

CLIMATE EXTREMES IN A GENERAL CLIMATE
MODEL WITH STOCHASTIC
PARAMETERIZATIONS

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Felipe Ignacio Tagle

February 2016

© 2016 Felipe Ignacio Tagle

ALL RIGHTS RESERVED

CLIMATE EXTREMES IN A GENERAL CLIMATE MODEL WITH STOCHASTIC PARAMETERIZATIONS

Felipe Ignacio Tagle, Ph.D.

Cornell University 2016

This work employs techniques from extreme value theory to evaluate the representation of temperature and precipitation extremes in two climate models and an observational dataset. The climate models correspond to the general climate model, the NCAR Community Atmosphere Model version 4 (CAM4), with two stochastic parameterizations of sub-grid scale processes: the stochastic kinetic energy backscatter (SKEBS) scheme and the stochastically perturbed parameterization tendency (SPPT) scheme. The observational dataset is version 7 of the satellite-based Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) 3B42 research product, developed at the National Aeronautics and Space Administration Goddard Space Flight Center. Temperature extremes are described in terms of the 95th percentile (20-yr return level) of the distribution of annual extremes of near-surface temperature, while precipitation extremes are characterized in terms of the analogous percentile with respect to daily precipitation amounts, in addition to less extreme precipitation statistics. The distribution of annual extremes is assumed to be well-approximated by the Generalized Extreme Value (GEV) distribution, which is fit at each gridpoint using a “block maxima” approach. A Bayesian hierarchical approach is used to estimate the parameters of the distribution of annual extremes in the 3B42 dataset.

CAM4 overestimates warm and cold extremes over land regions, particu-

larly over the Northern Hemisphere when compared against observations and reanalysis. The addition of a stochastic parameterization generally produces a warming of both warm and cold extremes relative to the unperturbed configuration, however, neither of the proposed parameterizations meaningfully reduce the biases in the simulated temperature extremes of CAM4. Similarly, the precipitation response to the use of stochastic parameterizations is remarkably muted, particularly that to SPPT. SKEBS is shown to enhance the dry bias of annual precipitation in CAM4 over the central contiguous United States, and also exacerbates the shortfall of moderate precipitation extremes over the same region. The 3B42 dataset shows severe overestimation of 20-yr return levels over eastern Asia.

The analysis of the 3 parameters that define the GEV distribution enhances the understanding of the behavior of extremes, revealing valuable information that may potentially help modeling centers improve the simulation of extremes in the climate models.

BIOGRAPHICAL SKETCH

Felipe Tagle began his undergraduate studies in Industrial Engineering at the Pontificia Universidad Catolica de Chile in 1995. After graduating with honors in 2001, he pursued a career in asset management in the Chilean financial industry, first as an international investment analyst at AFP Cuprum, one of the leading pension of the country, and later as a portfolio manager of international bond funds at the asset manager Larrain Vial Administradora General de Fondos. Eager to enhance his training in mathematical finance, he joined the Master's in Engineering program at Cornell University, with a Financial Engineering concentration, in 2007. As he interacted with the faculty at the Operations Research and Information Engineering department and learned of the research that was being done there he became interested in pursuing a doctoral degree at the same institution. In the Fall of 2009, he began his PhD studies at Cornell University intent on pursuing research in the field of mathematical finance, but during his second year in the program, he decided to collaborate with Gennady Samorodnitsky on an National Science Foundation grant aimed at studying the impact of stochastic parameterizations on extreme climate events, which became the topic of his dissertation.

to Olga, Maria Eugenia and Florencia; my past and my future.

ACKNOWLEDGEMENTS

The work contained in this dissertation would not have been possible without the help and support of many individuals. I would like to thank my advisor, Gennady Samorodnitsky, for convincing me that climate science was equally if not more interesting and challenging than mathematical finance, and for his insights during the course of my PhD studies. I would also like to thank my co-advisor Natalie Mahowald, for helping me through this transition to the atmospheric sciences, and Leonard Gross, whose brilliance, warmth and humility will always be a source of inspiration. It was a true honor to attend his Analysis and Functional Analysis courses in the mathematics department. A special thanks also to Mircea Grigoriu for his guidance and insights during the many weekly workgroup meetings.

Many thanks to Judith Berner, at the National Center for Atmospheric Research (NCAR), who has not only contributed significantly to the work herein but through her support and guidance during my several visits to NCAR has become a true mentor as I move from student to scientist. I would also like to thank the members of the Climate and Global Dynamics group and the Institute for Mathematics Applied to Geosciences at NCAR, particularly Joe Tribbia and Doug Nychka, for their hospitality during these visits.

A special thanks to my parents, without whose support none of this would have been possible. And to my daughter, Florencia, for her love despite the many years apart.

TABLE OF CONTENTS

Biographical Sketch	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	viii
List of Figures	ix
1 Introduction	1
2 Temperature Extremes	5
2.1 Quantile analysis	5
2.1.1 Introduction	5
2.1.2 Data and Experimental Design	8
2.1.3 Methodology	13
2.1.4 Results	18
2.1.5 Discussion and Conclusions	35
2.1.6 Appendix	40
2.2 Temporal evolution	41
2.2.1 Motivation	41
2.2.2 Data	43
2.2.3 Results and further remarks	44
3 Precipitation Extremes	52
3.1 Introduction	52
3.2 Data and Methods	55
3.3 Results	57
3.3.1 Annual precipitation extremes	57
3.3.2 Non-extreme precipitation metrics	63
3.3.3 Moderate extreme precipitation metrics	67
3.4 Discussion and Conclusions	69
4 Precipitation Extremes in TRMM 3B42	74
4.1 Extreme Value Theory	76
4.2 Data	78
4.2.1 TMPA	79
4.3 Model	81
4.3.1 First stage	81
4.3.2 Second stage	82
4.3.3 Third stage	85
4.4 Implementation	86
4.4.1 MCMC Implementation	87
4.5 Results	88
4.6 Summary and Conclusions	91

LIST OF TABLES

2.1	CMIP5 models used in the analysis. The respective temperature indices were downloaded from the ETCCDI indices archive (http://www.cccma.ec.gc.ca/data/climindex/).	45
2.2	CMIP5 models used in the analysis (continued).	46

LIST OF FIGURES

1.1	Schematic showing the range of time and horizontal scales of different atmospheric phenomena (from [60]).	3
2.1	Idealized Gaussian distributions of temperature T (top) for CAM (blue) and ERA (red) and the corresponding GEV probability density functions obtained by drawing 5000 samples of length 10000 from the distributions above and fitting GEV distributions to the resulting 5000 extreme observations. A standard Gaussian distribution is assumed for ERA T . The left panel considers T distributions with the same standard deviation, but the climatological mean of CAM is shifted by 1, while the right panel considers T distributions with the same mean, but the climatological standard deviation of CAM is 1.3. Vertical lines denote the mean of the GEV distribution. GEV distributions adjusted by their T_{max} (g, h) and T_{min} (i, j) climatologies.	20
2.2	GEV pdfs of CAM4 (blue) and ERA (red) with parameters values (left legend) set to the area-averaged estimates over land regions, with gridpoint estimates corresponding to averages of parameter values obtained from 500 resampled replicates of the annual extremes of daily maximum temperature. Location and 20-yr return level differences are indicated in the right legend, along with differences in a modified scale parameter, defined as $\Delta\tilde{\sigma}_s = -(\sigma_s - \sigma_{ERA})f(\xi_{ERA})$, with f obtained from equation (2.1), and s representing the CAM dataset. Probability density functions adjusted by the area-averaged maximum of the annual cycle $\max \bar{T}_{max}^{ac}$ (b) and $\max \bar{T}_{mean}^{ac}$ (c). Vertical lines denote 20-yr return levels.	21
2.3	CAM4 20-yr return level estimates of 1979-2010 annual extremes of daily maximum (left) and minimum (right) temperature, and the corresponding differences with estimates from ERA Interim (middle) and the HadEX2 observational dataset (bottom). Stippling indicates return level differences are significant at a 5% level.	23
2.4	CAM4 20-yr return level estimates (top) of 1979-2010 annual extremes of daily maximum (left) and minimum (right) temperature, and the corresponding differences with estimates of ERA Interim (bottom). Return level estimates are adjusted by the corresponding extremes of the annual cycle, $\max \bar{T}_{max}^{ac}$ and $\min \bar{T}_{min}^{ac}$	25
2.5	Differences of 20-yr return level estimates of 1979-2010 annual extremes of daily maximum (left) and minimum (right) temperature between CAM4 and the stochastic parameterization schemes, SKEBS (top), SPPT (bottom).	27

2.6	Continental regions adapted from the sub-continental definitions of [79] (as in Fig. 10.7 of [12]).	28
2.7	Same as in Fig. 2.5, except simulations have been adjusted by the respective extremes of the T_{max} and T_{min} annual cycles.	29
2.8	GEV pdfs, as in Fig. 2.2(a), of CAM (solid blue), and SKEBS (dashed blue) for (left) warm and (right) cold extremes over land, with respect to (top) ERA (red) and (bottom) HadEX2 (red). Parameter values in the HadEX2 comparison correspond to area-averages within 15-72.5°N, excluding Greenland.	30
2.9	GEV pdfs of T_{max} annual extremes, as in Fig. 2.8(a), with parameter values area-averaged over continental regions defined in Fig. 2.6.	32
2.10	As in Fig. 2.9, with T_{min} annual extremes.	33
2.11	GEV pdfs as in Fig. 2.9 and Fig. 2.10, of T_{max} annual extremes over Northern Hemisphere land regions (top), and T_{min} annual extremes over Southern Hemisphere land regions (bottom); adjusted by the respective extreme of the monthly mean temperature annual cycle, \bar{T}_{mean}^{ac}	34
2.12	GEV pdfs, as in Fig. 2.11, but adjusted by the respective extreme of the T_{max} and T_{min} annual cycle.	35
2.13	Differences of 20-yr return level estimates of 1979-2010 annual extremes of daily maximum (left) and minimum (right) temperature between ERA and the stochastic parameterization schemes, SKEBS (top), SPPT (bottom). Stippling indicates return level differences are significant at a 5% level.	40
2.14	Same as in Fig. 2.13, except simulations have been adjusted by the respective extremes of the T_{max} and T_{min} annual cycles.	40
2.15	Global area-averaged TXx, TXn, TNn, TNx ETCCDI indices of CAM4, SKEBS, SPPT, reanalysis and observations over the period 1979-2010. The shading represents the interquartile model spread (range between the 25th and 75th quantiles) of 26 models belonging to the CMIP5 archive. For each index, the left column displays the absolute values and the right column shows anomalies with respect to the reference period 1981 to 2000.	50
2.16	Decadal trends in global and regional area-averaged TXx, TXn, TNn and TNx indices of CAM4, SKEBS, SPPT and ERA, with regions defined in Fig. 2.6. Light blue dots denote analogous index decadal trends from the 5 other members of the 1° IPCC-AMIP CAM4 ensemble. Gray block denotes the 5-95th percentile range of decadal trends from 26 models of the CMIP5 archive.	51
3.1	The ratio of the multimodel ensemble mean of P_{20} estimates over P_{20} estimated from ERA-40 (Figure 7 from [53])	53

3.2	The ratio of the CMIP5 median of P_{20} over ERA-Interim P_{20} (Figure 1 from [54])	54
3.3	(left) 20-yr return level estimates of 1979-2010 annual extremes of daily precipitation of CAM4 and (right) and the ratio with those of ERA.	58
3.4	Ratios of 20-yr return level estimates of 1979-2010 annual extremes of daily precipitation of (left) SKEBS with (top) ERA and (bottom) CAM4, and (right) SPPT with (top) ERA and (bottom) CAM4.	59
3.5	GEV pdfs as in Fig. 2.2, but of annual extremes of daily precipitation, with (top) CAM (solid blue) and SKEBS (dashed blue) with respect to ERA (red) over (left) land and (right) oceans. (bottom) As above but with SPPT in place of SKEBS.	60
3.6	As in Fig. 3.5 but with the spatial domain restricted to 30°S-30°N.	62
3.7	Topographical map of the contiguous United States showing the regions defined for the National Climate Assessment (from [84]).	64
3.8	(top-left) 1979-2010 average of annual precipitation for observations, and anomalies, as percent, of (top-right) CAM4, (bottom-left) SKEBS and (bottom-right) SPPT. Observations correspond to the NOAA CPC 0.25x.025 Daily US Unified Gauge-Based Analysis of Precipitation data. Stippling in the panels of the stochastic parameterizations indicates differences with CAM4 are statistically significant at the 5% level.	65
3.9	As in 3.8, but with respect to precipitation frequency.	66
3.10	As in 3.8, but with respect to precipitation intensity.	67
3.11	As in 3.8, but with respect to the number of wet days.	68
3.12	As in 3.8, but with respect to extreme precipitation fraction.	69
3.13	As in 3.8, but with respect to extreme event intensity.	70
4.1	Location parameter estimates based on the (left) Bayesian hierarchical framework (posterior mean) and (right) maximum likelihood estimation performed at each grid-point, for JJA.	89
4.2	Posterior means of (left) location, (middle) scale, and (right) shape parameter estimates based on Bayesian hierarchical framework, for the (top) TRMM 3B42V7 dataset and the (bottom) APHRO dataset, for JJA.	90
4.3	(left) JJA 20-yr return level estimates obtained from posterior means of GEV parameter values for APHRO, and (right) the respective differences with the TRMM dataset.	91

CHAPTER 1

INTRODUCTION

In recent decades, there has been increasing interest in the analysis of extreme climate events given their significant impact on human and natural systems [59, 33]. Such events typically account for thousands of deaths and billions of dollars in damages globally each year, as population and infrastructure continue to expand in areas that are vulnerable to extremes such as flooding, storm damage, and extreme heat or cold [33]. Simulations from global coupled climate models are the primary tools for forecasting potential future changes in extreme climate statistics [53]. Evaluations of the models participating in the Coupled Model Intercomparison Project Phase 3 (CMIP3) and 5 (CMIP5) have shown that temperature extremes are reasonably represented, as compared to reanalysis and observations [35], with greater uncertainties in the simulation of cold extremes [53, 54].

Owing to the multitude of extreme events in the climate system and how the extreme nature of a climate phenomenon is usually dependent on the affected region [86], most studies of climate extremes rely on the use of extreme indices [89, 1, 79, 42]. Other studies focus on more extreme climate statistics, typically relying on results from extreme value theory to approximate the distribution of annual extremes [53, 54, 13, 94, 96]. In this dissertation, we follow the latter approach, and use tools from extreme value theory to investigate the representation of temperature and precipitation extremes in both climate models and observational datasets.

The climate models on which we focus are experiments using stochastic parameterization, but in order to introduce the concept of stochastic parameter-

ization we must begin by understanding the structure of a numerical climate model. Figure 1.1 depicts the range of time and horizontal scale of different atmospheric phenomena, all of which are important for weather and climate. As explained in [60], there is no significant *spectral gap*, that is, relevant atmospheric processes occur at all scales. Furthermore, strong interactions exist between the phenomena at different scales, therefore these interactions must be represented in a numerical model aimed at simulating the evolution of Earth's weather or climate. However, despite the dramatic increases in computational power in recent decades, computer resources remain finite, which inevitably leads to numerical models having a finite resolution. The shaded region in the figure shows the typical space and time scales of phenomena that can be resolved in a present day climate model. The processes that fall outside of the region – the unresolved processes – cannot be neglected. Instead, they are represented by sub-grid models or so-called physical parameterizations, or simply *parameterizations*. Thus, the *dynamical core* is the part of the model that solves the governing fluid and thermodynamic equations on resolved scales, while the parameterizations represent sub-grid scale processes and other processes not included in the dynamical core such as radiative transfer [60].

Presently, these parameterizations are typically deterministic, relying on bulk formulas, based on the notion that the properties of the unresolved subgrid scales are determined by the large scales. However, studies have shown that resolved states are associated with many possible unresolved states, suggesting the use of stochastic methods for numerical weather and climate prediction that potentially allow a proper representation of the uncertainties, a reduction of systematic biases, and improved representation of long-term climate variability [37].

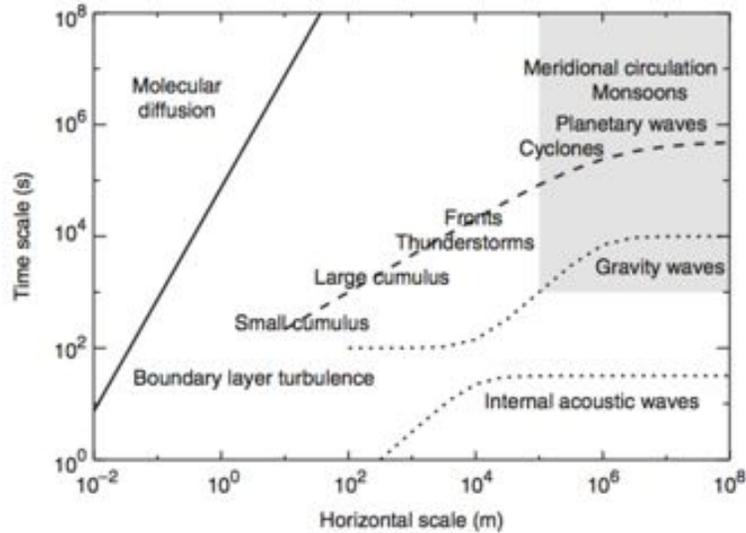


Figure 1.1: Schematic showing the range of time and horizontal scales of different atmospheric phenomena (from [60]).

While several studies have highlighted the improvements in the representation of certain atmospheric phenomena due the use of stochastic parameterizations (i.e. [48, 4, 7]), no studies thus far have addressed the impact on climate extremes. Here we examine the role of two stochastic parameterizations of sub-grid scale processes, the stochastic kinetic energy backscatter (SKEBS) scheme and the stochastically perturbed parameterization tendency (SPPT) scheme, on the simulation of temperature and precipitation extremes.

The focus of chapter 2 is on temperature extremes. While there are many ways to characterize such extremes, we follow the recent literature and express temperature extremes in terms of an upper percentile of the distribution of annual extremes, specifically, the 95th percentile, which in the climate and hydrology literature is commonly referred to as the 20-yr return level. This distribution, according to extreme value theory, is well approximated by the Generalized Extreme Value distribution, which we fit to the sample of annual extremes

following the “block maxima” approach. The latter work has been published, under the title of “Temperature Extremes in the Community Atmosphere Model with Stochastic Parameterizations”, in the January 2016 issue of the Journal of Climate. In light of the increasing interest in the temporal evolution of climate extremes, we also provide a brief analysis of the temporal evolution of temperature extremes, as simulated by these stochastic parameterizations, based on a subset of a popular family climate indices.

Chapter 3 examines precipitation extremes. The first half of the chapter repeats the analysis based on 20-yr return levels of chapter 2, but with respect to the distribution of annual extremes of daily precipitation, while the second half focuses on less extreme as well as non-extreme aspects of the precipitation distribution, based on precipitation statistics commonly featured in the literature.

The final chapter, chapter 4, broadens the focus of the previous chapters, as it investigates the representation of extremes in a satellite-based precipitation product, TRMM 3B42 version 7, using a rain gauge-based dataset as reference. It also differs in the methodology used to estimate the parameters of distribution of extremes. Due to the short-sample size of the TRMM dataset, we use a technique from extreme value theory based on exceedances above a high threshold, in addition to a Bayesian hierarchical model in order to borrow strength across space. But perhaps what is most noteworthy is the use of a high performance computing technique to carry out the Markov Chain Monte Carlo simulation, that exploits the computing capabilities of graphical processing units (GPU) to achieve speed ups of up to 50x.

CHAPTER 2

TEMPERATURE EXTREMES

2.1 Quantile analysis

2.1.1 Introduction

In recent decades, there has been increasing interest in the analysis of extreme climate events given their significant impact on human and natural systems [59, 33]. Such events typically account for thousands of deaths and billions of dollars in damages globally each year, as population and infrastructure continue to expand in areas that are vulnerable to extremes such as flooding, storm damage, and extreme heat or cold [33]. The Intergovernmental Panel for Climate Change (IPCC) in their Fifth Assessment Report [42] concluded that most global land areas have experienced significant warming of both maximum and minimum temperature extremes since about 1950. Simulations from global coupled climate models are the primary tools for forecasting potential future changes in extreme climate statistics [53]. Because an important aspect of the evaluation of the reliability of these forecasts is an assessment of the models' ability to simulate observed climate extremes, the release of a new generation of climate models is usually accompanied by studies focusing on this topic [53, 54, 81]. Evaluations of the models participating in the Coupled Model Intercomparison Project Phase 3 (CMIP3) and 5 (CMIP5) have shown that temperature extremes are reasonably represented, as compared to reanalysis and observations [35], with greater uncertainties in the simulation of cold extremes [53, 54]. However, performance in representing temperature extremes is strongly dependent on the

choice of verification dataset, particularly reanalyses, as discrepancies between these can be as large as the intermodel spread between the CMIP ensemble [81].

Owing to the multitude of extreme events in the climate system and how the extreme nature of a climate phenomenon is usually dependent on the affected region [86], most studies of climate extremes rely on the use of extreme indices [89, 1, 79, 42]. Other studies focus on more extreme climate statistics, typically relying on results from extreme value theory to approximate the distribution of annual extremes [53, 54, 13, 94, 96]. In this study we follow the latter approach and evaluate the performance of the NCAR Community Atmosphere Model version 4 (CAM4) [70], in simulating observed temperature extremes, as measured by 20-yr return levels, against reanalysis and observational datasets, and furthermore, investigate whether the observed discrepancies are climatological in nature. Indeed, it is not unusual for climate models to exhibit systematic errors in mean temperatures; for instance, a longstanding error that is present in the NCAR Community Earth System Model (CESM) is the so-called “warm bias over land”, which refers to temperatures over land being too warm in summer [70]. Here, we will examine to what extent differences in mean temperature between the model and the verification datasets explain the observed discrepancies in temperature extremes.

Despite the continuing increase of computing power, which allows climate models to be run with ever-higher resolution, many important physical processes (e.g. tropical convection, gravity wave drag, micro-physical processes) are still not resolved [80, 37]. Some subgrid-scale processes are altogether unrepresented or represented very crudely, for example, flow over mountains is a source of turbulent kinetic energy at small scales, but in the model it is rep-

resented by a drag coefficient [72, 80]. Since fine-scale climate processes have been shown to regulate not only the mean, but also the tails of the daily temperature and precipitation distributions [29], numerous studies have highlighted the importance of horizontal grid resolution on the simulation of climate extremes [96, 75, 47, 58]. Stochastic-dynamics prediction is an alternative way to represent the effect of fluctuating subgrid-scale processes [72]. The omission of variability of unresolved subgrid-scale processes has been proposed as one reason for persistent biases across different models [46, 74, 7]. For instance, [7] showed that including a stochastic representation leads to comparable improvements in the Northern Hemispheric circulation as increasing horizontal resolution.

One such stochastic parameterization is the stochastic kinetic energy backscatter scheme (SKEBS), whose origin lies in large-eddy simulation modeling [67]. It is based on the rationale that a small fraction of the model dissipated energy interacts with the resolved-scale flow and acts as a systematic forcing. Recently, it was adapted by [80] to numerical weather prediction. Its impact on weather and seasonal forecasts are reported, for example, in [4, 8, 6, 5], [73] and [30]. On seasonal scales, in integrations with the European Centre for Medium-Range Weather Forecasts (ECMWF) model, SKEBS has been shown to reduce the bias in the circulation over the North American continent and improve the occurrence of Northern Hemispheric blocking [48, 4]. In the tropics, SKEBS positively influences the representation of convectively coupled waves [7]. Another stochastic parameterization is the stochastically perturbed parameterization tendency (SPPT) scheme [14, 73] which samples the physical tendencies from an assumed subgrid-scale probability density function. The present study will provide the first assessment of the impact of stochastic parameterizations

on climate extremes.

The paper is organized as follows. The datasets and a brief review of the stochastic parameterizations is presented in section 2.1.2. Section 2.1.3 describes the methodology, while section 2.1.4 compares the model-simulated extremes with the observational evidence. Section 2.1.5 provides a discussion and concludes.

2.1.2 Data and Experimental Design

CAM is the atmospheric component of the Community Climate System Model (CCSM) and the new CESM, developed at the National Center for Atmospheric Research (NCAR), under the support of the National Science Foundation and the Department of Energy. In this study we consider a simulation of CAM with prescribed SSTs and sea ice according to the Atmospheric Model Inter-comparison Project (AMIP) protocol [39]. Specifically, we use a so-called 1° IPCC-AMIP simulation, spanning the years 1979-2010, at a horizontal resolution of $0.9^\circ \times 1.25^\circ$. Simulated temperature extremes are evaluated against ERA Interim (hereafter ERA) [28], the latest global atmospheric reanalysis produced by ECMWF, covering the years 1979 to present, on a $1.5^\circ \times 1.5^\circ$ regular grid. Maximum and minimum 2m temperature since previous post-processing, at a 12h time step, from 1979 to 2010, were downloaded from the ECMWF archive (<http://apps.ecmwf.int/datasets/data/interim-full-daily/>), from which daily temperature extremes were constructed as the maximum and minimum of the 2 respective 12h daily values. For each dataset, annual extremes of daily maximum and minimum 2m temperature are computed at each gridpoint over land.

Although reanalyses offer the advantages of gridded output with global spatial coverage, they are nonetheless observationally-constrained model output. Their output may be classified into 4 categories depending on the relative influence of the observational data and the numerical model [49]. Near-surface temperature belongs to the “type B” category, indicating that observational data exists that directly affects its value, however, the model component still exercises considerable influence. Significant discrepancies in extreme temperature statistics between reanalyses have been documented in [53, 54] and [81], particularly for cold extremes. However, ERA has been shown to adequately capture recent temperature extreme trends over Europe [24]. Moreover, in a comparison of several reanalysis products, [31] reported that temperature extremes in ERA exhibit the highest temporal and spatial correlations with those of gridded observations over the past 30 years.

As an additional verification dataset, we consider the gridded land-based HadEX2 observational dataset. HadEX2 consists of the comprehensive set of indices of temperature and precipitation extremes defined by the Expert Team on Climate Change Detection Indices (ETCCDI), which are calculated directly from station data and interpolated onto a regular grid using a modified version of Shepard’s angular distance weighting interpolation algorithm (see [32] for details). In many countries, these provide the only publicly available information about temperature and precipitation extremes. Because the HadEX2 indices are derived solely from station data, they are free from biases originating from model specification error, as is potentially the case for reanalysis products; however, direct comparison with model output becomes encumbered by the fact that the latter corresponds to area averages, not point values. Systematic biases introduced by this spatial-scale mismatch, however, should be minor given the

smoothness of temperature fields. The dataset is available on a $2.5^\circ \times 3.75^\circ$ grid, with a temporal coverage of 1901-2010. We use the TXx and TNn indices, corresponding to the annual extremes of daily maximum and minimum 2m temperature but restrict the domain of analysis to those gridpoints with complete temporal coverage of the 1979-2010 period.

To investigate the impact of stochastic parameterizations on temperature extremes two experiments were performed.

