical and future simulations of the monthly maximum value of daily maximum temperatures (TXx) in Section 3.2 we use ERA-Interim (Dee et al., 2011) as a reference dataset.

Multi-model ensemble simulations The Climate Model Intercomparison Project (CMIP5) has been designed to allow for multi-model comparison and evaluation studies (Taylor et al., 2012). Although large model spread, biases and uncertainties remain in the ensemble projections (Knutti and Sedláček, 2013), for example with respect to extremes (Sillmann et al., 2013b), the water (Mueller et al., 2011b; Mueller and Seneviratne, 2014), and land carbon cycle (Anav et al., 2013), the archive of standardised scenario-driven model experiments provides one of the main avenues to study climate variability and change (e.g. (Stocker et al., 2013)), including present and future climate extremes (Sillmann et al., 2013a; Seneviratne et al., 2016). We use one ensemble member from 37 individual models or model variants (Table S1) to avoid unequal sample sizes in the multi-model ensembles. Furthermore, this choice is made to assess variability in land-atmosphere coupling *across* models, because individual ensemble members from the same model show comparably small spread in land-atmosphere coupling and present-day and future land-atmosphere coupling are highly correlated (Supplementary Online Figure S1, metric and definition is provided below). This

TABLE 7.1.: Datasets used for model evaluation

Name of dataset	Variable	Type / Group	Provider & Reference
LandFlux-EVAL[a]	ET	Ensemble Median	Mueller et al. (2013)
LandFlux-EVAL[a]	ET	Median of Reanalyses	Mueller et al. (2013)
LandFlux-EVAL[a]	ET	Median of LSMs	Mueller et al. (2013)
LandFlux-EVAL[a]	ET	Median of Diagnostic datasets	Mueller et al. (2013)
PRUNI[a,b]	ET	Diagnostic	Sheffield et al. (2006, 2010)
MPIBGC[a,b]	ET	Diagnostic	Jung et al. (2011)
CSIRO[a,b]	ET	Diagnostic	Zhang et al. (2010)
GLEAM[a,b], V. 1A	ET	Diagnostic	Miralles et al. (2011a,b)
AWB[a,b]	ET	Diagnostic	Mueller et al. (2011a)
EI-ORCHIDEE[a,b]	ET	LSM	Krinner et al. (2005)
CRU-ORCHIDEE[a,b]	ET	LSM	Krinner et al. (2005)
VIC[a,b]	ET	LSM	Sheffield et al. (2006); Sheffield and Wood (2007)
GL-NOAH-PF[a,b]	ET	LSM	Rodell et al. (2004); Rui and Beaudoing (2016)
MERRA-LAND[a,b]	ET	LSM	Reichle et al. (2011)
ERA-Interim[a,b]	ET	Reanalysis	Dee et al. (2011)
CFSR[a,b]	ET	Reanalysis	Saha et al. (2010)
JRA-25[a,b]	ET	Reanalysis	Onogi et al. (2007)
MERRA[a,b]	ET	Reanalysis	Bosilovich (2008)
CRU-TS3.2[a]	T	Observations	Harris et al. (2014)
ERA-Interim reanalysis[a]	T	Reanalysis	Dee et al. (2011)
CFSR reanalysis[a]	T	Reanalysis	Saha et al. (2010)

[a] All T-ET combinations of marked datasets have been used to derive the ET-T constraint.

[b] Original individual datasets that contributed to the LandFlux-EVAL synthesis project (Mueller et al., 2013).

indicates that the large spread between models is dominated by variability *across* models, and thus land-atmosphere coupling is a model-inherent feature on climatological time scales (Supplementary Online Figures S1 and S2, see further discussion below). On shorter (e.g. annual or seasonal) time scales, models indeed show substantial variability in their land-atmosphere coupling (Sippel et al., 2016a), which could be used as a constraint in large single-model ensembles but is beyond the scope of the present study.

Data processing and analysis All datasets were remapped to a common $2.5°\text{x}2.5°$ spatial resolution for analysis and before computing T-ET coincidences. For model evaluation (Section 3.1), all computations and analyses are performed on a monthly temporal resolution and are restricted to the time period 1989-2005 due to data availability constraints of the ET reference datasets (Mueller et al., 2013). Thus, the reference period for model evaluation corresponds to the last 17 years of the 'historical' scenario in CMIP5 models. T-ET coincidences are computed based on monthly deseasonalised and linearly detrended time series of T and ET, and coincidence rates are calculated separately for each individual season. Only land pixels outside of desert regions following the Köppen-Geiger climate classification are considered (Kottek et al., 2006). The model evaluation is conducted based on all individual pixels, and additionally on area-averages for so-called IPCC-SREX regions (IPCC, 2012).

7.2.2. Diagnostic-based model evaluation using T-ET coupling

The T-ET link and the Vegetation-Atmosphere Coupling (VAC) Index An adequate characterisation of the coupling between soil moisture and temperature is key to model evaluation using observations-based datasets. This coupling is often diagnosed by correlation-based metrics such as for example between T and ET, $\rho_{(T,ET)}$ (Seneviratne et al., 2006; Lorenz et al., 2012), or the difference in the covariability of temperature and sensible heat, where the latter is calculated with and without accounting for soil moisture deficits (Miralles et al., 2012). Here, we aim to exploit the T-ET coupling by using a natural extension of $\rho_{(T,ET)}$ that focusses on the tails of T-ET dependedencies. Deseasonalised and detrended time series of ET (x_i^{ET}) and T (x_i^T, i denotes the time step), are partitioned into five distinct classes of Vegetation-Atmosphere Coupling (VAC) following (Zscheischler et al., 2015), resulting in a time series of discrete events x_i^{VAC}:

$$x_i^{VAC} = \begin{cases} a, & \text{if} \quad x_i^T < th_{lower}^T \quad \text{and} \quad x_i^{ET} < th_{lower}^{ET}, \\ b, & \text{if} \quad x_i^T > th_{upper}^T \quad \text{and} \quad x_i^{ET} > th_{upper}^{ET}, \\ c, & \text{if} \quad x_i^T > th_{upper}^T \quad \text{and} \quad x_i^{ET} < th_{lower}^{ET}, \\ d, & \text{if} \quad x_i^T < th_{lower}^T \quad \text{and} \quad x_i^{ET} > th_{upper}^{ET}, \\ 0 & \text{otherwise.} \end{cases}$$

Event thresholds th_{lower} and th_{upper} might be chosen relative to the variability of each time series by fixing the probability p to exceed or fall below a threshold through the choice of an appropriate quantile:

$$Pr[X > th_{upper}] = Pr[X < th_{lower}] = p \tag{7.1}$$

Taking time series length restrictions into account, we choose the 30th and 70th percentile as lower and upper thresholds in all time series (i.e. such that $Pr[X < th_{lower}] = Pr[X > th_{upper}] = 0.3$). Here, we focus on coincidences of *warm temperature anomalies* ('T-events': $x_i^T > th_{upper}^T$) with anomalies in ET ('ET-events', i.e. either $x_i^{ET} > th_{upper}^{ET}$ for VAC_b or $x_i^{ET} < th_{lower}^{ET}$ for VAC_c), we derive coincidence rates r_{VAC_b} by counting the number of VAC_b-events (see Quiroga et al. (2002); Donges et al. (2016) for earlier formulations of event coincidence analysis, and e.g. Rammig et al. (2015); Siegmund et al. (2016) for applications in an ecological context):

$$r_{VAC_b} = \frac{1}{N_0} \sum_{i=1}^{N} 1_{[b]}(x_i^{VAC})$$

Here, $1_A(x)$ is the indicator function, defined as $1_A(x) = 1$ if $x \epsilon A$ and $1_A(x) = 0$ otherwise, N denotes the length of the time series. Hence, we simply count coincidences of T and ET in a given category (e.g. positive T *and* positive ET for VAC_b) to get the average coincidence rate (r_{VAC_b}). N_0 acts as a normalisation constant and is chosen in our study such that $0 \leq r_{VAC_b} \leq 1$, i.e. we normalise with the total number of 'T-events', $N_0 = \sum_{i=1}^{N} 1_{[x^T > th_{upper}^T]}(x_i^T)$. Hence, if all (or none) of the 'T-events' in the time series would coincide with 'ET-events', then the average coincidence rates would be given by $r_{VAC_b} = 1$ (or $r_{VAC_b} = 0$). For independent time series, i.e. no coupling, r_{VAC_b} would approximate the occurrence rate of 'ET-events' in the time series (defined for VAC_b) that is governed by the chosen threshold, i.e. $r_{VAC_b} = \frac{1}{N}\sum_{i=1}^{N} 1_{[x_i^{ET} > th_{upper}^{ET}]}(x_i^{ET})$ (hence, $r_{VAC_b} \approx 0.3$ in our case). Coincidence rates r_{VAC_c} follow equivalently by replacing VAC_b with VAC_c and in the definition of 'ET-events' in the previous description. We compute r_{VAC_b} and r_{VAC_c} for all seasons but with an emphasis on the warmest season of the year. In this study, significance of coincidence rates is established by randomly permuting one time series with respect to the other 100 times. Hence, VAC-rates from models or observations-based benchmarks that fall outside the 5th to 95th percentile range of the VAC-rates obtained from randomly permuted time series are significantly different from independent data at the 0.1 level.

In other words, r_{VAC_b} gives the fraction of the highest 30% of temperatures that coincide with the highest 30% of ET (i.e., occurrence rate of 'energy-limited regimes'), while r_{VAC_c} denotes the fraction of the highest 30% temperatures that correspond with the lowest 30% ET (i.e., occurrence rate of 'water-limited regimes'). Figure 7.1c-d shows a simple example of monthly time series of T and ET simulated from two CMIP5 models and occurrences of VAC_b and VAC_c are highlighted, and Figure 7.1e-f shows the correlation of T and ET. Note that for the same region (area-average over Central Europe, CEU) and time of the year (monthly data for June, July, and August), one model produces predominantly energy-limited regimes (VAC_b, Figure 7.1c,e and compare to conceptual illustration in Figure 7.1a), whereas the other model produces predominantly water-limited regimes (VAC_c, Figure 7.1d,f and concept in Figure 7.1b).

We abbreviate the average occurrence rates r_{VAC_b} and r_{VAC_c} as VAC_b and VAC_c for convenience in the remainder of the paper. In comparison to more traditional coupling metrics, such as e.g. $\rho_{(T,ET)}$, VAC might be expected to yield similar results on very long time scales, whereas on shorter time scales the VAC index picks up non-linearities in the tails (e.g. during warm temperature anomalies). At the monthly time scale (as used in this study), VAC_b and VAC_c detect distinct non-linearities in models and observations in summer T-ET coupling e.g. in CEU: Supplementary Online Figure S3 shows that, by correlating VAC_b

with VAC_c derived from individual models, observations-based benchmarks, and from a two-dimensional Gaussian distribution, VAC_b and VAC_c rates in models and observations-based benchmarks exceed those that would be expected in random data. This deviation indicates that the warm tail is indeed different to the remainder of the distribution (we observe no such deviation for the cold tail, Supplementary Online Figure S3), and hence an evaluation metric that focuses on the tail such as the VAC index is indeed useful for our present purpose. In addition to the main text, the model evaluation is presented for $\rho_{(T,ET)}$ to demonstrate robustness to the chosen methodological approach (Supplementary Online Figure S4), and for the VAC index using a 90th percentile threshold (Supplementary Online Figure S4). Both alternatives show qualitatively similar results (see Results and Discussion section).

A constraint on T-ET coupling in multi-model ensembles In general, a constraint links an observations-based diagnostic with a key model output variable across multiple models (Cox et al., 2013), and thus can be used to reduce model uncertainties and spread. Here, we derive a T-ET coupling constraint as the uncertainty range from the 54 combinations of T-ET benchmarking datasets. A Gaussian kernel with reliable data-based bandwidth selection (Sheather and Jones, 1991) is fitted over all 54 1989-2005 coincidence rates (r_{VACc}) for each meteorological season and pixel (and each SREX region average). Throughout this paper, the 5th to 95th percentile range of the fitted Gaussian kernels is taken as the plausible range of observations, and the reduced (constrained) ensemble of CMIP5 simulations is obtained by retaining only those CMIP5 models that simulate T-ET coincidences that fall within this range of observational uncertainty.

7.3. Results and discussion

In this section, we first evaluate land-coupling in CMIP5 models explicitly against an observations-based ensemble of T-ET combinations and explore the link to temperature variability and extremes (Section 3.1). All model evaluation results are presented globally and exemplarily for Central Europe (CEU) as a region where global models and observations differ widely. Subsequently, we constrain the ensemble of CMIP5 models using each model's land-coupling as diagnosed through the VAC_c index and discuss implications for biases in simulated present-day temperature extremes and warming projections (Section 3.2).

7.3.1. Evaluation of land-atmosphere coupling in CMIP5 models and the link to temperature variability and extremes

Evaluation of T-ET coupling in CMIP5 models. Models and observations-based datasets show a relatively large spread in their representation of T-ET coupling, as expressed exemplarily in CEU through both VAC_b and VAC_c across various seasons (Figure 7.2a,b) or diagnosed through more traditional coupling metrics such as $\rho_{(T,ET)}$ (Supplementary Online Figure S4). Individual models indicate pronounced qualitative differences in the warm season, where some models point to energy-limited, whereas others indicate predominantly water-limited conditions (Figure 7.2a,b, and Figure 7.1, for an illustrative example). Observations-based T-ET datasets agree qualitatively, i.e., indicating energy-limited to neutral conditions in the CEU example, thus implying an overestimation of water-limited regimes in CEU in roughly 50% of CMIP5 models (Figure 7.2).

This pattern holds across most regions of the globe, as many CMIP5 models consistently overestimate occurrences of VAC_c regimes (and correspondingly underestimate VAC_b occurrences) in the warm season of the year (Figure 7.2c,d, see Supplementary Online Figure S5 for a definition of the warm season in each pixel). In mid-latitude and several tropical regions (e.g. Central North America, Central Europe, the Amazon, India, parts of Africa), more than 25% and up to 50% of CMIP5 models lie outside the observational range (Figure 7.2d). These discrepancies hold also if metrics that emphasise the whole distribution ($\rho_{(T,ET)}$) or more extreme parts of the tail (VAC based on a 90th percentile threshold) are used for model evaluation (Supplementary Online Figure S4, results for individual seasons are presented for VAC_c and VAC_b in Supplementary Online Figures S6 and S7, respectively). Moreover, the spread between the individual models' representation of land-atmosphere coupling strongly exceeds the spread in observational datasets, although different diagnostic, reanalyses and land surface model datasets are included in the observations-based ensemble (Figure 7.2e for CMIP5 model spread and Figure 7.2f for spread in observations-based benchmark datasets).

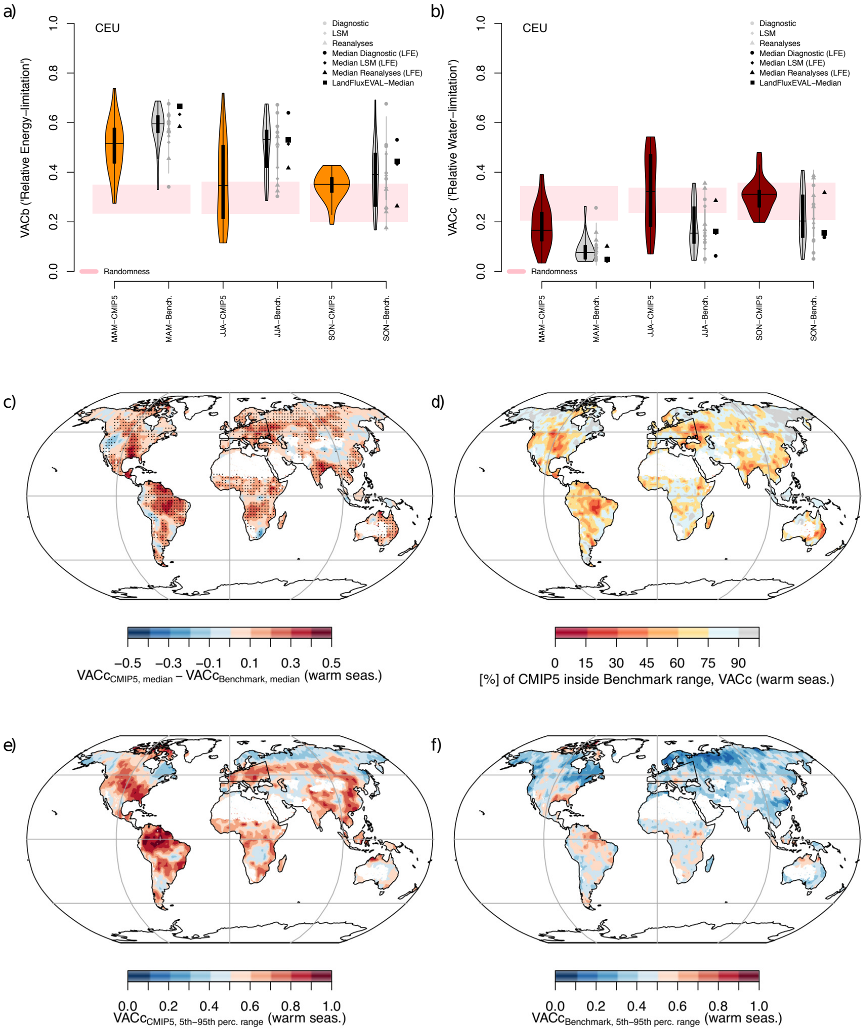


FIGURE 7.2.: Evaluation of T-ET coupling in global climate models. Full caption is continued on the following page.

Furthermore, the models' land-atmosphere coupling, as diagnosed here through the VAC-index, is a highly model-inherent feature, as different model variants or

FIGURE 7.2.: (continued) a, b) VACb and VACc coupling in the CMIP5 climate model ensemble and observations-based benchmarking datasets in Central Europe (CEU, 1989-2005, area-average) with systematic warm season differences (circles, diamonds, and triangles indicate diagnostic, land surface models, and reanalyses reference datasets, respectively). Randomness indicates the 5th to 95th percentile range obtained by randomly permutating both time series with respect to the other ($N = 100$ times) to obtain independent data. c) Difference in the VACc median of the CMIP5 ensemble and benchmarking datasets. d) Fraction of CMIP5 models that are inside the 5th-95th percentile spread of the benchmarking datasets. e, f) Range of VACc-occurrences (5th to 95th percentile range) in CMIP5 models (e) and in the ensemble of observations (f).

ensemble members from the same model generally lie relatively close to each other (Supplementary Online Figures S1 and S2). However, model-specific signatures of model output are not unusual, as diagnosed before for spatial patterns of temperature and precipitation (Knutti et al., 2013) or the statistical information content in carbon fluxes (Sippel et al., 2016b). Furthermore, present-day land-atmosphere coupling is strongly related to future land-atmosphere coupling in the individual models (Supplementary Online Figure S1). A detailed overview of VAC_c coupling in individual models and ensemble members relative to the benchmark datasets for Central Europe and Central North America is presented in Supplementary Online Figures S1 and S2. Despite regionally pronounced qualitative discrepancies, it should be noted that on a global scale, the distribution of water-limited and energy-limited patterns in models and observations agrees qualitatively (Supplementary Online Figure S8). Likewise, the findings of climatologically too pronounced water-limited regimes in individual models w.r.t. observations does not exclude the possibility of future changes in the coupling strength in transitional regions (Seneviratne et al., 2006) or of strong water limitations during extreme events in the real world (Miralles et al., 2012; Whan et al., 2015). To this end, an evaluation of the year-to-year variability of the coupling behaviour in larger ensembles of individual models, including very rare events, could constitute a topic for further study, as this study was restricted to relatively moderate events in a 16 year period (70th percentile threshold for the computation of the VAC-index) and one ensemble member per model. Besides, we also note that observations-based benchmark datasets show systematic (albeit smaller) differences in the representation of land-atmosphere coupling: Diagnostic datasets indicate more frequent energy-limited regimes (see e.g. Figure 7.2), and thus differ consistently to generally drier land surface models and reanalysis products, consistent with earlier findings (Santanello et al., 2015).

T-ET coincidences and the link to temperature variability and extremes. The representation of T-ET coupling as diagnosed through the VAC index largely determines the variability of temperatures at monthly and inter-annual time scales across the CMIP5 multi-model ensemble in CEU (Figure 7.3a) and in most regions of the globe except in some subarctic climates (Figure 7.3b). Therefore, this relationship is indicative for the strong influence of land-atmosphere coupling on surface climate. This is consistent with previous findings in Europe in models with and without land-atmosphere interactions (Seneviratne et al., 2006; Fischer and Schär, 2009; Fischer et al., 2012). An important result is that models that produce VAC_c indices within the range of benchmark datasets also produce a realistic near surface temperature variability, whereas models that fall too frequently in water-limited regimes also overestimate summer temperature variability (Figure 7.3a). Moreover, in mid-latitude and tropical regions, the state of the land surface is strongly associated with the mean and variability of temperature extremes at the daily time scale in the warmest season (TXx, Figure 7.3c,d). The link between between the representation of land-atmosphere coupling and simulated temperature extremes and variability in global climate models is consistent with earlier studies, which has been demonstrated for Europe in individual models (Seneviratne et al., 2006; Lorenz et al., 2012; Davin et al., 2016) and in ensembles of regional models (Fischer et al., 2012; Bellprat et al., 2013). Therefore, the relationship between T-ET coincidence rates and temperature extremes might offer an avenue to derive an explicit land-atmosphere coupling constraint (the likely root cause for biases) to alleviate biases in temperature variability and extremes in the multi-model CMIP5 ensemble.

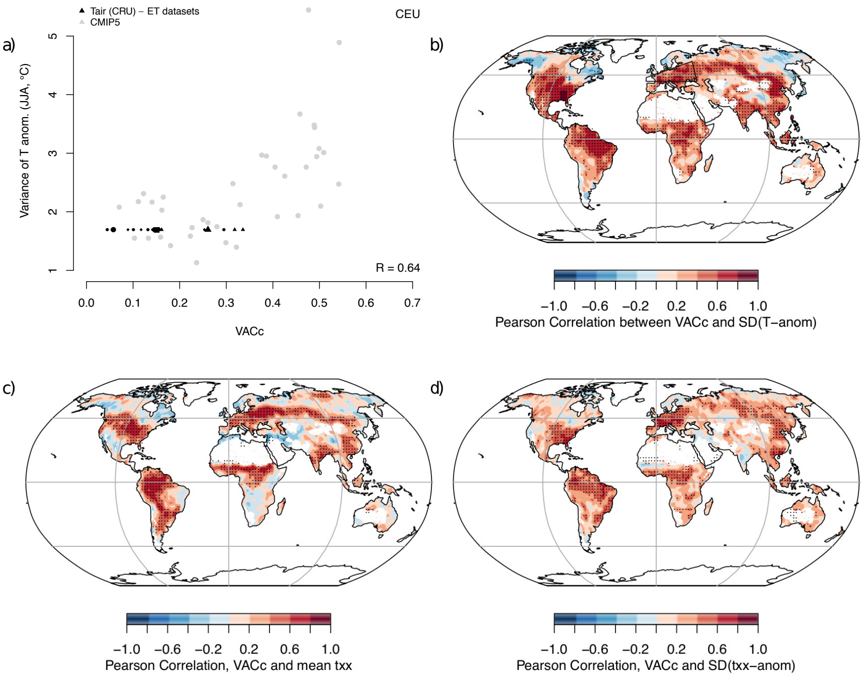


FIGURE 7.3.: a, b) Relationship between model-specific T-ET coupling (expressed through VACc) and model simulated variability of monthly temperature anomalies (JJA) in Central Europe (a), and globally (b). c, d) Relationship betweeen VACc-coupling and mean (c) and standard deviation (d) of simulated monthly maximum value of daily maximum temperature (TXx) in summer (JJA).

7.3.2. Analysis of constrained multi-model ensemble and implications for future climate projections

A constraint on land-atmosphere coupling in the CMIP5 ensemble. The association between T and ET in the constrained ensemble resembles the observations-based benchmarking datasets in T-ET coupling very well (shown as a bivariate density estimate in Figure 7.4a-b for CEU and CNA, respectively), whereas the unconstrained CMIP5 ensemble produces too many occurrences of VAC_c conditions in both CEU and CNA. Due to the intimate link between land-atmosphere coupling and temperature variability and extremes (see previous Section), we expect that the improvement in the representation of land-atmosphere coupling in the constrained ensembles yields a corresponding improvement in the representation of temperature extremes at the daily time scale in coupling-sensitive regions.

Coupling-sensitive regions are prone to warm season biases in climate models (Christensen and Boberg, 2012; Bellprat et al., 2013). In the present analysis, high biases in temperature extremes are indeed prevalent in the original (unconstrained) CMIP5 ensemble in these regions (Figure 7.4c,e). For example, the ensemble mean warm season TXx is overestimated by up to 5°C, and higher biases are detected in the 90th percentile of TXx in CNA, CEU or the Amazon (all biases in daily variables relative to ERA-Interim, see Figure 7.4c,e). In a CMIP5 ensemble constrained by the land-atmosphere coupling metric VAC_c, the representation of temperature extremes is improved in regions prone to coupling-induced biases (Figure 7.4d,f), i.e. both mean TXx and the 90th percentile of TXx are significantly reduced. The ensemble mean of present-day temperature extremes in other regions remains unchanged. Moreover, projected future temperature extremes are reduced in the constrained ensemble (Figure 7.5), similarly to present-day reductions in regions prone to present-day biases in land-atmosphere coupling. This is illustrated in Figure 7.5a for TXx (monthly area-averages in summer) in CEU, where the hot end of the original model ensemble is in fact never realised in observed temperatures. The application of the constraint thus not only affects mean TXx, but also reduces the spread of the model ensemble (Figure 7.5a,b). The reduction in ensemble mean and ensemble spread is retained for the entire 21st century (Figure 7.5a,b). Hence, this result reinforces that coupling-related biases are model-inherent features, i.e. models that simulate too many VAC_c-occurrences today (and associated high biases in extreme temperatures) are very likely to do so in the future. However, one should keep in mind that the reduction in ensemble mean and spread is confined to coupling-sensitive regions in CEU, CNA, and to some degree in the Amazon region (Figure 7.5c,d).

Our results imply that an accurate representation of land surface processes is crucially relevant for a correct simulation of temperature extremes, and more generally for simulated near-surface climate variability. Land-atmosphere coupling is thus an important source of bias in state-of-the-art global climate model simulations. By using an observations-based land-atmosphere coupling diagnostic

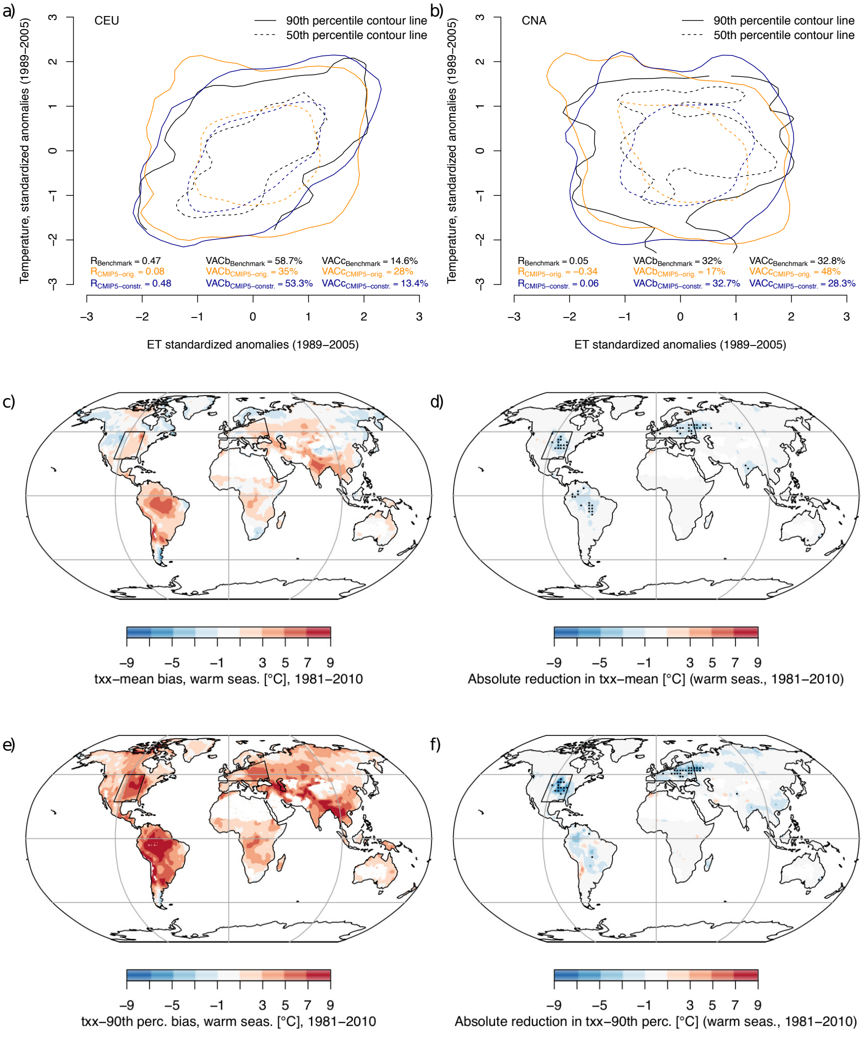


FIGURE 7.4.: a-b) Contour lines of bivariate kernel density estimates of T-ET relationship in the benchmarking datasets, the original and constraint CMIP5 ensemble for (a) Central Europe, and (b) Central North America (1989-2005, area-average). c, e) Biases in warm season (c) TXx mean, and (e) 90th percentile of TXx in the original CMIP5 ensemble, and (d, f) reduction in (d) TXx mean, and (f) 90th percentile TXx through the application of the land-coupling constraint. Regions with a significant reduction in (d) TXx mean, and (f) the across-model average in the 90th percentile of TXx according to a permutation significance test are stippled.

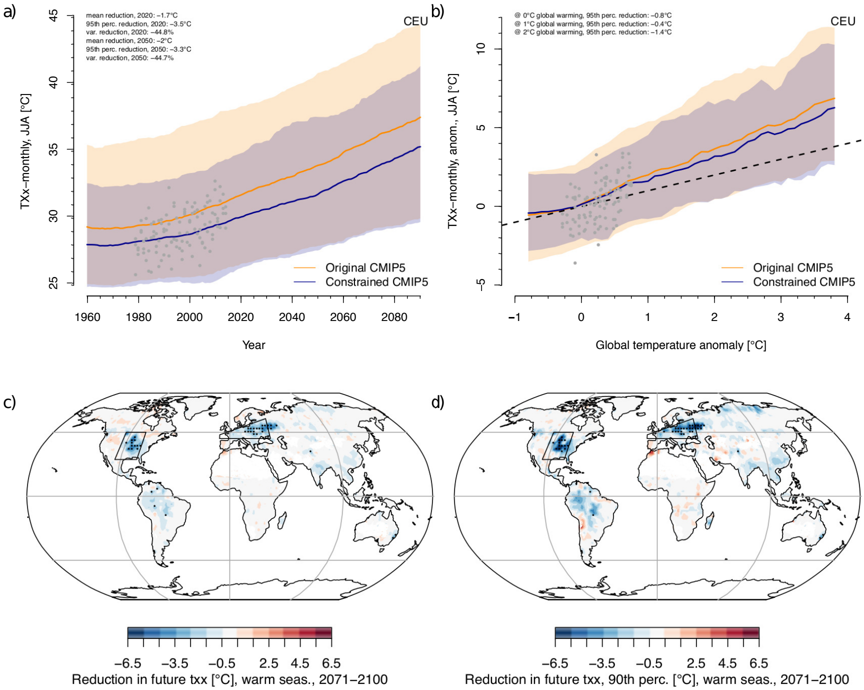


FIGURE 7.5.: Application of land coupling constraint to CMIP5 ensemble. a, b) Ensemble prediction of original and constrained multi-model ensemble for (a) future absolute TXx and (b) range of TXx anomalies relative to global mean temperature anomalies in each model, following Seneviratne et al. (2016). Envelopes indicate 5th to 95th percentile. c, d) Global maps of projected changes in simulated (c) mean TXx, and (d) 90th percentile of TXx in the VACc-constrained CMIP5 ensemble.

to constrain the multi-model CMIP5 ensemble, we have shown that biases in extremes in the large ensemble can be alleviated to a certain degree. As bias correction methodologies that take the physical causes for biases into account are still widely lacking (Ehret et al., 2012; Bellprat et al., 2013) and multivariate bias correction methods are currently in development (Cannon, 2016), the identification of models with a *physically plausible* representation of near-surface climate and land-atmosphere interactions at the regional scale might be crucial to extract accurate and relevant information about climate extremes in the context of climatic changes in the 21st century (Mitchell et al., 2016b; Schleussner et al., 2016b;

Seneviratne et al., 2016). For example, model selection for event attribution studies or a quantification of changes in univariate climate extremes is often based on a statistical performance criterion (Perkins et al., 2007; King et al., 2016; Otto et al., 2015). Our results indicate that these procedures could be further refined through incorporating observations-based diagnostics or constraints in order to analyse model simulations that are indeed 'right for the right reasons' (at least given physics-guided and observations-based relationships). Moreover, the impacts of climate and its extremes e.g. on human health or ecosystems (Mitchell et al., 2016a; Frank et al., 2015) are often inherently related to multiple climate variables (Ehret et al., 2012; Leonard et al., 2014). Therefore, simple constraints as motivated for instance in the present study might complement more conventional bias correction procedures (e.g. Hempel et al., 2013) to derive physically consistent estimates of climate impacts. This approach appears promising, because biases within climate models (i.e. in different variables) and across climate model ensembles are often correlated (e.g. Knutti, 2010; Mueller and Seneviratne, 2014; Sippel et al., 2016a). Hence, beyond soil moisture control on simulated temperature extremes as the present study's focus, related biases in other variables such as warm season precipitation or ET might be similarly relevant in this context. For example, VAC_c occurrences across the CMIP5 ensemble are negatively associated with precipitation and ET in the warm season in mid-latitude regions (Supplementary Online Figure S9) - both crucial variables in the water cycle that show pronounced summer low biases in CMIP5 models (Mueller and Seneviratne, 2014). Therefore, a constrained model ensemble with improved land-atmosphere coupling, a likely root cause of biases (Lorenz et al., 2012), might not only improve temperature extremes and variability, but additionally might reduce biases in associated variables such as ET or precipitation.

Is there a link between present-day land-atmosphere coupling and warming projections? We investigate whether the representation of land-atmosphere coupling in climate models affects the magnitude of 21st century warming (e.g. Fischer et al., 2012; Stegehuis et al., 2013). We first note that regions sensitive to land-atmosphere coupling in the CMIP5 model ensemble also show relatively strong warming in daily-scale temperature extremes (TXx), for example Central America or South and Central Europe (Figure 7.6a,b). More importantly, however, models that produce frequent VAC_c occurrences (water-limited regimes) tend to be associated with larger rates of warming in TXx, although it should be emphasised that this relationship is not simple or linear (Figure 7.6c,d, see also Fischer et al. (2012)). Conversely, this pattern reverses in boreal regions, where strongly energy-limited models (i.e. very few VAC_c occurrences) tend to produce larger warming. However, in boreal regions this apparent relationship likely stems from a spurious correlation with the individual models' background warming (i.e., warming in annual averages), as the correlation in fact disappears if the background warming is subtracted from summer warming (Supplementary Online Figure S10). In contrast, in mid-latitude regions warm season warming that exceeds annual average warming remains confined to the warm season. A multi-model projection constrained by a plausible representation of land-atmosphere coupling reduces differences in TXx estimates in a future climate relative to the present in coupling-sensitive regions such as Central Europe and Central North America by locally by around 0.5°C to 1°C - but this remains a regional effect (Figure 7.6e,f). These results are consistent with earlier studies that used an ensemble of regional models over Europe that used the standard deviation of temperatures as a constraint (Fischer et al., 2012).

7.4. Conclusions

In the present study, we have evaluated land-atmosphere coupling in state-of-the-art climate models with an ensemble of observations using a diagnostic based on coincidences of T and ET anomalies (the so called VAC index). While observations and models broadly agree on spatial patterns of land-atmosphere coupling, our results reveal that models differ widely in coupling-sensitive regions in the mid-latitudes and the tropics. Several models exhibit systematically too frequent coincidences of high temperature anomalies with negative ET anomalies (water-limited regimes) in mid-latitude regions in the warm season, and in several tropical regions year-round. Across the multi-model ensemble, we found a strong association of land-atmosphere coupling with simulated temperature variability and extremes. The spread between models largely explains differences in

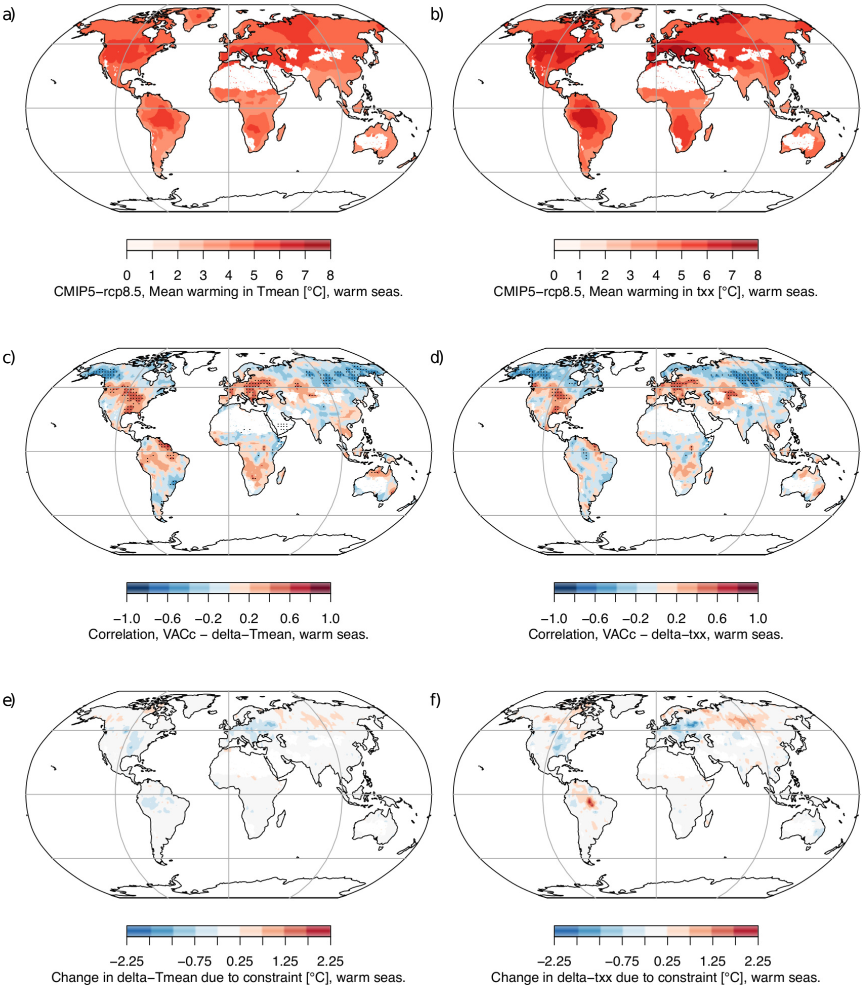


FIGURE 7.6.: a, b) Projected warming in warm season (a) mean temperature, and (b) TXx across the CMIP5 ensemble (RCP8.5 scenario, 2071-2100 relative to 1981-2010). c, d) Correlation between VACc in the warm season and the projected warming in (c) mean temperature, and (d) TXx. Stippling indicates significant correlations. e, f) Relative change in (e) mean warming and (f) TXx warming due to the application of the land-atmosphere coupling constraint, warming defined as 2071-2100 relative to 1981-2100.

simulated monthly temperature variability and daily extremes. We applied a land-atmosphere coupling constraint to the multi-model ensemble, which considerably improves the representation of land-atmosphere coupling in the ensemble, and reduces biases in temperature variability and extremes in present-day simulations in a physically consistent manner (Figure 7.4). Furthermore, the constraint leads to reduced variability and lower extreme temperatures in future projections. However, the overall projected changes in temperature extremes are not so strongly affected (reduction around $0.5 - 1.0$°C locally in regions that are sensitive to land-atmosphere coupling), because the models with overestimated land-atmosphere coupling display similar anomalies from the multi-ensemble mean in present and future. In conclusion, we selected models with a *physically plausible* representation of land surface processes (and near-surface climate) using observations-based constraints that are guided by physical considerations. This approach complements more traditional bias correction approaches and offers new avenues to obtain improved estimates of future climate impacts.

