DEVELOPMENT OF A GLOBALLY CALIBRATED STOCHASTIC WEATHER GENERATOR WITH THE PATTERN SCALING TECHNIQUE TO STUDY FUTURE CLIMATES

BY SARAH JANE WILSON KEMSLEY

JULY 2022

A thesis submitted in partial fulfilment of the requirements for a degree of

Doctor of Philosophy

to the

School of Environmental Sciences of the University of East Anglia

REGISTRATION NUMBER:100248275

ABSTRACT

The most sophisticated tools for studying future climates are General Circulation Models (GCMs). However, GCMs have biases, coarse resolution, and have not simulated all relevant scenarios. Stochastic weather generators (SWGs) produce synthetic daily time series of weather variables that, when calibrated with local observational data, can address the GCM bias and scale issues. If their parameters are perturbed using specific simulations, projections are still limited to scenarios simulated by the parent GCM. To solve this, the pattern scaling (PS) technique is applied for the first time to estimate SWG parameter perturbations for future scenarios or warming levels, by diagnosing GCM-simulated dependence of SWG parameters on global mean surface temperature (GMST).

A daily SWG is developed for multiple weather variables and calibrated using observations from a range of climates. It uses a Markov-chain gamma-distribution model for precipitation and a regression model conditioned on precipitation for temperatures. A new assessment of Markov-chain model order across Köppen climate regimes finds that optimal model order primarily depends on performance metric. Another innovation is a generalised linear model for mean wind speed, with further regression models for humidity and radiation. The model is collectively referred to as the **G**lobally **C**alibrated stochastic **W**eather **G**enerator (GCWG).

Response patterns of the key precipitation and temperature parameters are diagnosed globally from GCM projections to demonstrate how to combine PS with the GCWG. For example, linear regression has been used to determine the response of first-order Markov probabilities to increasing GMST, with R^2 scores typically exceeding 0.5 over land. Mean daily minimum temperatures show larger increases with GMST than maximums (increasing at 3.4°C/°C compared to 1.3°C/°C). GCWG parameters are hence perturbed by GCM-scale responses to construct local-scale time series for a range of scenarios with the potential to emulate those that have not yet been simulated by GCMs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

CDD/CWD – Consecutive Dry Days per year/Consecutive Wet Days per year

CMIP – Coupled Model Intercomparison Project

CSDI/CS10 – Cold Spell Duration Index/Cold Spell 10^{th} Percentile

DJF – December, January, February

ECAD – European Climate Assessment & Dataset

FAO – Food and Agriculture Organization

GCM – General Circulation Model

GCWG – Globally Calibrated stochastic Weather Generator

GLM – Generalised Linear Model

GMST – Global Mean Surface Temperature

GWL – Global Warming Level

IAV – Interannual Variability

IPCC – International Panel on Climate Change

IVA – Interannual Variability in mean precipitation Amount

IVO – Interannual Variability in precipitation Occurrence

JJA – June, July, August

MAM – March, April, May

MLR – Multiple Linear Regression

PET – Potential Evapotranspiration

PS – Pattern Scaling

RCP – Representative Concentration Pathway

RMSE – Root Mean Squared Error

SON – September, October, November

SSP – Shared Socioeconomic Pathway

SWG – Stochastic Weather Generator

WSDI/WS90 – Warm Spell Duration Index/Warm Spell 90^{th} Percentile

ACKNOWLEDGEMENTS

I would firstly like to thank my supervisors Tim Osborn, Steve Dorling, Craig Wallace, and Jo Parker. They have shown me guidance, patience, and support in making the transition from a Physics graduate to a confident researcher in Environmental Science. For that I am extremely grateful. I would also like to thank EnvEast (grant number NE/S007334/1) and Atkins for the opportunity and expertise offered throughout my journey.

I would also like to thank my bee girls – Lea, Indie, and Emily – for keeping me sane during some of the toughest times in my personal life, the newfound friendship and of course, the cocktails. My oldest friends: Sophie, Daisy, and Phoebe, have too kept me going throughout the madness of undertaking a PhD whilst simultaneously coping with the pandemic. I am grateful for the kindness, support (and cake) shared with Joanna and Ruth throughout. I would also like to thank my PhD cohort for the continued camaraderie and accepting my social flakiness.

Thank you to Jordan, my best friend, partner, and number one superfan. To describe our lives as a soap opera over the last few years would be quite an understatement; though you have continued to make sure I am reminded of my strengths and uplifted me in times of self-doubt throughout and I'm excited to start the next chapter of our lives together! For the love, kindness, friendship (and doughnuts), I thank you for it all.

Finally, I would like to thank my family (and of course my cat, Luna). My younger sisters have kept me humble by consistently surpassing my pre-PhD achievements - you can stop that now. Lastly, I am grateful for the support of my parents. Without their encouragement, words of wisdom, and free food the day before payday, I would not be completing this thesis. So, Dr Wilson and Dr Kemsley; I hope to join you as Dr Wilson Kemsley!

1 INTRODUCTION

Climate change over the last 150 years has exceeded the changes that can be expected due to natural variability, and it is with confidence that anthropogenic greenhouse gas and aerosol emissions can be held responsible (Chen et al., 2021; Stocker et al., 2013). Changes to the climate can have severe implications on several sectors, including agricultural and hydrological management (Arnell et al., 2013), human health and mortality (Luber and McGeehin, 2008), and on tourism, fisheries, and on modifications to land-usage (IPCC, 2022). Furthermore, there is wide agreement that changes to the frequency and severity of extreme events as a function of climate change are likely (Easterling et al., 2000; Seneviratne et al., 2021). It is therefore important that impact assessors, stakeholders who make decisions about adaptation and investment, and policymakers alike have access to climate projections that capture a range of potential future climate scenarios at high temporal and spatial scales. To aid with the construction of a tool that addresses these requirements, this thesis is part of a CASE studentship with industrial partner, Atkins.

General Circulation Models (GCMs) are the most sophisticated tools at our disposal that are used to construct projections representing a range of scenarios, including mitigated and high-forcing pathways (Tebaldi et al., 2021). However, GCMs are constrained by the computational cost, resulting in the parameterisation of several sub-grid cell processes and a spatial resolution that may be too low for impact assessments. This has resulted in the development of several downscaling techniques, including the computationally efficient stochastic weather generators. In instances where GCM resolution is more appropriate for impact assessments (e.g., simulations produced by the recent HighResMIP (Haarsma et al., 2016)), clarity of the forced changes, including their uncertainty, can nevertheless be limited by the low number of realisations. Although a range of possible scenarios representing changes in anthropogenic emissions and socio-economic factors have been explored by GCMs, it is difficult to discern the relative likelihoods of any given scenario, and uncertainty in the future climate cannot be addressed studying a few climate scenarios alone (Grübler and Nakicenovic, 2001). Pattern scaling has been widely used to emulate scenarios and time-periods that GCMs have not produced projections for, providing a technique that helps expand the number of projections available to impact planners and helps address some of the issues associated with using just a limited number of direct GCM simulations. Assuming a linear relationship between global surface temperature change and local climate change in a variable, pattern scaling has typically only been used to construct projections over long

temporal scales (e.g., monthly, seasonally, annually) and usually only addresses changes in local-scale mean climate, not variability.

While stochastic weather generators can be used to downscale GCM resolution, they are also computationally efficient tools that can produce long time series for a suite of weather variables at local scales and for a range of temporal scales (e.g., monthly, daily, sub-daily). They are commonly used to produce temporally consistent time series (between variables as inputs into hydrological, agricultural or ecological models (Semenov et al., 1998). Able to produce long time series with high temporal scales, stochastic weather generators can also be used to robustly assess the risk of extreme weather events. However, many stochastic weather generators (e.g., WGEN (Richardson and Wright, 1984), LARS-WG (Semenov and Barrow, 2002)) that are in broad use have been calibrated only for a sub-region of interest, frequently North America or Europe. These locations are not representative of all the different global climate regimes, where different configurations of the underlying statistical stochastic model may perform better. Furthermore, without some form of perturbation, stochastic weather generators are only able to produce time series that is statistically similar to an observed series.

In this thesis, the pattern scaling technique will be combined with a stochastic weather generator to address some of the aforementioned issues. This has led to the development of a computationally inexpensive tool to construct daily, local-scale time series for a suite of weather variables and a range of future climate scenarios, utilising the strengths of the pattern scaling technique and stochastic weather generators. The main aims of this thesis can be broken down into four primary objectives:

1. To develop a globally calibrated stochastic weather generator, assessed for multiple climatic regimes, and producing a suite of weather variables at a daily temporal scale, namely: precipitation, maximum and minimum temperature, mean wind speed, solar radiation and relative humidity. The stochastic weather generator developed here will be referred to as GCWG (**G**lobally **C**alibrated stochastic **W**eather **G**enerator).
2. To use the pattern scaling technique to diagnose the response of GCWG parameters to a changing climate as simulated by GCMs. Pattern scaling will relate these changes to global mean surface temperature (GMST) and will include parameters accounting for changes in future variability.

3. To apply the response patterns diagnosed in Objective 2 to perturb observed GCWG parameters and hence to generate local-scale and daily timescale future scenario data to better understand climate uncertainty.
4. To study weather extremes using the GCWG under present-day and future global warming levels (GWLs).

1.1 THESIS STRUCTURE

Chapter 2 is a literature review, comprised of four main sections. Observed and projected changes to the frequency and severity of extreme weather events are discussed in the first section. The second and third sections discuss GCMs, stochastic weather generators, and a selection of widely used downscaling techniques. Finally, the fourth section includes a review of pattern scaling strengths, limitations, and uses.

Chapters 3, 4 and 5 address Objective 1, focussing on the development of a globally calibrated stochastic weather generator (GCWG) using a global network of weather station observations and case studies representative of five different climatic regimes (Peel et al., 2007). A first-order Markov-chain gamma-distribution model for precipitation, conditional multiple linear regression model for maximum and minimum temperature, generalised linear model with a log-link function for mean wind speed, and multiple linear regression models for solar radiation and relative humidity are developed, using robust statistical testing models to validate choices.

Chapter 3 mainly comprises of a first-author publication in the *International Journal of Climatology*, titled "Selecting Markov chain orders for generating daily precipitation series across different Köppen climate regimes", assessing Markov-chain model-order performance (zeroth, first, second, and third) at reproducing observed precipitation globally using a range of metrics: Bayesian Information Criteria, the ability to reproduce observed distributions for wet- and dry-spell length and the interannual variability of precipitation occurrence. Further metrics are included in Chapter 3, including the ability to reproduce observed mean annual maximum 5-day precipitation and the interannual variability of mean precipitation.

Chapter 4 further validates the precipitation model using observed data from five stations, representative of the five Köppen climate regimes (tropical, dry, temperate, continental and polar). The conditional, multiple linear regression model for daily maximum

and minimum temperature is also introduced and validated. The ability of GCWG at reproducing univariate and multivariate extremes is also assessed.

Chapter 5 uses similar statistical tests as Chapter 4 to formulate and validate the generalised linear model for wind speed and the multiple linear regression models for solar radiation and relative humidity. The suites of observed and generated variables are also used to calculate evapotranspiration. This provides a good indicator into the proficiency of GCWG.

Chapters 6 and 7 address Objectives 2 and 3. The first section of Chapter 6 is a manuscript that is planned for submission to *Climatic Change*. Chapter 6 diagnoses global patterns of GCM-scale GCWG precipitation parameter responses to increasing GMST using two GCMs: IPSL-CM6A-LR (Boucher et al., 2020) and ACCESS-ESM1-5 (Ziehn et al., 2020). The parameters consider are the first-order Markov chain transition probabilities and the gamma parameters of the wet-day amount distributions. The scaling of these parameters enables incorporation of changes in the daily variability alongside the mean. The responses are used to perturb observed parameters at two case study locations and hence produce long precipitation time series under a range of global warming levels (GWLs): 1.5, 2.0 and 4.0°C above pre-industrial. Extreme precipitation events are also assessed at different GWLs.

Chapter 7 follows a similar trajectory to Chapter 6, though here GCWG temperature parameters have been pattern scaled using only ACCESS-ESM1-5 to derive the response patterns. Once more, spatial responses are used to perturb parameters calculated from observed station data at two case study locations to study local-scale future changes in maximum and minimum daily temperature. Univariate and multivariate (or compound) extremes at different GWLs are considered here.

Objective 4 is not addressed by any single chapter alone and, rather, is kept in mind throughout all chapters that address validation of methods used in stochastic weather generation and the construction of future projections. Stochastic weather generators can produce long running time series with computational efficiency, providing an opportunity to produce several realisations (i.e., multiple time series) of the climate for present-day and future climates. Long time series are required to study the changing risk of very rare events. Return periods, extreme indices, multivariate extremes, and wet- and dry-spell length distributions are all considered in the representation of extreme weather throughout this thesis.

To summarise, local-scale projections are of great importance to impact assessors and as inputs to hydrological, ecological, and agricultural models. While there have been great advancements in GCM resolution and downscaling techniques in recent years, further methodological developments can offer an opportunity to study a wider range of futures and further explore uncertainty in climate projections. In summary, the work presented in this thesis utilises the strengths of stochastic weather generators with the strengths of the pattern scaling technique to produce a local-scale, computationally efficient tool that can emulate a range of climate futures and which offers functional advantages over existing techniques.

2 LITERATURE REVIEW

The current changes experienced by Earth's climate are far greater than those expected to be seen through natural variability. Anthropogenic greenhouse gas and atmospheric aerosol emissions are held, with 95% certainty, responsible for the increase in mean global surface temperature (GMST) since pre-industrial times (Chen et al., 2021; Stocker et al., 2013). Alongside increases to the GMST, changes to the severity and frequency of extreme weather events have been recorded (Coumou and Rahmstorf, 2012) . How these extremes may continue to change in the future is of particular interest to impact assessors, as these extreme events can have severe adverse effects on several physical processes such as the hydrological cycle (Arnell et al., 2013), on a variety of sectors such as agriculture, tourism, forestry and fisheries (IPCC, 2022), and on human mortality and comfort (Luber and McGeehin, 2008). The effect climate change has and is projected to have on extreme weather events is discussed as part of this literature review.

The future behaviour of the climate and its response to varying greenhouse gas emissions is uncertain. The Paris Agreement, adopted in 2015 and formulated by the United Nations Framework Convention on Climate Change (UNFCCC), aims to unify efforts in limiting the increase in GMST to below 2°C compared with pre-industrial temperatures, with significant efforts to keep the increase to less than 1.5°C (Savaresi, 2016). Climate change scenarios have been developed to account for a range of possible futures, integrating changes in socioeconomic and environmental factors, climate policy and population. These scenarios range from mitigation pathways to those with high anthropogenic forcings (O'Neill et al., 2014). Of course, with several potential climate futures, there is much uncertainty regarding which pathway we will follow.

There are several different types of models that can be used to produce projections of climate change scenarios, including general circulation models (GCMs) and stochastic weather generators (SWGs). It is important that risk assessors have access to realistic and robust models to simulate present and future climates under a variety of emission scenarios. These simulations are invaluable tools for impact studies, providing information to allow users to ideally minimise the negative but exploit the positive consequences of climate change (Ribalaygua et al., 2013). For a climate change scenario to be used in impact assessments, there are several requirements, including (but not limited to):

- scenarios should have daily temporal resolution and site-specific spatial resolution (the order of a few kilometres);
- they should include a full suite of climate variables in their output;
- changes in means, variability and extremes in climate should be included;
- high-quality and long-running data.

GCMs, regional climate models (RCMs) and SWGs address different requirements. GCMs produce global, high-quality, long time series for a suite of weather variables, driven by anthropogenic changes to the atmosphere for a range of emissions and economic scenarios. While GCM projections are undoubtedly useful on a global scale, the models are computationally expensive, with spatial resolutions that are too low for use in impact assessments, alongside demonstrating significant biases in describing present-day climate (Sørland et al., 2018; Wu and Gao, 2020) arising from the parameterization of sub-grid scale processes such as convection, formation and microphysics of clouds, and local forcing (Eden and Widmann, 2014). RCMs were developed to address the course spatial resolution of GCMs and produce finer scale information that better represents sub-GCM grid-scale processes that is more suitable for impact assessments. However, it is thought that the use of RCMs leads to little improvement regarding the biases in the simulation of present-day climate, and the credibility of the models are highly dependent on the quality of the GCMs used to provide boundary conditions (Sørland et al., 2018; Wu and Gao, 2020). SWGs are an alternative, computationally inexpensive method of generating weather time series, capable of producing long-running data for a suite of weather variables based on observations (Wilson Kemsley et al., 2021). However, stochastic weather generators are not conditioned on circulation patterns, meaning they cannot be used without adjustment to study future climate. The strengths and weaknesses of GCMs, RCMs, and SWGs will be discussed in this chapter.

2.1 EXTREME WEATHER EVENTS AND CLIMATE CHANGE

An association between increasing GMST and changes in the frequency and severity of extreme weather events has been established (Seneviratne et al., 2021). This is of particular interest to impact assessors, with human society and terrestrial ecosystems facing their greatest threats with an increase in the severity and/or frequency of extreme weather events. Despite a rapid increase in world agricultural production over the past few decades, climate disasters are also thought to be the main cause of variations in production and huge contributors to food insecurity (Li et al., 2009). There is therefore a push to better understand

how extreme weather events, such as heat waves, droughts, floods, and cyclones, might change at different global warming levels (GWLs) due to their devastating impacts on society. Discerning trends in the observations of extreme events since pre-industrial times is one such way of studying the impact climate change is having on such events. Climate change projections can also be used to assess the effects further changes to the climate may have on extreme events in the future. Observed trends and projected changes in extremes from model projections participating in Phases 5 and 6 of the Coupled Model Intercomparison Project (CMIP5 and CMIP6 respectively) will be discussed in this chapter. All changes in GMST are discussed relative to the pre-industrial mean surface temperature (1850-1900).

Studying the effect changing GMST may have on extreme events is subject to certain caveats. Statistical analysis can be used to determine whether the frequency of occurrence or severity of an extreme event is significantly greater than expected in an unchanging climate (Coumou and Rahmstorf, 2012). Robust statistical analysis of extreme weather events requires:

- a comparable, defined extreme;
- careful selection of time series to reduce selection bias;
- high-quality, long-running data.

To address the first bullet point, the Expert Team on Climate Change Detection and Indices (ETCCDI) formulated 27 metrics that define comparable extremes that have been widely used to examine trends in temperature and precipitation extremes (Dunn et al., 2020). A full list of the indices can be found on http://cccma.seos.uvic.ca/ETCCDMI/list_27_indices.html. The second and third bullet points can be addressed using a range of the available climate projections and observations. However, underpinning much of the uncertainty into the study of changing extremes in future projections lies in the differences between climate models themselves (referred to as inter-model uncertainty), alongside uncertainty in future greenhouse gas and aerosol emissions. Uncertainties associated with projections will be discussed in detail in Section 2.4.

The 21st century has seen a remarkable number of record-breaking heatwaves across the world, with studies holding global warming responsible in many cases. Examples include a heatwave across Moscow with a death toll of up to 11,000 and a harvest loss of 30%, the most destructive Australian bushfires on record, and the hottest European summer

in 500 years (Coumou and Rahmstorf, 2012). It has been postulated that anthropogenic warming trends have, at least, doubled the chance of such record-breaking heatwaves occurring in Europe, with the probability of very hot summers also largely increasing (Stott et al., 2004). Alongside an increase in the intensity, frequency, and duration of global hot extremes, there has been a decrease in those of global cold extremes (Alexander et al., 2006). This has resulted in a higher proportion of warm night-time temperatures; following analyses of global weather stations, Alexander et al. (2006) found a greater than doubled frequency of the number of annual warm nights across north Africa and northern South America. In line with global changes, Skansi et al. (2013) found widespread decreases in extreme cold indices, alongside less marked increases in warm extremes following analyses of a network of observations from South American weather stations. Magnitudes of change in extreme temperatures are often much larger than changes in GMST (IPCC, 2021a), particularly over mid-latitude land regions. This is due to greater warming trends on land than on the ocean's surface, caused by the greater specific heat capacity of water than land.

CMIP5 and CMIP6 projections are in agreement that large increases in the temperature of hot and cold extremes will occur with a GMST increase of 1.5°C for a range of ETCCDI temperature indices. For example, less than half a degree of warming in GMST is responsible for changes in TXx (the monthly maximum value of daily maximum temperature) that exceed natural variability (Seneviratne et al., 2021). Warming in TNn (the monthly minimum value of daily minimum temperature) shows strong equator-to-pole amplification. Specifically, the Arctic sees an increase in warming of the coldest nights at a rate three times that of global warming. Increases in TXx are much more uniform over land, though increases in the temperature of hot extremes still show a rate of increase up to two times greater than global warming across mid-latitudes. Mueller et al. (2016) postulate that, even under moderate forcing pathways, more than half of the world's population will experience hot summers 9 out of 10 years by the year 2050, where a hot summer is defined as a summer "with higher mean temperatures than during the historically hottest summer" (Mueller et al., 2016). Tebaldi and Wehner (2018) find significant differences in increases in TXx under different forcing pathways. By the end of the 21st century, 92% of land surface areas are projected to experience present-day 20-year return values of TXx every other year under a high forcing pathway, as opposed to 62% under a mitigating scenario (Tebaldi and Wehner, 2018). This highlights the effect a forcing pathway has on resultant projections and some of the associated uncertainties.

Changes in extreme precipitation have been witnessed since pre-industrial times. Recent record-breaking examples of extreme precipitation include the Pakistani flood in 2010 – the worst flood in its recorded history, affecting up to one-fifth of its land and 20 million people, damages worth £1.3 billion following the wettest recorded autumn in Wales and England in 2000, and the highest levels of the Elbe River, Dresden, since records began in 1275 (Coumou and Rahmstorf, 2012). More recently, West Europe experienced widespread extreme precipitation in 2021 across Germany, Belgium, Luxembourg and neighbouring countries, resulting in catastrophic floods (Kreienkamp et al., 2021), with the intensity of 1- and 2-day precipitation in these regions having increased significantly as a function of climate change. A positive statistical relationship between the risk of great floods ("floods with discharges exceeding 100-year levels from basins larger than 200,000km^2" (Milly et al., 2002)) and increasing GMST has been found, alongside a substantial increase of extreme precipitation across the mid-latitudes. Where observed precipitation data is available, a larger proportion of weather stations across Africa show significant increases in extreme daily precipitation than decreases, though observations cover only a limited region of the continent. Sun et al. (2021) studied the changes in Rx1day and Rx5day indices (maximum 1-day and 5-day precipitation amounts respectively) from observed precipitation records globally. Statistically significant increases in Rx1day and Rx5day indices in the late 20th century have been noted in over two thirds of stations studied across much of the globe, including Asia, Europe and North America. Conversely, negative trends in Rx1day and Rx5day indices were not considered greater than could be expected by chance (Dunn et al., 2020). Trends in changes to extreme precipitation are less clear across South America. While evidence suggests an increase in extreme precipitation, the trends are not considered robust. This is due to data sparsity and spatially varying trends in several regions (Skansi et al., 2013).

Based on climate model projections, the risk of extreme precipitation is considered "very likely" to grow further as the climate continues to change in the future (Seneviratne et al., 2021). While the magnitude of changes to extreme precipitating events are dependent on climate scenario (Arnell and Gosling, 2016), CMIP5 models project increases in Rx1day indices regardless of forcing pathway. This is reiterated by the more recent CMIP6 projections. An increase of just 0.5°C in GMST shows significant increases in extreme precipitation; this is indicated by the 50-year return periods of Rx1day and Rx5day not overlapping between 1.5°C and 2.0°C of warming. Though most land areas can expect an increase in extreme precipitation, decreases are projected across subtropical ocean areas due

to changes in storm track shifts. In some model realisations, these the changes in storm tracks may result in decreases in extreme precipitation over nearby land areas (Seneviratne et al., 2021).

With projected increases in extreme precipitation events, it can be ascertained that the risk of flooding will also increase. However, extreme precipitation is not the only factor to consider in the study of flooding events. Modellers face several challenges in studying the changing risk of river floods, primarily relating to the inter-model uncertainty, alongside the accurate depiction of a range of other factors that contribute to floods, including storms, overflowing rivers and antecedent soil moisture. Despite this, there is confidence in an overall increase in the total global flood risk (Arnell and Gosling, 2016). Much of the changes to flood frequency (both increases and decreases) are found in south and east Asia. Increases to the frequency of great floods have also been projected to further increase substantially across Southeast Asia, Peninsular India, east Africa and the north Andes (Hirabayashi et al., 2013). He et al. (2022) quantifies GCM projected flood risk at a range of GWLs, finding decreasing return periods for 1 in 100-year floods for up to (for example) 96% of major basin areas in India by 4°C GMST increase. This would subsequently result in a 2765% increase to the human exposure of such events, contrasting increases of 239% to human exposure over the 1.5°C GWL (He et al., 2022). This is undoubtedly a calamitous increase between the GWLs, and reiterates benefits of conforming with the Paris Agreement's goal of limiting GMST increase to 1.5°C. With the water holding capacity of air increasing by approximately 7% per 1°C warming, even in locations where the frequency of precipitation occurrence may decrease, the amount of rainfall in any given event may increase. This may further exacerbate the risk of global flood risk (Trenberth, 2011).

An increase in GMST will lead to more evaporation, in turn increasing surface drying and potentially increasing the intensity and duration of droughts (Trenberth, 2011). Alongside a devastating impact on agricultural yield, prolonged drought can greatly impact ecosystem health, the economy, water supplies and the spread of vector borne diseases (Cole et al., 2002). These events are therefore considered one of the most dangerous extreme weather events for humans; for example, of all extreme events, droughts have resulted in the highest rate of mortality in the U.S. (Luber and McGeehin, 2008)). As such, it is important that impact assessors can consider how a change in drought severity may affect infrastructure and health, alongside how to reduce the environmental risk such events pose (Jones, 2001). As with floods, there are several factors that affect the occurrence of drought, including precipitation deficit, soil moisture deficit, air temperature and atmospheric

evaporative demand. Natural unforced variability, such as the El Niño-Southern Oscillation (ENSO) also significantly affects drought occurrence, with a decrease in global drought conditions under La Niña conditions (Trenberth et al., 2014). These factors present difficulties in the quantification of historical and future drought occurrence as a function of increasing GMST alone.