The stochastic kinetic-energy backscatter scheme

The stochastic kinetic-energy backscatter scheme (SKEBS) aims to represent model uncertainty arising from unresolved subgrid-scale processes by introducing random perturbations to streamfunction and potential temperature tendencies. SKEBS is based on the rationale that a small fraction of the model dissipated energy interacts with the resolved-scale flow and acts as a systematic forcing.

The scheme introduces at each time step and gridpoint additive perturbations to the streamfunction tendency

$$\dot{\psi}_p(\phi, \lambda, t) = \dot{\psi}_{\text{dyn}}(\phi, \lambda, t) + f(\phi, \lambda, t),$$

where $\dot{\psi}_{\text{dyn}}$ and $\dot{\psi}_p(\phi, \lambda, t)$ is the streamfunction tendency before and after perturbation and $f(\phi, \lambda, t)$ the perturbation tendency. Here, λ and ϕ denote longitude and latitude in physical space and t time. Furthermore, we let the perturbation tendency forcing be expressed in a triangularly truncated spherical

harmonics expansion:

$$f(\phi, \lambda, t) = \sum_{m=-N}^N \sum_{n=|m|}^N f_n^m(t) P_n^{|m|}(\cos \phi) e^{im\lambda}.$$

Here, m and n denote the zonal and total wavenumbers, N is the truncation wavenumber of the numerical model, and P_n is the associated Legendre function of degree n and order m . The spherical harmonics $Y_n^m = P_n^m e^{im\lambda}$ form an orthogonal set of basis functions on the sphere. If they are non-vanishing for at least one $n \leq N$ and do not follow a white-noise spectrum, the pattern perturbations will be spatially correlated in physical space.

Since the physical processes mimicked by this forcing have finite correlation times, temporal correlations are introduced by evolving each spectral coefficient as a first-order autoregressive (AR1) process:

$$f_n^m(t + \Delta t) = \alpha f_n^m(t) + \sqrt{(1 - \alpha)} g_n \epsilon(t),$$

where α is the linear autoregressive parameter determining the temporal decorrelation time, g_n the wavenumber-dependent noise amplitude and ϵ a Gaussian white-noise process with mean zero and variance η . The noise amplitude g_n is chosen to have power-law behavior, $g_n = bn^p$, and determine the variance spectrum of the forcing.

The pattern $f(\phi, \lambda, t)$ is interpreted as a streamfunction tendency forcing. In the case of perturbing potential temperature, a second perturbation pattern is created analogously, but with a different power-law behavior and potentially a different temporal correlation. The behavior of this scheme is determined by the following parameters: the exponent of the power law, p ; the wavenumber perturbation range, $n_1 - n_2$; and the amplitude of forcing energy, which determines the normalization constant b .

In the original implementation, the streamfunction pattern is subsequently weighted with the normalized total instantaneous dissipation rate from numerical dissipation, deep convection and gravity and mountain wave drags [80, 8] so that the perturbations are largest in regions with large dissipation and have little effect in regions where and when the dissipation is small. A simplified version of SKEBS assumes the dissipation rate to be a spatially and temporally constant, resulting in a state-independent (additive) stochastic forcing. This simplification relies on underlying model dynamics to determine which perturbations will grow and which ones will be damped [6]. Here, we use the simplified version with constant dissipation rate.

The stochastically perturbed parameterization tendency scheme

The stochastically perturbed parameterization tendency (SPPT) scheme is a revision of the original stochastic diabatic tendency scheme of [14] and perturbs the parameterized tendency of physical processes with multiplicative noise. It is based on the notion that – especially with increasing numerical resolution – the equilibrium assumption no longer holds and the subgrid-scale state should be sampled rather than represented by the equilibrium mean. Consequently, SPPT multiplies the accumulated physical tendencies \dot{x} of temperature, zonal and meridional winds and humidity (T, u, v, q) at each gridpoint and time step with a multiplicative random coefficient $r(\phi, \lambda, t)$:

$$X_p = (1 + r) \dot{x}, \quad \text{with } x = u, v, T, q.$$

Here, X_p is the perturbed parameterized tendency for the variables $x = u, v, T, q$ and $r(\phi, \lambda, t)$ a random pattern with spatial and temporal correlations. By design, the perturbations are large where the physical tendencies, and presumably their

uncertainty, is large and has very little effect where and when the tendencies are small. SPPT uses the same pattern generator as SKEBS (see above) but a different normalization.

The stochastic pattern evolves in spectral space as

$$r_n^m(t + \Delta t) = \alpha r_n^m(t) + \sqrt{(1 - \alpha)} g_n \epsilon(t),$$

where all variables are as defined above. The temporal correlations are given by the decorrelation time τ defining $\alpha = \exp(-\Delta t/\tau)$. The noise amplitudes are given as

$$g_n = F_0 \exp -L \frac{n(n+1)}{2} \quad \text{with} \quad F_0 = \left(\frac{\sigma^2 (1 - \alpha)^2}{2 \sum_{n=1}^N (2n+1) \exp(-Ln(n+1))} \right)^{1/2},$$

where L is a horizontal length scale defining the spatial correlations and σ^2 the perturbation variance at each gridpoint. The normalization constant F_0 is chosen so that the variance at any gridpoint, σ^2 , is given by the total variance in spectral space [93]. The resulting stochastic pattern follows at each gridpoint a Gaussian with mean zero and variance σ^2 .

2.1.3 Methodology

In this study we characterize the extreme behavior of the annual extremes of daily maximum (T_{max}) and minimum (T_{min}) temperature in terms of 20-yr return values. A T -year return value can be informally interpreted as the value that is exceeded by an annual extreme on average once every T years. More precisely, it is defined as the quantity that is exceeded in any given year with probability $p = 1/T$, which corresponds to the $(1 - p)$ quantile of the distribution of annual

extremes. We assume that the distribution of annual extremes can be approximated by a Generalized Extreme Value (GEV) distribution, given by

$$F(x; \mu, \sigma, \xi) = \begin{cases} \exp \left[- \left(1 + \xi \frac{x - \mu}{\sigma} \right)^{-1/\xi} \right] & \xi \neq 0, 1 + \xi(x - \mu)/\sigma > 0, \\ \exp \left[- \exp \left(- \frac{x - \mu}{\sigma} \right) \right] & \xi = 0 \end{cases}$$

with μ the location parameter, $\sigma > 0$ the scale parameter and ξ the shape parameter. The shape parameter controls the tail behavior of the distribution. If $\xi < 0$ the probability density function (pdf) is bounded above, with upper endpoint given by $\mu - \sigma/\xi$. If $\xi > 0$, the GEV distribution is heavy-tailed and its pdf decays “slowly” (polynomially) as $x \rightarrow \infty$; when $\xi = 0$ the tails of the pdf decay relatively faster. These cases give rise to 3 distinct families of extreme value distributions: the Weibull, Fréchet, and Gumbel family, respectively. We fit a GEV distribution, at each gridpoint over land, to the sample of annual extremes and derive the 20-year return value from the quantile function

$$z_T = \begin{cases} \mu - \sigma/\xi \left[1 - (-\log(1 - 1/T))^{-\xi} \right] & \xi \neq 0, \\ \mu - \sigma \log[-\log(1 - 1/T)] & \xi = 0, \end{cases} \quad (2.1)$$

evaluated at $T = 20$. For cold extremes, we fit the GEV distribution to the negative of the sample of annual T_{min} extremes and reverse the sign of the estimated return value. To compare model parameters and return level estimates with estimates from the verification datasets, the former estimates were regridded to match the coarser grids of the latter datasets.

The justification of the GEV distribution as an appropriate representation of the behavior of annual extremes follows from an important result in extreme value theory, which states that the limiting distribution of the maximum of a sufficiently large random sample belongs to only one of the 3 extreme value distributions [61]. However the daily observations of maximum and minimum

temperature, from which the samples of annual extremes are computed, exhibit features typical of environmental datasets such as an annual cycle and serial correlation that violate the assumptions defining a random sample. Indeed, the annual cycle negates the condition that the observations are identically distributed while the presence of serial correlation implies dependence among consecutive observations. Despite this, empirical studies have shown the GEV distribution to be a good candidate distribution to describe environmental extremes [52, 53, 54, 96].

Since the model simulations are forced with observed boundary conditions which exhibit trends over the simulations period, we use the non-parametric Mann-Kendall test [17] to investigate whether such trends are captured in the annual extremes of daily temperature. Unsurprisingly, the test confirms the existence of trends at many gridpoints, in agreement with the large number of observational studies that have identified trends in such temperature statistics at both global [32] and regional scales [99, 56, 15].

A common approach to address this form of non-stationarity is to assume that the parameters of the GEV distribution are time dependent [20]. [52] examined the time dependence of GEV parameters in a transient climate setting and found that a model that allowed for time-variation in the location and scale parameters, in the form of a linear and log-linear trend, respectively, best represented the transient behavior of temperature extremes. In a similar manner, we fit several models with and without time-varying parameters, and as in [52] likelihood ratio tests rejected models that assumed time-variation in the shape parameter, but unlike their work, the datasets did not provide evidence in support of a time-varying scale parameter. Consequently, we performed the anal-

yses described below assuming a GEV distribution with and without a linear trend in the location parameter and found that the differences were generally minor and did not alter qualitatively the conclusions. Therefore, in what follows we present the results pertaining only to the stationary GEV model.

[43] compared the short-sample performance of maximum likelihood and the method of L-moments in the estimation of the upper quantiles of the GEV distribution, and demonstrated that for a broad range of shape parameter values, the estimates by the latter method showed lower root-mean-square error relative to those of the former. Since then, the method of L-moments has been widely used in observational and simulation studies requiring the estimation of return values from short samples [53, 96]. However, this method does not permit the estimation of time-varying GEV parameters as the computation of L-moments requires that the random sample be identically distributed. The above non-stationary GEV distribution were therefore fit by maximum likelihood, and for consistency, it was also used in the estimation of the stationary GEV parameters. One of the benefits of maximum likelihood is that approximate standard errors of the estimated parameters can be obtained by using the inverse of the observed information matrix; but these approximations tend to be unreliable for small sample sizes [52].

An alternative approach to quantify the uncertainty in parameter estimates is through resampling [34]. In this study, we use a bootstrapping technique to generate 500 resampled replicates from the original sample of annual extremes, and to each replicate fit a GEV distribution and derive return levels, from which standard errors may be computed. Assuming that the original dataset is represented by an $m \times n \times N$ array, with m , n and N denoting longitude, latitude and

time, respectively, each new sample corresponds to one of $N m \times n$ matrices, in order to preserve the spatial dependence structure.

Several of the analyses rely on hypothesis tests performed at each grid point, as is the case, for example, of likelihood ratio tests where the validity of one model is measured relative to another, or in the evaluation of the statistical significance of observed trends as in the Mann-Kendall tests above. However, the interpretation of these results in a spatial context, where there is spatial dependence, is often misleading [64]. A field significance test is a popular statistical technique designed for the simultaneous evaluation of multiple hypothesis tests, usually specified over geographic areas. Such a test may be interpreted as a type of meta-test, as the data being tested are the results of individual or local tests, and the null hypothesis is that all of the individual null hypotheses are true [97]. Because of the complex dependence structures found in environmental datasets, the sampling distribution of test statistics such as the number of tests that are significant is difficult to derive analytically. Fortunately, good approximations may be obtained by means of resampling. We first compute the desired test statistic based on the results of the original collection of annual extremes, and subsequently generate 500 replicates of annual extremes, in the manner described above, and on each replicate perform the individual hypothesis tests and compute the test statistic. The procedure yields 500 values of the test statistic, from which the significance under the null hypothesis of the original value may be determined.

2.1.4 Results

Demonstration on a simple example

We begin the analysis with a simple example that helps illustrate how differences in the mean and variance of an idealized distribution of temperature affects the distribution of the associated extremes (Fig. 2.1). Similar schematics have been developed since the IPCC Third Assessment Report to characterize the effect of a changing climate on the daily temperature distribution (e.g. Fig. 2.32 in [36]). We assume that the distribution of temperature is well-described by a Gaussian distribution and consider a scenario with two distributions having the same variance but the mean of one is slightly larger than the other, as would be the case, for instance, of a warm bias of CAM relative to ERA, and another, where the means are the same but the distribution of CAM exhibits a higher variance. Analytically or through Monte Carlo simulation, we can derive the GEV distribution of the maximum and minimum of a large sample from each of these distributions, representing the distributions of the annual maximum and minimum of T_{max} and T_{min} , respectively. The GEV pdfs of warm and cold extremes for the first scenario are shown in Figs. 2.1(c) and (e), respectively. In each plot, the distributions of CAM and ERA are identical except that that of CAM is shifted to the right by a quantity equal to the difference of the means of T . Thus a shift in the distribution of T induces an identical shift in the distributions of both warm and cold extremes. In the second scenario, despite the means of T being the same, the pdf of CAM for warm extremes exhibits both a shift towards warmer temperatures and an increase in variability relative to that of ERA, while that of cold extremes is shifted towards colder temperatures with an identical increase in variability (Figs. 2.1(d) and (f)). An

aspect that we will examine repeatedly in this study is the role of differences in the T_{max} and T_{min} climatologies on the respective pdfs of extremes. Because we have implicitly neglected any form of non-stationarity, such as the presence of an annual cycle, these climatologies correspond here to the means of the respective GEV distributions. Adjusting for these mean differences in the first scenario (Figs. 2.1(g) and (i)) removes the shift observed above and renders the pdfs identical, while in the second, the mean differences are canceled, however the discrepancies in variability persist (Figs. 2.1(h) and (j)).

We apply these ideas to the GEV pdfs of T_{max} annual extremes of CAM4 and ERA over land regions, with parameters set to area-averages of grid-point estimates (Fig. 2.2). Parameter values are denoted in the panel. At this spatial scale, shape parameter estimates are remarkably similar across the two datasets, namely, -0.27 and -0.28 for CAM4 and ERA respectively, suggesting that the tail behavior is well captured by the model. The negative values indicate that the distributions of T_{max} annual extremes can be approximated by those from the Weibull family, which is characterized by a bounded upper tail. We note that the quantile function of the GEV distribution, as depicted in equation (2.1), for the case of $\xi \neq 0$, is a function of all 3 GEV parameters, but by holding ξ constant, quantile differences may be conveniently decomposed in terms of differences in location and scale parameter estimates. Furthermore, given the functional dependence of the variance of the GEV distribution (not shown) on only the shape and scale parameters, a constant shape parameter implies that differences in scale parameter estimates may be interpreted as differences in interannual variability. Therefore, differences in 20-yr return levels may be analyzed in terms of differences in the location parameter, which is measure of central tendency, and interannual variability of temperature extremes. Here we observe that the re-

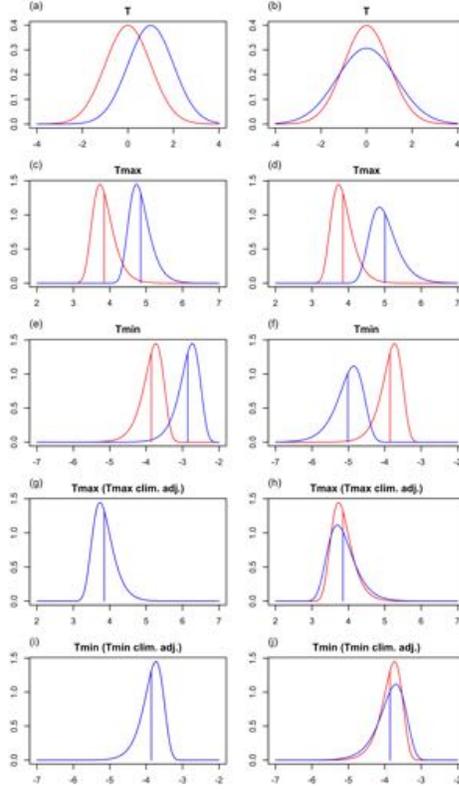


Figure 2.1: Idealized Gaussian distributions of temperature T (top) for CAM (blue) and ERA (red) and the corresponding GEV probability density functions obtained by drawing 5000 samples of length 10000 from the distributions above and fitting GEV distributions to the resulting 5000 extreme observations. A standard Gaussian distribution is assumed for ERA T . The left panel considers T distributions with the same standard deviation, but the climatological mean of CAM is shifted by 1, while the right panel considers T distributions with the same mean, but the climatological standard deviation of CAM is 1.3. Vertical lines denote the mean of the GEV distribution. GEV distributions adjusted by their T_{max} (g, h) and T_{min} (i, j) climatologies.

turn level estimate of CAM4 exceeds that of ERA by 2.25°C (Fig. 2.2(a)), which is largely explained by the difference of 1.88° in the location parameter estimates, while the contribution from the scale parameter difference is only 0.36°C . Note how this value does not correspond to the actual difference in scale parameters, rather, it is the difference scaled by a factor that depends on the value of the

shape parameter, which we assume to be the same for both datasets. However, due to the slight difference of 0.01 in the shape parameter estimates, the decomposition is not exact, as seen by the 0.01°C discrepancy between the actual 20-yr return level difference and the sum of the two contributions. Comparing the two distributions, the GEV distribution of CAM4 appears shifted to right and slightly wider than that of ERA, reflecting the larger magnitude of the location and scale parameter estimates in the CAM4 simulation.

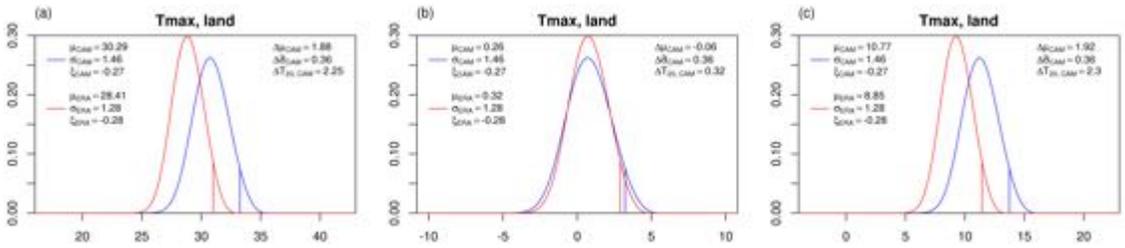


Figure 2.2: GEV pdfs of CAM4 (blue) and ERA (red) with parameters values (left legend) set to the area-averaged estimates over land regions, with gridpoint estimates corresponding to averages of parameter values obtained from 500 resampled replicates of the annual extremes of daily maximum temperature. Location and 20-yr return level differences are indicated in the right legend, along with differences in a modified scale parameter, defined as $\Delta\tilde{\sigma}_s = -(\sigma_s - \sigma_{ERA})f(\xi_{ERA})$, with f obtained from equation (2.1), and s representing the CAM dataset. Probability density functions adjusted by the area-averaged maximum of the annual cycle $\max \bar{T}_{max}^{ac}$ (b) and $\max \bar{T}_{mean}^{ac}$ (c). Vertical lines denote 20-yr return levels.

Adding a constant to a sample of annual extremes will result in an identical change in the value of the location parameter, as in the mean, while leaving the shape and scale parameters unchanged. Therefore, since the quantile function (2.1) is linear in the location parameter, return levels will reflect any systematic differences in the extreme temperatures between model and reanalysis. We consider two potential sources of systematic biases, T_{max} and T_{mean} climatologi-

cal differences. The T_{max} (T_{mean}) annual cycle is defined as the 1979-2010 average for each calendar month of monthly T_{max} (T_{mean}). Because annual warm extremes tend to coincide with the maximum of the warm extreme climatology and mean temperature climatology, adjusting for these climatologies can be done by subtracting from these the $\max \bar{T}_{max}^{ac}$ and $\max \bar{T}_{mean}^{ac}$, respectively. After adjusting for the T_{max} climatology (Fig. 2.2(b)) the distributions are quite similar, as the difference in location parameter estimates of Fig. 2.2(a) becomes negligible, and the difference of 2.25°C in the return values is reduced to 0.32°C, which coincides almost exactly with the discrepancy in scale parameters. Adjusting for the T_{mean} climatology, however, does little to improve the correspondence in return values between the two datasets, on the contrary, the difference is increased to 2.3°C, as the discrepancy in location parameters increases from 1.88°C to 1.92°C. The similarity in the distributional pattern depicted in the first panel of Fig. 2.2 with that of Fig. 2.1(d), together with the relative invariance of the location parameter difference to the adjustment in mean biases, provide the first indications that the distributions of T of CAM4 and ERA are related qualitatively as in Fig. 2.1(b).

Temperature extremes in CAM4, ERA and observations

Warm and cold extremes of CAM4 and differences with the corresponding temperature extremes of ERA and HadEX2 are displayed in Fig. 2.3. Significant positive differences are seen over land regions in comparison to ERA, in particular over midlatitudes in the Northern Hemisphere and in the subtropics of South America, with the exception of central Africa and the Arabian Peninsula where slight negative differences emerge (Fig. 2.3(c)). Over the Midwestern United States return levels exceed 40°C in CAM4, while the values in ERA are roughly

10-15°C smaller. The limited spatial coverage of the HadEX2 TXx index restricts the scope of the comparison primarily to the Northern Hemisphere. Positive differences of similar magnitude can be seen over central North America and western Eurasia, where return value differences between model and observations also differ by up to 10°C (Fig. 2.3(e)). However, coherent regions of colder warm extremes emerge over the Himalayas and North America extending into the Greenland ice sheet.

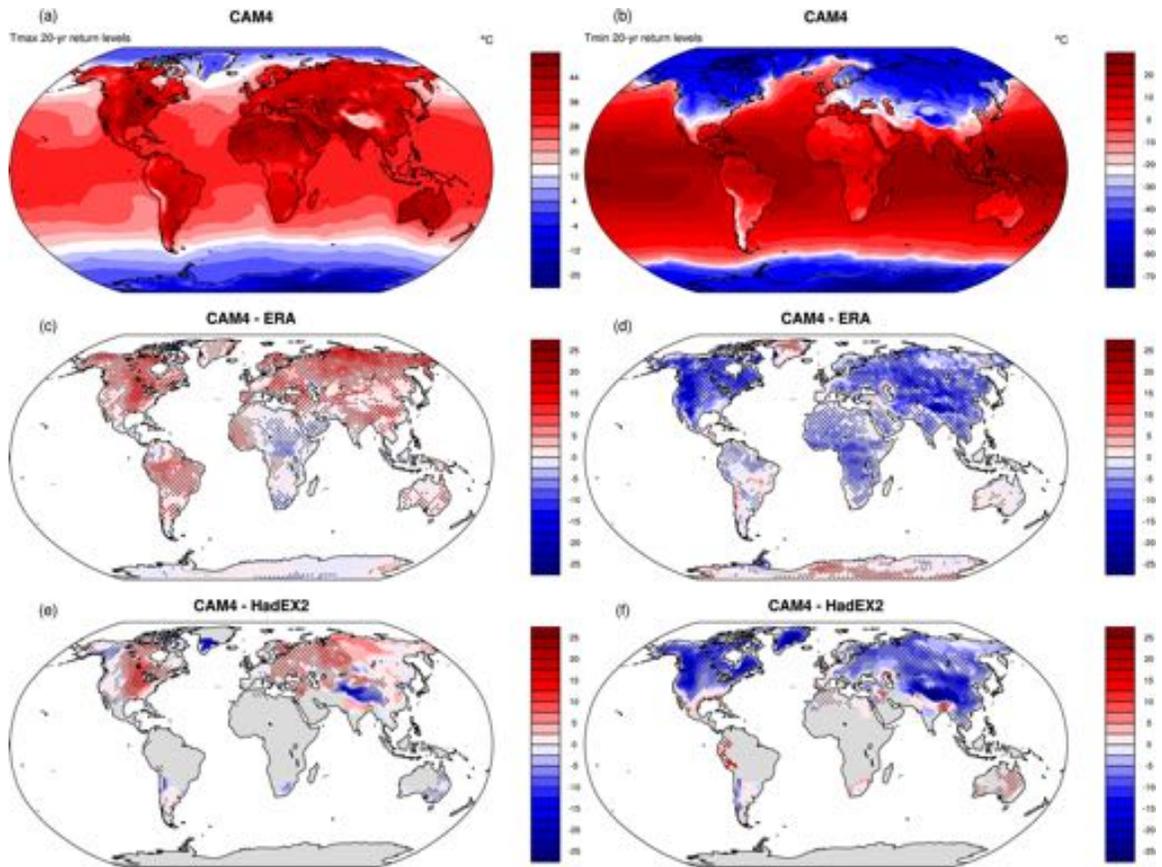


Figure 2.3: CAM4 20-yr return level estimates of 1979-2010 annual extremes of daily maximum (left) and minimum (right) temperature, and the corresponding differences with estimates from ERA Interim (middle) and the HadEX2 observational dataset (bottom). Stippling indicates return level differences are significant at a 5% level.

Cold extremes are well simulated over Australia and most of South America, but are significantly colder over most of the Northern Hemisphere (Fig. 2.3(d)). Area-averaged discrepancies over North America and Asia exceed 5°C . Similar differences appear in the comparison with HadEX2, with the exception of Greenland (Fig. 2.3(f)). Indeed, simulated cold extremes are generally in good agreement over ice-covered land regions relative to reanalysis, in contrast with HadEX2 where severe negative differences extend into the Greenland ice sheet. We note that the spatial coverage is better for the HadEX2 TNn index.

We investigate the degree to which the differences in warm and cold extremes between CAM4 and ERA can be explained by T_{max}/T_{min} climatological differences. Figure 2.4 displays CAM4 warm and cold extremes, adjusted by $\max \bar{T}_{max}^{ac}$ and $\min \bar{T}_{min}^{ac}$ respectively, and the differences with the corresponding adjusted ERA extremes. Warm extremes are in good agreement after the adjustment, although over the Midwestern United States, Australia and China the climate model indicates return values up to 6°C higher than ERA, but these values are not always statistically significant. In contrast, large coherent regions where CAM4 overestimates cold extremes remain after the adjustment, particularly over western Eurasia and North America. However, area-averaged differences of 0.8°C , 1.2°C , 2.1°C , over North America, Europe and Asia, respectively, highlight the notable reductions in cold extreme return-level discrepancies between the two datasets.