Part III.

Extreme events in terrestrial ecosystems: Drivers and attribution

8. Contrasting and interacting changes in spring and summer carbon cycle extremes in European ecosystems[1,2]

Abstract

Climate extremes have the potential to cause extreme responses of terrestrial ecosystem functioning. However, it is neither straightforward to quantify and predict extreme ecosystem responses, nor to attribute these responses to specific climate drivers. Here, we construct a factorial experiment based on a large ensemble of process-oriented ecosystem model simulations driven by a regional climate model (12.500 model-years in 1985-2010) in six European regions. Our aims are to (1) attribute changes in the intensity and frequency of simulated ecosystem productivity extremes (EPEs) to recent changes in climate extremes, CO_2 *concentration, and land-use, and to (2) assess the effect of timing and seasonal interaction on the intensity of EPEs. Evaluating the ensemble simulations reveals that (1) recent trends in EPEs are seasonally contrasting: Spring EPEs show consistent trends towards increased carbon uptake, while trends in summer EPEs are predominantly negative in net ecosystem productivity (i.e. higher net carbon release under drought and heat in summer) and close-to-neutral in gross productivity. While changes in climate and its extremes (mainly warming) and changes in* CO_2 *increase spring productivity, changes in climate extremes decrease summer productivity neutralizing positive effects of* CO_2*. Furthermore, we find that*

[1]This chapter is published as Sippel, S. M. Forkel, A. Rammig, K. Thonicke, M. Flach, M. Heimann, F. E. L. Otto, M. Reichstein and M. D. Mahecha. 2017. *Environmental Research Letters* **12**(7), 075006. doi:10.1088/1748-9326/aa7398.

[2]Supplementary material that complements this Chapter with more detailed explanations is available in Appendix E.

(2) drought or heat wave induced carbon losses in summer (i.e. negative EPEs) can be partly compensated by a higher uptake in the preceding spring in temperate regions. Conversely, however, 'carry-over' effects from spring to summer that arise from depleted soil moisture exacerbate the carbon losses caused by climate extremes in summer, and are thus undoing spring compensatory effects. While the spring-compensation effect is increasing over time, the carry-over effect shows no trend between 1985-2010. The ensemble ecosystem model simulations provide a process-based interpretation and generalization for spring-summer interacting carbon cycle effects caused by climate extremes (i.e. compensatory and carry-over effects). In summary, the ensemble ecosystem modelling approach presented in this paper offers a novel route to scrutinize ecosystem responses to changing climate extremes in a probabilistic framework, and to pinpoint the underlying eco-physiological mechanisms.

8.1. Introduction

Climate variability and extremes are key features influencing terrestrial ecosystem functioning (Smith, 2011; Reyer et al., 2013; Baldocchi et al., 2016). Climatic extremes directly propagate into the biosphere through various eco-physiological pathways, for instance affecting plant phenological events (Jentsch et al., 2009; Ma et al., 2015) or carbon cycling from regional to global scales (Knapp et al., 2002; Reichstein et al., 2013; Zscheischler et al., 2014a; Frank et al., 2015). Major climatic extreme events such as the European heat wave and drought 2003 (Ciais et al., 2005; Reichstein et al., 2007), or droughts in North America (Schwalm et al., 2012; Wolf et al., 2016), Australia (Ma et al., 2016) and the Amazon (Phillips et al., 2009; Lewis et al., 2011) consistently cause net carbon losses. However, because the number of directly observed large-scale extreme climate events and associated impacts on ecosystem productivity are rare, and because field experiments are often limited in extent and thus difficult to upscale to larger regions (Beier et al., 2012), crucial uncertainties remain in our understanding of processes that control these phenomena.

Climatic extreme events are changing in magnitude and frequency (Alexander et al., 2006; IPCC, 2012), and these occur in addition to more gradual climatic changes in, e.g., seasonal variation (Stine et al., 2009; Cassou and Cattiaux, 2016) and climate trends. These changes, in tandem with non-linear feedbacks or lagged effects (Frank et al., 2015), might impart decisive consequences for regional and global-scale carbon balances of terrestrial ecosystems (Reichstein et al., 2013).

For example, the extreme summer drought 2012 in the contiguous United States caused losses in carbon uptake in summer (Wolf et al., 2016) which were offset by warming-induced increases in spring carbon uptake, leading to a spring-summer compensation of the regional carbon balance (Figure 8.1). Furthermore, Wolf et al. (2016) hypothesised that earlier spring plant activity could have induced negative carry-over effects to summer productivity via soil-moisture deficits (Figure 8.1), as suggested before (Richardson et al., 2010). However, as the evidence for seasonal compensation of extremes in Wolf et al. (2016) is based on a single event only it remains uncertain whether such interacting effects can be expected generally for climate extremes in summer. Long time series allowing to comprehensively study additional *independent* climatic extreme events in spring and/or summer would be required as such lagged effects in ecosystem productivity could have simply occurred by chance.

Climate extremes may cause immediate or delayed responses in ecosystems (Frank et al., 2015), but not all climate extremes lead to an extreme ecosystem response (Smith, 2011). Therefore, systematic quantification and attribution of contemporary trends in ecosystem productivity extremes, including potential interactions of events, is required. Respective analysis on observations is often hindered by small sample sizes. Alternatively, large ensembles of climate-ecosystem model simulations might complement a 'case study type' assessment of extremes in the observational record because they allow to explore how climate variability and extreme events are related to extreme ecosystem responses (Ciais et al., 2005; Schwalm et al., 2012; Wolf et al., 2016). For example, multi-thousand member ensembles of climate simulations were used to analyse and attribute extreme climate events, such as the Russian heat wave 2010 (Otto et al., 2012), or to investigate the role of climate extremes in causing, e.g., floods (Pall et al., 2011; Schaller et al., 2016) and heat-health related issues (Mitchell et al., 2016a). This approach is appropriate when analysing the impact of climatic extreme events on ecosystem functions.

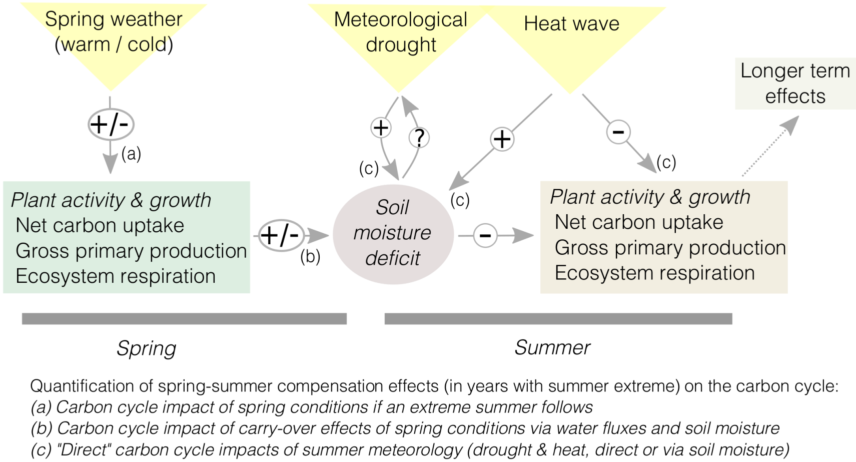


FIGURE 8.1.: Conceptual illustration of spring-summer interacting carbon cycle effects due to climate extremes. In years affected by summer heat and drought (Arrows (c)), warm spring conditions could potentially partly compensate for carbon losses in summer due to higher carbon uptake in spring (Arrow (a), associated with (+)). Conversely, however, warm spring conditions might lead to earlier soil moisture depletion (Arrow (b), associated with (+)) and thus a carry-over effect from spring to summer carbon cycling. Diagram modified from Sippel et al. (2016c).

This study investigates two main objectives: Our first objective is to systematically assess changes in EPEs in spring and summer using climate-ecosystem model ensemble simulations, and to attribute seasonal changes in EPEs to changes in climate extremes, atmospheric CO_2 and land-use change. Second, we focus on interactions between negative summer EPEs and the preceding spring conditions, and reinvestigate the outlined spring compensation and carry-over effects in years affected by negative summer EPEs on regional carbon cycling from a climate-ecosystem ensemble modeling perspective, and provide a model-based interpretation and generalisation of these effects.

8.2. Data and methods

The methodological workflow of the study is as follows: We use a large ensemble of bias-corrected regional climate model simulations (Section 8.2.1) to drive an ensemble of ecosystem model simulations (Section 8.2.2) for six eco-physiologically different European regions. Factorial model simulations are set up (Section 8.2.3) and used to disentangle climatic and non-climatic drivers of seasonal changes in EPEs, and to scrutinise respective spring-summer interacting carbon cycle effects (Section 8.2.4).

8.2.1. Regional climate model simulations and physically consistent bias correction

The core ingredient to the present study is an ensemble of regional climate simulations over Europe that cover 26 years of transient climate change (1985-2010) and 800 ensemble members in each year (i.e. 20,000 members in total) based on perturbed initial conditions. Climate model simulations have been generated through distributed computing on citizen scientists' computers (http://www.climateprediction.net/weatherathome), using the global general circulation model HadAM3P (1.875°x1.25°x15min resolution, 19 vertical levels) and a dynamically downscaled regional model version (HadRM3P, 0.44°x0.44°x5min resolution, Massey et al. (2015)) in atmosphere-only mode. Hence, the model is driven by observed sea surface temperatures, sea ice fractions, the solar cycle, and the observed atmospheric composition (greenhouse gases, aerosols, ozone, see Massey et al. (2015) for further details). The present experimental setup has been used to assess and attribute changes in climatic extreme events and its impacts in various sectors (Otto et al., 2012; Sippel and Otto, 2014; Schaller et al., 2016; Mitchell et al., 2016a), because the large available sample size allows to scrutinise even small changes in the odds of climatic extreme events. European summer climate in HadRM3P and other climate models is frequently too hot and dry (Massey et al., 2015). To alleviate this issue, we apply a resampling-based bias correction that preserves the physical consistency in the ensemble simulations (for details see Sippel et al. (2016a)): A Gaussian kernel fitted over 1985-2010 mean summer area-averaged temperatures in the ERA-Interim dataset (Dee et al., 2011) in each of the six European regions (Table E1) is used as a constraint for resampling 500 ensemble members in each year. The resampling procedure improves the representation of summer climate in HadRM3P substantially, but reduces the available sample size of the ensemble and cannot account for all possible biases (Sippel et al., 2016a).

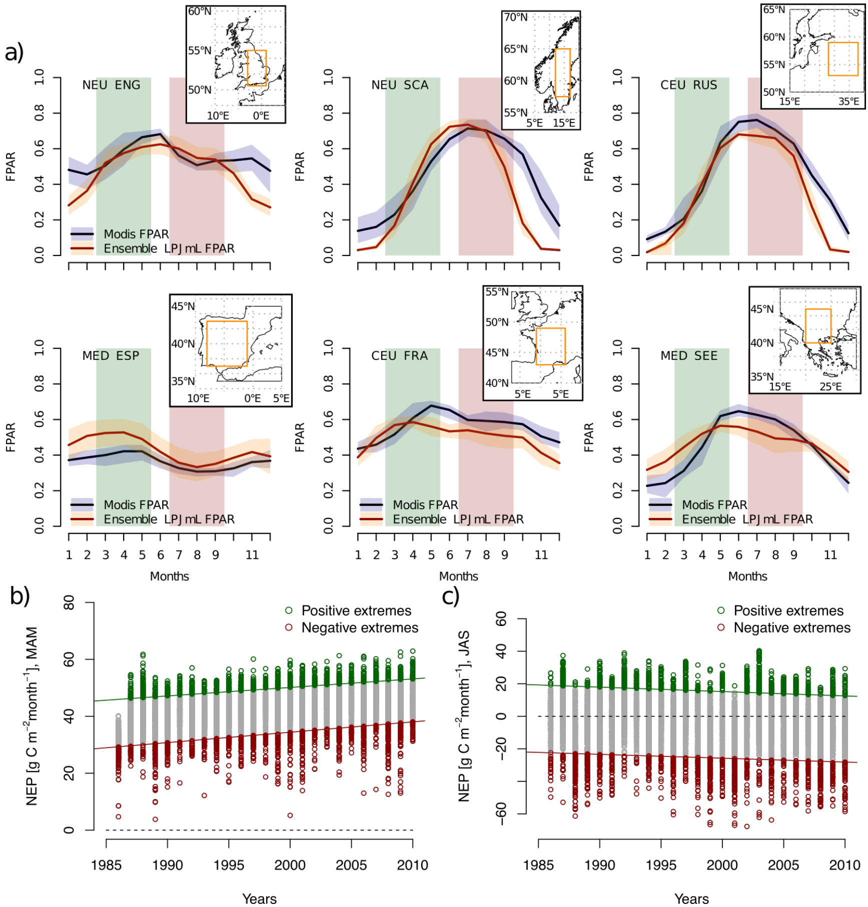


FIGURE 8.2.: a) Illustration of seasonal cycles in vegetation phenology (indicated by the Fraction of Absorbed Photosynthetically Active Radiation, FPAR) in satellite observations (MODIS) and in the ensemble of LPJmL model simulations ($\pm 2\sigma$ range) in all six regions studied in this paper. b, c) Identification of extremes in the response variable's distribution in the presence of trends in (b) spring and (c) summer: Quantile regression of the 10th and 90th conditional percentile against time.

8.2.2. Terrestrial ecosystem simulations: Model description

The process-based Lund-Potsdam-Jena managed Land dynamic model (LPJmL, Version 3.5) simulates terrestrial vegetation dynamics (growth, competition and mortality), land-atmosphere fluxes of carbon (gross and net primary productivity, ecosystem respiration) and water (evaporation, transpiration, interception) in natural ecosystems (Sitch et al., 2003) and under human land use (Bondeau et al., 2007). Carbon allocation in LPJmL follows the fully coupled photosynthesis and water balance scheme of the BIOME3 model (Haxeltine and Prentice, 1996), i.e. the photosynthetic light-use efficiency is subject to environmental controls via co-limiting light-limited enzyme regeneration and rubisco-limited enzyme-kinetic rates (ibid.). Respiration from plant compartments follows a modified Arrhenius relationship (Lloyd and Taylor, 1994). Heterotrophic decomposition of litter and soil carbon pools depends additionally on soil moisture and follows first-order kinetics (Sitch et al., 2003). LPJmL consists of 11 natural plant functional types and 13 crop functional types that differ in their bioclimatic limits and ecophysiological parameters. Here, we run LPJmL with an improved hydrology scheme (Gerten et al., 2004; Schaphoff et al., 2013), human land use (Bondeau et al., 2007), agricultural water use (Rost et al., 2008), and an improved phenology module (Forkel et al., 2014). Phenology and photosynthesis-related parameters have been optimised against remote sensing observations resulting in an improved simulation of natural vegetation greenness dynamics (Forkel et al., 2015). LPJmL ensemble simulations are performed at a monthly temporal and at 0.5° spatial resolution. The spinup procedure consists of 1,200 years by randomly concatenating individual ensemble members (sampled from the first ten available years, 1986-1995) with transient CO_2 concentration and land use.

Region selection All ensemble simulations are conducted for six individual regions in Europe that broadly sample the spectrum of variability of vegetation productivity in Europe (Figure E1), revealed from seasonal cycles in the satellite observed Fraction of Photosynthetically Active Radiation (FPAR) taken from the MODIS FPAR product (Myneni et al., 2002). Spring (March-May) and summer (July-September) cover very different seasonality patterns in FPAR (Figure 8.2).

The LPJmL ensemble reproduces seasonal dynamics of vegetation phenology at the regional scale and the regional gradient in FPAR dynamics (Figure 8.2).

8.2.3. Factorial model simulations

The factorial set of climate-ecosystem model simulations (Table 8.1) is based on a standard run ('All'), in which LPJmL is run with all drivers, including transient CO_2 concentrations and human land-use (Fader et al., 2010). Moreover, LPJmL is run separately for constant CO_2 ('CONSTCO2'), constant land-use ('CONSTLU'), and both constant CO_2 and land-use ('CONSTLUCO2'). In this factorial, ensemble-based setup the differences between these runs are used to disentangle and pinpoint climatic and non-climatic (CO_2, land-use) drivers of contemporary changes in EPEs (Section 8.3.1 and 8.3.2). Lastly, to investigate carry-over effects from spring conditions to EPEs in summer (Section 8.3.3), an additional LPJmL simulation driven by randomised spring climatic conditions ('SPRINGRAND') is conducted. This step consists of randomly concatenating members of the climate ensemble between summer and spring (on June 1st) within each year such that summer meteorology remains identical to the 'All' run, but spring conditions are different. Hence, the difference in summer carbon cycling between 'All' and 'SPRINGRAND' is driven by lagged effects from spring in the ecosystem model.

TABLE 8.1.: Overview over factorial model simulations[a].

Scenario name	CO_2	land-use	climate	Section
All	transient CO_2	transient land-use	transient climate	8.3.1–8.3.3
CONSTCO2	constant CO_2 [b]	transient land-use	transient climate	8.3.1–8.3.2
CONSTLU	transient CO_2	constant land-use[c]	transient climate	8.3.1–8.3.2
CONSTLUCO2	constant CO_2 [b]	constant land-use[c]	transient climate	8.3.1–8.3.2
SPRINGRAND	transient CO_2	transient land-use	transient climate, spring randomisation[d]	8.3.3

[a] Each factorial simulation is conducted for 1986-2010 climate and propagated through the entire climate ensemble.
[b] fixed to 345ppm in 1985
[c] fixed to 1985 land-use values
[d] Ensemble members have been randomly concatenated on June 1st in each year ('random spring', but meteorological summer and autumn are identical to the other scenarios).

8.2.4. Analysis methodology

Selection of extreme events All individual ensemble members are averaged to regional and seasonal means for further analysis. Ecosystem productivity extremes (EPEs) are sampled directly from the tail of the response variable distribution (sensu Smith (2011)), which is either gross primary productivity (GPP) or net ecosystem productivity (NEP) in the present study. Let $x_{i,t,s,fac}$ denote the response variable x ($x \in \{GPP, NEP\}$), an arbitrary ensemble member i, in year t, season s, and from any factorial run fac (region is not indexed separately to lighten the notation). Ensemble members in which the response variable exceeds or falls below a given threshold in the 'All'-simulations are labelled as positive and negative EPEs ($x^{+extreme}_{j,t,s,fac}$ and $x^{-extreme}_{j,t,s,fac}$, respectively). The index j runs only over ensemble members within in a given category ($^{-extreme}$ or $^{+extreme}$). In Section 8.3.1, an illustrative extreme value analysis is conducted by fitting a Generalised Pareto Distribution (GPD) to extremes in the response variable, where the GPD constitutes a suitable limit distribution for such peak-over-threshold selection of extreme values (Coles et al., 2001). These statistical fits are derived from the 'All' simulations separately for the response variables's lower and upper tails (negative and positive EPEs) using a 5th and 95th quantile threshold to identify EPEs, and separately for each season and two decadal periods (1986–1995 and 2001–2010).

In Section 8.3.2 and 8.3.3, a quantile regression of the 10th (90th) conditional percentile against time in the 'All' simulation is performed (Cade and Noon, 2003) to identify EPEs relative to time-dependent thresholds, thus accounting for potential trends in the 25-year period. This yields a selection of 1250 EPEs (out of 12.500 members) for each response variable, region, and season (see Figure 8.2b,c for an illustration).

Attribution to drivers of change In Section 8.3.1 (Figure 8.3), the effects of individual factors on changes in EPEs (CO_2: $\Delta x^{-extreme}_{\mathbf{CO2}_s}$, land-use: $\Delta x^{-extreme}_{\mathbf{LU}_s}$, climate: $\Delta x^{-extreme}_{\mathbf{climate}_s}$, indicated here exemplarily only for negative extremes) between both periods and in season s are teased out by computing the difference between both time periods of the averaged individual effects from the factorial simulations (averages over any specific dimension are denoted as $\overline{x}.$):

$$\begin{aligned}\Delta x^{-extreme}_{\mathbf{CO2}_s} &= (\overline{x}^{-extreme}_{\cdot,2001-2010,s,All} - \overline{x}^{-extreme}_{\cdot,2001-2010,s,CONSTCO2}) - \\ &\quad (\overline{x}^{-extreme}_{\cdot,1986-1995,s,All} - \overline{x}^{-extreme}_{\cdot,1986-1995,s,CONSTCO2}) \qquad (8.1)\\ \Delta x^{-extreme}_{\mathbf{LU}_s} &= (\overline{x}^{-extreme}_{\cdot,2001-2010,s,All} - \overline{x}^{-extreme}_{\cdot,2001-2010,s,CONSTLU}) - \\ &\quad (\overline{x}^{-extreme}_{\cdot,1986-1995,s,All} - \overline{x}^{-extreme}_{\cdot,1986-1995,s,CONSTLU}) \qquad (8.2)\\ \Delta x^{-extreme}_{\mathbf{climate}_s} &= (\overline{x}^{-extreme}_{\cdot,2001-2010,s,CONSTLUCO2} - \\ &\quad (\overline{x}^{-extreme}_{\cdot,1986-1995,s,CONSTLUCO2}). \qquad (8.3)\end{aligned}$$

In Section 8.3.2, the contribution of changes in CO_2, land-use and climate to trends in EPEs are estimated individually for each tail, response variable, region and season. We assumer linear trend slopes over the 25-year period and computed these in both tails separately (illustrated here for the negative tail, ${\beta_{\mathbf{All}}}_s^{-extreme}$),

$$ {\beta_{\mathbf{All}}}_s^{-extreme} = \frac{\Delta(x^{-extreme}_{\cdot,\cdot,s,All})}{\Delta t}. \tag{8.4}$$

The contribution of trends in CO_2 (${\beta_{\mathbf{CO2}}}_s^{-extreme}$), land-use (${\beta_{\mathbf{LU}}}_s^{-extreme}$), and climate (${\beta_{\mathbf{climate}}}_s^{-extreme}$) to changes in the response variable is determined from factorial model simulations, i.e.

$$ {\beta_{\mathbf{CO2}}}_s^{-extreme} = \frac{\Delta(x^{-extreme}_{\cdot,\cdot,s,All} - x^{-extreme}_{\cdot,\cdot,s,CONSTCO2})}{\Delta t} \tag{8.5}$$

$$ {\beta_{\mathbf{LU}}}_s^{-extreme} = \frac{\Delta(x^{-extreme}_{\cdot,\cdot,s,All} - x^{-extreme}_{\cdot,\cdot,s,CONSTLU})}{\Delta t} \tag{8.6}$$

$$ {\beta_{\mathbf{climate}}}_s^{-extreme} = \frac{\Delta(x^{-extreme}_{\cdot,\cdot,s,CONSTLUCO2})}{\Delta t}. \tag{8.7}$$

To further examine *climate-related* drivers of change in ecosystem productivity, we analyse the individual contribution of trends in temperature, precipitation and radiation to ${\beta_{\mathbf{climate}}}_s$). A simple statistical attribution framework is presented in the Supplementary Material to this Chapter based on the 'CONSTLUCO2' scenario.

Spring-summer interacting carbon cycle effects due to climate extremes
In Section 8.3.3, we identify all ensemble members that experience a negative EPE in summer (June-September, i.e., $x_{j,t,s,All}^{-extreme,JJAS}$) using a time-dependent 10th percentile threshold. To detect spring compensation effects, we analyse the preceding spring conditions in the identified ensemble members in terms of ecosystem productivity anomalies that might potentially alleviate carbon losses in summer. Furthermore, the contribution of carry-over effects from spring to negative summer EPEs (e.g. via soil moisture depletion) is disentangled using factorial model simulations by analysing the difference between the 'All' and 'SPRINGRAND' simulations, i.e. with identical summer meteorology in both factorial simulations, but randomised spring meteorology. Hence, we compute spring-summer carry-over effects as the difference between the identified negative summer EPEs in both scenarios.

8.3. Results

In this section, we firstly illustrate in one region how large ensembles of climate-ecosystem model simulations can be used to study EPEs (Section 8.3.1) and, secondly present a systematic assessment of spring and summer trends in EPEs and an attribution to drivers (Section 8.3.2). Lastly, we investigate spring-summer interacting carbon cycle effects due to climate extremes (Section 8.3.3).

8.3.1. An illustrative attribution analysis of ecosystem productivity extremes

The probability distributions of monthly GPP and NEP from the LPJmL ensemble in CEU-FRA for an earlier (1986-1995) and a more recent (2001-2010) period reveal an overall upward shift of GPP and NEP in spring but more nuanced changes in summer (Figure 8.3a for NEP and Figure E3a for GPP). To investigate these changes in more detail, we apply an extreme value analysis to the tails of the probability distributions in both periods. Return time plots (Figure 8.3b-e for NEP and Figure E3b-e for GPP) have been used widely in event attribution studies (National Academies of Sciences, Engineering, and Medicine, 2016) to scrutinise the tails of a distribution by plotting the magnitude of an extreme event

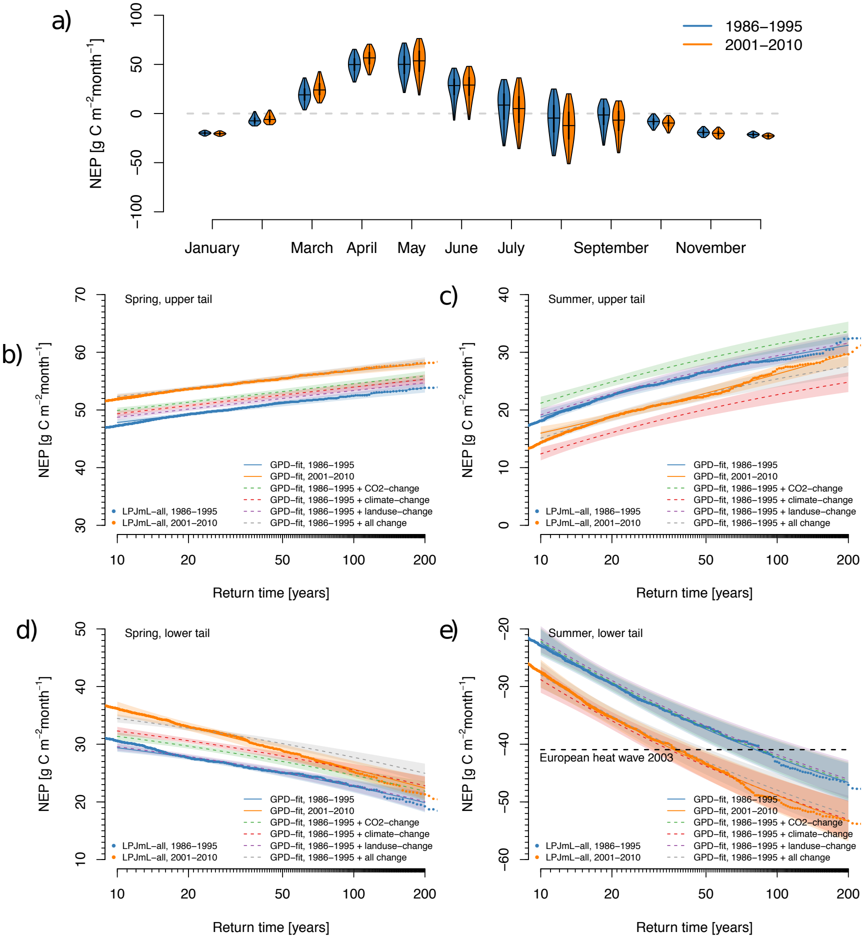


FIGURE 8.3.: a) Seasonal cycle of NEP distribution as simulated by the LPJmL-ensemble for 1986-1995 and 2001-2010 in the France subregion. (b–e) Return time plots of seasonal NEP extremes (i.e. plotting the magnitude of an extreme event as a function of return time) in spring (b,d) and summer (c,e) for the upper (b,c) and lower (d,e) tail of the distribution for 1985-1995 and 2001-2010 (solid blue and orange lines, respectively, derived from fitting a Generalized Pareto Distribution (GPD) to threshold exceeding extremes in each tail, c.f. Coles et al. (2001)). b–e) Differences between the blue and orange lines indicate how the likelihood of extremes occurring has changed between the two compared decades. To illustrate the relative importance of individual drivers, we also plot the effects of changes in NEP that are driven individually by CO_2, land-use, and climate from factorial model simulations depicted by the dashed lines, following Eqs. 8.1-8.3. Dots indicate individual ensemble members.

as a function of return time. Here, an event in the upper (lower) tail with an average return time of 20 years corresponds to a 95th (5th) percentile event when using annual data.

Differences between the blue and orange lines in Figure 8.3b-e indicate how the likelihood of EPEs occurring has changed between the two compared decades for a given season and extreme type. In spring, terrestrial ecosystems exhibit an increase in GPP and NEP under extreme conditions in the upper and lower tail of the distribution in the more recent period (both tails shifted upward for any given return time, Figure 8.3b,d and figure E3b,d). These increases are driven by a roughly equal positive contribution of climate and CO_2 changes in the upper tail, and a larger contribution of climate change in the lower tail, in particular for GPP (Figure E3d). Changes in the tails of the GPP distribution between both periods that are induced by individual drivers in the ecosystem model are largely additive, i.e. the average contribution of changes in CO_2, land-use, and climate added to the statistical model for the 1986-1995 tail matches the statistical fit for the 2001-2010 tail (Figure 8.3b-e and Figure E3).

Changes in summer GPP are close to neutral, because the negative response to climate change is compensated by a positive contribution of CO_2. NEP has significantly reduced (Figure 8.3c,e), predominantly due to negative climate effects. For illustration, the European heat wave and drought of 2003 (Figure 8.2e, dashed horizontal line) results in a roughly 1-in-80 year event in the 1985–1995 decade but is already a 1-in-35 year event in the recent period. While the difference between the two decades used in this study is not comparable to a counterfactual climate simulation as utilised in other attribution studies (Mitchell et al., 2016a) it is reasonable to assume that the main difference in the climate simulations and thus NEP simulations comes from anthropogenic climate change.

Because ecosystem responses to climate extremes are often highly nonlinear and asymmetric depending on the type of extreme, changes in the likelihood of EPEs as discussed here are likely different from risk ratios based on meteorological variables alone (Stott et al., 2004, 2013). This study therefore exemplifies a simulation of the whole chain of events from meteorology to ecosystem responses in extreme event attribution (Stone and Allen, 2005) and presents a framework for studying extreme ecosystem impacts.

8.3.2. Attribution of trends in ecosystem productivity extremes

Across all six European regions, trends towards increased gross productivity in spring for both positive and negative EPEs from 1986-2010 confirm a general upward shift in the GPP distribution (Figure 8.4a) that is driven by both climate and CO_2 changes. The pattern of an upward shift in spring is also found for NEP, but to a smaller extent that can be explained by a smaller sensitivity to recent changes in CO_2 and climate (Figure 8.4b). This is because recent climate change and CO_2 fertilisation are not only enhancing primary productivity in spring but also ecosystem respiration, causing a smaller net response. Positive GPP trends are generally more than twice as large as NEP trends, i.e. less than half of the increased carbon uptake remains in the system after increased respiratory losses are accounted for, which is a consistent pattern for both positive and negative EPEs.

In summer, the response of ecosystem productivity to recent climate change reverses (with few exceptions), but remains positive for CO_2 changes: Hence, predominantly negative ecosystem productivity responses to recent climate change are balanced by a positive response to CO_2 change, causing a mix of slightly increased (two regions), close-to-neutral (three regions) and reduced (one region) gross carbon uptake. Summer increases are confined to energy-limited regions in northern Europe (NEU-SCA and CEU-RUS) and more pronounced for the upper tail of GPP - because the response of positive EPEs to recent climate change is marginally positive (in contrast to the other regions, Figure 8.4a). Similar to spring, summer NEP trends are generally smaller in magnitude than GPP trends, and almost exclusively negative. The observed negative trends in summer ecosystem productivity and EPEs are most pronounced in water-limited regions in southern Europe (MED-SEE, MED-ESP, CEU-FRA) with relatively similar trend slopes in the upper and lower tail. The energy-limited regions in northern Europe experience reduced summer productivity under negative EPEs, but small increases in NEP under positive EPEs due to slightly different climate respones in the upper and lower tail (Figure 8.4b).

Overall, LPJmL ensemble simulations reveal that seasonally contrasting responses of EPEs to changing climate conditions will be a crucial factor in determining regional-scale carbon balances in the near future. Further analyzing climate-induced *trends* in spring and summer ecosystem productivity ($\beta_{\mathbf{climate}_{MAM}}$ and $\beta_{\mathbf{climate}_{JAS}}$) in the Supplementary Material to this Chapter reveals that climate-induced positive productivity trends are mostly driven by warming temperatures in spring, whereas the ecosystem response to summer warming is negative for NEP and GPP across Europe (except GPP in NEU-SCA).

8.3.3. Elucidating spring-summer interacting carbon cycle effects due to climate extremes

In 2012, the contiguous United States experienced a very warm spring followed by an extreme summer drought. Wolf et al. (2016) hypothesised that warmer spring conditions and elevated spring plant activity might have induced soil moisture deficits, thereby exacerbating the impacts of summer drought (Figure 8.1). Here, we analyse lagged effects in all ensemble members that experience extreme reductions in summer productivity[3] (negative EPEs). Specifically, we investigate

a) whether productivity losses induced by summer droughts are (increasingly) compensated by warmer spring conditions ('spring compensation', conceptual link (a) in Figure 8.1), and

b) whether spring-summer 'carry over effects' via soil moisture depletion further exacerbate negative EPEs in summer (conceptual link (b) in Figure 8.1)?

The conditional selection of summer extremes over NEU-ENG (Figure 8.5a) shows that negative summer extremes can be preceded by various ecosystem productivity conditions in spring (Figure 8.5a), i.e. there is no obvious deterministic link. However, there is indeed a *probabilistic* link between carbon cycling under summer extremes and the preceding spring productivity conditions, as four out of five European regions show -on average- increased spring GPP that compensates to a small extent for summer reductions (2.7-19.0% average compensation, Table 8.2), but smaller effects are observed for NEP (-4.3% to +7.8%). The MED-SEE region is an exception where summer extremes co-occur with on average reduced spring productivity (-4.6% in GPP and -10.4% in NEP of the summer

[3] The subregion over Spain is excluded from the analysis because seasonality in ecosystem productivity differs strongly from other European regions, see Figure 8.2

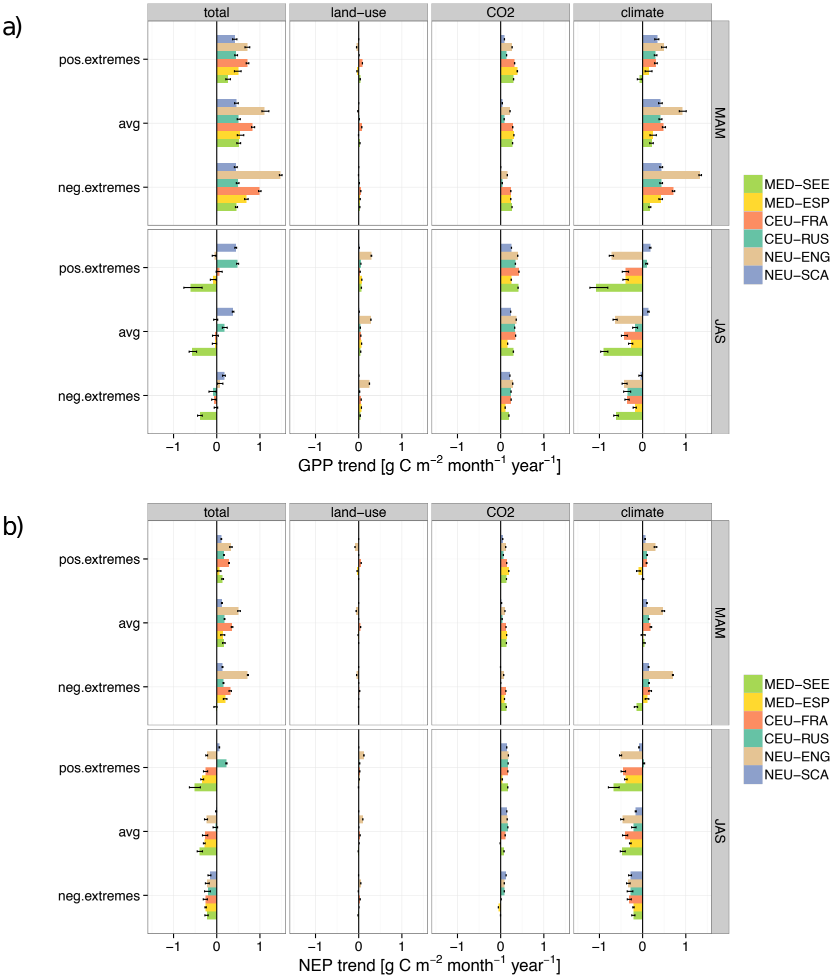


FIGURE 8.4.: Factorial attribution of spring (MAM) and summer (JAS) trends in EPEs in six European regions to changes in land-use, CO_2 and climate, for (a) GPP, and (b) NEP, as simulated by LPJmL.

anomaly are in addition lost in spring). Moreover, elevated ecosystem productivity in spring (GPP, and less so NEP) is increasingly compensating reductions in summer productivity in all European regions over the past 25 years (Figure 8.6), albeit average spring compensation of negative EPEs in summer can only account for a fraction of the summer anomaly. These trends might be a consequence of seasonally contrasting trend slopes (Section 8.3.1 and Section 8.3.2).