Observations show an increase in length and intensity of droughts across regions such as Europe and West Africa. Other regions, however, have experienced decreases in frequency and severity, such as Central America and north western Australia (Seneviratne et al., 2021). An increase in drying across low- and mid- latitudes is currently projected by several climate models under current and increasing greenhouse gas emissions (Trenberth et al., 2014). This is reiterated by the findings of Dai (2012), where it was found that over 82% of CMIP5 models project a decrease in soil-moisture content in the top-10cm across Europe, southern Africa, southeast Asia, Australia and much of the Americas and the Middle East from 5 to 15% by the end of the 21st century. Vogel et al. (2020) find increases in hot *and* dry multivariate extremes (defined as having a 1% likelihood of occurrence) over many land-surface areas as GMST increases. Concurrent extreme hot and dry events can exacerbate the conditions required for drought. Hotspots of increasing occurrence in such events are present across northern South America and southern Africa. For example, at a GWL of 1.5°C, approximately 3 clusters of hot and dry extremes are projected per year. By 4.0°C, this has increased to 9 clusters per year. It is known that drought and fire regimes interact closely, with, for example, Flannigan et al. (2000) finding that the contributing factors toward wildfire risk across the United States are projected to increase up to as much as 50% by 2060, and Kilpeläinen et al. (2010) similarly noting projected increases to the annual number of forest fire alarm days by the end of the 21st century.

There have been several studies attempting to assess the impact climate change might have on extreme storms, such as tropical cyclones, extratropical cyclones, and severe convective storms. Tropical cyclones typically cause the highest property damage and loss of all extreme events (Trenberth, 2011). There has been much speculation over whether increasing sea surface temperatures (SSTs) as a result of anthropogenic emissions, alongside other environmental changes, might increase cyclone activity (Knutson et al., 2019). Quantifying and attributing the impacts climate change may have on extreme storms presents several difficulties due to the short-lived, rare and localised behaviour of such events, alongside obtaining high-quality, accurate, and temporally consistent observations (Knutson et al., 2019; Knutson et al., 2010). However, it has been speculated that over the

last 40 years, an increasing intensity of cyclones has been witnessed, with almost a doubling in the number of the most severe tropical cyclones – though this trend is believed to be within natural variability (Mendelsohn et al., 2012). Although the frequency of tropical cyclones may lie within the expected occurrences due to natural variability, there is a suspected increase in the intensity of such storms, with the damage caused increasing. Maximum wind speeds associated with tropical cyclones are related to increased greenhouse gas warming and are suspected to increase with GMST (Elsner, 2020). However, with growing population and infrastructure in coastal, cyclone-prone areas, it is unlikely that climate change is a singular factor in this increased damage. Several factors currently hinder the study of cyclones and future climate change. These include limited historical records, under-reporting of low-damage storms and subjective recordings of damage and intensity.

2.2 CLIMATE PROJECTIONS

2.2.1 GENERAL CIRCULATION MODELS

Attempts at understanding the climate go back as far as ancient Greece. Whilst computational models were not possible until the early 1900s, conceptual and radiative-convective models have been developed and theorized for centuries. Many theories and ideas have since been confirmed using computational models and observations, such as Halley's theory of circulation and the preservation of atmospheric equilibrium, providing the foundations for early computational climate exploration (Edwards, 2011) and paving the way for the sophisticated GCMs in use today.

GCMs are presently the most powerful and sophisticated tools used for simulating future climate response to anthropogenic forcing. GCMs are primarily used to study the large-scale effects of climate change and are successful at simulating features of large-scale atmospheric circulations. GCMs have been used to study the Earth's pre-industrial climate, confirming that anthropogenic greenhouse gas emissions and aerosols are warming the atmosphere (Roeckner et al., 1999). GCMs are global, numerical models used to simulate the physical processes of the Earth's atmosphere (and oceans, unless they are an atmosphere only GCM), and can simulate atmospheric circulatory patterns with the earth and atmosphere divided into a three-dimensional grid. These complex models use partial differential equations to represent many conservation laws (e.g., atmospheric mass, momentum, energy, and water vapour), representations of sea ice, atmospheric radiation and convection, alongside several other important processes (Dubrovský, 1997). CMIP6 (the

most up to date Coupled Model Intercomparison Project) GCMs depict the Earth and its atmosphere using between 14 and 102 atmospheric layers above sea-level, and with horizontal resolutions generally between 100 and 250km, though in some higher-scale models, resolutions can reach up to 60km (IPCC, 2021b) . Although GCM resolution has come a long way in recent years, even the highest resolution GCMs produce climate variables on a scale that is inappropriate for impact assessments, and are therefore often described as representing area-averaged values for each grid-cell (Wilks, 1999a), as opposed to local-scale climate. Furthermore, GCMs are known to imperfectly simulate the occurrence of extreme events due to their course resolution and the parameterisation of the physical drivers of such events (Bellprat and Doblas-Reyes, 2016).

GCMs have simulated Earth's climate under past, present, and future conditions, with the capacity to study both transient and steady-state emission scenarios. Using these projections, GCMs can be used to study future greenhouse gas forcing scenarios with variable emissions. Currently, GCM-based studies often focus on long-term projections as the signal-to-noise ratio is robust, although closer time scales (i.e., decadal) are of more interest to impact assessors (Cabré et al., 2010). However, running these coupled climate models for many different scenarios and time scales is currently not computationally viable or practical, as simulations tend to push computational capacities to their limits. Although the resolution of GCMs has greatly improved in recent years (IPCC, 2021b), impact and hydrological assessments require data at a much finer scale. There has been (and remains) a mismatch in scale between GCM outputs and the spatial scale required for hydrological assessments (Doblas-Reyes et al., 2021; Wood et al., 2004).

Despite the accuracy of GCMs at simulating changes in the climate system over large spatial scales, it is smaller-scale and regional details, such as, for example, steep topography, physical properties related to cloud formation and microphysics, and local-scale forcings, that are poorly portrayed due to the models' limiting low temporal and horizontal resolutions (Ribalaygua et al., 2013). These processes often require parameterization and thus are a major cause of uncertainty in GCM simulations (Haarsma et al., 2016). Furthermore, it is local-scale changes that are of particular interest to impact modellers as it is these changes in local climate that are the determining factors in consequences to agricultural, ecological and hydrological systems (Wilks, 1999a) as opposed to the global scale changes. This coarse resolution is therefore a major weakness of GCMs and as such means they are unable to properly simulate certain physical processes, such as fine scale features of the hydrological

cycle, which require a much higher resolution, and cannot be used directly in impact assessments.

To better represent some of these finer-scale processes, a selection of modelling centres partook in the High Resolution Model Intercomparison Project (HighResMIP) resulting in projections with resolutions of up to 25km (Haarsma et al., 2016) as part of the recent Coupled Model Intercomparison Project 6 (CMIP6). HighResMIP simulations provide better descriptions of cloud formation, circulation characteristics, changes to water availability, and extreme events such as tropical cyclones (Roberts et al., 2020). However, higher resolutions of course result in a greater computational cost. A multitude of ensemble members are required to capture the unforced variability that occurs on top of the forced climate signal. Due to the increased computational power required, several ensembles have not yet been simulated by these higher resolution models. Alongside the requirement of several ensemble members to diagnose unforced from forced climate signals, several simulations that are representative of a wide range of scenarios is ever present. The added computational complexity of HighResMIP models has resulted in a smaller range of simulated scenarios. This unfortunately does not address the assertion that a wide range of possibilities ought to be considered to capture uncertainties in modelling the future climate, particularly for distant futures (Jones et al., 2016).

There has been a significant effort to develop methods of downscaling GCM information to provide site-specific, higher resolution data that is more suitable for use in impact assessments (Maraun et al., 2010; Wilby and Wigley, 1997; Wilks and Wilby, 1999). Several methods have been developed to downscale GCM projections to local scales that are more appropriate for impact assessments. These can broadly be categorised into two techniques:

- dynamical downscaling,
- statistical downscaling.

Dynamical downscaling is usually in the form of the previously introduced RCMs. RCMs are numerical models with higher resolutions than GCMs, nested into a single GCM with boundary conditions defined by the driving GCM. Statistical downscaling techniques generally can be broken down into several categories; “perfect prognosis” (PP), “weather-typing”, “stochastic weather generators” and “model output statistics” (MOS) (Maraun et al., 2010; Wilby et al., 2002b) PP techniques develop statistical relationships between area-

averaged and local-scale variables from observational data which can hence be applied to the GCM large-scale data. MOS is similar to PP, though relationships are instead determined between local-scale observations and *simulated* area-averaged data. Weather-typing methods are also a regression-based approach, though relationships are instead established between area-averaged variables that represent large scale circulation, and local-scale surface level variables.

2.2.2 DYNAMICAL DOWNSCALING

There are many techniques that can be categorised as dynamical downscaling methods. One of the most prevalent and widely discussed dynamical downscaling methods is the "nesting technique", resulting in the aforementioned RCMs. Regional climate models (sometimes also referred to as limited-area models, LAMs) are used to simulate weather variables over sub-global domains (e.g., Europe) with much higher resolutions than GCMs, resulting in better representation of several of the aforementioned sub-GCM grid-cell processes. GCM projections are used to supply an RCM with time dependent and initial boundary conditions (Caya and Laprise, 1999; Maraun et al., 2010). RCMs hence produce projections that are physically consistent with external forcings, though with much higher resolutions than the driving GCM (Wilby et al., 2002b). The resultant resolution of an RCM is usually 5 - 50km; much higher than the usual 100 – 250km of present-day GCMs. Increasing GCM time step from (typically) 30 minutes to 5 minutes (or less) usually accompanies the increase in spatial resolution to maintain numerical stability (Maraun et al., 2010).

As expected with a higher resolution, RCMs outperform GCMs in representing hydrological cycles, orographic effects, land-sea contrasts, and land-surface characteristics. The benefits of using an RCM are particularly noteworthy in the representation of precipitation in mountainous regions and locations with steep topographic changes (Frei et al., 2003). The greatest benefits contributed by RCMs in comparison to the driving GCMs is regarding the representation of precipitation (Frei et al., 2006). Mesoscale and fine-scale physical processes, such as precipitating weather systems and fronts are also represented better by RCMs than GCMs due to the increased resolution (Sørland et al., 2018) alongside better depictions of river flow (Pastén-Zapata et al., 2020). RCMs generally produce larger amounts of precipitation in comparison to their driving GCM, resulting in better representations of daily precipitation distributions.

It is evident that RCMs add significant value compared to GCMs at modelling local-scale climate (Rummukainen, 2016), though RCMs are still relatively computationally intensive while only yielding results that are typically not significantly better than simpler, statistical approaches (Wilks, 2010). Although the resolution of RCMs is much greater than GCMs, resolutions may still remain too course for practitioners such as flood planners and agricultural modellers that require local- or point-scale data (Vrac and Naveau, 2007; Wilby et al., 2002b). RCMs also include biases in modelling present-day climate, much like GCMs. The skill of downscaling precipitation is also somewhat variable by region, season, and the intensity and duration of precipitation events. It is also important to note that a 30-year (for example) RCM simulation represents only one possible realisation of a 30-year period. Natural variability is therefore a key source of uncertainty in the estimation of extreme weather events from RCM projections (Maraun et al., 2010). This method of dynamical downscaling is therefore constrained by the availability of RCM simulations.

2.2.3 STATISTICAL DOWNSCALING

Statistical downscaling is a popular alternative for producing time series with higher resolutions than GCMs due to the relative ease of use and good performance that is comparable with the dynamical approach (Eden and Widmann, 2014). Statistical downscaling approaches typically establish statistical links between weather over large spatial scales (GCM projections) and local-scale data (weather station observations) (Maraun et al., 2010). Observed daily means, medians and variances are generally reproduced well using statistical regression approach. However, low-frequency variations such as seasonal totals are reproduced with less skill and, much like dynamical downscaling, variances of variables – particularly precipitation – are underrepresented (Wilby et al., 2002a; Wilks and Wilby, 1999). Statistical downscaling methods are also considered poor regarding accurate simulation of extreme precipitation events (Hashmi et al., 2011). Unlike the dynamical approach, statistical downscaling requires relatively little GCM driving data and does not stretch computational resources. Furthermore, in the study of present climates, statistical techniques tend to outperform dynamical ones due to avoiding any GCM-induced errors (Ribalaygua et al., 2013). There are several statistical techniques that can be used to establish relationships between station scale data and GCM (area-averaged) scale data; here the prominently used regression based (i.e., PP, MOS, and weather-typing) and stochastic weather generator techniques will be discussed.

There are several traditional regression-based downscaling methods, such as linear regression, canonical correction analysis, and principal component analysis that can all be used to downscale GCM-scale data. PP approaches determine relationships between observed, large-scale predictors and observed, local-scale predictands. However, observations for weather variables at GCM-scales are relatively rare, and as such reanalysis data tends to be used instead. The relationships between the observed (or reanalysis) large-scale and local-scale variables can hence be applied to the future GCM-scale projections to derive simulated, local-scale time series. Modern PP methods tend to include a variable representing noise in the regression to better represent variability and extremes (Maraun et al., 2010).

Model output statistics is an alternative approach that is thought to better represent low-frequency variations than traditional PP methods. In contrast to the statistical regression approach, MOS determines a relationship between *simulated* area-averaged predictors (from a GCM) and observed, local-scale predictands (Eden and Widmann, 2014; Maraun et al., 2010). The relationship can hence be applied to a simulated predictor for a future GCM projection. MOS methods widely outperform regression downscaling methods but show differences in skill that are dependent on location and season. Similar caveats exist with MOS to statistical regression downscaling. To apply both regression techniques (MOS and statistical regression methods) to a climate change context, the predictors must capture the climate change signal and the statistical relationship determined from the fitting of the regression is assumed to remain constant under a changed climate, though it is thought that MOS has advantages over the statistical regression in both regards. Simulated precipitation (used as a predictor in MOS) is physically consistent with other atmospheric variables, making it less likely that any climate change signal is missed. An additional benefit of MOS is the use of relationships determined between simulated and observed variables over the same spatial scale to provide a correction for the model data.

An alternative approach that links local-scale variables to large-scale variables that represent circulation is known as weather-typing. Local-scale meteorological variables can be categorised into different weather states that are based on large-scale upper-air variables (Vrac et al., 2007; Wilby et al., 2002b). Simply, a statistical relationship between these weather states and a local, surface variable must be determined. Future scenarios are constructed in one of two ways. Observed variable distributions can be resampled, conditioned on circulations patterns produced by a GCM, or alternatively, Monte Carlo techniques can be used to simulate a synthetic sequence of weather patterns and then

resampling from observed data. This approach is attractive as it maintains sensible linkages between large scale climate drivers and local weather, upholding validity for a range of environmental variables and has applications for multi-sites (Vrac and Naveau, 2007; Wilby et al., 2002b). However, weather-typing downscaling techniques are typically inadequate in comparison to alternative methods in representing the occurrence of rare or extreme events. A further, major downfall of this approach is in its representation of precipitation changes. Changes in precipitation produced by changes to the frequency of the defined weather patterns are rarely consistent with GCM predicted changes.

An alternative approach is the use of stochastic weather generators. SWGs can be used to downscale GCM data using the "change-factor" approach. A relationship between present (i.e., the control period) and a GCM time series representative of a future climate scenario can be determined and hence applied to observed, weather station scale data (Wilks, 2010). Multiplicative change factors are widely used to identify changes in precipitation statistics (e.g., first-order Markov probabilities);

$$\frac{P^{Fut}}{P^{Obs}} = \frac{P^{GCMFut}}{P^{GCMCon}} \tag{2.1}$$

where P is a general variable (Kilsby et al., 2007). Obs, Con, Fut, and GCM suffixes refer to the observed, control period, future and GCM-projected data respectively. The factor of change, α, (the righthand side of the above equation) is calculated on a calendar month basis to account for intra-annual seasonality in the climate change signal (Fatichi et al., 2011) and gives

$$P^{Fut} = \alpha P^{Obs} \tag{2.2}$$

Additive change factors,

$$T^{Fut} - T^{Obs} = T^{GCMFut} - T^{GCMCon} \tag{2.3}$$

can be determined for all other weather variables (T), this time giving

$$T^{Fut} = T^{Obs} + \alpha \tag{2.4}$$

Upon determining the changes to the weather variable statistics, these can be input into the stochastic weather generator to produce future, station scale data. This method is not restricted to GCMs – RCMs be used similarly, potentially providing better matches to local climate characteristics (Fatichi et al., 2011).

A *combination* of regression and stochastic weather generator techniques can also be used in downscaling GCM projections. Semenov and Barrow (1997) present a multiple linear regression approach for use with the stochastic weather generator, LARS-WG. Relationships between anomalies in local scale predictands (e.g., mean temperature and precipitation) and regional scale predictors (GCM grid-box scale), including mean sea level pressure, temperature, and precipitation, can be calculated (Barrow et al., 1996). Relationships are determined over a control period (e.g., 1960 – 1990), usually for each month. The relationship can be applied to GCM derived changes, providing local scale, future parameters. The parameters can hence be used to drive a weather generator, such as LARS-WG. However, there are caveats with this method. Regression models generally perform better for temperature than precipitation. For both variables, regression models do not appear to perform well in explaining the variance of the observed data. The performance of this technique is very dependent on how accurately the regression models capture the variability of the observed data.

An alternative example of such an approach is outlined by Wilks (2010), applied to the input parameters of a Markov-chain SWG (see more in Section 2.3). The conditional probabilities used in a SWG are, for example, a wet day following a wet day (p_{11}), a wet day following a dry day (p_{01}) and the average wet-day precipitation amount (μ). In this example, the parameters can be predicted using

$$p_{11} = a + bZ + cZ^2 \tag{2.5}$$

$$p_{01} = d + eZ + fZ^2 \tag{2.6}$$

$$\mu = g + hZ + iZ^2 \tag{2.7}$$

where Z is vorticity at a nearby GCM grid-cell (Wilks, 2010) and a, b, c, d, e, f, g, h and i are regression coefficients. This example has been given due to GCMs typically overestimating the probabilities of precipitation on an area-averaged scale, and underestimating precipitation amounts (Wilks, 1999a), though there are several other regression techniques that have been proposed for different variables (Rummukainen, 1997).

Equations 2.5 – 2.7 are fit through combining the observed daily precipitation record with gridded reanalysis data (consistent with a chosen GCM). Changes in daily precipitation can hence be modelled through calculation from GCM simulated vorticity time series. However, variations in vorticity simulate insufficient variability in daily precipitation amounts – several extensions and adjustments to this method have since been made to

incorporate further predictors, such as sea level pressure and surface specific humidity. Rummukainen (1997) notes that, in order to reduce the risk of omitting predictors that contain useful information, many predictors should first be considered in the regression. However, inclusion of more predictors increases the risk of overfitting regressions, and there is generally little consensus about which predictors are most appropriate to capture climate change signals (Eden and Widmann, 2014).

A final example of the application of regression-based downscaling techniques to a stochastic weather generator is the Statistical DownScaling Model (SDSM) developed by Wilby et al. (2000) for use in assessing local climate change impacts. Described as a hybrid of the regression-based and stochastic weather generator approaches to downscaling GCM projections, large-scale circulation patterns and atmospheric variables (e.g., circulation, humidity, vorticity), similar to the MOS technique, are used to condition the parameters of a local-scale stochastic weather generator (Hassan et al., 2014; Wilby et al., 2002b). This is similar to equations 2.5 – 2.7, where vorticity was used as a predictor. Prior to downscaling, a range of predictors are assessed from reanalysis datasets, based on correlations with the predictand to ensure appropriate variables are considered. This reduces the chance of overfitting the model, the risk of not including a predictor that contains valuable information, and importantly, is flexible based on the desired variable.

2.3 STOCHASTIC WEATHER GENERATORS

Whilst stochastic weather generators can be used for GCM downscaling, they are a weather simulation technique in their own right. SWGs aim to reproduce the random nature of weather and are site-specific models that are not conditioned by circulation patterns. Unlike GCMs, stochastic weather generators are non-predictive and produce weather time series that are statistically similar to observed data for a given location. Meteorological events and physical processes are not considered in stochastic models. This method of generating sequences of weather based on statistical properties of observed data was initially developed to generate synthetic series of precipitation for hydrological modelling purposes (Semenov and Barrow, 2002) and has been used for decades, with the earliest examples developed in the 1960s (Gabriel and Neumann, 1962). By the 1980s, stochastic weather generators were being used to produce a range of climate variables, including maximum and minimum temperatures and solar radiation (Semenov et al., 1998). They are now used to produce long, temporally consistent records that are useful in many applications, including water resource engineering, flood analysis, agricultural assessments and ecosystem models.

Conversely to GCMs, SWGs produce time series across small spatial scales (often spanning just a few kilometres, or for individual points) and are computationally economical. SWGs also have the capacity to model climate variables at a high temporal resolutions, typically simulating a suite of variables on a daily to sub-daily scale (Ailliot et al., 2015), and due to their computationally inexpensive design, can produce data that spans multiple decades, meaning their output is more appropriate for climate impact studies than GCMs (Wilby and Wigley, 1997) whilst providing the opportunity for the robust study of rare and extreme events. Historically, SWGs do not consider spatial coherence. For example, two weather stations that are geographically close may synthesise weather that is contradictory in nature. However, there has been research into ensuring spatial coherence between generated time series that show promising results (such tools are typically referred to as multisite SWGs) (Peleg et al., 2017; Steinschneider and Brown, 2013; Dawkins et al., 2022; Serinaldi and Kilsby, 2012). These methods span varying spatial and temporal scales, and utilise a range of stochastic techniques, including Markov chains and Generalized additive models for precipitation distribution location, scale and shape. Multisite SWGs will however not be considered here, as this study's focus is on the subsequent application of pattern scaling to a SWG.

Stochastic weather generators generally produce a time series of precipitation first, followed by the simulation of any other weather variables, including (but not limited to) daily temperature, mean wind speed, relative humidity, and solar radiation. It is common for stochastic weather generators to produce daily time series, though methods for generating sub-daily time series have also been discussed (Peleg et al., 2017). Due to the conditioning of other weather variables on daily precipitation, precipitation is generally referred to as a "primary" variable. Temperatures (mean, diurnal range, maximum and/or minimum) are widely referred to as "secondary" variables due to their generation following and conditioned on precipitation, but often prior to the simulation of any other variables (referred to collectively as "tertiary" variables).

Markov-chain stochastic weather generators are widely used to generate daily precipitation time series. Alternative approaches, such as series-type weather generators and the Neyman-Scott Rectangular Pulse (NSRP) can also be used to generate sequences of precipitation. These models require the determining of a suite of site-specific parameters based on observations (or, where lacking observations, reanalysis data) and aim to produce synthetic time series that are statistically similar to the input data (Dubrovský, 1997). There are several methods that can be used to validate the effectiveness of the model, such as

information criterion and statistical testing. Markov-chain, series-type and NSRP SWGs will be discussed in turn here.

The earliest recorded stochastic weather generators are chain-dependent in nature, with precipitation modelled as a Markov process (Gabriel and Neumann, 1962). Initially designed to assist hydrological impact assessments, they were implemented as a simple method of modelling precipitation occurrence and amount. Markov chain weather generators are typically used to produce the driving occurrence of precipitation for a suite of weather variables, including solar radiation, minimum and maximum daily temperatures. In these SWGs, daily precipitation status (i.e., wet or dry) and amount (if a wet day is simulated) are first synthesised, followed by any other weather variables. Although changes in several other weather variables are projected as a function of changing climate, methods used to simulate precipitation time series will be a focus of this literature review. This is due to wide literature documenting different methods of generating daily precipitation in comparison to the secondary and tertiary variables; the relative complexity in producing accurate depictions of precipitation under present-day and future climates and due to the ultimate scaling of precipitation parameters in Chapter 6.

The most widely used Markov-chain SWGs use 2-states (i.e., wet or dry) to generate precipitation occurrence through conditioning the probability of a wet day on the status of the previous day(s). Although 2-state chains are most frequently used, there has been research into using a greater number of states to represent different precipitation amounts (Gregory et al., 1993). First order models are most commonly used, where order refers to the number of previous days considered in the calculation of transition probabilities (first orders consider precipitation occurrence on the day previous), although other orders are discussed in literature (Jimoh and Webster, 1996; Schoof and Pryor, 2008; Wilson Kemsley et al., 2021). For a first-order model, two independent transition probabilities are calculated from observed precipitation data:

$$P_{00} = \frac{n_{00}}{n_{00} + n_{01}} \tag{2.8}$$

$$P_{11} = \frac{n_{11}}{n_{11} + n_{10}} \tag{2.9}$$

where P_{00} is the probability of a dry day following a dry day and P_{11} is the probability of a wet day following a wet day. Probabilities are typically calculated monthly. Indices 0 and 1 refer to dry (0) or wet (1) days. The n values are sums from the observed data; n_{01} is the

total number of wet days preceded by dry days, n_0 the number of wet days, and so forth for the remaining n values. It is important to note that $P_{00} + P_{01} = 1$ and $P_{10} + P_{11} = 1$ (Schoof and Pryor, 2008) (where, for example P_{01} is the probability of a wet day following a dry day). On a wet day, an amount of precipitation is usually taken independently from a two-parameter gamma distribution, typically fitted monthly. Markov-chain SWGs that use a gamma distribution to simulate *amount* of precipitation on wet days are referred to as Markov-chain gamma-distribution models.