In summary we note that in large areas, particularly in the Northern Hemisphere, excluding the Greenland ice-sheet and the Himalayas, the climate model generally overestimates both warm and cold extremes. This is consistent with our hypothesis that CAM4 exhibits greater variability in T relative to ERA, as

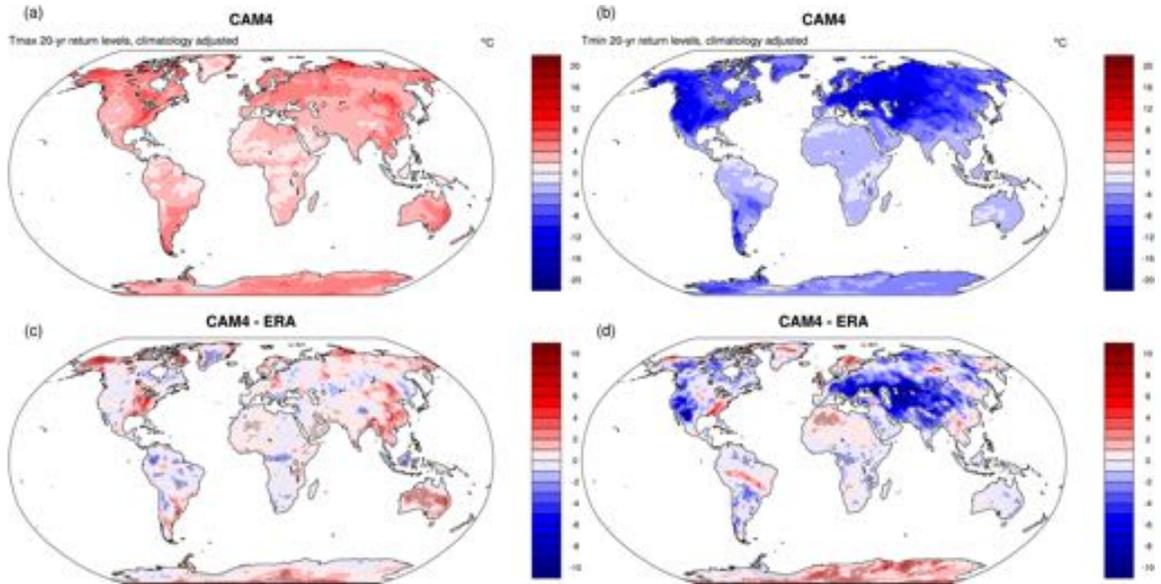


Figure 2.4: CAM4 20-yr return level estimates (top) of 1979-2010 annual extremes of daily maximum (left) and minimum (right) temperature, and the corresponding differences with estimates of ERA Interim (bottom). Return level estimates are adjusted by the corresponding extremes of the annual cycle, $\max \bar{T}_{max}^{ac}$ and $\min \bar{T}_{min}^{ac}$.

this overestimation is qualitatively depicted in panels (d) and (f) of Fig. 2.1. Adjusting by the respective extreme climatologies leads to substantial reductions in the noted discrepancies although slight differences remain, specially for cold extremes.

Impact of stochastic parameterizations on extreme events

Next, we analyse the impact of adding a stochastic parameterization on the model-issues discussed above. Overall, warm and cold extremes are quite similar to those of unperturbed CAM4 (Fig. 2.13). The signature of SPPT and SKEBS on extreme events is remarkably similar, specially when considering that SKEBS

is most active in the extratropics, while SPPT has the largest impact on the near surface fields in the tropics [8, 5]. This suggests that the stochastic parameterizations may excite modes of variability already present in CAM4 rather than impose their specific characteristics on the distribution of extreme events, which is further substantiated in a forthcoming study. Discrepancies are apparent only by direct comparison with CAM4 (Fig. 2.5). Warm extremes simulated by SKEBS are considerably warmer over parts of the Northern Hemisphere, particularly over western Asia (Fig. 2.5(a)), in some parts in excess of 8°C. Replacing SKEBS with SPPT results in a nearly identical spatial pattern, except that the positive differences are less pronounced (Fig. 2.5(c)). Cold extremes for both SKEBS and SPPT relative to CAM4, present a less coherent spatial pattern, but statistically significant positive differences are found over most continents, which help reduce the overestimation of cold extremes mentioned above and thus lead to a better agreement with reanalysis.

To assess if these differences between CAM4 and the stochastic parameterizations are collectively statistically significant, or could be the result of sampling variability perhaps due to the limited sample sizes, we test the null-hypotheses that the CAM4 and SKEBS, and CAM4 and SPPT, simulations are realizations from a single data-generating process with the same extreme value distribution. This analysis is carried out at each gridpoint, by performing a likelihood-ratio test involving the likelihood of a GEV distribution fit separately to each time series of annual extremes, and the likelihood of a GEV distribution fit to the concatenated time series. A field significance test, as described in section 2.1.3, is then performed to assess whether the results of the individual tests are significant at a regional level. We note that under the null hypothesis the concatenated time series represents a sample drawn from the same GEV distribution, there-

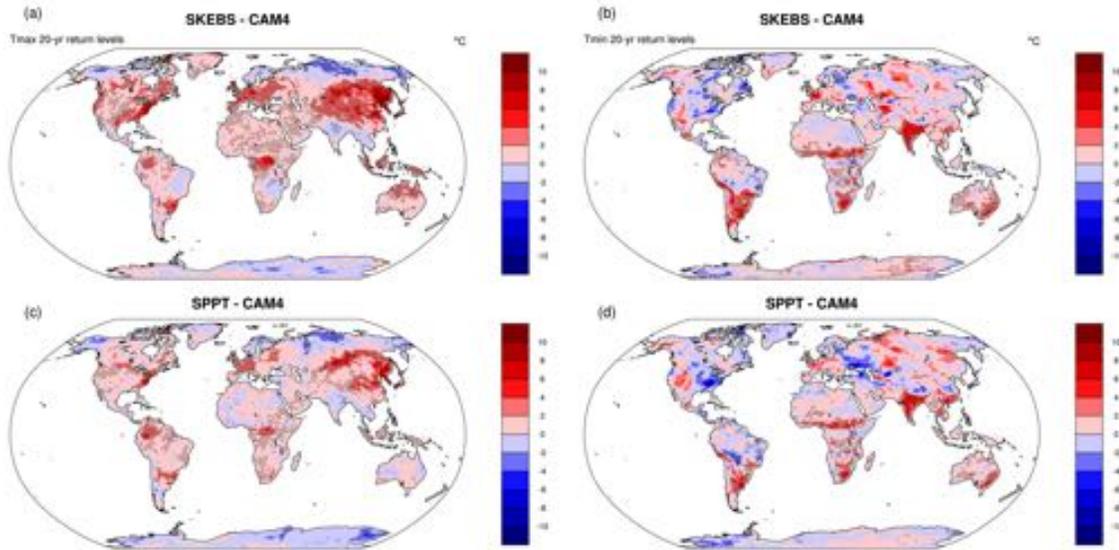


Figure 2.5: Differences of 20-yr return level estimates of 1979-2010 annual extremes of daily maximum (left) and minimum (right) temperature between CAM4 and the stochastic parameterization schemes, SKEBS (top), SPPT (bottom).

fore, in the resampling stage, replicates are generated by selecting at random $2N$ $m \times n$ matrices from the concatenated $m \times n \times 2N$ array. We use seven continental regions: North America, South America, Europe, Africa, Asia, Australasia and Antarctica; displayed in Fig. 2.6, each a combination of sub-continental scale regions defined in [79]. The null hypotheses that warm extremes simulated by the two pairs of simulations belong to the same GEV distribution are strongly rejected in all regions except Antarctica. The analogous hypotheses for cold extremes are strongly rejected in all regions.

We further assess, as was done earlier for the return level discrepancies between CAM4 and ERA, to what degree the observed warm and cold extreme differences between reanalysis and the stochastic parameterizations can be attributed to the respective climatologies (Fig. 2.14). The spatial patterns after ad-

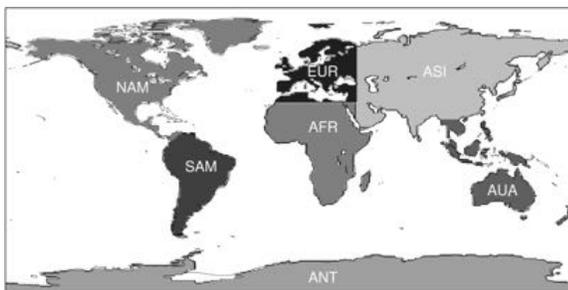


Figure 2.6: Continental regions adapted from the sub-continental definitions of [79] (as in Fig. 10.7 of [12]).

justing for the respective extreme climatologies closely resemble those of CAM4 relative to reanalysis. A comparison with CAM4 is displayed in Fig. 2.7. Warm extremes of both parameterizations show very close agreement with those of CAM4, as do cold extremes in the Southern Hemisphere. However, a heterogeneous spatial pattern of areas of warm and cold differences, though mostly not statistically significant, emerge for cold extremes over the Northern Hemisphere, with the exception of a large area of warmer extremes over Northwestern Asia, which is most clearly depicted in the SPPT and CAM4 comparison.

Overall, the effect of introducing a stochastic parameterization, with its tendency to enhance the overestimation of warm extremes in CAM4, while mildly reducing that of cold extremes, is conceptually consistent with panels (c) and (e) of Fig. 2.1, as the first effect may be represented by a distribution of warm extremes shifted towards warmer temperatures, while the second, by a similar though more moderate shift of the distribution of cold extremes. This would suggest that SKEBS introduces a warm bias in the distribution of T of CAM4, rather than augment its variability.

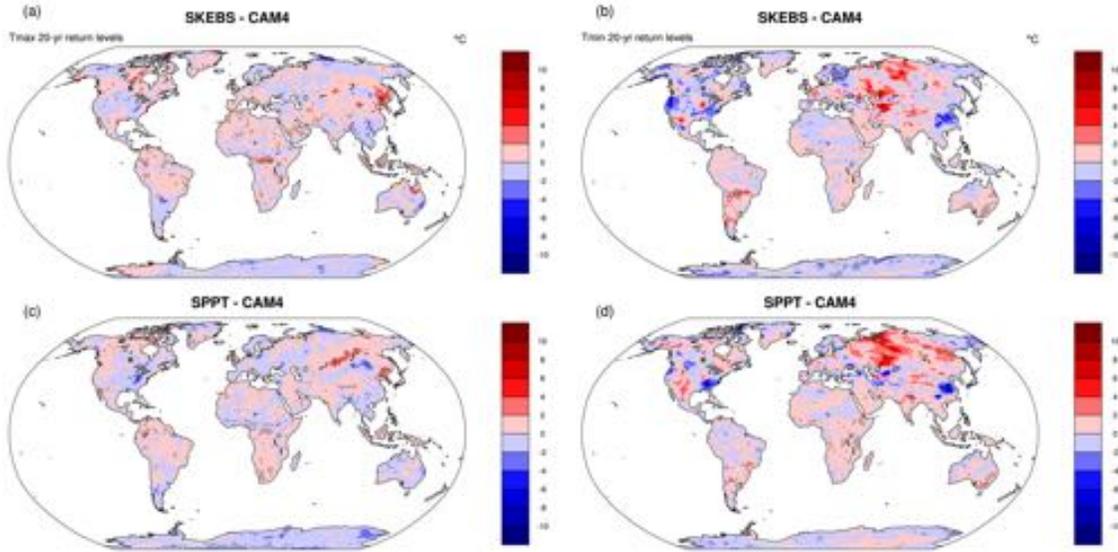


Figure 2.7: Same as in Fig. 2.5, except simulations have been adjusted by the respective extremes of the T_{max} and T_{min} annual cycles.

Regional analysis of GEV pdfs

In this section we aim to characterize the discrepancies of temperature extremes observed above in terms of distributional differences in global and regional-scale GEV pdfs, as was done in section 2.1.4. Figures 2.8(a) and (b) display the pdfs of CAM4, SKEBS and ERA for T_{max} and T_{min} annual extremes, respectively, over land regions. The GEV parameters at each gridpoint correspond to averages over the 500 resampled estimates. Due to space considerations, we present results only for SKEBS as those of SPPT are comparable. Differences in parameter estimates are highly significant, as indicated by 2 sample t-tests [97], except for most of the shape parameter estimates. Henceforth all parameter differences should be interpreted as statistically significant at a 5% level except where otherwise noted. As was noted in section 2.1.4, at these scales the shape parameter estimates for warm extremes exhibit little variation across datasets; here we con-

firm that this similarity extends to cold extremes as well, indicating that the tail behavior is well captured by the models in a more general sense. This close agreement between shape parameter estimates supports the use of the decomposition of return level differences in terms of location and scale discrepancies, as discussed above.

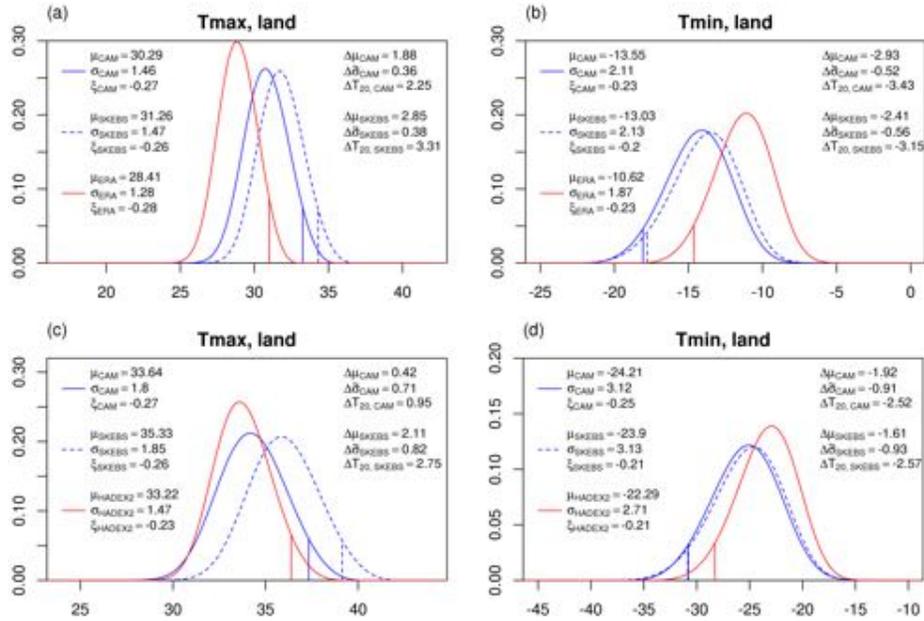


Figure 2.8: GEV pdfs, as in Fig. 2.2(a), of CAM (solid blue), and SKEBS (dashed blue) for (left) warm and (right) cold extremes over land, with respect to (top) ERA (red) and (bottom) HadEX2 (red). Parameter values in the HadEX2 comparison correspond to area-averages within 15-72.5°N, excluding Greenland.

Figure 2.8(a) adds to Fig. 2.2(a) the SKEBS GEV pdf curve, which appears displaced to the right of that of CAM4, with only a minor increase in the scale parameter. Thus the overestimation of warm extremes relative to CAM4 reported in Fig. 2.5(a) can be interpreted as the result of a shift in the distribution of CAM4 towards higher temperatures, rather than an increase in interannual variability. The displacement gives rise to a return level difference of 3.3°C against ERA compared to the 2.5°C difference of CAM4. Disagreements of simi-

lar magnitude apply to cold extremes, with differences in return levels of -3.4°C and -3.2°C , for CAM4 and SKEBS, respectively, relative to ERA, which are related predominantly to discrepancies in location parameter values. Despite the considerable increase in interannual variability in the models and reanalysis, as compared with warm extremes, the distributional differences between CAM4 and ERA are consistent with Fig. 2.1(f), while the slight displacement to the right of the SKEBS distribution is consistent with Fig. 2.1(e).

In Figs. 2.8(c) and (d) we repeat the analysis above using HadEX2 as the reference dataset. In order to minimize potential distortions arising from observational uncertainties, the averages of the GEV parameters were taken over land regions within $15\text{-}72.5^{\circ}\text{N}$, excluding Greenland. Despite the difference in land definition, the comparison with HadEX2 yields qualitatively similar results. However, due to its Northern-hemispheric focus, extremes show significantly higher interannual variability across the three datasets. Discrepancies in scale parameter estimates also increase in magnitude, particularly for cold extremes, which leads to these differences playing a more important role in explaining differences in return level estimates.

A more detailed examination of regional differences in warm extremes between the models and ERA is presented in Fig. 2.9. The regions correspond to those displayed in Fig. 2.6, but we exclude Antarctica. With the exception of Africa, where there is a notable similarity between the various pdfs, the distributional pattern observed for warm extremes (Fig. 2.8(a)) is largely reproduced across the regions. Discrepancies in return levels are more severe over the Northern Hemisphere, but display greater uniformity compared to those of the Southern Hemisphere. For instance, differences between CAM and ERA

range between 3.55-3.68°C in the former, compared to -0.23-2.94°C in the latter. The hemispheric differences in the magnitude of return level discrepancies are tied to larger differences in both location and scale parameter estimates. Note that mild discrepancies in shape parameter estimates arise in some regions, such as in Europe and Australasia, which diminish the accuracy of the decomposition. However, it is of little concern as in such cases the variation in return levels is still primarily driven by location parameter differences.

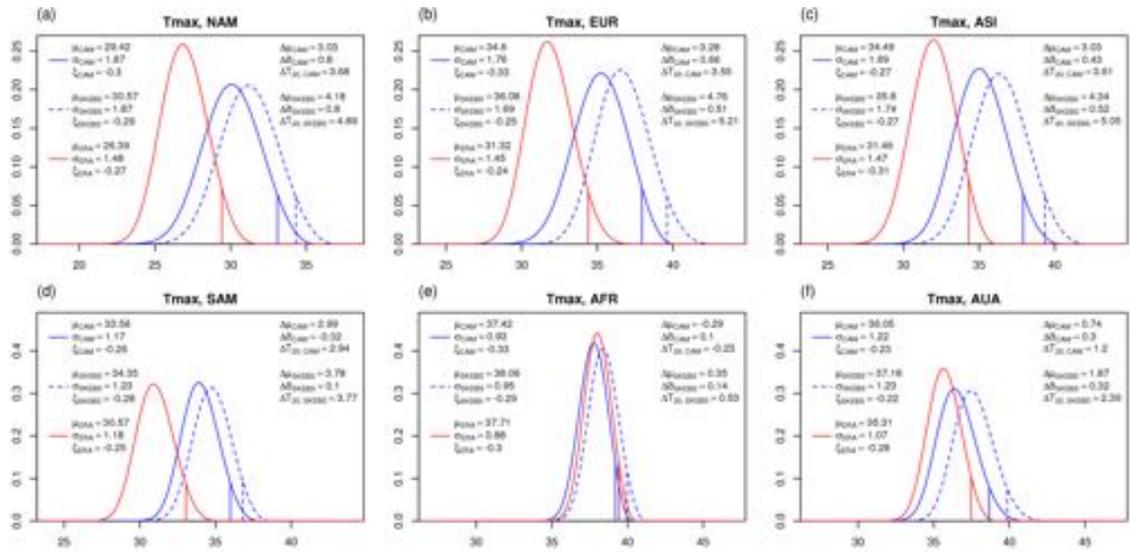


Figure 2.9: GEV pdfs of T_{max} annual extremes, as in Fig. 2.8(a), with parameter values area-averaged over continental regions defined in Fig. 2.6.

The distributional patterns for cold extremes over the continental regions in Fig. 2.10 are qualitatively similar to that observed at the global scale (Fig. 2.8(a)), with minor differences over South America and Australasia, where the discrepancy in return levels between CAM4 and ERA is so small that the right shift induced by SKEBS either cancels it, or enhances it after it switches sign. The signature of SKEBS on the distribution of annual cold extremes is almost indistinguishable from that of CAM4 over the Northern Hemisphere, particularly

over North America, where the close agreement in the location and scale parameter estimates is such that the difference in return levels results is due largely by the slight discrepancy in the shape parameter estimates.

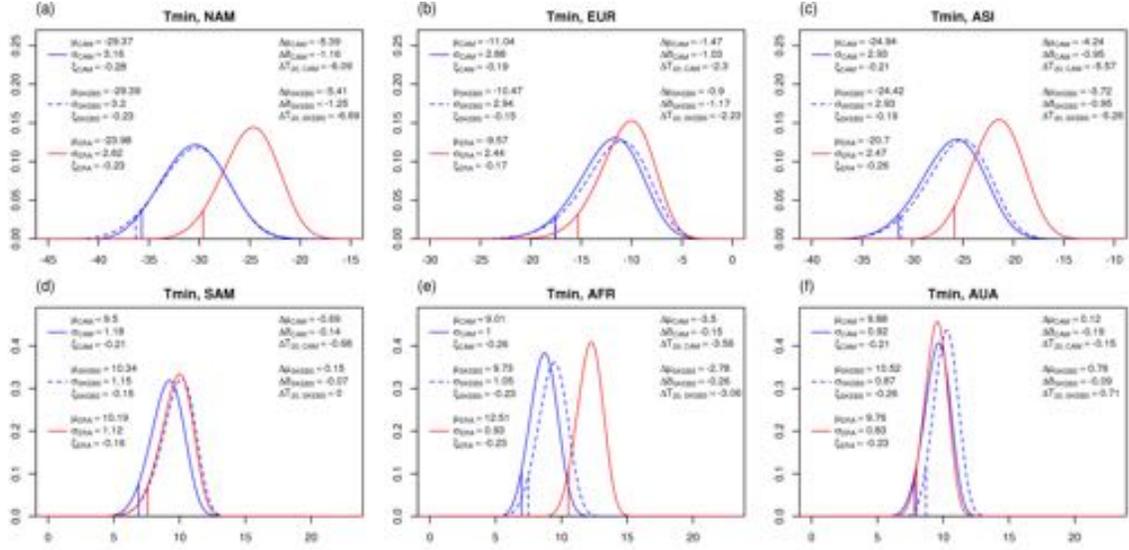


Figure 2.10: As in Fig. 2.9, with T_{min} annual extremes.

In general, the impact of SKEBS on the distribution of annual extremes of CAM4 is largely limited to changes in the location of the distribution, with minor effects on the tail behavior and interannual variability. As noted above, this is consistent with SKEBS shifting the distribution of T of CAM4 towards warmer temperatures (Fig. 2.1(a)). We explore this further by adjusting the regional pdfs examined above of both warm and cold extremes, by their respective T_{mean} climatologies, i.e. $\max \bar{T}_{mean}^{ac}$ and $\min \bar{T}_{mean}^{ac}$ respectively. We note that such an adjustment affects only the location of the distributions. In Fig. 2.11, for space reasons, we focus only on Northern Hemisphere warm extremes and Southern Hemisphere cold extremes. Over the Northern Hemisphere, the reductions in location parameter differences between CAM4 and ERA are at most 1.6°C , but the reductions between SKEBS and ERA are consistently larger, with

values in excess of 2°C in Asia and Europe, represented by shifts of the SKEBS distributions towards colder temperatures relative to ERA. This effect considerably reduces the location parameter differences between the SKEBS and CAM4 pdfs to within 0.5°C . Over the Southern Hemisphere, the adjustment also leads consistently to an improved agreement in location parameter estimates between SKEBS and CAM4. These results provide further evidence that SKEBS introduces a warm bias in the distribution of T of CAM4.

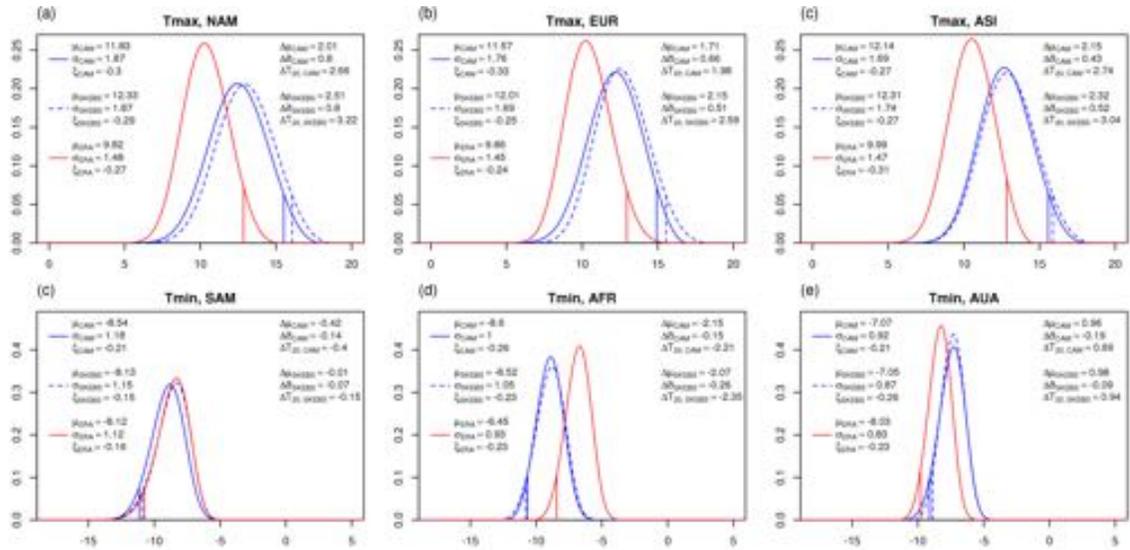


Figure 2.11: GEV pdfs as in Fig. 2.9 and Fig. 2.10, of T_{max} annual extremes over Northern Hemisphere land regions (top), and T_{min} annual extremes over Southern Hemisphere land regions (bottom); adjusted by the respective extreme of the monthly mean temperature annual cycle, \bar{T}_{mean}^{ac} .

In Fig. 2.12 we extend the analysis of the previous figure by adjusting the warm and cold extreme distributions by their respective T_{max} and T_{min} climatologies. Across extremes and regions, the adjustment largely cancels the location discrepancies between the CAM4 and reanalysis distributions, with values less than 0.5°C , and similar reductions apply to the discrepancies between SKEBS

and CAM4. The resulting discrepancies in return values relative to reanalysis fall within 0.8°C . Note how the adjustment by the annual extreme climatology has the compound effect on the SKEBS distribution of correcting for both the warm bias in T relative to CAM4, and the enhanced variability of CAM4 relative to ERA.

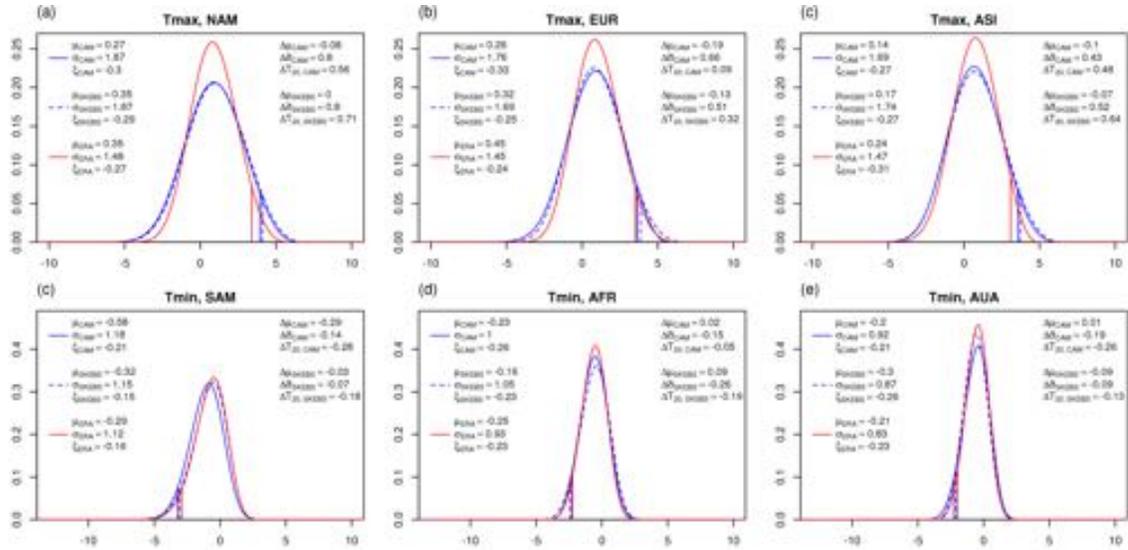


Figure 2.12: GEV pdfs, as in Fig. 2.11, but adjusted by the respective extreme of the T_{max} and T_{min} annual cycle.