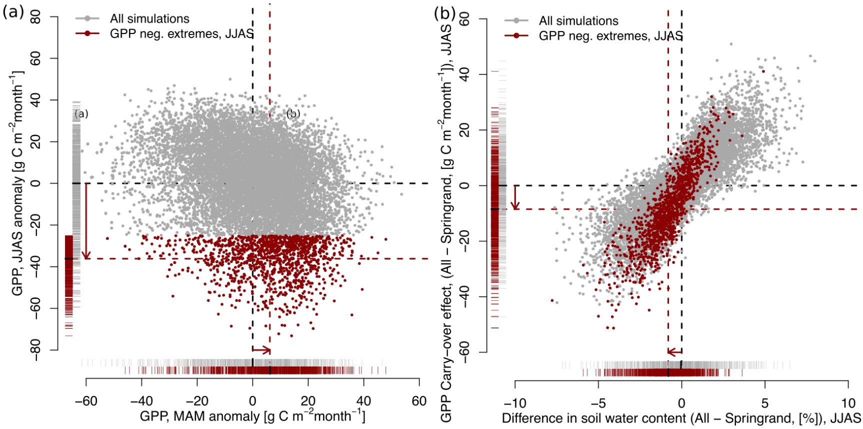


FIGURE 8.5.: Spring-summer interacting carbon cycle effects due to climate extremes illustrated in one region (NEU-ENG). a) Association of spring anomalies with summer (June-September) anomalies in individual ensemble members. b) Differences in soil water content explain spring-summer carry-over effects in the carbon cycle. The average spring compensation and contribution of carry-over effects to negative EPEs in summer are indicated by a horizontal red arrow in (a) and vertical red arrow in (b), respectively. Marginal distributions are plotted at the edge of each plot as individual ticks for all ensemble members (gray) and negative summer extremes (red).

Is there a causal link between spring carbon cycling and summer extremes? Carry-over effects from spring to summer contribute on average 8.3-23.5% for GPP (6.0-19.1% for NEP) to the magnitude of extreme productivity reductions in summer (negative EPEs, see Figure 8.5b for an illustration). This carry-over contribution is revealed by analysing differences in summer EPEs in the 'All' and 'SPRINGRAND' simulations (Table 8.1), where summer meteorology is identical but spring conditions randomised in the latter simulation. Hence, summer ecosystem productivity extremes would be less severe if they would have been preceded by *random* spring conditions. The carry-over effects simulated by LPJmL are due to soil moisture depletion, because differences in soil moisture content explain a large fraction of the magnitude of carry-over effects across all regions (Table E2, Figure 8.5b for NEU-ENG). These carry-over effects have been largely stable over the last 25 years (Figure 8.6).

In summary, the analyses presented here provide an independent process model explanation and generalisation of the observed seasonal compensation mechanism (Wolf et al., 2016). However, we find that the average spring compensation of summer extremes is relatively small for GPP, almost neutral for NEP, and even negative (spring amplification of summer extreme) in MED-SEE for both GPP and NEP. Conversely, carry-over effects from spring to summer extremes via soil moisture play an important role in shaping simulated EPEs and exacerbate carbon cycle impacts on average. Hence, a substantial contribution of compensation effects (as observed for the 2012 US event, Wolf et al., 2016) cannot generally be expected at present in Europe, and the role of these effects remains to be quantified on larger spatial scales, including uncertain long term legacy effects of climate extremes (Anderegg et al., 2015). Furthermore, positive compensation trends as found for recent years (Figure 8.6) cannot continue indefinitely, simply because there are natural limits to shifts in ecosystem phenology (Körner and Basler, 2010) and plant physiological responses to warming (Norby and Luo, 2004).

8.4. Discussion

The results of our study provide evidence that EPEs in European ecosystems show a seasonally contrasting response to changes in climate when investigated using a large ensemble of ecosystem model simulations. Spring climatic changes tend to shift the GPP and NEP distribution upwards (including extremes in the upper and lower tails), whereas climatic changes in summer, most notably warming, lead to approximately neutral (GPP) or even negative trends (NEP), i.e. intensified carbon losses under climate extremes. Further, summer carbon losses as a result of climate extremes are partly compensated by a higher uptake in the preceding spring in temperate regions, but these spring compensatory effects are largely undone through a negative carry-over effect from spring to summer via depleted soil moisture, which further exacerbates summer carbon losses. Hence, our analyses

TABLE 8.2.: Spring compensation of summer extremes in GPP and contribution of dynamical effects.

Region	Variable	Summer extreme	Spring compensation			Carry-over effect		
		Mean (gC m^{-2} $month^{-1}$)	Mean (gC m^{-2} $month^{-1}$)	Mean[a] (% of summer anomaly)	Trend[a,b] (% $year^{-1}$)	Mean (gC m^{-2} $month^{-1}$)	Mean (% of summer anomaly)	Trend[b] (% $year^{-1}$)
NEU-SCA	GPP	-24.0	5.6	19.0	1.7 *	-2.8	11.5	-0.1
NEU-ENG	GPP	-36.1	6.2	13.2	2.1 *	-8.5	23.5	-0.4 *
CEU-RUS	GPP	-38.2	5.7	11.3	1.0 *	-3.3	8.3	0.0
CEU-FRA	GPP	-33.7	1.0	2.7	1.7 *	-4.8	14.4	-0.1
MED-SEE	GPP	-34.3	-2.1	-4.6	0.7 *	-3.1	8.6	0.1
NEU-SCA	NEP	-19.5	2.0	7.8	0.3 *	-1.9	9.1	0.2 *
NEU-ENG	NEP	-27.2	1.5	3.8	1.2 *	-5.1	19.1	-0.8 *
CEU-RUS	NEP	-32.2	1.4	3.3	0.5 *	-2.0	6.0	0.0
CEU-FRA	NEP	-27.3	-1.6	-4.3	1.0 *	-2.9	11.1	-0.4 *
MED-SEE	NEP	-24.9	-3.5	-10.4	0.4 *	-2.1	8.7	-0.4 *

[a] The sign is reversed for the computation of spring compensation relative to the summer anomaly for ease of understanding.
[b] Significance of the trend slopes at the 5% confidence level is indicated by an asterics (*).

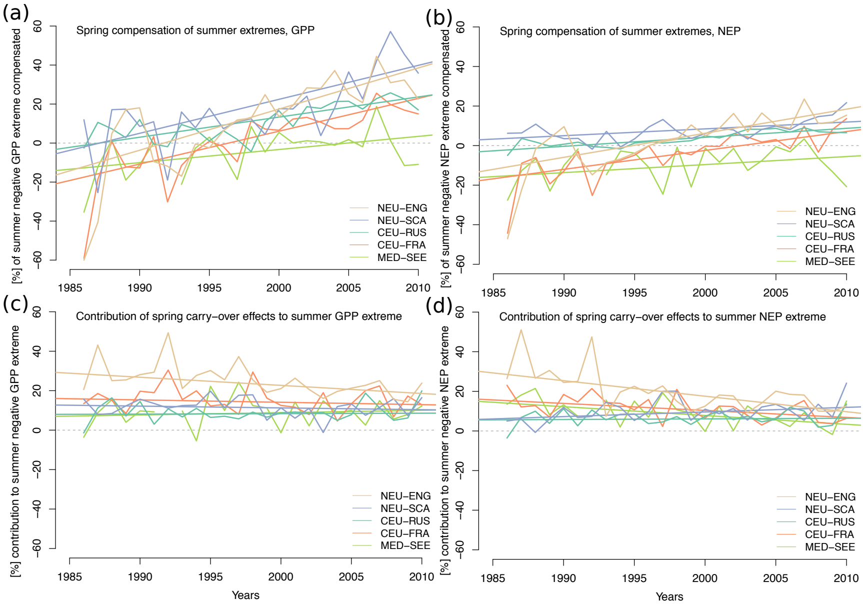


FIGURE 8.6.: Trends and interannual variability in spring-summer interacting carbon cycle effects due to climate extremes in summer. a, c) Spring compensation of negative summer EPEs in (a) GPP, and (b) NEP. b, d) Contribution of spring-summer carry-over effect to negative summer EPEs in (b) GPP, and (b) NEP.

provide a model generalisation and interpretation of seasonal compensation and carry-over effects of carbon-cycle extremes.

However, the results of the present analysis might be confined by the fact that the underlying climate ensemble is based on just one regional climate model and uncertainties related to simulated trends, changes in (individual) climate variables, potential feedback mechanisms, and the applied bias correction remain (Massey et al., 2015; Sippel et al., 2016a).

Ecosystem models are derived from well-established theory of plant-atmosphere carbon exchange (Bonan, 2015), and are widely analysed in the context of climate extremes (Ciais et al., 2005; Reichstein et al., 2007; Zscheischler et al., 2014a). Nonetheless, the results presented here can still be influenced by scale mismatches, where models scale carbon assimilation from leaf to the ecosystem scale (Rogers et al., 2017), or ecophysiological processes are simulated without considering a diurnal cycle and averaged over 0.5° grid cell size.

Moreover, a possible caveat of the present study is that ecosystem and carbon cycle models tend to overestimate the response of terrestrial carbon cycling to drought conditions if compared to observations-based datasets (Huang et al., 2016). LPJmL and related earlier versions have been shown to overestimate the sensitivity of ecosystem productivity to precipitation deficits in central European regions as compared to tree ring data (Babst et al., 2013; Rammig et al., 2015), albeit qualitative responses are largely captured (Rammig et al., 2015). Temperature extremes that are not associated with precipitation deficits are not affected (ibid.). On the continental scale in Europe, extremes in LPJmL simulated GPP respond more sensitively to climate extremes than data-driven products, but in a qualitatively consistent way considering for example the ratio between positive and negative GPP extremes in Europe (Zscheischler et al., 2014c). In this context, comparing the upper and lower tail of simulated ecosystem productivity in this study (Figure 8.3c vs. Figure 8.3e) reveals that extreme carbon losses in the lower tail are larger in magnitude than gains due to positive EPEs for a given return period (slopes in the return time plots in the lower tail exceed those in the upper tail for both GPP and NEP). This asymmetry in EPEs is consistent with analyses at the continental and global scale in observations-based products (Zscheischler et al., 2014c). Van Oijen et al. (2014) compares ecosystem productivity simulations and the vulnerability to precipitation deficits to satellite observations of vegetation greenness and finds that LPJmL (and other vegetation meodels) largely reproduce spatial patterns across Europe. Furthermore, the LPJmL version used in the present study incorporates a phenology scheme that improves phenological dynamics and variability of FPAR (Forkel et al., 2015), and thus might overcome one of the previously identified key weaknesses of earlier LPJmL versions (Mahecha et al., 2010a).

Nonetheless, the analysis and attribution of simulated EPEs ignores a number of ecosystem processes and potential feedbacks between these, as these are missing in the LPJmL ecosystem model (e.g. wind disturbance, pests, nitrogen and phosphorous limitations) and generally many ecosystem processes and feedbacks during climatic extreme events are still unknown or uncertain (Reichstein et al., 2013; Frank et al., 2015). Hence, model improvements can only be conducted in synthesis with improving our process understanding of climatic extreme events. Therefore, dedicated ecosystem manipulation experiments (Knapp et al., 2002; Jentsch et al., 2007; Beier et al., 2012) will be crucial to evaluate and scrutinise model predictions.

Despite these caveats, we argue that the analyses and tools presented here are useful to investigate specific hypotheses related to extremes in terrestrial ecosystems. Our approach allows a physically consistent probabilistic assessment of extremes in ecosystem productivity. Because the outlined probabilities and return times of EPEs are based on one ecosystem model, they should not be taken at face value, but rather be regarded as an approach to scrutinise model sensitivities and attribute drivers behind contemporary changes in ecosystem risk on decadal time scales.

An application of the analysis metrics developed for this study to other process-oriented ecosystem models or data-driven approaches (Tramontana et al., 2016) could be one way to sample respective ecosystem model uncertainties, and to further scrutinise various hypotheses about interacting and contrasting contemporary changes in the frequency and intensity of ecosystem productivity extremes. Thereby, our suggested ensemble analyses might complement state-of-the-art ecosystem risk assessments (Van Oijen et al., 2014; Rolinski et al., 2015) and possibly guide ecosystem manipulation experiments towards pinpointing the most relevant and uncertain drivers of contemporary change in ecosystem extremes.

8.5. Conclusion

In this paper, we illustrate large ensemble simulations of ecosystem productivity as a useful tool to explore variability and change in EPEs from a probabilistic perspective. The approach allows to identify the drivers of changes in EPEs using attribution-type analyses (Stott et al., 2013) and to analyse interacting carbon cycle effects caused by climate extremes (i.e. compensatory and carry-over effects). We find contrasting trends in spring vs. summer carbon cycle extremes in six eco-physiologically different European regions. A recent upward shift in the distribution of spring ecosystem productivity (including extremes) can be attributed to recent climate warming and CO_2 increases, whereas in summer, ecosystem extremes are intensifying for NEP (i.e. more carbon lost to the atmosphere under drought and heat conditions) and roughly stable for GPP, despite a positive response to increasing CO_2. Despite these overarching trends, regional differences are emerging, in that water-limited regions in South Europe show smaller trends in spring, hence benefitting to a smaller degree from warming, while negative trends in summer net ecosystem productivity and its extremes are least pronounced in temperature-limited northern regions.

Furthermore, spring GPP increasingly compensates negative EPEs in summer GPP in four out of five European regions. However, this compensation occurs only partly, on average in the range of 2.7–19.0% of the summer anomaly, but depends on the definition of extremes (Figure 8.5). Spring compensation effects and trends are smaller but mostly positive for NEP. Conversely, spring-summer carry-over effects exacerbate carbon cycle losses under summer extremes (contribution of 8-23% in GPP and 6-19% in NEP to summer anomaly), thereby counterbalancing and undoing positive compensation-related effects. Therefore, we expect that climate extremes increasing in frequency and intensity (IPCC, 2012) might further exacerbate legacy effects of ecosystem extremes in the long term beyond the actual events (Anderegg et al., 2015).

Part IV.

9. Conclusions and outlook

A few years after contemplating about return times of severe storm surges that have been causing disastrous impacts on coastal villages in the past, Hauke Hain, the Dykemaster and central character in Theodor Storm's Schimmelreiter (quoted in the Introduction of this thesis), finished the construction of a new, flatter dyke at some sections of the coastline. However, a few years further on, he finds the dykes and himself amidth a century flood and storm surge that he had not expected in that intensity:

'"Der Wind ist umgesprungen!" rief er "nach Nordwest, auf halber Springfluth! Kein Wind; - wir haben solchen Sturm noch nicht erlebt!" (...) *Da sank aufs Neu' ein großes Stück des Deiches vor ihm in die Tiefe, und donnernd stürzte das Meer sich hinterdrein; (...) dann ritt er an den Abgrund, wo unter ihm die Wasser, unheimlich rauschend, sein Heimathsdorf zu überfluthen begannen; (...) aber unten auf dem Deiche war kein Leben mehr, als nur die wilden Wasser, die bald den alten Koog fast völlig überfluthet hatten.'*

These quotes round off the tale of the Schimmelreiter[1] ('The Rider on the White Horse'). In short, a century flood and storm surge unfolds that severely threatens the dykes that protect the coastal village. However, Hauke Haien decides not to cut off 'his' newly constructed dyke, which might have, according to the legend, relieved pressure from the old dykes and possibly inundated uninhabited lands. Eventually, the old dykes break, causing a severe flood disaster in the coastal village.

This tale serves as a reminder and analogy of two rather general but crucial points that might help to put the findings of this thesis into a broader context:

[1]Theodor Storm. 1888/2011. Der Schimmelreiter ('The Rider on the White Horse'). ISBN 3458362169.

First, climate extremes are an essential part of climatic variability. Hence, incidences of climate extremes will recur with certainty in the future, but a comprehensive physical understanding or prediction of these events still constitutes an enormous scientific challenge (Zhang et al., 2014). Nonetheless, analysing the statistics of climate extremes that occurred in the past (e.g. Alexander et al., 2006), and scrutinising the drivers behind these events (Otto et al., 2016), one might be able to learn about the statistical properties and probabilities of climate extremes occurring, and how these might be changing due to various and interacting drivers. The present dissertation contributes in this context by scrutinising statistical quantification methodologies (Part I), and tools to improve the interpretation and bias correction of climate model ensemble simulations (Part II).

Second, the impact of climate extremes on socio-economic and ecological systems is often highly nonlinear (e.g. in the case of dyke breaks) and mediated by various external drivers or system properties, and management decisions (see e.g. IPCC (2012) for an in-depth discussion, or Reichstein et al. (2013) with a focus on terrestrial ecosystems). In this context, this thesis presents tools for an explicit impact assessment of climate extremes in the terrestrial biosphere (Part III) that can be used for instance to disentangle seasonally interacting drivers and nonlinear effects of climate extremes in ecosystem carbon cycling.

The overarching objective of the present PhD thesis is to improve the quantification of, and contributing to the understanding of climate extremes and their impact on ecosystem-atmosphere interactions. To achieve these goals, the thesis explores a wide range of generic methodological considerations (Part I), approaches to enable sound process-oriented model ensemble simulations using observation-based constraints (Part II), towards an attribution of ecosystem impacts arising from climate extremes (Part III). Overall, the thesis lays out a comprehensive framework for systematically quantifying and attributing the impacts of climate extremes in the terrestrial biosphere using joint analyses of observations and model ensembles. While the generic methodological issues in Part I (Chapters 2–4) are of general interest for everyone dealing with the robust quantification of climate extremes and variability in relation to some reference period, Part II (Chapters 5–7) is more specific to improve the interpretation and bias correction of model-ensemble simulations through observation-based constraints. In contrast to conventional statistical bias correction, these approaches retain the physical consistency of the original model simulations, which is of crucial relevance for any assessment of climate extremes or their impacts in the terrestrial biosphere. Finally, Part III (Chapters A and 8) illustrates an assessment of extreme responses in ecosystem productivity to climate extremes using an ensemble of climate-ecosystem model simulations. These analyses are used to (1) attribute trends in the intensity of ecosystem productivity extremes to various drivers, and (2) disentangle effects of seasonally interacting carbon cycle effects due to climate extremes.

Overall, the thesis shows that firstly, scrutinising statistical methods and diagnostics, and evaluating observation-based constraints on model ensembles, are key to an improved understanding as well as quantification of climate extremes and their impacts. Secondly, a consequent probabilistic interpretation of climate-ecosystem model ensemble simulations offers novel perspectives on the mechanistic pathways and interacting effects of terrestrial ecosystem responses to climate extremes. In the following, I derive an outlook on the implications for each of the tree Parts of this thesis (Sections 9.1–9.3), and also highlight future research needs emerging from the findings of this thesis. Finally, I adopt a data-driven perspective on ecosystem functioning and explore illustratively how climate model ensemble simulations could be combined with purely data-driven ecosystem models to complement process-oriented ecosystem model simulations (Section 9.4).

9.1. Statistical quantification of extremes

In Part I of this thesis, I revisited several methodological choices that contribute to a robust quantification of spatially aggregated climate extremes in observational or simulated gridded datasets (Chapters 2–4). Chapter 2 and 3 showed that conventional statistical methodologies that are based on a reference period standardisation of gridded data can impose substantial biases outside the reference period on spatially or temporally aggregated estimates of climate extremes. For example, the occurrence of 'two-sigma extremes' could be overestimated by 48.2% compared to a reference period of 30 years in randomly Gaussian distributed data in the absence of any trends. This phenomenon occurs because the statistical estimators of mean and standard deviation, used for the standardisation procedure, are statistically dependent on the reference period but independent from periods outside the reference period.

Hence, these findings highlight that assessing the robustness of methodological choices is a crucial first step in analyses of climate extremes in spatio-temporal datasets. In particular, as these analyses often attract media attention, and are sometimes related to anthropogenic climate change[2], robust and transparent methods for quantification are simply essential to enable informed public discourse.

Several implications follow specifically from the normalisation-induced biases that I will briefly illustrate in a few examples that go beyond the issues discussed in Part I:

- The use of 30-year reference periods to derive grid-cell based statistics such as the sample mean or sample standard deviation for normalisation does not warrant accurate analyses of spatio-temporal climate extremes, although 30-year reference periods might appear as common practice (WMO, 1989; Donat et al., 2017).

- Standard climate datasets that are used widely in the climate or climate impact community are computed from gridded monthly station-based anomalies in a common reference period such as the CRU-TS datasets (Harris et al., 2014). Therefore, some noteworthy properties of these datasets, such as reduced spatial variability within the reference period (Tingley, 2012), might stem at least partly from the anomaly-based generation of these datasets. For example, consider independent and identically distributed Gaussian data in a surrogate 'true and perfect climate' in any given month:

$$X_{t,m,i} \sim \mathcal{N}(\mu, \sigma^2), \tag{9.1}$$

[2] http://www.nytimes.com/2012/08/07/science/earth/extreme-heat-is-covering-more-of-the-earth-a-study-says.html

where i is a grid cell index over a large number of grid cells and t is an arbitrary time step in an arbitrary month m. For any given reference period length n_{ref}, the reference period sample means in each grid cell and month follow $\hat{\mu}_{m,i} \sim \mathcal{N}(\mu_{\text{ref},m,i}, \frac{\sigma^2}{n_{ref}})$, and anomalies are calculated such as $X_{anom,t} = X_t - \hat{\mu}_{ref}$ (subscript i and m are dropped for convenience). Hence, generating a gridded anomaly dataset that is based on a reference period that does not cover the full time period will lead to a spatially different distribution inside and outside the reference period, respectively:

$$X_{anom,t\in\{obase\}} \sim \mathcal{N}(0, (1 + \frac{1}{n_{ref}})\sigma^2), \text{ and} \tag{9.2}$$

$$X_{anom,t\in\{ref\}} \sim \mathcal{N}(0, (1 - \frac{1}{n_{ref}})\sigma^2). \tag{9.3}$$

Obviously, differences in the variability of the spatial distributions could impose deleterious biases on the detection and quantification of climate extremes in gridded datasets. In further research, phenomena of this kind could be further investigated and potentially addressed analytically as shown in Chapters 2 and 3.

- Statistical bias correction methodologies typically calibrate a relationship between a simulated variable (e.g. T_{mod}) and the corresponding observed variable (T_{obs}) based on a long-term probability distribution (Maraun, 2016) in a given reference period for which observations are available. This relationship is then often extrapolated to future simulations (see e.g. Hempel et al., 2013). Assuming uncorrelated transient climate model runs and some 'perfect and true' observations, consider a simple example of monthly mean bias correction: We assume that $X_{mod,t,m,i} \sim \mathcal{N}(\mu_{mod,t,m,i}, \sigma^2_{mod,t,m,i})$, and $X_{obs,t,m,i} \sim \mathcal{N}(\mu_{obs,t,m,i}, \sigma^2_{obs,t,m,i})$ with the subscripts t, m, i denoting time, month, and grid cell as above.

  Subtracting the mean bias defined as the difference between the sample means of the observations and model simulations in the reference period yields statistically bias corrected time series (for each month m and grid cell i, subscripts dropped for convenience):

$$X_{mod-cor,t} = X_{mod,t} - \hat{\mu}_{mod} + \hat{\mu}_{obs} = X_{anom,t} + \hat{\mu}_{obs} \tag{9.4}$$

Because $\hat{\mu}_{obs} \sim \mathcal{N}(\mu_{obs}, \frac{\sigma^2_{obs}}{n_{ref}})$, and with Eq. 9.2 and 9.3 above, it follows:

$$X_{mod-cor,t\in\{obase\}} \sim \mathcal{N}(\hat{\mu}_{obs}, (1+\frac{2}{n_{ref}})\sigma^2_{mod}), \text{ and} \tag{9.5}$$

$$X_{mod-cor,t\in\{ref\}} \sim \mathcal{N}(\hat{\mu}_{obs}, \sigma^2_{mod}). \tag{9.6}$$

Therefore, care is needed when analysing spatio-temporal variability or extremes in datasets that have been bias corrected or otherwise statistically pre-processed based on a fixed reference period, which remains a standard method for bias correcting climate data for state-of-the-art impact assessments (e.g. Hempel et al., 2013; Frieler et al., 2016; Mitchell et al., 2017). Similarly, if anomalies are derived relative to a sample mean (e.g. typically used for precipitation), this procedure might not only inflate the variability in the out-of-base period, but might also increase the long-term averages (see Chapter 3 for a detailed discussion), and the examples shown here could be easily extended to this analogous case.

In summary, Chapter 2–3 and the brief examples presented here highlight that any assessments of extreme events in gridded spatio-temporal datasets of climate or ecosystem variables require careful statistical pre-processing and robust detection metrics. Therefore, comprehensive testing and benchmarking of detection algorithms and detection metrics is crucial, particularly as new and more complex indicators are being developed. For example, these include metrics to quantify magnitude and extent of climate extremes simultaneously (Russo et al., 2014, 2015), or to detect multivariate extreme events (Flach et al., 2016). In this context, climate model ensemble simulations that feature large sample sizes might serve as a useful test bed for assessing the statistical robustness of methodological approaches. Chapter 4 illustrates such a benchmarking approach, and compares an empirical analysis of an ensemble simulation to inferences about climate extremes based on extreme value theory, but from smaller available sample sizes. This analysis showed that model ensemble simulations can inform parameter choices for inferences about extreme values in observations that are inherently limited in spatial and temporal extent. In conclusion, integrating methodological sanity checks of this kind into future research practice would provide an important step towards robust science.

9.2. Observation-based constraints to improve the simulation of climate extremes and ecosystem impacts

Part II of this thesis is dedicated to the development and application of methods that constrain climate model ensemble simulations using observational data to reduce biases in climate variables and their extremes, while retaining physical consistency across multiple climate variables. Bias correction is a crucial step for assessing climate impacts (e.g. Frieler et al., 2016; Ahlström et al., 2017), and in particular the impacts of climate extremes (e.g. Chapter 5), but conventional statistical bias correction methods are not ideally suited for this task due to their physical inconsistency and inability to retain feedbacks or the multivariate correlation structure of climate variables (Ehret et al., 2012; Sippel et al., 2016a).

Chapter 5 details a novel bias correction methodology designed to minimise biases in regional climate model ensemble simulations, including simulated climate extremes and ecosystem impacts, while physical consistency is preserved. The method uses the distribution of observed temperatures as a resampling constraint. Using a similar concept, Chapter 7 shows that biases in temperature extremes in multi-model ensembles can be reduced if the ensemble is constrained by suitably chosen diagnostic metrics that are based on observational datasets (here: a land-atmosphere coupling metric). The application of observation-based constraints to climate model ensembles is conceptually similar to the concept of emerging constraints in the climate system (Hall and Qu, 2006; Cox et al., 2013; Wenzel et al., 2014), but instead of seeking relationships between present-day observables and future metrics, we here seek diagnostics that are related to processes that determine present-day model biases. While it is often not straightforward to pinpoint the physical origin of model biases, addressing biases jointly using suitable constraints might constitute a physically consistent way forward, because model biases affect variables not independently from each other. Overall, Part II of this thesis showed that constraint-based approaches that reduce biases in climate model ensembles by resampling or reweighting individual ensemble members constitute a useful avenue for investigating the impacts of climate extremes, and one that might complement conventional statistical bias correction.

However, several methodological caveats remain at present that could be addressed in future research:

- Biases in climate models stem from uncertainties in the representation of many different processes and thus differ among regions, seasons and variables. Therefore, it is unlikely that there exists one or a small set of constraints that can address climate model biases globally in a 'one fits all' manner. For example, Chapter 7 showed that an observation-based land-atmosphere coupling constraint can significantly reduce biases in a multi-model ensemble in regions that are sensitive to the representation of land-atmosphere coupling, but the constraint has no effect elsewhere. Hence, applications of constraint-based bias correction methods presented in this thesis require thorough testing and evaluation. Addressing this caveat by a more objective analysis of constraints for bias correction might further probe the possibilities and limitations of this method for analysing climate-impact simulations. Such enquiries could be embedded in systematic model-observation comparisons, and might benefit from high quality observational datasets as demonstrated by Massonnet et al. (2016).

- At present, conventional statistical bias-correction that is widely used for impact assessments (e.g. Hempel et al., 2013; Frieler et al., 2016) and constraint-based bias correction appear almost mutually exclusive. Therefore, it might be worthwhile to explore whether observation-based constraints that screen out plausible from implausible ensemble members could be combined with conventional bias correction approaches. For example, a two-step approach is conceivable, in which first constraints are applied to pinpoint physically implausible simulations, and second remaining biases that do not compromise physical consistency could be removed statistically. These approaches thus might assist and improve impact assessments in the context of 1.5° vs. 2°C global climate targets (Schleussner et al., 2016b; Mitchell et al., 2016b). Ultimately, however, bias correction methods cannot and should not replace model development and improvement.

9.3. Extremes events in terrestrial ecosystems: Drivers and attribution

In Part III, extensive climate-ecosystem ensemble simulations are scrutinised to investigate how climate extremes, and changes in their occurrence frequency and intensity, might affect carbon cycling in European terrestrial ecosystems. The analyses reveal a seasonally contrasting response of terrestrial ecosystem productivity extremes to recent changes in climate: Spring gross and net carbon uptake is increasing, while extremes in ecosystem productivity in summer point towards reduced carbon uptake, hence higher net carbon release under drought and heat conditions. Moreover, spring-summer interacting carbon cycle effects due to climate extremes are disentangled, which include spring compensatory effects for summer extremes in the ecosystem carbon balance, and contrariwise, carry-over effects from spring to summer that might exacerbate the effect of climate extremes in summer (see Chapter A for a discussion). Chapter 8 shows that both effects play out as crucial factors in the response of European terrestrial ecosystems to climate extremes. However, in the context of spring-summer interactions under climate extremes, it remains an important research question as to whether spring compensatory or carry-over effects will dominate ecosystem responses to these events in the future. In other words, whether ecosystems can sustain their ecophysiological functioning under more intense or frequent future climate extremes through simple temporal compensatory effects, or whether temporal shifts under climate extremes might exacerbate adverse impacts on ecosystem functioning, including potential long-term legacy effects, remains an open research question of high relevance.

The analysis presented in Chpater 8 might serve as a blueprint of how climate-ecosystem ensemble simulations might constitute a useful framework to investigate specific hypotheses related to the response of ecosystem carbon cycling to climate extremes. Climate model ensemble simulations are increasingly becoming available to the scientific community, e.g. to assess future climate targets (Mitchell et al., 2017) or historical runs (see e.g. Angélil et al. (2017) for a description). These ensembles now comprise a larger number of models and are often computed in a spatial resolution that is suitable for the analysis of impacts of climate extremes in various sectors (Mitchell et al., 2017). Therefore, the analysis methods in Chapter 8 also illustrate how ensemble-based impact simulations could be investigated and interpreted in related sectors in future research, such as agricultural or heat-health related impacts (e.g. Mitchell et al., 2017).

Specifically with regard to a comprehensive probing into ecosystem responses to climate extremes, a variety of important research questions and methodological challenges remain to be addressed:

- Climate variability and extremes affect terrestrial ecosystems through complex pathways (Reichstein et al., 2013; Frank et al., 2015). For example, ecosystem carbon uptake under warmer spring conditions prior to extremely dry and hot summers might compensate *temporally* for summer carbon losses (Wolf et al., 2016). Moreover, Jung et al. (2017) showed that inter-annual variability in global land carbon uptake arises from *spatially* compensating responses to local water availability. Such spatial compensation of ecosystem responses might be also expected under climate extremes. Therefore, comprehensive, long-term climate-ecosystem ensembles might serve as an invaluable resource for testing hypotheses of this kind, and might complement observational datasets that are limited in time and space with a probabilistic dimension (Allen and Stainforth, 2002).

- As some climate extremes are changing in frequency and magnitude (IPCC, 2012), the question to what extent ecosystem impacts of climate extremes can be attributed to climatic drivers becomes relevant (Hansen et al., 2016; Otto, 2016). Chapter 8 has shown illustratively how various climatic- and non-climatic drivers affect changes in the odds of ecosystem productivity extremes, but this analyses is contingent on one process-oriented ecosystem model. Therefore, using ensembles of several process-oriented ecosystem models, carefully evaluated against observation-based datasets, could potentially further pinpoint the drivers behind changes in ecosystem productivity extremes for a broader set of models, different biomes, and for future periods - towards addressing the challenge of attributing not only climate extremes, but also extremes in climate impacts (Hansen et al., 2016; Otto, 2016).

The research outlook presented above is based on process-oriented ecosystem models that are designed to simulate ecosystem functioning based on established ecological theory (Bonan, 2015). However, these models can represent ecosystem functioning only to a limited degree, and for instance often overestimate the response to drought (Rammig et al., 2015; Huang et al., 2016). In recent years, data-driven models based on observational datasets and derived through machine learning or statistical techniques have emerged as a complement to process-oriented models (Jung et al., 2011; Tramontana et al., 2016; Jung et al., 2017). Therefore, combining climate model ensembles with data-driven models (see example provided in Section 9.4) might overcome limitations that are inherent to process models–which might thus further contribute to robust scientific answers to the research questions sketched above.

9.4. A data-driven perspective on terrestrial ecosystem productivity based on satellite-derived vegetation productivity proxies

In the final section of this thesis, I illustrate that elaborating the suggested research avenues is straightforward. I explore if one can obtain a purely data-driven perspective on the quantification of terrestrial ecosystem productivity, and their climate drivers, using the ensemble based approach elaborated in Chapter 8. Here, I first train several statistical models of the Fraction of Absorbed Photosynthetically Active Radiation (FPAR) as a satellite-observed vegetation greenness proxy (Gobron et al., 2010) using meteorological observations as predictors. In a second step, I drive the empirical models forward using the regional climate model ensemble and pinpoint the contribution of individual climate variables to recent trends in vegetation greenness.

Methodology The method is largely based on Chapter 8 (and Supplementary Material), but instead of statistically emulating a process-oriented ecosystem model, we use FPAR obtained from the MODIS satellite (MOD15A2, Myneni et al., 2002) in conjunction with gridded reanalysis fields of meteorological variables as predictors. The methodology is briefly reviewed here:

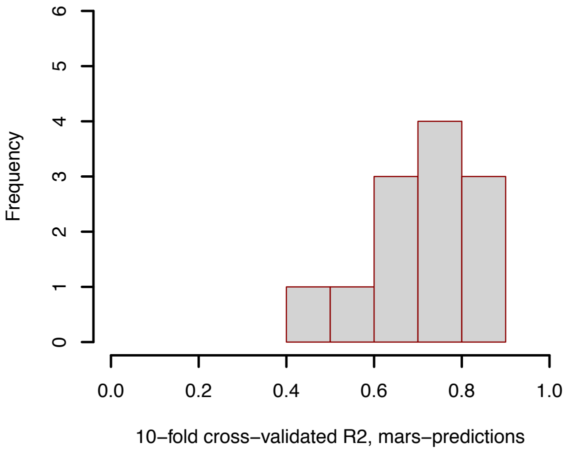


FIGURE 9.1.: Evaluation of empirical estimates of seasonal-scale anomalies in satellite-observed FPAR. Histograms of observations vs. model predictions as determined by R^2, for each region and season after 10-fold pixel-based cross-validation.

1. We train several *additive* regression models (i.e. assuming no interactions between predictor variables), stratified by plant functional type in the time period 2001-2012, using multivariate adaptive regression splines (Friedman et al., 2001):

$$\text{sys}_s = f(\text{env}) = \sum_{m=1}^{k} g_{Tair_i}(\text{Tair}_m) + \sum_{m=1}^{k} g_{Precip_m}(\text{Precip}_m) + \sum_{m=1}^{k} g_{Radiation_m}(\text{Radiation}_m). \tag{9.7}$$

The target variable (sys_s) is seasonal FPAR in 0.5° pixels across the Northern hemisphere, and a separate model is trained for each dominant natural plant functional type in each region (see Table E1 based on MODIS-derived land cover (MOD12Q1, Friedl et al., 2010)) and each season (denoted by subscript s: spring or summer; 12 models in total). Meteorological variables (temperature, precipitation, short-wave radiation) in monthly resolution (subscript m; k denotes the total number of months considered in the statistical model, here: $k = 7$) from ERA-Interim (Dee et al., 2011) in the three concurrent months (i.e. March–May for spring) and four preceding months (Nov–Feb for spring) are used as predictors (explained in detail in Eq. E.3, Supplementary Material to Chapter 8).

2. 10-fold cross-validation of the derived models and aggregation of the predictions to regional averages (see Table E1 for region definition) shows that 10 out of 12 models exceed R^2-values of 0.6 (Figure 9.1), i.e. the meteorological predictors explain a significant fraction of year-to-year variations in FPAR on a regional level.

3. The regional climate model ensemble simulations presented in Chapter 8 are used to predict the statistical FPAR models forward (i.e. separately for the dominant plant functional type in each region and season), and to calculate contributions of individual variables to recent trends in FPAR. The linear trend contributions are calculated as:

$$\frac{\Delta\ \mathrm{sys}_s}{\Delta t} = \underbrace{\sum_{m=1}^{k} \frac{\Delta g_{Tair_m}(\mathrm{Tair}_m)}{\Delta t}}_{\text{Tair contrib.: } \beta_{\mathbf{Tair}}} + \underbrace{\sum_{m=1}^{k} \frac{\Delta g_{Precip_m}(\mathrm{Precip}_m)}{\Delta t}}_{\text{Precip contrib.: } \beta_{\mathbf{Precip}}} + \ldots \quad (9.8)$$

Results and Conclusion The analysis presented here reveals positive trends in FPAR across all six regions in spring (the trend is very small in one Mediterranean region, MED-ESP), but in summer FPAR trends are small with two out of six regions showing negative trends (Figure 9.2). Hence, FPAR trends derived through combining a data-driven model with climate ensemble forcing confirm a seasonally contrasting response of European ecosystems to recent climatic changes (shown in Chapter 8). The strongest positive trends in summer FPAR are observed in forest-covered regions in Northern Europe (NEU-SCA and CEU-RUS).

Positive trends in spring FPAR are driven almost exclusively by warmer concurrent (i.e. spring) temperatures, with a small but consistently positive contribution of warmer preceding (winter) temperatures. For summer FPAR, the contribution of warmer concurrent (summer) temperatures is small (and negative in two regions), and concurrent changes in short-wave radiation appear to increase FPAR by a small amount in all regions. Interestingly, a negative contribution of

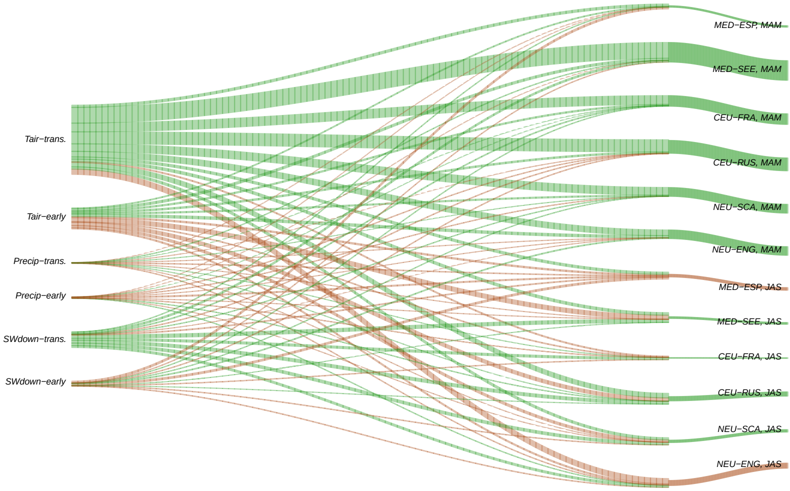


FIGURE 9.2.: Climatic variables that drive recent changes in $FPAR$ across six European regions in spring and summer derived from empirical model predictions. The width of the link between the env-variable (left) and $FPAR$ in each region (right) indicates the contribution of the individual driver (e.g. β_{Tair}, β_{Precip}, etc.). Note that brown colours indicate a negative trend contribution of the respective driver, whilst green colours indicate a positive contribution (trans. - at time of seasonal anomaly, early - before seasonal flux anomaly).

warming in the preceding spring temperatures appears to reduce summer FPAR consistently across all regions. This feature might further point at a crucial role of carry-over effects for ecosystem functioning in a warming climate (potentially induced via soil moisture depletion), which has been observed and discussed earlier for boreal forests at high Northern latitudes (Buermann et al., 2013).