A key criticism of the first order approach is that it has a limited 'memory' of rare events, such as exceptionally long sequences of wet or dry weather, thus underestimating their occurrence (Semenov and Barrow, 1997). As a result, the interannual variance of precipitation can be poorly represented (i.e., suffering from "overdispersion"). Wilby et al. (2002) suggested additions to the calculation of first-order transition probabilities to better represent low-frequency statistics (such as seasonal totals and variances) through the inclusion of parameters that represent climate forcings such as the North Atlantic Oscillation index or Sea Surface Temperature anomaly. The inclusion of additional parameters to account for forcings bestowed no advantage over a model with no forcings in terms of daily diagnostics, though monthly statistics were much better represented by the forced parameters in several locations. It is important to note that overdispersion was not entirely eliminated, and remains an issue with Markov-chain SWGs (Wilby et al., 2002; Wilks and Wilby, 1999).

An alternative approach at improving the representation of interannual variance of precipitation occurrence may be with the use of higher order Markov chains (Jones and Thornton, 1993; Lennartsson et al., 2008). The aforementioned Markov-chain model is simple to generalise to higher orders. The precipitation status in a second order model is conditioned on the precipitation status of the previous two days, requiring the calculation of eight transition probabilities. The order can be increased further, with the number of transition probabilities required increasing exponentially, with the total number generalised as 2^{m+1}, where m is the order of model. In a global comparison of model-order performance, Wilson Kemsley et al. (2021) noted differences in performance dependent on assessment metric. For example, third-order models commonly reproduced the observed distribution of dry-spell lengths most accurately, while first-orders better captured the observed distributions of wet-spell length. Differences in climate regime were also noted. However, it is documented that, should a low order model reasonably fit observed data, better estimates

of future weather are obtained while minimising the risk of overfitting. These findings reflect the requirement for a flexible approach regarding model-order choice (Coe and Stern, 1982).

Following the generation of a wet day, a Markov-chain gamma-distribution model attributes a random amount of precipitation from a gamma distribution that has been fitted monthly to the observed data. The two-parameter gamma distribution has several strengths and is considered a good fit for most global precipitation distributions. The distribution however, is not without weaknesses, and though it is the most commonly used, several studies have documented the use of other distributions instead (Lennartsson et al., 2008), such as the exponential or mixed exponential, skewed normal, or empirical distributions (Qian et al., 2005; Soltani and Hoogenboom, 2003a). The gamma distribution typically matches observed data with greater accuracy than exponential or skewed normal distributions (Semenov et al., 1998). However, a failing with the 2-parameter gamma distributions is that they have been shown to underestimate the occurrence of small precipitation amounts and are thought to overestimate the probability of large precipitation amounts (Semenov et al., 1998). However, the two -parameter gamma distribution has also been shown to misrepresent distributions with heavy tails (Furrer and Katz, 2008), showing performance that is variable based on site. Using the mixed exponential distribution *may* reduce overdispersion in a SWG, though the fitting procedure is considerably more complicated than gamma distributions. Empirical distributions have greater flexibility and can therefore capture the shape of precipitation intensities with better accuracy than the aforementioned distributions. However, they require more parameters than the simple shape and scale parameters necessary to fit the gamma distribution, and therefore increase the complexity of scaling parameters with GMST (see Chapter 8 for future work relating to this point).

To summarise, Markov-chain gamma-distribution weather generators have several strengths. Firstly, the widely used first-order models only require the calculation of four independent parameters per month (i.e., P_{00}, P_{11} , wet-day shape and scale parameters). This is a particularly attractive strength of this approach towards generating daily precipitation time series when considering Objective 2, discussed in more detail in Chapter 1. Secondly, Markov-chain models are flexible regarding the order (and number of states) used to generate precipitation occurrence. Should reproducing dry-spell length accurately, for example in drought assessments, be the focus of a study, higher orders (e.g., third order) may be used instead of first order (Wilson Kemsley et al., 2021). There is also flexibility regarding the choice of distribution used to model precipitation amount. While there are

advantages to alternative distributions, the two-parameter distribution will be used here to model intensity of rainfall on wet days due to its flexibility, simplicity, overall good performance globally, and wide use in approximating daily precipitation distributions in literature (e.g., Wilks, D.S., 1999a; Wilson Kemsley et al., 2021; Wilby and Wigley, 2002) . It is for these reasons that Markov-chain gamma-distribution weather generators are the primary focus of this section and will be used in the development of a globally calibrated weather generator in this thesis.

Series-type stochastic weather generators function slightly differently, and were developed to better represent long dry and wet spells than the Markov-chain approach (Racsko et al., 1991). In a series-type weather generator, precipitation status is determined using empirical or semi-empirical probability distributions for lengths of dry and wet spells that have been fitted to observed data. A wet spell is defined as a sequence of wet days followed and preceded by dry days (conversely for dry spells). These parameters are particularly important for assessing the impacts of climate change on crop growth and water resources. Series-type weather generators will be discussed in a comparison between series-type and Markov-chain stochastic weather generators (see Table 2.1)

The NSRP model is another alternative method to generate sequences of precipitation. As with Markov and series-type weather generators, parameters are usually calculated monthly to account for seasonal variations. The NSRP model is flexible in its structure, and parameters relate to the physical processes underlying precipitation occurrence, such as convective rain cells (Cowpertwait et al., 1996). The occurrence of precipitation events (referred to as “storm origins”) are modelled as a Poisson process, with wet periods defined by the origins of the storm event. Parameters are also estimated for the waiting time for the origin of a “rain cell” (a rectangular pulse) following a storm origin, the lifetime of a rain cell, the number of rain cells attributed to a given storm origin, and the intensity of each rain cell (Cowpertwait, 2004, 1994). The duration of a rain cell is taken independently and randomly from an exponential distribution, with Poisson rates and durations typically modelled hourly. Rain cells may overlap, resulting in precipitation that is summed over concurrent cells. Extensions have been made to the NSRP model to account for “heavy” and “light” rainfall events and the simulation of coherent precipitation events over multiple sites (Cowpertwait, 2004). While NSRP models show great accuracy in reproducing observed precipitation, they will not be considered further here. This is due to a relatively complex fitting procedure with many parameters in comparison to the alternative

Markov and series-type methods, alongside the requirement of hourly precipitation observations, which may be sparser in global coverage than daily records.

SWG input parameters are calculated from observed data, and therefore the resultant synthetic time series is representative of the observed climate that has been used to calibrate the generator (Qian et al., 2005). The quality of the observed data that is used as to calculate parameters is therefore one determinant of the accuracy of the simulated data (Soltani and Hoogenboom, 2003b), and therefore it is recommended that, to determine robust statistical parameters, at least 20 years of site-specific historical daily data is required (Semenov and Barrow, 1997). This raises an issue that GCMs do not face – in many locations across the globe, 20 years of daily observations are not readily available. This is essential to realistically model the temporal variability. Aside from potentially lacking data for various locations across the globe, there are several possible systematic errors in measuring daily climate variables, including typographical and transmission errors, and wind-caused undercatch of precipitation by gauges (Hulme, 1992).

Four prominent (Soltani & Hoogenboom, 2003) stochastic weather generators (WGEN (Richardson and Wright, 1984), SIMMETEO (Soltani and Hoogenboom, 2003a), AAFC-WG and LARS-WG) use a range of methods to produce weather time series. The methodologies used by these weather generators will be outlined, with key characteristics compared in Table 2.1. Note that only Markov and series-type weather generators are included in Table 2.1. This is due to a greater wealth of literature critiquing and comparing these types of SWGs, alongside a simpler fitting procedure than for example, NSRP models. Precipitation status and amount are arguably the most important variables to reproduce correctly. This is because several SWGs condition other climate variables on the precipitation, and this is a particularly important variable for use in hydrological, ecological and agricultural impact assessments. In many of the weather generators discussed in Table 2.1, solar radiation and maximum and minimum daily temperatures are considered stochastic processes with daily means and standard deviations that are conditioned on precipitation status.

Table 2.1 – Four prominent stochastic weather generators, how they generate precipitation and the conditioning of maximum and minimum temperature and solar radiation on precipitation.

	PRECIPITATION		
	STATUS	**DISTRIBUTION**	**MIN & MAX TEMP, SOLAR RADIATION**
WGEN	*Defined as >0mm.* 1st order 2-state Markov chain applied to previous day's dry/wet status. Different transition probabilities determined for each month.	Precipitation amount follows a 2-parameter gamma distribution. Gamma parameters calculated for each month.	*Not conditioned on precipitation status*
SIMMETEO	1st order 2-state Markov chain used to calculate transitional probabilities for each month from monthly data.	Precipitation amount follows a 2-parameter gamma distribution. Gamma parameters calculated for each month.	*Conditioned on precipitation status*
AAFC-WG	2nd order 2-state Markov chain applied to previous day's dry/wet status. Different transition probabilities determined for each month.	Precipitation amount follows empirical distribution estimated from log-transformed precipitation amounts on wet days bimonthly.	*Conditioned on precipitation status*
LARS-WG	*Defined as >0mm.* Parameters are calculated for each month from lengths of dry and wet series (semi-empirical distribution).	Precipitation amount follows a semi-empirical distribution (with 10 intervals) with parameters calculated for each month. Amount of precipitation depends on amount from previous day's status.	*Conditioned on precipitation status.*

WGEN is a Markov-chain stochastic weather generator that uses a first order 2-state structure to determine precipitation status, and a two-parameter gamma distribution to determine precipitation amount with parameters and probabilities (equations 2.8 and 2.9)

calculated monthly. On wet days, precipitation amounts are randomly taken from the two-parameter gamma distribution independently of the sequence of precipitation (Parlange and Katz, 2000). Daily precipitation amounts are considered independent and identically distributed variables each month. Following the simulation of status and (if relevant) amount of precipitation, daily minimum and maximum temperatures are simulated. In WGEN, minimum temperature is not conditioned on precipitation status and is obtained by fitting cosine functions to both the means and the coefficient of variations of observed data for 28-day periods throughout the year. Maximum temperature follows a similar process, although unlike minimum temperature, is conditioned on precipitation status, thus fitting different cosine functions for wet and dry days (Semenov et al., 1998).

In a study across several different ecoregion divisions, WGEN was found to perform well at reproducing some variables and poorly at others (Semenov et al., 1998). WGEN accurately reproduced the mean numbers of wet days, although the accuracy of this was variable by location. Semenov et al. (1998) found that inter-annual variance in monthly precipitation simulated by WGEN tended to slightly underestimate the observed. This is thought to be due to the handling of precipitation events as independent from the monthly gamma distribution. It is also thought that rare events do not affect transition probabilities enough to capture the frequency of their occurrence, and as such the shapes of distributions for wet- and dry spell-length are not reproduced accurately enough. Despite this, WGEN was successful at reproducing the observed number of heavy precipitation events, although it is important to note that only a small number of weather stations have been considered.

Regarding temperatures, WGEN showed good accuracy with respect to the mean and variability of maxima and minima, although this performance (particularly in reproducing maxima) is highly dependent on location. Despite WGEN's relatively good performance at reproducing observed temperatures, solar radiation appears to often be overestimated and monthly means frequently differ significantly to the observed. Smaller standard deviations and variances in solar radiation were also found. Many of WGEN's weaknesses at modelling global climates are thought to stem from its original calibration for climates in the USA (Qian et al., 2008).

SIMMETEO is also a Markov-type weather generator that follows a slightly different approach to WGEN, producing daily weather variable time series from monthly statistics for precipitation, daily maximum and minimum temperatures, and solar radiation. Input parameters for SIMMETEO include the mean number of wet days, amount of

precipitation, solar radiation, maximum and minimum temperatures calculated monthly from observed data, as opposed to daily values (Bannayan and Crout, 1999). This is an advantage SIMMETEO has over WGEN – monthly records tend to be more readily available than daily data (Geng et al., 1986). In a given month, transition probabilities are calculated using the total precipitation and number of wet days. On wet days, precipitation amount is modelled once again by a 2-parameter gamma distribution.

To generate temperature and solar radiation, monthly climatic statistics for each variable are required. Means and standard deviations are estimated using first-order Fourier series requiring three climatic coefficients: the annual average, the amplitude (half of the difference between the maximum and minimum) and the month at which the maximum is reached, for each variable. Minimum temperature is generated independently of precipitation status, whereas maximum temperature and solar radiation require a wet- or dry-day status. Values are sampled from normal distributions, with lag-1 auto- and cross-correlations maintained between temperatures and solar radiation (Soltani and Hoogenboom, 2003a).

In a study of SIMMETEO's performance across five sites in Iran, SIMMETEO was found to perform similarly to WGEN regarding the mean number of wet days, although with poorer performance in calculating variances in precipitation occurrence (Soltani and Hoogenboom, 2003a). Despite this, SIMMETEO was successful in generating the number of heavy precipitation events. No statistical difference was found between generated and observed monthly means in precipitation amount. SIMMETEO appeared to perform worse than WGEN with respect to generating daily maximum temperatures, although successfully reproduced minimum temperatures. Regarding solar radiation, SIMMETEO outperformed WGEN significantly, with most simulated means statistically similar to observed data.

Conversely to WGEN and SIMMETEO, LARS-WG is a series-type weather generator, developed at Long Ashton Research Station to overcome limitations of Markov-type models with regards to precipitation occurrence (Semenov and Barrow, 2002). Semi-empirical distributions of wet and dry spell lengths with ten intervals are fitted to an observed time series. Sequences of wet and dry spells are chosen from this distribution to determine precipitation status. On wet days, precipitation amount is determined from another semi-empirical distribution fitted to observed data. Separate semi-empirical distributions (for wet and dry spell lengths and wet day precipitation) are fitted for each month. The flexibility of the semi-empirical distribution allows the approximate fitting of a

variety of shapes by adjusting the histogram intervals. Minimum and maximum temperatures are conditioned on precipitation status and are determined through fitting Fourier series to means and standard deviations of observed data monthly, with separate Fourier series being fit for wet and dry days. Both minimum and maximum temperatures are fit to normal distributions. Solar radiation is also conditioned on precipitation status, with wet and dry semi-empirical distributions calculated for each month.

Multiple studies have studied LARS-WG's ability to reproduce observed data across a range of different ecoregions (Qian et al., 2008, 2005; Semenov, 2008; Semenov et al., 1998). LARS-WG performs relatively well at reproducing observed precipitation across all locations. Much like WGEN, LARS-WG performs well with respect to reproducing the number of heavy precipitation events, although in comparison WGEN, LARS-WG reproduces the full observed precipitation distribution more accurately. Regarding the length of wet and dry spells, LARS-WG outperforms WGEN. This is thought to be because rare events, such as unusually long wet or dry series, are not detected using the Markov-chain approach, whereas empirical distributions represent all wet and dry spell lengths present in the observed record. Monthly mean precipitation is also reproduced well, although generated inter-annual variance is lower than in observed data. Alongside accurately reproducing precipitation, the mean monthly maximum and minimum temperatures are reproduced well. Small errors in the reproduced distribution of temperatures arise through fitting a daily curve to average monthly values, arising from fitting the Fourier series. It is also important to note that in several locations globally, maximum and minimum daily temperatures do not follow normal distributions (Qian et al., 2008). This results in variable performance regarding the reproduction of variances in temperature (Semenov et al., 1998).

AAFC-WG (developed at Agriculture and Agri-Food Canada) was developed for impact assessors in Canada's agricultural industry (Qian et al., 2008). Much like WGEN and SIMMETEO, AAFC-WG follows a Markov-type model, although using a second order 2-state structure instead of first order to determine precipitation status. However, unlike SIMMETEO and WGEN's two-parameter gamma distribution approach, AAFC-WG uses empirical distributions to determine the *amount* of precipitation on a rainy day. The empirical distributions are determined bimonthly from logarithm-transformed precipitation amounts. Solar radiation, minimum temperatures and maximum temperatures are each conditioned on the precipitation status of the given day. Empirical distributions for solar radiation and temperature are estimated from residual series; means and standard deviations

vary daily through a spline interpolation of monthly means and standard deviations of observed daily data (Qian et al., 2008), with different distributions for wet and dry days.

It was found that AAFC-WG simulates the distributions of wet and dry spells as accurately LARS-WG – with both distributions showing no statistical difference between observed and generated data across Canada (Qian et al., 2008). AAFC-WG was also successful in reproducing the mean and associated root-mean squared error in lengths and of wet and dry spells. AAFC-WG also outperformed LARS-WG at reproducing the observed probability distributions of temperature. Whilst the methods AAFC-WG use look promising, it is important to recognise that it has only been tested across a small number of Canadian weather stations.

2.4 PATTERN SCALING

Whilst it is known that the climate is changing, the magnitude and regional patterns of climate change remain ambiguous whilst societal, economic and technological changes are also uncertain (Herger et al., 2015). It has been long known that there may be an infinite number of 'pertinent, plausible, alternative futures' (IPCC-TGICA, 1999), with some studies arguing that there is no basis for the relative likelihood of such futures to be quantified (Grübler and Nakicenovic, 2001). Mitchell et al. (1999), also states that "even with a perfect projection of concentrations, there would be considerable uncertainties in the resulting climate response". Uncertainties in future projections may arise from choice of scenario, biases within models themselves, differences in simulating the radiative forcing and subsequent climate response, and internal variability (Lee et al., 2021) with Santer et al. (1990) describing each GCM as "an individual species within the genus of climate models". Uncertainties arising between models themselves may involve their treatment of clouds, sea ice and land surface processes, resolution used, and ocean representation (Santer et al., 1990). Uncertainties pertaining to the simulation of radiative forcings may arise through differing conversions of emissions to concentrations, derivations of the forcings themselves, and modelling the resultant climate response (Mitchell et al., 1999), alongside further uncertainties regarding the pathway of future societal, technological, and economic developments (Herger et al., 2015).

The ability to study a wide range of possible futures, and hence assess the associated risks and attempt to quantify some of the abovementioned uncertainty, is of importance to impact modelers and other end-users (Herger et al., 2015; Lynch et al., 2017). It is also

known that long-term GCM projections are more robust than short-term anthropogenically forced simulations due to a more robust signal-to-noise ratio. However, impact assessors tend to need projections on shorter times scales, such as the early- and mid-21st century. With the assertion that the uncertainty in global futures cannot be fully addressed by the study of a select number of climate scenarios, the importance of developing a method to study a wider range of different climate scenarios with several realisations is emphasised.

Pattern scaling (PS) provides a computationally efficient technique that has primarily been used to address two major uncertainties; the underlying scenario used, and the structural uncertainties associated with individual model's climate sensitivities (Lee et al., 2021). PS can be used to construct additional climate change projections using GCM projections, and allows for climate change patterns to be scaled to represent a variety of GWLs, including scenarios with different sensitivities to climate forcing, different concentrations of greenhouse gas emissions, and for time periods where simulations have not been undertaken or are unavailable (Osborn et al., 2016). These simulations may be more useful to impact assessors, who require a variety of climate projections to be used independently or to analyse effects of mitigation policies compared to reference scenarios (Tebaldi and Arblaster, 2014).

A computationally inexpensive technique, PS was originally developed to create projections from equilibrium responses of a GCM to a doubling of CO_2 concentration, however more recently, PS has been used to construct a wide range of transient climate simulations from climate model projections (Osborn et al., 2016; Santer et al., 1990). This technique postulates that, for a given time period and scenario, the change in a climatic variable can be expressed as linear function of the global mean surface temperature change (Cabré et al., 2010; Tebaldi and Arblaster, 2014). A linear relationship is calculated between the GCM-scale changes in a variable, V, and the GMST anomaly, ΔT,

$$\Delta V_{xijy} = \Delta T_{xy} V'_{ij} \tag{2.10}$$

at each grid-cell (denoted by i), month/season (j), period (y), and forcing scenario (x), where V' is a fixed, spatial, response pattern (Kennett and Buonomo, 2006). The response pattern is derived in units of local change in variable V per degree Celsius of warming. For temperature changes, the units tend to be in $\frac{degrees}{degrees}$, whereas precipitation changes are usually expressed as a percentage change per degree, $\frac{\%}{degrees}$.

Two different methods have been discussed widely in the literature to determine the response pattern. The time-slice method (sometimes referred to as the "delta" method) determines the response pattern by studying two different, discrete time periods (epochs) for a given forcing scenario. To determine the response pattern, the difference in a variable's anomaly between a future and a reference period is divided by the change in GMST between the two periods for each GCM grid cell (Mitchell, 2003). The anomaly in variable V is calculated at a regional scale (i.e., at a grid cell defined by i in equation 2.10) with epochs usually around a century apart. It is common practice to choose a baseline epoch at the end of the 20th century, and a future epoch at the end of the 21st century to ensure the signal is from forced change and not internal variability (Tebaldi and Arblaster, 2014). The future epoch is often chosen based on a specified GMST change relative to preindustrial levels. Although computationally efficient, the time-slice method lacks adaptability to additional predictors and has less skill in estimating trends than the alternative linear regression method (Lynch et al., 2017). Results are also variable depending on the choice of epoch.

Linear least squares regression is an alternative method that can be used to obtain the response of variable V to changing GMST. This method is considered more robust and has hence been favoured in recent literature (Lustenberger et al., 2014; Lynch et al., 2017). Opposed to comparing epochs, the full GCM simulated time series can be utilised, improving the signal-to-noise ratio, which is primarily caused by natural variability. This also reduces associated mean-squared error. Due to the use of data across a continuous time series which may include different rates of warming, linear regression is also assumed to be more applicable to the study of presently unmodelled periods and scenarios (Mitchell, 2003). Ruosteenoja et al. (2007) suggest using ensemble means with different initial conditions to determine the response pattern, instead of using individual GCM responses. By using several different realisations of the same GCM, noise caused by internal variability may be further reduced and the strength of the climate change signal further strengthened. Although linear regression is considered a more robust method of obtaining the response pattern, some features, such as temperature change at high latitudes, cannot be approximated by linear relationships.

The merits of this technique are clear, with both PS and the similar "time-shift" approach (not to be confused with the time-slice method for obtaining response patterns) discussed in the recent IPCC Assessment Report 2021 (Doblas-Reyes et al., 2021). The time-shift approach defines a target GWL (a common GWL is 2°C, for example) and finds a

(typically) 20- to 30-year period wherein the GMST increase of the GCM matches the specified GWL (Herger et al., 2015; Matte et al., 2019; Vautard et al., 2014). This method maintains physical consistencies between variables that are *not* maintained by PS, as well as preserving space-time covariance. Despite this, Matte et al. (2019) found high correlations between regional responses obtained by PS and the time-shift approaches over Europe. Although there are benefits to the time-shift approach, its major drawback is that it cannot be used to study GWLs beyond those simulated by GCMs. For example, RCP2.6 scenarios may not be used to estimate certain GWLs reached by RCP8.5 as RCP2.6 simply does not reach the levels of GMST increase of RCP8.5. Furthermore, the time-shift approach does not consider effects if climate responses are a function of both increasing GMST *and* the rate of warming, which may decrease the skill of the technique for long time scales.

While PS provides impact assessors and risk modellers with a computationally inexpensive tool to study a range of alternative scenarios from a limited number of GCMs (Tebaldi and Arblaster, 2014), there are two main causes of uncertainty in determining the response pattern. Both uncertainties, it is important to note, are also present with most emulators of future climate based on GCM results, including the previously mentioned time-shift approach, and are not necessarily unique to PS. The first cause of uncertainty in determining regional climate response to GMST change is once more associated with the simulations themselves, referred to as "inter-model uncertainty" (Giorgi, 2008). This arises from the aforementioned differences between GCM simulations, and is considered by several studies a dominant cause of uncertainty in the construction of regional climate responses (Kittel et al., 1998; Tebaldi and Arblaster, 2014; Tebaldi and Knutti, 2018).

The second is caused by the internal variability in the climate system itself. Several methods have been explored to reduce the impact internal variability has on identifying the response pattern. Using several initial-condition ensembles, as opposed to just one realisation, provides a better identification of the pattern from background internal variability (Mitchell, 2003). The same GCMs simulated under different forcing scenarios (e.g., pooled SSP3-7.0 and SSP5-8.5 data) can be used simultaneously to diagnose the response pattern, once again reducing the impact of internal variability (Osborn et al., 2018). Although this is a cause of uncertainty in determining the response pattern, several studies (Lustenberger et al., 2014; Osborn et al., 2018; Tebaldi et al., 2021) note that the error caused by internal variability is less than the inter-model uncertainty.

A critical assumption of pattern scaling is that there is a linear relationship between local climate change and the GMST increase. This assumption has been assessed in several studies. It has been found that this assumption is mostly valid regarding estimation of regional temperature changes (Cabré et al., 2010; Tebaldi and Arblaster, 2014). Comparisons between the time-shift approach and pattern scaling have been used to assess this assumption of linear scalability. Matte et al. (2019) employed this approach to assess the scalability of temperature and precipitation over Europe, finding higher correlations between the two methods for temperature than precipitation. The assumption of linear scalability with regards to precipitation is known to be weaker than temperature; this is thought to be due to differences in the parameterization of cumulus convection and cloud formation (Matte et al., 2019; Santer et al., 1990) and due to large interdecadal variability in regional precipitation patterns that tend to dominate the climate signal (Cabré et al., 2010). Osborn et al. (2018) showed that some of the poorer performance in the scaling of precipitation responses was due to unforced variability. Using a four-member initial condition ensemble (as opposed to one) much improved the performance of precipitation scaling, though errors still exist. Residual poor performance could be attributed to remaining unforced variability in the four-member ensemble mean. Despite this, precipitation remains approximately linearly scalable with global mean temperature with a good degree of accuracy (Tebaldi and Arblaster, 2014), particularly at longer time scales (Matte et al., 2019). The effect of internal variability on precipitation can be reduced by scaling means over longer periods, e.g., over 30-years. For both temperature and precipitation, Giorgi (2005) found that the non-linear fraction of the climate change signal tends to decrease with the magnitude of the signal. This asserts that the assumption of linearity holds increasingly well as the climate change signal becomes more pronounced (Giorgi, 2005; Kennett and Buonomo, 2006).