2.1.5 Discussion and Conclusions

This work investigates the skill of the general circulation model CAM4 in simulating annual extremes of near-surface temperature. Previous studies have demonstrated that the extreme statistics of temperature over the late 20th century in many general circulation models and observationally-constrained datasets agree reasonably well, although there is substantial sensitivity on the choice of the reference dataset [53, 54, 81]. In general, these studies report signif-

icant improvements in the agreement after correcting for systematic differences in the mean conditions of temperature extremes.

Temperature extremes are evaluated in terms of 20-yr return levels, and compared against those estimated from the ERA Interim reanalysis and the gridded land-based HadEX2 observational dataset. Our results indicate that CAM4 overestimates both warm and cold extremes over land regions, particularly over the Northern Hemisphere when compared against reanalysis. These differences appear to be more pronounced than those found by [53] in a similar assessment using the ensemble mean of models participating in the IPCC AR4 diagnostic exercise and ERA40, and the more recent [54] using the CMIP5 multi-model ensemble median and ERA Interim. Similar spatial patterns, though less spatially coherent, emerge relative to the HadEX2 dataset. Colder warm extremes, however, arise in high-terrain regions, such as Greenland and the Tibetan Plateau, but these discrepancies are likely the result of biases in the station data measurements, as similar biases were reported in [81] when comparing CMIP5 and HadEX2 TXx climatologies.

We interpret these return level discrepancies in terms of differences in global and regional-scale GEV pdfs, obtained by spatially averaging gridpoint GEV parameter estimates. At these scales, estimates of the shape parameter, which controls the tail behavior of the GEV distribution, are very similar across all datasets for both cold and warm extremes. This allows the analysis to focus on distributional differences based on the two remaining GEV parameters, namely, the location and scale parameters, where the first represents a measure of central tendency while the second, in the present context, is intimately related to the interannual variability of annual extremes. CAM4 GEV pdfs exhibit more

extreme return values relative those of the verification datasets in agreement with the overestimation found at the gridpoint level. The discrepancies originate primarily from differences in location parameter estimates, as represented by a shift of the CAM4 pdfs towards higher (lower) temperatures for warm (cold) extremes. Interannual variability is significantly higher for cold extremes in all datasets. CAM4 consistently overestimates the magnitude of the scale parameter, indicating that interannual variability in annual temperature extremes in CAM4 is too large when compared to observations and reanalysis.

The differences in location parameters are negligible if the annual extremes of the model and reanalysis are adjusted by their respective climatologies of monthly extremes. When adjusted in this way, the agreement between the global and regional GEV pdfs is very good, although the overestimation of the scale parameter persists, particularly for cold extremes over Northern-hemispheric regions. We stress that the match with reanalysis data can only be achieved when adjusting with the bias of the T_{max}/T_{min} climatologies, which is an extreme statistic. When debiasing with monthly mean temperature the discrepancy between reanalysis and model remains, confirming that it is not the mean warm/cold bias over land which explains the differences in return values. We note that this overestimation of temperature extremes is also reflected in the diurnal temperature range between both datasets. Considering differences of the 1979-2010 average of the monthly mean diurnal cycle between CAM4 and ERA at the month where the T_{max} climatology is extreme (not shown), a good correspondence exists with warm extreme return level differences over land (spatial correlation of 0.58); though the relationship does not hold as well at the T_{min} climatology extreme.

To evaluate the impact of missing subgrid-scale variability the effect of two stochastic parameterization schemes was studied: a stochastic kinetic energy backscatter (SKEBS) scheme and stochastically perturbed parameterization tendency (SPPT) scheme. Including a stochastic parameterization noticeably increased the magnitude of warm extremes, while reducing that of cold extremes. This is contrary to the effect of adding a white or red noise process to a linear system, which would result in an increase in variability and consequently higher return levels for both T_{max} and T_{min} annual extremes [9, 38]. Since CAM4 already overestimates extremes, the effect of adding a stochastic parameterization is beneficial for cold extremes but adverse for warm extremes. However, neither of the parameterizations schemes meaningfully reduce the overestimation of temperature extremes in CAM4. Unexpectedly, the impact of the two schemes was very similar, although SKEBS is typically most active in the mid-latitudes, while SPPT tends to have the biggest impact in the tropics, where convection leads to large tendencies, and hence perturbations, in the physical parameterizations. Although different in nature, both schemes seem to excite modes of variability in CAM4 in such a way that the response in the extremes is the same. [48] and [4] provide an example of another process, namely Northern-Hemispheric blocking, where adding a stochastic parameterization does not change its structure, rather only its relative frequency.

Our findings can be best summarized using the schematics shown in Fig. 2.1. Comparing the model and reanalysis, the distributional differences between annual warm and cold extremes closely resemble Figs. 2.1(d) and (f), respectively, suggesting that the distribution of T in the model exhibits too much variability relative to that of reanalysis. In contrast, the addition of a stochastic parameterization to the model does not induce significant changes on either the shape

or scale parameter; instead the distributions of extremes are shifted towards warmer temperatures, as displayed conceptually in Figs. 2.1(c) and (e). This suggests that the stochastic schemes introduce a systematic bias in the mean, rather than enhance the variability of T .

Adjusting for biases in the extremes, leads to negligible differences in the location parameters of both perturbed and unperturbed simulations and results in extremes comparing much better with reanalysis. However, adjusting for the bias in mean temperature has a muted effect on the discrepancies in extremes between the models and reanalysis. We conclude that CAM4 misses an important aspect of temperature extremes, namely, the mean statistics of temperature extremes. These can be easily adjusted for in historical data, but not necessarily for projections in a changing climate, thus, more attention should be given to this aspect when using climate models for predicting extremes.

The analysis of temperature extremes conducted in this study was based on annual extremes of daily temperature observations, but recently a large body of research has focused on less extreme measures of extreme climate behavior, usually based on climate indices. Such a family of indices was recently defined by the ETCCDI and formed the basis of the HadEX2 observational dataset, which was featured prominently in the latest IPCC assessment report as providing a critical line of evidence into how climate extremes have been evolving in response to anthropogenic forcings. An assessment of the performance of the stochastic parameterization schemes examined here in reproducing temperature and precipitation extremes represented by these indices will be the subject of the next chapter.

2.1.6 Appendix

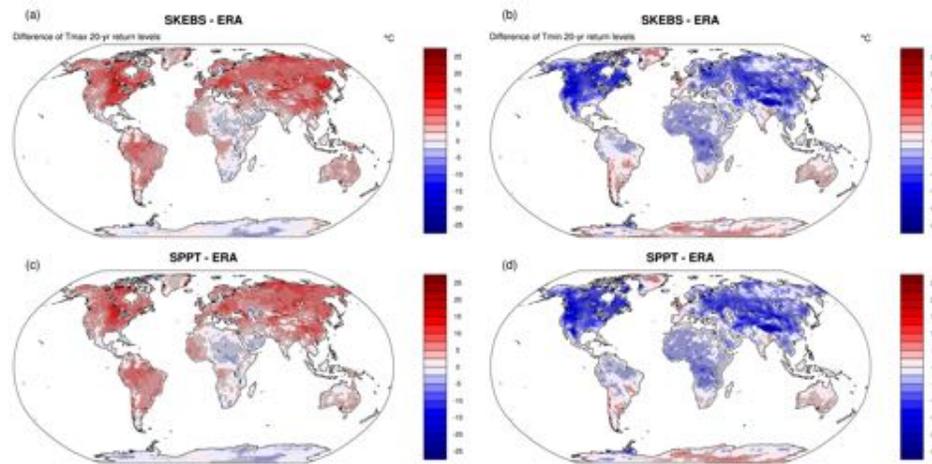


Figure 2.13: Differences of 20-yr return level estimates of 1979-2010 annual extremes of daily maximum (left) and minimum (right) temperature between ERA and the stochastic parameterization schemes, SKEBS (top), SPPT (bottom). Stippling indicates return level differences are significant at a 5% level.

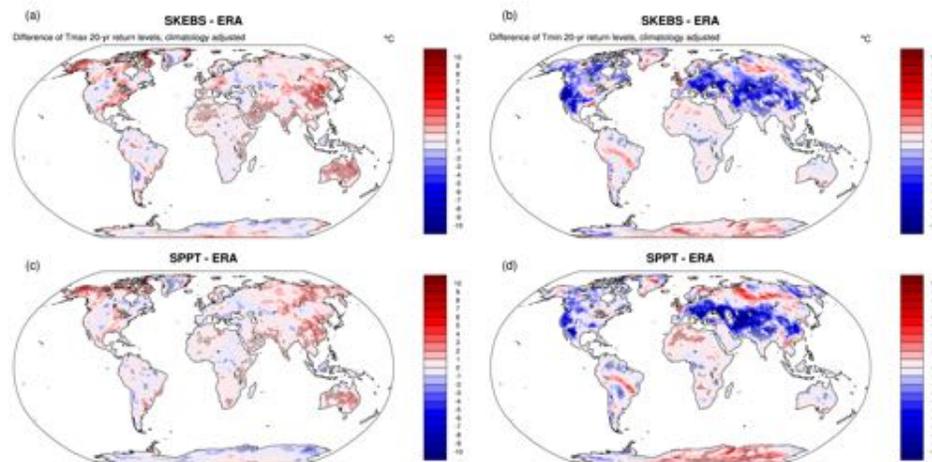


Figure 2.14: Same as in Fig. 2.13, except simulations have been adjusted by the respective extremes of the T_{max} and T_{min} annual cycles.

2.2 Temporal evolution

2.2.1 Motivation

Indices of climate extremes derived from observational datasets have become an important line of evidence into the response of temperature and precipitation extremes to anthropogenically enhanced greenhouse forcing. The Intergovernmental Panel on Climate Change (IPCC), in its Fifth Assessment Report (AR5, [42]), concluded that since the middle of the twentieth century, most land areas have experienced significant warming of both maximum and minimum temperature extremes, based on the evidence provided by [32] which showed that at a global scale, the number of cold days and nights have decreased while the numbers of warm days and nights have increased. These and other indices, such as the hottest or coldest day of a year, and the annual maximum 1 day or 5 day precipitation rates, 27 in total, were recently defined by the World Meteorological Organization Commission for Climatology/Climate Variability and Predictability/Joint Technical Commission for Oceanography and Marine Meteorology Expert Team on Climate Change and Detection Indices (ETCCDI), as part of an internationally coordinated effort to improve the observed data availability and analysis of variables that represent the more extreme aspects of climate [1, 32]. The ETCCDI indices, however, capture moderate aspects of temperature and precipitation extremes [57, 100], with the advantage that changes in their characteristics are more readily identified as they are typically based on events with reoccurrence times of a year or shorter [54]. In some countries where restrictions on data sharing exist, these indices represent the only publicly available information on climate extremes provided by local weather services [31].

In addition to monitoring present and assessing future changes in climate extremes [89, 1, 79], these indices have also found application in the evaluation of climate models [81, 83, 82] and observationally-based datasets [31, 24]. [81] showed that the climate models participating in the Coupled Model Inter-comparison Project Phase 5 (CMIP5) are generally able to match the temporal evolution of these indices, after adjusting for mean differences, as compared to reanalysis, but deficiencies exist, particularly with respect to extreme precipitation intensity. The difficulties in simulating precipitation extremes have been well documented [26, 85], and are typically associated with the misrepresentation of subgrid-scale processes. There is evidence that increasing the horizontal resolution of a model leads to improvements in the simulation of climate extremes [96, 95, 58], as fine-scale processes become explicitly resolved; however, at a significant computational cost, and the required retuning of the physical parameterizations may lead to an exacerbation of existing deficiencies [2].

An alternative approach to improving the representation of such processes is by means of stochastic parameterizations [73]. Numerous studies have reported on the effect of stochastic parameterizations on the quality of model simulations i.e. [48, 4, 7], however, little work has been devoted to examining their impact on climate extremes. In the previous section we investigated the performance of the Community Atmosphere Model version 4 (CAM4) with stochastic parameterizations in representing temperature extremes, as described by 20-yr return levels, and found that the presence of a stochastic parameterization produced a shift in the distributions of annual warm and cool temperature extremes towards warmer temperatures. Since CAM4 overestimates both warm and cold extremes, compared with reanalysis and observations, this resulted in more severe warm extremes, but a better agreement in cold extremes. In this section, we

extend this work by studying the temporal evolution of temperature extremes as represented by a subset of the ETCCDI indices. More specifically, we consider the temporal evolution of the TXx and TNn indices, representing the warmest and coldest day of the year, as well as the TXn and TNx indices, representing the coldest day and warmest night of the year, as simulated by CAM4 and the stochastic parameterization experiments and compare it with that of the ERA-Interim reanalysis product. Furthermore, to place this analysis within a broader context, we compare the temporal evolution of the latter models against that of the current generation of climate models, represented by the CMIP5 archive.

Note that section 2.1 investigated the effects of stochastic parameterizations on temperature extremes in a stationary setting, that is, under the assumption that the distributional properties of these did not change in time. By examining the temporal evolution of annual temperature extremes, we are evaluating precisely how this assumption is violated.

This content of this section is as follows. First, we provide a brief description of the datasets and climate indices in section 2.2.2, followed by the results and further remarks in 2.2.3.

2.2.2 Data

As in 2.1, we use as control experiment the 1° IPCC-AMIP simulation of CAM4 and consider two stochastic parameterizations schemes: the stochastic kinetic energy backscatter (SKEBS) scheme and the stochastically perturbed parameterization tendency (SPPT) scheme. We use here the same verification datasets, ERA Interim, developed at the ECMWF and the gridded land-

based HadEX2 observational dataset. To provide a context for the model-based temperature extremes, we also consider indices computed from the CMIP5 archive [88], which are available in the ETCCDI extremes indices archive (www.cccma.ec.gc.ca/data/climdex/). We use 26 CMIP5 models (see Tables 2.2.2 and 2.2.2), including only one ensemble member from each model, typically the first one. As in [83], the historical simulations from 1971–2005 are concatenated with simulations using the RCP4.5 forcing scenario to cover the analysis period.

We focus on a subset of the ETCCDI temperature indices, the absolute indices TXx, TXn, TNn and TNx, corresponding to the annual maximum of daily maximum temperature (TX), the annual minimum of TX, the annual minimum of daily minimum temperature (TN), and the annual maximum of TN, respectively. The indices are calculated using the R package `climdex.pcic`. For comparison, the model and reanalysis indices are regridded to the $3.75^\circ \times 2.5^\circ$ HadEX2 grid, using the first-order conservative remapping procedure implemented in the Climate Data Operators (<https://code.zmaw.de/projects/cdo>). Because of the temporal and spatial coverage limitations in HadEX2, which vary across indices, we apply masks that exclude from the analyses those gridpoints with less than 30 years of data. Anomalies are computed with respect to the reference period 1981-2000.

2.2.3 Results and further remarks

The globally area-averaged temperature indices of the models, reanalysis and observations over the period 1979-2010 are displayed in Fig. 2.15. The discrep-

Table 2.1: CMIP5 models used in the analysis. The respective temperature indices were downloaded from the ETCCDI indices archive (<http://www.cccma.ec.gc.ca/data/climindex/>).

	Model	Institution
1	ACCESS1-0	Commonwealth Scientific and Industrial Research Organisation and Bureau of Meteorology, Australia
2	BCC-CSM1-1	Beijing Climate Center, China Meteorological Administration, China
3	BCC-CSM1-1-M	Beijing Climate Center, China Meteorological Administration, China
4	BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University, China
5	CanCM4	Canadian Centre for Climate Modelling and Analysis, Canada
6	CanESM2	Canadian Centre for Climate Modelling and Analysis, Canada
7	CCSM4	National Center for Atmospheric Research (NCAR), USA
8	CMCC-CM	Centro Euro-Mediterraneo per i Cambiamenti Climatici, Italy
9	CNRM-CM5	Centre National de Recherches Meteorologiques, France
10	CSIRO-Mk3-6-0	Commonwealth Scientific and Industrial Research Organization, Australia

Table 2.2: CMIP5 models used in the analysis (continued).

Model	Institution
11 GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory, USA
12 GFDL-ESM2M	NOAA Geophysical Fluid Dynamics Laboratory, USA
13 GISS-E2-R	NASA Goddard Institute for Space Studies, USA
14 HadCM3	Met Office Hadley Centre, UK
15 HadGEM2-ES	Met Office Hadley Centre, UK
16 INMCM4	Institute for Numerical Mathematics, Russia
17 IPSL-CM5A-LR	Institut Pierre-Simon Laplace, France
18 IPSL-CM5B-LR	Institut Pierre-Simon Laplace, France
19 MIROC4h	AORI (Atmosphere and Ocean Research Institute), NIES (National Institute for Environmental Studies), JAMSTEC (Japan Agency for Marine-Earth Science and Technology), Japan
20 MIROC5	AORI, NIES, JAMSTEC, Japan
21 MIROC-ESM	AORI, NIES, JAMSTEC, Japan
22 MIROC-ESM-CHEM	AORI, NIES, JAMSTEC, Japan
23 MPI-ESM-LR	Max Planck Institute for Meteorology, Germany
24 MPI-ESM-MR	Max Planck Institute for Meteorology, Germany
25 MRI-CGCM3	Meteorological Research Institute, Japan
26 NorESM1-M	Norwegian Climate Centre, Norway

ancies between ERA and HadEX2 are consistent with those found in [31]. In 2.1 it was shown that the influence of stochastic parameterizations was that of a shift in the distributions of annual extremes towards warmer temperatures, with a stronger response in warm extremes. Here we see this reflected in the consistently positive spread in the TXx index between the two stochastic parameterization schemes and CAM; though it is less apparent in the TNn index, as the average spread is only 0.5°C and 0.2°C , for SKEBS and SPPT, respectively. The values of the TXx index, representing the warmest day of the year, are usually samples from the distribution of TX during the warmest month of the year, and likewise for the TNx index, representing the warmest night of the year, but with respect to the distribution of TN. The resemblance in the spreads between the stochastic parameterizations and CAM in both indices suggests that these generate similar responses in the distribution of both TX and TN at that time of the year. Examination of the spreads in the TNn and TXn indices leads to analogous conclusions, but with respect to the coldest month of the year.

[31] investigated the effect on these indices of the spatial scale mismatch between the area averages of climate models and reanalysis, with the point values of the HadEX2 dataset, over regions similar to the ones defined by the masks described above. The authors concluded that because the climate model and reanalysis grid-point values represent spatial averages of daily values rather than spatial averages of local annual extremes, as in the case of HadEX2, systematic biases arise of the order of $2\text{-}3^{\circ}\text{C}$. In practice, simulations that coincide with HadEX2 in their representation of these indices, would show temperatures corresponding to the warmest day and night to be approximately $2\text{-}3^{\circ}$ colder than the respective values of HadEX2, while those corresponding to the coldest day and night would be $2\text{-}3^{\circ}$ warmer. Therefore, the HadEX2 TXx and TNx

indices located near the upper end of the CMIP5 interquartile range suggests that the ensemble reasonably simulates these indices, whereas the respective CAM-simulated indices fall either on or above the HadEX2 line, indicating an overestimation of these, with the stochastic parameterizations furthering it. In contrast, both the CMIP5 ensemble and the model simulations are cold biased in TNn and TXn relative to HadEX2, with the model simulations more severely so as they fall over the lower end of the CMIP5 interquartile range.

The spreads between the various datasets are considerably reduced after normalizing the indices by their respective 1981-2000 average, in agreement with [81] and [31]. Note the substantial difference in variability of the indices based on TN compared to those of TX. The abrupt spikes and drops in the year 2010 are likely the result of the reduction in the number of gridpoints comprising the global average, which range from 7% to 20%.

Decadal trends of the indices from the models, reanalysis and HadEX2 were also investigated at the grid-point level. These trends were computed using the non-parametric Theil-Sen slope estimator [78], with the Mann-Kendall test [97] to establish significance. The choice of this approach over the traditional linear regression technique was motivated by its robustness to outliers and because the distributional properties of annual extremes tend to violate the normality assumption embedded in the latter. Where necessary, the time series were also prewhitened according to procedure described in [92], and more recently in [101], to account for the effect of serial correlation on the hypothesis test of no trend. While differences in spatial patterns were discernible between the various models, we could not distinguish these differences from internal climate variability, defined as the unforced climate variability intrinsic to a given

climate state [51]. It arises from atmospheric, oceanic, land, and cryospheric processes and their coupled interactions, and it is known to generate substantial fluctuations in temperature trends at the grid-point level (see for instance Figure 4 and 5 in [51]).

The impact of internal climate variability was confirmed when computing regional and global decadal trends by area-averaging grid-point index values (Fig. 2.16). We computed these trends for the other 5 members of CAM4 1° IPCC-AMIP ensemble, and found that for every index the global-level decadal trends corresponding to the stochastic parameterizations were within the range of values spanned by the CAM4 ensemble, and only few regional trends fell outside this range, suggesting that the impact of stochastic parameterizations on temperature extreme decadal trends is minor, to the extent that it is indistinguishable from the normal variability inherent to the climate system.

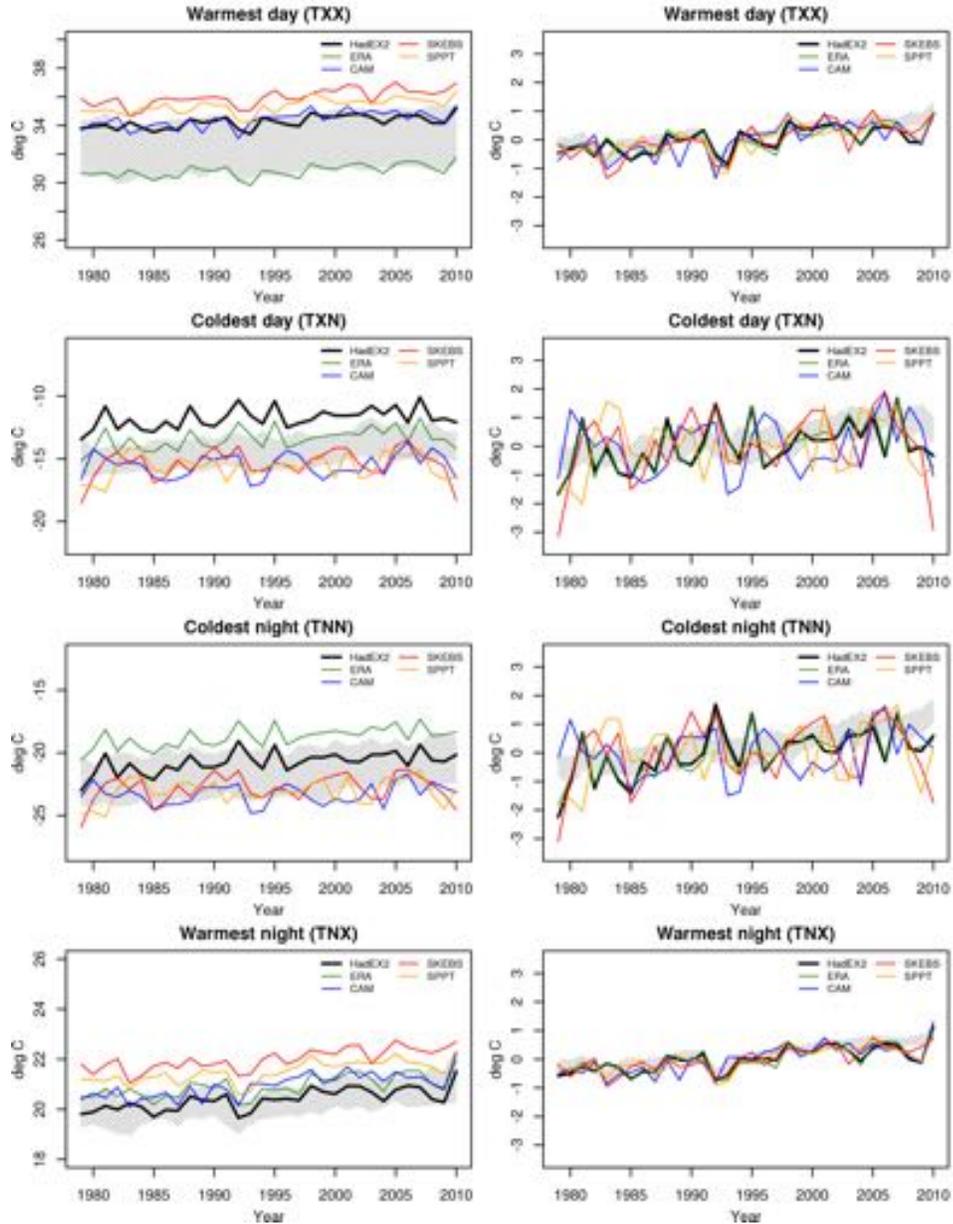


Figure 2.15: Global area-averaged TX_x, TX_n, TN_n, TN_x ETCCDI indices of CAM4, SKEBS, SPPT, reanalysis and observations over the period 1979-2010. The shading represents the interquartile model spread (range between the 25th and 75th quantiles) of 26 models belonging to the CMIP5 archive. For each index, the left column displays the absolute values and the right column shows anomalies with respect to the reference period 1981 to 2000.

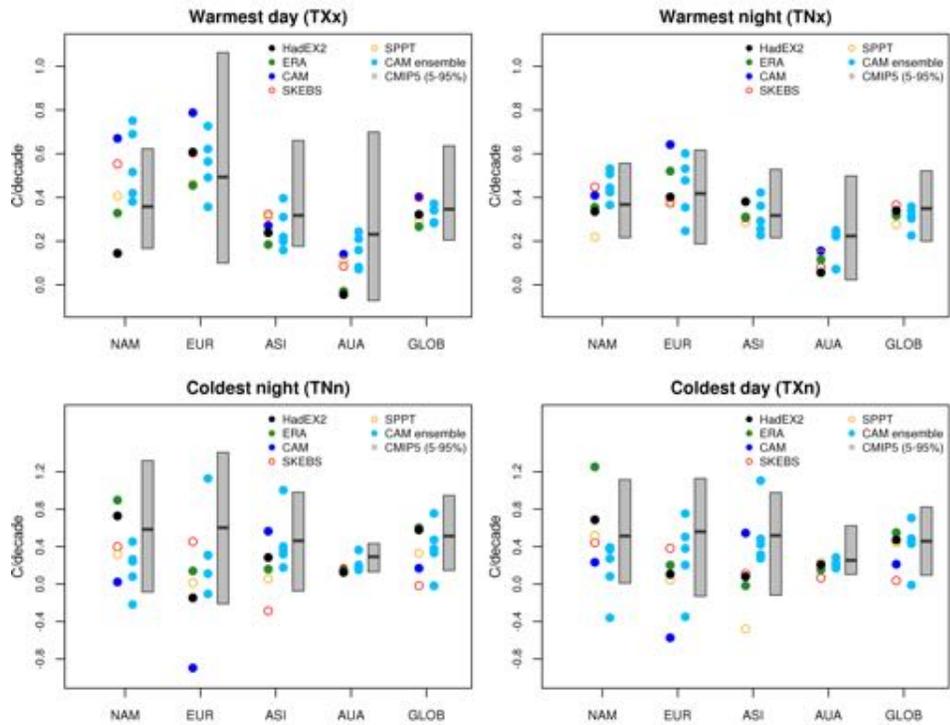


Figure 2.16: Decadal trends in global and regional area-averaged TXx, TXn, TNn and TNx indices of CAM4, SKEBS, SPPT and ERA, with regions defined in Fig. 2.6. Light blue dots denote analogous index decadal trends from the 5 other members of the 1° IPCC-AMIP CAM4 ensemble. Gray block denotes the 5-95th percentile range of decadal trends from 26 models of the CMIP5 archive.