In conclusion, the illustrative data-driven analysis shows that a combination of data-driven climate-ecosystem relationships with a regional climate model ensemble yields estimates of drivers and contemporary trends in ecosystem productivity that are consistent with process-oriented model results presented in Chapter 8. Thus, a combination of data-driven models with regional or global climate model ensembles might provide a productive route for future research into the impacts of climate extremes and variability on terrestrial ecosystem functioning that is complementary to process-oriented ecosystem model ensembles (Section 9.3).

This final section only intended to concretely illustrate one of the many potential avenues that could be followed using the methodology developed in this thesis. Overall, I see highest potential in systematically expanding the approaches to global analyses and different system responses by combining climate model ensembles with both process-oriented *and* data-driven models of ecosystem functioning. Thereby, explicit links between climate extremes, recent changes in their frequency and intensity, the underlying drivers, and ecosystem impacts can be disentangled towards systematic 'end-to-end' attribution studies (Stone and Allen, 2005), thus bridging a crucial research gap that probabilistically links anthropogenic emissions and other climate drivers to observed impacts of climate and its extremes (Hansen and Stone, 2016; Otto, 2016). Other points could be investigating explicit climate extremes to ecosystem impact relationships under large-scale modes of climatic variability such as El Niño, including interacting and compensatory effects of ecosystem responses to climate extremes across space and time (e.g. Wolf et al., 2016; Jung et al., 2017), and extending these approaches beyond ecosystem carbon cycling, e.g. towards agricultural impacts or biodiversity patterns.

10. Bibliography

Abbot, C. 1914. The solar constant of radiation. *Science* **39**(1001), 335–348.

Adeel, Z., U. Safriel, D. Niemeijer, and R. White. 2005. Ecosystems and human well-being: desertification synthesis. *Technical report*. World Resources Institute (WRI).

Adler, D. 2005. *vioplot: Violin plot*. R package version 0.2.

AghaKouchak, A., D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian. 2012. *Extremes in a changing climate: detection, analysis and uncertainty*. Vol. 65. Springer Science & Business Media.

Ahlström, A., G. Schurgers, and B. Smith. 2017. The large influence of climate model bias on terrestrial carbon cycle simulations. *Environmental Research Letters* **12**(1), 014004.

Alexander, L., and S. Perkins. 2013. Debate heating up over changes in climate variability. *Environmental Research Letters* **8**(4), 041001.

Alexander, L. V. 2016. Global observed long-term changes in temperature and precipitation extremes: A review of progress and limitations in IPCC assessments and beyond. *Weather and Climate Extremes* **11**, 4–16.

Alexander, L. V., and J. M. Arblaster. 2009. Assessing trends in observed and modelled climate extremes over australia in relation to future projections. *International Journal of Climatology* **29**(3), 417–435.

Alexander, L., X. Zhang, T. Peterson, J. Caesar, B. Gleason, A. Klein Tank, M. Haylock, D. Collins, B. Trewin, F. Rahimzadeh *et al.* 2006. Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres* **111**(D5), D05109.

Allan, R. P., B. J. Soden, V. O. John, W. Ingram, and P. Good. 2010. Current changes in tropical precipitation. *Environmental Research Letters* **5**(2), 025205.

Allen, C. D., A. K. Macalady, H. Chenchouni, D. Bachelet, N. McDowell, M. Vennetier, T. Kitzberger, A. Rigling, D. D. Breshears, E. T. Hogg *et al.* 2010. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management* **259**(4), 660–684.

Allen, M. 2003. Liability for climate change. *Nature* **421**(6926), 891–892.

Allen, M. R., and D. A. Stainforth. 2002. Towards objective probabalistic climate forecasting. *Nature* **419**(6903), 228–228.

Allen, M. R., and W. J. Ingram. 2002. Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**(6903), 224–232.

Anav, A., P. Friedlingstein, M. Kidston, L. Bopp, P. Ciais, P. Cox, C. Jones, M. Jung, R. Myneni, and Z. Zhu. 2013. Evaluating the land and ocean components of the global carbon cycle in the cmip5 earth system models. *Journal of Climate* **26**(18), 6801–6843.

Anderegg, W. R., C. Schwalm, F. Biondi, J. J. Camarero, G. Koch, M. Litvak, K. Ogle, J. D. Shaw, E. Shevliakova, A. Williams *et al.* 2015. Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models. *Science* **349**(6247), 528–532.

Angélil, O., D. A. Stone, M. Tadross, F. Tummon, M. Wehner, and R. Knutti. 2014. Attribution of extreme weather to anthropogenic greenhouse gas emissions: sensitivity to spatial and temporal scales. *Geophysical Research Letters* **41**(6), 2150–2155.

Angélil, O., D. Stone, M. Wehner, C. J. Paciorek, H. Krishnan, and W. Collins. 2017. An independent assessment of anthropogenic attribution statements for recent extreme temperature and rainfall events. *Journal of Climate* **30**(1), 5–16.

Angert, A., S. Biraud, C. Bonfils, C. Henning, W. Buermann, J. Pinzon, C. Tucker, and I. Fung. 2005. Drier summers cancel out the CO_2 uptake enhancement induced by warmer springs. *Proceedings of the National Academy of Sciences of the United States of America* **102**(31), 10823–10827.

Babst, F., B. Poulter, P. Bodesheim, M. D. Mahecha, and D. C. Frank. 2017. Improved tree-ring archives will support earth-system science. *Nature Ecology & Evolution* **1**, 0008.

Babst, F., B. Poulter, V. Trouet, K. Tan, B. Neuwirth, R. Wilson, M. Carrer, M. Grabner, W. Tegel, T. Levanic *et al.* 2013. Site-and species-specific responses of forest growth to climate across the european continent. *Global Ecology and Biogeography* **22**(6), 706–717.

Babst, F., M. Carrer, B. Poulter, C. Urbinati, B. Neuwirth, and D. Frank. 2012. 500 years of regional forest growth variability and links to climatic extreme events in europe. *Environmental Research Letters* **7**(4), 045705.

Baldocchi, D. 2008. 'Breathing' of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany* **56**(1), 1–26.

Baldocchi, D., E. Falge, L. Gu, R. Olson, D. Hollinger, S. Running, P. Anthoni, C. Bernhofer, K. Davis, R. Evans *et al.* 2001. Fluxnet: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society* **82**(11), 2415–2434.

Baldocchi, D., Y. Ryu, and T. Keenan. 2016. Terrestrial carbon cycle variability. *F1000Research* **5**, 2371.

Bao, J., S. C. Sherwood, L. V. Alexander, and J. P. Evans. 2017. Future increases in extreme precipitation exceed observed scaling rates. *Nature Climate Change* **7**, 128–132.

Barnett, A. G., J. C. van der Pols, and A. J. Dobson. 2005. Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology* **34**(1), 215–220.

Barriopedro, D., E. Fischer, J. Luterbacher, R. Trigo, and R. García-Herrera. 2011. The hot summer of 2010: redrawing the temperature record map of Europe. *Science* **332**(6026), 220–224.

Bastos, A., C. Trigo, R. Trigo, and S. W. Running. 2014. Analysing the spatio-temporal impacts of the 2003 and 2010 extreme heatwaves on plant productivity in Europe. *Biogeosciences* **11**, 3421.

Bauch, M. 2016. Regen, Hunger, Erosion während der Dantean Anomaly. *Die Gegenwart des Mittelalters. Das Magazin der Max Weber Stiftung*. Vol. 01/2016. Deutsche Geisteswissenschaftliche Institute im Ausland. Pp. 6–9.

Bayley, G. V., and J. M. Hammersley. 1946. The 'effective' number of independent observations in an autocorrelated time series. *Supplement to the Journal of the Royal Statistical Society* **8**(2), 184–197.

Beer, C., M. Reichstein, P. Ciais, G. Farquhar, and D. Papale. 2007. Mean annual GPP of Europe derived from its water balance. *Geophysical Research Letters* **34**(5), L05401.

Beer, C., P. Ciais, M. Reichstein, D. Baldocchi, B. Law, D. Papale, J.-F. Soussana, C. Ammann, N. Buchmann, D. Frank *et al.* 2009. Temporal and among-site variability of inherent water use efficiency at the ecosystem level. *Global Biogeochemical Cycles* **23**(2), 1–13.

Beer, C., U. Weber, E. Tomelleri, N. Carvalhais, M. Mahecha, and M. Reichstein. 2014. Harmonized european long-term climate data for assessing the effect of changing temporal variability on land–atmosphere CO_2 fluxes. *Journal of Climate* **27**(13), 4815–4834.

Beier, C., C. Beierkuhnlein, T. Wohlgemuth, J. Penuelas, B. Emmett, C. Körner, H. Boeck, J. H. Christensen, S. Leuzinger, I. A. Janssens *et al.* 2012. Precipitation manipulation experiments–challenges and recommendations for the future. *Ecology Letters* **15**(8), 899–911.

Bellprat, O., and F. Doblas-Reyes. 2016. Attribution of extreme weather and climate events overestimated by unreliable climate simulations. *Geophysical Research Letters* **43**(5), 2158–2164.

Bellprat, O., S. Kotlarski, D. Lüthi, and C. Schär. 2013. Physical constraints for temperature biases in climate models. *Geophysical Research Letters* **40**(15), 4042–4047.

Betts, R. A., C. D. Jones, J. R. Knight, R. F. Keeling, and J. J. Kennedy. 2016. El Niño and a record CO_2 rise. *Nature Climate Change* **6**(9), 806–810.

Betts, R. A., O. Boucher, M. Collins, P. M. Cox, P. D. Falloon, N. Gedney, D. L. Hemming, C. Huntingford, C. D. Jones, D. M. Sexton *et al.* 2007. Projected increase in continental runoff due to plant responses to increasing carbon dioxide. *Nature* **448**(7157), 1037–1041.

Bindoff, N. L., P. A. Stott, M. AchutaRao, M. R. Allen, N. Gillett, D. Gutzler, K. Hansingo, G. Hegerl, Y. Hu, S. Jain *et al.* 2013. Detection and attribution of climate change: from global to regional. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Black, M. T., D. J. Karoly, and A. D. King. 2015. The contribution of anthropogenic forcing to the Adelaide and Melbourne, Australia, heat waves of january 2014. *Bulletin of the American Meteorological Society* **96**(12), S145–S148.

Boberg, F., and J. H. Christensen. 2012. Overestimation of Mediterranean summer temperature projections due to model deficiencies. *Nature Climate Change* **2**(6), 433–436.

Boé, J., and L. Terray. 2008. Uncertainties in summer evapotranspiration changes over Europe and implications for regional climate change. *Geophysical Research Letters* **35**(5), L05702.

Bonan, G. 2015. *Ecological climatology: concepts and applications*. Cambridge University Press.

Bond-Lamberty, B., and A. Thomson. 2010. Temperature-associated increases in the global soil respiration record. *Nature* **464**(7288), 579–582.

Bondeau, A., P. C. Smith, S. Zaehle, S. Schaphoff, W. Lucht, W. Cramer, D. Gerten, H. Lotze-Campen, C. Müller, M. Reichstein *et al.* 2007. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology* **13**(3), 679–706.

Bosilovich, M. G. 2008. NASA's modern era retrospective-analysis for research and applications: Integrating Earth observations. *IEEE Earthzine* **1**(1-4), 82367.

Budyko, M. 1974. *Climate and Life*. Academic Press, New York.

Buermann, W., P. R. Bikash, M. Jung, D. H. Burn, and M. Reichstein. 2013. Earlier springs decrease peak summer productivity in North American boreal forests. *Environmental Research Letters* **8**(2), 024027.

Büntgen, U., R. Brázdil, K.-U. Heussner, J. Hofmann, R. Kontic, T. Kyncl, C. Pfister, K. Chromá, and W. Tegel. 2011a. Combined dendro-documentary evidence of Central European hydroclimatic springtime extremes over the last millennium. *Quaternary Science Reviews* **30**(27), 3947–3959.

Büntgen, U., W. Tegel, K. Nicolussi, M. McCormick, D. Frank, V. Trouet, J. O. Kaplan, F. Herzig, K.-U. Heussner, H. Wanner *et al.* 2011b. 2500 years of European climate variability and human susceptibility. *Science* **331**(6017), 578–582.

Burke, E. J. 2011. Understanding the sensitivity of different drought metrics to the drivers of drought under increased atmospheric co2. *Journal of Hydrometeorology* **12**(6), 1378–1394.

Burke, E. J., and S. J. Brown. 2008. Evaluating uncertainties in the projection of future drought. *Journal of Hydrometeorology* **9**(2), 292–299.

Burke, M., S. M. Hsiang, and E. Miguel. 2015. Global non-linear effect of temperature on economic production. *Nature* **527**(7577), 235–239.

Buser, C. M., H. Künsch, D. Lüthi, M. Wild, and C. Schär. 2009. Bayesian multi-model projection of climate: bias assumptions and interannual variability. *Climate Dynamics* **33**(6), 849–868.

Cade, B. S., and B. R. Noon. 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* **1**(8), 412–420.

Cannon, A. J. 2016. Multivariate bias correction of climate model outputs: Matching marginal distributions and inter-variable dependence structure. *Journal of Climate* **29**(19), 7045–7064.

Cao, L., G. Bala, K. Caldeira, R. Nemani, and G. Ban-Weiss. 2010. Importance of carbon dioxide physiological forcing to future climate change. *Proceedings of the National Academy of Sciences* **107**(21), 9513–9518.

Carvalhais, N., M. Forkel, M. Khomik, J. Bellarby, M. Jung, M. Migliavacca, M. u, S. Saatchi, M. Santoro, M. Thurner, U. Weber, B. Ahrens, C. Beer, A. Cescatti, J. T. Randerson, and M. Reichstein. 2014. Global covariation of carbon turnover times with climate in terrestrial ecosystems. *Nature* **514**(7521), 213–217.

Cassou, C., and J. Cattiaux. 2016. Disruption of the European climate seasonal clock in a warming world. *Nature Climate Change* **6**, 589–594.

Cattiaux, J., R. Vautard, C. Cassou, P. Yiou, V. Masson-Delmotte, and F. Codron. 2010. Winter 2010 in Europe: a cold extreme in a warming climate. *Geophysical Research Letters* **37**(20), L20704.

Chambers, J. Q., J. I. Fisher, H. Zeng, E. L. Chapman, D. B. Baker, and G. C. Hurtt. 2007. Hurricane katrina's carbon footprint on us gulf coast forests. *Science* **318**(5853), 1107–1107.

Chapin III, F. S., G. M. Woodwell, J. T. Randerson, E. B. Rastetter, G. M. Lovett, D. D. Baldocchi, D. A. Clark, M. E. Harmon, D. S. Schimel, R. Valentini *et al.* 2006. Reconciling carbon-cycle concepts, terminology, and methods. *Ecosystems* **9**(7), 1041–1050.

Christensen, J., E. Kjellström, F. Giorgi, G. Lenderink, and M. Rummukainen. 2010. Weight assignment in regional climate models. *Climate Research* **44**(2), 179–194.

Christensen, J., F. Boberg, O. Christensen, and P. Lucas-Picher. 2008. On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters* **35**(20), L20709.

Christensen, J. H., and F. Boberg. 2012. Temperature dependent climate projection deficiencies in CMIP5 models. *Geophysical Research Letters* **39**(24), L24705.

Christiansen, B. 2015. The role of the selection problem and non-gaussianity in attribution of single events to climate change. *Journal of Climate* **28**(24), 9873–9891.

Christidis, N., P. A. Stott, A. A. Scaife, A. Arribas, G. S. Jones, D. Copsey, J. R. Knight, and W. J. Tennant. 2013. A new HadGEM3-A-based system for attribution of weather-and climate-related extreme events. *Journal of Climate* **26**(9), 2756–2783.

Ciais, P., C. Sabine, G. Bala, L. Bopp, V. Brovkin, J. Canadell, A. Chhabra, R. DeFries, J. Galloway, M. Heimann *et al.* 2014. Carbon and other biogeochemical cycles. Pp. 465–570. *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Ciais, P., M. Reichstein, N. Viovy, A. Granier, J. Ogée, V. Allard, M. Aubinet, N. Buchmann, C. Bernhofer, A. Carrara *et al.* 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* **437**(7058), 529–533.

Coles, S., J. Bawa, L. Trenner, and P. Dorazio. 2001. *An introduction to statistical modeling of extreme values*. Vol. 208. Springer.

Collins, M. 2007. Ensembles and probabilities: a new era in the prediction of climate change. *Philosophical Transactions of the Royal Society, A* **365**(1857), 1957–1970.

Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason, R. S. Vose, G. Rutledge, P. Bessemoulin *et al.* 2011. The twentieth century reanalysis project. *Quaterly Journal of the Royal Meteorological Society* **137**(654), 1–28.

Cook, B., T. Ault, and J. Smerdon. 2015. Unprecedented 21st century drought risk in the American Southwest and Central Plains. *Science Advances* **1**(1), e1400082.

Cook, E. R., C. A. Woodhouse, C. M. Eakin, D. M. Meko, and D. W. Stahle. 2004. Long-term aridity changes in the western United States. *Science* **306**(5698), 1015–1018.

Coumou, D., and A. Robinson. 2013. Historic and future increase in the global land area affected by monthly heat extremes. *Environmental Research Letters* **8**(3), 034018.

Coumou, D., and S. Rahmstorf. 2012. A decade of weather extremes. *Nature Climate Change* **2**(7), 491–496.

Cox, P. M., D. Pearson, B. B. Booth, P. Friedlingstein, C. Huntingford, C. D. Jones, and C. M. Luke. 2013. Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature* **494**(7437), 341–344.

Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell. 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**(6809), 184–187.

Crowther, T., K. Todd-Brown, C. Rowe, W. Wieder, J. Carey, M. Machmuller, B. Snoek, S. Fang, G. Zhou, S. Allison *et al.* 2016. Quantifying global soil carbon losses in response to warming. *Nature* **540**(7631), 104–108.

Curry, C., J. Sillmann, D. Bronaugh, K. Alterskjaer, J. Cole, D. Ji, B. Kravitz, J. Kristjánsson, J. Moore, H. Muri *et al.* 2014. A multimodel examination of climate extremes in an idealized geoengineering experiment. *Journal of Geophysical Research: Atmospheres* **119**(7), 3900–3923.

Dai, A. 2011. Drought under global warming: a review. *Wiley Interdisciplinary Reviews: Climate Change* **2**(1), 45–65.

Davin, E. L., E. Maisonnave, and S. I. Seneviratne. 2016. Is land surface processes representation a possible weak link in current regional climate models?. *Environmental Research Letters* **11**(7), 074027.

Davin, E. L., S. I. Seneviratne, P. Ciais, A. Olioso, and T. Wang. 2014. Preferential cooling of hot extremes from cropland albedo management. *Proceedings of the National Academy of Sciences* **111**(27), 9757–9761.

Dee, D., S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, P. Bauer *et al.* 2011. The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137**(656), 553–597.

Deutscher Wetterdienst 2014. Seasonal bulletin on the climate in WMO Region VI - Europe and Middle East. *Technical report*. Deutscher Wetterdienst.

Dirmeyer, P. A. 2011. The terrestrial segment of soil moisture–climate coupling. *Geophysical Research Letters* **38**(16), L16702.

Dirmeyer, P. A., Y. Jin, B. Singh, and X. Yan. 2013. Trends in land–atmosphere interactions from CMIP5 simulations. *Journal of Hydrometeorology* **14**(3), 829–849.

Dittus, A. J., D. J. Karoly, S. C. Lewis, and L. V. Alexander. 2015. A multi-region assessment of observed changes in the areal extent of temperature and precipitation extremes. *Journal of Climate* **28**(23), 9206–9220.

Dole, R., M. Hoerling, J. Perlwitz, J. Eischeid, P. Pegion, T. Zhang, X.-W. Quan, T. Xu, and D. Murray. 2011. Was there a basis for anticipating the 2010 Russian heat wave?. *Geophysical Research Letters* **38**(6), L06702.

Donat, M. G., A. L. Lowry, L. V. Alexander, P. A. O'Gorman, and N. Maher. 2016. More extreme precipitation in the world's dry and wet regions. *Nat. Clim. Change* **6**, 508–513.

Donat, M. G., A. L. Lowry, L. V. Alexander, P. A. O'Gorman, and N. Maher. 2017. Addendum: More extreme precipitation in the world's dry and wet regions. *Nature Climate Change* **7**(2), 154–158.

Donat, M. G., L. V. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar. 2013a. Global land-based datasets for monitoring climatic extremes. *Bulletin of the American Meteorological Society* **94**(7), 997–1006.

Donat, M., L. Alexander, H. Yang, I. Durre, R. Vose, R. Dunn, K. Willett, E. Aguilar, M. Brunet, J. Caesar *et al.* 2013b. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset. *Journal of Geophysical Research: Atmospheres* **118**(5), 2098–2118.

Donges, J., C.-F. Schleussner, J. Siegmund, and R. Donner. 2016. Event coincidence analysis for quantifying statistical interrelationships between event time series. *The European Physical Journal Special Topics* **225**(3), 471–487.

Donlon, C. J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer. 2012. The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment* **116**, 140–158.

Drake, B. G., M. A. Gonzàlez-Meler, and S. P. Long. 1997. More efficient plants: a consequence of rising atmospheric co2?. *Annual Review of Plant Biology* **48**(1), 609–639.

Easterling, D. R., G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns. 2000. Climate extremes: observations, modeling, and impacts. *Science* **289**(5487), 2068–2074.

Ehret, U., E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert. 2012. Should we apply bias correction to global and regional climate model data?. *Hydrology and Earth System Sciences* **16**(9), 3391–3404.

Emori, S., and S. Brown. 2005. Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophysical Research Letters* **32**(17), L17706.

Fader, M., S. Rost, C. Müller, A. Bondeau, and D. Gerten. 2010. Virtual water content of temperate cereals and maize: Present and potential future patterns. *Journal of Hydrology* **384**(3), 218–231.

Falkowski, P., R. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, K. Hibbard, P. Högberg, S. Linder *et al.* 2000. The global carbon cycle: a test of our knowledge of earth as a system. *Science* **290**(5490), 291–296.

Fasullo, J. T., and K. E. Trenberth. 2008. The annual cycle of the energy budget. part ii: Meridional structures and poleward transports. *Journal of Climate* **21**(10), 2313–2325.

Fatichi, S., C. Pappas, and V. Y. Ivanov. 2015. Modeling plant–water interactions: an ecohydrological overview from the cell to the global scale. *Wiley Interdisciplinary Reviews: Water* **3**, 327–368.

Fay, M. P., and P. A. Shaw. 2010. Exact and asymptotic weighted logrank tests for interval censored data: the interval R package. *Journal of Statistical Software*.

Fedorov, A. V., and S. G. Philander. 2000. Is El Niño changing?. *Science* **288**(5473), 1997–2002.

Feng, S., and Q. Fu. 2013. Expansion of global drylands under a warming climate. *Atmospheric Chemistry and Physics* **13**(10), 081–10.

Field, C. B., M. J. Behrenfeld, J. T. Randerson, and P. Falkowski. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**(5374), 237–240.

Fischer, E., and R. Knutti. 2013. Robust projections of combined humidity and temperature extremes. *Nature Climate Change* **3**(2), 126–130.

Fischer, E., J. Rajczak, and C. Schär. 2012. Changes in European summer temperature variability revisited. *Geophysical Research Letters* **39**(19), L17706.

Fischer, E. M., and C. Schär. 2009. Future changes in daily summer temperature variability: driving processes and role for temperature extremes. *Climate Dynamics* **33**(7-8), 917–935.

Fischer, E. M., and R. Knutti. 2015. Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nature Climate Change* **5**(6), 560–564.

Fischer, E., S. Seneviratne, D. Lüthi, and C. Schär. 2007. Contribution of land-atmosphere coupling to recent european summer heat waves. *Geophysical Research Letters* **34**(6), L06707.

Fisher, R. A. 1925. Applications of Student's distribution. *Metron* **5**(3), 90–104.

Fisher, R. A., and L. H. C. Tippett. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 24. Cambridge Univ Press. Pp. 180–190.

Flach, M., F. Gans, A. Brenning, J. Denzler, M. Reichstein, E. Rodner, S. Bathiany, P. Bodesheim, Y. Guanche, S. Sippel *et al.* 2016. Multivariate anomaly detection for Earth observations: A comparison of algorithms and feature extraction techniques. *Earth System Dynamics Discussions*. Under Review.

Foken, T. 2008a. The energy balance closure problem: an overview. *Ecological Applications* **18**(6), 1351–1367.

Foken, T. 2008b. *Micrometeorology*. Springer Science & Business Media.

Forkel, M., M. Migliavacca, K. Thonicke, M. Reichstein, S. Schaphoff, U. Weber, and N. Carvalhais. 2015. Co-dominant water control on global inter-annual variability and trends in land surface phenology and greenness. *Global Change Biology* **21**, 3414–3435.

Forkel, M., N. Carvalhais, S. Schaphoff, M. Migliavacca, M. Thurner, K. Thonicke *et al.* 2014. Identifying environmental controls on vegetation greenness phenology through model–data integration. *Biogeosciences* **11**(23), 7025–7050.

Frank, D., M. Reichstein, M. Bahn, K. Thonicke, D. Frank, M. D. Mahecha, P. Smith, M. Velde, S. Vicca, F. Babst *et al.* 2015. Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts. *Global Change Biology* **21**, 2861–2880.

Friedl, M. A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang. 2010. MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* **114**(1), 168–182.

Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti. 2014. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate* **27**(2), 511–526.

Friedman, J., T. Hastie, and R. Tibshirani. 2001. *The elements of statistical learning*. Vol. 1. Springer Series in Statistics. Springer, Berlin.

Frieler, K., R. Betts, E. Burke, P. Ciais, S. Denvil, D. Deryng, K. Ebi, T. Eddy, K. Emanuel, J. Elliott *et al.* 2016. Assessing the impacts of 1.5 c global warming–simulation protocol of the inter-sectoral impact model intercomparison project (isimip2b). *Geoscientific Model Development Discussions*. Under Review.

Fritsch, F. N., and R. E. Carlson. 1980. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis* **17**(2), 238–246.

Fu, Y. H., H. Zhao, S. Piao, M. Peaucelle, S. Peng, G. Zhou, P. Ciais, M. Huang, A. Menzel, J. Peñuelas *et al.* 2015. Declining global warming effects on the phenology of spring leaf unfolding. *Nature* **526**(7571), 104–107.

Galton, F. 1886. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland* **15**, 246–263.

Ganopolski, A., C. Kubatzki, M. Claussen, V. Brovkin, and V. Petoukhov. 1998. The influence of vegetation-atmosphere-ocean interaction on climate during the mid-holocene. *Science* **280**(5371), 1916–1919.

García-Herrera, R., J. Díaz, R. Trigo, J. Luterbacher, and E. Fischer. 2010. A review of the European summer heat wave of 2003. *Critical Reviews in Environmental Science and Technology* **40**(4), 267–306.

Gasparrini, A., Y. Guo, M. Hashizume, E. Lavigne, A. Zanobetti, J. Schwartz, A. Tobias, S. Tong, J. Rocklöv, B. Forsberg *et al.* 2015. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet* **386**(9991), 369–375.

Gerten, D., S. Schaphoff, U. Haberlandt, W. Lucht, and S. Sitch. 2004. Terrestrial vegetation and water balance-hydrological evaluation of a dynamic global vegetation model. *Journal of Hydrology* **286**(1), 249–270.

Gettelman, A., and R. B. Rood. 2016. Simulating the atmosphere. Pp. 61–85. *Demystifying Climate Models*. Springer.

Ghil, M., P. Yiou, S. Hallegatte, B. Malamud, P. Naveau, A. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, V. Kossobokov *et al.* 2011. Extreme events: dynamics, statistics and prediction. *Nonlinear Processes in Geophysics* **18**(3), 295–350.

Gilleland, E., and R. W. Katz. 2011. New software to analyze how extremes change over time. *Eos, Transactions American Geophysical Union* **92**(2), 13–14.

Giorgi, F., C. Jones, G. R. Asrar *et al.* 2009. Addressing climate information needs at the regional level: the cordex framework. *World Meteorological Organization (WMO) Bulletin* **58**(3), 175.

Gneiting, T., and A. E. Raftery. 2005. Weather forecasting with ensemble methods. *Science* **310**(5746), 248–249.

Gneiting, T., and M. Katzfuss. 2014. Probabilistic forecasting. *Annual Review of Statistics and Its Application* **1**, 125–151.

Gobron, N., A. Belward, B. Pinty, and W. Knorr. 2010. Monitoring biosphere vegetation 1998–2009. *Geophysical Research Letters* **37**(15), L15402.

Golyandina, N., and A. Zhigljavsky. 2013. *Singular Spectrum Analysis for time series*. Springer.

Graham, R. A., and R. H. Grumm. 2010. Utilizing normalized anomalies to assess synoptic-scale weather events in the western united states. *Weather and Forecasting* **25**(2), 428–445.

Graven, H., R. Keeling, S. Piper, P. Patra, B. Stephens, S. Wofsy, L. Welp, C. Sweeney, P. Tans, J. Kelley *et al.* 2013. Enhanced seasonal exchange of CO_2 by Northern ecosystems since 1960. *Science* **341**(6150), 1085–1089.

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), 716–721.

Grimm, V., and S. F. Railsback. 2012. Pattern-oriented modelling: a 'multi-scope' for predictive systems ecology. *Philosophical Transactions of the Royal Society B* **367**(1586), 298–310.

Grotjahn, R., R. Black, R. Leung, M. F. Wehner, M. Barlow, M. Bosilovich, A. Gershunov, W. J. Gutowski Jr, J. R. Gyakum, R. W. Katz *et al.* 2016. North American extreme temperature events and related large scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics* **46**(3-4), 1151–1184.

Gruber, N., and J. N. Galloway. 2008. An Earth-system perspective of the global nitrogen cycle. *Nature* **451**(7176), 293–296.

Grumm, R. H., and R. Hart. 2001. Standardized anomalies applied to significant cold season weather events: Preliminary findings. *Weather and Forecasting* **16**(6), 736–754.

Gudmundsson, L., J. Bremnes, J. Haugen, and T. E. Skaugen. 2012. Technical note: Downscaling RCM precipitation to the station scale using quantile mapping–a comparison of methods. *Hydrology and Earth System Sciences* **9**(5), 6185–6201.

Guha-Sapir, D., F. Vos, R. Below, and S. Ponserre. 2011. Annual disaster statistical review 2010. *Technical report*. Centre for Research on the Epidemiology of Disasters.

Haerter, J., S. Hagemann, C. Moseley, and C. Piani. 2011. Climate model bias correction and the role of timescales. *Hydrology and Earth System Sciences* **15**(3), 1065–1079.

Hall, A., and X. Qu. 2006. Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys Res Lett* **33**(3), 1–4.

Hall, D., and K. Rao. 1999. *Photosynthesis*. 6th edition. ed. Cambridge University Press.

Hansen, G., and D. Stone. 2016. Assessing the observed impact of anthropogenic climate change. *Nature Climate Change* **6**(5), 532–537.

Hansen, G., D. Stone, M. Auffhammer, C. Huggel, and W. Cramer. 2016. Linking local impacts to changes in climate: a guide to attribution. *Regional Environmental Change* **16**(2), 527–541.

Hansen, J., M. Sato, and R. Ruedy. 2012. Perception of climate change. *Proc. Natl Acad. Sci. USA* **109**(37), E2415–E2423.

Harris, I., P. Jones, T. Osborn, and D. Lister. 2014. Updated high-resolution grids of monthly climatic observations–the CRU TS3.10 Dataset. *International Journal of Climatology* **34**(3), 623–642.

Hart, R. E., and R. H. Grumm. 2001. Using normalized climatological anomalies to rank synoptic-scale events objectively. *Monthly Weather Review* **129**(9), 2426–2442.

Hauser, M., R. Orth, and S. I. Seneviratne. 2016. Role of soil moisture vs. recent climate change for the 2010 heat wave in Western Russia. *Geophysical Research Letters*.

Haxeltine, A., and I. C. Prentice. 1996. BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. *Global Biogeochemical Cycles* **10**(4), 693–709.

Haylock, M., N. Hofstra, A. Klein Tank, E. Klok, P. Jones, and M. New. 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres* **113**(D20), D20119.

Hays, J. D., J. Imbrie, N. J. Shackleton *et al.* 1976. Variations in the Earth's orbit: pacemaker of the ice ages. *Science* **194**(4270), 1121–1132.

Heimann, M., and M. Reichstein. 2008. Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature* **451**(7176), 289–292.

Held, I. M., and B. J. Soden. 2006. Robust responses of the hydrological cycle to global warming. *Journal of Climate* **19**(21), 5686–5699.

Hempel, S., K. Frieler, L. Warszawski, J. Schewe, and F. Piontek. 2013. A trend-preserving bias correction–the ISI-MIP approach. *Earth System Dynamics* **4**(2), 219–236.

Herring, S. C., M. P. Hoerling, T. C. Peterson, and P. A. Stott. 2014. *Explaining Extreme Events of 2013 from a Climate Perspective*. Vol. 95. American Meteorological Society.

Hijmans, R. J. 2015. *raster: Geographic Data Analysis and Modeling*. R package version 2.3-33.

Hirschi, M., S. I. Seneviratne, V. Alexandrov, F. Boberg, C. Boroneant, O. B. Christensen, H. Formayer, B. Orlowsky, and P. Stepanek. 2011. Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience* **4**(1), 17–21.

Hoerling, M., J. Eischeid, A. Kumar, R. Leung, A. Mariotti, K. Mo, S. Schubert, and R. Seager. 2014. Causes and predictability of the 2012 Great Plains drought. *Bulletin of the American Meteorological Society* **95**(2), 269–282.

Hoffmann, H., and T. Rath. 2012. Meteorologically consistent bias correction of climate time series for agricultural models. *Theoretical and Applied Climatology* **110**(1-2), 129–141.

Högberg, P., A. Nordgren, N. Buchmann, A. F. Taylor, A. Ekblad, M. N. Högberg, G. Nyberg, M. Ottosson-Löfvenius, and D. J. Read. 2001. Large-scale forest girdling shows that current photosynthesis drives soil respiration. *Nature* **411**(6839), 789–792.

Horton, J. L., and S. C. Hart. 1998. Hydraulic lift: a potentially important ecosystem process. *Trends in Ecology & Evolution* **13**(6), 232–235.

Huang, J., H. Yu, X. Guan, G. Wang, and R. Guo. 2015. Accelerated dryland expansion under climate change. *Nature Climate Change* **6**, 166–171.

Huang, Y., S. Gerber, T. Huang, and J. W. Lichstein. 2016. Evaluating the drought response of CMIP5 models using global gross primary productivity, leaf area, precipitation, and soil moisture data. *Global Biogeochemical Cycles* **30**(12), 1827–1846.

Hulme, M. 1996. Recent climatic change in the world's drylands. *Geophysical Research Letters* **23**(1), 61–64.

Huntingford, C., P. D. Jones, V. N. Livina, T. M. Lenton, and P. M. Cox. 2013. No increase in global temperature variability despite changing regional patterns. *Nature* **500**(7462), 327–331.

Huntingford, C., T. Marsh, A. A. Scaife, E. J. Kendon, J. Hannaford, A. L. Kay, M. Lockwood, C. Prudhomme, N. S. Reynard, S. Parry *et al.* 2014. Potential influences on the United Kingdom's floods of winter 2013/14. *Nature Climate Change* **4**(9), 769–777.

Huybers, P., and W. Curry. 2006. Links between annual, Milankovitch and continuum temperature variability. *Nature* **441**(7091), 329–332.

Ilie, I., P. Dittrich, N. Carvalhais, M. Jung, A. Heinemeyer, M. Migliavacca, J. Morison, S. Sippel, J.-A. Subke, M. Wilkinson, and M. D. Mahecha. 2016. Reverse engineering model structures for soil and ecosystem respiration: the potential of gene expression programming. *Geoscientific Model Development Discussions*. Under Review.

Ingram, W. 2016. Extreme precipitation: Increases all round. *Nature Climate Change* **6**(5), 443–444.

IPCC 2012. Summary for policymakers. In *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the Intergovernmental Panel on Climate Change*. C. Field, V. Barros, T. Stocker, Q. Dahe, D. Dokken, K. Ebi, M. Mastrandrea, K. Mach, G. Plattner, S. Allen, M. Tignor, and P. Midgley. (Eds.). Cambridge University Press.

IPCC 2014. Summary for policymakers. Pp. 1–32.. In *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. C. Field, V. Barros, D. Dokken, K. Mach, M. Mastrandrea, T. Bilir, K. Chatterjee, K. Ebi, Y. Estrada, R. Genova, B. Girma, E. Kissel, A. Levy, S. MacCracken, M. Mastrandrea, and L. White. (Eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jentsch, A., J. Kreyling, and C. Beierkuhnlein. 2007. A new generation of climate-change experiments: events, not trends. *Frontiers in Ecology and the Environment* **5**(7), 365–374.

Jentsch, A., J. Kreyling, J. Boettcher-Treschkow, and C. Beierkuhnlein. 2009. Beyond gradual warming: extreme weather events alter flower phenology of European grassland and heath species. *Global Change Biology* **15**(4), 837–849.

Ji, F., Z. Wu, J. Huang, and E. P. Chassignet. 2014. Evolution of land surface air temperature trend. *Nature Climate Change* **4**(6), 462–466.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1994. *Continuous univariate distributions. Volume 1*. John Wiley & Sons, New York.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1995. *Continuous univariate distributions. Volume 2*. John Wiley & Sons, New York.

Jones, P. W. 1999. First-and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review* **127**(9), 2204–2210.

Jones, R. 2004. Generating high resolution climate change scenarios using PRECIS. *Technical report*. Met Office.

Jung, M., M. Reichstein, C. R. Schwalm, C. Huntingford, S. Sitch, A. Ahlström, A. Arneth, G. Camps-Valls, P. Ciais, P. Friedlingstein *et al.* 2017. Compensatory water effects link yearly global land co2 sink changes to temperature. *Nature* **541**(7638), 516–520.

Jung, M., M. Reichstein, H. A. Margolis, A. Cescatti, A. D. Richardson, M. A. Arain, A. Arneth, C. Bernhofer, D. Bonal, J. Chen *et al.* 2011. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research: Biogeosciences (2005–2012)* **116**(G3), G00J07.