It is therefore advised that higher forcing scenarios should be interpolated to study lower scenarios (Mitchell, 2003) and not the other way around. This is thought to be because weaker forcing causes less well-defined patterns of linear coefficients, which, when extrapolated, lack the characteristics present in high forcing responses (Osborn et al., 2018). Although it is possible to obtain and extrapolate a response pattern from low-end warming scenarios, it is recommended that, when using pattern scaling to make projections under high-end warming scenarios, the patterns should be diagnosed from strong forcing scenarios. Lynch et al. (2017) find that signal-to-noise ratios are lower for patterns produced using RCP4.5 than RCP8.0. Osborn et al., (2018) suggest that, up until approximately 3.5^{o}C of

warming, using a combination of warming scenarios (e.g., RCP2.6, 4.5 and 6.0) to determine the response pattern, followed by a high-end warming scenario (RCP8.5) to determine the response beyond.

It is also assumed that the climate system has a linear response to anthropogenic forcing (Mitchell, 2003) and that the climate system and anthropogenic forcing are independent of each other. The validity of this assumption is hindered by short lived, regionally differentiated species, such as anthropogenic aerosols or black carbon (Shiogama et al., 2013; Tebaldi and Arblaster, 2014). While this assumption may not always be true, estimation errors in patterns caused by changes due to internal variability are considered small on a global scale (Lynch et al., 2017).

Several studies have further assessed the accuracies of PS and the validities of its assumptions. The assumption that the change in any climatic variable's spatial pattern for a given scenario remains constant with respect to global temperature change is critical. Cabré et al. (2010) use the time-slice method to assess the validity of this assumption for changes to the annual mean temperature and precipitation, producing spatial patterns for three different epochs. The assumption of scalability was evident for temperature and normalised precipitation changes, with some exceptions in Northern Hemisphere high latitudes. Tebaldi and Arblaster (2014) find that patterns produced by different Representative Concentration Pathways (RCPs) were highly correlated with each other, supporting the assumption that patterns generated by different forcing scenarios are not significantly different. The strongly mitigated RCP2.6, however, was excluded from determining multi-model patterns due to decreased correlations in patterns produced from different centuries. Lynch et al. (2017) note that differences in temperature change patterns produced by different RCPs (8.5 and 4.5) using linear regression were larger than differences using the time-slice approach. Statistically significant differences in patterns were found across mid-high latitudes (the Arctic, land areas bordering the Mediterranean, and the subtropical South Pacific (Lynch et al., 2017)).

Presently, PS is typically used to construct scenarios studying changes to the mean climate and do not consider changes in variability, typically scaling seasonal or annual means in temperature and precipitation. Examples include Cabré et al. (2010) constructing patterns for changes to annual and seasonal mean temperature and precipitation with changing GMST. Lynch et al. (2017) compared the delta-change and least squares regression approaches for the construction of patterns using seasonal and annual means in surface air

temperature. Alexeef et al. (2018), Mitchell (2003) and Tebaldi and Arblaster (2014) are all further examples of scaling mean statistics. Few studies have considered the incorporation of changing variability with GMST, with literature limited in comparison to the scaling of means (Herger et al., 2015; Osborn et al., 2016).

PS is also considered more limited in the study of changes to extreme events, though there has been some research into scaling statistics relating to extreme events. This is likely due to the focus of most PS studies on constructing scenarios relating to mean climate and not the variability. One method used to study and assess changes to extremes was introduced by Lustenberger et al. (2014) wherein extreme temperature indices were scaled with GMST. The scaled relationship has variable results in accuracy, highly dependant on the scaled index. There is evidence that the relationship between some indices (such as cold-spell days index, CWFI, TX10P) and GMST change violates the assumption of linearity, and therefore should not be scaled. As with scaling precipitation and temperature, the uncertainty due to inter-model differences is once again greater than the uncertainty induced by internal variability.

2.5 CHAPTER 2 SUMMARY

With undeniable evidence that our climate is changing, and little certainty regarding which pathway anthropogenic emissions and socio-economic factors will follow in the future, there is an ever-pressing need for sophisticated models that can construct a variety of projections to better understand the range of uncertainties we may face with regards to future climate and weather extremes. GCMs are the most sophisticated tools at our disposal to study the impact changing anthropogenic emissions will have on the Earth's climate. While they provide detailed pictures of global changes the Earth may face, they are limited by their computational requirements, resulting in parameterization of several small-scale processes and therefore a course resolution. However, local-scale projections are much more useful for impact assessors and as inputs to hydrological, agricultural and ecological models. Furthermore, to fully understand the range of uncertainty associated with a changing climate, a wide range of scenarios (and several realisations of such scenarios) must be considered; more than those presently simulated by GCMs (Alexeeff et al., 2018). These uncertainties arise from the definition of scenarios themselves, inter-model differences in the derivation of the climates response to anthropogenic emissions and parameterization, internal variability, and biases within the models themselves.

To address the mismatch between GCM-scale data and impact assessment requirements, a range of downscaling techniques have been introduced. A dynamical downscaling technique, RCMs are one such approach that can be used to construct regional climate projections. RCMs represent several sub-GCM scale features with greater accuracy than GCMs, including orographic effects, precipitation patterns in areas with steep topography, mesoscale processes such as precipitating weather fronts, and land-sea contrasts. However, despite the significant added value RCMs provide, RCMs tend to yield results that are similar to the computationally more efficient statistical approaches. RCMs are also dependent on the driving GCM, and biases in the simulation of present-day climate are still present. Statistical downscaling methods are computationally effective whilst showing similar skill in downscaling GCM simulations to the dynamical methods. There are a range of statistical downscaling approaches, including perfect prognosis, model output statistics, and stochastic weather generators. While downscaling techniques provide users with fine-resolution time series, uncertainties associated with modelling the future climate are not addressed, as only scenarios simulated by GCMs can be downscaled.

Prior to their introduction as a downscaling technique, SWGs were introduced and are widely used as computationally efficient tools that can simulate local-scale time series from observational or reanalysis data. They can produce high-quality, temporally consistent, long-running time series for a suite of weather variables at a local-scale, addressing a key requirement of the data needed for the robust statistical analysis of extreme weather events and with a resolution that is suitable for impact assessments. There are several methods that can be used to stochastically synthesise a suite of climate variables, though most approaches are dependent on their generation of precipitation. Markov chains, series-type weather generators, and NSRP are different approaches to the stochastic modelling of precipitation. Modelling of precipitation has been focussed on in this review, though Chapters 4 and 5 address the generation of secondary and tertiary variables in greater detail. While SWGs are computationally efficient tools at modelling present day climate, they are probabilistic, not deterministic and they cannot be used to simulate time series for different scenarios without some perturbation of the parameters to represent future climates.

The construction of a range of scenarios, including those that are not simulated by GCMs, is necessary to better understand the wide range of uncertainties in modelling the future climate. Pattern scaling is a computationally efficient technique that has been introduced to construct projections for a range of scenarios or time-periods that have not been simulated by GCMs, or scenarios with presently only a limited number of ensembles.

A linear relationship between local climate change and changing GMST is assumed, with local responses of a variable to global climate change diagnosed from a GCM, normalised by GMST change. However, PS has typically been used to diagnose the response of seasonal (or annual) means in temperature and precipitation and has limitations in the study of extreme events.

The advantages of both stochastic weather generators and pattern scaling will be utilised in this thesis through the development of a globally calibrated, statistical tool, hereafter referred to as the **G**lobally **C**alibrated stochastic **W**eather **G**enerator (GCWG). The development of the GCWG in reproducing a range of observed climates (Belda et al., 2014) is presented in Chapters 3 to 5, capable of generating daily times series of precipitation, maximum and minimum temperature, mean wind speed, solar radiation and relative humidity. The response of GCWG precipitation and temperature parameters (see Chapters 3 and 4) to external forcing will be diagnosed from GCMs and hence used to perturb parameters calculated from weather station observations. The perturbed GCWG is therefore capable of generating local scale, long running time series for a suite of weather variables under any specified GWL representative of a range of scenarios, whilst avoiding GCM induced errors in modelling point-scale climate. Furthermore, the simulation of several realisations at a specified GWL or scenario with computational efficiency is possible using the GCWG. Unlike the discussed time-shift method, the use of PS provides a unique opportunity to produce long time series at a specified GMST level, rather than an average over a relatively small period (e.g., 30 years) to better quantify uncertainties related to the changing risk of weather extremes. To reiterate, the strengths of a stochastic weather generator and pattern scaling both address issues with using GCM projections to study the future climate alone, and both techniques will therefore be utilised in the construction of the GCWG to simulate future climate scenarios. Perturbing the parameters of the GCWG through PS will first address the issue regarding the limited number of scenarios (and the number of realisations) produced by GCMs. The perturbed GCWG also addresses the mismatch between GCM resolution and the scale required for impact assessments, while additionally providing an opportunity to produce several realisations of local climate with long-time series, and hence study the associated changing risk to extremes.

3 STOCHASTIC PRECIPITATION GENERATION WITH MARKOV CHAINS OF DIFFERENT ORDERS

Precipitation is considered a primary variable in the development of the **G**lobally **C**alibrated stochastic **W**eather **G**enerator (GCWG). This is due to the conditioning of secondary and tertiary variables on the precipitation of the day and the use of precipitation amounts as predictors. In this thesis, a Markov-chain gamma-distribution approach to generating stochastic precipitation sequences has been adopted in the GCWG following a comparison of methods presented Chapter 2, Section 2.3. To reiterate, the strengths of a Markov-chain gamma-model include their simplicity and computational efficiency, their common usage resulting in rich literature aiding in the development of a precipitation generator, flexibility regarding model-order, and finally, owing to the simplicity of the technique, it is relatively straightforward to diagnose the response of the parameters to global mean surface temperature (GMST) changes from climate model projections.

For a first-order Markov-chain model (the most widely used in literature), only 4 parameters are required to capture precipitation occurrence and the wet-day gamma distribution per month. Although series-type weather generators can approximate a wide variety of shapes effectively, many more parameters are required (for example, LARS-WG, a series-type weather generator, uses 21 per month) (Semenov and Barrow, 2002). Diagnosing the responses of Markov-chain gamma-distribution parameters to GMST increase is therefore also simpler than the parameters from alternative approaches. It is also apparent that, while first-order models are the most prevalent in literature, there are merits to using higher (and lower) orders, varying with geographic location and season (Hosseini et al., 2011). This provides some flexibility regarding the modelling of precipitation. Generally, zeroth order models require less computational power than higher orders, while second- and third-order models may have the capacity to better represent long dry spells, a known deficiency with first-order models (Lennartsson et al., 2008). Model-order performance is also variable on season, with Chin (1977) finding that higher orders are preferential in the winter season across the USA. Hybrid-order Markov-chain models have been suggested to accurately represent dry climates, with first-order dependence for wet spells, and higher-order dependence for dry-spells (Chen and Brissette, 2014; Wilks, 1999b)
.

In this thesis, secondary (maximum and minimum temperatures, see Chapter 4) and a selection of tertiary variables (mean wind speed, solar radiation and relative humidity, see Chapter 5) will be conditioned on daily precipitation. Therefore, shortcomings in simulating precipitation occurrence may lead to inaccuracies in secondary and tertiary variables. This emphasises the need for an educated choice in model order used to simulate occurrence and amount of precipitation as accurately as possible. This chapter covers the methods used to simulate statistically similar time series to observed precipitation records, including a global assessment of model-order performance.

Many widely used stochastic weather generators have been calibrated for a specific study location. This limits their applicability to regions with different climate characteristics to the calibration site. The main section of this chapter (Section 3.1) has already been published under Wilson Kemsley et al. (2021) in *International Journal of Climatology* and compares the performance of zeroth, first, second and third order models at reproducing observed precipitation statistics at over 40,000 weather stations, covering five climatic regimes. Additional evaluations and analyses not included in that paper are provided here in Section 3.2. Results will be used to inform stochastic weather generator users of which model order performs best in each climatic regime. A Markov-chain-gamma-distribution model has been coded in Python to stochastically generate precipitation time series. Code has been written for zeroth, first, second, and third order chains.

3.1 SELECTING MARKOV CHAIN ORDERS FOR GENERATING PRECIPITATION SERIES ACROSS DIFFERENT KÖPPEN CLIMATE REGIMES

3.1.1 INTRODUCTION

Stochastic weather generators are a technique used to produce synthetic rainfall time series with high spatial and temporal resolutions. Computationally inexpensive tools, they can be used to produce long time series for use in hydrological and agricultural risk assessments when the record length or quality of representative observational data are inadequate. For example, measurement time series are often too short to robustly estimate the probability of extreme events, such as long wet- or dry-spells. Stochastic weather generators were initially developed to address these issues (Semenov et al., 1998), though they have since been applied to a broader range of problems (e.g., perturbing their

parameters so they produce synthetic time series under future climates (Eames et al., 2011)) and to other climate variables.

Precipitation is one of the most important variables for assessing risks affecting crop growth and the hydrological cycle. Precipitation is also often considered as the primary variable when stochastically modelling other weather variables (e.g., solar radiation, maximum and minimum temperatures) which can be conditioned on the precipitation status (Richardson, 1981). It is important that impact and risk assessors have access to the most accurate high-resolution models (Dubrovský, 1997), as generated data is often used in place of insufficient observed records as an input to hydrological, ecological and agronomic studies (Larsen and Pense, 1982). Therefore, it is particularly important to ensure accurate modelling of daily precipitation.

Markov chains are a commonly used stochastic approach to modelling daily precipitation. Simulated occurrence of rain is conditional on the previous day(s)' precipitation status. The order of a Markov chain model refers to the number of previous days considered, i.e., a first-order model conditions precipitation status on the status of one previous day. Several of the widely used Markov-type weather generators (for example WGEN (Richardson and Wright, 1984) SIMMETEO (Soltani and Hoogenboom, 2003a) and AAFC WG(Qian et al., 2005)) use the same model order regardless of geographical location. Furthermore, many of these weather generators have been designed, implemented and tested for climates local to where they were produced, meaning they may not be optimal at generating realistic rainfall time series in different climates(Semenov et al., 1998). Despite the common use of first-order models in climate and hydrological impact studies, their limited memory of extremes has been criticised (Semenov and Barrow, 1997), and though it is accepted that different sites require different orders, first-orders remain prevalent (Lennartsson et al., 2008). However, the stochastic dependence on previous days' precipitation is dependent on particular meteorological drivers (Chin, 1977), with first-order models commonly misrepresenting important meteorological properties (Ailliot et al., 2015).

Although Markov-chain models are the focus of this study, there are several other methods used to stochastically model precipitation occurrence, including Markov renewal processes (MRP) (Foufoula-Georgiou and Lettenmaier, 1987) and series-based approaches (e.g. LARS-WG (Qian et al., 2005)). Both methods use observed wet or dry-spell lengths to generate synthetic precipitation time series differently to Markov-chain models. In MRP

models, the probability of precipitation depends on the number of days since the last precipitation event, whereas series-based weather generators draw lengths of wet and dry spells from semi-empirical distributions (Semenov et al., 1998).

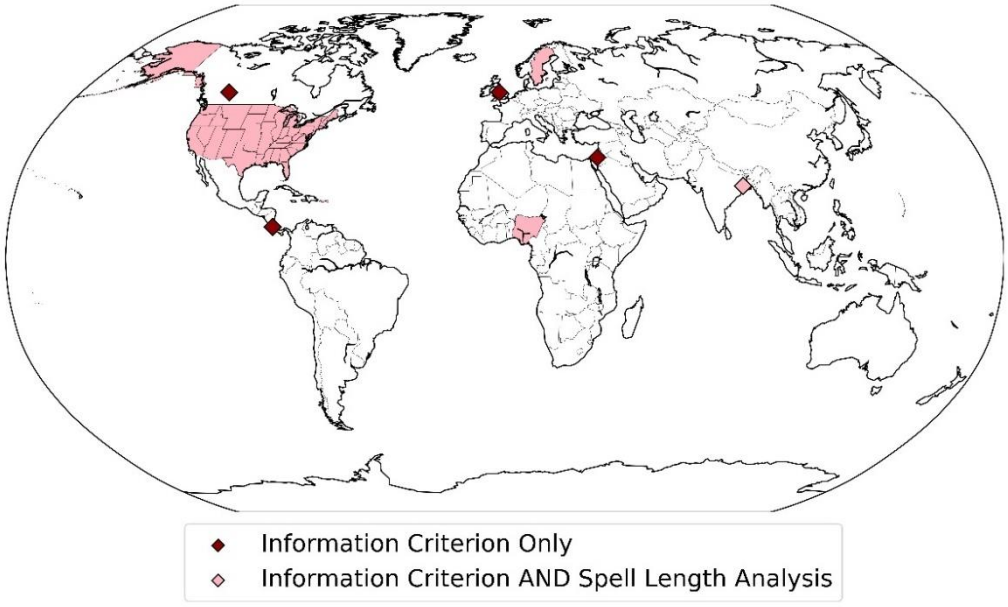


Figure 3.1 – Global locations of previous Markov chain model-order assessments used to inform this study.

Akaike (Tong, 1975) and Bayesian Information Criteria (Schwarz, 1978) are commonly used to select model order, balancing the model fit with the number of parameters. There have been several local or regional studies which have assessed the best model-order, including across the USA (Schoof and Pryor, 2008), Costa Rica (Harrison and Waylen, 2000), England (Gates and Tong, 1976), Sweden (Lennartsson et al., 2008), Canada, Israel, India and Nigeria (Jimoh and Webster, 1996) using Information Criteria and other methods, including spell length analysis (Figure 3.1). Although many of these studies reiterate that a first-order model is adequate, seasonal and spatial variations in model-order choice were identified in all of them. Further studies critique this first-order assumption (Gates and Tong, 1976) and the use of an Information Criterion as the sole method of model-order selection (Hosseini et al., 2011).

Further evidence suggests that using a first-order model, regardless of location, may not always reproduce observed weather accurately. Low order models are known to underestimate interannual and inter-seasonal precipitation variances (suffering from overdispersion (Katz and Parlange, 1998)), despite often being chosen by Information

Criteria (Harrison and Waylen, 2000). Higher-order models may more accurately reproduce variances in weather, though the risk of overfitting becomes more prominent.

A common use of precipitation generators is to estimate the probabilities of extreme events, such as long wet or dry spells, due to their ability to simulate long time series (Yaoming et al., 2004). Although Information Criteria are the most common methods used to assess model-order performance, they do not test the ability to accurately reproduce the distribution of wet- and dry-spell lengths, despite their importance for many applications. It is known (Lennartsson et al., 2008) that first-order models often do not reproduce dry spells accurately even though they are favoured by Information Criteria. This inability is an established issue with many Markov-type weather generators (Semenov and Barrow, 2002). Wet spells are important for hydrological modelling, impacting flood risk and soil erosion, whilst knowledge of dry-spell behaviour impacts agricultural and environmental planning, preparation for drought and irrigation infrastructure (Ochola and Kerkides, 2003). Therefore, a method to assess the ability of different model orders to reproduce observed wet- and dry-spell distributions is needed.

This chapter extends the existing knowledge base in several respects that are important for providing better guidance to users and developers of Markov chain weather generators. First, we undertake an assessment of preferential model order using *multiple* metrics of model performance. Alongside the Bayesian Information Criteria (BIC), we also quantify the ability of different model orders to reproduce observed distributions of wet- and dry-spell length, and the interannual variability of precipitation occurrence (referred to as IVO). In each case, we consider the relative performance of models with orders 0, 1, 2 and 3. Higher orders are not considered in this study due to their increased risk of overfitting and greater computational demand, reducing usability. Second, we undertake a global assessment using more than 44,000 weather stations across most land areas. Third, using the global assessment, we consider whether the choice of model order is systematically dependent on the climate regime of each location.

Local climates can be classified into regimes according to several different characteristics, such as annual precipitation, interannual temperature and precipitation variances, cloud cover etc. (Belda et al., 2014). Here we use the Köppen climate classification which is one such classification in widespread use (Köppen, 1900), using monthly temperature and precipitation statistics to identify different climatic regimes across the world (Figure 3.2).

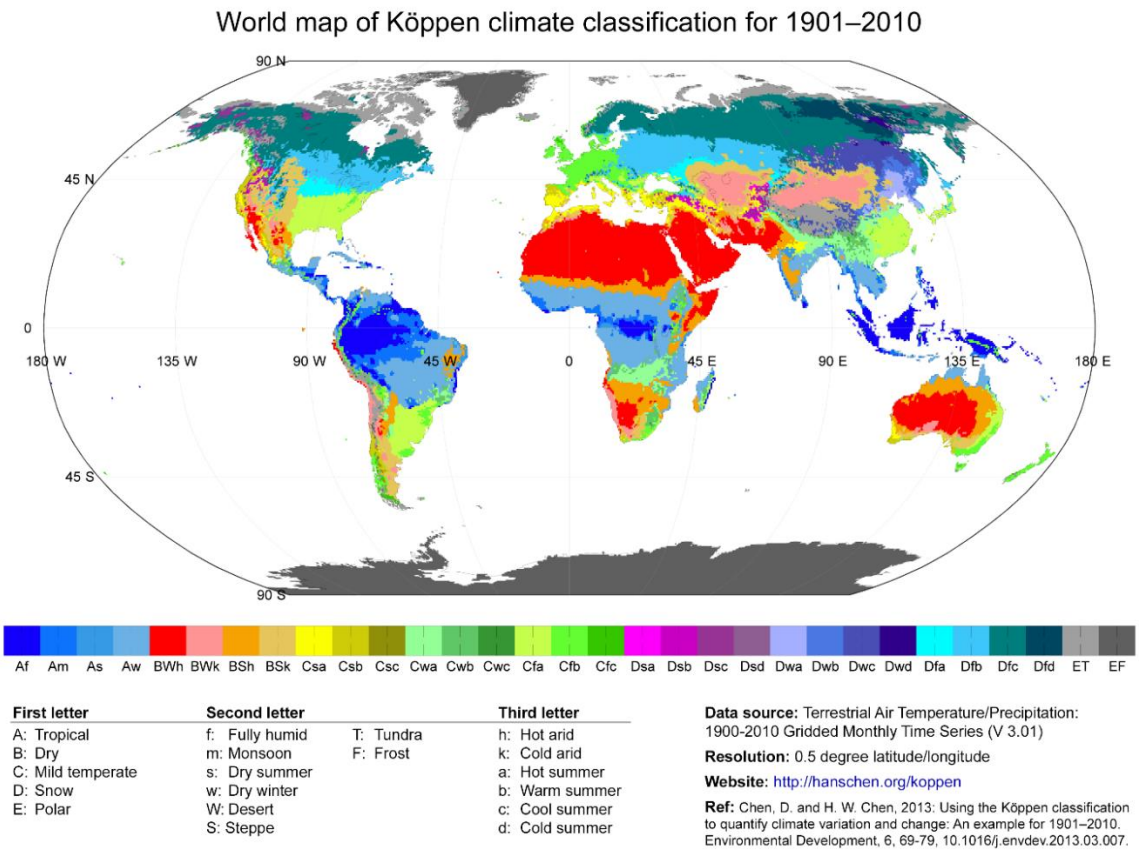


Figure 3.2 – The Köppen climate classification according to Chen and Chen (2013). There are some differences in terminology used for each class; in the current study, we denote class C as "temperate" and class D as "continental". Figure from Chen and Chen (2013)

3.1.2 DATA

We used daily precipitation data from the Global Historical Climatology Network Daily (GHCN-D) to fit Markov chain properties and evaluate the models. GHCN-D quality control procedures flag potentially inaccurate or inconsistent records, estimated to affect approximately 0.3% of the data. Flagged data was removed, and only weather stations with at least 20 cumulative years of daily precipitation remaining (Soltani and Hoogenboom, 2003b) were used, resulting in a total of 44071 stations. Historical records were capped at their most recent 30 years, to reduce any artefacts that may arise from using records of different lengths.

Each weather station was allocated a climate classification, determined by the centre of the Köppen grid cell which the longitude and latitude of the weather station was closest to. Locations are allocated to one of five overarching classes: tropical, dry, temperate, continental or polar. Within each class, regimes can be subcategorized further based on climatic behaviour. Our study applied each evaluation method to observed data in each overarching regime to determine the most appropriate model order for each climatic zone. Köppen classification data was taken from Chen and Chen (2013) on a 0.5°x0.5° grid.

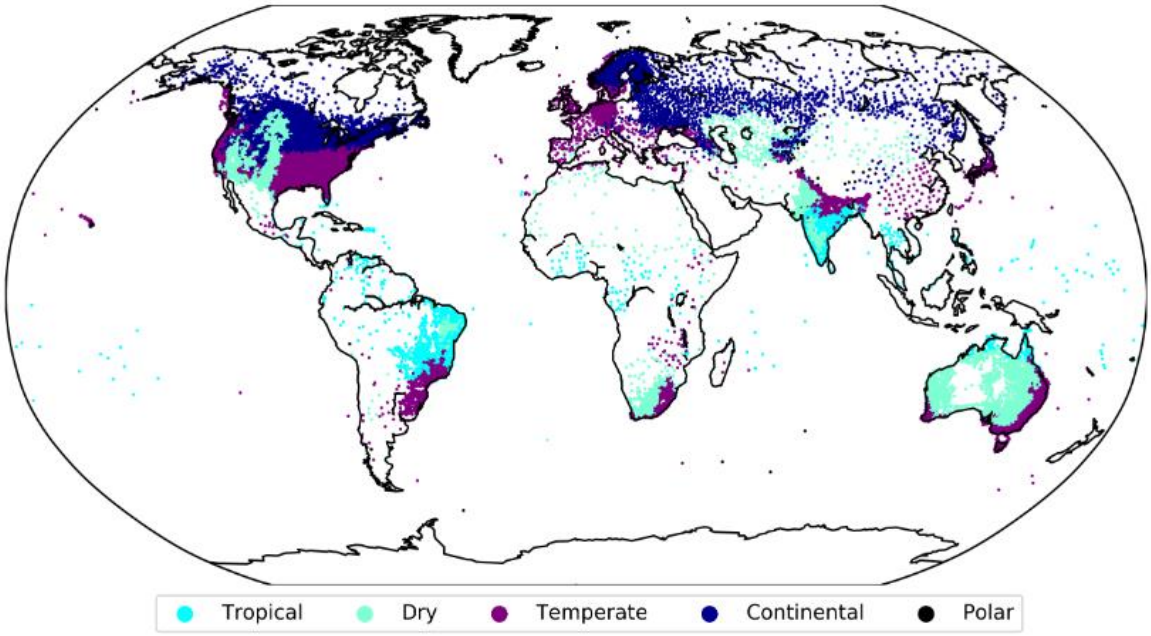


Figure 3.3 – All available weather stations meeting the outlined criteria and their associated overarching climate classification.