CHAPTER 3

PRECIPITATION EXTREMES

3.1 Introduction

In the previous chapter we examined the effect of stochastic parameterizations on temperature extremes, here we shift the focus to their impact on precipitation extremes. We begin by reviewing the recent literature on such climate extremes, starting with the study by [53], which investigated the simulation of precipitation extremes based on 20-yr return levels (P_{20}) of 1981-2000 annual extremes of daily accumulation in models participating in the Coupled Model Intercomparison Project Phase 3 (CMIP3). The authors showed that 20-yr precipitation extremes were reasonably simulated in the moderate and high latitudes but considerable uncertainties emerged over the Tropics and subtropical regions. In particular, the amplitude of zonally-averaged extremes was higher in the Tropics for most of the reanalysis products than in any of the 16 models. This underestimation over the Tropics is reflected in the ensemble mean of 20-yr extremes as a narrow band along the equator of weaker extremes compared to those of the ERA-40 reanalysis [91] (Fig. 3.1). The models simulate, on average, however, more intense precipitation extremes in the generally dry regions of northern Africa and off the sub-tropical west coasts of Africa and North and South America. As an indication of the uncertainty stemming from internal variability, differences in zonally-averaged extremes between different simulations conducted with the same model were in general much smaller than intermodel differences, while sampling variability, as measured by 95% bootstrap confidence intervals computed from observationally-constrained datasets, was

also shown to be small compared to the latter differences.

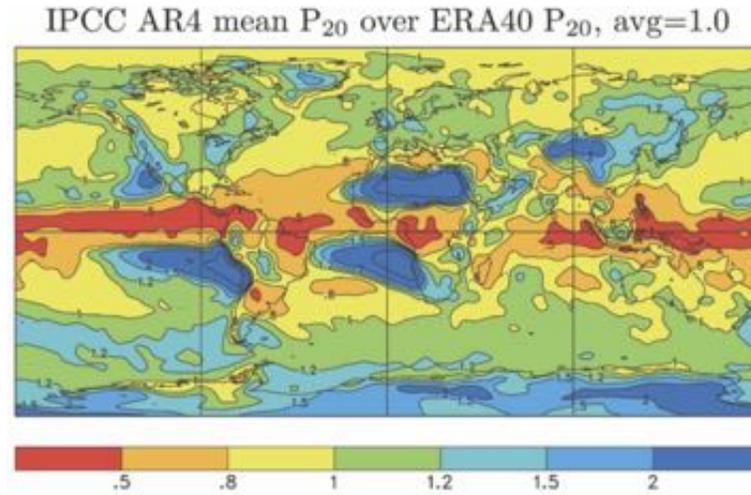


Figure 3.1: The ratio of the multimodel ensemble mean of P_{20} estimates over P_{20} estimated from ERA-40 (Figure 7 from [53])

[54] provided an update of the analysis in [53], using the latest generation of climate models participating in the Coupled Model Intercomparison Project Phase 5 (CMIP5). As in [53], precipitation extremes were characterized in terms of 20-yr return levels of annual extremes of daily precipitation amounts, but the annual extremes corresponded to those of 1986-2005. The performance of CMIP5 models in simulating precipitation extremes was found to be comparable to that of the CMIP3 ensemble. The CMIP5 multi-model median of P_{20} compares reasonably well to the ERA-Interim reanalysis over the extratropics, but considerable discrepancies emerge over the Tropics and subtropical regions (Fig. 3.2). The narrow band of weaker band along the equator persists, as well as the overestimation of precipitation extremes over northern Africa and off the western coasts of Africa and North and South America.

[26] explains that realistically simulating the regional patterns, temporal variations, and correct combination of frequency and intensity of precipita-

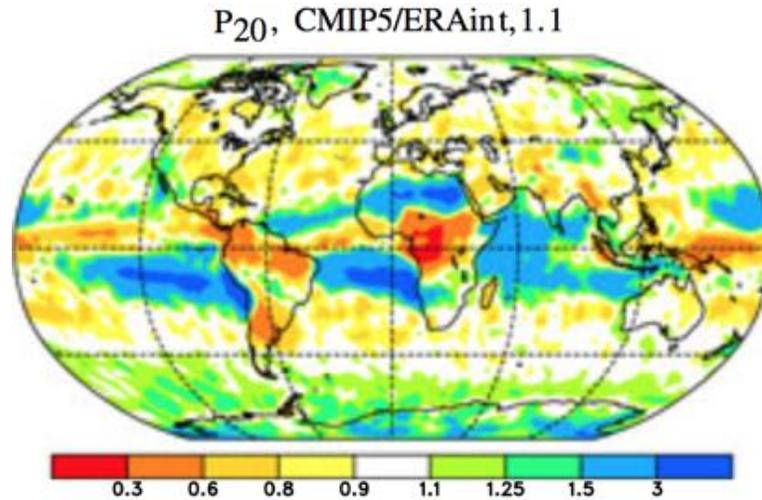


Figure 3.2: The ratio of the CMIP5 median of P_{20} over ERA-Interim P_{20} (Figure 1 from [54])

tion is quite challenging for coupled global climate models (CGCMs) given the complexity of precipitation processes in the atmosphere, which include cloud microphysics, cumulus convection, planetary boundary layer processes, large-scale circulations, and many others. Further compounding the difficulty is that precipitation is episodic as opposed to smoothly varying like temperature and other climate variables, and it can also have different types (e.g., convective versus stratiform) and phases (i.e., solid versus liquid).

The author notes further that a widespread problem among numerical weather and climate models is that these tend to precipitate too frequently at reduced intensity, even though precipitation amount is reasonable, caused partly by the frequent firing of moist convection, whereas in nature the convective inhibition processes often allow atmospheric instability to accumulate before intense convection begins. Various approaches have been applied to alleviate this “drizzling” problem in models, but a full solution has proven to be difficult.

Another outstanding problem described in [26] is the so-called double in-

terrestrial convergence zone (ITCZ) phenomenon, characterized by two perennial zonal belts of maximum precipitation straddling the equator over the central and eastern Pacific Ocean in most CGCMs. In nature it is only during boreal spring that a weak rainfall maximum emerges south of the equator in the eastern Pacific. These equatorial rainfall and SST biases have been linked to erroneous cloud feedbacks in the tropical Pacific, errors in equatorial west-east SST gradients and marine stratus clouds over the Peruvian coast, and other air-sea interactions in CGCMs.

The analysis of temperature extremes as simulated by experiments with stochastic parameterizations in chapter 2 is largely repeated here for precipitation extremes. But consistent with the literature that focuses on more moderate aspects of precipitation extremes, we extend the analysis to include some of the metrics commonly used in those studies, in addition to non-extreme metrics that capture general aspects such as annual precipitation amounts and frequency.

This chapter is organized as follows. Section 3.2 describes the datasets and methods used to compute the extreme and non-extreme precipitation metrics. Section 3.3 presents the results while section 3.4 provides a discussion and concludes.

3.2 Data and Methods

As in chapter 2, we use as control experiment the first ensemble member of the CAM 4 1° IPCC-AMIP simulations, which spans the years 1979-2010, at a horizontal resolution of $0.9^\circ \times 1.25^\circ$. This simulation follows the AMIP protocol [39], and so uses observed monthly averaged distributions of sea surface tempera-

ture and sea ice as boundary conditions. We also consider the two stochastic parameterization schemes used in the previous chapters, namely, the stochastic kinetic energy backscatter (SKEBS) scheme and the stochastically perturbed parameterization tendency (SPPT) scheme (see 2.1.2 for details), as well as the reanalysis ERA-Interim (hereafter ERA) against which the simulated annual extremes are compared. Recall that reanalysis products, with their global coverage and gridded output make them ideal candidates as verification datasets for climate model output, however, uncertainties in reanalysis precipitation fields are substantial as observations are not directly assimilated in these products [54]. At each grid point, the time series of annual maximum of daily precipitation, which by definition corresponds to the RX1day ETCCDI index, is computed using the R package `climdex.pcic`. A Generalized Extreme Value distribution (GEV) is fit to this index at each gridpoint and 20-yr return values are computed in the manner described in 2.1.3. To enable a comparison with ERA, simulated return values and GEV parameter estimates are regridded to match the latter's coarser grid.

The set of moderate extreme statistics follow [84, 40] and are based on the concept of a wet day, defined as one in which the daily precipitation amount exceeds the average of the annual 95th percentile of daily precipitation of some reference dataset. From this definition 3 metrics can be derived: the 1979-2010 average of (a) the number of wet days; (b) the extreme precipitation fraction, defined as total wet day precipitation divided by the total precipitation amount; and (c) the extreme precipitation intensity, defined as the total wet day precipitation divided by the number of wet days. Here we use the CPC precipitation dataset (see below) as the reference dataset from which to compute the average 95th percentiles.

Following [27], we consider as non-extreme metrics the mean frequency, intensity and total precipitation amount over the 1979-2010 period. More specifically, we define a rainy day as one in which the daily precipitation amount exceeds 1 mm, and define precipitation frequency over a given period as the ratio of the number of rainy days and the total number of days in the period; and precipitation intensity as the ratio of the total precipitation amount and number of rainy days.

Moderate and non-extreme statistics from 1979-2010 are validated over the contiguous United States (CONUS) against the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) $0.25^\circ \times 0.25^\circ$ US Unified Precipitation data (<http://www.esrl.noaa.gov/psd/>), which provides high resolution gridded precipitation data over the US, derived from station data, from 1948 to 2006, with an extension to the present through a real-time version. The spatial domain spans 20.125° - 49.875° N and 230.125° - 304.875° E. Daily precipitation amounts are regridded to match the model $0.9^\circ \times 1.25^\circ$ grid, prior to computing the 95th percentiles and statistics at each grid point.

3.3 Results

3.3.1 Annual precipitation extremes

Figure 3.3 displays the 20-yr return values of annual precipitation extremes of CAM4 and the ratio of these over those of ERA. CAM4 P_{20} values are consistently simulated over the extratropics, except over Antarctica where the magnitude of CAM4 extremes is twice that of ERA in some areas, in agreement with

the findings of [53] and [54]. However, considerable discrepancies emerge over the Tropics and subtropical regions, with a spatial pattern resembling the comparison of CMIP5 and ERA, except that the area of more intense precipitation over northern Africa extends further east into the Arabian peninsula and Indian ocean in CAM4 (contrast Fig. 3.2 to Fig. 3.3(b)).

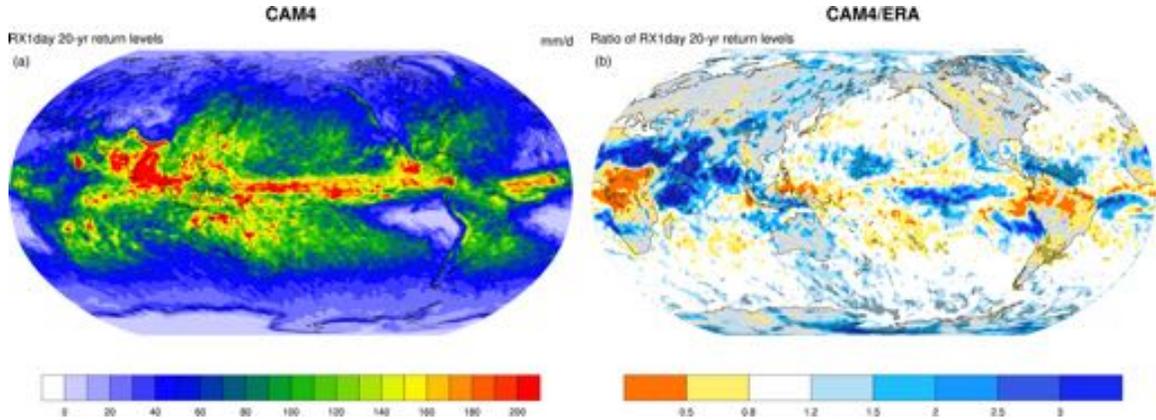


Figure 3.3: (left) 20-yr return level estimates of 1979-2010 annual extremes of daily precipitation of CAM4 and (right) and the ratio with those of ERA.

The response of precipitation extremes to the perturbed tendencies of the stochastic parameterization schemes is not easily discernable when comparing P_{20} ratios of these experiments and the CAM4 control simulation over ERA (Fig. 3.4). Differences in the simulation of precipitation extremes appear to be largely confined to the equatorial Pacific Ocean. The SKEBS and SPPT P_{20} ratios with CAM4 in the denominator, however, reveal two distinct spatial patterns over the Tropics. Over land regions, precipitation extremes appear to be invariant to the effect of stochastic parameterizations, in contrast with temperature extremes. The response over the tropical oceans in SKEBS is characterized by a narrow band of weaker extremes along the equator, and more intense extremes off the equator, particularly in the western equatorial Pacific. In contrast, SPPT

simulated precipitation extremes are remarkably similar to those of CAM4, with the exception of small areas off the western coasts of South and North America where the SPPT exhibits weaker extremes.

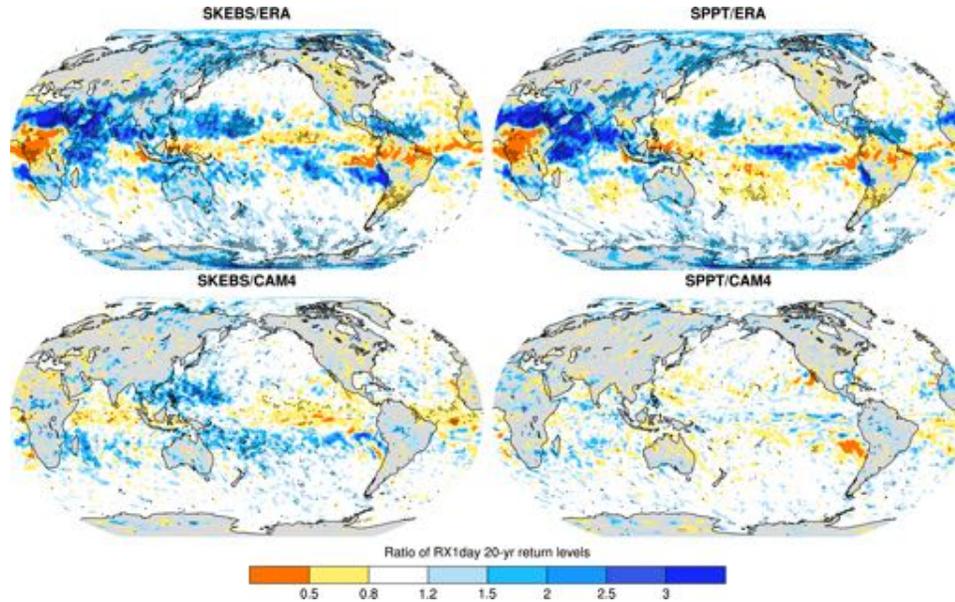


Figure 3.4: Ratios of 20-yr return level estimates of 1979-2010 annual extremes of daily precipitation of (left) SKEBS with (top) ERA and (bottom) CAM4, and (right) SPPT with (top) ERA and (bottom) CAM4.

As in 2.1.4, we examine the distributional differences in global GEV pdfs that underlie the observed differences in return values between the datasets. Figure 3.5 displays the GEV pdfs of CAM4, SKEBS, SPPT and ERA annual precipitation extremes over land and ocean regions. Unlike temperature extremes, precipitation extremes exhibit an unbounded upper tail over both land and oceans, as indicated by positive shape parameter values in the reanalysis of approximately 0.2 in both cases, typical for these types of extremes (e.g. [50]). This stability in the shape parameter estimates across land and oceans is also seen in the CAM4 estimates, despite it systematically underestimating the tail behavior relative to reanalysis, at nearly half of its value. With values nearly identical to the con-

trol simulation, it appears neither of the stochastic parameterization schemes meaningfully modifies the tail behavior of the latter.

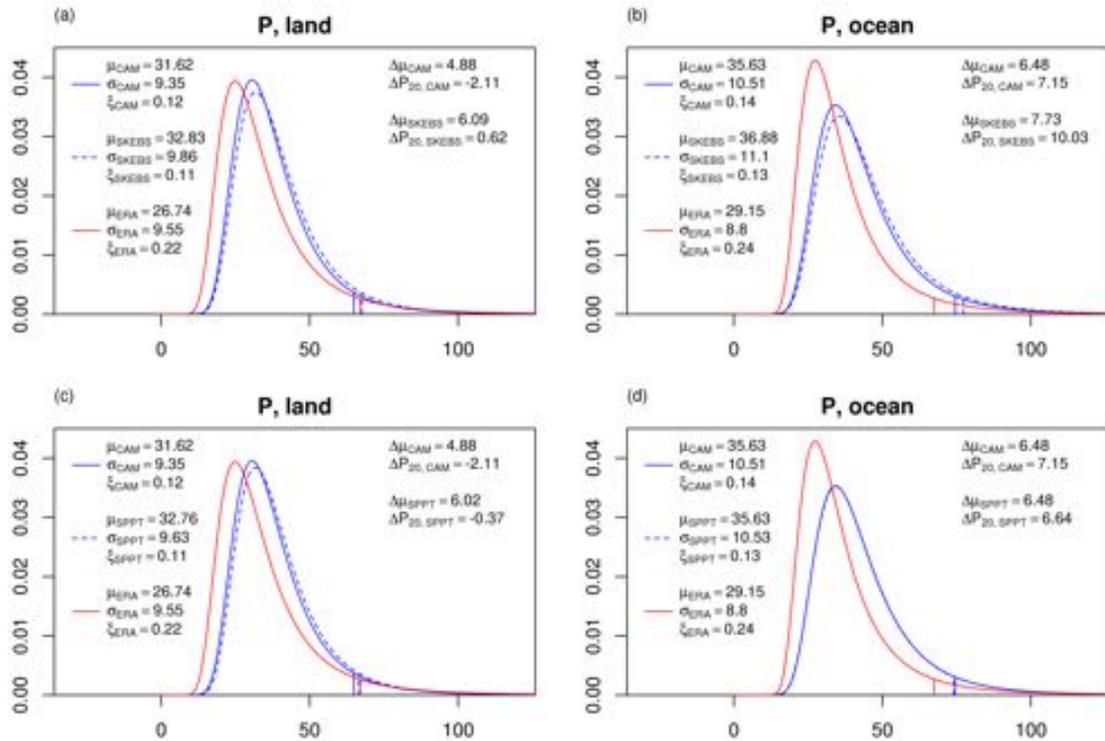


Figure 3.5: GEV pdfs as in Fig. 2.2, but of annual extremes of daily precipitation, with (top) CAM (solid blue) and SKEBS (dashed blue) with respect to ERA (red) over (left) land and (right) oceans. (bottom) As above but with SPPT in place of SKEBS.

Over land regions, there is good agreement in the scale parameter estimates between the model simulations and reanalysis, but large positive biases in the location parameter estimates exist, in excess of 4 mm/day, which the stochastic parameterizations tend to make more severe by roughly 1 mm/day. Despite these positive biases, which in isolation would tend to produce an overestimation of identical magnitude in return levels relative to ERA, as they imply a shift of the distribution towards more intense precipitation, the heavier tail of the reanalysis distribution compensates for this effect and results in simulated

return levels that are comparable to those of ERA, namely, -0.37 mm/day and 0.62 mm/day for SPPT and SKEBS, respectively.

Over oceans, however, scale parameter estimates from simulated extremes differ considerably from those of observationally-constrained extremes. Interestingly, the magnitude of these estimates increases relative to the values over land in the models, for instance, from 9.4 to 10.5 mm/day in CAM4, while it decreases in ERA, from 9.6 to 8.8 mm/day. Consistent with the grid-point return value comparison in Fig. 3.4, the GEV distribution of SPPT-simulated extremes over oceans is nearly indistinguishable from that of CAM4. SKEBS-simulated precipitation extremes, in contrast, exhibit higher interannual variability relative to CAM4, as represented by slightly higher scale parameter. The positive bias in the location parameter of CAM4 becomes more pronounced over oceans, as does in SKEBS. The greater amplitude of the location parameter biases together with the positive biases in scale parameter estimates result in more intense annual extremes, up to 10 mm/day higher than ERA in the SKEBS simulation.

We have seen that discrepancies between simulated and observationally-constrained precipitation extremes are largely confined to the tropical and subtropical regions. It seems appropriate then to repeat the pdf analysis above focusing specifically on these regions, in particular, restricting the spatial domain to 30°S-N (Fig. 3.6). The effect on the annual extreme distributions of this spatial restriction is a general increase in the magnitude of the GEV parameters. Notable is the near doubling of the shape parameters of ERA, from 0.22 (land) and 0.24 (ocean) to 0.42 in both cases, while the increases in the simulated extremes are more moderate, such as that of CAM4, from 0.12 to 0.15 over

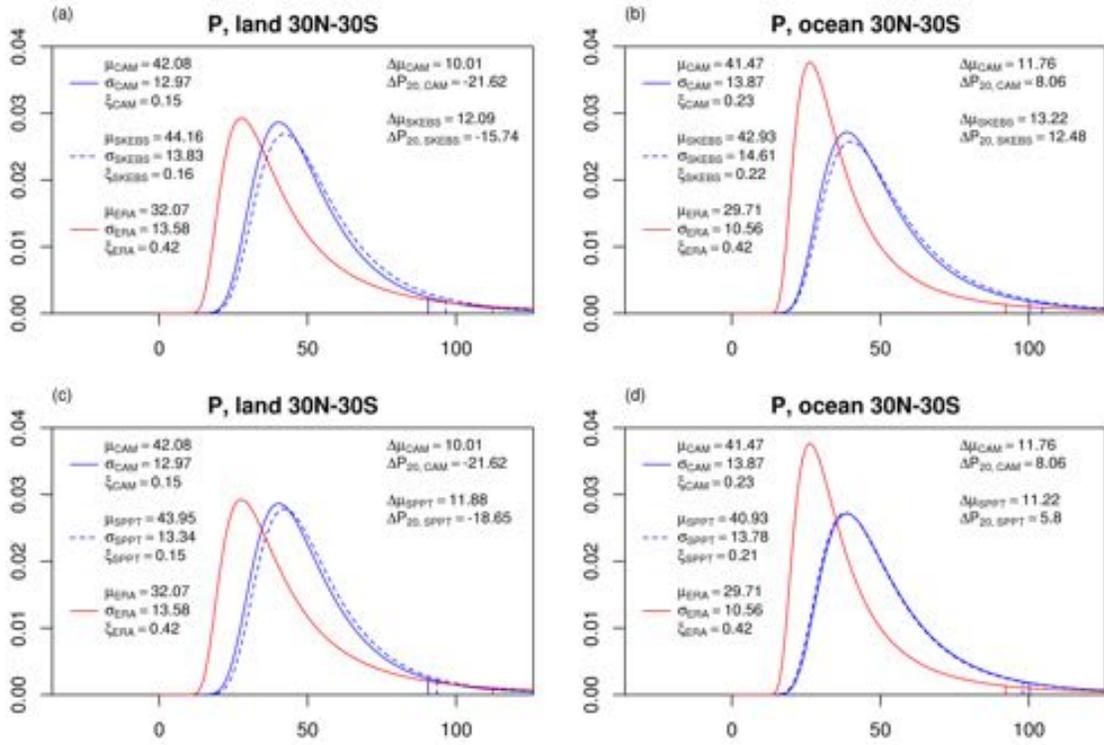


Figure 3.6: As in Fig. 3.5 but with the spatial domain restricted to 30°S-30°N.

land regions. Qualitatively, however, most of the remarks made for the spatially unrestricted case still apply. For instance, the scale parameter estimates are comparable across models and reanalysis over land regions, and the simulated location parameter estimates still exhibit a positive bias relative to reanalysis. But the increases in the latter parameter values are such that the positive bias of CAM4 increases to 10 mm/day, while that of SKEBS and SPPT, to 12.1 and 11.9 mm/day, respectively. The large discrepancies in shape parameters, however, overwhelm the effect of the location parameter biases resulting in substantial negative biases in 20-yr extremes, reaching -21.6 mm/day in CAM4, consistent with the dry bias observed over central Africa and South America in Fig. 3.3.

Over oceans, the positive biases in location parameter estimates are comparable to those over land, however, as in the spatially unrestricted case, the differences in the scale parameter estimates between the models and reanalysis are such that together with the discrepancies in the tail behavior between these datasets, results in return levels that exceed those ERA by 8.1 mm/day in the CAM4, 12.5 mm/day and 5.8 mm/day in SKEBS and SPPT respectively. The dry biases off the eastern coasts of North and South America in the P_{20} ratio of SPPT and CAM4 in Fig. 3.3 are consistent with the smaller difference in return levels relative to ERA, while the more intense precipitation extremes, particularly over the equatorial western Pacific, simulated by SKEBS are here represented by large location and scale parameter biases compared to CAM4, which translate into a larger difference in 20-yr extremes.

3.3.2 Non-extreme precipitation metrics

In this and the following section we move away from the asymptotic theory of annual extremes, with its reliance on the GEV distribution, to other aspects of the precipitation distribution. We also adopt a more regional focus, that of CONUS, as opposed to the above global focus. The choice of CONUS was motivated primarily by the availability of high-quality gauge-based observational datasets, one such dataset being CPC used here, allowing us to dispense with the aforementioned potential inaccuracies of the reanalysis precipitation fields. For convenience in the analysis below, we use the U.S. regional definitions (Fig. 3.7) from the National Climate Assessment [69] as in [84]. Our analysis here centers on the most basic of precipitation statistics, namely, annual precipitation, precipitation frequency and intensity, defined in section 3.2.

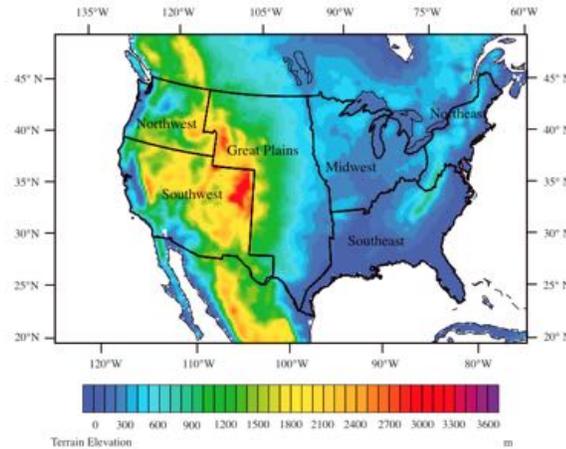


Figure 3.7: Topographical map of the contiguous United States showing the regions defined for the National Climate Assessment (from [84]).