Kala, J., M. G. De Kauwe, A. J. Pitman, B. E. Medlyn, Y.-P. Wang, R. Lorenz, and S. E. Perkins-Kirkpatrick. 2016. Impact of the representation of stomatal conductance on model projections of heatwave intensity. *Scientific Reports* **6**(23418), 1–7.

Kamae, Y., H. Shiogama, M. Watanabe, and M. Kimoto. 2014. Attributing the increase in Northern Hemisphere hot summers since the late 20th century. *Geophysical Research Letters* **41**(14), 5192–5199.

Katz, R. W. 2010. Statistics of extremes in climate change. *Climatic Change* **100**(1), 71–76.

Katz, R. W., and B. G. Brown. 1992. Extreme events in a changing climate: variability is more important than averages. *Climatic Change* **21**(3), 289–302.

Katz, R. W., P. F. Craigmile, P. Guttorp, M. Haran, B. Sanso, and M. L. Stein. 2013. Uncertainty analysis in climate change assessments. *Nature Climate Change* **3**(9), 769–771.

Keenan, T. F., J. Gray, M. A. Friedl, M. Toomey, G. Bohrer, D. Y. Hollinger, J. W. Munger, J. O'Keefe, H. P. Schmid, I. S. Wing *et al.* 2014. Net carbon uptake has increased through warming-induced changes in temperate forest phenology. *Nature Climate Change* **4**(7), 598–604.

Kharin, V. V., and F. W. Zwiers. 2000. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean gcm. *Journal of Climate* **13**(21), 3760–3788.

Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl. 2007. Changes in temperature and precipitation extremes in the ipcc ensemble of global coupled model simulations. *Journal of Climate* **20**(8), 1419–1444.

Kharin, V. V., F. Zwiers, X. Zhang, and M. Wehner. 2013. Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change* **119**(2), 345–357.

King, A. D., M. T. Black, S.-K. Min, E. M. Fischer, D. M. Mitchell, L. J. Harrington, and S. E. Perkins-Kirkpatrick. 2016. Emergence of heat extremes attributable to anthropogenic influences. *Geophysical Research Letters* **43**(7), 3438–3443.

Klein Tank, A., J. Wijngaard, G. Können, R. Böhm, G. Demarée, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen *et al.* 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. *International Journal of Climatology* **22**(12), 1441–1453.

Knapp, A. K., P. A. Fay, J. M. Blair, S. L. Collins, M. D. Smith, J. D. Carlisle, C. W. Harper, B. T. Danner, M. S. Lett, and J. K. McCarron. 2002. Rainfall variability, carbon cycling, and plant species diversity in a mesic grassland. *Science* **298**(5601), 2202–2205.

Knutson, T. R., F. Zeng, and A. T. Wittenberg. 2013. The extreme March-May 2012 warm anomaly over the Eastern United States: Global context and multimodel trend analysis. *Bulletin of the American Meteorological Society* **94**(9), S13.

Knutti, R. 2008. Should we believe model predictions of future climate change?. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **366**(1885), 4647–4664.

Knutti, R. 2010. The end of model democracy?. *Climatic Change* **102**(3-4), 395–404.

Knutti, R., and J. Sedláček. 2013. Robustness and uncertainties in the new cmip5 climate model projections. *Nature Climate Change* **3**(4), 369–373.

Knutti, R., D. Masson, and A. Gettelman. 2013. Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters* **40**(6), 1194–1199.

Kodra, E., and A. R. Ganguly. 2014. Asymmetry of projected increases in extreme temperature distributions. *Scientific Reports* **4**, 5884.

Kodra, E., K. Steinhaeuser, and A. R. Ganguly. 2011. Persisting cold extremes under 21st-century warming scenarios. *Geophysical Research Letters* **38**(8), L08705.

Koenker, R. 2015. *quantreg: Quantile Regression*. R package version 5.11.

Köppen, W. 1900. Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt. *Geographische Zeitschrift* **6**(11. H), 593–611.

Körner, C. 2003. Slow in, rapid out–carbon flux studies and Kyoto targets. *Science* **300**(5623), 1242–1243.

Körner, C. 2017. A matter of tree longevity. *Science* **355**(6321), 130–131.

Körner, C., and D. Basler. 2010. Phenology under global warming. *Science* **327**(5972), 1461–1462.

Koster, R. D., P. A. Dirmeyer, Z. Guo, G. Bonan, E. Chan, P. Cox, C. Gordon, S. Kanae, E. Kowalczyk, D. Lawrence *et al.* 2004. Regions of strong coupling between soil moisture and precipitation. *Science* **305**(5687), 1138–1140.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel. 2006. World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* **15**(3), 259–263.

Krinner, G., N. Viovy, N. de Noblet-Ducoudré, J. Ogée, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice. 2005. A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles* **19**(1), 1–33.

Kuglitsch, F. G., A. Toreti, E. Xoplaki, P. M. Della-Marta, C. S. Zerefos, M. Türkeş, and J. Luterbacher. 2010. Heat wave changes in the eastern Mediterranean since 1960. *Geophysical Research Letters* **37**(4), L04802.

Lacis, A. A., G. A. Schmidt, D. Rind, and R. A. Ruedy. 2010. Atmospheric co2: Principal control knob governing EarthÕs temperature. *Science* **330**(6002), 356–359.

Le Quéré, C., M. R. Raupach, J. G. Canadell, G. Marland, L. Bopp, P. Ciais, T. J. Conway, S. C. Doney, R. A. Feely, P. Foster *et al.* 2009. Trends in the sources and sinks of carbon dioxide. *Nature Geoscience* **2**(12), 831–836.

Le Tertre, A., A. Lefranc, D. Eilstein, C. Declercq, S. Medina, M. Blanchard, B. Chardon, P. Fabre, L. Filleul, J.-F. Jusot *et al.* 2006. Impact of the 2003 heatwave on all-cause mortality in 9 French cities. *Epidemiology* **17**(1), 75–79.

Leakey, A. D., E. A. Ainsworth, C. J. Bernacchi, A. Rogers, S. P. Long, and D. R. Ort. 2009. Elevated CO_2 effects on plant carbon, nitrogen, and water relations: six important lessons from FACE. *Journal of Experimental Botany* **60**(10), 2859–2876.

Lehmann, J., D. Coumou, and K. Frieler. 2015. Increased record-breaking precipitation events under global warming. *Climatic Change* **132**(4), 501–515.

Lenderink, G., and E. Van Meijgaard. 2008. Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geoscience* **1**(8), 511–514.

Leonard, M., S. Westra, A. Phatak, M. Lambert, B. van den Hurk, K. McInnes, J. Risbey, S. Schuster, D. Jakob, and M. Stafford-Smith. 2014. A compound event framework for understanding extreme impacts. *Wiley Interdisciplinary Reviews: Climate Change* **5**(1), 113–128.

Lewis, S. L., P. M. Brando, O. L. Phillips, G. M. van der Heijden, and D. Nepstad. 2011. The 2010 Amazon drought. *Science* **331**(6017), 554–554.

Li, C., E. Sinha, D. E. Horton, N. S. Diffenbaugh, and A. M. Michalak. 2014. Joint bias correction of temperature and precipitation in climate model simulations. *Journal of Geophysical Research: Atmospheres* **119**(23), 13–153.

Liu, C., and R. P. Allan. 2013. Observed and simulated precipitation responses in wet and dry regions 1850–2100. *Environmental Research Letters* **8**(3), 034002.

Lloyd, J., and J. Taylor. 1994. On the temperature dependence of soil respiration. *Functional Ecology* **8**(3), 315–323.

Lorenz, E. N. 1963. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* **20**(2), 130–141.

Lorenz, R., E. Davin, and S. Seneviratne. 2012. Modeling land-climate coupling in Europe: Impact of land surface representation on climate variability and extremes. *Journal of Geophysical Research: Atmospheres* **117**(D20), D20109.

Lovelock, J. E., and A. J. Watson. 1982. The regulation of carbon dioxide and climate: Gaia or geochemistry. *Planetary and Space Science* **30**(8), 795–802.

Lucas, H. S. 1930. The great European famine of 1315, 1316, and 1317. *Speculum* **5**(4), 343–377.

Luo, Y. 2007. Terrestrial carbon–cycle feedback to climate warming. *Annu. Rev. Ecol. Evol. Syst.* **38**, 683–712.

Luo, Y., J. T. Randerson, G. Abramowitz, C. Bacour, E. Blyth, N. Carvalhais, P. Ciais, D. Dalmonech, J. B. Fisher, R. Fisher *et al.* 2012. A framework for benchmarking land models. *Biogeosciences* **9**(10), 3857–3874.

Luterbacher, J., and C. Pfister. 2015. The year without a summer. *Nature Geoscience* **8**(4), 246–248.

Luterbacher, J., D. Dietrich, E. Xoplaki, M. Grosjean, and H. Wanner. 2004. European seasonal and annual temperature variability, trends, and extremes since 1500. *Science* **303**(5663), 1499–1503.

Luterbacher, J., J. Werner, J. E. Smerdon, L. Fernández-Donado, F. González-Rouco, D. Barriopedro, F. C. Ljungqvist, U. Büntgen, E. Zorita, S. Wagner *et al.* 2016. European summer temperatures since roman times. *Environmental Research Letters* **11**(2), 024001.

Ma, X., A. Huete, J. Cleverly, D. Eamus, F. Chevallier, J. Joiner, B. Poulter, Y. Zhang, L. Guanter, W. Meyer *et al.* 2016. Drought rapidly diminishes the large net CO_2 uptake in 2011 over semi-arid Australia. *Scientific Reports*.

Ma, X., A. Huete, S. Moran, G. Ponce-Campos, and D. Eamus. 2015. Abrupt shifts in phenology and vegetation productivity under climate extremes. *Journal of Geophysical Research: Biogeosciences* **120**(10), 2036–2052.

Mahecha, M. D., M. Reichstein, M. Jung, S. I. Seneviratne, S. Zaehle, C. Beer, M. C. Braakhekke, N. Carvalhais, H. Lange, G. Le Maire *et al.* 2010a. Comparing observations and process-based simulations of biosphere-atmosphere exchanges on multiple timescales. *Journal of Geophysical Research: Biogeosciences* **115**(G2), G02003.

Mahecha, M. D., M. Reichstein, N. Carvalhais, G. Lasslop, H. Lange, S. I. Seneviratne, R. Vargas, C. Ammann, M. A. Arain, A. Cescatti *et al.* 2010b. Global convergence in the temperature sensitivity of respiration at ecosystem level. *Science* **329**(5993), 838–840.

Maraun, D. 2012. Nonstationarities of regional climate model biases in european seasonal mean temperature and precipitation sums. *Geophysical Research Letters* **39**(6), L06706.

Maraun, D. 2013. Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate* **26**(6), 2137–2143.

Maraun, D. 2016. Bias correcting climate change simulations-a critical review. *Current Climate Change Reports* **2**(4), 211–220.

Maraun, D., F. Wetterhall, A. Ireson, R. Chandler, E. Kendon, M. Widmann, S. Brienen, H. Rust, T. Sauter, M. Themeßl *et al.* 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics* **48**(3), 1–34.

Martins, E. S., and J. R. Stedinger. 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research* **36**(3), 737–744.

Massey, N., R. Jones, F. Otto, T. Aina, S. Wilson, J. Murphy, D. Hassell, Y. Yamazaki, and M. Allen. 2015. weather@home-development and validation of a very large ensemble modelling system for probabilistic event attribution. *Quarterly Journal of the Royal Meteorological Society* **141**(690), 1528–1545.

Massonnet, F., O. Bellprat, V. Guemas, and F. J. Doblas-Reyes. 2016. Using climate models to estimate the quality of global observational data sets. *Science*.

Mastrandrea, M. D., C. B. Field, T. F. Stocker, O. Edenhofer, K. L. Ebi, D. J. Frame, H. Held, E. Kriegler, K. J. Mach, P. R. Matschoss *et al.* 2010. Guidance note for lead authors of the IPCC Fifth Assessment Report on consistent treatment of uncertainties. *Technical report*. Internation Panel on Climate Change.

McGregor, G., P. Bessemoulin, K. Ebi, and B. Menne. 2010. Heat waves and health: guidance on warning-system development. *Technical report*. World Meteorological Organization, World Health Organization.

Mearns, L. O., R. W. Katz, and S. H. Schneider. 1984. Extreme high-temperature events: changes in their probabilities with changes in mean temperature. *Journal of Climate and Applied Meteorology* **23**(12), 1601–1613.

Medlyn, B. E., R. A. Duursma, D. Eamus, D. S. Ellsworth, I. C. Prentice, C. V. Barton, K. Y. Crous, P. De Angelis, M. Freeman, and L. Wingate. 2011. Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biology* **17**(6), 2134–2144.

Meehl, G. A., and C. Tebaldi. 2004. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science* **305**(5686), 994–997.

Meehl, G. A., T. Karl, D. R. Easterling, S. Changnon, R. Pielke Jr, D. Changnon, J. Evans, P. Y. Groisman, T. R. Knutson, K. E. Kunkel *et al.* 2000. An introduction to trends in extreme weather and climate events: observations, socioeconomic impacts, terrestrial ecological impacts, and model projections. *Bulletin of the American Meteorological Society* **81**(3), 413–416.

Menzel, A., T. H. Sparks, N. Estrella, E. Koch, A. Aasa, R. Ahas, K. Alm-Kübler, P. Bissolli, O. Braslavská, A. Briede *et al.* 2006. European phenological response to climate change matches the warming pattern. *Global Change Biology* **12**(10), 1969–1976.

Middleton, N., and D. Thomas. 1992. *World Atlas of Desertification: United Nations Environmental Programme*. Arnold.

Milborrow, S. 2015. *earth: Multivariate Adaptive Regression Splines*. R package version 4.2.0. Derived from mda:mars by Trevor Hastie and Rob Tibshirani.

Millennium Ecosystem Assessment 2005. Dryland systems. *Ecosystems and human well-being: current state and trends* Pp. 623–662.

Milly, P., and K. Dunne. 2016. Potential evapotranspiration and continental drying. *Nature Climate Change* **6**, 946–949.

Milly, P. C., and K. A. Dunne. 2011. On the hydrologic adjustment of climate-model projections: The potential pitfall of potential evapotranspiration. *Earth Interactions* **15**(1), 1–14.

Min, E., W. Hazeleger, G. Van Oldenborgh, and A. Sterl. 2013. Evaluation of trends in high temperature extremes in north-western Europe in regional climate models. *Environmental Research Letters* **8**(1), 014011.

Miralles, D. G., A. J. Teuling, C. C. van Heerwaarden, and J. V.-G. de Arellano. 2014. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nature Geoscience* **7**(5), 345–349.

Miralles, D., M. v. den Berg, A. Teuling, and R. d. Jeu. 2012. Soil moisture-temperature coupling: A multiscale observational analysis. *Geophysical Research Letters* **39**(21), L21707.

Miralles, D., R. De Jeu, J. Gash, T. Holmes, and A. Dolman. 2011a. Magnitude and variability of land evaporation and its componentsat the global scale. *Hydrology and Earth System Sciences* **15**, 967–981.

Miralles, D., T. Holmes, R. De Jeu, J. Gash, A. Meesters, and A. Dolman. 2011b. Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences* **15**(2), 453–469.

Mitchell, D., C. Heaviside, S. Vardoulakis, C. Huntingford, G. Masato, B. P. Guillod, P. Frumhoff, A. Bowery, D. Wallom, and M. Allen. 2016a. Attributing human mortality during extreme heat waves to anthropogenic climate change. *Environmental Research Letters* **11**(7), 074006.

Mitchell, D., K. AchutaRao, M. Allen, I. Bethke, U. Beyerle, A. Ciavarella, P. M. Forster, J. Fuglestvedt, N. Gillett, K. Haustein *et al.* 2017. Half a degree additional warming, prognosis and projected impacts (HAPPI): background and experimental design. *Geoscientific Model Development* **10**(2), 571.

Mitchell, D., R. James, P. M. Forster, R. A. Betts, H. Shiogama, and M. Allen. 2016b. Realizing the impacts of a 1.5°c warmer world. *Nature Climate Change*.

Mora, C., A. G. Frazier, R. J. Longman, R. S. Dacks, M. M. Walton, E. J. Tong, J. J. Sanchez, L. R. Kaiser, Y. O. Stender, J. M. Anderson *et al.* 2013. The projected timing of climate departure from recent variability. *Nature* **502**(7470), 183–187.

Mueller, B., and S. Seneviratne. 2014. Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophysical Research Letters* **41**(1), 128–134.

Mueller, B., M. Hirschi, and S. I. Seneviratne. 2011a. New diagnostic estimates of variations in terrestrial water storage based on ERA-interim data. *Hydrological Processes* **25**(7), 996–1008.

Mueller, B., M. Hirschi, C. Jimenez, P. Ciais, P. Dirmeyer, A. Dolman, J. Fisher, M. Jung, F. Ludwig, F. Maignan *et al.* 2013. Benchmark products for land evapotranspiration: Landflux-EVAL multi-data set synthesis. *Hydrology and Earth System Sciences* **17**, 3707–3720.

Mueller, B., S. Seneviratne, C. Jimenez, T. Corti, M. Hirschi, G. Balsamo, P. Ciais, P. Dirmeyer, J. Fisher, Z. Guo *et al.* 2011b. Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophysical Research Letters* **38**(6), L06402.

Myhre, G., D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza *et al.* 2013. Anthropogenic and natural radiative forcing. *Technical report*. International Panel on Climate Change.

Myneni, R., S. Hoffman, Y. Knyazikhin, J. Privette, J. Glassy, Y. Tian, Y. Wang, X. Song, Y. Zhang, G. Smith *et al.* 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of the Environment* **83**(1), 214–231.

Mystakidis, S., E. L. Davin, N. Gruber, and S. I. Seneviratne. 2016. Constraining future terrestrial carbon cycle projections using observation-based water and carbon flux estimates. *Global Change Biology* **22**, 2198–2215.

National Academies of Sciences, Engineering, and Medicine 2016. *Attribution of extreme weather events in the context of climate change*. National Academies Press, Washington D.C.

Nicholls, N. 1996. Long-term climate monitoring and extreme events. Pp. 101–115. *Long-Term Climate Monitoring by the Global Climate Observing System*. Springer.

NOAA National Centers for Environmental Information 2016. Climate at a glance: Global time series. *Technical report*. NOAA.

Nogaj, M., P. Yiou, S. Parey, F. Malek, and P. Naveau. 2006. Amplitude and frequency of temperature extremes over the north atlantic region. *Geophysical Research Letters* **33**(10), L10801.

Norby, R. J., and Y. Luo. 2004. Evaluating ecosystem responses to rising atmospheric CO_2 and global warming in a multi-factor world. *New Phytologist* **162**(2), 281–293.

O'Gorman, P. A. 2015. Precipitation extremes under climate change. *Current Climate Change Reports* **1**(2), 49–59.

O'Gorman, P. A., and T. Schneider. 2009. The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proceedings of the National Academy of Sciences* **106**(35), 14773–14777.

Onogi, K., J. Tsutsui, H. Koide, M. Sakamoto, S. Kobayashi, H. Hatsushika, T. Matsumoto, N. Yamazaki, H. Kamahori, K. Takahashi *et al.* 2007. The JRA-25 reanalysis. *Journal of the Meteorological Society of Japan* **85**(3), 369–432.

Orlowsky, B., and S. I. Seneviratne. 2012. Global changes in extreme events: regional and seasonal dimension. *Climatic Change* **110**(3), 669–696.

Orlowsky, B., and S. Seneviratne. 2013. Elusive drought: uncertainty in observed trends and short-and long-term cmip5 projections. *Hydrology and Earth System Sciences* **17**(5), 1765–1781.

Orth, R., J. Zscheischler, and S. I. Seneviratne. 2016. Record dry summer in 2015 challenges precipitation projections in Central Europe. *Scientific Reports* **6**(28334), 1–8.

Otto, F. E. 2016. Extreme events: The art of attribution. *Nature Climate Change* **6**(4), 342–343.

Otto, F. E., G. J. van Oldenborgh, J. Eden, P. A. Stott, D. J. Karoly, and M. R. Allen. 2016. The attribution question. *Nature Climate Change* **6**(9), 813–816.

Otto, F. E., K. Haustein, P. Uhe, C. A. Coelho, J. A. Aravequia, W. Almeida, A. King, E. Coughlan de Perez, Y. Wada, G. Jan van Oldenborgh *et al.* 2015. Factors other than climate change, main drivers of 2014/15 water shortage in southeast Brazil. *Bulletin of the American Meteorological Society* **96**(12), S35–S40.

Otto, F., N. Massey, G. Oldenborgh, R. Jones, and M. Allen. 2012. Reconciling two approaches to attribution of the 2010 russian heat wave. *Geophysical Research Letters* **39**(4), L04702.

Pall, P., M. Allen, and D. Stone. 2007. Testing the Clausius–Clapeyron constraint on changes in extreme precipitation under CO_2 warming. *Climate Dynamics* **28**(4), 351–363.

Pall, P., T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. Hilberts, D. Lohmann, and M. R. Allen. 2011. Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature* **470**(7334), 382–385.

Pan, Y., R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips, A. Shvidenko, S. L. Lewis, J. G. Canadell *et al.* 2011. A large and persistent carbon sink in the world's forests. *Science* **333**(6045), 988–993.

Parker, D. E. 2011. Recent land surface air temperature trends assessed using the 20th century reanalysis. *Journal of Geophysical Research: Atmospheres* **116**(D20), D20125.

Parmesan, C., T. L. Root, and M. R. Willig. 2000. Impacts of extreme weather and climate on terrestrial biota. *Bulletin of the American Meteorological Society* **81**(3), 443–450.

Perkins, S., A. Pitman, N. Holbrook, and J. McAneney. 2007. Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *Journal of Climate* **20**(17), 4356–4376.

Perkins, S., L. Alexander, and J. Nairn. 2012. Increasing frequency, intensity and duration of observed global heatwaves and warm spells. *Geophysical Research Letters* **39**(20), L20714.

Peterson, T. C., X. Zhang, M. Brunet-India, and J. L. Vázquez-Aguirre. 2008. Changes in North American extremes derived from daily weather data. *Journal of Geophysical Research: Atmospheres* **113**(D7), D07113.

Phillips, O. L., L. E. Aragão, S. L. Lewis, J. B. Fisher, J. Lloyd, G. López-González, Y. Malhi, A. Monteagudo, J. Peacock, C. A. Quesada *et al.* 2009. Drought sensitivity of the Amazon rainforest. *Science* **323**(5919), 1344–1347.

Piani, C., and J. Haerter. 2012. Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophysical Research Letters* **39**(20), L20401.

Piani, C., D. Frame, D. Stainforth, and M. Allen. 2005. Constraints on climate change from a multi-thousand member ensemble of simulations. *Geophysical Research Letters* **32**(23), D24108.

Piani, C., G. Weedon, M. Best, S. Gomes, P. Viterbo, S. Hagemann, and J. Haerter. 2010. Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *Journal of Hydrology* **395**(3), 199–215.

Pierce, D. 2014. *ncdf4: Interface to Unidata netCDF (version 4 or earlier) format data files*. R package version 1.12.

Pinty, B., M. Clerici, I. Andredakis, T. Kaminski, M. Taberner, M. Verstraete, N. Gobron, S. Plummer, and J.-L. Widlowski. 2011. Exploiting the MODIS albedos with the Two-stream Inversion Package (JRC-TIP): 2. fractions of transmitted and absorbed fluxes in the vegetation and soil layers. *Journal of Geophysical Research: Atmospheres* **116**(D9), D09106.

Pitman, A. 2003. The evolution of, and revolution in, land surface schemes designed for climate models. *International Journal of Climatology* **23**(5), 479–510.

Pope, V., M. Gallani, P. Rowntree, and R. Stratton. 2000. The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics* **16**(2), 123–146.

Prein, A. F., R. M. Rasmussen, K. Ikeda, C. Liu, M. P. Clark, and G. J. Holland. 2016. The future intensification of hourly precipitation extremes. *Nature Climate Change* **7**, 48–52.

Quesada, B., R. Vautard, P. Yiou, M. Hirschi, and S. I. Seneviratne. 2012. Asymmetric European summer heat predictability from wet and dry southern winters and springs. *Nature Climate Change* **2**(10), 736–741.

Quiroga, R. Q., T. Kreuz, and P. Grassberger. 2002. Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Physical Review E* **66**(4), 041904.

R Development Core Team 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. Version 2.10.0.

R Development Core Team 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rahmstorf, S., and D. Coumou. 2011. Increase of extreme events in a warming world. *Proc. Natl Acad. Sci. USA* **108**(44), 17905–17909.

Rammig, A., M. Wiedermann, J. Donges, F. Babst, W. von Bloh, D. Frank, K. Thonicke, and M. Mahecha. 2015. Coincidences of climate extremes and anomalous vegetation responses: comparing tree ring patterns to simulated productivity. *Biogeosciences* **12**, 373–385.

Ramos, A. M., R. M. Trigo, and M. L. R. Liberato. 2014. A ranking of high-resolution daily precipitation extreme events for the Iberian Peninsula. *Atmospheric Science Letters* **15**(4), 328–334.

Reichle, R. H., R. D. Koster, G. J. De Lannoy, B. A. Forman, Q. Liu, S. P. Mahanama, and A. Touré. 2011. Assessment and enhancement of MERRA land surface hydrology estimates. *Journal of Climate* **24**(24), 6322–6338.

Reichstein, M., E. Falge, D. Baldocchi, D. Papale, M. Aubinet, P. Berbigier, C. Bernhofer, N. Buchmann, T. Gilmanov, A. Granier *et al.* 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology* **11**(9), 1424–1439.

Reichstein, M., M. Bahn, P. Ciais, D. Frank, M. D. Mahecha, S. I. Seneviratne, J. Zscheischler, C. Beer, N. Buchmann, D. C. Frank, D. Papale, A. Rammig, P. Smith, K. Thonicke, M. van der Velde, S. Vicca, A. Walz, and M. Wattenbach. 2013. Climate extremes and the carbon cycle. *Nature* **500**(7462), 287–295.

Reichstein, M., M. D. Mahecha, P. Ciais, S. I. Seneviratne, E. M. Blyth, N. Carvalhais, and Y. Luo. 2011. Elk–testing climate–carbon cycle models: a case for pattern–oriented system analysis. *iLEAPS Newsletter* **11**, 14–21.

Reichstein, M., P. Ciais, D. Papale, R. Valentini, S. Running, N. Viovy, W. Cramer, A. Granier, J. Ogee, V. Allard *et al.* 2007. Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis. *Global Change Biology* **13**(3), 634–651.

Reyer, C. P., S. Leuzinger, A. Rammig, A. Wolf, R. P. Bartholomeus, A. Bonfante, F. de Lorenzi, M. Dury, P. Gloning, R. Abou Jaoudé *et al.* 2013. A plant's perspective of extremes: terrestrial plant responses to changing climatic variability. *Glob. Chang. Biol.* **19**(1), 75–89.

Richardson, A. D., T. A. Black, P. Ciais, N. Delbart, M. A. Friedl, N. Gobron, D. Y. Hollinger, W. L. Kutsch, B. Longdoz, S. Luyssaert *et al.* 2010. Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **365**(1555), 3227–3246.

Rodell, M., P. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich *et al.* 2004. The global land data assimilation system. *Bulletin of the American Meteorological Society* **85**(3), 381–394.

Rogers, A., B. E. Medlyn, J. S. Dukes, G. Bonan, S. Caemmerer, M. C. Dietze, J. Kattge, A. D. Leakey, L. M. Mercado, Ü. Niinemets *et al.* 2017. A roadmap for improving the representation of photosynthesis in Earth system models. *New Phytologist* **213**(1), 22–42.

Rohde, R., R. Muller, R. Jacobsen, S. Perlmutter, A. Rosenfeld, J. Wurtele, J. Curry, C. Wickham, and S. Mosher. 2013. Berkeley Earth temperature averaging process. *Geoinformatics and Geostatistics: An Overview* **1**(2), 1–13.

Rolinski, S., A. Rammig, A. Walz, W. von Bloh, M. Van Oijen, and K. Thonicke. 2015. A probabilistic risk assessment for the vulnerability of the european carbon cycle to weather extremes: the ecosystem perspective. *Biogeosciences* **12**(6), 1813–1831.

Root, B., P. Knight, G. Young, S. Greybush, R. Grumm, R. Holmes, and J. Ross. 2007. A fingerprinting technique for major weather events. *Journal of Applied Meteorology* **46**(7), 1053–1066.

Rost, S., D. Gerten, A. Bondeau, W. Lucht, J. Rohwer, and S. Schaphoff. 2008. Agricultural green and blue water consumption and its influence on the global water system. *Water Resources Research*.

Rousseeuw, P. J., and M. Hubert. 2011. Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **1**(1), 73–79.

Rui, H., and H. Beaudoing. 2016. *README Document for Global Land Data Assimilation System Version 2 (GLDAS-2) Products*. Goddard Earth Sciences (GES) Data and Information Services Center.

Russo, S., A. Dosio, R. G. Graversen, J. Sillmann, H. Carrao, M. B. Dunbar, A. Singleton, P. Montagna, P. Barbola, and J. V. Vogt. 2014. Magnitude of extreme heat waves in present climate and their projection in a warming world. *Journal of Geophysical Research: Atmospheres* **119**(22), 12500–12512.

Russo, S., J. Sillmann, and E. M. Fischer. 2015. Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environmental Research Letters* **10**(12), 124003.

Saha, S., S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer *et al.* 2010. The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society* **91**(8), 1015.

Santanello, J. A., C. D. Peters-Lidard, S. V. Kumar, C. Alonge, and W.-K. Tao. 2010. Diagnosing the local land-atmosphere coupling in models and observations. *iLEAPS Newsletter* **9**, 22–24.

Santanello, J. A., J. Roundy, and P. A. Dirmeyer. 2015. Quantifying the land–atmosphere coupling behavior in modern reanalysis products over the US Southern Great Plains. *Journal of Climate* **28**(14), 5813–5829.

Santos-Fernández, E. 2012. *Multivariate statistical quality control using R*. Vol. 14. Springer Science & Business Media.

Schaller, N., A. L. Kay, R. Lamb, N. R. Massey, G. J. Van Oldenborgh, F. E. Otto, S. N. Sparrow, R. Vautard, P. Yiou, I. Ashpole *et al.* 2016. Human influence on climate in the 2014 southern England winter floods and their impacts. *Nature Climate Change* **6**(6), 627–634.

Schaller, N., F. Otto, G. van Oldenborgh, N. Massey, S. Sparrow, and M. Allen. 2014. The heavy precipitation event of May-June 2013 in the upper Danube and Elbe basins. In *Explaining Extreme Events of 2013 from a Climate Perspective*. S. Herring, M. Hoerling, T. Peterson, and P. Stott. (Eds.). Vol. 95. Bulletin of the American Meteorological Society. Pp. S69–S72.

Schaphoff, S., U. Heyder, S. Ostberg, D. Gerten, J. Heinke, and W. Lucht. 2013. Contribution of permafrost soils to the global carbon budget. *Environmental Research Letters* **8**(1), 014026.

Schär, C., P. Vidale, D. Lüthi, C. Frei, C. Häberli, M. Liniger, and C. Appenzeller. 2004. The role of increasing temperature variability in European summer heatwaves. *Nature* **427**(6972), 332–336.

Scheffran, J., M. Brzoska, J. Kominek, P. M. Link, and J. Schilling. 2012. Disentangling the climate-conflict nexus: empirical and theoretical assessment of vulnerabilities and pathways. *Review of European Studies* **4**(5), 1.

Schiermeier, Q. 2011. Extreme measures. *Nature* **477**(7363), 131–132.

Schimel, D., R. Pavlick, J. B. Fisher, G. P. Asner, S. Saatchi, P. Townsend, C. Miller, C. Frankenberg, K. Hibbard, and P. Cox. 2015. Observing terrestrial ecosystems and the carbon cycle from space. *Global Change Biology* **21**(5), 1762–1776.

Schlesinger, W. H., and S. Jasechko. 2014. Transpiration in the global water cycle. *Agricultural and Forest Meteorology* **189**, 115–117.

Schleussner, C.-F., J. F. Donges, R. V. Donner, and H. J. Schellnhuber. 2016a. Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized countries. *Proceedings of the National Academy of Sciences* **113**(33), 9216–9221.

Schleussner, C.-F., J. Rogelj, M. Schaeffer, T. Lissner, R. Licker, E. M. Fischer, R. Knutti, A. Levermann, K. Frieler, and W. Hare. 2016b. Science and policy characteristics of the Paris Agreement temperature goal. *Nature Climate Change*.

Schneider, U., A. Becker, A. Meyer-Christoffer, M. Ziese, and B. Rudolf. 2011. Global precipitation analysis products of the GPCC. *Technical report*. Global Precipitation Climatology Centre (GPCC), Deutscher Wetterdienst, Offenbach a. M., Germany.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf. 2014. GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology* **115**(1-2), 15–40.

Schulze, E.-D. 2006. Biological control of the terrestrial carbon sink. *Biogeosciences* **3**(2), 147–166.

Schwalm, C. R., C. A. Williams, K. Schaefer, A. Arneth, D. Bonal, N. Buchmann, J. Chen, B. E. Law, A. Lindroth, S. Luyssaert *et al.* 2010. Assimilation exceeds respiration sensitivity to drought: A FLUXNET synthesis. *Global Change Biology* **16**(2), 657–670.

Schwalm, C. R., C. A. Williams, K. Schaefer, D. Baldocchi, T. A. Black, A. H. Goldstein, B. E. Law, W. C. Oechel, R. L. Scott *et al.* 2012. Reduction in carbon uptake during turn of the century drought in western North America. *Nature Geoscience* **5**(8), 551–556.

Screen, J. A. 2014. Arctic amplification decreases temperature variance in northern mid-to high-latitudes. *Nature Climate Change* **4**, 577–582.

Sellers, P., R. Dickinson, D. Randall, A. Betts, F. Hall, J. Berry, G. Collatz, A. Denning, H. Mooney, C. Nobre *et al.* 1997. Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science* **275**(5299), 502–509.

Seneviratne, S. I. 2012. Climate science: Historical drought trends revisited. *Nature* **491**(7424), 338–339.

Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär. 2006. Land–atmosphere coupling and climate change in Europe. *Nature* **443**(7108), 205–209.

Seneviratne, S. I., M. G. Donat, A. J. Pitman, R. Knutti, and R. L. Wilby. 2016. Allowable CO_2 emissions based on regional and impact-related climate targets. *Nature* **529**, 477–483.

Seneviratne, S. I., M. G. Donat, B. Mueller, and L. V. Alexander. 2014. No pause in the increase of hot temperature extremes. *Nature Climate Change* **4**(3), 161–163.

Seneviratne, S. I., M. Wilhelm, T. Stanelle, B. Hurk, S. Hagemann, A. Berg, F. Cheruy, M. E. Higgins, A. Meier, V. Brovkin *et al.* 2013. Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment. *Geophysical Research Letters* **40**(19), 5212–5217.

Seneviratne, S. I., N. Nicholls, D. Easterling, C. M. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi *et al.* 2012. Changes in climate extremes and their impacts on the natural physical environment. *Managing the risks of extreme events and disasters to advance climate change adaptation* Pp. 109–230.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling. 2010a. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews* **99**(3), 125–161.

Seneviratne, S., T. Corti, E. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. Teuling. 2010b. Climate change and soil moisture-climate interactions: Using new diagnostics to identify hot spots of land-atmosphere coupling. *iLEAPS Newsletter* **9**, 18–21.

Sheather, S. J., and M. C. Jones. 1991. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B Statistical Methodology* **53**(3), 683–690.

Sheffield, J., and E. F. Wood. 2007. Characteristics of global and regional drought, 1950–2000: Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle. *Journal of Geophysical Research: Atmospheres* **112**(D17), D17115.

Sheffield, J., E. F. Wood, and F. Munoz-Arriola. 2010. Long-term regional estimates of evapotranspiration for Mexico based on downscaled ISCCP data. *Journal of Hydrometeorology* **11**(2), 253–275.

Sheffield, J., E. F. Wood, and M. L. Roderick. 2012. Little change in global drought over the past 60 years. *Nature* **491**(7424), 435–438.

Sheffield, J., G. Goteti, and E. F. Wood. 2006. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J Climate* **19**(13), 3088–3111.

Sherwood, S., and Q. Fu. 2014. A drier future?. *Science* **343**(6172), 737–739.

Sherwood, S. C., and M. Huber. 2010. An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences* **107**(21), 9552–9555.

Siegmund, J. F., M. Wiedermann, J. F. Donges, and R. V. Donner. 2016. Impact of temperature and precipitation extremes on the flowering dates of four German wildlife shrub species. *Biogeosciences* **13**(19), 5541–5555.

Sillmann, J., M. Croci-Maspoli, M. Kallache, and R. W. Katz. 2011. Extreme cold winter temperatures in europe under the influence of north atlantic atmospheric blocking. *Journal of Climate* **24**(22), 5899–5913.

Sillmann, J., V. V. Kharin, F. W. Zwiers, X. Zhang, and D. Bronaugh. 2013a. Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *Journal of Geophysical Research: Atmospheres* **118**(6), 2473–2493.

Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh. 2013b. Climate extremes indices in the cmip5 multimodel ensemble: Part 1. model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres* **118**(4), 1716–1733.

Sippel, S., and F. E. L. Otto. 2014. Beyond climatological extremes-assessing how the odds of hydrometeorological extreme events in South-East Europe change in a warming climate. *Climatic Change* **125**(3-4), 381–398.

Sippel, S., D. Mitchell, M. T. Black, A. J. Dittus, L. Harrington, N. Schaller, and F. E. Otto. 2015a. Combining large model ensembles with extreme value statistics to improve attribution statements of rare events. *Weather and Climate Extremes* **9**, 25–35.

Sippel, S., F. Otto, M. Forkel, M. Allen, B. Guillod, M. Heimann, M. Reichstein, S. Seneviratne, K. Thonicke, and M. D. Mahecha. 2016a. A novel bias correction methodology for climate impact simulations. *Earth System Dynamics* **7**(1), 71.

Sippel, S., H. Lange, M. Mahecha, M. Hauhs, P. Bodesheim, T. Kaminski, F. Gans, and O. Rosso. 2016b. Diagnosing the dynamics of observed and simulated ecosystem gross primary productivity with time causal information theory quantifiers. *PLOS ONE* **11**(10), e0164960.

Sippel, S., J. Zscheischler, and M. Reichstein. 2016c. Ecosystem impacts of climate extremes crucially depend on the timing. *Proceedings of the National Academy of Sciences* **113**(21), 5768–5770.

Sippel, S., J. Zscheischler, M. Heimann, F. E. Otto, J. Peters, and M. D. Mahecha. 2015b. Quantifying changes in climate variability and extremes: Pitfalls and their overcoming. *Geophysical Research Letters* **42**(22), 9990–9998.

Sippel, S., P. Walton, and F. E. Otto. 2015c. Stakeholder perspectives on the attribution of extreme weather events: An explorative enquiry. *Weather, Climate, and Society* **7**(3), 224–237.

Sitch, S., B. Smith, I. C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J. Kaplan, S. Levis, W. Lucht, M. T. Sykes *et al.* 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology* **9**(2), 161–185.