It is clear from Figure 3.3 that there is not global coverage, and that some locations, such as the USA and most of Europe, have a significantly higher density of weather stations than others (such as Africa and South America). Nevertheless, even the least sampled class – polar – has over 400 stations with at least 20 years of observed data.

3.1.3 METHOD

3.1.3.1 MARKOV MODEL FITTING

A 2-state Markov chain-gamma model was used to simulate the occurrence and amount of precipitation, with 2-states referring to a day being either wet or dry. A day was described as "wet" if precipitation was above 0.1mm. This method of stochastic modelling is loosely based on the weather generator WGEN (Richardson and Wright, 1984). The probability of a wet day is conditional on the historical precipitation status. A k-th order

Markov-chain refers to the number of conditional k previous days. For the widely used first-order, 2-state model, the transition probabilities are:

$$p_{ij} = \frac{n_{ij}}{\sum_{m=0}^{1} n_{im}} \tag{3.1}$$

where i and j can represent wet (1) or dry (0) days. For example, p_{01} is the probability of a wet day following a dry day, and n_{01} the number of wet days following dry days (calculated from a historical dataset). This process can be extended to other orders:

$$p_i = \frac{n_i}{\sum_{m=0}^{1} n_m}$$

$$p_{ijq} = \frac{n_{ijq}}{\sum_{m=0}^{1} n_{ijm}}$$

$$p_{ijql} = \frac{n_{ijql}}{\sum_{m=0}^{1} n_{ijqm}} \tag{3.2}$$

for orders 0, 2 and 3 respectively. The number of transition probabilities calculated therefore increases exponentially and can be generalised to 2^{k+1} (where k is model order). Transition probabilities were calculated for each month at each weather station, resulting in $12(2^{k+1})$ transition probabilities for each station and each model order. However, the number of independent transition probabilities is half this number, i.e., $12(2^k)$, because $p_1 = 1 - p_0$.

It is a common assumption that precipitation amount is conditionally independent of precipitation occurrence (Richardson, 1981). Upon generating a wet day, a random precipitation amount is independently taken from the corresponding month's two-parameter gamma distribution. For each month, shape (α) and scale (β) parameters were calculated from wet-day only data for each station. Thom estimators (Thom, 1958) were used for calculating the shape,

$$\alpha = \frac{1 + \left(1 + \frac{4D}{3}\right)^{\frac{1}{2}}}{4D} \tag{3.3}$$

and the scale,

$$\beta = \frac{\bar{x}}{\alpha} \tag{3.4}$$

parameters, with sample statistic

$$D = \ln(\bar{x}) - \frac{1}{n}\sum_{i=1}^{n} \ln(x_i) \tag{3.5}$$

where x refers to amounts of precipitation and n the number of wet days. This method is used in place of moment estimators, which are considered "inefficient" compared with the Thom estimators that make better use of the information in a dataset (Wilks, 2011).

3.1.3.2 MARKOV MODEL EVALUATION

Information Criteria are methods that can be used to determine how suitable different order Markov chains are for modelling precipitation occurrence, based on calculations of log-likelihood functions from the transition probabilities (Schoof and Pryor, 2008). Here, we chose the Bayesian Information Criterion (BIC) over the Akaike Information Criterion (AIC) because Katz (1981) showed that AIC estimators can be inconsistent, and also BIC is less prone to asymptotic bias and more widely used in recent literature (Harrison and Waylen, 2000; Schoof and Pryor, 2008). This method seeks to find the best model containing the fewest parameters (i.e., minimal k). For a 2-state Markov chain model of order k, BIC values are given by

$$BIC(k) = -2L_k + 2^k[\ln(N)] \tag{3.6}$$

where L_k is the log-likelihood function and N is the number of days in the historical record used to calculate the transition probabilities. L_k is calculated from the estimated transition probabilities using the functions given by Schoof and Pryor (2008) for each model order. The model order that minimises BIC is chosen as the best order.

As there are 12 sets of transition probabilities for each station, 12 BIC values for each model order were determined and compared enabling an evaluation of annual and seasonal model-order dependence. The mode of the model orders that minimise BIC across each of the 12 months was taken as the annual model order. Seasonal model-order choices

were also studied, using the mode of the model orders that minimise BIC across June, July, August (JJA) and December, January, February (DJF).

In addition to BIC, three other assessment criteria are considered, providing additional information on model suitability that cannot be gained by BIC alone. Each order's ability to reproduce the distribution of wet and dry spell lengths was compared. A wet spell was defined as a period of wet days preceded and followed by a dry day (and vice versa for a dry spell). For each model order, 50 years of precipitation data were generated and the length of every wet and dry spell in the generated data determined.

Kernel density estimation (KDE) can be used to determine a non-parametric probability density function of a random variable (Guidoum, 2015). KDE was used here to estimate probability density distributions (Rajagopalan et al., 1997) for wet- and dry-spell lengths (see Figure 3.4 for an example). Upon determining distributions for each model-order, the root-mean-squared-difference (RMSD) between the observed distribution and the distributions produced by each Markov chain was calculated. A bandwidth of 2 days was used for KDE and a test across the climate regimes indicated that the results are not sensitive to this choice. In addition to comparing the full spell length KDE distributions, we also evaluate the performance of each model order at reproducing four percentiles (75, 90, 95 and 99th) in the tail of the distributions. These percentiles are calculated directly from the underlying spell length data, rather than from the KDE distributions to avoid any dependence on the choice of KDE bandwidth.

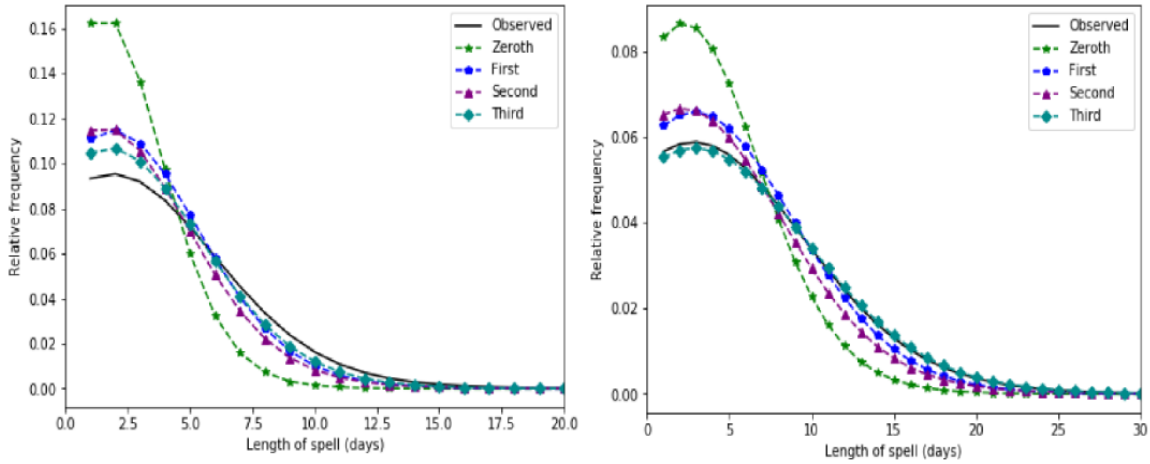


Figure 3.4 – Probability density distribution of the length of dry spells (left) and wet spells (right) produced by different model orders over a 50-year period at a weather station in Highcroft, Tasmania (GHCN-D station ID ASN00094028)

Finally, each models' ability to reproduce the IVO was tested. Once again, 50 years of daily precipitation occurrence was generated. The total number of wet days in each season or year was recorded, and the standard deviation of these 50 values (number of wet days per season/year over the 50-year period) was calculated. The model order producing the smallest absolute difference between the generated and observed standard deviation was deemed to perform best.

3.1.4 RESULTS

The methods outlined in section 3 were applied to 44071 weather stations that met criteria detailed in section 3. Results from these weather stations were aggregated into 837 5° latitude by 5° longitude cells (Table 3.1) to reduce biasing area-average results to locations with dense station coverage (e.g., USA, Figure 3.3). This also increased the weighting of locations with sparse data coverage, for example across Africa and South America. The climate regime of a grid-cell was allocated by taking the modal classification from each individual station in the cell. This methodology was also applied to obtain an overall model-order for each grid-cell and each method of analysis used.

Table 3.1 – Number of weather stations and corresponding grid cells in each climate regime.

REGIME	NO. OF STATIONS	NO. OF CELLS
Tropical	4140	173
Dry	8555	139
Temperate	20389	192
Continental	10552	229
Polar	435	104
TOTAL	**44071**	**837**

3.1.4.1 BAYESIAN INFORMATION CRITERION

The Bayesian Information Criterion was the first method used to analyse model-order performance. The BIC optimisation most often selected a first-order model to minimise BIC across all regimes (Table 3.2), in agreement with previous literature (Schoof and Pryor, 2008) but extending this finding to more regions, regimes and seasons. Although a first-order model may be allocated overall to each regime, there are clear behavioural differences between the zones. For tropical and dry regimes, a sizable minority of cells show

zeroth- or second-order dependence. A sizable minority of temperate cells (17.1%) are minimised by a second-order model, however unlike tropical and dry regimes, there are very few cells minimised by zeroth-order model. Conversely, first-order models almost universally minimise BIC for continental and polar cells (99.2% and 91.8% respectively).

Table 3.2 – Gridded comparison of Markov model-order choices for each of the Köppen climate regimes based on each model order's ability to minimise BIC. Values shown are the % of grid cells within each climate regime where the mode of individual stations' best model-order is equal to 0, 1, 2 or 3.

	0			1			2			3		
REGIME	**Ann.**	**DJF**	**JJA**	**Ann.**	**DJF**	**JJA**	**Ann.**	**DJF**	**JJA**	**Ann.**	**DJF**	**JJA**
Tropical	11.8	17.4	18.5	56.2	53.2	53.8	30.3	27.5	24.7	1.7	3.8	3.9
Dry	18.2	16.1	24.5	63.6	67.8	55.9	17.5	15.4	19.6	0.7	0.7	0.0
Temp.	3.2	4.4	5.5	80.7	80.1	77.9	17.1	15.5	16.6	0.0	0.0	0.0
Cont.	0.4	0.0	0.4	99.2	99.2	98.7	0.4	0.8	0.8	0.0	0.0	0.0
Polar	4.1	3.1	4.1	91.8	84.5	86.6	3.1	11.3	8.2	1.0	1.0	1.0

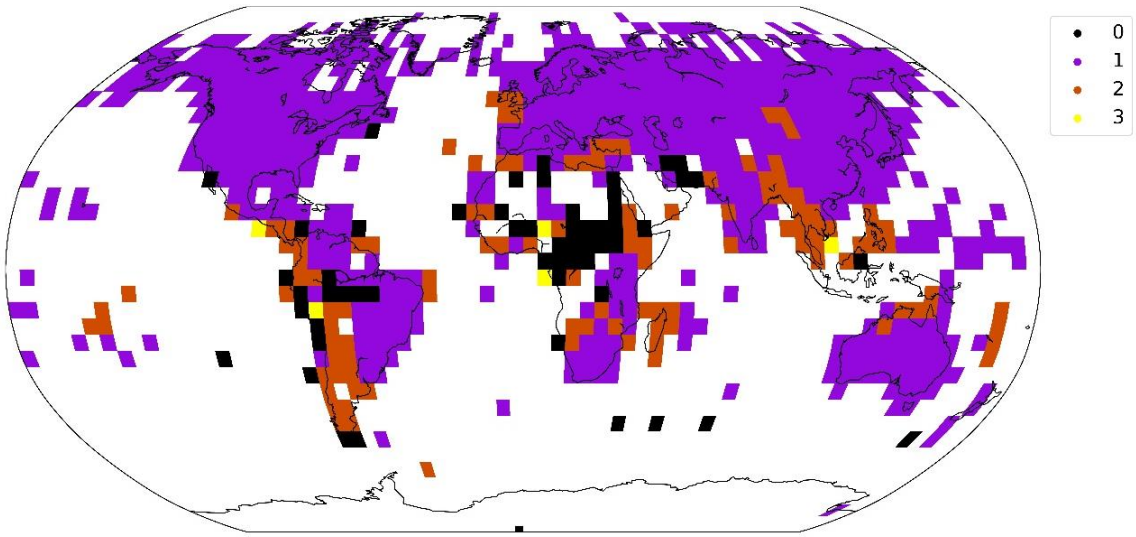


Figure 3.5 – Model-order choices for each 5°x5° grid cell, with colour representing model order choice as determined by Bayesian Information Criterion

These differences in behaviour may arise from spatial variation (Figure 3.5). The majority of the northern hemisphere extra-tropics are best represented by a first-order model. There is much more variation across the southern hemisphere and the tropics, where zeroth- and second-order models are much more prevalent. For example, Chilean and Peruvian cells

do not show first-order dependence, but zeroth or second. Central African cells also show zeroth-order dependence, whereas Mainland Southeast Asia has many cells with second-order dependence.

Central Africa primarily has tropical humid climates (fully humid, monsoon or with dry winters: Af, Am and Aw, respectively) alongside hot dry steppe and desert climates to the north (BSh and BWh respectively – Figure 3.2). Cells in tropical humid (Af) regions were found to show both first- and second-order dependence in equal measures (43.4%). Although dry desert (BWh) cells showed overall preference for first-order models (49.0%), a significant minority (33.7%) showed zero-order dependence. These differences in climatic zone account for some of the spatial variation noted in Figure 3.5. Cold arid dry steppe (BSk) cells are most often found in the northern hemisphere and show differences in behaviour to the overarching regime (dry). BIC was minimised by a first-order model in 91% of BSk cells – almost 30% more than the total dry group. The BSk cells without first-order dependence mostly occur in the South American steppe and instead show mostly second-order dependence. Sub-regimes across temperate, continental and polar regimes performed similarly to their aggregated behaviour.

Several geographical regions have no usable GHCN-D data as indicated by an absence of cells in their location. These regions include Northern and South-West Africa, Maritime Southeast Asia, and the Middle East. These locations are often tropical and dry and are located around the equator or southern hemisphere. This is a limitation with using in situ observed data and has the potential to affect overall model-order preferences.

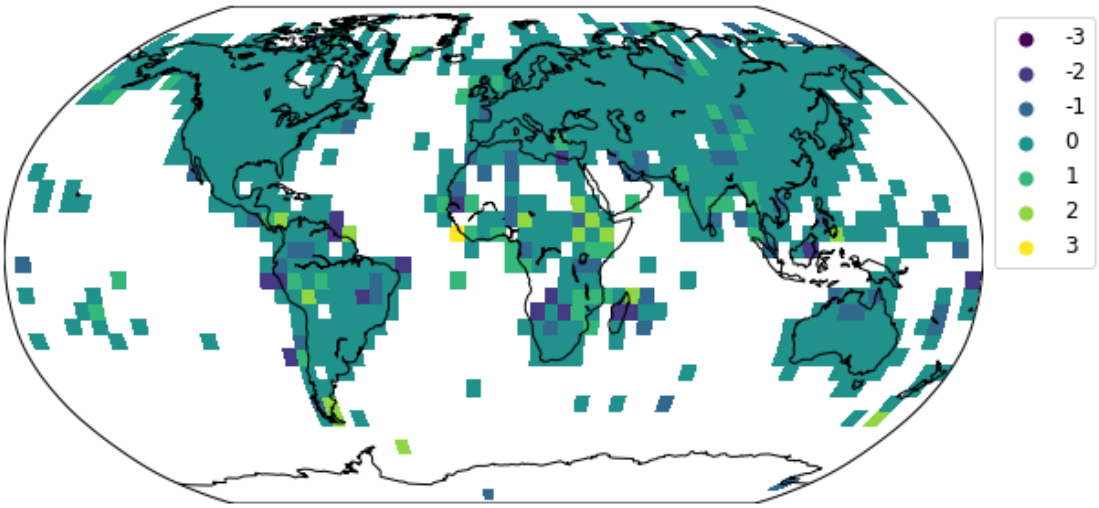


Figure 3.6 – Model-order choice for JJA (BIC) minus the model-order choice for DJF (BIC).

The same overall picture emerges when model orders with the best BIC are considered on a seasonal basis (Table 3.2), i.e., that first-order models are selected most often in all Köppen regimes but zeroth-order models are a notable minority in tropical and dry regimes and second-order models are a notable minority in tropical, dry and temperate regimes. Nevertheless, some seasonal differences are apparent and can be seen in the model-orders for DJF subtracted from JJA (Figure 3.6). Whilst the same model order is selected by BIC across most of the northern hemisphere in both seasons, it is once again cells in the tropics and southern hemisphere that experience the most variation. This variation is shown predominantly in tropical and dry regions once again (Table 3.2), although these fluctuations are minor. Dry climates exhibit the most noticeable seasonal fluctuation, with first-order minimising BIC in 67.8% cases in DJF but only 55.9% in JJA. This is reflected in the increased number of zeroth-order cells in JJA.

3.1.4.2 SPELL-LENGTH ANALYSIS

Although first-order models most commonly minimise BIC across all climate regimes, this is not always the case when other metrics are used to assess model-order performance.

Table 3.3 – Gridded comparison of Markov model-order choices for each of the Köppen climate regimes based on each model's ability to reproduce the interannual variability of precipitation occurrence (IVO) and distributions of wet-spell length and dry-spell length. Values shown are the % of grid cells within each climate regime where the mode of individual stations' best model-order is equal to 0, 1, 2 or 3

	0			1			2			3		
REGIME	**Wet**	**Dry**	**IVO**	**Wet**	**Dry**	**IVO**	**Wet**	**Dry**	**IVO**	**Wet**	**Dry**	**IVO**
Tropical	5.1	0.0	1.7	21.3	6.3	7.9	25.8	8.6	41.0	47.8	85.1	49.4
Dry	0.0	1.4	0.0	67.8	9.2	11.2	7.7	10.6	63.9	24.5	78.7	25.2
Temp.	1.1	1.1	0.0	58.6	6.7	13.3	6.6	3.9	58.6	33.7	88.2	28.2
Cont.	0.0	0.4	0.0	89.9	5.2	25.6	0.0	0.0	53.9	10.1	94.4	21.4
Polar	0.0	3.2	1.0	69.1	4.4	17.5	4.1	1.1	54.6	26.8	93.2	26.8

Across all regimes except tropical, a first-order model most commonly outperforms others at reproducing the distribution of wet-spell lengths (Table 3.3). First-order models outperform the other models for wet-spell length at a large majority (89.9%) of continental

regime grid cells, whereas sizeable minorities of grid cells are best represented by third-order models across dry, temperate and polar regimes (24.5%, 33.7% and 26.8% respectively). In the tropical regime, by contrast, third-order models most often perform the best, followed by the second and then first -order models. Across all regimes, zero-order models rarely or never reproduce wet-spell length distributions best.

Though third-order models only outperform other orders for wet-spell lengths in tropical regimes, they are dominant as the best order for reproducing dry-spell length distributions (Table 3.3) across *all* regimes. This supports the work of Lennartson et. al. (2008), who found that higher order models better reproduce the distribution of very long dry spells in Sweden. There is less spatial variation of optimal model order for dry spells than with BIC and wet spells, reiterated by the high percentage (at least 78% across each regime) of cells being best represented by a 3rd order Markov chain (Table 3.3). For tropical and dry climates, there is a monotonic increase in preference for a model as the model's order increases (hence with the lowest percentages for zeroth-order models). For the other climate regimes, more than 88% of grid cells have third-order models as the best performing for dry spells, and first-order models are the next most frequent.

Table 3.4 – The gridded median percentage difference between observed and generated dry-spell length percentiles for each of the Köppen climate regimes and each model order.

		PERCENTILE			
	MODEL ORDER	**75th**	**90th**	**95th**	**99th**
TROPICAL	**0**	-6.5	-20.0	-25.3	-38.5
	1	14.5	3.4	-4.0	-19.9
	2	0.0	0.0	-4.6	-15.2
	3	3.8	8.3	1.2	-8.0
DRY	**0**	-20.5	-29.3	-34.2	-38.3
	1	5.8	-5.8	-13.1	-24.0
	2	-15.3	-16.3	-20.1	-23.6
	3	5.0	-3.8	-9.3	-17.7
TEMPERATE	**0**	-26.4	-33.4	-36.3	-39.3
	1	4.8	-1.6	-6.7	-16.2
	2	-25.0	-18.1	-16.5	-18.4
	3	3.5	0.3	-0.8	-6.0
CONTINENTAL	**0**	-25.0	-33.5	-35.4	-39.9
	1	3.0	-3.4	-7.9	-13.7
	2	-25.0	-17.9	-16.7	-17.0
	3	3.0	0.4	-1.0	-5.4
POLAR	**0**	-25.0	-33.3	-38.5	-46.7
	1	4.1	-3.1	-8.2	-18.2
	2	-25.0	-16.7	-18.0	-20.7
	3	0.0	0.0	0.0	-7.7

The ability of each model order to reproduce percentiles of the dry-spell length distributions was also studied. Percentage differences between the observed and generated dry-spell length 75^{th}, 90^{th}, 95^{th} and 99^{th} percentiles were calculated for each station, with the median percentage difference across grid cells given in Table 3.4. Medians were taken across regimes as opposed to means to reduce the influence of outliers.

Whilst each model order underestimates the 99^{th} percentile of observed dry spells, there is a large improvement from zeroth order to third. Second- and first-order models tend to underestimate by a similar percentage, with third orders having as little as a 5.4% underestimation in continental regimes. This supports the work of Lennartson et. al. (2008), suggesting that lower order models do not accurately reproduce the distribution of very long dry spells. The largest median bias in these extreme (99^{th} percentile) dry spells occurs in the dry Köppen regime, but even here the bias is much less when considering something slightly less extreme (e.g., 95^{th} percentile) providing that a third order model is used. In all regimes, there is again a large improvement in reproducing the 95^{th} percentile from zeroth to third order, though in this case second-order models tend to have larger median biases than first order. This is particularly noticeable in temperate, continental, and polar regimes. Third-order models continue to have the smallest percentage differences at the 75^{th} and 90^{th} percentiles except for the tropical regime where the second-order model performs best.

The dominant patterns in the results are: (1) at the 75^{th} percentile, first and third order models slightly overestimate dry-spell lengths whereas the other two orders underestimate them; (2) the median bias in dry-spell lengths mostly decreases monotonically as the percentile increases, ending in the underestimation of the 99^{th} percentile spell lengths by all model orders and (3) with few exceptions the biases are smallest for the third-order model, though the first-order model performs nearly as well for the 75^{th} and 90^{th} percentiles in most regimes.

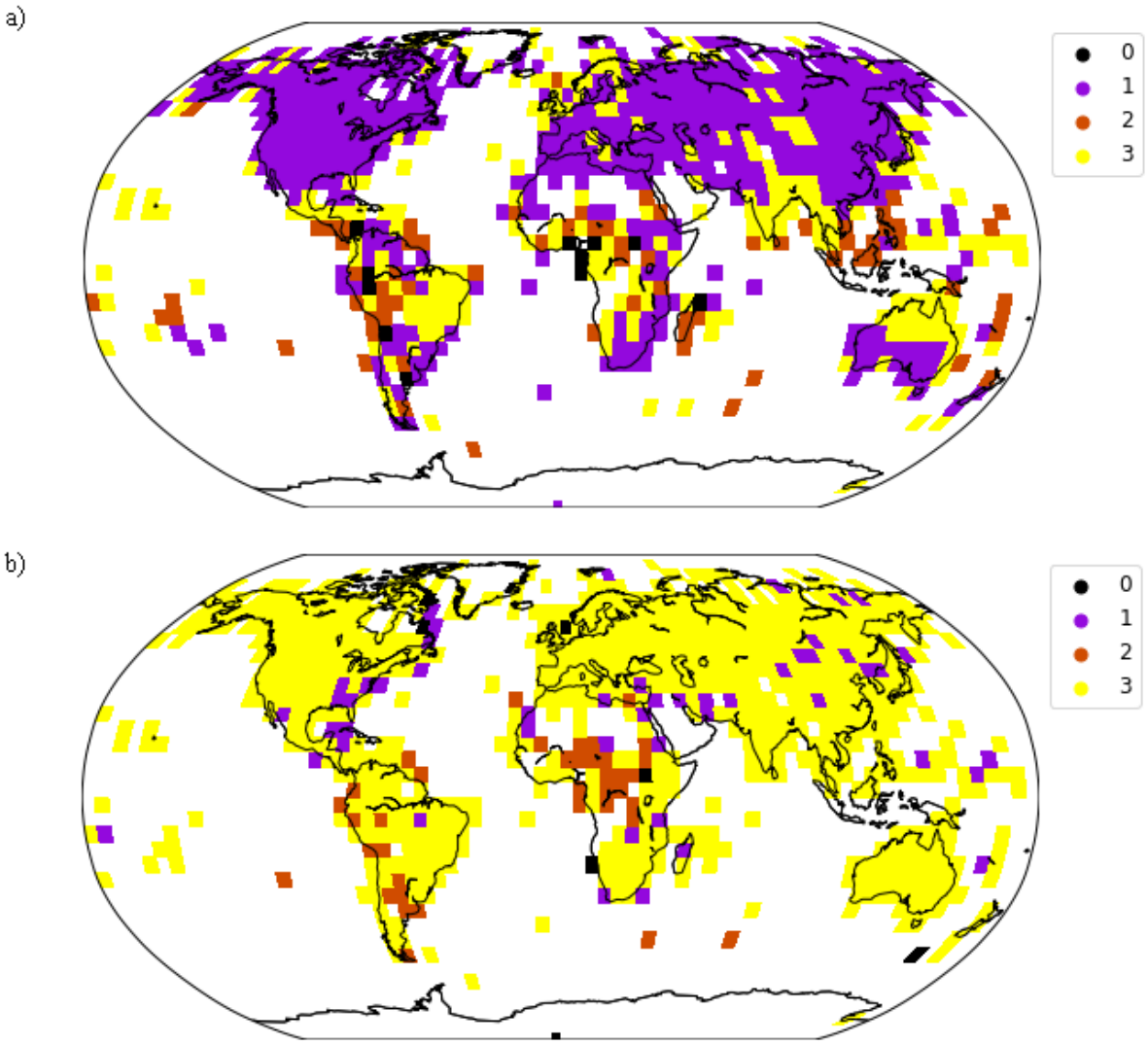


Figure 3.7 – Gridded model-order choices using spell length distribution analysis for (a) wet spells and (b) dry spells.