Figure 3.8 displays the 1979–2010 average of total annual precipitation over CONUS of the CPC dataset, and that of CAM4, SKEBS and SPPT expressed in terms of percent anomalies. The upper-left panel presents the well-known spatial distribution of precipitation over CONUS, as depicted in CPC, consisting of dry conditions on an annual basis extending through the North and Southwest, and wetter conditions in the North and Southeast. CAM4 exhibits a strong dry bias along the central United States, a slight wet bias towards the east coast, and a stronger wet bias towards the west except along the coastline where an intense dry bias develops. This spatial pattern remains largely unchanged in the simulations with stochastic parameterizations, particularly for SPPT, where statistically significant differences are limited to parts of the Midwest. The SKEBS simulation, however, presents a more severe dry bias over the central US, but these drier conditions relative to CAM4, as they extend over parts of the western US, represent an improvement in the simulation of this statistic.

Simulated and observed precipitation frequency over the 1979-2010 period

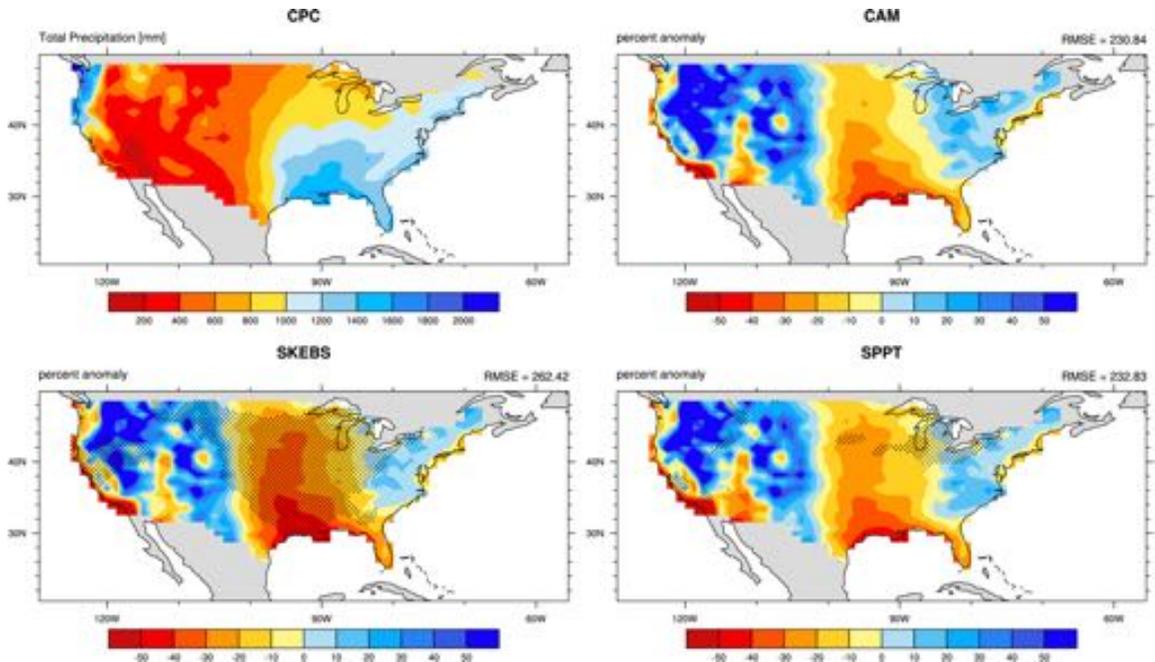


Figure 3.8: (top-left) 1979-2010 average of annual precipitation for observations, and anomalies, as percent, of (top-right) CAM4, (bottom-left) SKEBS and (bottom-right) SPPT. Observations correspond to the NOAA CPC 0.25x.025 Daily US Unified Gauge-Based Analysis of Precipitation data. Stippling in the panels of the stochastic parameterizations indicates differences with CAM4 are statistically significant at the 5% level.

is displayed in Figure 3.9. Observed precipitation frequency can be broadly characterized by an east-west gradient, with less frequent precipitation over the drier west and more frequent precipitation in the east. CAM4 simulated precipitation occurs generally too frequently relative to observations, which, as noted above, is a common shortcoming of climate models. The most severe overestimation occurs over the Great Plain and stretches westward over parts of the North and Southwest. While the perturbed tendencies in the SPPT simulation do not seem to effect the frequency of precipitation in CAM4, those of SKEBS lead to better agreement with observations as less frequent precipitation is generated over large parts of the central and western US.

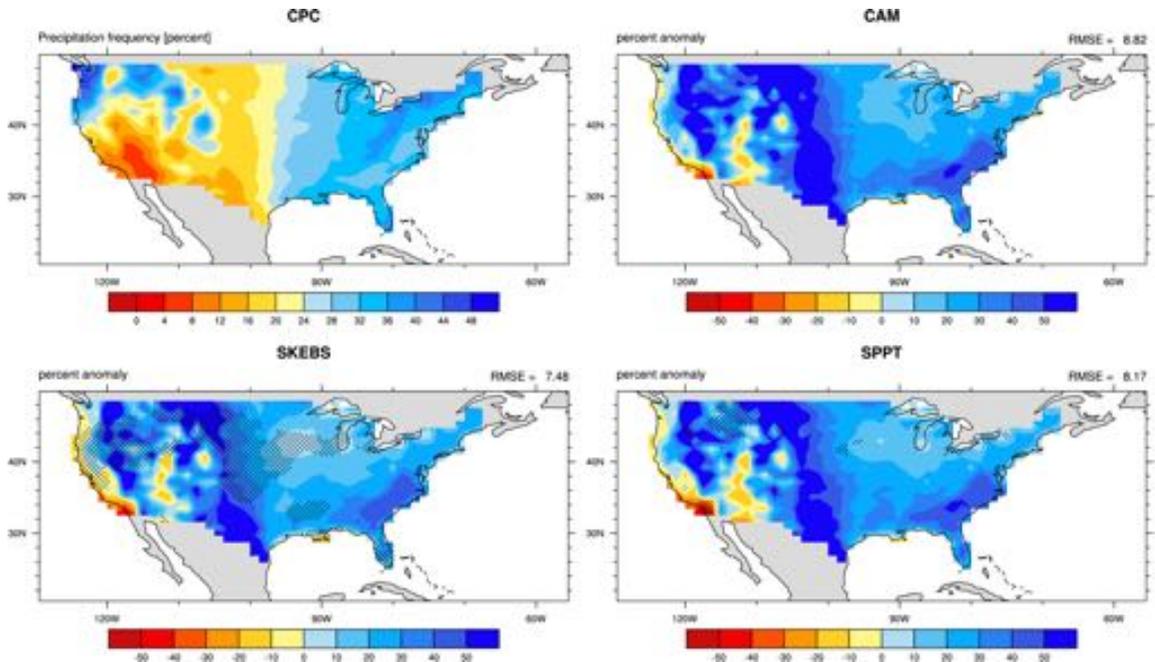


Figure 3.9: As in 3.8, but with respect to precipitation frequency.

The last of the non-extreme precipitation statistics, precipitation intensity, is displayed in Figure 3.10. The spatial pattern of the observations closely resembles that of total precipitation, as does the spatial pattern of anomalies in CAM4. Again, the simulation of precipitation intensity in SPPT is largely unchanged from that of CAM4, which was to be expected given the similarity in total precipitation and frequency. Despite the slight improvements in the simulation of precipitation frequency, the dry bias over the central US leads to a worsening of the representation of precipitation intensity in SKEBS, with a RMSE of 3 mm/day compared to 2.7 mm/day in CAM4.

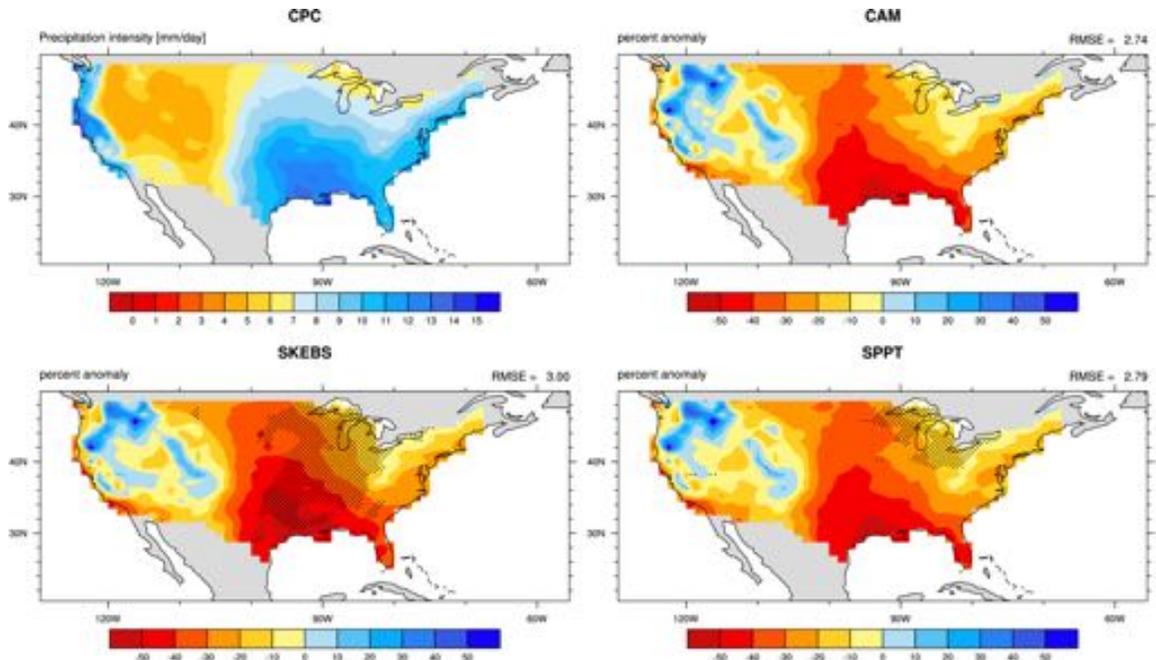


Figure 3.10: As in 3.8, but with respect to precipitation intensity.

3.3.3 Moderate extreme precipitation metrics

Figure 3.11 displays the 1979-2010 average number of wet days, as defined in section 3.2, for CPC, CAM4 and the two stochastic parameterization schemes. By construction, the average number of wet days in the CPC dataset is constant across the spatial domain. Notable differences arise with respect to CAM4, with positive anomalies exceeding 50% over much of western CONUS, while the number of wet days are less than half of the observed quantities over the southern Great Plains and Southeast. Overall, the spatial pattern resembles that of precipitation intensity in Fig. 3.8, but with more severe anomalies. As with the non-extreme precipitation metrics, the response of the number of wet days to the SPPT perturbations is hardly discernible. In contrast, the response to the SKEBS perturbations is an expansion of the area of negative anomalies further

north across the Great Plains, with most values statistically different from the CAM4 anomalies at the 5% significance level.

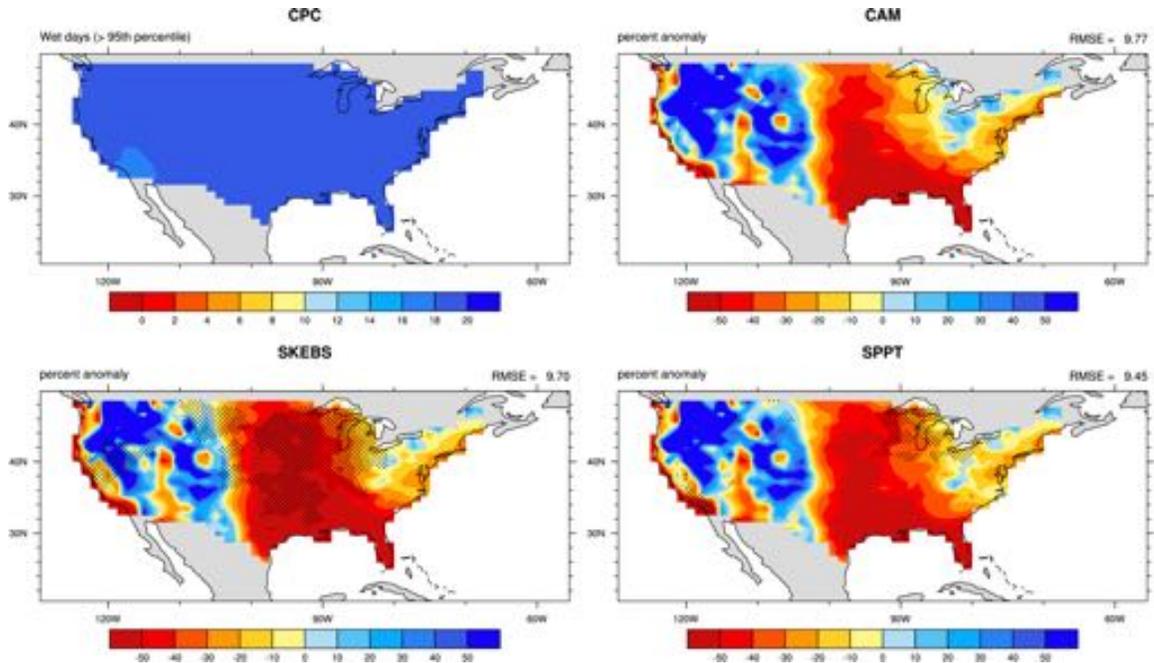


Figure 3.11: As in 3.8, but with respect to the number of wet days.

Simulated and observed extreme precipitation fraction is displayed in Figure 3.12. This fraction, as depicted in the CPC dataset, is fairly constant, within a 0.4-0.5 range, over most of the eastern part of the country. In contrast, over the Great Plains and the Southwest, values exceed 50%, and over the driest areas of the Southwest, the fraction increases above 80%. The spatial patterns and magnitude of the CAM4 anomalies are similar to those of the number of wet days, though the positive anomalies are not as extreme. The region of extreme negative anomalies also extends further north in the SKEBS simulation, but unlike the wet day count case, the differences with CAM4 are generally not statistically significant. The SPPT anomalies again agree closely with CAM4.

Wet day intensity in CPC over the 1979-2010 period closely resembles the

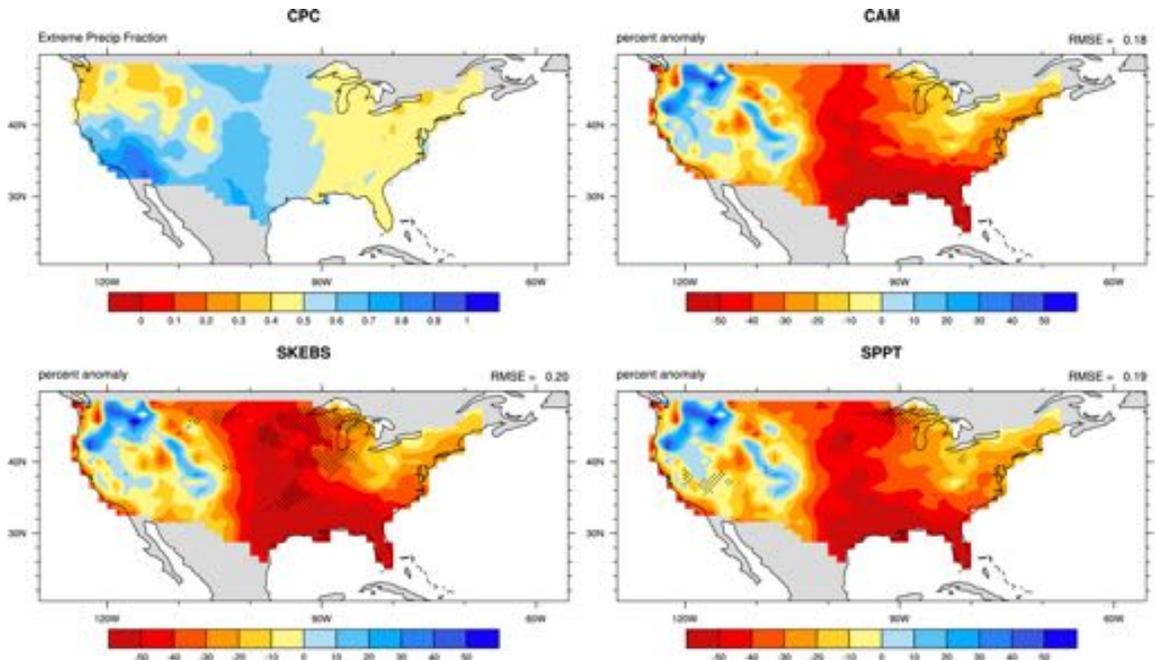


Figure 3.12: As in 3.8, but with respect to extreme precipitation fraction.

spatial pattern of total precipitation intensity (Fig. 3.13). Interestingly however, there is good agreement with CAM4, as well as the stochastic parameterization simulations, as most anomalies fall within the $\pm 10\%$ range.

3.4 Discussion and Conclusions

This study documents the performance of the NCAR Community Atmosphere Model CAM4 with stochastic parameterizations in simulating 1979-2010 precipitation extremes and non-extreme precipitation statistics, thereby extending the efforts of [87] that focused on temperature extremes. Precipitation extremes are evaluated in terms of 20-yr return levels of annual extremes of daily precipitation as well as moderate extreme statistics, based on the notion of a wet day, de-

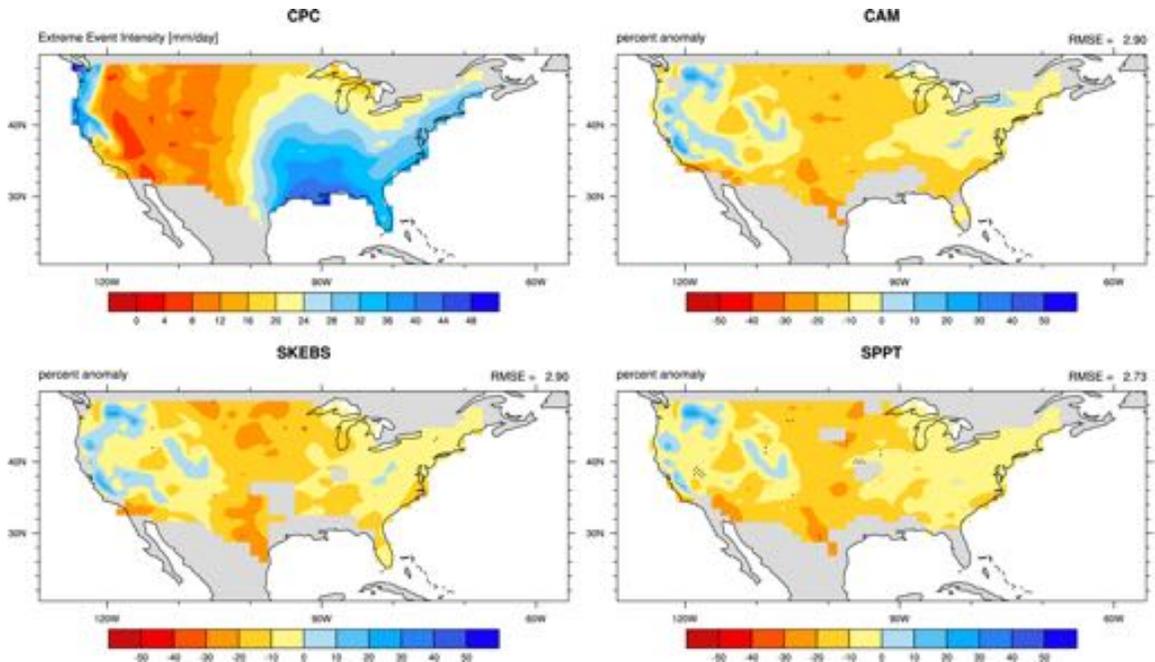


Figure 3.13: As in 3.8, but with respect to extreme event intensity.

defined as one in which the daily precipitation amount exceeds the average of the annual 95th percentile of daily precipitation of some reference dataset. The non-extreme statistics refer to the commonly analysed mean annual precipitation amounts, precipitation frequency and intensity (e.g. [26], [2]). Simulated 20-yr extremes are compared with those of ERA-Interim (ERA) reanalysis, while the moderate extreme and non-extreme statistics are evaluated against the NOAA CPC 0.25×0.25 Daily US Unified Gauge-Based Analysis of Precipitation data.

Our results indicate that CAM4 20-yr extremes are, in general, comparable in their magnitude and spatial patterns to the extremes simulated by the multimodel ensembles CMIP3 and CMIP5 relative to observationally-constrained datasets. CAM4 20-yr return levels show good agreement with ERA in the extratropics, but notable discrepancies arise in the Tropics and subtropical regions, with a band of weaker extremes along the equator, which is particularly

severe over central Africa, together with coherent areas of intense precipitation extremes over northern Africa, off the western coasts of North and South America, and off the coasts of Africa. These discrepancies may be related to difficulties in the representation of deep convection in the model [54].

The response in precipitation extremes, as measured by 20-yr return levels, to the stochastic parameterizations was, to our surprise, rather muted. In fact, the only discernible differences with the control simulation were confined to the equatorial Pacific Ocean. SPPT precipitation extremes are remarkably similar to those of CAM4, with the exception of small areas off the western coasts of South and North America where the SPPT exhibits weaker extremes. The response to the SKEBS scheme is slightly stronger, characterized by a narrow band of drier extremes along the equator, and more intense extremes off the equator, particularly in the western equatorial Pacific. Further work is required to understand this behavior of SKEBS over the tropics. But more generally, given this muted impact on precipitation extremes, it remains to be determined whether the perturbed tendencies of these stochastic parameterizations affect precipitation-related process that self-compensate in the upper quantiles of the distribution of annual extremes, or whether these processes themselves are invariant to the type of perturbations induced by these parameterizations.

As in [87], we interpret these return level discrepancies in terms of differences in GEV pdfs obtained from spatially averaging grid-point GEV parameter estimates. The weaker precipitation extremes over tropical land regions in the CAM4 simulation relative to the reanalysis (Fig. 3.3) are consistently represented when comparing global GEV pdfs, by a return level 2 mm/day lower than that of ERA. But what we learn from this analysis is that this difference is

explained primarily by a difference in the shape parameter estimates between the two datasets, which implies a difference in tail behavior. In each case, the shape parameter is positive, indicating an unbounded upper tail, but the shape parameter in ERA is nearly double the magnitude of that of CAM4. The heavier tail of ERA compensates for the effect of a higher location parameter estimate in CAM4, which, with nearly identical values in the scale parameter estimates, results in a higher return level for ERA. The GEV pdfs of the stochastic parameterizations are in close agreement with that of CAM4, more so than that of SPPT, in agreement with the grid-point return level estimates of Fig. 3.4. Over oceans, the discrepancies in return levels between CAM4 and ERA, particularly over the Tropics, observed in Fig. 3.3, translate into a higher return level in the CAM4 GEV pdf. This positive difference can be traced primarily to a large difference in scale parameter estimates favoring more intense precipitation extremes in CAM4, since the positive biases in location parameter estimates are similar to those over land and the change in shape parameter estimates is minor. When restricting the spatial domain to 30°S-30°N, the differences in parameter estimates are generally amplified, but the conclusions are qualitatively the same.

In light of the uncertainties in the representation of precipitation in reanalysis products ([53], [54], [81]) we chose to use an observational dataset, CPC, for the analysis of moderate extreme and non-extreme statistics. A more comprehensive spatial analysis would have required a satellite-based observational dataset, but these suffer from a short temporal domain due to the recent start of the satellite era. Indeed, we could have used such a dataset in the analysis of 20-yr return levels above, but that would have meant using a different technique to estimate the distribution of precipitation extremes because of the short sample size, and this was beyond the scope of this study. But more importantly,

that would have impaired the comparability of our results with recent studies involving model ensembles (e.g. [53], [54]).

With respect to the non-extreme statistics, SKEBS is shown to enhance the dry bias in CAM4 that extends over central CONUS and has a muted impact on the frequency of precipitation, leading to a more severe shortfall in the intensity of precipitation over the latter region. On the other hand, the differences in the simulation of these statistics between SPPT and CAM4 are for the most part not statistically significant, from which can be concluded that the perturbations associated with the SPPT scheme not only have little effect on the tails of the precipitation distribution but also on the body of the distribution.

The simulated number of wet days display spatial patterns much like those of total annual precipitation (Fig. 3.8), but with more severe negative anomalies over central CONUS in CAM4 and SPPT, representing shortfalls in the number of wet days exceeding 50% of those in CPC, which are exacerbated in SKEBS. Interestingly, the average wet day intensity is comparable across the simulations to that of CPC. In other words, while the number of wet days is underestimated in the models, once such an event occurs its intensity is on average comparable to that of CPC, which is consistent then with the underestimation seen in the wet day precipitation fraction. Future work applying tools from extreme value theory should help in determining whether this behavior is due to differences in the tail characteristics (shape parameter) between the simulated and observed precipitation distributions.

CHAPTER 4

PRECIPITATION EXTREMES IN TRMM 3B42

Precipitation is a fundamental component of the climate system, playing a key role in the coupling of the water and energy cycles. The large spatial and temporal variability of precipitation intensity, type, and occurrence makes measurements based on conventional surface instrumentation difficult [44]. Rain gauges provide the most direct measurement of point precipitation at the surface [71], however, their representativeness is limited to a small radius around the instrument [21]. Dense networks of gauges are therefore required for adequate representation, which are not feasible over much of the globe, particularly over oceanic and developing regions [45]. Weather radars address some of the issues of rain gauge coverage, like providing a spatial measurement of precipitation rather than point measurements, but they are limited in extent and number [55], and suffer from beam blockage in mountainous regions, anomalous propagation errors, and imprecise backscatter to rain rate relationships [44]. In contrast, satellite-based measurements offer an unrivaled vantage point to obtain a comprehensive description of global precipitation [55].

Several satellite-based precipitation products are available with varying spatial and temporal characteristics (see Table II in [55]). One of the most important of these is the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA), belonging to the latter category and developed at the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) [45]. Launched on 27 November 1997, TRMM is a joint US–Japan satellite mission to provide a detailed and comprehensive dataset of the four-dimensional distribution of rainfall and latent heating over tropical and

subtropical oceans and continents (40°N-S) [63]. Of the many TRMM data products available at the GSFC Earth Sciences Data and Information Services Center (GES DISC), the 3-hourly 3B42 TMPA product is one of the most popular because of its high spatial and temporal resolution. These products have been used extensively in applications ranging from climate model validation to water resource management [2, 26, 62]. But because satellite-based precipitation estimates are derived from indirect precipitation measurements, quantification of the errors associated with these estimates is critical to its usefulness in applications [19]. Several studies have been devoted to assessing the spatial error structures of the version 7 TMPA products (e.g [19, 18, 16, 22, 41, 68]), but the metrics used typically do not capture the extremes of the precipitation distribution. In this final chapter, we evaluate precipitation extremes, as represented by the TMPA version 7 research product, based on an analysis of 20-yr return levels of daily precipitation, against the APHRODITE observational dataset over the Monsoon Asia region, using a Bayesian hierarchical model to fit the extreme value parameters.