Smith, A. B., and R. W. Katz. 2013. US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Natural hazards* **67**(2), 387–410.

Smith, M. D. 2011. An ecological perspective on extreme climatic events: a synthetic definition and framework to guide future research. *Journal of Ecology* **99**(3), 656–663.

Stahl, K., H. Hisdal, J. Hannaford, L. Tallaksen, H. Van Lanen, E. Sauquet, S. Demuth, M. Fendekova, and J. Jordar. 2010. Streamflow trends in Europe: evidence from a dataset of near-natural catchments. *Hydrology and Earth System Sciences* **14**, p–2367.

Stark, J. D., C. J. Donlon, M. J. Martin, and M. E. McCulloch. 2007. OSTIA: An operational, high resolution, real time, global sea surface temperature analysis system. *OCEANS 2007-Europe*. IEEE. Pp. 1–4.

Stegehuis, A. I., A. J. Teuling, P. Ciais, R. Vautard, and M. Jung. 2013. Future European temperature change uncertainties reduced by using land heat flux observations. *Geophysical Research Letters* **40**(10), 2242–2245.

Stephenson, D. B., H. Diaz, and R. Murnane. 2008. *Definition, diagnosis, and origin of extreme weather and climate events*. Vol. 348. Cambridge University Press: New York.

Stine, A., P. Huybers, and I. Fung. 2009. Changes in the phase of the annual cycle of surface temperature. *Nature* **457**(7228), 435–440.

Stocker, T., D. Qin, G. Plattner, M. Tignor, S. Allen, J. Boschung *et al.* 2013. Climate Change 2013: The Physical Science Basis. Working Group 1 (WG1) Contribution to the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report (AR5). Cambridge, United Kingdom and New York, NY. *Technical report*. International Panel on Climate Change (IPCC).

Stone, D. A., and M. R. Allen. 2005. The end-to-end attribution problem: from emissions to impacts. *Climatic Change* **71**(3), 303–318.

Stone, D. A., M. R. Allen, P. A. Stott, P. Pall, S.-K. Min, T. Nozawa, and S. Yukimoto. 2009. The detection and attribution of human influence on climate. *Annual Review of Environment and Resources* **34**, 1–16.

Stone, D., M. Auffhammer, M. Carey, G. Hansen, C. Huggel, W. Cramer, D. Lobell, U. Molau, A. Solow, L. Tibig *et al.* 2013. The challenge to detect and attribute effects of climate change on human and natural systems. *Clim. Chang.* **121**(2), 381–395.

Stott, P. A., and P. Walton. 2013. Attribution of climate-related events: understanding stakeholder needs. *Weather* **68**(10), 274–279.

Stott, P. A., D. A. Stone, and M. R. Allen. 2004. Human contribution to the European heatwave of 2003. *Nature* **432**(7017), 610–614.

Stott, P. A., M. Allen, N. Christidis, R. M. Dole, M. Hoerling, C. Huntingford, P. Pall, J. Perlwitz, and D. Stone. 2013. Attribution of weather and climate-related events. Pp. 307–337. *Climate Science for Serving Society*. Springer, Dordrecht Heidelberg New York London.

Stott, P. A., N. Christidis, F. E. Otto, Y. Sun, J.-P. Vanderlinden, G. J. van Oldenborgh, R. Vautard, H. von Storch, P. Walton, P. Yiou *et al.* 2016. Attribution of extreme weather and climate-related events. *Wiley Interdisciplinary Reviews: Climate Change* **7**(1), 23–41.

Sullivan, J., and L. Sanders. 1974. *Method for Obtaining Wet-bulb Temperatures by Modifying the Psychometric Formula*. US Department of Commerce, National Oceanic and Atmospheric Administration, Environmental Data Service, Center for Experiment Design and Data Analysis.

Sun, F., M. L. Roderick, and G. D. Farquhar. 2012. Changes in the variability of global land precipitation. *Geophysical Research Letters* **39**(19), L19402.

Swann, A. L., F. M. Hoffman, C. D. Koven, and J. T. Randerson. 2016. Plant responses to increasing CO_2 reduce estimates of climate impacts on drought severity. *Proceedings of the National Academy of Sciences* **113**(36), 10019–10024.

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), 485–498.

Tebaldi, C., and R. Knutti. 2007. The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **365**(1857), 2053–2075.

Teuling, A. J., S. I. Seneviratne, R. Stöckli, M. Reichstein, E. Moors, P. Ciais, S. Luyssaert, B. van den Hurk, C. Ammann, C. Bernhofer *et al.* 2010. Contrasting response of european forest and grassland energy exchange to heatwaves. *Nature Geoscience* **3**(10), 722–727.

Teutschbein, C., and J. Seibert. 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology* **456**, 12–29.

Thibault, K. M., and J. H. Brown. 2008. Impact of an extreme climatic event on community assembly. *Proceedings of the National Academy of Sciences* **105**(9), 3410–3415.

Thompson, W. R. 1935. On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *The Annals of Mathematical Statistics* **6**(4), 214–219.

Thornthwaite, C. W. 1948. An approach toward a rational classification of climate. *Geographical Review* **38**(1), 55–94.

Thornthwaite, C. W., and J. Mather. 1955. The water balance. *Technical Report 1*. Centerton, Drexel Institute of Technology-Laboratory of Climatology. Publications in Climatology.

Tingley, M. P. 2012. A bayesian ANOVA scheme for calculating climate anomalies, with applications to the instrumental temperature record. *J. Clim.* **25**(2), 777–791.

Tingley, M. P., and P. Huybers. 2013. Recent temperature extremes at high northern latitudes unprecedented in the past 600 years. *Nature* **496**(7444), 201–205.

Tollefson, J. 2016. Global warming already driving increases in rainfall extremes. *Nature News*.

Toms, J., and M. Lesperance. 2003. Piecewise regression: a tool for identifying ecological thresholds. *Ecology* **84**(8), 2034–2041.

Tramontana, G., M. Jung, C. R. Schwalm, K. Ichii, G. Camps-Valls, B. Ráduly, M. Reichstein, M. A. Arain, A. Cescatti, G. Kiely *et al.* 2016. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences* **13**, 4291–4313.

Trenberth, K. E. 1999. Conceptual framework for changes of extremes of the hydrological cycle with climate change. *Climatic Change* **42**(1), 327–339.

Trenberth, K. E. 2011. Changes in precipitation with climate change. *Climate Research* **47**(1-2), 123–138.

Trenberth, K. E., A. Dai, G. Van Der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, and J. Sheffield. 2014. Global warming and changes in drought. *Nature Climate Change* **4**(1), 17–22.

Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons. 2003. The changing character of precipitation. *Bulletin of the American Meteorological Society* **84**(9), 1205–1217.

Trenberth, K. E., J. T. Fasullo, and J. Kiehl. 2009. Earth's global energy budget. *Bulletin of the American Meteorological Society* **90**(3), 311–323.

Van den Brink, H., and G. Können. 2011. Estimating 10000-year return values from short time series. *International Journal of Climatology* **31**(1), 115–126.

van den Hurk, B., H. Kim, G. Krinner, S. I. Seneviratne, C. Derksen, T. Oki, H. Douville, J. Colin, A. Ducharne, F. Cheruy, N. Viovy, M. J. Puma, Y. Wada, W. Li, B. Jia, A. Alessandri, D. M. Lawrence, G. P. Weedon, R. Ellis, S. Hagemann, J. Mao, M. G. Flanner, M. Zampieri, S. Materia, R. M. Law, and

J. Sheffield. 2016. LS3MIP (v1.0) contribution to CMIP6: the land surface, snow and soil moisture model intercomparison project – aims, setup and expected outcome. *Geoscientific Model Development* **9**(8), 2809–2832.

Van Huijgevoort, M., H. Van Lanen, A. Teuling, and R. Uijlenhoet. 2014. Identification of changes in hydrological drought characteristics from a multi-gcm driven ensemble constrained by observed discharge. *Journal of Hydrology* **512**, 421–434.

Van Oijen, M., J. Balkovic, C. Beer, D. Cameron, P. Ciais, W. Cramer, T. Kato, M. Kuhnert, R. Martin, R. Myneni *et al.* 2014. Impact of droughts on the carbon cycle in European vegetation: a probabilistic risk analysis using six vegetation models. *Biogeosciences* **11**(22), 6357–6375.

Van Oldenborgh, G., A. Van Urk, and M. Allen. 2012. The absence of a role of climate change in the 2011 Thailand floods. *Bulletin of the American Meteorological Society* **93**(7), 1047–1049.

Vautard, R., A. Gobiet, D. Jacob, M. Belda, A. Colette, M. Déqué, J. Fernández, M. García-Díez, K. Goergen, I. Güttler *et al.* 2013. The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project. *Climate Dynamics* **41**(9-10), 2555–2575.

von Buttlar, J., J. Zscheischler, and M. D. Mahecha. 2014. An extended approach for spatiotemporal gapfilling: dealing with large and systematic gaps in geoscientific datasets. *Nonlinear Processes in Geophysics* **21**, 203–215.

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

Wang, C., L. Zhang, S.-K. Lee, L. Wu, and C. R. Mechoso. 2014. A global perspective on CMIP5 climate model biases. *Nature Climate Change* **4**(3), 201–205.

Wang, K., and R. E. Dickinson. 2012. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics*.

Warszawski, L., K. Frieler, V. Huber, F. Piontek, O. Serdeczny, and J. Schewe. 2014. The inter-sectoral impact model intercomparison project (ISI–MIP): Project framework. *Proceedings of the National Academy of Sciences of the United States of America* **111**(9), 3228–3232.

Weedon, G., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo. 2014. The WFDEI meteorological forcing data set: WATCH forcing data methodology applied to ERA-Interim reanalysis data. *Water Resources Research* **50**(9), 7505–7514.

Weedon, G., S. Gomes, P. Viterbo, W. Shuttleworth, E. Blyth, H. Österle, J. Adam, N. Bellouin, O. Boucher, and M. Best. 2011. Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *Journal of Hydrometeorology* **12**(5), 823–848.

Weiner, J. 2014. *riverplot: Sankey or ribbon plots*. R package version 0.3.

Wenzel, S., P. M. Cox, V. Eyring, and P. Friedlingstein. 2014. Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 earth system models. *Journal of Geophysical Research: Biogeosciences* **119**(5), 794–807.

Westra, S., L. V. Alexander, and F. W. Zwiers. 2013. Global increasing trends in annual maximum daily precipitation. *Journal of Climate* **26**(11), 3904–3918.

Wetter, O., C. Pfister, J. P. Werner, E. Zorita, S. Wagner, S. I. Seneviratne, J. Herget, U. Grünewald, J. Luterbacher, M.-J. Alcoforado *et al.* 2014. The yearlong unprecedented European heat and drought of 1540–a worst case. *Climatic Change* **125**(3-4), 349–363.

Whan, K., J. Zscheischler, R. Orth, M. Shongwe, M. Rahimi, E. O. Asare, and S. I. Seneviratne. 2015. Impact of soil moisture on extreme maximum temperatures in Europe. *Weather and Climate Extremes* **9**, 57–67.

Wickham, H. 2009. *ggplot2: elegant graphics for data analysis*. Springer New York.

Wigley, T. M. 2009. The effect of changing climate on the frequency of absolute extreme events. *Climatic Change* **97**(1-2), 67–76.

Williams, A. P., C. D. Allen, A. K. Macalady, D. Griffin, C. A. Woodhouse, D. M. Meko, T. W. Swetnam, S. A. Rauscher, R. Seager, H. D. Grissino-Mayer *et al.* 2013. Temperature as a potent driver of regional forest drought stress and tree mortality. *Nature Climate Change* **3**(3), 292–297.

WMO 1989. Calculation of monthly and annual 30-year standard normals. *Technical report*. World Meteorological Institution.

WMO 2011. Weather extremes in a changing climate: Hindsight on foresight. World Meteorological Organization.

Wofsy, S., M. Goulden, J. Munger, S.-M. Fan, P. Bakwin, B. Daube, S. Bassow, and F. Bazzaz. 1993. Net exchange of CO_2 in a mid-latitude forest. *Science* **260**(5112), 1314–1317.

Wolf, S., T. F. Keenan, J. B. Fisher, D. D. Baldocchi, A. R. Desai, A. D. Richardson, R. L. Scott, B. E. Law, M. E. Litvak, N. A. Brunsell *et al.* 2016. Warm spring reduced carbon cycle impact of the 2012 US summer drought. *Proceedings of the National Academy of Sciences* **113**(21), 5880–5885.

Wood, A. W., L. R. Leung, V. Sridhar, and D. Lettenmaier. 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change* **62**(1-3), 189–216.

Xoplaki, E., J. Gonzalez-Rouco, J. Luterbacher, and H. Wanner. 2003. Mediterranean summer air temperature variability and its connection to the large-scale atmospheric circulation and ssts. *Climate Dynamics* **20**(7-8), 723–739.

Xu, X. T., S. L. Piao, X. H. Wang, A. P. Chen, P. Ciais, and R. B. Myneni. 2012. Spatio-temporal patterns of the area experiencing negative vegetation growth anomalies in China over the last three decades. *Environmental Research Letters* **7**(3), 1–9.

Yang, Y., H. Guan, O. Batelaan, T. R. McVicar, D. Long, S. Piao, W. Liang, B. Liu, Z. Jin, and C. T. Simmons. 2016. Contrasting responses of water use efficiency to drought across global terrestrial ecosystems. *Scientific Reports* **6**(23284), 1–8.

Zaehle, S., P. Friedlingstein, and A. D. Friend. 2010. Terrestrial nitrogen feedbacks may accelerate future climate change. *Geophysical Research Letters* **37**(1), L01401.

Zander, K. K., W. J. Botzen, E. Oppermann, T. Kjellstrom, and S. T. Garnett. 2015. Heat stress causes substantial labour productivity loss in Australia. *Nature Climate Change* **5**(7), 647–651.

Zhang, X., G. Hegerl, F. W. Zwiers, and J. Kenyon. 2005. Avoiding inhomogeneity in percentile-based indices of temperature extremes. *Journal of Climate* **18**(11), 1641–1651.

Zhang, X., G. Hegerl, S. Seneviratne, R. Stewart, F. Zwiers, and L. Alexander. 2014. WCRP grand challenge: Understanding and predicting weather and climate extremes. *Technical report*. World Climate Research Program. White Paper.

Zhang, Y., R. Leuning, L. B. Hutley, J. Beringer, I. McHugh, and J. P. Walker. 2010. Using long-term water balances to parameterize surface conductances and calculate evaporation at 0.05 spatial resolution. *Water Resources Research* **46**(5), W05512.

Zhou, Y., and G. Ren. 2011. Change in extreme temperature event frequency over mainland China, 1961- 2008. *Climate Research* **50**(2-3), 125–139.

Zieba, A. 2010. Effective number of observations and unbiased estimators of variance for autocorrelated data an - overview. *Metrology and Measurement Systems* **17**(1), 3–16.

Zscheischler, J., A. Michalak, C. Schwalm, M. Mahecha, D. Huntzinger, M. Reichstein, G. Berthier, P. Ciais, R. Cook, B. El-Masri *et al.* 2014a. Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data. *Global Biogeochemical Cycles* **28**(6), 585–600.

Zscheischler, J., M. D. Mahecha, J. von Buttlar, S. Harmeling, M. Jung, A. Rammig, J. T. Randerson, B. Schölkopf, S. I. Seneviratne, E. Tomelleri *et al.* 2014b. A few extreme events dominate global interannual variability in gross primary production. *Environmental Research Letters* **9**(3), 035001.

Zscheischler, J., M. D. Mahecha, S. Harmeling, and M. Reichstein. 2013. Detection and attribution of large spatiotemporal extreme events in Earth observation data. *Ecological Informatics* **15**, 66–73.

Zscheischler, J., M. Reichstein, S. Harmeling, A. Rammig, E. Tomelleri, and M. D. Mahecha. 2014c. Extreme events in gross primary production: a characterization across continents. *Biogeosciences* **11**, 2909–2924.

Zscheischler, J., R. Orth, and S. I. Seneviratne. 2015. A submonthly database for detecting changes in vegetation-atmosphere coupling. *Geophysical Research Letters* **42**(22), 9816–9824.

Zwiers, F. W., and H. von Storch. 1995. Taking serial correlation into account in tests of the mean. *J. Clim.* **8**(2), 336–351.

A. Ecosystem impacts of climate extremes crucially depend on the timing[1]

> '*The year 1540 was unprecedented in centuries. It was dreadful, bright, and hot. Bright weather and heat [...] lasted for 29 weeks, in which rain fell on not more than 6 days [...]. Meadows and forests were yellow from the heat and the earth opened large cracks; at several locations grapes and vine withered, many forests burned, fountains and springs dried out completely. [...] (But) there was an abundance of corn and a lot of delicious wine.*'

Translated from German, a contemporary witness describing the contrasting impacts of a mega-heat and drought event of 1540 in Europe (Wetter et al., 2014).

The impacts of climate extremes have always been of crucial importance to human societies, but they also play a key role in affecting structure and functioning of ecosystems. Whether there are any impacts at all, and how these impacts manifest themselves, critically depends on the timing, magnitude, extent, and type of the climate anomaly. Although many studies have been undertaken to investigate the impacts of climate extremes on ecosystem functioning, attempts to build an overarching framework have had little success so far and many open questions remain (Frank et al., 2015). A study published in *Proceedings of the National Academy of Sciences of the United States of America* (Wolf et al., 2016) provides new insights into the question of how impacts of climate extremes occurring during different periods of the year can interact and counteract each other.

Wolf et al. (2016) investigated the year 2012 and its impacts on terrestrial carbon fluxes in the continental United States, an extreme year in which a record

[1]This chapter is published as Sippel, S., J. Zscheischler, and M. Reichstein. 2016. *Proceedings of the National Academy of Sciences of the United States of America* **113**(21), 5768–5770. doi:10.1073/pnas.1605667113.

warm spring was followed by a severely dry and hot summer (Knutson et al., 2013; Hoerling et al., 2014). The authors analysed three independent streams of observational data and data-driven models, and demonstrated that losses in net carbon uptake during summer were largely offset by unusual carbon gains in spring caused by its record-exceeding warmth and early arrival. In this way, the continental United States remained a carbon sink despite the exceptional drought that spanned most of the country. This news is good and suggests that warmer springs can alleviate the devastating impacts of summer droughts (Figure A1). The bad news, however, might follow suit: Because ecosystem fluxes of carbon and water are tightly coupled through plant stomata, higher spring carbon uptake might lead to an earlier depletion of soil water resources through increased evapotranspiration, thus amplifying extreme temperatures in the summer. Wolf et al. (2016) hypothesise that this effect has exacerbated the 2012 summer drought and contributed to elevated surface heating, and thereby highlight the important role that land-atmosphere feedbacks could play during climate extremes. However, it cannot be excluded that a less warm spring would have depleted soil water resources less rapidly, rendering the impacts of the rainfall deficit in summer less severe. These important questions have not been answered definitely and deserve more detailed investigations. It is critical to disentangle the different counteracting feedbacks, not least because events such as the year 2012 in the United States might occur more often in the future.

The authors arrive at their synthesis by combining so-called ‘bottom-up’ with ‘top-down’ approaches. A network of local flux tower measurements of carbon and water exchange across the United States on land was complemented with photosynthetic carbon uptake derived from satellite remote sensing and an atmospheric inverse model that estimates net carbon uptake using atmospheric measurements of CO2 concentrations. The study thus provides empirical evidence both at the ecosystem and continental scales that two different but prevalent types of climate extremes in temperate ecosystems can have compensatory impacts on the carbon cycle. The results complement previous analyses indicating that dry summers offset increases in vegetation carbon uptake driven by warmer springs in the Northern Hemisphere extratropics (Angert et al., 2005). Empirical insights into carbon and water cycle dynamics aside, however, the study highlights important scientific questions related to (i) disentangling the extent, magnitude, and relevant components that contributed to a compensation of climate extreme-related impacts and (ii) understanding and quantifying plant-soil-atmosphere feedbacks in a warmer world.

A.1. Identifying carbon cycle components that cancel out

To advance our process understanding about impacts of events such as the extreme spring and summer in 2012 in the United States, it is important to understand which components of the carbon cycle contributed to the observed compensation. The net ecosystem carbon flux is the difference between the plant's photosynthetic carbon uptake and carbon losses through ecosystem respiration. During the summer of 2012, reductions in photosynthetic carbon uptake exceeded the reduction in respiratory carbon flux, consistent with previous observations during droughts (Ciais et al., 2005). Despite the observed surplus in gross carbon uptake in spring, annual gross carbon uptake remained substantially below average across the continental United States. Surprisingly, annual net carbon uptake in the continental United States was still close to average, which highlights the role of ecosystem respiration in shaping the impacts of climate extremes on net carbon uptake. Ecosystem respiration increased in spring only moderately, whereas its decrease in summer was large. Grasping how individual carbon cycle components react to climate extremes and implementing these processes into mechanistic models may thereby lead to better constrained carbon projections (Friedlingstein et al., 2014). On a different note, Wolf et al. (2016) find that high spring uptake, particularly in the eastern temperate forests, prevented the United States from shifting from a carbon sink to a carbon source. The spatially nonuniform signal demonstrates how the impacts of climate extremes differ between ecosystems and illustrates the challenge associated with finding general response patterns to climate extremes (Frank et al., 2015). What can be beneficial for one ecosystem might be devastating for another (Teuling et al., 2010). Global climate models indicate that warm spring temperatures similar to the temperatures in 2012 lie at the cooler end of the temperature distribution in the second half of the 21st century (Wolf et al., 2016). In contrast, severe summer droughts will remain rare but

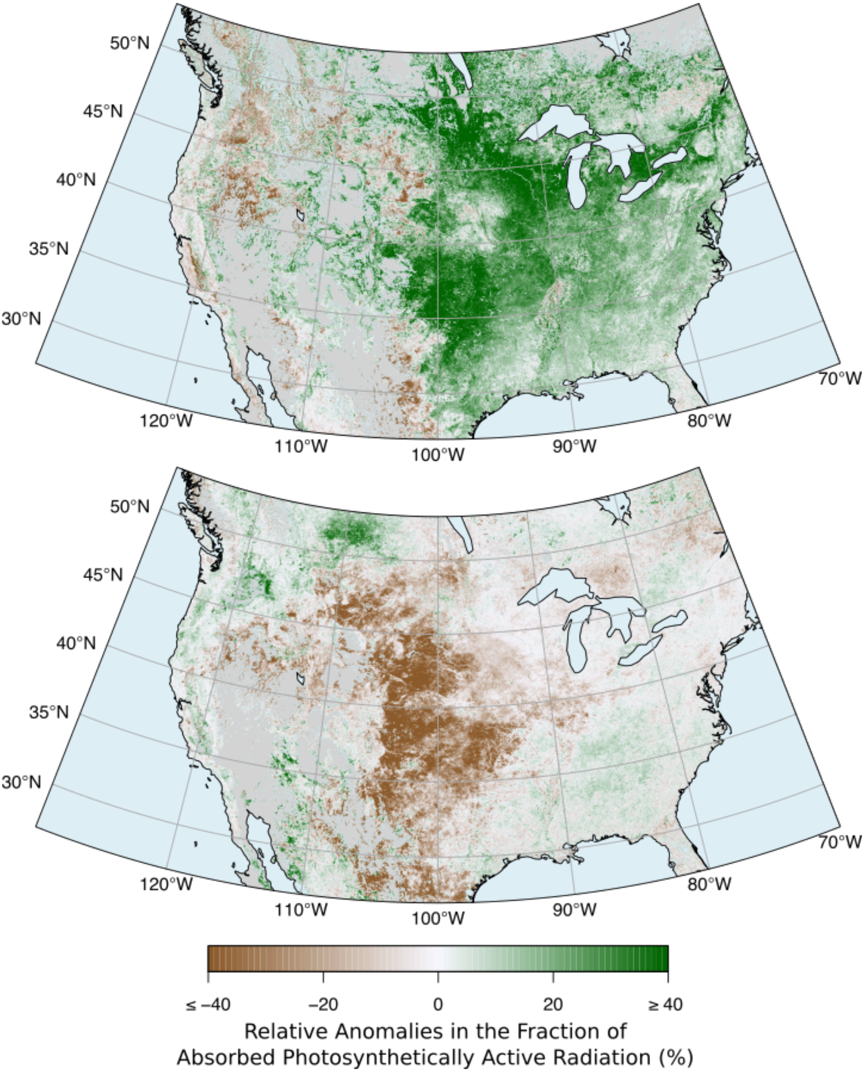
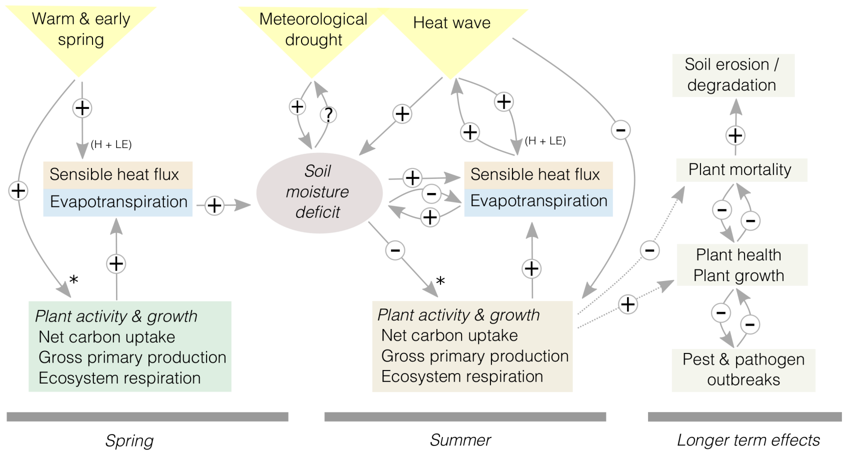


FIGURE A1.: Early spring gains (a, March-April) and late summer reductions (b, July-August) in the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR; %), an indicator for vegetation activity, in the year 2012 relative to 2001-2014. Grid cells with a long-term mean FAPAR below 10% are shown in gray.

impacts will likely be exacerbated by hotter temperatures (Williams et al., 2013). Over the past few decades, net carbon uptake in temperate forests has increased because of warmer spring temperatures that induce a lengthening of the growing season (Menzel et al., 2006; Keenan et al., 2014). Overall warmer springs might thus offset some of the adverse impacts of hot summer droughts. However, to start leaf unfolding, temperate forests also require a sufficient degree of winter chilling, which implies that the observed warming-induced changes might not simply follow spring temperatures in the future (Körner and Basler, 2010; Fu et al., 2015). Wolf et al. (2016) have disentangled temporal and spatial components of ecosystem carbon impacts of the anomalous year 2012, but experimentalists and modelers will have to work together to figure out whether positive impacts of more favourable spring conditions or adverse impacts of dry and hot summers will prevail under future climate conditions.

A.2. The role of plant-soil-atmosphere feedbacks in enhancing summer heat

That land-atmosphere feedbacks can strongly influence the magnitude of extreme heatwaves and droughts has long been acknowledged (Seneviratne et al., 2010a). Dry soils can exacerbate extremely high temperatures, whereas wet soils impede the development of extreme heat waves through evaporative cooling (Miralles et al., 2014). The role of plants in these feedback mechanisms is much less well understood. Warmer conditions, accompanied by higher radiation, generally lead to higher photosynthetic activity, particularly in the energy-limited areas that span most of the United States. More photosynthetic activity induces higher evapotranspiration rates, thereby depleting soil water. If photosynthesis is strongly enhanced in spring and soil water is not replenished through precipitation, in tandem with high summer temperatures (e.g., through a blocking event), the dry soils will enhance the summer heat because more of the incoming radiation is translated into sensible heat (Seneviratne et al., 2010a). Quantifying the different contributions of vegetation, lack of precipitation, and spring temperatures to the resulting concurrent drought and heatwave in summer is challenging (Figure A2). To disentangle the impacts of enhanced photosynthesis in spring on summer drought and summer temperatures, one could conduct, for example, factorial model runs with and without vegetation. Wolf et al. (2016) show in their study that neither seasons (spring and summer) nor carbon, water, and energy fluxes should be interpreted separately when analyzing the impacts of climate extremes. On the one hand, the authors see depletion of soil moisture through early vegetation activity in a warm spring potentially amplifying summer heating, a typical lagged direct effect of an extremely warm spring (Frank et al., 2015). On the other hand, spring and summer, and photosynthesis and respiration, compensate each other with respect to the net annual effect on the carbon cycle, leading to a near-neutral same-year carbon balance. Can one thus speak of an overall reduced net carbon impact of the 2012 drought? The future will tell, because lagged and indirect effects can be important. Mechanisms for such effects include, for instance, depending on the ecosystem, plant mortality, pathogen dynamics, or soil erosion and degradation (Allen et al., 2010; Reichstein et al., 2013). If 2012 conditions become more frequent in the future, in concert with potential mitigation effects through elevated CO_2 (Leakey et al., 2009), the competition between plant populations induced by vegetation dynamics may lead to either enhanced carbon storage (e.g., in woody vegetation) or depletion. Thus, for understanding the ‘true’ integral effect of a year like 2012, it is important that we monitor and analyse subsequent years, which is possible thanks to long-term observations established by the respective research networks [e.g., AmeriFlux, Europe’s Integrated Carbon Observation System (ICOS), and the National Ecological Observatory Network (NEON)]. In addition, even longer term archives (e.g., in tree rings or lake sediments) should provide complementary information in terms of the time scale and processes involved.



* Climate effects on ecosystem carbon fluxes are shown only in qualitative terms. Individual fluxes might be affected differently by climate extremes (see text).

FIGURE A2.: Conceptual framework of potential plant-soil-atmosphere feedbacks and ecosystem impacts. Solid arrows indicate direct impacts (positive or negative), and dashed arrows show hypothesised longer term effects of summer drought. H, sensible heat; LE, latent heat or evapotranspiration.

B. Supplementary Material for Chapter 2

B.1. Guide to the artificial normalisation example

We provide the original source code that was used to carry out the artificial normalisation example shown in Figure 2.1 in a step-by-step guide using the R Statistical Programming Environment (R Development Core Team, 2013). We first generate an artificial dataset containing 10,000 time series, where each time series consists of $n = 60$ independent and identically distributed Gaussian variables. As stated in the main text, this can be understood as an analogy to a spatio-temporal temperature dataset that comprises 60 years of data across 10,000 geographical grid cells. Subsequently, each time series is centered and scaled with estimates of the mean and standard deviation as derived from a reference period of length $n_{ref} = 30$ (here, the first 30 values of each time series are chosen). For each time point t, we then count the number of σ-extremes in the original Gaussian data and the normalised data (Figure 2.1). Lastly, the proposed correction (for a formal derivation see Section B.2) leads to the corrected normalised time series shown in Figure 2.2a. A more detailed tutorial and R-code for normalisation and correction is available under https://github.com/sebastian-sippel/normalization.

```
# Define parameters for normalisation example:
nref = 30;          # Length of reference period
ngridcells = 10000; # Number of independent grid cells
sigma = 2;          # Sigma threshold
```

```r
# Generate Gaussian time series each of which consists of 60 values:
data.orig = sapply(1:ngridcells, FUN=function(x) rnorm(60));
# Estimate the mean and standard deviation of each time series
# based on the reference period (first 30 values):
mean.estimate = sapply(1:ngridcells, FUN=function(x) mean(data.orig[1:nref,x]));
sd.estimate = sapply(1:ngridcells, FUN=function(x) sd(data.orig[1:nref,x]));
# Generate anomalies, and normalise each time series with its sample mean
# and sample standard deviation:
data.anom = sapply(1:ngridcells, FUN=function(x) data.orig[,x]-mean.estimate[x]);
data.norm = sapply(1:ngridcells, FUN=function(x) data.anom[,x]/sd.estimate[x]);
```

```r
# count +2sigma events throughout each time series and at each time step, for the
# original and normalised data:
data.orig.2sigma.extremes = apply(X=data.orig, 1, function(x) length(which(x > 2)));
data.norm.2sigma.extremes = apply(X=data.norm, 1, function(x) length(which(x > 2)));
# Compute the corrected number of sigma extremes:
# Out-of-base period:
data.norm.2sigma.extremes.obase.cor = apply(X=data.norm, MARGIN=c(1),
FUN=function(x) length(which((x / sqrt(1+1/nref)) > qt(pnorm(sigma), df=nref-1))));
# Reference period:
data.norm.2sigma.extremes.ibase.cor = apply(X=data.norm, MARGIN=c(1),
FUN=function(x) length(which(((x*x)*nref/((nref-1)*(nref-1))) > qbeta(pnorm(sigma),
shape1 = 0.5, shape2 = nref/2-1))));
```

```
# Plot the number of sigma extremes:
plot(data.norm.2sigma.extremes, col='darkred', pch=8)
points(data.orig.2sigma.extremes, col='black', pch=8)
points(x = 1:nref, data.norm.2sigma.extremes.ibase.cor[1:nref], col
    ='darkblue',
pch=8)
points(x = c(1:60)[-(1:nref)], data.norm.2sigma.extremes.obase.cor
    [-(1:nref)],
col='darkgreen', pch=8)
legend('topleft', c('Conventional normalisation', 'i.i.d. Gaussian
    variables',
 'Normalisation + correction, reference period', 'Normalisation +
    correction,
 out-of-base'), col=c('darkred', 'black', 'darkblue', 'darkgreen'),
     pch=8)
```

B.2. Normalisation-induced changes to stationary and independent Gaussian time series

At any grid cell i, time series of the form $X_{t,i}; t = 1, ..., n; i = 1, ..., k$ are normalised to yield standardised ‘z-scores’ with respect to a defined reference period as a subset of the full record:

$$z_{t,i} = \frac{X_{t,i} - \hat{\mu}_{ref,i}}{\hat{\sigma}_{ref,i}} \quad . \tag{B.1}$$

In this example, each sample in each time series $X_{t,i}$ is drawn independently from a Gaussian distribution with the expected value $E[X_{t,i}] = \mu_i$ and the variance given by $Var(X_{t,i}) = \sigma_i^2$. Thus, the estimators $\hat{\mu}_i$ for the mean μ_i and the estimator $\hat{\sigma}_i^2$ for the variance σ_i^2 satisfy (Von Storch and Zwiers, 2001) in each grid cell

$$\hat{\mu}_i = \frac{1}{n}\sum_{t=1}^{n} X_{t,i} \sim \mathcal{N}(\mu_i, \frac{\sigma_i^2}{n}) \quad \text{and} \tag{B.2}$$

$$\hat{\sigma}_i^2 = \frac{1}{n-1}\sum_{t=1}^{n}(X_{t,i} - \hat{\mu}_i)^2 \sim \sigma_i^2 \chi_{n-1}^2 \frac{1}{n-1} \quad . \tag{B.3}$$

Hence, the collection of sample means $\hat{\mu}_{ref,i}$ follows a normal distribution with expected value $E[\hat{\mu}_{ref,i}] = \mu_i$ and variance $Var(\hat{\mu}_{ref,i}) = \frac{\sigma_i^2}{n_{ref}}$ (Eq. B.2) across grid cells. Here we show that this widely used normalisation approach changes the statistical properties of the distribution across grid cells. This extends an issue previously discussed (Zhang et al., 2005), but here we are not confined to percentile-based estimates of temperature extremes. In the following subsections we distinguish normalisation in the reference period (where the estimators are dependent on the samples) from the normalisation in the out-of-base period, where the estimators are independent from the samples.

In the following sections we consider each grid cell independently. In order to improve readability, we therefore omit the index i for the grid cells and simply write X_t.

B.2.1. Normalisation in the out-of-base period

At any time t in the (independent) out-of-base period, the anomalies are given by the random variable

$$X_{anom,t} = X_t - \hat{\mu}_{ref} \quad , \tag{B.4}$$

with different realisations across grid cells. Consequently, anomalies that are generated by subtracting the reference period (that is, independent) sample mean follow again a Gaussian distribution, because the difference between two Gaussian variables $X = X_1 - X_2$ is Gaussian distributed (Johnson et al., 1994) with $\mu = \mu_1 - \mu_2$ and variance $\sigma^2 = \sigma_1^2 + \sigma_2^2$, i.e.,

$$X_{anom,t} \sim \mathcal{N}(0, \sigma^2(1 + \frac{1}{n_{ref}})) \quad . \tag{B.5}$$

Please note that the increase in variance caused by deriving anomalies and implied by Eq. B.5 holds for any distribution with finite variances, i.e. not only Gaussian distributions.

Dividing anomalies by the estimated standard deviation ('standardizing') yields standardised 'z-scores':

$$z_t = \frac{X_{anom,t}}{\hat{\sigma}_{ref}} \quad . \tag{B.6}$$

Following Eq. B.3, the 'z-scores' are characterised by Student's t-distribution with $\nu = n - 1$ degrees of freedom (cf. the definition of the t-distribution (Fisher, 1925)), which is scaled by the variance inflation given in Eq. B.5:

$$z_t \sim \sqrt{1 + \frac{1}{n_{ref}}} \cdot t(n_{ref} - 1) \quad . \tag{B.7}$$

Hence, after normalisation, we expect the grid cell values at any given time step t in the out-of-base period to follow a scaled t-distribution (Eq. B.7), rather than the Gaussian distribution as implied in earlier reports (Hansen et al., 2012; Coumou and Robinson, 2013). Although the t-distribution converges against the Gaussian distribution for a large number of degrees of freedom (i.e. increasing n_{ref}, see Figure 2.1 and Figure B1), its tails are considerably heavier even for a relatively large number of degrees of freedom. This well-known distribution allows us to derive a correction based on quantiles for normalised z-scores that can be constructed by adjusting the 'σ-extreme' of interest using Eq. B.7 (see Figure B1 for an illustration). For example, the probability of a 2σ-extreme in a Gaussian distribution corresponds to a 2.12σ event in the scaled t-distribution (for $n_{ref} = 30$, Section B.1).

B.2.2. Normalisation in the reference period

In the reference period, the estimators of mean and variance are not independent from the samples. This fact causes the underestimation of extremes in the reference period, as illustrated for instance in Figure 2.1. In this subsection, we first discuss the changes induced to the distribution by deriving anomalies (i.e. Eq. B.4), and secondly demonstrate how changes induced by normalisation according to Eq. B.6 in the reference period can be analytically corrected.