Spatial variations are once again important for model-order performance at reproducing both wet- and dry-spell length distributions (Figure 3.7), though not as prominently as seen earlier with BIC. The northern hemisphere extratropics are again almost universally best described by the same model-order (first for wet spells, third for dry spells). However, unlike BIC and dry spells, there is notable variation across Europe in model-order performance for wet spells (where third order appears prominent in the north-west, but first order elsewhere). This variation in model order across Europe contributes to third-order models outperforming other orders across a sizable minority of stations in temperate regimes (33.7%). However, it is across the tropics and southern hemisphere where most variation is present, primarily in tropical and dry regimes. Despite third- and first- order models

representing wet spells best in tropical and dry regimes respectively, a majority of tropical grid cells are represented best by an order other than third and a sizable minority (33.2%) of dry regime grid cells are represented by a model order other than first. While third-order models almost entirely outperform others across Brazil, Northern Australia and India, there is much more variation across Central Africa, despite being in the same regime (tropical).

The wet-spell distribution across the Aw (tropical with dry winters) regime (Figure 3.2), present across much of Brazil and Central Africa, is reproduced best by a third-order model in 61.1% of cells. This is higher than the overall 47.8% of tropical grid cells represented by a third-order model. Dry, hot-arid steppe (BSh) regions, present to the north of Central Africa, and across Australia and India, also often show third-order dependence (51.4%) despite first-order models most commonly outperforming others in dry regime grid cells. Sub-regimes for temperate, continental and polar climates follow similar patterns to their aggregated classifications for both wet- and dry-spell distributions.

For both wet- and dry-spell distributions, there is notably less variation in model-order performance across continental regimes than other classifications. It is important to note that continental climates (category D, termed "Snow", in Figure 3.2) are almost exclusively present in the northern hemisphere extra-tropics, with each other regime present across the tropics and in the southern hemisphere. This potentially explains the greater variation in the other regimes.

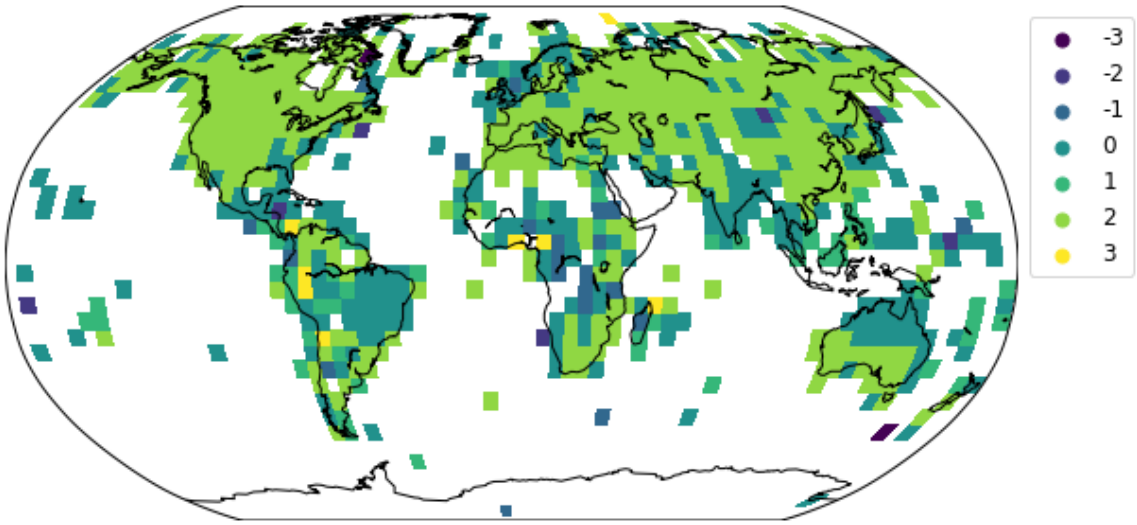


Figure 3.8 – Model-order choice for dry spells minus the model-order choice for wet spells.

The difference between model-order performance at reproducing wet-spell and dry-spell distributions was studied (Figure 3.8) by subtracting the model-order choice for wet-spell distributions from the choice for dry-spell distributions. For example, should a grid cell be best represented by a third-order model for dry spells and a first-order model for wet spells, the difference is +3. A difference of +2 is widespread across most of the northern hemisphere. This is as expected, with first-order models commonly outperforming all others for wet spells and third order for dry spells. It is important to note that for many locations, including Brazil, Central America, North Australia, India, Europe and Central Africa, the same model-order performs best for both wet and dry spells (a difference of 0, Figure 3.8).

3.1.4.3 INTERANNUAL VARIABILITY OF PRECIPITATION OCCURRENCE

In this section we evaluate the performance of the models at generating year-to-year variability in the number of wet days (i.e., precipitation occurrence). In all regimes, second-order models most commonly reproduce the IVO, with the exception of tropical regime cells, where third order most commonly outperforms others (Table 3.3). In all regimes (except tropical), the IVO in at least 50% of cells is represented best by a second order model. Although third-order models perform best in tropical regimes (49.4%), the number of cells represented best by a second-order models follows closely (41%). As with BIC and spell-length analysis, zeroth-order models rarely (or never) outperform others.

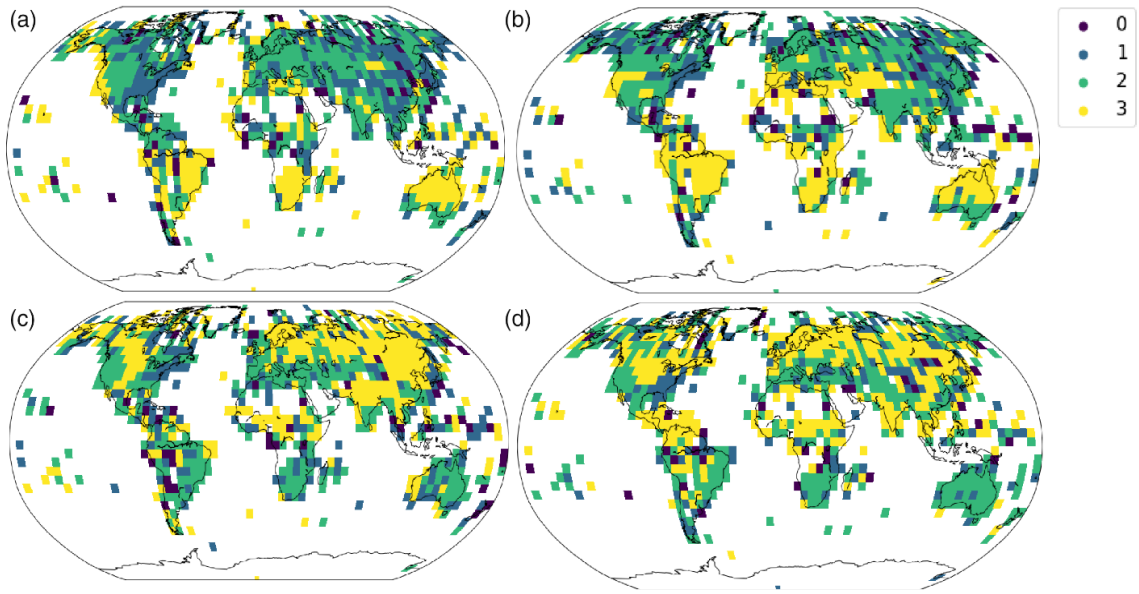


Figure 3.9 – Model-order choices for each 5°x5° grid cell, with colour representing model-order choice as determined by interannual variability of seasonal precipitation occurrence for (a) March, April, May (MAM), (b) June, July, August (JJA), (c) September, October, November (SON) and (d) December, January, February (DJF).

There are some notable differences in model-order performance at reproducing the interannual variability of precipitation occurrence for individual seasons across the regimes. Much like annual IVO, second-order models most commonly outperform others, with zeroth order rarely performing best (though with larger minorities than the annual occurrence). For all seasons in tropical regimes, third-order models consistently outperform others. While a small minority of cells are best represented by a first-order model annually (25.6% maximum, found in continental regimes), in JJA, first-order models outperform third-order models in both continental and polar regimes (by 27% and 34.1% respectively).

There is less coherence in the spatial pattern of annual model-order choice for reproducing the IVO (not shown here) compared with BIC and spell-length analysis (Figure 3.5 and Figure 3.7). Nevertheless, the individual seasons do have coherent patterns in the best-performing model order (Figure 3.9), but these patterns are somewhat opposite between seasons resulting in some cancellation of coherent patterns for the annual results. There is a stark contrast in the performance of third-order models at reproducing the seasonal IVO between DJF and JJA across the continental regime. In DJF, third-order models outperform others in 54.4% of continental cells, whereas in JJA this drops to only 6.8%. This is reflected in Figure 3.9. In JJA and MAM much of the northern hemisphere is represented best by second- and first-order models, whereas in DJF and SON, many of these cells in are instead

represented by a third-order model. These cells are mostly continental and temperate. The reverse is true for South America, where third orders are prevalent in JJA and MAM and second orders in DJF and SON. In JJA, third-order models outperform all others in both tropical and dry regimes (48.3 % and 47.1% respectively), with a sizable minority of second-order cells. However, in DJF, third-order models perform best in only 27.5% of cells, with an increased percentage of second-order cells (49.3%). Further seasonal differences can be noted across Australia, India and southern Africa.

It is widely known that a common limitation of Markov chain models is in their underestimation of the observed interannual variability (Wilks, 2010; Wilks and Wilby, 1999). Here, we find that all model orders tend to underestimate the observed variability, with zeroth-order models consistently underestimating more than higher-order models. For each model order, this underestimation is similar across all regimes and tends to decrease as model order increases. However, the improvement stops at second order, and here we find little difference between second and third orders. This agrees with previous findings that lower order models may be more prone to overdispersion (Katz and Parlange, 1998). The IVO tended to be underestimated by a greater percentage by all model orders in locations that had higher observed variability (China, South America, Pacific Islands) and less so in areas with lower observed variability (Australia, India, West USA).

3.1.5 SUMMARY

It is apparent that although each metric highlights different behaviour across the regimes, with stark variation between tropical/dry and continental regimes, the model order with the highest percentage of grid cells is the same across all regimes except tropical.

Table 3.5 – Model-orders for each metric with highest percentage of grid cells reproducing the observed data the best.

	BIC	WET SPELL	DRY SPELL	IVO
TROPICAL	1	3	3	3
DRY	1	1	3	2
TEMPERATE	1	1	3	2
CONTINENTAL	1	1	3	2
POLAR	1	1	3	2

The model order with the highest percentage of grid cells in each regime can be said to perform best (Table 3.5). Bayesian Information Criterion and dry-spell length analysis each favour a single model order globally, despite behavioural differences in each regime. For wet spells, each regime favours a first-order model, except tropical which prescribes third order. The Bayesian Information Criterion was satisfied by a first-order model across all regimes. A third-order model performs best at reproducing dry spells across all climate regimes (especially for the higher percentiles of the dry-spell length distribution), while second order reproduces the IAV of precipitation occurrence best in all regimes except tropical (once again prescribing third order).

Spatial variation was noted using each method of analysis. Much of the northern hemisphere contain cells with a high percentage following the overall metric's model order prescription, regardless of their climate classification. It is across the tropics and southern hemisphere (South America, Central Africa and Mainland Southeast Asia) that shows the most variation from the model order for each method of analysis. These locations are often categorised into tropical and dry climates.

There are several other points of note. Whilst a third-order model best reproduces observed dry-spell distributions, first-order models are chosen as second best across 76.9% of cells. At the 95^{th} and 99^{th} percentile, while third-order models underestimate the observed percentiles the least, first order performs similarly at the lower percentiles. Third- and first-order models perform second best at reproducing wet-spell distributions across 40.0% and 54.4% of cells respectively. Third-order models also require the calculation of eight independent transition probabilities (per month) while a first-order model calculates only

two, thus the higher order model requires more computing power and, most importantly, increases the possibility of over-fitting a model. Although second-order models most commonly outperform others at reproducing the IAV of precipitation occurrence, first- and third-order models represent a sizable minority of stations in all regimes. It is important to note that while using a higher order model reduces the underestimation of IAV, all orders underestimate it, with little difference between second- and third-order models.

There is a noticeable dip in how frequently second-order models perform best at reproducing wet- and dry-spell distributions (Table 3.3). Here, unlike BIC (Table 3.2), model-order performance is not penalised for the number of parameters used. A potential cause of this dip could be that the percentage of third-order models is boosted by also being representative of the performance of model orders higher than three, which are not considered here. For example, when reproducing wet-spell distributions, there may be decaying performance across higher model orders which, in this study, might be encapsulated by the percentage of third-order grid cells.

However, there are limitations due to data coverage across much of northern and southwest Africa, the Middle East, Indonesia and the Philippines (Figures 3.5 to 3.9). Much of this domain falls into tropical and dry regimes, potentially impacting the overall model-order choice of these climate types. Aggregating data into grid cells also loses some of the detail from sub-categorising the weather stations. In this analysis, continental regimes are almost completely represented by Dfb and Dfc climates (i.e., those with continental humid climates; Figure 3.2), with 78.6% of continental cells falling into one of these sub-categories. There is no representation of Dsa or Dsd (continental with dry summers) climates using the gridded approach at 5° resolution. Thus, the overall model-order choice for continental climates may not be representative of all sub-categories within the regime. A similar issue is noted across temperate and polar regimes. Only one grid-cell represents Csc and Cwc cells (climates with temperate, cool summers with dry summers and winters respectively), whilst over half of the temperate cells represent Cfa and Cfc (temperate, fully humid climates).

3.1.6 CONCLUSION

The ability of four Markov chain weather generator models to generate realistic daily precipitation time series was assessed by four different methods across 44071 weather stations globally. Each weather station was allocated a regime based on Köppen's climate classification system (Chen and Chen, 2013): tropical, dry, temperate, continental or polar.

To extend previous results, the performance of the different order models was assessed using four metrics: the widely used Bayesian Information Criterion, and the less commonly considered abilities to reproduce observed wet- and dry-spell lengths and the IAV in precipitation occurrence.

Analysis was undertaken for each weather station, and then performance was aggregated on a grid-cell basis to avoid the assessment being dominated by densely observed regions. Model performance measures were aggregated into 837 5°x5° cells based on the longitudes and latitudes of the weather stations. Depending on which metric was used, different model-orders performed the best. First-order models most frequently minimised BIC and best reproduced observed wet-spell distributions across each regime except tropical. Whilst this agrees with other preceding local or regional studies, including across the USA (Schoof and Pryor, 2008), Nigeria (Jimoh and Webster, 1996) and Costa Rica (Harrison and Waylen, 2000), this finding has now been demonstrated on a near-global scale, and included a range of different climate regimes. Unlike other regimes, third-order models best reproduced observed wet-spell distributions across tropical regime locations. Third-order models also best reproduced dry-spell distributions across all regimes, strengthening previous evidence that a low-order model may not accurately reproduce extreme dry periods (Lennartsson et al., 2008). This is strengthened further by the ability of third-order models to reproduce most successfully the observed 99th percentile (and, in most cases, the 95th percentile) of dry-spell length distributions. Second-order models most commonly reproduced the IAV in precipitation occurrence best in each regime, once again with the exception of tropical regimes where third orders perform best.

Although the most frequently selected model order remained the same across different climate regimes for each metric (except for tropical regimes), interesting variations within and between regimes were found. Tropical and dry regimes showed the most deviation from overall behaviour. Third-order models reproduced both wet- and dry-spell distributions best across Brazil, India, North Australia and Europe, regardless of the climatic region. Central Africa, Mainland Southeast Asia and western South America also showed notable geographical variation of model-order performance for each metric. The behaviour of climate classification sub-categories often reflected the aggregated performance but there were some exceptions primarily across dry and tropical zones. However, many sub-categories were not present or were only represented by a few grid cells in this study, resulting in temperate and continental regimes being more representative of some sub-

categories than others. This leads to the notion that analysis of model order selection may be more beneficial on a finer spatial scale, i.e., sub-continental, or with greater inclusion of other sub-categories. This may better reflect the noted spatial variance.

This study has informed the next chapters on developing a global stochastic weather generator to synthesise daily time series of a suite of climatic variables, including maximum daily temperature and cloud cover. There are also several ways in which these results can be applied to industry and future research. As weather variables produced by a stochastic generator are frequently used as input into hydrological and agricultural models, our results show that it would be beneficial to use different model-orders to generate precipitation data depending on the purpose of the model. For example, studies focussing on extended dry periods may favour third-order models due to their superior ability to reproduce the upper tail of the dry-spell distribution, whereas a first-order model acts as a good, computationally efficient "all-rounder" for other studies. Tropical climates are an exception, where third-order models reproduce distributions of both wet and dry spells and the IAV of precipitation occurrence best overall. This study can also inform users on the most appropriate model-order to use based on the location of a specific region. For example, generating realistic wet-spell distributions across Europe requires different model-orders to those favoured in the eastern USA, despite both regions falling into the temperate classification (Figures 3.2 and 3.7a).

3.2 SUPPLEMENTARY MATERIAL

Analysis not included in Wilson Kemsley et al. (2021) have also been performed to better understand model-order performance across the globe. The main section of this chapter focussed on the performance at generating precipitation *occurrence*, whereas these additional tests evaluate the performance also at generating precipitation *amounts*. The amount of precipitation is an important metric to accurately reproduce, as it has implications on extremes such as flooding and drought. The metrics used here (annual maximum 5-day precipitation and interannual variability of annual or seasonal mean precipitation) depend upon both the generation of dry-wet sequences (i.e., by the order of the Markov chain) and upon the generation of amounts on wet days (drawn from the gamma distribution fitted to the observed wet-day amounts). Therefore, the outcome is dependant not only on the order of the Markov chain but also on the suitability of the gamma distribution to represent the precipitation amounts.

3.2.1 MEAN ANNUAL MAXIMUM 5-DAYPRECIPITATION

Each models' ability to reproduce the mean annual maximum 5-day precipitation was tested. Similarly to Wilson Kemsley et al. (2021), 50-years of daily precipitation occurrence was generated. On days with precipitation >0.1mm, an amount of precipitation is randomly sampled from a gamma distribution that has been fit separately for each month to the observed wet-day precipitation amounts. The mean values of each (overlapping) 5-day period were computed, and then the maximum 5-day mean during each calendar year was selected. These 50 annual maxima were averaged. These values, for each model order, were compared to the mean annual maximum 5-day precipitation of the observed record. The absolute difference between the generated and observed values was taken. The model order producing the smallest absolute difference was deemed to perform best.

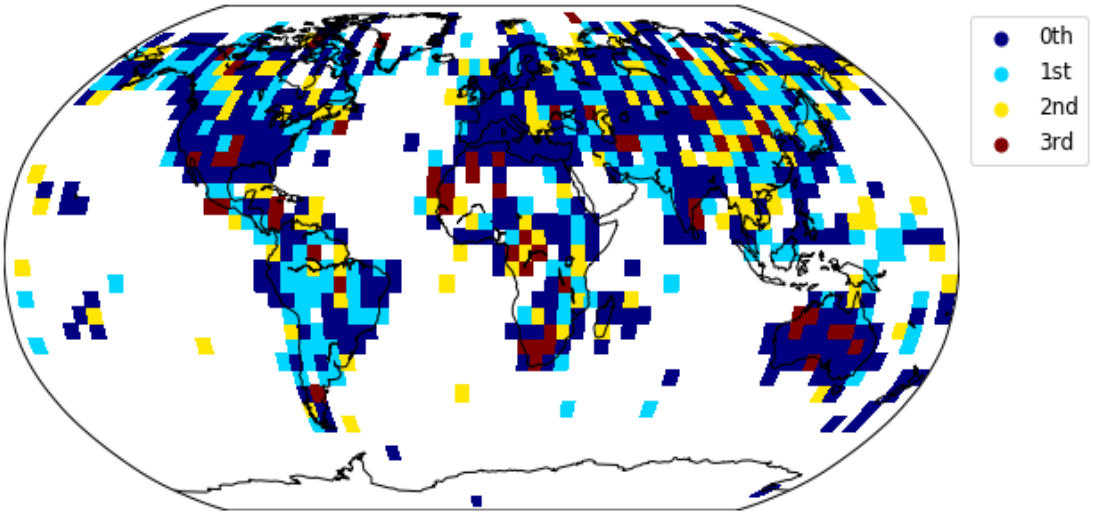


Figure 3.10 – Model-order choice for performance at reproducing annual maximum mean 5-day precipitation.

Table 3.6 – Gridded comparison of Markov model-order choices for each of the Köppen climate regimes based on each model's ability to reproduce the observed annual maximum mean 5-day precipitation. Values shown are the % of grid cells within each climate regime where the mode of individual stations' best model-order is equal to 0, 1, 2 or 3.

	MODEL ORDER			
	0	**1**	**2**	**3**
Tropical	34.4	28.7	25.2	11.6
Dry	30.6	21.8	17.5	30.2
Temp.	35.0	29.2	23.0	13.8
Cont.	34.3	31.2	25.9	8.6
Polar	40.3	31.2	21.8	6.7

A zeroth-order model was found to reproduce the annual maximum mean 5-day precipitation best across all climatic regimes (Table 3.6). However, unlike previous methods of analysis, differences between model-orders are less clearly distinguished. For each regime, only 34 – 41% of cells are represented best by a zeroth-order model, closely followed by first-order models with between 21% and 32% of cells. Unlike the previous methods, spatial patterns in model-order choice for best representing 5-day maximum precipitation are also less clear (Figure 3.10).

The difference in performance between each model order is also significantly smaller for this analysis than for BIC and wet-/dry-spell length analysis. This is expected because annual maximum 5-day precipitation depends partially on the distribution of wet spells (which is influenced by model order) and partly on the gamma distribution of wet-day rainfall amounts (which is independent of model order). The other performance metrics depend solely on the sequencing of wet and dry days, which is different between model orders.

3.2.2 INTERANNUAL VARIABILITY OF MEAN PRECIPITATION

Table 3.7 – Gridded comparison of Markov model-order choices for each of the Köppen climate regimes based on each model's ability to reproduce the observed interannual variability of mean precipitation amount. Values shown are the % of grid cells within each climate regime where the mode of individual stations' best model-order is equal to 0, 1, 2 or 3.

	MODEL ORDER			
	0	**1**	**2**	**3**
Tropical	12.5	22.6	36.3	28.6
Dry	5.0	23.0	46.8	25.2
Temp.	3.3	15.3	64.4	16.9
Cont.	3.0	30.9	47.9	18.2
Polar	6.5	28.3	39.1	26.0

Section 3.1.4.3 studied the ability of each model order to reproduce the interannual variability in precipitation *occurrence* (IVO). Here, the capabilities of each model order to reproduce the interannual variability in mean precipitation *amount* (IVA) has been assessed using 50-years of generated time series. While it can be expected that the results here complement those in Table 3.3, there are some subtle differences.

Second-order models most commonly reproduce IVA best across each regime (Table 3.7) followed by third-order models in tropical, dry, and temperate regimes, and first order in continental and polar. Third-order models typically reproduce IVO best in tropical regimes, differing from IVA, whereas the highest performing model order remains consistent for each other regime. Reproducing IVA considers occurrence and *amount* of precipitation. This difference in model-order performance could be explained through the fitted monthly gamma distributions. Conversely to Figure 3.9, Figure 3.11 shows that some

of the largest changes in seasonal model-order choice can be seen in South America and sub-Saharan Africa; both largely tropical regions.

In temperate regimes, most cells (64.4%) are best represented by a second-order model. In all other regimes, a second-order model outperforms others in no more than 48% of cells. In dry, temperate, and polar regimes, first- and third-order models perform similarly (23% and 25.2%, 15.3% and 16.9%, and 28.3% and 26.0% respectively). As with BIC and spell-length analysis, zeroth-order models rarely outperform other orders, reproducing the IVA best across 12.5% of tropical cells and far fewer cells in the other regimes. The spatial pattern of model-order performance for reproducing the interannual variability of annual mean precipitation is shown in Figure 3.11and is less spatially coherent than with BIC and spell-length analysis over most land-surface areas.

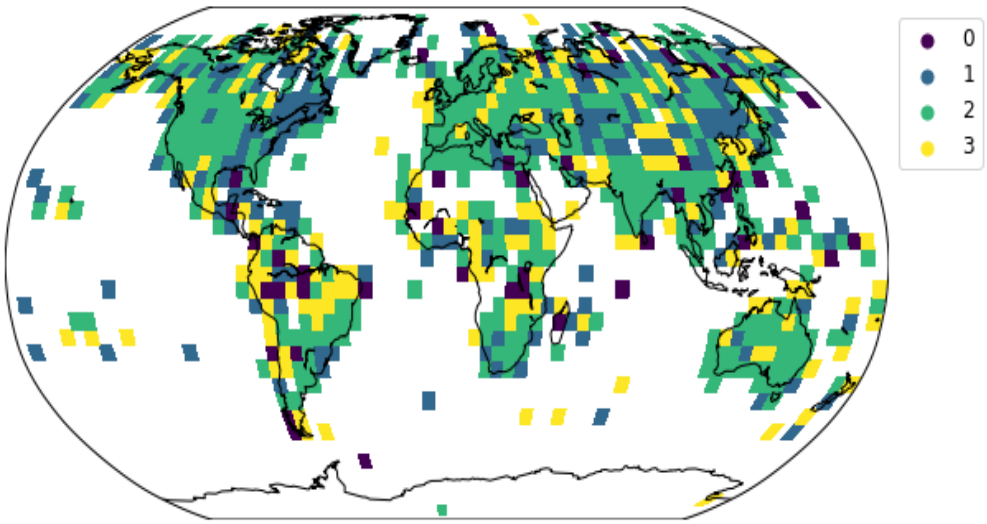


Figure 3.11 – Model-order choices for each 5°x5° grid cell, with colour representing model-order choice as determined by interannual variability comparison.