This work is organized as follows. Section 4.1 provides an overview of the techniques used from extreme value theory; section 4.2 describes in some detail the TRMM dataset and its component datasets, as well as the gauge-based dataset used as reference; section 4.3 describes the Bayesian hierarchical model and section 4.4 offers details about the model implementation. Lastly, section 4.5 presents the results and 4.6 provides a summary and concludes.

4.1 Extreme Value Theory

The analysis of temperature extremes in chapter 2 illustrates the use of a common technique in extreme value theory, where observations are partitioned into blocks – usually of yearly length, if the observations are indexed by time – and a GEV distribution is subsequently fit to the collection of the most extreme observations within each block. When the extreme observations correspond to the maximum values, the approach is referred to as the “block maxima” approach, otherwise, “block minima.” Since all but one observation per block is discarded, this approach is often deemed wasteful of potentially useful information to the inference procedure. An alternative technique that does not rely on blocking is the peaks-over-threshold approach, which focuses on observations that exceed a high threshold. These are typically modeled using a generalized Pareto distribution, whose survival function is of the form

$$\bar{H}(y) = \begin{cases} (1 + \xi \frac{y}{\sigma})^{-1/\xi} & \text{if } \xi \neq 0, \\ \exp(-\frac{y}{\sigma}) & \text{if } \xi = 0. \end{cases}$$

with a scale parameter $\sigma > 0$ and $y \leq 0$. Note that y has an upper bound $-\sigma/\xi$ if $\xi < 0$. A different yet equivalent characterization of extremes which unifies the above techniques is derived from the theory of point processes. Under this formulation, the exceedances above the threshold occur according to a Poisson process, while the excesses of those exceedances are assumed to follow a generalized Pareto distribution. More specifically, if Y_1, Y_2, \dots are a series of independent and identically distributed random variables that are well-behaved in the extreme value sense, i.e. there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$P\{(\max(Y_1, \dots, Y_n) - b_n)/a_n \leq z\} \rightarrow G(z),$$

with G non-degenerate, with lower and upper endpoints z_- and z_+ , respectively, then the sequence of point processes N_n on \mathbb{R}^2 , given by

$$N_n = \{(i/(n+1), (Y_i - b_n)/a_n) : i = 1, \dots, n\},$$

converges on regions $[0, 1] \times [u, \infty)$, $u > z_-$, to a non-homogeneous Poisson process with intensity measure on $A = [t_1, t_2] \times [z, z_+)$, of the form

$$(t_2 - t_1) \left(1 + \xi \frac{z - \mu}{\sigma}\right)^{-1/\xi}.$$

In practice, given n observations, a high threshold u is chosen, for which the Poisson approximation is deemed to be valid and A is set to $[0, n/n_y] \times [u, \infty)$, with n_y the number of observations per year. The normalizing sequences $\{a_n\}$ and $\{b_n\}$, can be absorbed into the location and scale parameters of the model so that the sequence of point processes can be defined as $N_n = \{(i/n_y, Y_i), i = 1, \dots, n\}$. Then, $N_n(A)$ is assumed to be well approximated by a Poisson process with intensity measure

$$\frac{n}{n_y} \left(1 + \xi \frac{z - \mu}{\sigma}\right)^{-1/\xi},$$

where the term n/n_y ensures that the parameters of the point process likelihood correspond the GEV parameters of the distribution of annual maxima.

Let N_u denote the number of observations that exceed u , and label them as $y_{(1)}, \dots, y_{(N_u)}$, then the likelihood function can be expressed as

$$L(\mu, \sigma, \xi | y_1, \dots, y_n) \propto \exp \left[-\frac{n}{n_y} \left(1 + \xi \frac{u - \mu}{\sigma}\right)^{-1/\xi} \right] \prod_{k=1}^{N_u} \frac{1}{\sigma} \left(1 + \xi \frac{y_{(k)} - \mu}{\sigma}\right)^{-1/\xi - 1}. \quad (4.1)$$

We introduce this point process model as the first stage model for the exceedances of daily precipitation amounts, placing the spatial structure in the second stage specification of the parameters μ , σ and ξ .

4.2 Data

From each of the datasets we use daily precipitation amounts and focus on the summer season (June, July, August)(JJA). Because our extreme value analysis is based on exceedances above a certain high hreshold, at each gridpoint we set such a threshold at the 0.95 empirical quantile of the non-zero measurements following [23, 95, 90]. Exceedances typically occur in clusters, so to ensure that the temporal independence assumption holds, a common desclustering method is employed where the start of a new cluster depends on the number of consecutive threshold deficits. Here, we set the number of threshold deficits to 2, and retain the maximum of each cluster for subsequent analysis.

As the verification dataset, we use the high-resolution gridded daily precipitation dataset from the Asian Precipitation–Highly Resolved Observational Data Integration Towards Evaluation of Water Resources (APHRODITE) project [98]. According to NCAR’s Climate Data Guide (<https://climatedataguide.ucar.edu>), it is the only long-term continental-scale daily product that contains a dense network of daily rain-gauge data for Asia, with a number of valid stations representing 2.3 to 4.5 times the data available through the GTS network. We use the Moonsoon Asia regional product covering 60-150°E and 15S-55°N, at a spatial resolution of 0.25×0.25 , corresponding to a 360×280 longitude/latitude grid. The spatial resolution is confined to 15S-50°N (360×260) to match that of 3B42V7. Similarly, in the time dimension, daily observations are available from 1951-2007, but are restricted to 1998-2007. Only land regions and grid-points with 5 or more clusters are considered, providing 33,975 grid-points for analysis.

4.2.1 TMPA

The TMPA provides a 3-hourly, nearly real-time, gridded precipitation product (3B42RT) with near-global spatial coverage (60°N-60°S) at a spatial resolution of $0.25^\circ \times 0.25^\circ$, and a gauge-adjusted, post-real time research version product (3B42), with a more limited coverage of 50°N-50°S. The final version of these products (version 7) was made available in June 2011, covering the years 1998-2014. The recently launched Global Precipitation Measurement (GPM) mission, initiated by NASA and the Japan Aerospace Exploration Agency (JAXA) will act as a successor to the TRMM mission. In this study we focus on the research version (3B42V7).

The following description of the primary sources of input data for the TMPA products follows [45] closely. These primary sources are provided by two different sets of sensors. The first involves passive microwave data collected by low earth orbit satellites, which include the Microwave Imager (TMI) on TRMM, the Special Sensor Microwave Imager (SSM/I) on Defense Meteorological Satellite Program (DMSP) satellites, Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) on Aqua, and the Advanced Microwave Sounding Unit-B (AMSU-B) on the NOAA satellite series. A strong physical relationship exists between such data and the hydrometeors that result in surface precipitation, however, significant gaps exist in the spatial coverage.

The passive microwave fields of view (FOVs) from TMI, AMSR-E, and SSM/I are converted to precipitation estimates at the TRMM Science Data and Information System (TSDIS) using sensor-specific versions of the Goddard Profiling Algorithm, which aims to reconstruct the observed radiances for each FOV by selecting the best combination of thousands of numerical modelgen-

erated microwave channel upwelling radiances. The associated vertical profiles of hydrometeors are then used to provide an estimated surface precipitation rate.

Passive microwave FOVs from AMSU-B are converted to precipitation estimates at the National Environmental Satellite, Data, and Information Service (NESDIS). Precipitation rates are obtained from ice water path estimates using relationships derived from cloud-model data. A drawback of the AMSU-B algorithm is that it cannot distinguish between precipitation and non-precipitation when clouds lack the ice phase. The above multichannel conically scanning passive microwave sensors share the limitation over land, however, over oceans, these can sense liquid hydrometeors in addition to the precipitation from ice-bearing clouds.

The second major source of input data for the TMPA is infrared data collected by the international constellation of geosynchronous earth orbit (GEO) satellites and merged by the CPC of the NOAA/National Weather Service. Although the GEO-IR data provides superior time-space coverage compared to passive microwave data, the infrared brightness temperatures (T_b) correspond primarily to cloud-top temperatures resulting in precipitation estimates which have been shown to correlated poorly with fine space/time scale precipitation.

The research product incorporates three additional data sources: the TRMM Combined Instrument (TCI) estimate, which employs data from both TMI and the TRMM precipitation radar, the GPCP monthly rain gauge analysis developed by the Global Precipitation Climatological Center (GPCC) and the Climate Assessment and Monitoring System (CAMS) monthly rain gauge analysis developed by CPC.

The main changes of version 7 of the TMPA relative to version 6 are the incorporation of additional satellite observations, the use of the new Global Precipitation Climatology Centre (GPCC) “full” gauge analysis whenever available, and the GPCC “monitoring” gauge analysis since 2010; and a new IR brightness temperature dataset for the period before the start of the CPC 4km Merged Global IR Dataset [45, 19].

Details about the TMPA algorithm can be found in [45].

4.3 Model

4.3.1 First stage

Let Y_{it} denote the daily accumulated precipitation at grid-point i and time t , and assume that the values that exceed the location specific threshold u_i follow the point process model outlined in section 4.1, with parameters μ_i , σ_i , and ξ_i . As was noted in chapter 2, [43] revealed that for small sample sizes, parameter values for the GEV are more reliably estimated by probability-weighted moments techniques than by likelihood methods. Subsequently, [66] added that uncertainties in the estimation of upper quantiles of the GEV distribution were linked to poor estimates of the shape parameter and proposed a new likelihood to restrain the values of this parameter which resulted in quantile estimates that were comparable or superior to that of [43]. [23] found that placing the prior in [66], a beta density with support $[-0.5, 0.5]$ and mean 0.1, on the likelihood (4.1) resulted in more stable shape parameter estimates. Here, as a first approach, we use (4.1) as the first-stage likelihood.

4.3.2 Second stage

We propose the following specification for the GEV parameters,

$$\begin{aligned}
 \mu_i | \beta_\mu, \phi_i^\mu, \sigma_\mu^2 &\sim N(X_i^\top \beta_\mu + \phi_i^\mu, \sigma_\mu^2) \\
 \log(\sigma_i) | \beta_\sigma, \phi_i^\sigma, \sigma_\sigma^2 &\sim N(X_i^\top \beta_\sigma + \phi_i^\sigma, \sigma_\sigma^2) \\
 \xi_i | \beta_\xi, \phi_i^\xi, \sigma_\xi^2 &\sim N(X_i^\top \beta_\xi + \phi_i^\xi, \sigma_\xi^2),
 \end{aligned} \tag{4.2}$$

with $X_i = (1, lon_i, lat_i, elev_i)$ a vector of covariate information specifying the longitude, latitude and elevation of location i , and β the corresponding vector, for each parameter, of regression coefficients. Modelling interest focuses on $\phi_i = (\phi_i^\mu, \phi_i^\sigma, \phi_i^\xi)$, a vector of spatial random effects. The choice of a Gaussian specification for the parameters, as in [23] and [77], is due to its computational convenience in the Markov Chain Monte Carlo sampling framework. The variances σ_μ^2 , σ_σ^2 , and σ_ξ^2 are chosen to be small, in order for the variability of the parameter values to be captured primarily by the spatial random effects.

The collection of vectors of spatial random effects $\{\phi_i\}$ will provide the desired spatial dependence that will allow the model to borrow strength across neighboring locations. We use conditionally autoregressive (CAR) models [10] given that we are dealing with areal data and because their conditional specification makes them ideally suited for Gibbs sampling. We provide a brief description of these and their multivariate extension below following [3].

Conditionally Autoregressive Models

In this section we provide a modern treatment of the auto-normal scheme of [10], now commonly referred to as the Gaussian CAR model, based on the recent text [3] (hereafter BCG). Given random variables Y_1, \dots, Y_n associated with areal

measurements, let

$$Y_i|y_j, j \neq i \sim N\left(\sum_j b_{ij}y_j, \tau_i^2\right), \quad i = 1, \dots, n. \quad (4.3)$$

By Brook's Lemma it can be shown that resulting joint distribution is of the form

$$p(y_1, \dots, y_n) \propto \exp\left\{-\frac{1}{2}\mathbf{y}'D^{-1}(I - B)\mathbf{y}\right\} \quad (4.4)$$

where $B = \{b_{ij}\}$ and D is diagonal with $D_{ii} = \tau_i^2$. Clearly (4.4) is a joint multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma_{\mathbf{y}} = (I - B)^{-1}D$. To ensure the symmetry of $\Sigma_{\mathbf{y}}$, an additional condition is imposed,

$$\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2}, \quad \forall i, j. \quad (4.5)$$

Note that B need not be symmetric. BCG define a *proximity matrix* W to encode the neighborhood structure of the set of areal units, with entries w_{ij} spatially connecting sites i and j . Many schemes are possible, the most straightforward being the binary 0-1, where $w_{ij} = 1$ if i and j are neighbors in some sense, denoted by $i \sim j$, and 0 otherwise. It may be that i and j share a boundary or vertex, or more generally, for a given i, j may be one of K -nearest neighbors. The entries of W can also represent a measure of distance, such as a decreasing function of intercentroidal distance between the units. Most often, W is constructed so as to be symmetric, although this need not be the case, for instance, under a K -nearest neighbor scheme with irregular areal units. Assuming we have at our disposal such a symmetric matrix W , we can then set $b_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \tau^2/w_{i+}$, where $w_{i+} := \sum_j w_{ij}$. Then (4.5) is satisfied and (4.4) becomes

$$p(y_1, \dots, y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\mathbf{y}'(D_w - W)\mathbf{y}\right\} \quad (4.6)$$

where D_w is diagonal with $(D_w)_{ii} = w_{i+}$.

But now $(D_w - W)\mathbf{1} = \mathbf{0}$, that is, Σ_y^{-1} is singular and (4.6) defines an improper distribution. The impropriety becomes more evident if we rewrite (4.6) as follows,

$$p(y_1, \dots, y_n) \propto \exp \left\{ -\frac{1}{2\tau^2} \sum_{i \neq j} w_{ij} (y_i - y_j)^2 \right\}, \quad (4.7)$$

since one can add a constant to all of the Y_i 's and (4.7) is unaffected. The Y_i are not centered, but a constraint such as $\sum_i Y_i = 0$ would suffice to achieve the needed centering. This provides as an illustration of a model with a proper conditional specification, but improper joint distribution. This specification is often referred to in the literature as the *intrinsically autoregressive* (IAR) model.

It is clear that (4.6) cannot be used as the likelihood function of any spatial process, so its use is limited to a prior distributional specification. In fact, this model belongs to the class of *pairwise-difference* priors described in [11]. In practice, the IAR specification is often attached to random spatial effects at the second stage of a hierarchical model.

BGC point out that the impropriety in (4.6) can be remedied in an obvious way by redefining $\Sigma_y^{-1} = D_w - \rho W$ and choosing $\rho \in \mathbb{R}$ judiciously so that Σ_y^{-1} is non-singular. It can be shown that it suffices to have $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$, where $\lambda_{(1)} < \dots < \lambda_{(n)}$ are the ordered eigenvalues of $D_w^{-1/2} W D_w^{-1/2}$. We note that since $\text{tr}(D_w^{-1/2} W D_w^{-1/2}) = 0 = \sum_i \lambda_{(i)}$, then necessarily $\lambda_{(1)} < 0$ and $\lambda_{(n)} > 0$. Alternatively, we can replace the adjacency matrix W by its scaled version $\tilde{W} := \text{diag}(1/w_{i+})W$. Then Σ_y^{-1} can be written as $M^{-1}(I - \alpha \tilde{W})$, with M diagonal, and if $|\alpha| < 1$ then $I - \alpha \tilde{W}$ is non-singular. It can be shown that Σ_y^{-1} is diagonally dominant and symmetric, hence positive definite as desired.

The multivariate extensions of the CAR model consider $\boldsymbol{\phi}^\top = (\phi_1, \dots, \phi_n)$, where each $\phi_i = (\phi_{i1}, \dots, \phi_{ip})^\top$, is $p \times 1$. Most multivariate CAR (MCAR) models

derive from [65]. Analogous to the univariate case, the conditional distributions are of the form

$$p(\boldsymbol{\phi}_i | \boldsymbol{\phi}_{j \neq i}, \Gamma_i) = N \left(\sum_{i \sim j} B_{ij} \boldsymbol{\phi}_j, \Gamma_i \right), \quad i, j = 1, \dots, n$$

where Γ_i and B_{ij} are $p \times p$ matrices. Using a multivariate analogue of Brook's Lemma one obtains the joint distribution

$$p(\boldsymbol{\phi} | \{\Gamma_i\}) \propto \exp \left\{ -\frac{1}{2} \boldsymbol{\phi}^\top \Gamma^{-1} (I - \widetilde{B}) \boldsymbol{\phi} \right\}$$

with Γ block-diagonal with blocks Γ_i and \widetilde{B} is $np \times np$ with (i, j) -th block B_{ij} . Again symmetry of $\Gamma^{-1}(I - \widetilde{B})$ is required, but a convenient special case sets $B_{ij} = b_{ij} I_{p \times p}$, which yields a symmetry condition $b_{ij} \Gamma_j = b_{ji} \Gamma_i$. As in the univariate case, we assume $b_{ij} = w_{ij}/w_{i+}$ and $\Gamma_i = \Sigma/w_{i+}$.

Setting $\widetilde{B} = B \otimes I$, and $\Gamma = D_w^{-1} \otimes \Sigma$, with D_w as above, we may express $\Gamma^{-1}(I - \widetilde{B})$ as $(D_w - W) \otimes \Sigma^{-1}$.

Here we take $p = 3$ and assume a binary scheme for the proximity matrix W , with $w_{ij} = 1$ if i and j share a common border and 0 otherwise.

4.3.3 Third stage

We assign conjugate priors to the β 's and Σ . The β 's are given independent multivariate normal priors, with the zero mean, except for the intercept terms, which are set to the areal mean of the grid-point maximum likelihood parameter estimates, while the covariances are assumed diagonal with large variances. Σ is assigned an Inverse-Wishart prior with 3 degrees of freedom and a diagonal scale matrix with entries reflecting the variability of the maximum likelihood parameter estimates, as in [23].

4.4 Implementation

Model fitting is carried out with Markov Chain Monte Carlo methods. In particular, we use Gibbs sampling for all model parameter updating, with the exception of the GEV parameters, for which the Metropolis-Hasting algorithm is used.

It is common in the literature for these MCMC methods to be implemented in R, or other high-level languages such as Matlab, that provide built-in data structures and functionality that allow for their relative quick implementation, in addition to scores of community-developed packages (in the case of R) that build on the basic functionality as well as on that of each other to further facilitate model implementation. The inevitable cost of these additional layers of functionality is computational speed. While current computer processors, even the most basic, have multiple cores that enable the parallel execution of multiple Markov chains, the processing time of a single chain of an MCAR model such as the one described here, of length 10,000 would typically be upwards of 12 hours (e.g. [23],[77]). For several decades, computing devices relied primarily on increases in the processor's clock speed for improved performance, but in recent years, fundamental limitations in the fabrication of integrated circuits have slowed the pace of increase in clock speed, forcing manufacturers to adopt multicore architectures to enhance performance [76]. However, the graphical processing unit (GPU) is specialized for compute-intensive, highly parallel computation, with its focus on graphics rendering which devotes more transistors to data processing rather than data caching and flow control as in the central processing unit (CPU) [25]. This architecture allows the GPU to achieve orders of magnitude differences in floating-point operations per second, relative to the

CPU [25]. Since the introduction of general-purpose computing platforms for GPU's, like NVIDIA's CUDA in 2006, GPU computing has seen wide application in many scientific disciplines such as medical imaging, computational fluid dynamics and environmental sciences [76]. In this work, we implement the MCMC methods using the NVIDIA CUDA C programming language on a computer with an NVIDIA GeForce GT 750M GPU. We achieved considerable improvements in execution time, as the MCMC fitting over the the Moonsoon Asia region in the CUDA C implementation ran in only 35 minutes, representing a roughly 50× speed up compared to same implementation in R.

We used two parallel chains for each dataset of length 10,000, with thinning every third iteration, after a burn in period of 10,000 iterations. Trace plots of parameters indicated good convergence to the the respective stationary distributions.

4.4.1 MCMC Implementation

- We sample $\theta_i = (\mu_i, \log(\sigma_i), \xi_i)$ independently using a Metropolis-Hasting step for each $i = 1, \dots, n$.

$$p(\theta_i | \mathbf{Y}, \beta, \phi_i, \Sigma_\theta) \propto \left(\prod_t L(y_{it}; \mu_i, \sigma_i, \xi_i) \right) N(\beta X_i + \phi_i, \Sigma_\theta).$$

where $\beta = (\beta_\mu, \beta_\sigma, \beta_\xi)$, $\phi_i = (\phi_i^\mu, \phi_i^\sigma, \phi_i^\xi)$ and Σ_θ is a diagonal matrix with elements σ_μ^2 , σ_σ^2 , and σ_ξ^2 . L is as in (4.1).

- We sample from β_μ , β_σ , and β_ξ independently. We specify the full conditional distribution of β_μ below; those β_σ and β_ξ are analogous.

$$\begin{aligned} p(\beta_\mu | \{\mu_i\}, \{\phi_i^\mu\}, \sigma_\mu^2, \mu_{\beta_\mu}, \Sigma_{\beta_\mu}) &\propto \left(\prod_i N(X_i \beta_\mu + \phi_i^\mu, \sigma_\mu^2) \right) N(\mu_{\beta_\mu}, \Sigma_{\beta_\mu}) \\ &\sim N(m, \Lambda) \end{aligned}$$

with $m = \Lambda(1/\sigma_\mu^2 X^\top(\mu - \phi^\mu) + \Sigma_{\beta_\mu}^{-1} \mu_{\beta_\mu})$, $\Lambda^{-1} = 1/\sigma_\mu^2 X^\top X + \Sigma_{\beta_\mu}^{-1}$, $\mu = (\mu_1, \dots, \mu_n)$, $\phi^\mu = (\phi_1^\mu, \dots, \phi_n^\mu)$, and X an $n \times 4$ matrix given by stacking the X_i 's.

- We update each ϕ_i as follows,

$$\begin{aligned} p(\phi_i | \phi_{j \neq i}, \theta_i, \beta) &\propto N(\theta_i; \beta X_i + \phi_i, \Sigma_\theta) N\left(\sum_{i \sim j} w_{ij}/w_{i+}, \Sigma/w_{i+}\right) \\ &\sim N(m, \Lambda) \end{aligned}$$

with $m = \Gamma\left[\Sigma_\theta^{-1}(\theta_i - \beta X_i) + \Sigma(\sum_{i \sim j} w_{ij} \phi_j)\right]$ and $\Lambda^{-1} = w_{i+} \Sigma^{-1} + \Sigma_\theta^{-1}$.

- Lastly, sampling from the full conditional distribution Σ amounts to sampling from an Inv-Wishart,

$$\begin{aligned} p(\Sigma | \phi, \nu_0, \Omega_0) &= N(0, \phi^\top (D_w - W) \otimes \Sigma^{-1} \phi) \text{Inv-Wishart}_{\nu_0}(\Omega_0) \\ &\sim \text{Inv-Wishart}_{\nu_0+n}\left(\sum_i \sum_j (D_w - W)_{ij} \phi_j \phi_i^\top + \Omega_0\right) \end{aligned}$$

4.5 Results

Figure 4.1 displays 1998-2007 location parameter estimates obtained from the Bayesian hierarchical fitting and by maximum likelihood at each gridpoint. The benefits of using the latter procedure are evident. The parameter values exhibit a greater degree of spatial smoothness, and more importantly, much of the dry areas of northern China now have reasonable parameter values assigned to them, thanks to the borrowing of strength across the spatial domain. We expect that a richer neighborhood structure than the basic binary scheme used here may achieve a more complete spatial coverage.

Prior to exploring differences in return levels of precipitation extremes, we examine differences in the individual parameter values in Figure 4.2, as represented by the respective posterior means. Considerable differences arise in

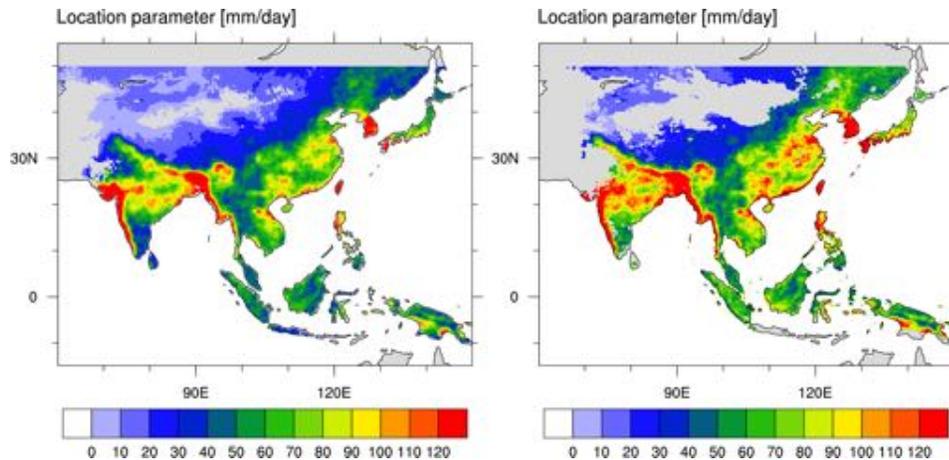


Figure 4.1: Location parameter estimates based on the (left) Bayesian hierarchical framework (posterior mean) and (right) maximum likelihood estimation performed at each grid-point, for JJA.

the location and scale parameter estimates, as indicated by their area-averages: 46.4 mm/day and 19.6 mm/day for the TRMM location and scale, respectively, compared to 22 mm/day and 8.1 mm/day for APHRO. In each case, the area-average of TRMM is more than double that of APHRO. The location parameter estimates in TRMM agree well with those of APHRO in the drier northern regions, but along the coastlines, TRMM estimates exceed 120 mm/day whereas the corresponding values in APHRO fail to surpass 100 mm/day. Greater discrepancies arise when comparing scale parameter estimates, with most of the southern and eastern parts of the Moonsoon Asia region exceeding 15 mm/day. There is better agreement in shape parameter values, although the spatial pattern in APHRO shows great spatial smoothness.