The generation of anomalies in the reference period in analogy to Eq. B.4 reduces the variability across grid cells to $Var(X_{anom,t}) = \sigma^2(1 - \frac{1}{n_{ref}})$. Note that this result does not only hold for the Gaussian distribution but for any distribution with finite second moments:

Lemma: Let there be k independent random variables X_i, each having the same standard deviation σ (i.e., $Var(X_i) = \sigma^2$). Assume we draw n samples i.i.d from each random variable and let $X_{i,j}$ be such a sample ($i = 1, ..., k$ and $j = 1, ..., n$). Let further $\hat{\mu}(X)_i$ denote the sample mean of X_i estimated from those n samples. Then

$$\hat{Cov}(X_{i,j}, \hat{\mu}(X)_i) \rightarrow \frac{\sigma^2}{n} \quad \text{for each} \quad j = 1, ..., n \quad \text{as} \quad k \rightarrow \infty. \tag{B.8}$$

Proof: W.l.o.g. let $j = 1$. Then

$$\hat{Cov}(X_{i,1}, \hat{\mu}(X)_i) =$$

$$= \frac{1}{k-1}\sum_{i=1}^{k}(X_{i,1}-\bar{X}_{i,1})(\frac{1}{n}\sum_{j=1}^{n}X_{i,j}-\overline{\frac{1}{n}\sum_{j=1}^{n}X_{i,j}}) \tag{B.9a}$$

$$= \frac{1}{k-1}\sum_{i=1}^{k}(X_{i,1}-\frac{1}{k}\sum_{l=1}^{k}X_{l,1})(\frac{1}{n}\sum_{j=1}^{n}X_{i,j}-\frac{1}{nk}\sum_{m=1}^{k}\sum_{j=1}^{n}X_{m,j}) \tag{B.9b}$$

$$= \frac{1}{k-1}\sum_{i=1}^{k}(X_{i,1}-\frac{1}{k}\sum_{l=1}^{k}X_{l,1})\frac{1}{n}\sum_{j=1}^{n}(X_{i,j}-\frac{1}{k}\sum_{m=1}^{k}X_{m,j}) \tag{B.9c}$$

$$= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{k-1}\sum_{i=1}^{k}(X_{i,1}-\frac{1}{k}\sum_{l=1}^{k}X_{l,1})(X_{i,j}-\frac{1}{k}\sum_{m=1}^{k}X_{m,j}) \tag{B.9d}$$

$$= \frac{1}{n}(\frac{1}{k-1}\sum_{i=1}^{k}(X_{i,1}-\frac{1}{k}\sum_{l=1}^{k}X_{l,1})^2+$$

$$\sum_{j=2}^{n}\frac{1}{k-1}\sum_{i=1}^{k}(X_{i,1}-\frac{1}{k}\sum_{l=1}^{k}X_{l,1})(X_{i,j}-\frac{1}{k}\sum_{l=1}^{k}X_{l,j})) \tag{B.9e}$$

$$= \frac{1}{n}\hat{\sigma}^2+\frac{n-1}{n}\hat{Cov}(X_{.,1},X_{.,j\neq 1}) \tag{B.9f}$$

$$\rightarrow \frac{\sigma^2}{n} \quad \text{as} \quad k\rightarrow\infty \tag{B.9g}$$

$\square$

With the above lemma it follows (assuming that k, the number of grid cells, is very large):

$$
\begin{aligned}
Var(X_{anom,t}) &= Var(X_t - \hat{\mu}_{ref}(X)) \\
&= Var(X_t) - 2Cov(X_t, \hat{\mu}_{ref}(X_t)) + \frac{Var(X_t)}{n_{ref}} \\
&= Var(X_t) - 2\frac{1}{n_{ref}}\sum_{s=1}^{n_{ref}} Cov(X_t, X_s) + \frac{Var(X_t)}{n_{ref}} \\
&= \sigma^2 - 2\frac{\sigma^2}{n_{ref}} + \frac{\sigma^2}{n_{ref}} \\
&= \sigma^2(1 - \frac{1}{n_{ref}}) \quad .
\end{aligned}
$$

A subsequent standardisation of anomalies following Eq. B.6 in the reference period changes the sample distribution across grid cells qualitatively to a non-Gaussian distribution. The resulting distribution follows a beta-distribution (Thompson, 1935; Johnson et al., 1995)

$$
(\frac{X_{anom,t}}{\hat{\sigma}_{ref}})^2 \sim n_{ref} Beta(0.5, \frac{n_{ref}-1}{2}) \quad . \tag{B.10}
$$

Alternatively, the distribution of standardised anomalies within the reference period has been described as a 'tau-distribution' (Thompson, 1935), where τ is defined as $\tau = \frac{X_{anom,t}}{\hat{\sigma}_{ref}}$. Here, tau is related to a t-distribution with $\nu = n_{ref} - 2$ degrees of freedom by $\tau = t_\nu \sqrt{\frac{n_{ref}-1}{n_{ref}-2+t_\nu^2}}$. Similarly to above, the transformation given by Eq. B.10 can be used to adjust the detection of normalised extremes within the reference period by quantile adjustments (see Figure B1). From the quantile-quantile plots shown in Figure B1 it becomes obvious that a normalisation across time-invariant Gaussian data yields an underestimation of extremes in the reference period (a), and an overestimation in the out-of-base (independent) period (b).

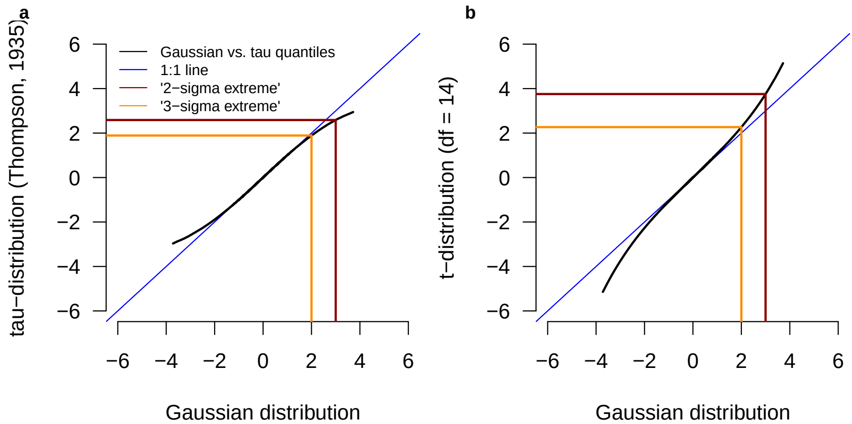


FIGURE B1.: Proposed analytical correction for normalisation-induced artefacts. Quantile-quantile plots of original Gaussian distributions vs. a) *tau*-distribution and b) the corresponding t-distribution after normalisation. The reference period length was chosen as $n = 15$ for illustration purposes. The simple quantile correction proposed is illustrated for the normalisation within a reference period (a) and in the out-of-base (independent) period (b) for 2σ and 3σ extremes.

B.3. Monte Carlo simulations

In order to test how specific features that are present in climatic data might affect the biases in normalised tails in the detection of spatially aggregated extremes, we conduct a variety of Monte-Carlo type simulations.

Each simulation is set up as follows:

- Generate $k = 10^5$ time series, each of which with $n = 130$ data points, drawn independently from a Gaussian distribution (exception: autocorrelated time series, see below).
- Define a reference period length of $n_{ref} = 30$, which has been used in climatological studies (Hansen et al., 2012) (exception: experiment using a variable reference period length, see below).
- Define remaining 100 data points in each time series as the out of base period.

- Detect extremes by counting 'σ extremes' in normalised and original time series for each time step t.
- Calculate the biases in the tails as relative differences (in percent) between the conventionally normalised time series (Eq. 2.2) and the original time series (i.e. without normalisation).

First, we test how the length of the reference period influences biases in the tails. It can be seen from the analytical argument put forward in section B.2 that the biases in the normalised tails are a function of sample size in the reference period. To illustrate this, we vary the length of the reference period (Figure B2a). The biases are decreasing for longer reference periods. However, in practical terms relatively large sample sizes in the reference period are needed in order to detect relatively rare events with small biases if the conventional normalisation scheme is used.

Second, we assess the effect of autocorrelation on the biases in the normalised tails. Autocorrelation is a feature frequently present in climatic time series (Zwiers and von Storch, 1995), and hence should be accounted for in statistical analyses. We simulate time series from an AR(1) process as

$$X_{AR1}(t) = \alpha X_{AR1}(t-1) + Z(t), \tag{B.11}$$

with white noise innovations $Z \sim \mathcal{N}(0, \tau^2)$. The model's parameter α determines the strength of the autocorrelation and is varied in the range $0 \leq \alpha \leq 0.9$. The overestimation of extremes strongly increases for autocorrelated data, which urges for caution in using a normalisation scheme in such time series. The reason for the stronger overestimation compared to the standard normalisation procedure is three-fold: Firstly, the variance of the sample mean of autocorrelated data (Zieba, 2010) is larger as compared to Eq. B.2:

$$\hat{\mu}_{X_{AR1},ref} = \mathcal{N}(0, [n + 2\sum_{k=1}^{n-1}(n-k)\rho_k]\frac{\sigma^2}{n^2}), \tag{B.12}$$

where ρ_k denotes the autocorrelation coefficient of the AR(1) model.

Secondly, the standard variance estimator (Eq. B.3) is biased for autocorrelated data (Bayley and Hammersley, 1946). The construction of an unbiased variance estimator is possible (Zieba, 2010), but requires the autocorrelation structure to be known exactly. Thirdly, the normalised distributions follow Student's t-distribution (as above), if the variance and mean estimates are derived from an independent sample. Hence, these three issues are causing the drastically increasing biases seen in Figure B2b for autocorrelated data.

Furthermore, trends and changing variance are common features in climatic time series (Ji et al., 2014; Huntingford et al., 2013; Screen, 2014). We test empirically how changes in the mean or variance in the independent period are changing the detection biases in normalised extremes. To do so, we add various offsets in the range $-1 \leq \delta \leq +2[\sigma]$. Similarly, we change the variance in the out-of-base period to $0.5 \leq \sigma \leq 2$. Subsequently, the relative difference between the standard normalisation scheme and the true number of extremes is calculated (Figure B2c). Our Monte-Carlo simulations reveal that normalisation biases (as discussed in the main text) are not constant under changes of the mean and variance of the time series. Although an analytical treatment is possible (see Section B.4), this empirical exercise allows to illustrate the sensitivity of the biases to both sign and magnitude of trends and changes in variance. Positive changes in the mean or variance are reducing the observed biases in the upper tail of the distribution, because any positive σ extreme would 'shift' towards the center of the distribution in this case. However, negative trends or changes in variance would induce the opposite effect and lead to a drastic overestimation in the upper tail. These results are equally applicable to the lower tail if the sign of the trend is reversed. We conclude that any assessment of extremes or the tails of normalised climatic data across different spatial or temporal domains needs to take potential non-stationarities into account.

B.4. Normalisation bias in non-stationary and independent time series

This section is motivated by the fact that normalisation-induced biases are sensitive to trends or changes in variance (see Section B.3). Here, we outline a correction method that takes such non-stationarities into account. Consider any random variable $X_{orig} \sim \mathcal{N}(\mu_{ref}, \sigma^2_{ref})$, from which $\hat{\mu}_{ref}$ and $\hat{\sigma}^2_{ref}$ are esti-

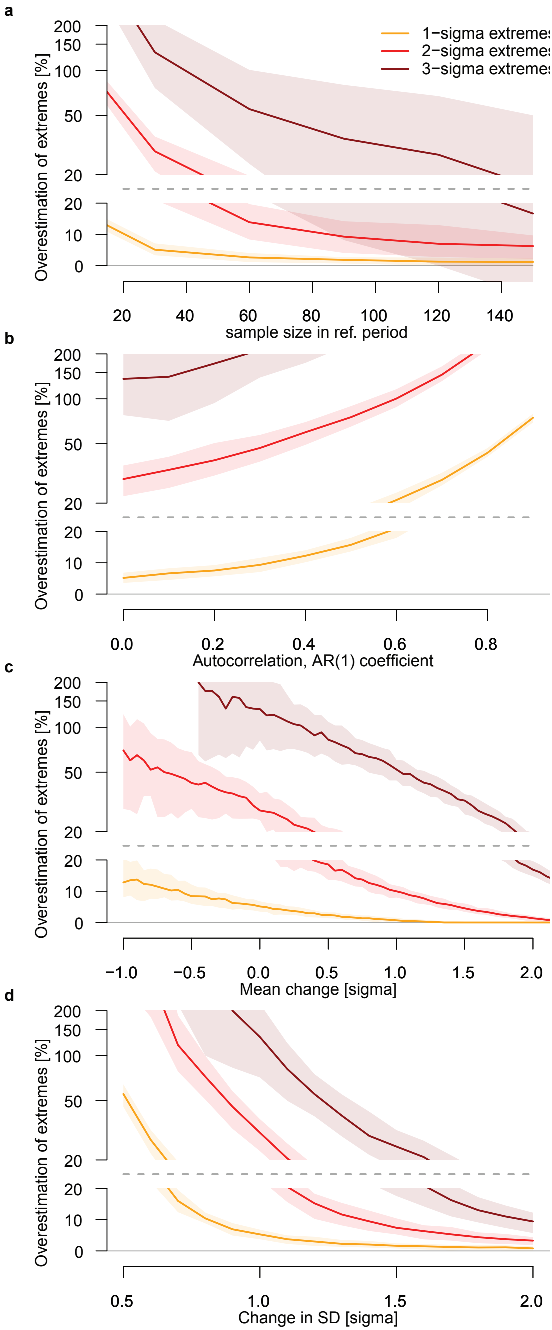


FIGURE B2.: Full caption is displayed on the next page.

FIGURE B2.: (continued) Sensitivity tests of normalisation-induced biases in the tails. Monte-Carlo type simulations are conducted to show how the biases in the upper tail are affected by a) varying sample size, b) different degrees of autocorrelation, c,d) trends and changing variance in the out-of-base period, respectively.

mated. Assume that at any time t outside the reference period the mean changes to $\mu_{t,obase} = \mu_{ref} + \delta_t$ and the standard deviation changes to $\sigma_{obase} = \lambda \cdot \sigma_{ref}$.

Non-stationarity in the out-of-base period would change the Gaussian distribution to

$$X_t \sim \mathcal{N}(\mu + \delta_t, \lambda^2 \cdot \sigma^2). \tag{B.13}$$

The generation of anomalies for Gaussian data is given in Eq. B.4 and the sample means follow Eq. B.2. Put together, this yields a distribution of anomalies across grid cells given by

$$X_{anom,t} \sim \mathcal{N}(\delta_t, \sigma^2(\lambda^2 + \frac{1}{n_{ref}}). \tag{B.14}$$

Accordingly, and similar to Eq. B.5, the spatial aggregation for the detection of extremes in the tails would result in a broader (but qualitatively unchanged) distribution. A search for non-adjusted σ extremes becomes hence inadequate.

However, the subsequent standardisation of non-stationary and independent time series is more important for biases in the tails. A generalisation of Student's t-distribution is the non-central t-distribution (Johnson et al., 1995), which is skewed and results from Eq. B.6, if $X_{anom,t}$ is replaced by a random Gaussian variable with non-zero mean (Von Storch and Zwiers, 2001). Hence, a standardisation of non-stationary Gaussian time series based on Eq. B.6 yields a spatial distribution of

$$\frac{X_{anom,t}}{\hat{\sigma}_{ref}} = \sqrt{\lambda^2 + \frac{1}{n_{ref}}} \cdot \frac{\left[\frac{X_{anom}-\delta_t}{\sqrt{\lambda^2+\frac{1}{n_{ref}}}} + \frac{\delta_t}{\sqrt{\lambda^2+\frac{1}{n_{ref}}}}\right]}{\hat{\sigma}_{ref}} \tag{B.15}$$

$$\Rightarrow z = \frac{X_{anom,t}}{\hat{\sigma}_{ref}} \sim \sqrt{\lambda^2 + \frac{1}{n_{ref}}} \cdot t'(\nu = n-1, ncp = \frac{\delta_t}{\sqrt{\lambda^2 + \frac{1}{n_{ref}}}}) \quad . \tag{B.16}$$

This can be seen as a centering and scaling of the enumerator in Eq. B.15 to yield a unit normal variable and an additive non-centrality-parameter. Hence, the division by the estimates of the standard deviation $\hat{\sigma}_{ref}$ yields a scaled version of the non-central t-distribution (Eq. B.16), implying $k = n_{ref} - 1$ degrees of freedom. Therefore, an analytical correction similar to Section B.2 can be constructed if the change in location and scale outside the reference period can be estimated (see also Figure 2.2). However, since estimates of trends or variance changes are made on relatively short time series, and because these are not independent from the estimated mean or variability, some minor biases remain (Figure 2.2). These biases are negligible if only the mean has changed, and they are much smaller than biases in the tails induced by an uncorrected normalisation procedure if variance changes are estimated as well. Nevertheless, we argue for some caution if very rare events are to be detected based on the application of a normalisation transformation.

B.5. Subtraction of trend components before computing standard deviation estimates

Several previous papers have used detrending procedures before estimating the standard deviation in a reference period (Coumou and Robinson, 2013; Huntingford et al., 2013). This data preprocessing step is assumed to avoid an overestimation of variability due to potential trends in time series in the (arbitrarily chosen) reference period. Others have used the period 1951-1980 as the reference, because this period is widely assumed to be associated with largely stationary temperatures (Hansen et al., 2012). The removal of trends before computing the

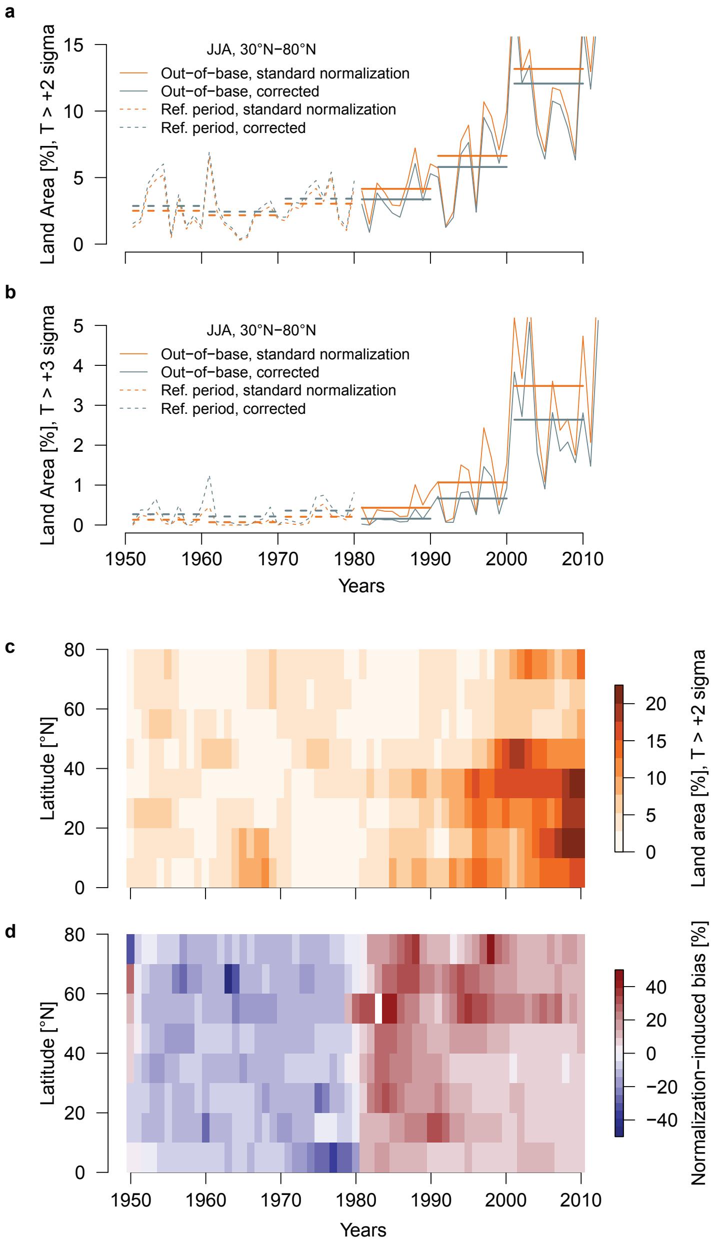


FIGURE B3.: Full caption is displayed on the next page.

FIGURE B3.: (continued) Increase in normalised hot temperature extremes in a spatio-temporal dataset (20th Century Reanalysis). a,b) Time series of fraction of extratropical Northern hemisphere land area covered by positive monthly 2σ (a) and 3σ (b) extremes in summer (reference period: 1951-1980). Horizontal lines indicate decadal averages for the conventional normalisation procedure (light blue) and our proposed correction (orange). c) Zonal evolution of fraction of land area covered by monthly positive 2σ extremes in Northern hemisphere summer. d) Zonal evolution of relative biases induced by the conventional normalisation approach. In all panels, the time series have been detrended before estimating the estimate of the standard deviation in the reference period (1951-1980).

standard deviation of each time series reveals only very minor changes both in terms of the overall increase in extremes and the preprocessing-induced biases. We estimate trends in each time series using Singular Spectrum Analysis as described in the Methodology section of the main paper, but other methodologies are likewise applicable. Next, we standardise each time series with the standard deviation estimates computed from detrended series and reproduce Figure 2.3 from the main paper (Figure B3).

To test the sensitivity of the biases and extremes to the choice of reference period, we repeat the previous analysis by normalizing the data based on mean and detrended SD estimates calculated for 1921-1950 (Figure 2.4). Although the choice of reference period influences the absolute number of σ extremes (because 1951-1980 had been warmer than 1921-1950), the biases that are induced by the normalisation procedure are still in a similar magnitude (Figure 2.4).

B.6. Asymmetry in temperature distributions

Another important question to test is whether recent estimates of asymmetry (Kodra and Ganguly, 2014) in seasonal extreme value distributions might be affected by subtracting a 'historical climatology', estimated from each time series. For this purpose, we follow the methodology of an earlier study (Kodra and Ganguly, 2014) but with i.i.d. Gaussian variables:

- We generate 60 seasons with each 90 days in $k = 10,000$ time series (that is, in analogy to spatial replicates)

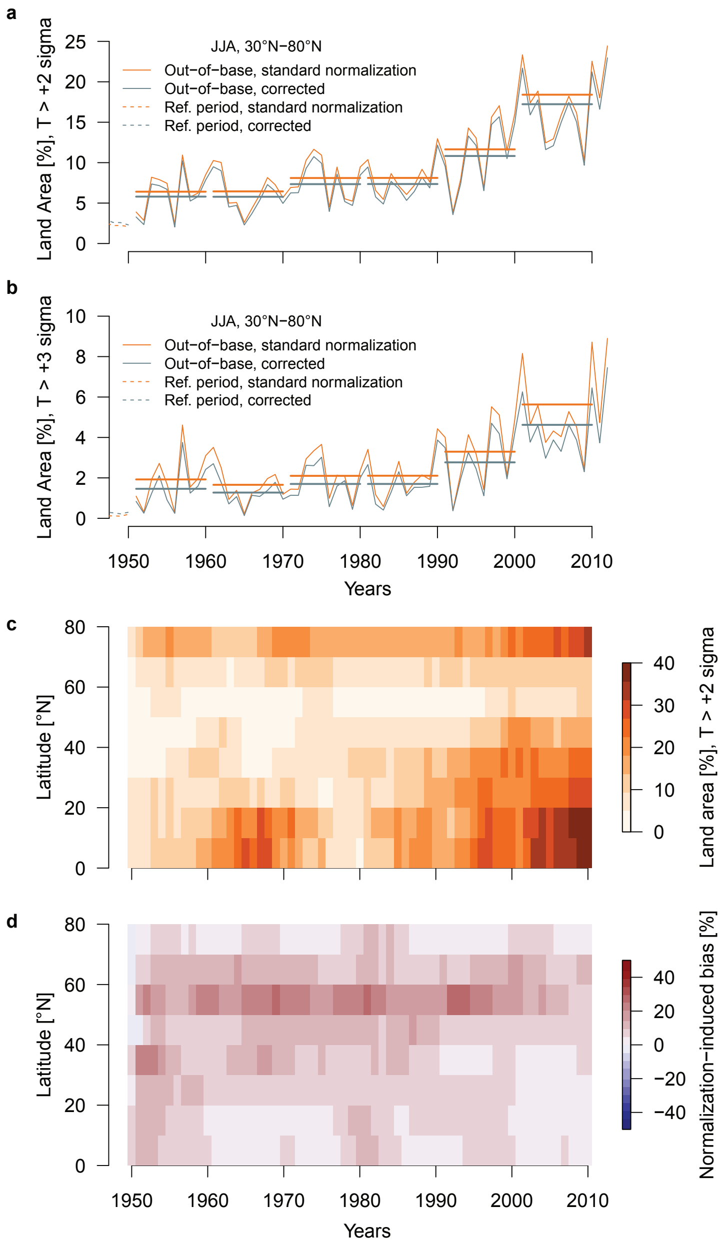


FIGURE B4.: Full caption is displayed on the next page.

FIGURE B4.: (continued) Increase in normalised hot temperature extremes in a spatio-temporal dataset (20th Century Reanalysis). a,b) Time series of fraction of extratropical Northern hemisphere land area covered by positive monthly 2σ (a) and 3σ (b) extremes in summer (reference period: 1921-1950). Horizontal lines indicate decadal averages for the conventional normalisation procedure (light blue) and our proposed correction (orange). c) Zonal evolution of fraction of land area covered by monthly positive 2σ extremes in Northern hemisphere summer. d) Zonal evolution of relative biases induced by the conventional normalisation approach. In all panels, the time series have been detrended before estimating the estimate of the standard deviation in the reference period (1921-1950).

- For each season, we only retain the maximum value. This procedure yields a distribution that can be approximated by a Weibull type extreme value distribution (Coles et al., 2001)
- Now, each time series is split into a historical and future period (first and second half of the time series, respectively)
- Following Kodra and Ganguly (2014), we compute the mean of the 'historical' period and subtract it from each times series.
- Subsequently, percentiles of the future and historical period are computed across all time series, and percentile-wise differences between the future and historical period are analysed (Figure B5a)
- We compare the so-derived percentile-wise changes to simply generating the differences between future and historical percentiles without the previous transformation (Figure B5a)

As shown in Section B.2, this procedure invariably leads to an inflation (reduction) of the variance in the surrogate 'future' ('historical') period. Hence, the upper tail of the 'future' extreme value distribution has increased, whereas the lower tail has decreased relative to untransformed changes (see red and grey lines in Figure B5a). However, since extreme value distributions are skewed, the change in variability also explains the observation of increased assymetry, if the changes in both tails are compared (Figure B5b). This increased asymmetry is not observed if the analysis is conducted without subtracting historical means (grey line in Figure B5b). These results are shown for extreme value distributions generated by retaining the highest value in each season, but would apply equally if only seasonal minima were retained (but with reversed changes in asymmetry).

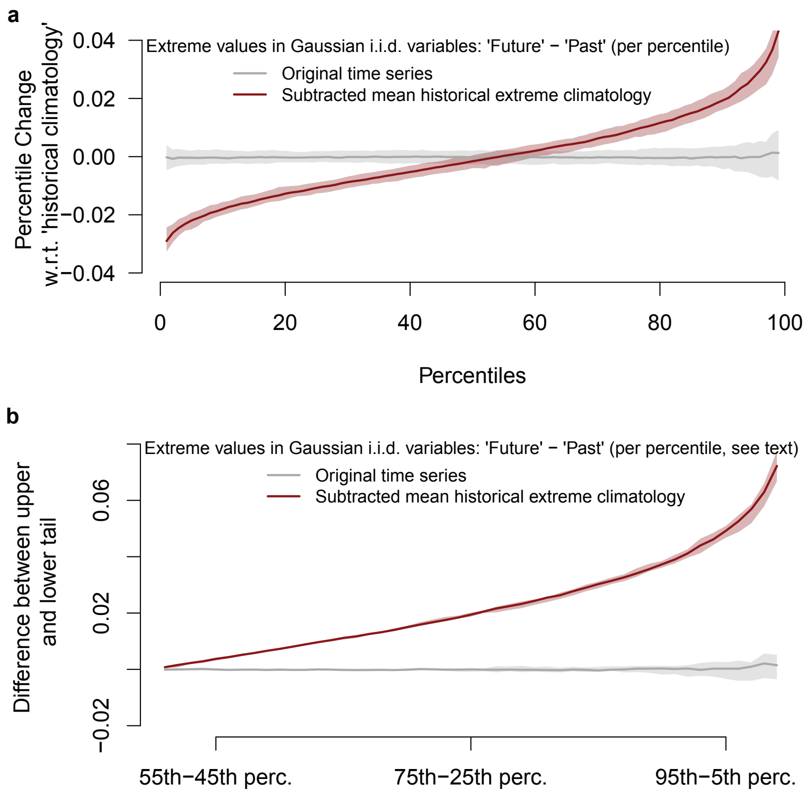


FIGURE B5.: Spurious increase in asymmetry due to data pre-processing. a) Percentile-wise changes across a large number of time series, expressed as the difference between a 'historical' and 'future' period. Induction of asymmetry occurs only if a historical mean climatology is estimated and subtracted from each time series. b) Like above, but differences in symmetric percentiles between the upper and lower tail, further illustrating induced asymmetry in the upper tail. Results are likewise applicable to the lower tail (with reversed asymmetry), if extreme value distribution are generated from minimum values.

C. Supplementary Material for Chapter 3

C.1. Analytical approximation of the expected value for the normalisation-induced bias

Assumptions and Notation:

- Assume independent and identically distributed (i.e., stationary) variables $X_{t,i}$ with mean given by $\mathbf{E}(X) = \mu$ and variance $\mathbf{Var}(X) = \sigma^2$. Let the subscripts t and i denote time and grid cell index, respectively. Note that in real-world applications, the biases could be estimated analytically by allowing for different sample means and variances across space.

- Let t_{oob} be an arbitrary time step in the 'out-of-base' (independent) period, and t_{ref} as an arbitrary time step inside the reference period. Let n_{ref} denote the length of the reference period.

- Let $\Delta_{bias} = \mathbf{E}(\frac{X_{t_{oob},i}}{\hat{\mu}_{\text{ref},i}}) - 1$ denote the relative change induced by normalisation by the mean of an independent reference period (i.e., 'normalisation bias', $X_{t_{oob},i}$ is not contained in $\mu_{\text{ref},i}$).

Our objective is to find an analytical approximation of the expected value for the artificially induced relative change (Δ_{bias}) by dividing a random variable $X_{t_{oob},i}$ as defined above by a sample mean estimated from different samples ('reference samples') drawn from the same distribution ($\hat{\mu}_{\text{ref,i}} = \frac{1}{n} \sum_{t_{\text{ref}}=1}^{n_{\text{ref}}} X_{t_{\text{ref}},i}$, where $\mathbf{E}(\hat{\mu}_{\text{ref,i}}) = \mu$), i.e.

$$\Delta_{bias} = \mathbf{E}(\frac{X_{t_{oob},i}}{\hat{\mu}_{\text{ref},i}}) - 1 \approx f(\mu, \sigma, n_{\text{ref}}). \tag{C.1}$$

Clearly, for large n_{ref} this quantity should go to 0. Because $X_{t,i}$ and $\hat{\mu}_{\text{ref},i}$ are independent, we can write,

$$\Delta_{bias} = \mathbf{E}(X_{t,i})\mathbf{E}(\frac{1}{\hat{\mu}_{\text{ref},i}}) - 1 = \mu\mathbf{E}(\frac{1}{\hat{\mu}_{\text{ref},i}}) - 1. \tag{C.2}$$

If we subsitute $\hat{\mu}_{\text{ref},i} = \mu(1 + \epsilon_{\text{ref},i})$, where $\mathbf{E}(\epsilon_i) = 0$, $\mathbf{Var}(\epsilon_i) = \frac{\sigma^2}{\mu^2 n_{\text{ref}}}$ (because $\epsilon_{ref,i} = \frac{\hat{\mu}_{ref,i}}{\mu} - 1$, and $\mathbf{E}(\hat{\mu}_{ref,i}) = \mu$ and $\mathbf{Var}(\hat{\mu}_{ref,i}) = \frac{\sigma^2}{n_{\text{ref}}}$), and the subscript ref has been dropped from ϵ_i for convenience, we get

$$\Delta_{bias} = \mu\mathbf{E}(\frac{1}{\mu(1+\epsilon_i)}) - 1 = \mathbf{E}(\frac{1}{1+\epsilon_i}) - 1. \tag{C.3}$$

A Taylor expansion around the function $g(x) = \frac{1}{1+x}$ at $x = 0$ yields

$$g(x) = \frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 - x^5 + ... \tag{C.4}$$

We will see below that the convergence criterion $\epsilon_i < |1|$ of the Taylor series is met in practically relevant cases, but it should be noted that convergence cannot be ensured in all theoretically conceivable cases. Using Taylor expansion, Δ_{bias} can be approximated, making use of the linearity of the expectation operator $\mathbf{E}()$ and of the fact that $\mathbf{E}(\epsilon_i) = 0$ and $\mathbf{E}(\epsilon_i^2) = \mathbf{Var}(\epsilon_i) = \frac{\sigma^2}{\mu^2 n_{\text{ref}}}$ by definition,

$$\Delta_{bias} = \mathbf{E}(\frac{1}{1+\epsilon_i}) - 1 \tag{C.5a}$$

$$= \mathbf{E}(1 - \epsilon_i + \epsilon_i^2 - \epsilon_i^3 + \epsilon_i^4 - \epsilon_i^5 + ...) - 1 \tag{C.5b}$$

$$= \frac{\sigma^2}{\mu^2 n_{\text{ref}}} - \mathbf{E}(\epsilon_i^3) + \mathbf{E}(\epsilon_i^4) - \mathbf{E}(\epsilon_i^5) + ... \tag{C.5c}$$

This expression yields a sum over the central moments of the distribution of ϵ_i's. For a symmetric probability distribution (recall that ϵ_i denote the deviations of the sample means in a reference period around the underlying true mean), $E(\epsilon_i^k) = 0$, where k is any uneven exponent $k \in \mathbb{N}$. Eq. C.5a then reduces to

$$\Delta_{bias} = \frac{\sigma^2}{\mu^2 n_{\text{ref}}} + \mathbf{E}(\epsilon_i^4) + \mathbf{E}(\epsilon_i^6) + ... \tag{C.6}$$

As long as $\epsilon_i < |1|$ is fulfilled, the quadratic term dominates both Eq. C.5a and Eq. C.6. The present analytical approximation (both Eq. C.5a and Eq. C.6) provides the important insights that

1) normalisation with a ‘reference period sample mean’ leads to an artificial increase of spatial averages in the out-of-base period, i.e. the bias is always positive in the out-of-base period, $\Delta_{bias} > 0$, and

2) that $\Delta_{bias} \propto (\frac{\sigma}{\mu}\frac{1}{\sqrt{n_{\text{ref}}}})^2$, i.e. the square of the coefficient of variation in the distribution of sample means (i.e., $c_v[\hat{\mu}_{\text{ref,i}}] = \frac{\sigma}{\mu\sqrt{n_{\text{ref}}}}$).

For any fixed n_{ref}, the amplitude of the normalisation-induced biases only depends on the square of the ratio $\frac{\sigma}{\mu}$. We verify below numerically that this approximation works well for random variables $X_{t,i}$ drawn from

i. a Gaussian distribution,

ii. a Generalised Extreme Value distribution with two different choices for the shape parameter ($\xi = 0$, ‘Gumbel distribution’, and $\xi \neq 0$).

C.1.1. Gaussian distribution

Assume $X_{t,i} \sim \mathcal{N}(\mu, \sigma^2)$, the distribution of the sample mean deviations from the true mean will follow $\epsilon_i \sim \mathcal{N}(0, \frac{\sigma^2}{\mu^2 n_{\text{ref}}})$. If we substitute with $\epsilon_i = \frac{\sigma}{\mu}\frac{1}{\sqrt{n_{\text{ref}}}}Y$, where $Y \sim \mathcal{N}(0,1)$ in Eq. C.6, the above expression reduces to

$$\Delta_{bias} = \frac{\sigma^2}{\mu^2 n_{\text{ref}}} + (\frac{\sigma}{\mu}\frac{1}{\sqrt{n_{\text{ref}}}})^4 \mathbf{E}(Y^4) + (\frac{\sigma}{\mu}\frac{1}{\sqrt{n_{\text{ref}}}})^6 \mathbf{E}(Y^6) + ... \quad \text{(C.7)}$$

Because higher-order moments of a standard normal distributed random variable are well-known and can be derived analytically (Johnson et al., 1994, i.e., $\mathbf{E}(Y^4) = 3$, $\mathbf{E}(Y^6) = 15$), an analytical expression of the normalisation-induced bias becomes straightforward:

$$\Delta_{bias} \approx \frac{\sigma^2}{\mu^2 n_{\text{ref}}} + 3(\frac{\sigma}{\mu}\frac{1}{\sqrt{n_{\text{ref}}}})^4 + 15(\frac{\sigma}{\mu}\frac{1}{\sqrt{n_{\text{ref}}}})^6. \quad \text{(C.8)}$$

A comparison of Eq. C.8 (i.e. the first three terms in the Taylor approximation) to numerical simulations shows that the analytical approximation works well

(Figure C1a). Furthermore, the estimation of mean and standard deviation from the empirical time series to calculate the expected value for the biases is unbiased and show surprisingly little variation (Figure C1b) even for a relatively small number of grid cells, where random variation in stationary time series becomes considerable (Figure C1b).

However, one important caveat is that Eq. C.3 and the subsequent approximation only works as long as $\epsilon_i < |1|$ is fulfilled. How likely is a violation of this criterion? Numerical simulations for $n_{\mathrm{ref}} = 30$ appear to be very stable for any $\frac{\mu}{\sigma} > 0.8$ in the $X_{t,i}$'s, i.e. corresponding roughly to a $C_v[\hat{\mu}_{\mathrm{ref},i}] \approx 0.2$. For such a choice of C_v the chance of $|\epsilon_i| \geq 1$ corresponds to a -5σ event with a probability of roughly 1 to 3.5 million. Given that the observed $\frac{\mu}{\sigma}$ ratios are considerably larger than the values tested here even in the driest regions of the world, we conclude that the approximation can be used for the vast majority, if not all, practical considerations.

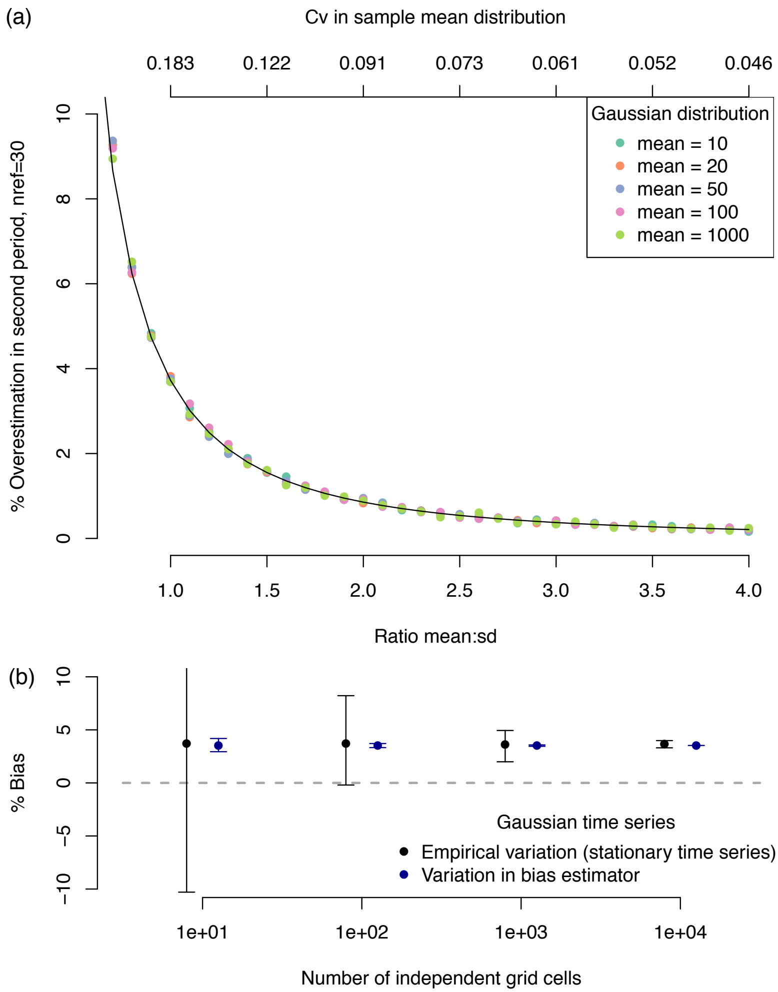


FIGURE C1.: a) Ratio of mean to sd vs. normalisation-induced bias in a Gaussian distribution for numerical simulations with various mean values (dots), and the derived analytical approximation (black line). The reference period length is taken as $n_{ref} = 30$, and numerical simulations are conducted with $n = 10^5$ grid cells with each 60 time steps. b) Analytical estimates of biases as calculated from sample mean and sample standard deviation following Eq. 3.1 in the main text (dark blue) for a given number of independent grid cells ($\frac{\mu}{\sigma} = 1$, $n_{\text{ref}} = 30$). For comparison, the magnitude of random changes in stationary time series (i.e. empirical variation in the quantity Δ_{bias}, following Eq. C.1) with $n_{\text{ref}} = 30$ and $n_{\text{obase}} = 30$ is shown in black. Error bars indicate the 5th to 95th percentile in repeated numerical simulations.