There are also differences in model-order performance at reproducing the seasonal IVA (Figure 3.12). Much like IVO, second-order models most commonly outperform others, with zeroth orders rarely performing best. Alongside South America and sub-Saharan Africa, the most notable seasonal differences in model-order performance are present across much of the northern hemisphere. In MAM/JJA, much of the northern hemisphere is represented by a second order best, whereas in SON/DJF the preference for third order is more widespread. This is reflected by third-order models outperforming others in many continental cells in SON and DJF (63.1% and 54.2%, respectively), which are primarily present in the northern hemisphere.

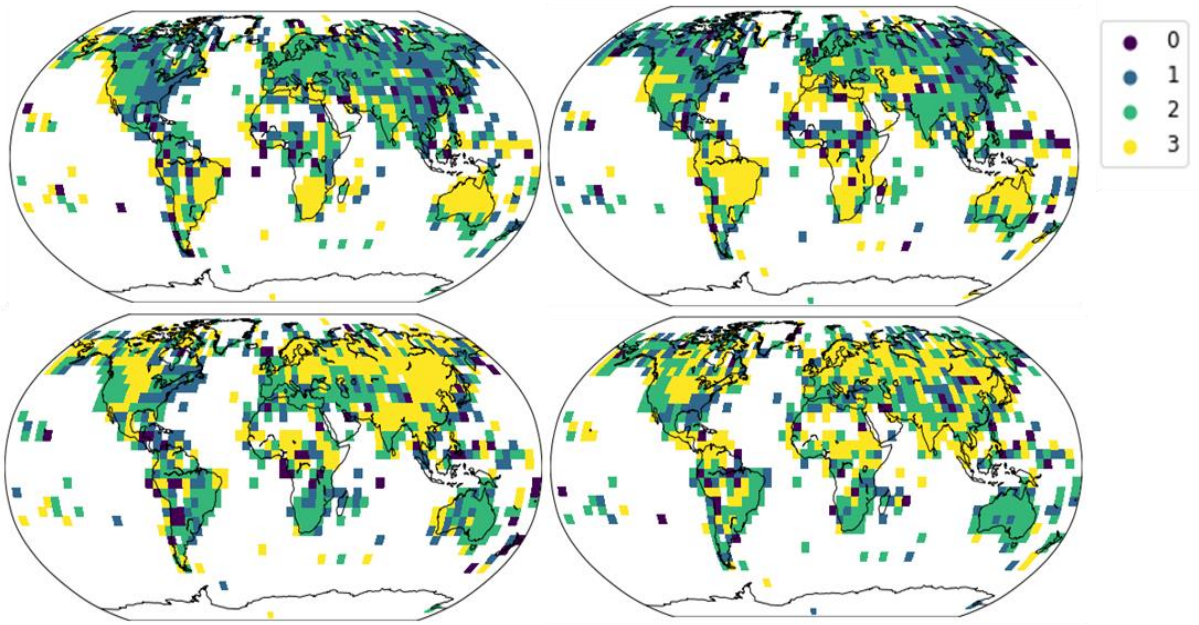


Figure 3.12 – Model-order choices for each 5°x5° grid cell, with colour representing model-order choice as determined by inter-seasonal variability comparison for a) March, April, May (MAM), b) June, July, August (JJA), c) September, October, November (SON) and d) December, January, February (DJF).

The reverse is true for South America, where third orders are prevalent in MAM/JJA, and second orders in SON/DJF. These cells are commonly in tropical or dry regimes. In JJA, third order models outperform all others in tropical and dry regimes (47.6% and 45.3%, respectively) with a sizable minority of second-order cells. Similarly, dry regimes in MAM are best represented by third-order models (42.5%) with similar proportions of tropical cells represented by second (35.7%) and third (33.3%) orders. Further seasonal differences can be noted across Australia and southern Africa.

The magnitude of IVA underestimation varies by model order. Figure 3.13 reiterates the known limitation of Markov-chain models' underestimation of interannual variability (we have shown here both in occurrence and amount). While most points lie just below the generated = observed line, correlation coefficients are less than 0.4 for all model orders studied. The ratio between generated and observed IVA values has also been calculated for each weather station studied in section 3.1.4 (Figure 3.14). Second-order models have the highest proportion of stations that fall into the 0.9-1.04 bin. Zeroth-order models have many stations falling into bins with a ratio of 0.59 and less. For first, second and third orders, the IVA is underestimated by between 36 and 40%. Each model order rarely produces IVAs that exceed observed variability.

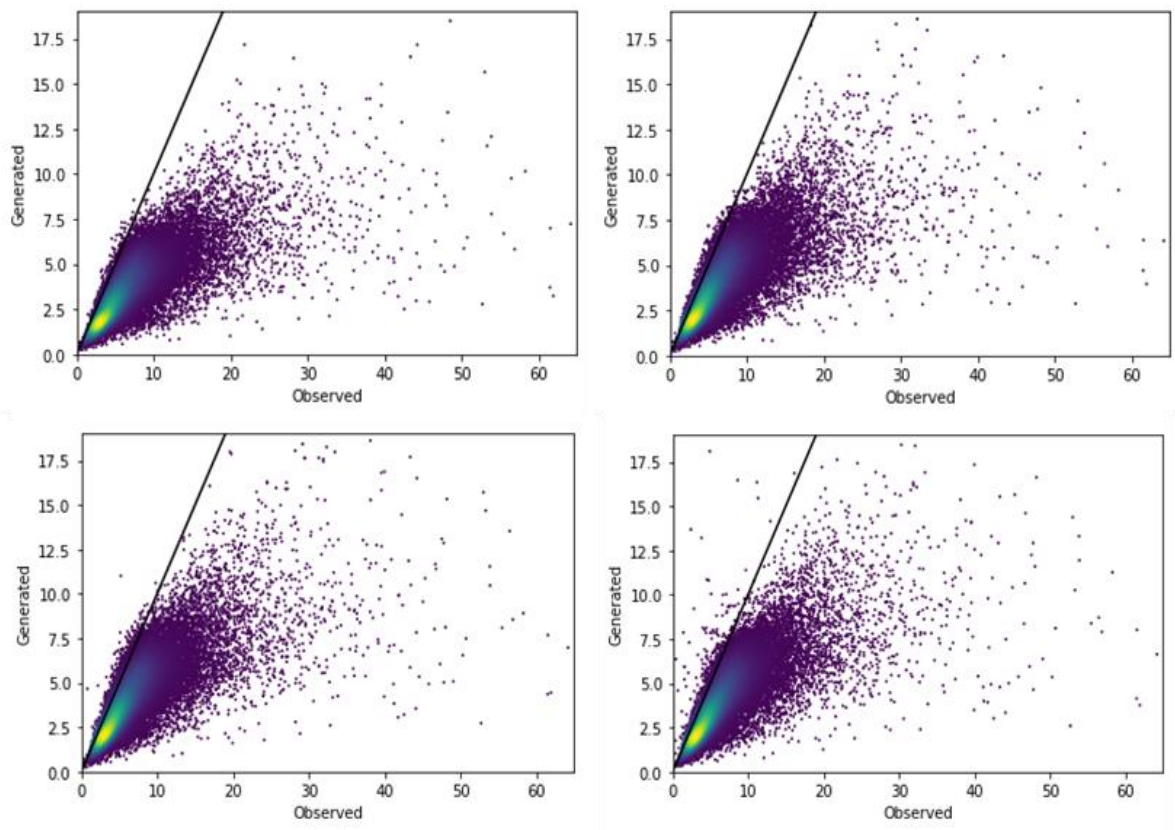


Figure 3.13 – Heat-scatter plots showing the generated against observed interannual variability for each weather station studied using model orders (top left) zero, (top right) one, (bottom left) two and bottom (right) three, with correlation coefficients 0.27, 0.31, 0.32 and 0.33 respectively.

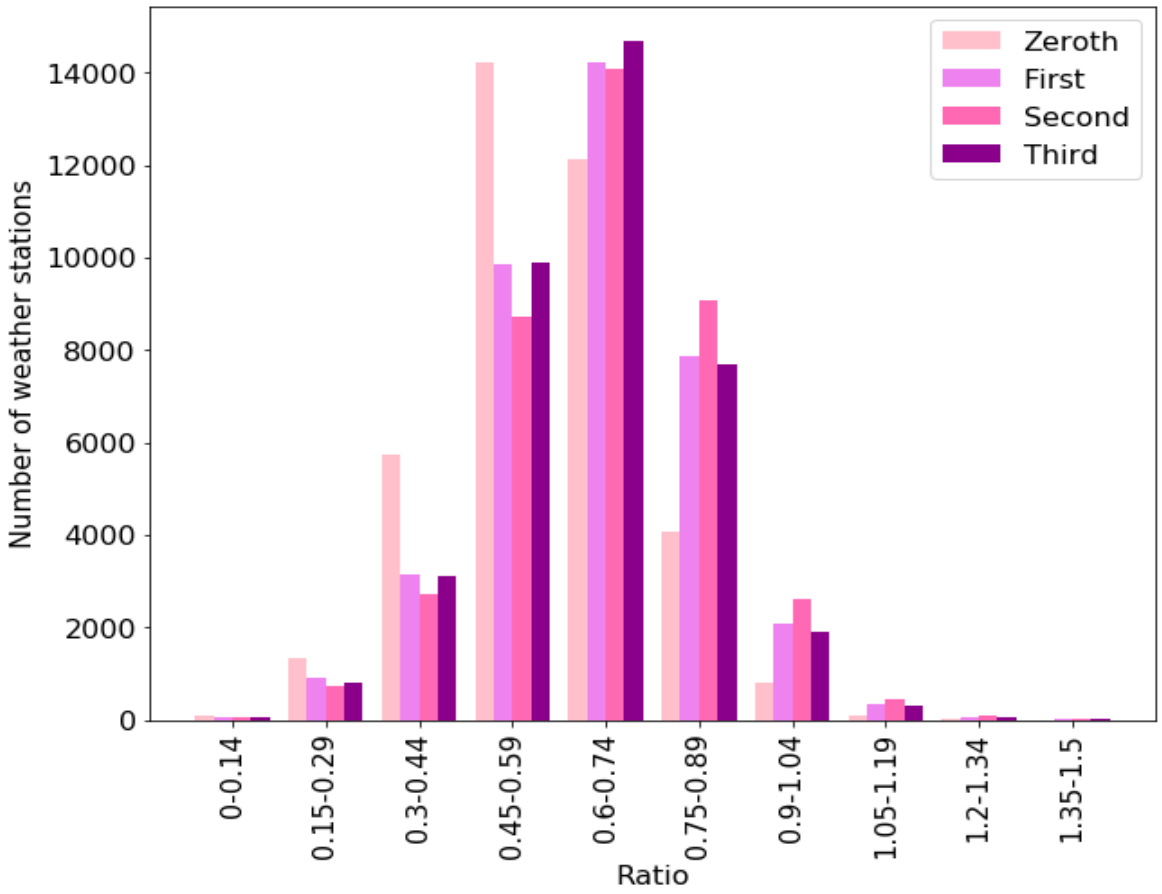


Figure 3.14 - Histogram showing the ratios between observed and generated values in IVA where the ratio is $\frac{V_{gen}}{V_{obs}}$ *where V is the interannual variability of mean precipitation for generated time series using each model order.*

3.3 CHAPTER 3 SUMMARY

It has been shown that different methods of evaluation prescribe different model orders, with exceptions for precipitation in tropical regimes. Following the results presented in Wilson Kemsley et al. (2021) (section 3.1), first-order models are recommended as a suitable "all-rounder". This is due to their computational efficiency, strong performance at reproducing the distribution of wet-spell lengths across all regimes and their similar performance to other model orders at reproducing IVO and IVA. When studying extreme events (for example, hydrological drought and flood) it is important that the distribution of wet and dry spell lengths is accurately captured. If there is greater emphasis on variability and extreme dry-spell lengths, a third-order model is recommended. This chapter provides the buildings blocks for the GCWG in providing a model for generating precipitation occurrence and amount.

Chapters 4 through 7 will build upon the results here, conditioning the secondary (maximum and minimum temperatures) and tertiary (mean wind speed, solar radiation, and relative humidity) on precipitation occurrence and/or amount. The GCWG will primarily consider first-order models unless stated otherwise. This is in part due to their good performance in several metrics, but also due to their computational simplicity. As outlined in Chapter 1, the precipitation input parameters for the GCWG further developed in Chapters 4 and 5 will be perturbed using the pattern scaling technique globally. For precipitation, the perturbed parameters in question will be the monthly first-order Markov probabilities (P_{00} and P_{11}) and wet-day gamma-distribution parameters (α and β).

4 STOCHASTIC WEATHER GENERATOR PRIMARY AND SECONDARY VARIABLE CALIBRATION AND VALIDATION

Following the generation of daily precipitation time series, the **G**lobally **C**alibrated stochastic **W**eather **G**enerator (GCWG) simulates the secondary and tertiary variables. The secondary variables (daily maximum and minimum temperature) are conditioned on the generated precipitation time series. To generate the tertiary variables, the GCWG requires a time series of daily precipitation *and* temperature values. The tertiary variables considered in the GCWG are mean daily wind speed, mean daily solar radiation, and maximum/mean daily relative humidity. These variables have been chosen as they can be used to estimate evapotranspiration. Inaccuracies in the generation of precipitation time series will negatively impact the secondary and tertiary variables. Likewise, inaccuracies in the generation of temperature time series may further negatively impact the simulation of the tertiary variables. This chapter therefore focusses on further validating the Markov-chain gamma-precipitation model outlined in Chapter 3, and the calibration and testing of a multivariate linear regression model at generating daily temperature time series.

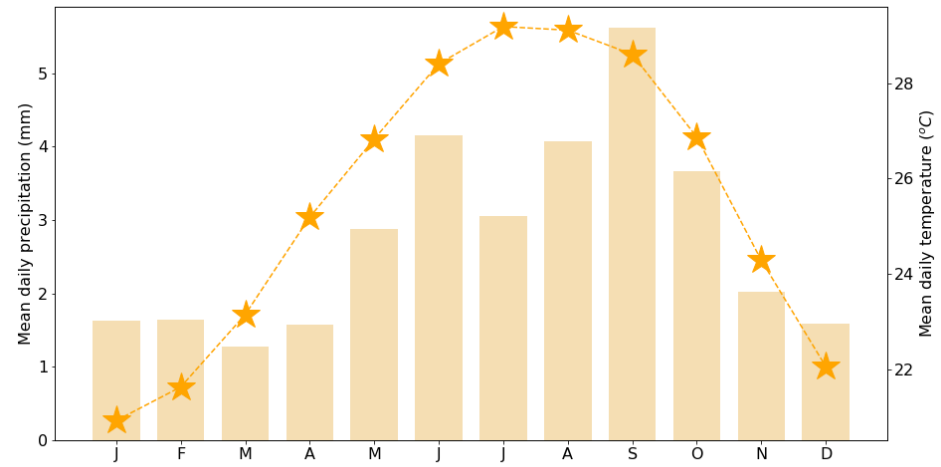
While Chapter 3 widely focussed on model order assessment, this chapter will assess the use of the previously described Markov-chain gamma-distribution model at reproducing observed precipitation at a selection of weather stations using three methods of analysis. Correlations between observed and stochastically simulated precipitation statistics will be studied alongside the Mann-Whitney U test used to compare monthly total precipitation, and extreme value analysis. A multivariate linear regression model, with regressions conditioned on the four precipitation statuses introduced in Chapter 3 (WW, WD, DW, and DD, where DW refers to a dry day preceding a wet day), will be used to stochastically produce daily temperature time series. This model will also be validated in this chapter using statistical testing and comparisons between observed and generated extreme indices. Finally, the accuracy of generated concurrent extremes (in precipitation and temperature) will be assessed.

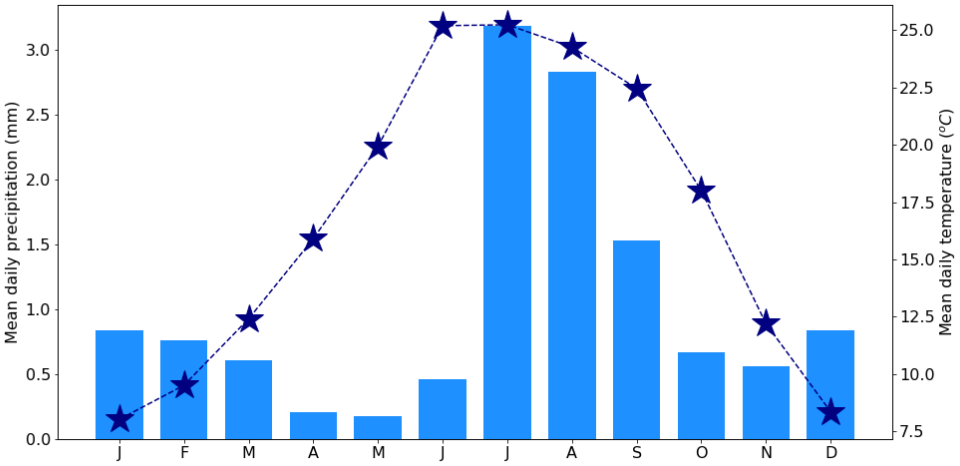
Table 4.1 – Locations and regimes of weather stations used to calibrate the GCWG. Each location contains records for precipitation, maximum and minimum temperature, and mean daily wind speed.

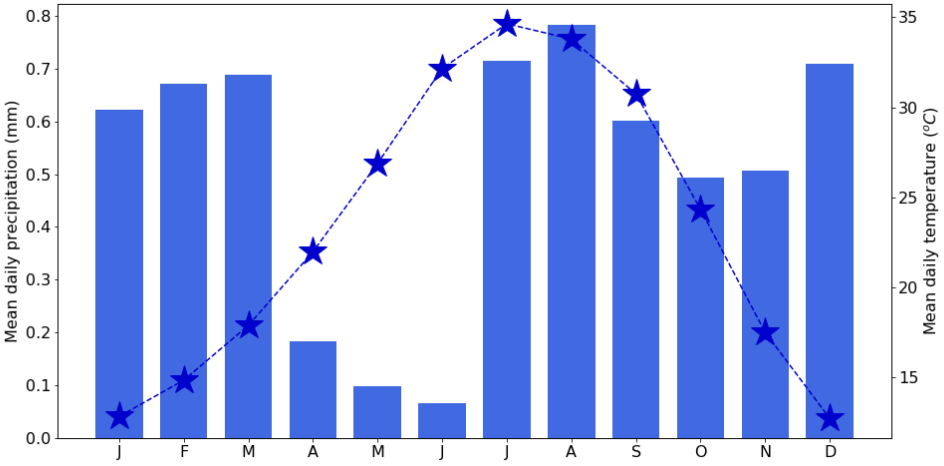
LOCATION	LAT, LON	RECORD LENGTH (YEARS)	REGIME	SOURCE
Key West, Florida	24.6, -81.6	66	Tropical savanna, Aw	NOAA
Fort Huachuca	31.6, -110.3	54	Hot desert, BWh	NOAA
Phoenix, Arizona	33.4, -112.0	62	Hot desert, BWh	NOAA
Reykjavik, Iceland	64.1, -21.8	57	Subpolar oceanic, Cfc	ECAD
Tallinn, Estonia	59.4, 24.7	56	Humid continental, Dfb	ECAD
Aasiaat, Greenland	68.4, -52.8	63	Polar, ET	ECAD

A range of weather stations across the five overarching Köppen climatic regimes (introduced in Chapter 3) have been used to calibrate and validate the combined precipitation and temperature model, shown in Table 4.1 and Figure 4.1. The use of a small sample of weather stations covering a range of climatic regimes is a widely used method to validate the use of a stochastic weather generator (Brunner et al., 2021; Dabhi et al., 2021). Typically, studies have used weather stations used that are geographically close (i.e., in one country or continent), and have not considered a range of globally representative observations. Long, high quality, observational time series with the inclusion of a wide range of climatic variables are limited in spatial coverage, and several locations used to calibrate the model are present in the USA or Europe. However, to ensure the most globally representative model possible whilst working within the constraints of limited spatial data coverage, locations have been chosen to span a range of the climatic regimes discussed in Chapter 3. The reasoning behind this is that a "hot desert" location in the USA ought to have similar precipitation and temperature characteristics as a "hot desert" station elsewhere in the world.

Observed datasets have been accessed from the National Centers for Environmental Information (NOAA) and the European Climate Assessment & Dataset (ECAD) (Klein Tank et al., 2002) projects. In the following sections and chapters, weather stations in the same climatic regime have been used interchangeably, i.e., Phoenix has been used in Section 4.1 and Fort Huachuca in 4.2; both stations are categorised as "hot desert" (BWh) locations and are geographically close (both in Arizona, USA). The weather station used is based on data availability and quality.


Mean daily precipitation (mm)
Mean daily temperature (°C)
0
1
2
3
4
5
22
24
26
28
J
F
M
A
M
J
J
A
S
O
N
D


Mean daily precipitation (mm)
Mean daily temperature (°C)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
J
F
M
A
M
J
J
A
S
O
N
D


Mean daily precipitation (mm)
Mean daily temperature (°C)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
15
20
25
30
35
J
F
M
A
M
J
J
A
S
O
N
D

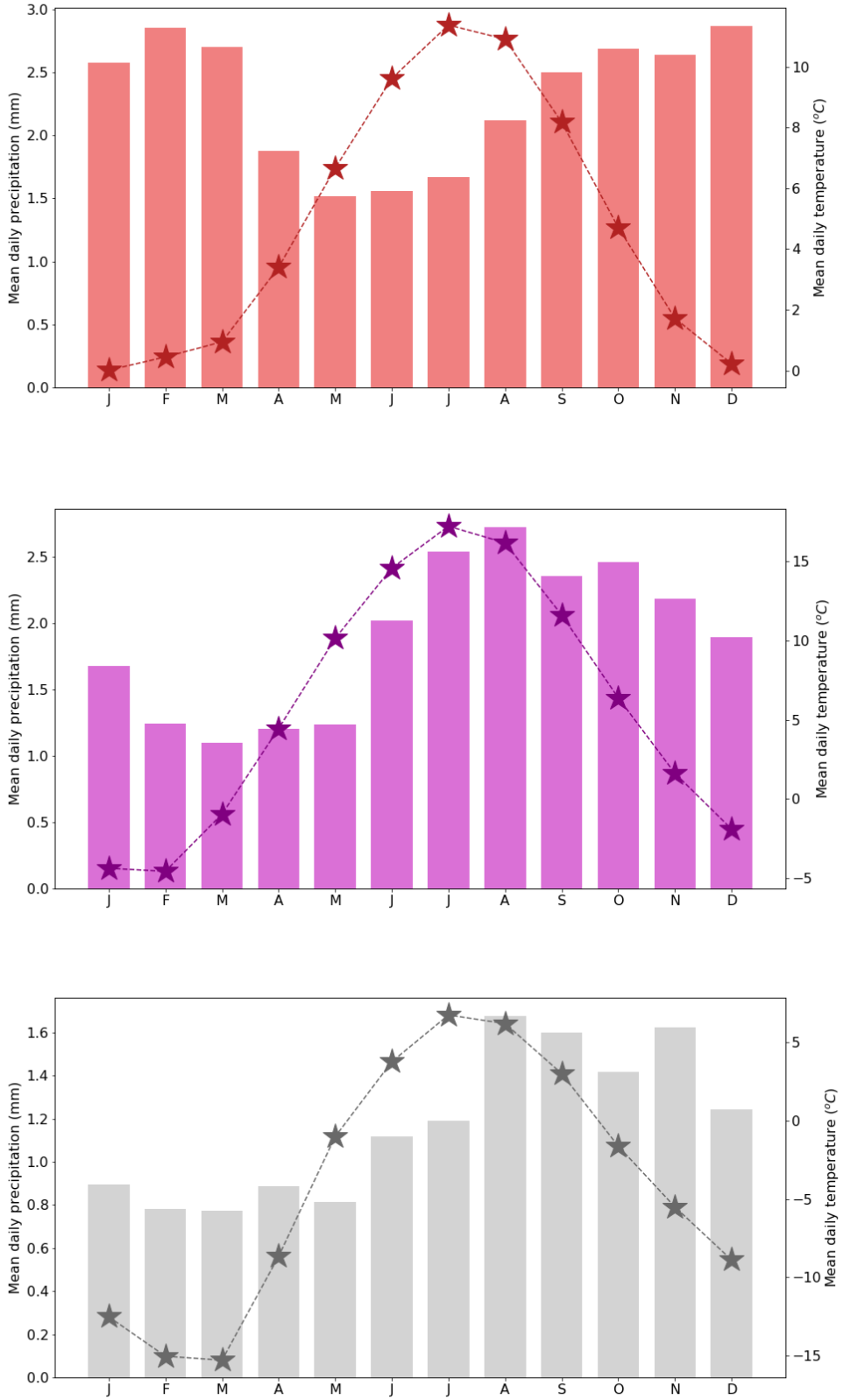


Figure 4.1 – The mean daily temperature (right-side y-axis) and mean daily precipitation (left-side y-axis) by month for the six weather stations outlined in Table 4.1 from top to bottom) Key West, Fort Huachuca, Phoenix, Reykjavik, Tallinn, and Aasiaat. Bars represent mean daily precipitation and stars mean temperature. Note different y-axes on

Figure 4.1 shows the climates in each location used to calibrate the GCWG. As each station lies in the Northern Hemisphere, seasonal cycles in temperature are similar at each location, though with very large differences in magnitude and Key West, lying closest to the tropics and therefore showing a cycle smallest in seasonal changes. Cycles in precipitation show greater differences, with several locations showing greatest rainfall during the warmer months, contrasting with Reykjavik showing greatest precipitation in the cooler months; Key West showing a non-sinusoidal precipitation pattern; and Reykjavik and Tallinn with daily mean precipitation >1mm most of the year.