Figure 4.3 displays the 20-yr return level estimates for APHRO, and the differences with the respective estimates in TRMM. While most 20-yr precipitation extremes in the regions, as depicted in the APHRO dataset, are within 60 mm/day, and only few regions, mostly along coastal regions, exceed the 100

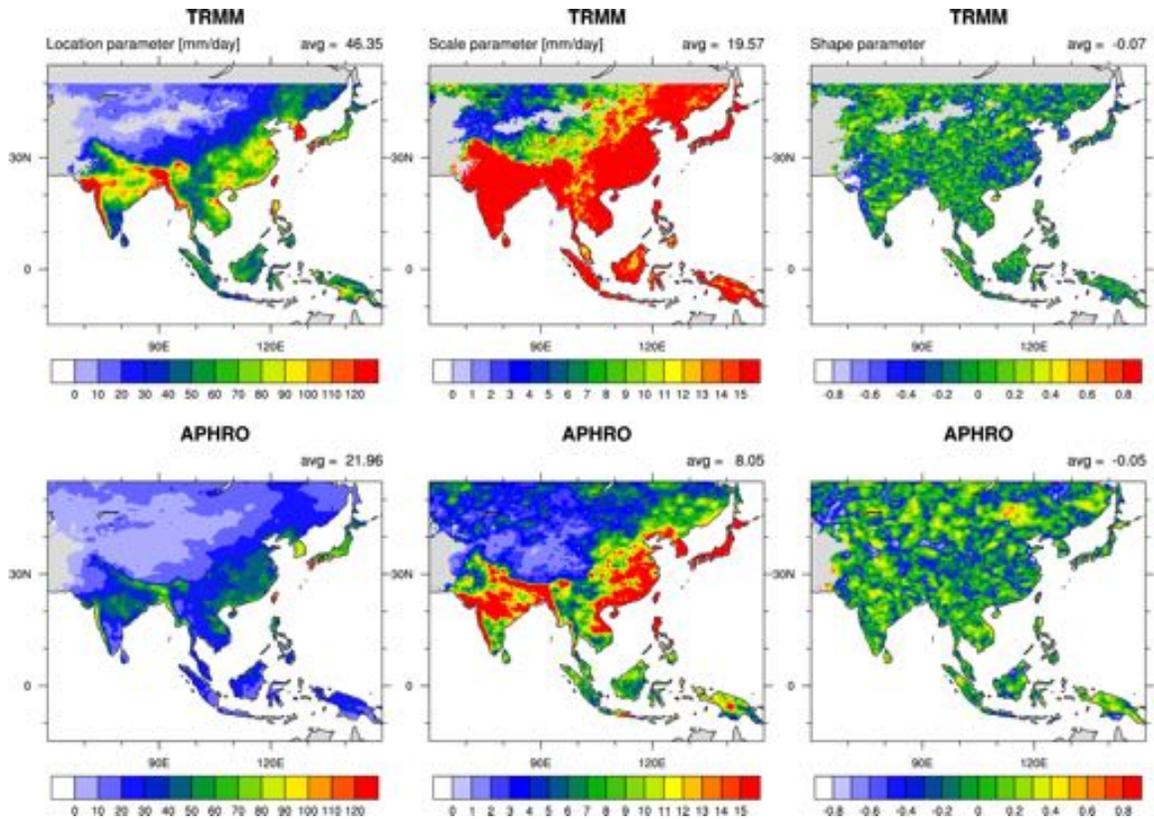


Figure 4.2: Posterior means of (left) location, (middle) scale, and (right) shape parameter estimates based on Bayesian hierarchical framework, for the (top) TRMM 3B42V7 dataset and the (bottom) APHRO dataset, for JJA.

mm/day level, TRMM return levels are on average 50 mm/day higher than those of APHRO. Over those regions where APHRO return levels achieve their maximum values, TRMM return levels are usually 100 mm/day higher. This overestimation calls for extreme caution when using this dataset for validation of precipitation extremes in climate models.

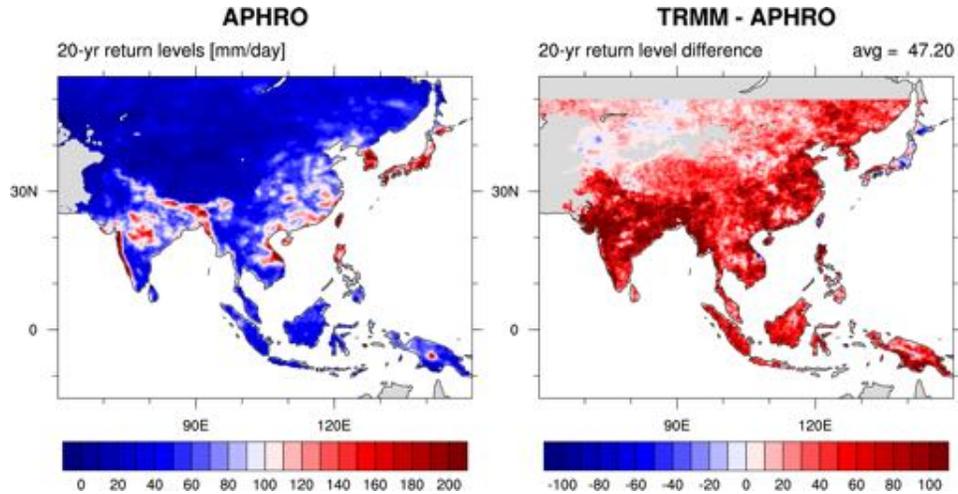


Figure 4.3: (left) JJA 20-yr return level estimates obtained from posterior means of GEV parameter values for APHRO, and (right) the respective differences with the TRMM dataset.

4.6 Summary and Conclusions

This study evaluates precipitation extremes in terms of 20-yr return levels of annual extremes in the satellite-based TRMM 3B42 version 7 research product against the APHRODITE Monsoon Asia gauge-based dataset over the period 1998-2007. Parameter values are obtained by fitting a Bayesian hierarchical model to daily precipitation amounts. The model is implemented using NVIDIA's CUDA C language.

The spatial coverage gain obtained by using the Bayesian hierarchical model, in contrast to pointwise maximum likelihood estimation, is substantial in the arid regions north of the Qinghai-Tibetan plateau. In addition, the Bayesian fitting gives the GEV parameter estimates a desirable level of smoothness across the spatial domain.

20-yr levels exhibit a dramatic overestimation relative to the gauge-based

dataset estimates in the south and eastern parts of the region, while they are consistently represented in the drier areas of the north. This overestimation is primarily due to differences in location and scale parameter estimates, as the shape parameter estimates are in reasonable agreement, indicating that the distribution of annual extremes, as represented by the 3B42 dataset is shifted towards more intense extremes and that these exhibit too much interannual variability. Studies over the region point to good agreement when comparing mean statistics with gauge-based datasets. For instance, over China, [18] reported a correlation coefficient (CC) of 0.93 and relative bias (RB) of -0.05% for 2008-2011 daily mean precipitation with respect to the China Meteorological Administration's national daily precipitation analysis. This suggests that the bias in location parameter estimates are not the result of biases in the mean structure, instead, the errors in the TMPA algorithm appear to be largely confined to the tails of the precipitation distribution.

This study was limited to the Monsoon Asia region over JJA. Future efforts will extend this analysis to DJF and other regions with high-quality gauge-based precipitation products such as CONUS with the CPC dataset, and Europe with the E-OBS dataset from the European Climate Assessment and Dataset project. Since this product will serve as version 0 of the recently launched Global Precipitation Measurement mission (GPM) [44] we hope that these results will assist GPM developers in improving the retrieval algorithms so that this dataset, already in use in a broad range of applications, can also provide an adequate reference for precipitation extreme studies.

BIBLIOGRAPHY

- [1] Lisa V Alexander and Julie M Arblaster. Assessing trends in observed and modelled climate extremes over Australia in relation to future projections. *International Journal of Climatology*, 29(3):417–435, 2009.
- [2] Julio T Bacmeister, Michael F Wehner, Richard B Neale, Andrew Gettelman, Cecile Hannay, Peter H Lauritzen, Julie M Caron, and John E Truesdale. Exploratory high-resolution climate simulations using the Community Atmosphere Model (CAM). *Journal of Climate*, 27(9):3073–3099, 2014.
- [3] Sudipto Banerjee, Bradley P Carlin, and Alan E Gelfand. *Hierarchical modeling and analysis for spatial data*. CRC Press, 2014.
- [4] J Berner, FJ Doblas-Reyes, TN Palmer, G Shutts, and A Weisheimer. Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366(1875):2559–2577, 2008.
- [5] J Berner, KR Fossell, S-Y Ha, JP Hacker, and C Snyder. Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Monthly Weather Review*, 143:1295–1320, 2015.
- [6] J Berner, S-Y Ha, JP Hacker, A Fournier, and C Snyder. Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Monthly Weather Review*, 139(6):1972–1995, 2011.
- [7] J Berner, Thomas Jung, and TN Palmer. Systematic model error: the impact of increased horizontal resolution versus improved stochastic and deterministic parameterizations. *Journal of Climate*, 25(14):4946–4962, 2012.
- [8] J Berner, GJ Shutts, M Leutbecher, and TN Palmer. A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *Journal of the Atmospheric Sciences*, 66(3):603–626, 2009.
- [9] Judith Berner. Linking nonlinearity and non-Gaussianity of planetary wave behavior by the Fokker-Planck equation. *Journal of the atmospheric sciences*, 62(7):2098–2117, 2005.

- [10] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236, 1974.
- [11] Julian Besag, Peter Green, David Higdon, and Kerrie Mengersen. Bayesian computation and stochastic systems. *Statistical science*, pages 3–41, 1995.
- [12] N.L. Bindoff, P.A. Stott, K.M. AchutaRao, M.R. Allen, N. Gillett, D. Gutzler, K. Hansingo, G. Hegerl, Y. Hu, S. Jain, I.I. Mokhov, J. Overland, J. Perlwitz, R. Sebbari, and X. Zhang. *Detection and Attribution of Climate Change: from Global to Regional*, book section 10, page 867952. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- [13] SJ Brown, J Caesar, and Christopher AT Ferro. Global changes in extreme daily temperature since 1950. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 113(D5), 2008.
- [14] Roberto Buizza, M Milleer, and TN Palmer. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 125(560):2887–2908, 1999.
- [15] Melissa S Bukovsky. Temperature trends in the NARCCAP regional climate models. *Journal of Climate*, 25(11):3985–3991, 2012.
- [16] Yancong Cai, Changjie Jin, Anzhi Wang, Dexin Guan, Jiabing Wu, Fenghui Yuan, and Leilei Xu. Comprehensive precipitation evaluation of TRMM 3B42 with dense rain gauge networks in a mid-latitude basin, northeast, china. *Theoretical and Applied Climatology*, pages 1–13, 2015.
- [17] Richard Chandler and Marian Scott. *Statistical methods for trend detection and analysis in the environmental sciences*. John Wiley & Sons, 2011.
- [18] Sheng Chen, Yang Hong, Qing Cao, Jonathan J Gourley, Pierre-Emmanuel Kirstetter, Bin Yong, Yudong Tian, Zengxin Zhang, Yan Shen, Junjun Hu, et al. Similarity and difference of the two successive v6 and v7 trmm multisatellite precipitation analysis performance over china. *Journal of Geophysical Research: Atmospheres*, 118(23):13–060, 2013.
- [19] Sheng Chen, Yang Hong, Jonathan J Gourley, George J Huffman, Yudong Tian, Qing Cao, Bin Yong, Pierre-Emmanuel Kirstetter, Junjun Hu, Jill

- Hardy, et al. Evaluation of the successive v6 and v7 trmm multisatellite precipitation analysis over the continental united states. *Water Resources Research*, 49(12):8174–8186, 2013.
- [20] Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.
- [21] Bruno Collischonn, Walter Collischonn, and Carlos Eduardo Morelli Tucci. Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. *Journal of Hydrology*, 360(1):207–216, 2008.
- [22] Francesco Lo Conti, Kuo-Lin Hsu, Leonardo Valerio Noto, and Soroosh Sorooshian. Evaluation and comparison of satellite precipitation estimates with reference to a local area in the Mediterranean Sea. *Atmospheric Research*, 138:189–204, 2014.
- [23] Daniel Cooley and Stephan R Sain. Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of agricultural, biological, and environmental statistics*, 15(3):381–402, 2010.
- [24] RC Cornes and PD Jones. How well does the ERA-Interim reanalysis replicate trends in extremes of surface temperature across Europe? *Journal of Geophysical Research: Atmospheres*, 118(18):10–262, 2013.
- [25] NVIDIA Corporation. *CUDA C Programming guide*. 2015.
- [26] Aiguo Dai. Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, 19(18):4605–4630, 2006.
- [27] Aiguo Dai, Xin Lin, and Kuo-Lin Hsu. The frequency, intensity, and diurnal cycle of precipitation in surface and satellite observations over low- and mid-latitudes. *Climate dynamics*, 29(7-8):727–744, 2007.
- [28] DP Dee, SM Uppala, AJ Simmons, P Berrisford, P Poli, S Kobayashi, U Andrae, MA Balmaseda, G Balsamo, P Bauer, et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, 2011.
- [29] Noah S Diffenbaugh, Jeremy S Pal, Robert J Trapp, and Filippo Giorgi. Fine-scale processes regulate the response of extreme events to global climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44):15774–15778, 2005.

- [30] FJ Doblas-Reyes, A Weisheimer, M Déqué, Noel Keenlyside, M McVean, JM Murphy, P Rogel, D Smith, and TN Palmer. Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 135(643):1538–1559, 2009.
- [31] Markus G Donat, Jana Sillmann, Simon Wild, Lisa V Alexander, Tanya Lippmann, and Francis W Zwiers. Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *Journal of Climate*, 27(13):5019–5035, 2014.
- [32] MG Donat, LV Alexander, H Yang, I Durre, R Vose, RJH Dunn, KM Willett, E Aguilar, M Brunet, J Caesar, et al. 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, 2013.
- [33] David R Easterling, Gerald A Meehl, Camille Parmesan, Stanley A Changnon, Thomas R Karl, and Linda O Mearns. Climate extremes: observations, modeling, and impacts. *Science*, 289(5487):2068–2074, 2000.
- [34] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*, volume 57. CRC press, 1994.
- [35] G. Flato, J. Marotzke, B. Abiodun, P. Braconnot, S.C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason, and M. Rummukainen. *Evaluation of Climate Models*, book section 9, page 741866. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- [36] CK Folland, TR Karl, JR Christy, R Clarke, G Gruza, J Jouzel, M Mann, J Oerlemans, M Salinger, and S Wang. Observed climate variability and change. in: Climate change 2001: the scientific basis. *Contribution of Working Group I to the IPCC Third Assessment Report*, Cambridge University Press, Cambridge, pages 99–181, 2001.
- [37] Christian L. E. Franzke, Terence J. O’Kane, Judith Berner, Paul D. Williams, and Valerio Lucarini. Stochastic climate theory and modeling. *Wiley Interdisciplinary Reviews: Climate Change*, 6(1):63–78, 2015.
- [38] Crispin W Gardiner. *Stochastic Methods, A Handbook for the Natural and Social Sciences*, volume 13 of *Springer Series in Synergetics*. Springer-Verlag Berlin Heidelberg, 4 edition, 2009.

- [39] W Lawrence Gates. AMIP: The atmospheric model intercomparison project. *Bulletin of the American Meteorological Society*, 73(12):1962–1970, 1992.
- [40] F Giorgi, E-S Im, E Coppola, NS Diffenbaugh, XJ Gao, L Mariotti, and Y Shi. Higher hydroclimatic intensity with global warming. *Journal of Climate*, 24(20):5309–5324, 2011.
- [41] Alemseged Tamiru Haile, Emad Habib, Mohamed Elsaadani, and Tom Rientjes. Inter-comparison of satellite rainfall products for representing rainfall diurnal cycle over the Nile basin. *International journal of applied earth observation and geoinformation*, 21:230–240, 2013.
- [42] D.L. Hartmann, A.M.G. Klein Tank, M. Rusticucci, L.V. Alexander, S. Bronnimann, Y. Charabi, F.J. Dentener, E.J. Dlugokencky, D.R. Easterling, A. Kaplan, B.J. Soden, P.W. Thorne, M. Wild, and P.M. Zhai. *Observations: Atmosphere and Surface*, book section 2, page 159254. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- [43] JRM Hosking, James R Wallis, and Eric F Wood. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261, 1985.
- [44] Arthur Y Hou, Ramesh K Kakar, Steven Neeck, Ardeshir A Azarbarzin, Christian D Kummerow, Masahiro Kojima, Riko Oki, Kenji Nakamura, and Toshio Iguchi. The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, 95(5):701–722, 2014.
- [45] George J Huffman, David T Bolvin, Eric J Nelkin, David B Wolff, Robert F Adler, Guojun Gu, Yang Hong, Kenneth P Bowman, and Erich F Stocker. The trmm multisatellite precipitation analysis (tmpa): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8(1):38–55, 2007.
- [46] Thomas Jung, G Balsamo, P Bechtold, ACM Beljaars, M Köhler, MJ Miller, J-J Morcrette, A Orr, MJ Rodwell, and AM Tompkins. The ECMWF model climate: Recent progress through improved physical parametrizations. *Quarterly Journal of the Royal Meteorological Society*, 136(650):1145–1160, 2010.
- [47] Thomas Jung, MJ Miller, TN Palmer, P Towers, N Wedi, D Achuthavarier, JM Adams, EL Altshuler, BA Cash, JL Kinter Iii, et al. High-resolution

- global climate simulations with the ECMWF model in project athena: Experimental design, model climate, and seasonal forecast skill. *Journal of Climate*, 25(9):3155–3172, 2012.
- [48] Thomas Jung, TN Palmer, and GJ Shutts. Influence of a stochastic parameterization on the frequency of occurrence of north pacific weather regimes in the ECMWF model. *Geophysical research letters*, 32(23), 2005.
- [49] Eugenia Kalnay, M Kanamitsu, R Kistler, W Collins, D Deaven, L Gandin, Mo Iredell, S Saha, G White, J Woollen, et al. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, 77(3):437–471, 1996.
- [50] Richard W Katz, Marc B Parlange, and Philippe Naveau. Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8):1287–1304, 2002.
- [51] JE Kay, C Deser, A Phillips, A Mai, C Hannay, G Strand, JM Arblaster, SC Bates, G Danabasoglu, J Edwards, et al. The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 2014.
- [52] Viatcheslav V Kharin and Francis W Zwiers. Estimating extremes in transient climate change simulations. *Journal of Climate*, 18(8):1156–1173, 2005.
- [53] Viatcheslav V Kharin, Francis W Zwiers, Xuebin Zhang, and Gabriele C Hegerl. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *Journal of Climate*, 20(8):1419–1444, 2007.
- [54] Viatcheslav V Kharin, FW Zwiers, X Zhang, and M Wehner. Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, 119(2):345–357, 2013.
- [55] Chris Kidd and George Huffman. Global precipitation measurement. *Meteorological Applications*, 18(3):334–353, 2011.
- [56] AMG Klein Tank and GP Können. Trends in indices of daily temperature and precipitation extremes in Europe, 1946-99. *Journal of Climate*, 16(22):3665–3680, 2003.
- [57] AMG Klein Tank, FW Zwiers, and X Zhang. Guidelines on analysis of

extremes in a changing climate in support of informed decisions for adaptation, World Meteorological Organization. *Geneva, Switzerland*, 56, 2009.

- [58] Pushkar Kopparla, Erich M Fischer, Cécile Hannay, and Reto Knutti. Improved simulation of extreme precipitation in a high-resolution atmosphere model. *Geophysical Research Letters*, 40(21):5803–5808, 2013.
- [59] Kenneth E Kunkel, Roger A Pielke Jr, and Stanley A Changnon. Temporal fluctuations in weather and climate extremes that cause economic and human health impacts: A review. *Bulletin of the American Meteorological Society*, 80(6):1077–1098, 1999.
- [60] Peter H Lauritzen, Christiane Jablonowski, Mark A Taylor, and Ramachandran D Nair. *Numerical techniques for global atmospheric models*, volume 80. Springer Science & Business Media, 2011.
- [61] M Ross Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer, 1983.
- [62] Xiang-Hu Li, Qi Zhang, and Chong-Yu Xu. Suitability of the TRMM satellite rainfalls in driving a distributed hydrological model for water balance computations in Xinjiang catchment, Poyang lake basin. *Journal of Hydrology*, 426:28–38, 2012.
- [63] Zhong Liu, Dana Ostrenga, William Teng, and Steven Kempler. Tropical rainfall measuring mission (trmm) precipitation data and services for research and applications. *Bulletin of the American Meteorological Society*, 93(9):1317–1325, 2012.
- [64] Robert E Livezey and WY Chen. Statistical field significance and its determination by monte carlo techniques. *Monthly Weather Review*, 111(1):46–59, 1983.
- [65] KV Mardia. Multi-dimensional multivariate gaussian markov random fields with application to image processing. *Journal of Multivariate Analysis*, 24(2):265–284, 1988.
- [66] Eduardo S Martins and Jery R Stedinger. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3):737–744, 2000.

- [67] PJ Mason and DJ Thomson. Stochastic backscatter in large-eddy simulations of boundary layers. *Journal of Fluid Mechanics*, 242:51–78, 1992.
- [68] Saber Moazami, Saeed Golian, M Reza Kavianpour, and Yang Hong. Comparison of PERSIANN and V7 TRMM Multi-satellite Precipitation Analysis (TMPA) products with rain gauge data over iran. *International journal of remote sensing*, 34(22):8156–8171, 2013.
- [69] National Assessment Synthesis Team (NAST). Climate change impacts on the united states the potential consequences of climate variability and change. 2000.
- [70] Richard B Neale, Jadwiga Richter, Andrew Conley, Sungsu Park, Andrew Gettelman, et al. Description of the NCAR community atmosphere model (CAM 4.0). *NCAR Tech. Note NCAR/TN-485+ STR*, 2010.
- [71] Mark New, Martin Todd, Mike Hulme, and Phil Jones. Precipitation measurements and trends in the twentieth century. *International Journal of Climatology*, 21(15):1889–1922, 2001.
- [72] Tim N Palmer. A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society*, 127(572):279–304, 2001.
- [73] TN Palmer, R Buizza, F Doblas-Reyes, Thomas Jung, M Leutbecher, GJ Shutts, M Steinheimer, and A Weisheimer. *Stochastic parametrization and model uncertainty*. European Centre for Medium-Range Weather Forecasts, 2009.
- [74] TN Palmer and Antje Weisheimer. Diagnosing the causes of bias in climate models—why is it so hard? *Geophysical & Astrophysical Fluid Dynamics*, 105(2-3):351–365, 2011.
- [75] Sara A Rauscher, Erika Coppola, Claudio Piani, and Filippo Giorgi. Resolution effects on regional climate model simulations of seasonal precipitation over Europe. *Climate dynamics*, 35(4):685–711, 2010.
- [76] Jason Sanders and Edward Kandrot. *CUDA by example: an introduction to general-purpose GPU programming*. Addison-Wesley Professional, 2010.
- [77] Huiyan Sang and Alan E Gelfand. Hierarchical modeling for extreme val-

- ues observed over space and time. *Environmental and Ecological Statistics*, 16(3):407–426, 2009.
- [78] Pranab Kumar Sen. Estimates of the regression coefficient based on Kendall’s tau. *Journal of the American Statistical Association*, 63(324):1379–1389, 1968.
- [79] Sonia I Seneviratne, Neville Nicholls, David Easterling, Clare M Goodess, Shinjiro Kanae, James Kossin, Yali Luo, Jose Marengo, Kathleen McInnes, Mohammad Rahimi, et al. 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*, pages 109–230, 2012.
- [80] Glenn Shutts. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 131(612):3079–3102, 2005.
- [81] J Sillmann, VV Kharin, X Zhang, FW Zwiers, and D Bronaugh. 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, 2013.
- [82] J Sillmann, VV Kharin, FW Zwiers, X Zhang, D Bronaugh, and MG Donat. Evaluating model-simulated variability in temperature extremes using modified percentile indices. *International Journal of Climatology*, 34(11):3304–3311, 2014.
- [83] Jana Sillmann, Markus G Donat, John C Fyfe, and Francis W Zwiers. Observed and simulated temperature extremes during the recent warming hiatus. *Environmental Research Letters*, 9(6):064023, 2014.
- [84] Deepti Singh, Michael Tsiang, Bala Rajaratnam, and Noah S Diffenbaugh. Precipitation extremes over the continental united states in a transient, high-resolution, ensemble climate model experiment. *Journal of Geophysical Research: Atmospheres*, 118(13):7063–7086, 2013.
- [85] Graeme L Stephens, Tristan L’Ecuyer, Richard Forbes, Andrew Gettleman, Jean-Christophe Golaz, Alejandro Bodas-Salcedo, Kentaroh Suzuki, Philip Gabriel, and John Haynes. Dreary state of precipitation in global models. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 115(D24), 2010.

- [86] David B Stephenson, HF Diaz, and RJ Murnane. *Definition, diagnosis, and origin of extreme weather and climate events*, volume 348. Cambridge University Press: New York, 2008.
- [87] Felipe Tagle, Judith Berner, Mircea D Grigoriu, Natalie M Mahowald, and Gennady Samorodnitsky. Temperature extremes in the community atmosphere model with stochastic parameterizations. *Journal of Climate*, 29, 2015.
- [88] Karl E Taylor, Ronald J Stouffer, and Gerald A Meehl. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4):485–498, 2012.
- [89] Claudia Tebaldi, Katharine Hayhoe, Julie M Arblaster, and Gerald A Meehl. Going to the extremes. *Climatic change*, 79(3-4):185–211, 2006.
- [90] L Tomassini and D Jacob. Spatial analysis of trends in extreme precipitation events in high-resolution climate model results and observations for germany. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 114(D12), 2009.
- [91] Sakari M Uppala, PW Killberg, AJ Simmons, U Andrae, V Bechtold, M Fiorino, JK Gibson, J Haseler, A Hernandez, GA Kelly, et al. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612):2961–3012, 2005.
- [92] Xiaolan L Wang and Val R Swail. Changes of extreme wave heights in northern hemisphere oceans and related atmospheric circulation regimes. *Journal of Climate*, 14(10):2204–2221, 2001.
- [93] Anthony Weaver and Philippe Courtier. Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, 127(575):1815–1846, 2001.
- [94] Michael F Wehner. Predicted twenty-first-century changes in seasonal extreme precipitation events in the parallel climate model. *Journal of climate*, 17(21):4281–4290, 2004.
- [95] Michael F Wehner, Kevin A Reed, Fuyu Li, Julio Bacmeister, Cheng-Ta Chen, Christopher Paciorek, Peter J Gleckler, Kenneth R Sperber, William D Collins, Andrew Gettelman, et al. The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. *Journal of Advances in Modeling Earth Systems*, 2014.

- [96] Michael F Wehner, Richard L Smith, G Bala, and Phillip Duffy. The effect of horizontal resolution on simulation of very extreme us precipitation events in a global atmosphere model. *Climate dynamics*, 34(2-3):241–247, 2010.
- [97] Daniel S Wilks. *Statistical methods in the atmospheric sciences*, volume 100. Academic press, 2011.
- [98] Akiyo Yatagai, Kenji Kamiguchi, Osamu Arakawa, Atsushi Hamada, Natsuko Yasutomi, and Akio Kitoh. APHRODITE: Constructing a long-term daily gridded precipitation dataset for asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society*, 93(9):1401–1415, 2012.
- [99] PM Zhai and XH Pan. Change in extreme temperature and precipitation over northern China during the second half of the 20th century. *Acta Geographica Sinica*, 58(S1):1–10, 2003.
- [100] Xuebin Zhang, Lisa Alexander, Gabriele C Hegerl, Philip Jones, Albert Klein Tank, Thomas C Peterson, Blair Trewin, and Francis W Zwiers. Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6):851–870, 2011.
- [101] Xuebin Zhang and Francis W Zwiers. Comment on applicability of prewhitening to eliminate the influence of serial correlation on the mann-kendall test by sheng yue and chun yuan wang. *Water resources research*, 40(3), 2004.