C.1.2. Generalised extreme value distribution

We investigate whether in Eq. C.5a the higher-order terms in the Taylor approximation can be ignored in practical applications, where an assumption of Gaussianity might not hold. Here, we test this for the Generalised Extreme Value distribution as an appropriate model for annual maxima as investigated in the main manuscript with two different choices for the distribution's shape parameter (ξ).

i. Gumbel distribution We first assume, in analogy to the paragraph above, independent and identically distributed (i.e., stationary) random variables drawn from a Generalised Extreme Value distribution with zero shape parameter ('Gumbel distribution', $X_{t,i} \sim GEV(\mu', \sigma', \xi = 0)$, where μ', σ' and $\xi = 0$ denote the GEV's location, scale and shape parameter, respectively, see e.g. Johnson et al., 1995). The expected values for mean (μ) and variance (σ^2) of a GEV are given by $\mu = \mu' + \sigma'\gamma$, where γ denotes Euler's constant.

Folllowing Eq. C.5a, we can readily derive an analytical expression for the expected value of the normalisation-induced bias, i.e.

$$\Delta_{bias} = \frac{\sigma^2}{\mu^2 n_{\text{ref}}} - \mathbf{E}(\epsilon_i^3) + \mathbf{E}(\epsilon_i^4) - \mathbf{E}(\epsilon_i^5) + ... \tag{C.9}$$

$$= (\frac{\pi}{\sqrt{6 n_{\text{ref}}}(\frac{\mu'}{\sigma'} + \gamma)})^2 - \mathbf{E}(\epsilon_i^3) + \mathbf{E}(\epsilon_i^4) - \mathbf{E}(\epsilon_i^5) + ... \tag{C.10}$$

Here, we note again that the quadratic term dominates the expression. If we make the simplifying assumption that the sample means $\hat{\mu}_{\text{ref},i}$ for $n_{\text{ref}} = 30$ follow (approximately) a Gaussian distribution (the assumption is only needed for the higher order terms of the Taylor expansion), we can derive an analytical approximation for the normalisation-induced bias by insertion and in analogy to above, i.e.

$$\Delta_{bias} \approx (\frac{\pi}{\sqrt{6 n_{\text{ref}}}(\frac{\mu'}{\sigma'} + \gamma)})^2 + (\frac{\sigma}{\mu}\frac{1}{\sqrt{n_{\text{ref}}}})^4 \mathbf{E}(Y^4) + ... \tag{C.11}$$

$$\approx (\frac{\pi}{\sqrt{6 n_{\text{ref}}}(\frac{\mu'}{\sigma'} + \gamma)})^2 + 3(\frac{\pi}{\sqrt{6 n_{\text{ref}}}(\frac{\mu'}{\sigma'} + \gamma)})^4. \tag{C.12}$$

Hence, we find that the magnitude of the bias estimates is proportional to the ratio of scale to location parameter ($\frac{\sigma'}{\mu'}$) for any fixed reference period length (but also the proportionality to the square of the ratio of standard deviation to mean remains, i.e. Eq. 3.1 (or Eq. C.11)). The analytical approximation can be verified by numerical simulation using GEV-distributed random variables, and is found to fit the data very well (Figure C2a). Furthermore, an estimator of the expected value of the biases by only estimating the mean and standard deviation of empirical time series (i.e., using the first term in the Taylor approximation) can be derived easily and is found to work reliable also for a small number of independent grid cells (Figure C2c).

ii. GEV distribution with $\xi \neq 0$ Here, we test whether the analytical argument from above can be extended to Generalised Extreme Value distributions with $\xi \neq 0$. In practical applications of the GEV to observed maximum precipitation, a shape parameter of $\xi \approx 0.2$ is often found (Van den Brink and Können, 2011), therefore we test here for $X_{t,i} \sim \mathrm{GEV}(\mu', \sigma', \xi = 0.2)$. The expected values for mean (μ) and variance (σ^2) of a GEV, when $0 > \epsilon < 1$, are given by $\mu = \mu' + \sigma' \frac{\Gamma(1-\xi)-1}{\xi}$ and $\sigma^2 = (\sigma')^2 \dfrac{(g_2 - g_1^2)}{\xi}$, where $g_k = \Gamma(1 - k\xi)$, $k = 1, 2$, and $\Gamma(t)$ is the gamma function (Johnson et al., 1995).

Hence, the (dominant) quadratic term in the Taylor approximation in Eq. C.5a reads,

$$\Delta_{bias} \approx \frac{(g_2 - g_1^2)}{n_{\mathrm{ref}} \xi [\frac{\mu'}{\sigma'} + \dfrac{\Gamma(1-\xi)-1}{\xi}]^2}. \tag{C.13}$$

The approximation works again very well in numerical simulations (Figure C2b), and shows that if $\xi \neq 0$, there is a dependency on ξ, n_{ref}, and again the ratio of $\frac{\sigma'}{\mu'}$, which determine the magnitude of the normalisation-induced bias. Please note that the approximation works similarly well for random variables drawn from a GEV-distribution with negative shape parameter ($\xi = -0.2$, not shown).

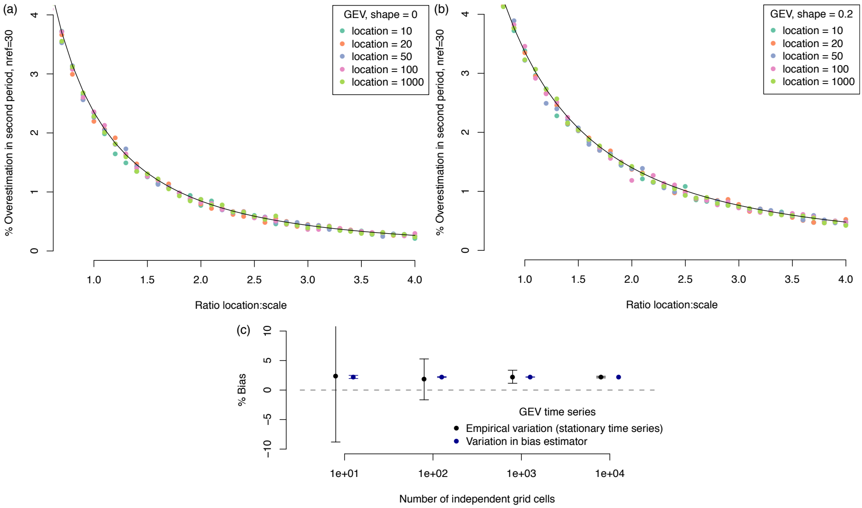


FIGURE C2.: a) Ratio of location to scale parameter vs. normalisation-induced bias in a Generalised Extreme Value distribution for the analytical approximation (black line) and numerical simulations with various location parameter values (dots), with a) zero shape parameter, and b) $\xi = 0.2$. Reference period length is taken as $n_{\mathrm{ref}} = 30$, and numerical simulations are conducted with $n = 10^5$ grid cells with each 60 time steps. c) Analytical estimates of biases as calculated from sample mean and sample standard deviation following Eq. 3.1 in the main text (dark blue) for a given number of independent grid cells drawn from a GEV distribution ($\frac{\mu'}{\sigma'} = 1$, $\xi = 0$, $n_{\mathrm{ref}} = 30$). For comparison, the magnitude of random changes in stationary time series (i.e. empirical variation in the quantity Δ_{bias}, following Eq. C.1) with $n_{\mathrm{ref}} = 30$ and $n_{\mathrm{obase}} = 30$ is shown in black. Error bars indicate the 5th to 95th percentile in repeated numerical simulations.

C.1.3. Short remark on non-stationarity in the out-of-base period

Many real-world precipitation time series show non-stationarities due to climatic variations (O'Gorman, 2015) that are typically diagnosed as relative changes in the precipitation amount. Hence, we can ask whether and how any 'real change in the expected value' outside the reference period can be disentangled from the normalisation-induced bias. Given the analytical approximation above, we can show that the highlighted normalisation-induced bias scales non-stationarities in the out-of-base period in a multiplicative way:

Let c denote any change between the reference period expected value and some future period (e.g. assume one is interested in global or latitudinal changes in a past and future climatic period), i.e. such that $\mathbf{E}(X_{t_{\mathrm{ref}},i}) = c\mathbf{E}(X_{t_{\mathrm{oob}},i})$, then the bias ($\Delta_{\mathrm{bias}}$, after accounting for the 'real change') would simply scale with the relative change (Δ denotes the total apparent change):

$$\Delta = c\mathbf{E}(\frac{X_{t,i}}{\hat{\mu}_{\mathrm{ref},i}}) - 1 \tag{C.14}$$

$$= c\mathbf{E}(\frac{1}{1+\epsilon_i}) - 1 \tag{C.15}$$

$$= \underbrace{c-1}_{\text{true change}} + \underbrace{c[\frac{\sigma^2}{\mu^2 n_{\mathrm{ref}}} - \mathbf{E}(\epsilon_i^3) + \mathbf{E}(\epsilon_i^4) - \mathbf{E}(\epsilon_i^5) + ...]}_{\Delta_{\mathrm{bias}}} \tag{C.16}$$

From Eq. C.16, it is straightforward to see that for any multiplicative changes in the expected value of the out-of-base variables, the normalisation-inudced bias scales with the change. Hence, this expression implies that to detect the 'true change c' between two periods, the normalisation-induced bias has to be accounted for, i.e.

$$c = \frac{\Delta + 1}{1 + \Delta_{\mathrm{bias}}}. \tag{C.17}$$

C.2. Comparison between aridity-based and precipitation-based definition of dryness

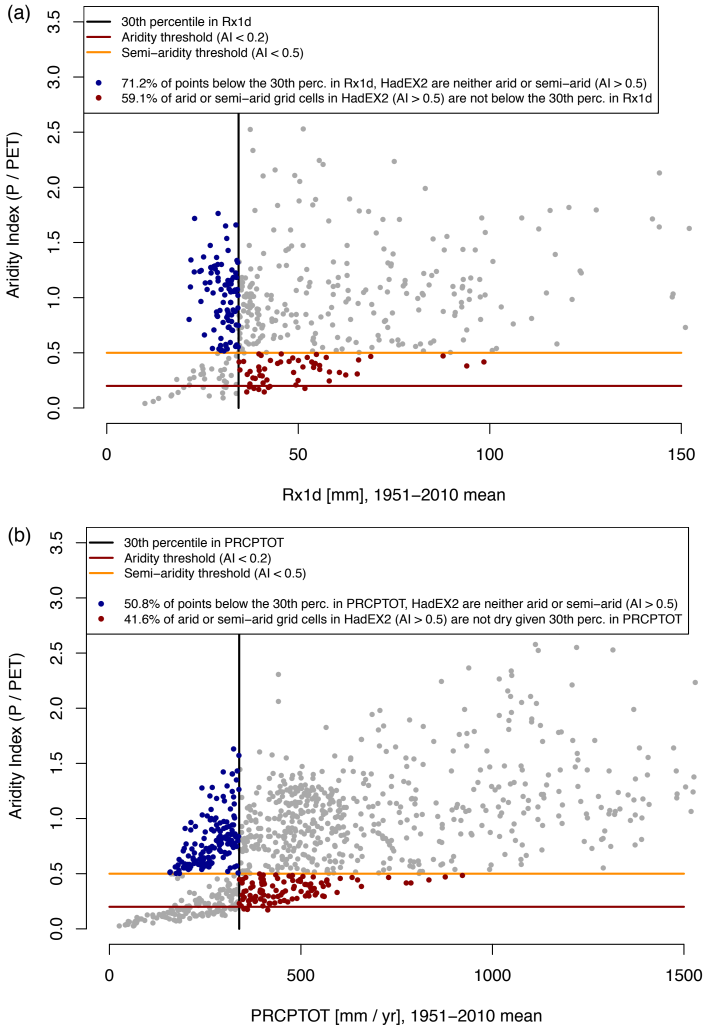


FIGURE C3.: Relationship between annual-maximum daily rainfall (Rx1d from HadEX2-GHCNDEX merged dataset) and aridity (a), and precipitation totals (PRCPTOT from HadEX2-GHCNDEX merged dataset) and aridity (b). Potential evapotranspiration is taken from the CRU-TS3.23 dataset (Harris et al., 2014)

D. Supplementary Material for Chapter 6

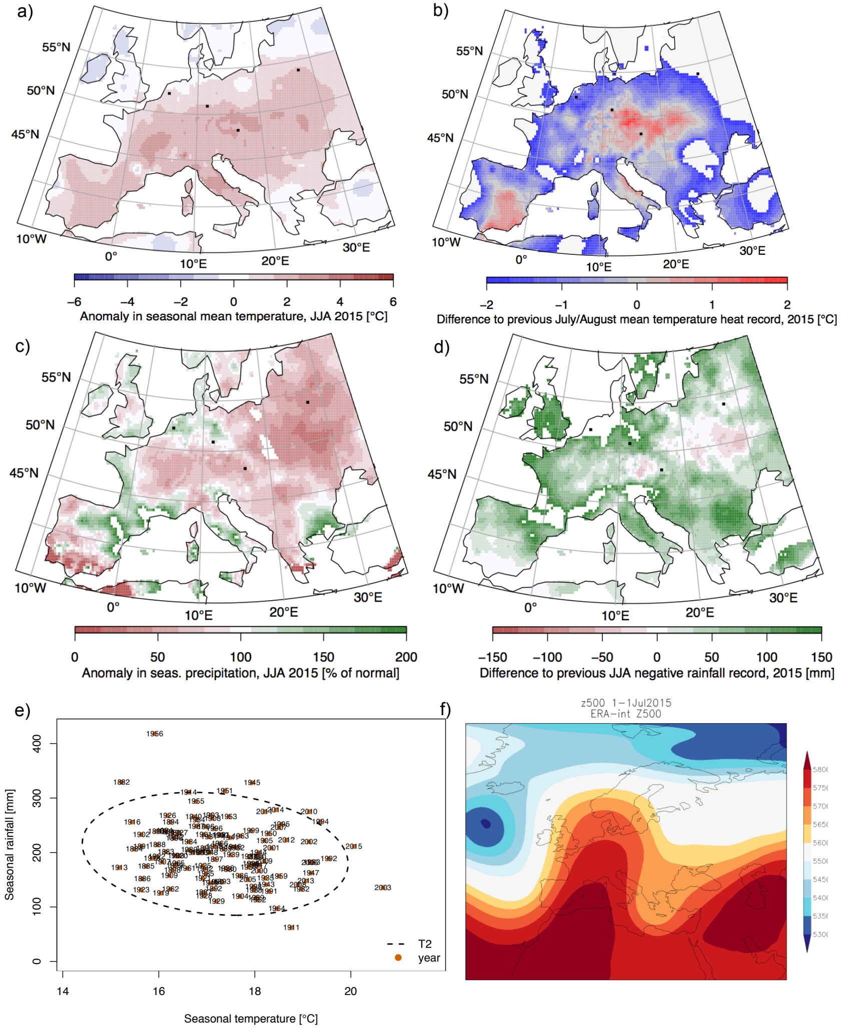


FIGURE D1.: a, c) Anomalies in European summer 2015 seasonal mean temperature and precipitation, respectively. b,d) Differences to previous July/August records in mean temperature and low rainfall, respectively, relative to 1950-2014 as in the main text. e) Seasonal temperature versus seasonal rainfall in Vienna, Austria. The ellipse denotes a quantile of 5% multivariate extremes computed by Hotelling's T^2 control chart (Santos-Fernández, 2012) using robust mean and covariance estimates (Rousseeuw and Hubert, 2011). f) 500 hPa Geopotential heights over Europe on July 1st 2015.

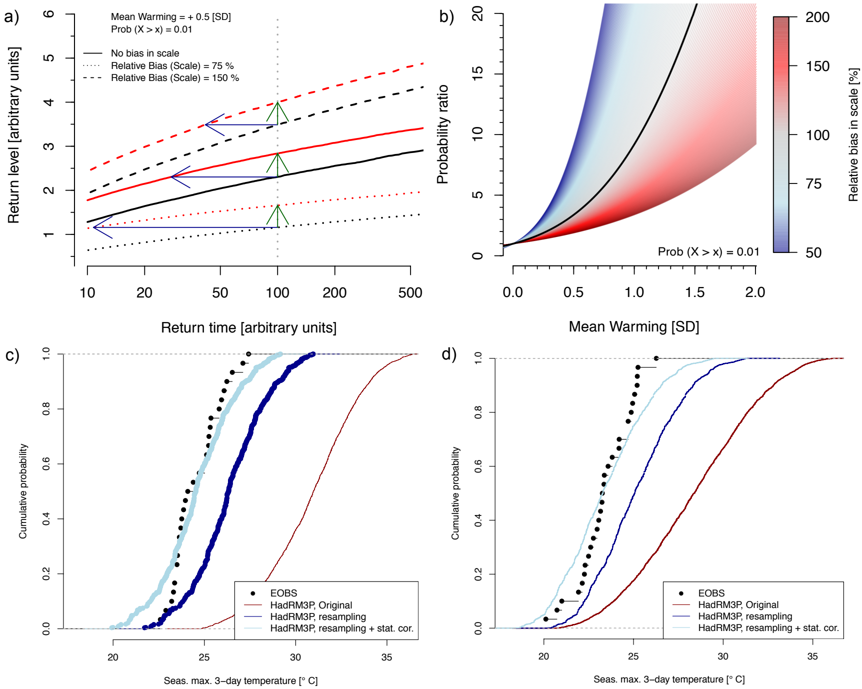


FIGURE D2.: Conceptual example of the relevance of biases for attribution statements (a-b) and the effect of resampling on cumulative distribution function in Vienna (c) and Jena (d). a) Return time plots for Gaussian distributed random variables that differ only in their standard deviation. A mean shift of 0.5σ results in an equal change in return levels, but leads to substantially different changes in return times across the three sets. b) Probability ratio for different levels of mean warming and biases in σ. c, d) Cumulative distribution function of Tair3d,max in Vienna (c) and Jena (d) in the "original", "resampled", and "resampled+mean-adjustment" simulations.

E. Supplementary Material for Chapter 9

E.1. Supplementary Methods

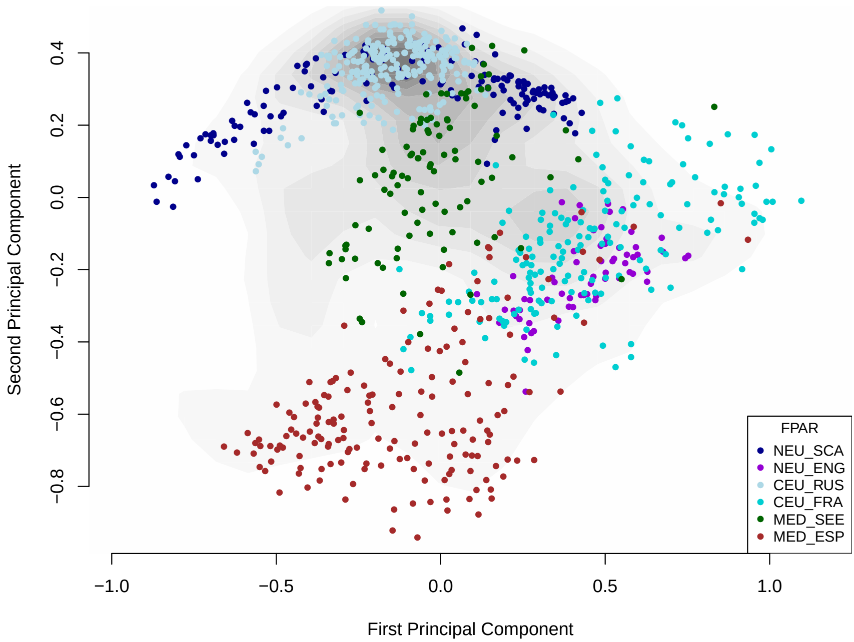


FIGURE E1.: Principal component analysis of FPAR mean season cycles over all European grid cells. Grey shading indicates the projection of all European grid cells onto the first and second principal component, all grid cells within the analysed regions are indicated by the individual colours. Selected European regions broadly cover the spectrum of variability in European seasonal cycles of vegetation activity as indicated by FPAR.

TABLE E1.: Overview over the six European regions that are scrutinised in this study.

Region Short Name	Full name	Longitude Range	Latitude Range	Dominant natural ecosystem type[a]	% land cover	% of Natural
NEU-SCA	Scandinavia	12° - 17°E	57.5° - 65°N	Evergreen needleleaf forest (ENF)	68.7%	70.9%
NEU-ENG	England	-3° - 1°E	50.5° - 55°N	Grasslands	17.5%	41.3%
CEU-RUS	North-West Russia	28° - 38°E	53° - 59°N	Mixed forests (MF)	54.0%	85.7%
CEU-FRA	France	-1° - 6°E	43° - 49°N	MF	9.8%	50.1%
MED-SEE	South-East Europe	20° - 25°E	40° - 45°N	Deciduous broadleaf forest (DBF)	13.4%	50.3%
MED-ESP	Spain	-8° - -1°E	37° - 43°N	Woody Savannas	34.7%	55.9%

[a] Based on MODIS PFT classification (Friedl et al., 2010).

E.1.1. Attribution of trends in ecosystem productivity to individual climatic drivers

In addition to factorial model simulations outlined in Chapter 8, we describe a statistical framework that can be used to pinpoint the individual contributions of trends in climatic variables to trends in ecosystem productivity. Our goal is to derive a statistical model that emulates the process-oriented ecosystem model. This approach is presented here in an illustrative manner and complements Chapter 8.

Assume that a relationship (named mapping hereafter) between external forcing variables (*env*) and response variables in the biosphere (*sys*, $sys \in \{GPP, NEP\}$, following notation in Chapter 8) exists for any season in the ecosystem model:

$$\text{sys}_s = f(\text{env}) = f(\text{Tair}_m, \text{Tair}_{m-1}, ..., \text{Tair}_{m-k}, \text{Precip}_m, \text{Precip}_{m-1}, ..., \text{Precip}_{m-k}, \text{Radiation}_m, \text{Radiation}_{m-1}, ..., \text{Radiation}_{m-k}) \tag{E.1}$$

Here, the subscript s denotes any given season (e.g. spring ecosystem carbon uptake), m denotes individual months and respective temporal lags in the climatic variables, and k is the total numbers of lags considered in the statistical model. For example, if the target variable is spring ecosystem productivity in spring (sys_{MAM}), we train a model based on the environmental variables in individual months in spring and before:

$$sys_{MAM} = f(\text{env}_{may}, \text{env}_{april}, \text{env}_{march}, \text{env}_{feb}, ...). \tag{E.2}$$

Assuming an additive approximation of the system yields:

$$\text{sys}_s = f(\text{env}) = \sum_{m=1}^{k} g_{Tair_m}(\text{Tair}_m) + \sum_{m=1}^{k} g_{Precip_m}(\text{Precip}_m) + \sum_{m=1}^{k} g_{Radiation_m}(\text{Radiation}_m). \tag{E.3}$$

If the additive approximation holds, we can directly investigate the contribution of individual drivers to long-term trends in ecosystem productivity by considering the linear trend slopes of individual terms, i.e.

$$\frac{\Delta \text{ sys}_s}{\Delta t} = \sum_{m=1}^{k} \underbrace{\frac{\Delta g_{Tair_m}(\text{Tair}_m)}{\Delta t}}_{\text{Tair contrib.: } \beta_{\mathbf{Tair}}} + \sum_{m=1}^{k} \underbrace{\frac{\Delta g_{Precip_m}(\text{Precip}_m)}{\Delta t}}_{\text{Precip contrib.: } \beta_{\mathbf{Precip}}} + \ldots \tag{E.4}$$

$$= \beta_{\mathbf{Tair}_m} + \beta_{\mathbf{Precip}_m} + \ldots \tag{E.5}$$

The β's in Eq. E.5 are the contribution of trends in individual climate variables to the overall trends in the ecosystem response variable and directly comparable to β_{CO_2} and β_{LU} in Chapter 8.

To illustrate this approach, we use 'multivariate adaptive regression splines' (MARS, Friedman et al. (2001)) as an additive regression technique, i.e. sensu Eq. E.3. In MARS, the $g_m()$'s are so-called hinge-functions that allow one break-point ('knot') per entering variable, which yields piecewise linear relationships with no higher-order terms or interactions allowed. As an additional control, we use multiple linear regression, i.e. where all individual contributions g_m are linear regression coefficients. We train statistical models individually for each regression technique, each of the six regions (see Chapter 8), spring (MAM) and summer (JAS), and separately for two terrestrial ecosystem fluxes (GPP and NEP) as response variables at the seasonal time scale. In each of the examples presented here, monthly climate variables (temperature, precipitation, and short-wave radiation) in the three concurrent months (e.g. March-May for spring, 'transient' in Figure E4), and four lagged months (Nov-Feb for spring, 'early' in Figure E5) are considered as predictors. The LPJmL ensemble simulations with constant land-use and constant CO_2 ('$CONSTLUCO2$', Chapter 8) are used for training and evaluation of the statistical models. Training and prediction of MARS models uses the R package *earth* (Milborrow, 2015).

Evaluation of additive approximation of LPJmL All statistical models are 10-fold cross-validated against an independent out-of-sample set of ensemble members. The so-derived additive approximation of area-averaged GPP and NEP fluxes in LPJmL works relatively well (e.g. in summer, cross-validated R^2 of the predictions exceed 0.8 in 11 out of 12 cases, in spring cross-validated R^2 exceeds

0.8 in eight out of 12 cases, see Figure E2). However, there is some seasonality in the goodness of fits: In summer, where most regions are moisture limited in LPJmL, the performances of the predictions are very high ($R^2 > 0.9$), whereas in spring the predictions only reach very high values in strongly temperature-limited boreal regions such as NEU-SCA or CEU-RUS.

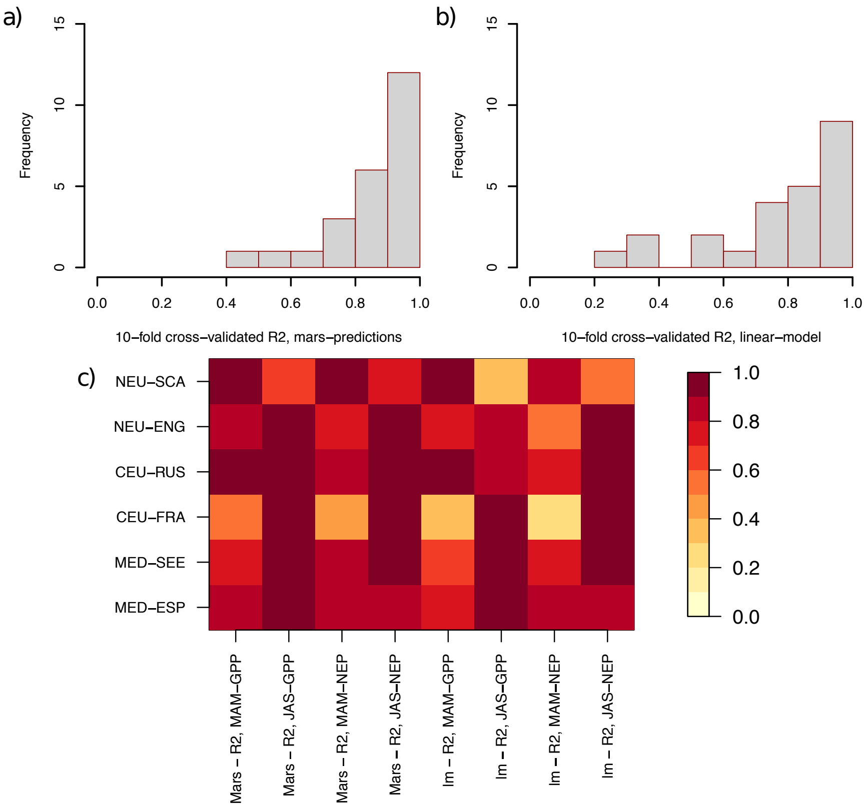


FIGURE E2.: Evaluation of additive approximation of LPJmL-simulated carbon fluxes for trend attribution. Histograms of (a) 10-fold cross-validated MARS-predictions and (b) 10-fold cross-validated predictions based on multiple linear regression (shown for comparison). c) Overview matrix of 10-fold cross-validated R^2 values from MARS-predictions for each of the six regions, both seasons and GPP/NEP predictions.

E.1.2. Analysis methodology and Code

All statistical analyses have been performed in R R Development Core Team (2008). We use the packages *ncdf4* and *raster* for data processing Pierce (2014); Hijmans (2015), *earth* for training and prediction of MARS models Milborrow (2015), *extRemes* and *quantreg* for analysing data and selection of extremes Gilleland and Katz (2011); Koenker (2015), and *riverplot*, *vioplot* and *ggplot2* Weiner (2014); Adler (2005); Wickham (2009) for data visualisation.

E.2. Supplementary Results

E.2.1. Attribution of trends in ecosystem productivity to driving climate variables

The illustrative statistical attribution scheme presented here shows that spring warming is the main climatic factor for increasing trends in spring GPP in the LPJmL ensemble simulations (Figure E4). In summer, warming temperatures decrease GPP in the LPJmL ensemble in five out of six regions, and regions in Southern Europe are additionally affected by a negative precipitation trend that decreases summer GPP (Figure E4). For NEP, as discussed in Chapter 8, trends are generally weaker (Figure E4): The positive contribution of warming spring temperatures to spring NEP trends is found only marginally in three regions in North Europe (NEU-SCA, NEU-ENG, and CEU-RUS), while warming summer temperatures cause a strongly decreasing NEP trend in summer across all regions. In summary, a contrasting response to warming temperatures in spring vs. summer is responsible for the contrasting response of ecosystem productivity to recent changes in climate (discussed in detail in Chapter 8).

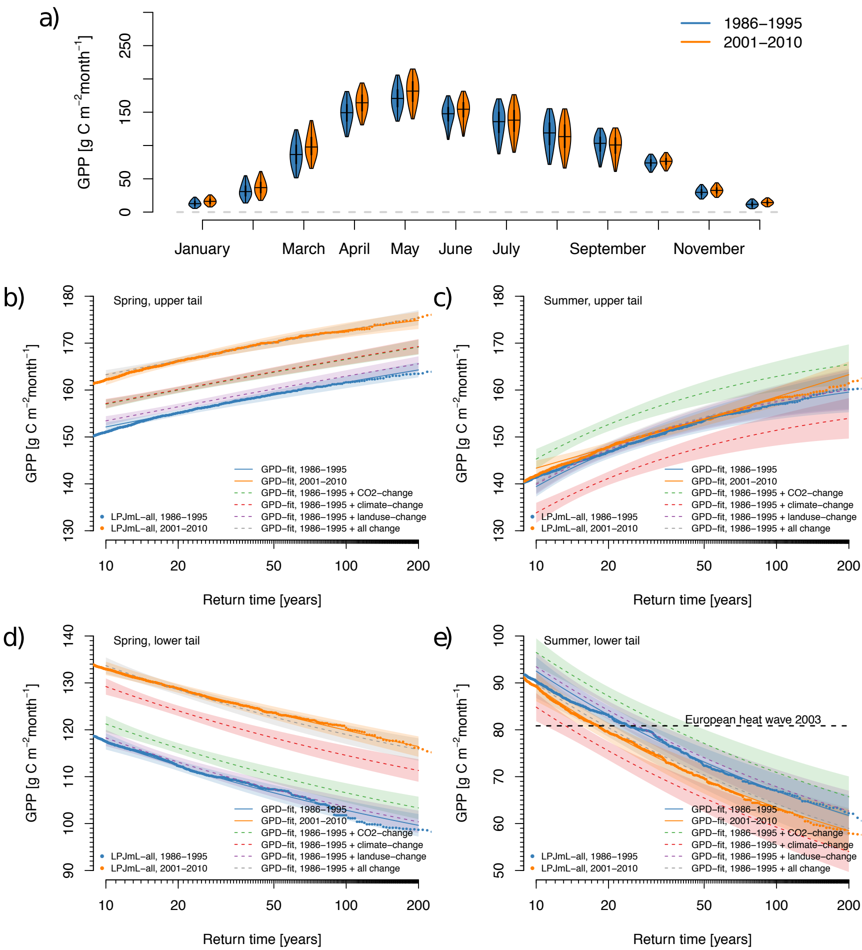


FIGURE E3.: a) Seasonal cycle of GPP distribution as simulated by the LPJmL-ensemble for 1986-1995 and 2001-2010 in the France subregion. b–e) Return time plots of carbon flux extremes in spring (b,d) and summer (c,e) for the upper (b,c) and lower (d,e) tail of the GPP distribution for 1985-1995 and 2001-2010 (solid lines). Dotted lines indicate the changes imposed to the 1986-1995 GPP tail by adding the average individual contribution of CO_2-, climate-, and land use-driven changes between the two periods.

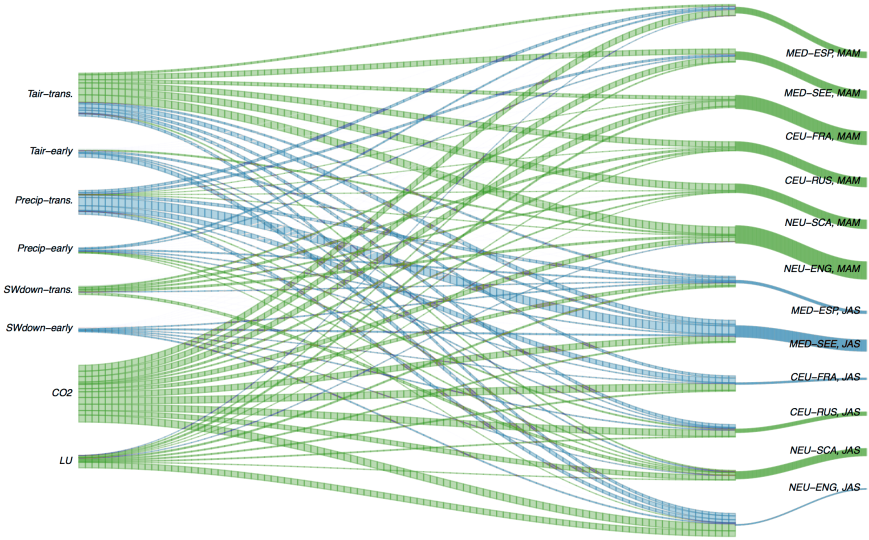


FIGURE E4.: Climatic variables that drive recent changes in GPP across six European regions in spring and summer in LPJmL simulations. The width of the link between the env-variable (left) and GPP in each region (right) indicates the contribution of the individual driver (e.g. β_{Tair}, β_{Precip}, etc.). Note that blue colours indicate a negative trend contribution of the respective driver, whilst green colours indicate a positive contribution (trans. - at time of seasonal anomaly, early - before seasonal flux anomaly).

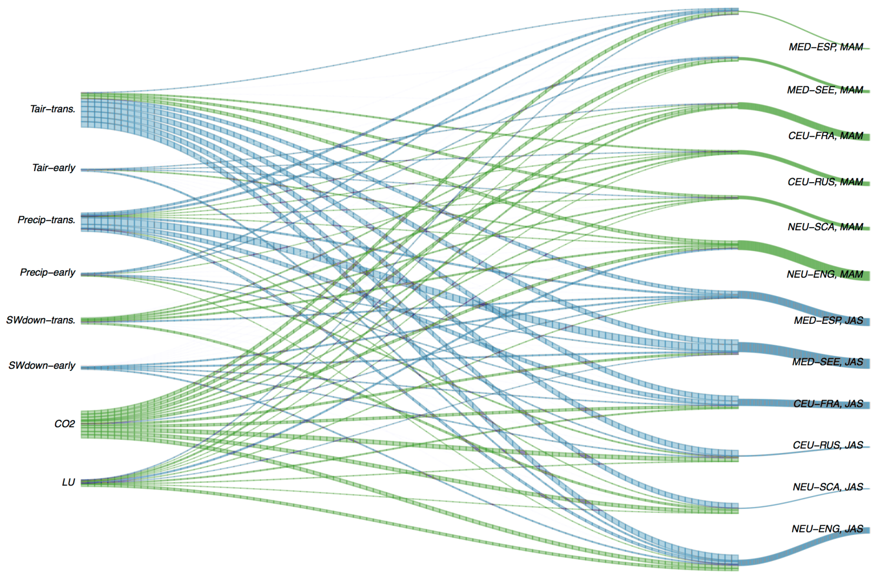


FIGURE E5.: Same as Figure E4, but for NEP.

E.2.2. Interacting carbon cycle effects due to climate extremes

TABLE E2.: Correlation of spring-summer carry-over effects with soil moisture. Carry-over effects in this study are inferred from the differences between negative extremes in summer GPP and GPP under the same meteorological forcing in summer but with randomised spring conditions (i.e. $\Delta GPP = GPP_{j,JJAS,t,All}^{-extreme,JJAS} - GPP_{j,JJAS,t,SPRINGRAND}^{-extreme,JJAS}$, where ΔGPP denotes carry-over effect). Therefore, we test whether differences in soil moisture content between both runs can explain the observed carry-over effects. Carry-over effects in NEP follow similarly. $SWC1$ and $SWC2$ denote soil water content in soil layers 1 and 2, respectively, and $\rho()$ is the Pearson correlation coefficient.

Region	$\rho(\Delta GPP, \Delta SWC1)$	$\rho(\Delta GPP, \Delta SWC2)$	$\rho(\Delta NEP, \Delta SWC1)$	$\rho(\Delta NEP, \Delta SWC2)$
NEU-SCA	0.78	0.87	0.83	0.90
NEU-ENG	0.52	0.85	0.53	0.88
CEU-RUS	0.80	0.92	0.82	0.90
CEU-FRA	0.32	0.90	0.35	0.91
MED-ESP	0.13	0.88	0.22	0.59
MED-SEE	0.19	0.92	0.31	0.91