4.1 FURTHER VALIDATION OF THE STOCHASTIC PRECIPITATION MODEL

The choice of a first-order Markov-chain gamma model for simulating precipitation occurrence in the GCWG has been outlined in Chapter 3. This section discusses the performance of a first-order Markov-chain gamma-distribution model at reproducing observed precipitation at an individual station basis, using observed data from a sample of weather stations presented in Table 4.1. Third order Markov chains have also been used in some instances to provide a comparison, due to their improved performance at reproducing distributions of dry-spell length (Chapter 3, Table 3.3) and for comparison.

In this section, independent training and test datasets have been used for each site, both taken from the corresponding observed record. While this is not widely seen in literature for the validation of a stochastic weather generator, this ensures the statistical testing is robust. Two thirds of the observed record have been used to calculate the precipitation parameters (referred to as the training dataset). In all instances this gives a large enough sample to accurately estimate Markov probabilities (Soltani and Hoogenboom, 2003b). The remaining third of the observed record has been held back and is used as the test dataset for comparison with a stochastically generated 300-year precipitation time series. This provides an unbiased estimate of the skill of the precipitation model. For the rest of this sub-section, “observed” refers to the test dataset (one third of the dataset), while “generated” (or “simulated”) has been produced using parameters derived from the training dataset (two thirds of the dataset). Data has not been split into thirds chronologically. Instead, every third year is held back for use as the test dataset. This is to ensure any trends from increasing global or regional climate change are captured.

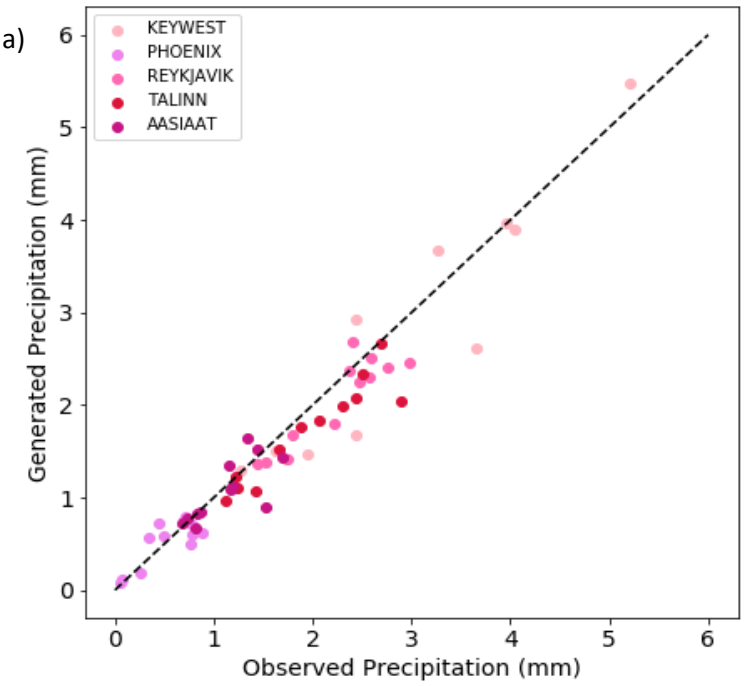
4.1.1 STOCHASTIC WEATHER GENERATOR PERFORMANCE AT REPRODUCING OBSERVED MEAN PRECIPITATION

Scatter plots between various observed and generated precipitation metrics have been produced for the five weather stations (Vu and Mishra, 2020). The means and standard deviations in daily precipitation (averaged over all days, wet *and* dry), the mean precipitation *only* on wet days, and the mean number of wet days have all been calculated monthly for the observed and 300-years of generated data. High correlations between observed and generated data indicate good performance at capturing the annual cycle of the observed data.

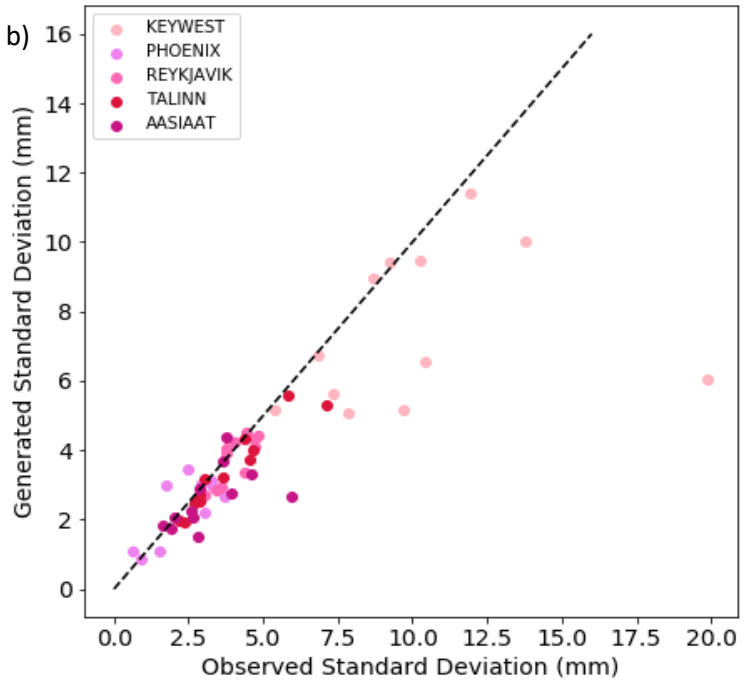
Table 4.2 – Pearson's Correlation Coefficient between simulated and observed values for a range of metrics. The mean daily precipitation (and standard deviation thereof), mean wet-day precipitation, and the mean number of wet days are calculated for each month from the observed and generated datasets.

		PEARSON'S CORRELATION COEFFICIENT (SIMULATED AGAINST OBSERVED)			
Weather Station	**Years in Training Set**	**Mean Wet + Dry-day Precip**	**Std. Wet+ Dry-day Precip**	**Mean Wet-day Precip**	**No. of Wet days**
Key West, Florida	44	0.94	0.25	0.57	0.98
Phoenix, Arizona	42	0.79	0.75	0.71	0.65
Reykjavik, Iceland	38	0.91	0.84	0.95	0.78
Tallinn, Estonia	38	0.90	0.96	0.97	0.99
Aasiaat, Greenland	42	0.73	0.55	0.69	0.95
	Mean	0.85	0.67	0.78	0.87

a)


KEYWEST
PHOENIX
REYKJAVIK
TALINN
AASIAAT
Generated Precipitation (mm)
Observed Precipitation (mm)
0
1
2
3
4
5
6

b)


KEYWEST
PHOENIX
REYKJAVIK
TALINN
AASIAAT
Generated Standard Deviation (mm)
Observed Standard Deviation (mm)
0
2
4
6
8
10
12
14
16
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

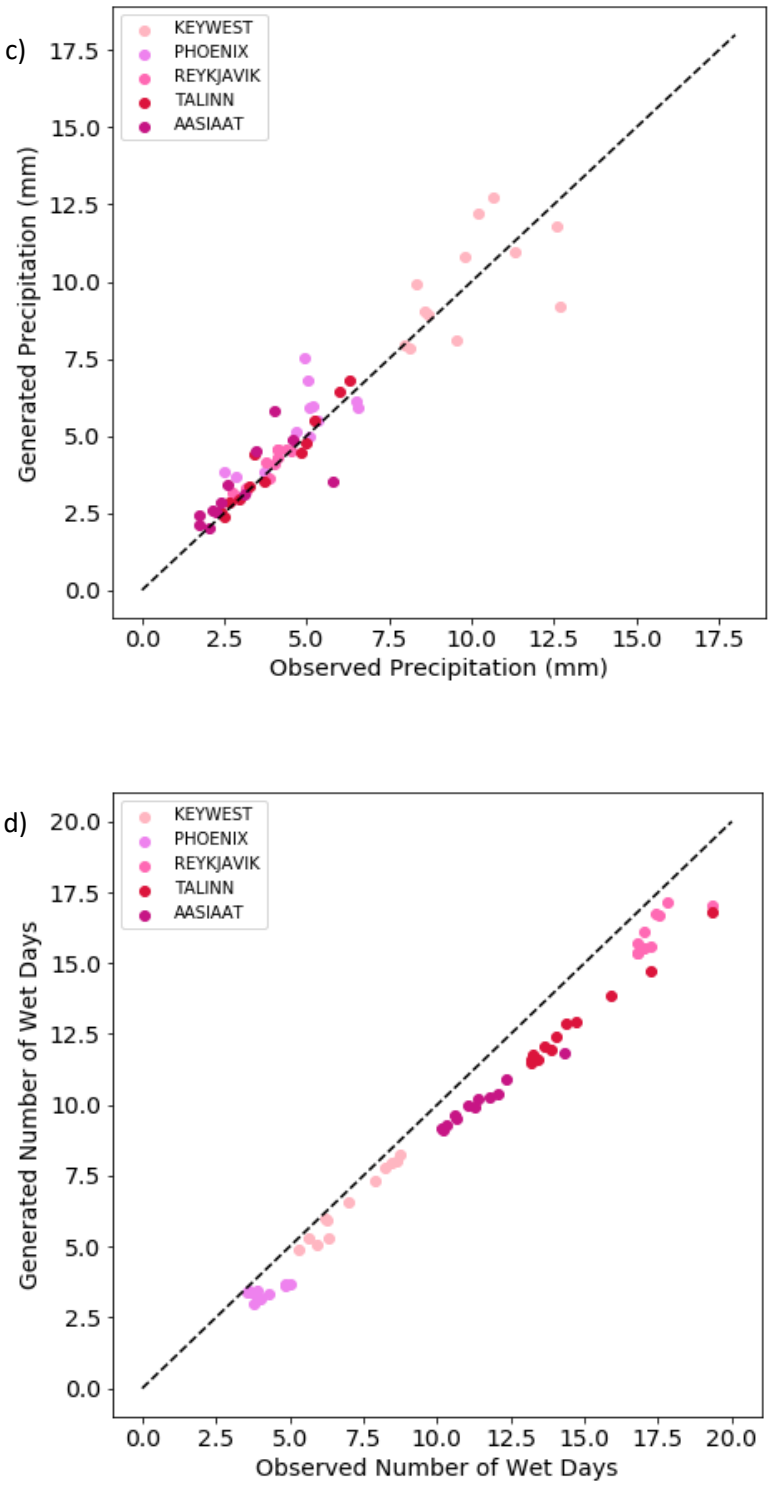


Figure 4.2 – Scatter plots between observed and synthetic precipitation statistics over each station and each month for: a) mean daily precipitation (all days); b) standard deviation in daily precipitation (all days); c) mean daily precipitation (wet days only) and d) the mean number of wet days per month.

For each metric, most locations show strong positive correlations between the observed and generated values (with values shown in Table 4.2). Scatter graphs (Figure 4.2) illustrate not just the correlations but also the magnitudes of these four statistics. Mean daily precipitation (on wet *and* dry days, i.e., all days) is reproduced well, with high correlation coefficients in all locations (>0.7). Standard deviations of daily amounts (on all days) are also well reproduced in Reykjavik, Tallinn, Phoenix, and Aasiaat. The correlation between observed and generated standard deviations in daily precipitation is weaker in Key West, with generated values generally estimating lower variability than observed, and one outlier month (with very high observed standard deviation) weakening the correlation. A lower correlation is also found for the mean wet-day precipitation in Key West (Table 4.2) than at other locations and there is greater scatter for this statistic, shown in Figure 4.2. However, Figure 4.2d shows that the generated number of mean wet days per month in Key West is highly correlated to the observed values (reflected by a correlation coefficient of 0.98). This indicates that it is likely not the occurrence of precipitation that is weakening the performance, but the *amount* of precipitation on wet days, therefore suggesting that the fitting of the gamma distribution is the likely cause of lower correlations in wet-day statistics. Although there are fewer wet days per month in Key West, they tend to have a greater amount of precipitation and greater variability (Figure 4.2c) than locations such as Reykjavik, Tallinn and Aasiaat. There will therefore be a smaller sample size available to calculate the shape and scale parameters despite the heaviest rainfall, perhaps resulting in unrealistic estimations and therefore the smallest correlations.

The highest correlations between observed and generated precipitation are found in the mean wet *and* dry day precipitations alongside the mean number of wet days per month, with mean correlations of 0.85 and 0.87 respectively (averaged over all sites). This is likely due to the good performance of the Markov model at reproducing the occurrence of wet days in comparison to the weaker performance of the gamma distribution at accurately reproducing the daily amounts, alongside smaller samples (in Key West and Phoenix) used to calculate gamma parameters. When a third-order model is used in place of first order, these correlations decrease slightly (a maximum of a 6% decrease) in all locations except Key West, where there is a 1% increase, indicating little difference between the two model orders at reproducing daily mean statistics at these sites. This supports the conclusions found in Chapter 3, whereby a first-order model has been chosen due to its all-round high performance alongside its computational efficiency.

If the test and training datasets are the same, utilising the full observed record to calculate GCWG input parameters (resulting in between 58 and 64 years of data), correlations between monthly statistics are very high for all metrics, largely increasing (where possible) on those shown in Table 4.2. This highlights the requirement of long, high-quality data, and issues that may arise in locations with less consistent precipitation, though it is important to note that testing generated data against a training dataset will of course improve validation metrics alongside decreasing the robustness of the comparison.

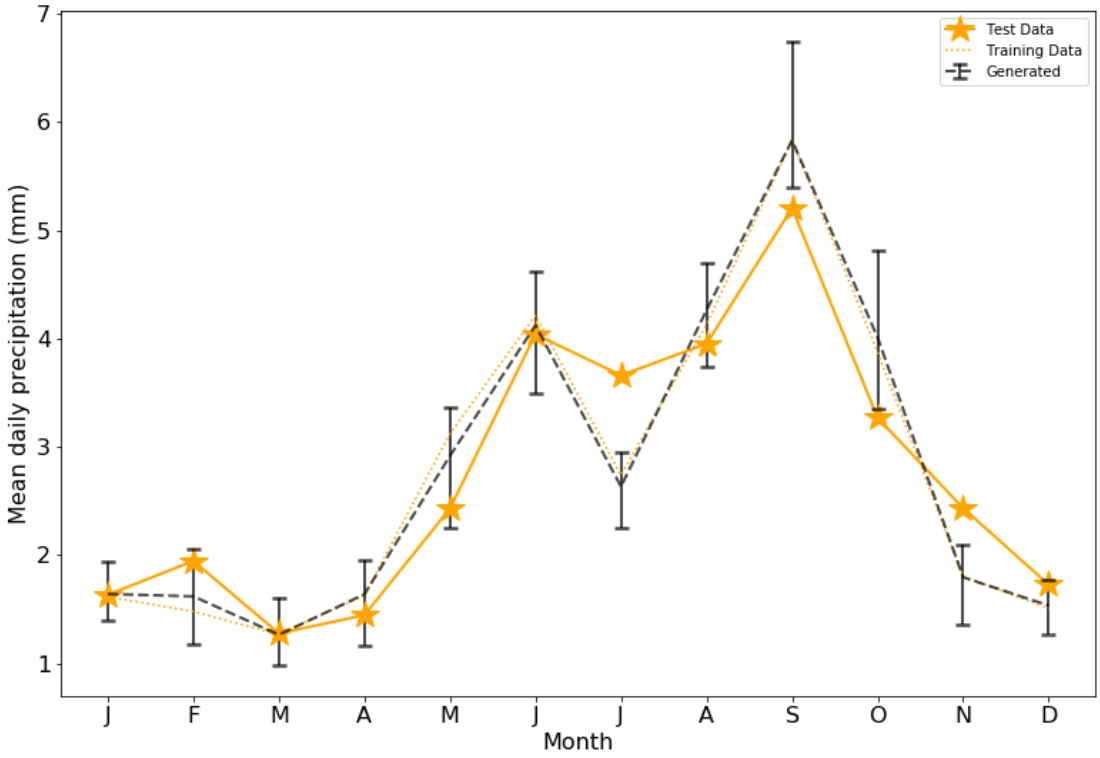
Instead of one 300-year generated time series, ten 30-year periods have been simulated (i.e., ten realisations of the present-day climate at each site). As before, training and test datasets differ, with a larger training than test dataset (in the same 2:1 ratio). The monthly means in daily precipitation are calculated for the observed (test) dataset and generated datasets. It can be expected that, if the observed dataset and ten generated datasets are realisations of the same climate system, observed monthly means will generally lie within the range of generated monthly means. Data has been produced using first- and third-order models for comparison, based on results presented in Chapter 3, wherein a first-order model was preferred for several metrics, though long dry-spells and tropical regimes widely prescribed a third-order model.

Table 4.3 – The number of months where the observed mean daily precipitation lies between the maximum and minimum generated values.

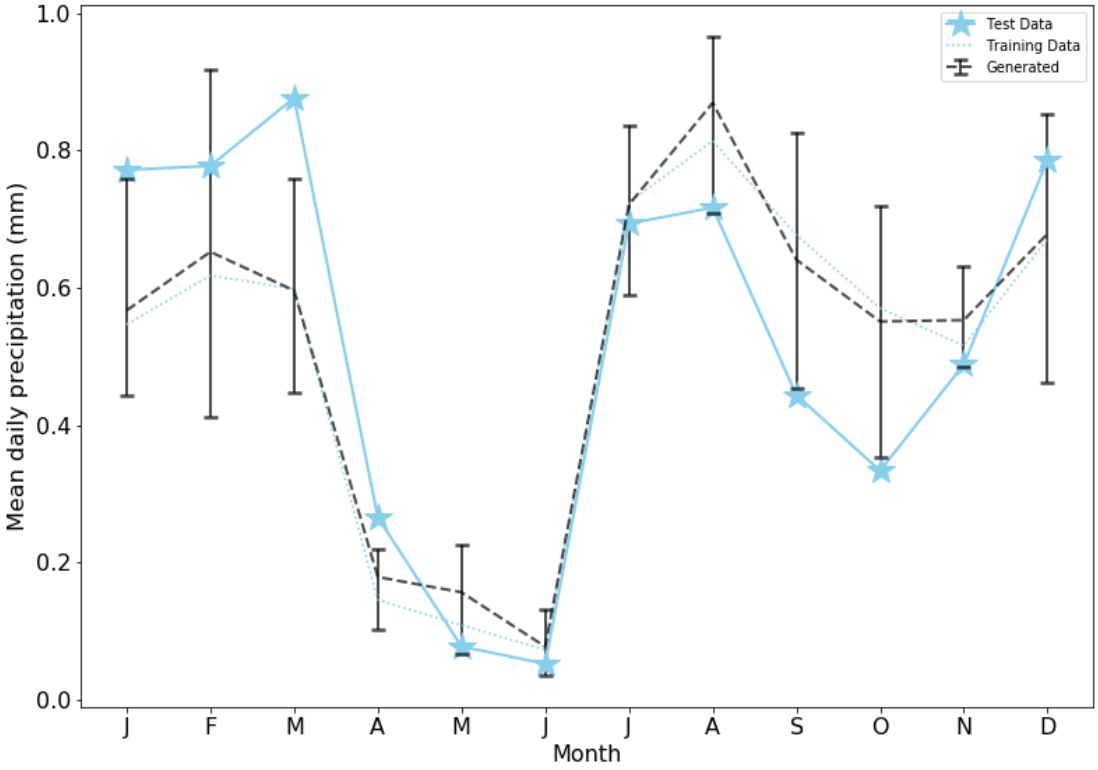
LOCATION	1ST ORDER	3RD ORDER
Key West	9	9
Phoenix	8	6
Reykjavik	10	7
Tallinn	10	6
Aasiaat	8	6

Using a first-order model, performance in all locations is generally very good, with the observed mean falling between the upper and lower generated means in at least 7 months at all locations (Table 4.3). First orders outperform third orders in all locations except Key West, where model orders perform equally. This further supports the findings presented in Chapter 3, wherein a first-order model was recommended as an appropriate all-rounder for all climatic regimes. Categorised as tropical, it is not all that surprising that first and third orders show similar results in Key West, where several metrics showed better performance with a third-order model than in other regimes (Chapter 3, Table 3.5).

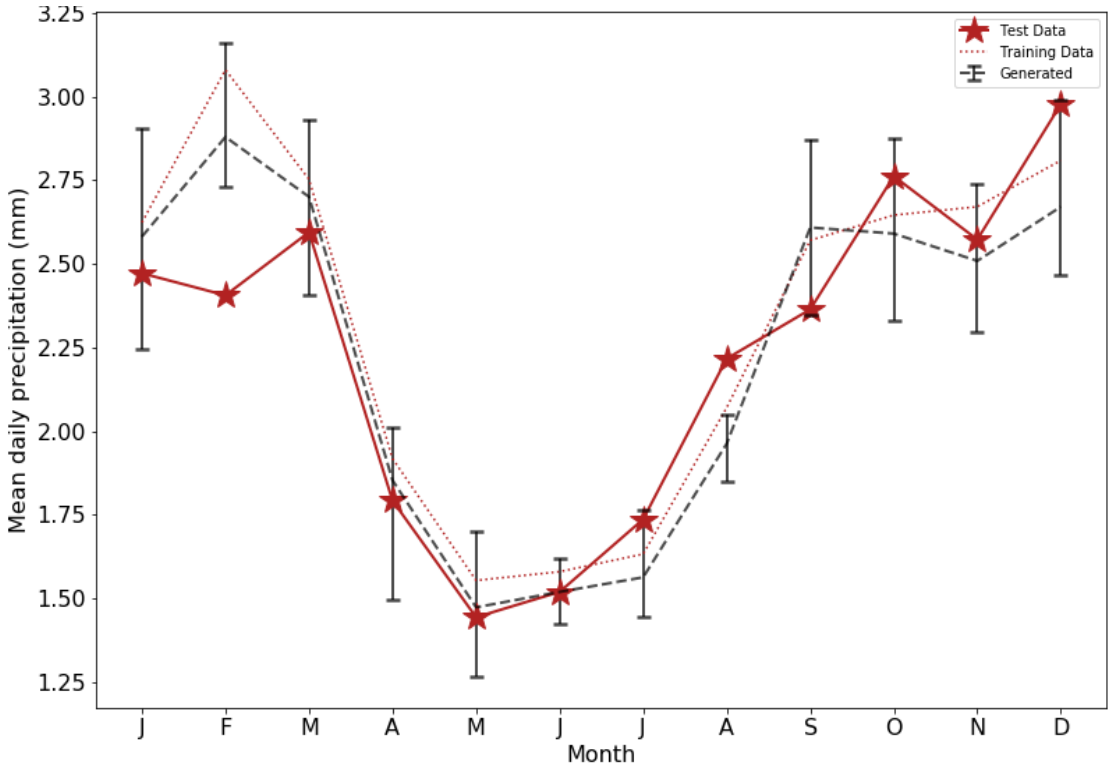
a)



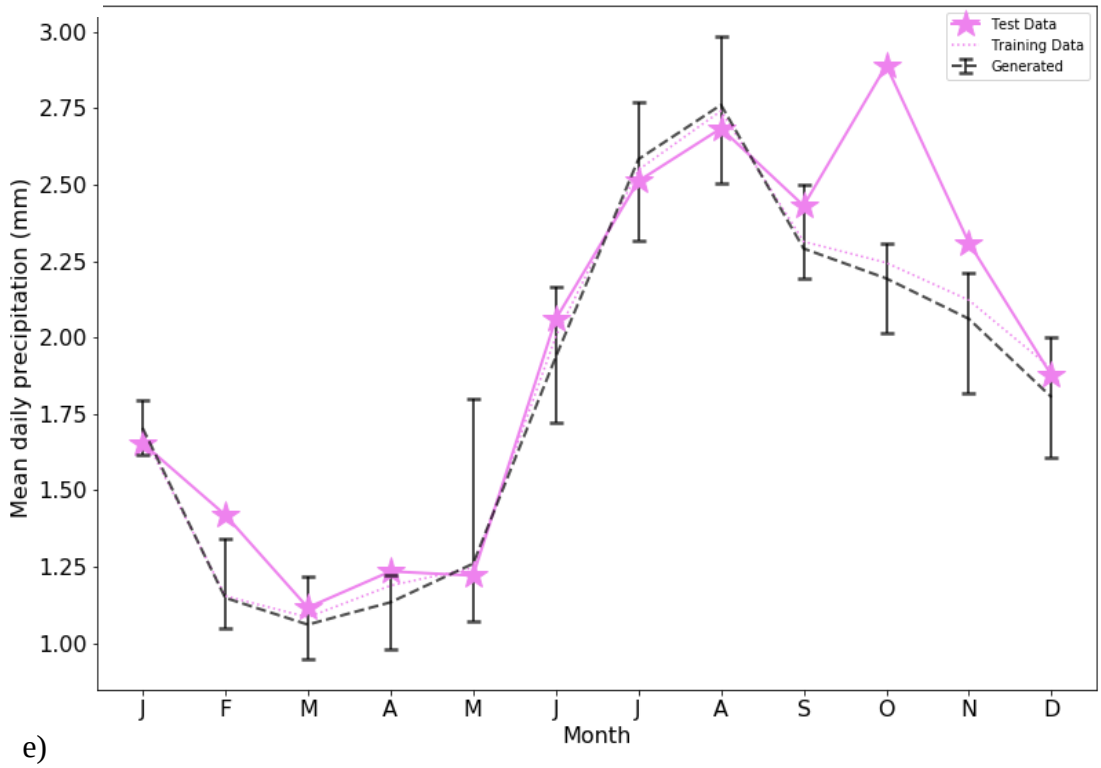
b)


Test Data
Training Data
Generated
Mean daily precipitation (mm)
1.0
0.8
0.6
0.4
0.2
0.0
J
F
M
A
M
J
J
A
S
O
N
D
Month

c)


Test Data
Training Data
Generated
Mean daily precipitation (mm)
3.25
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
J
F
M
A
M
J
J
A
S
O
N
D
Month

d)



e)

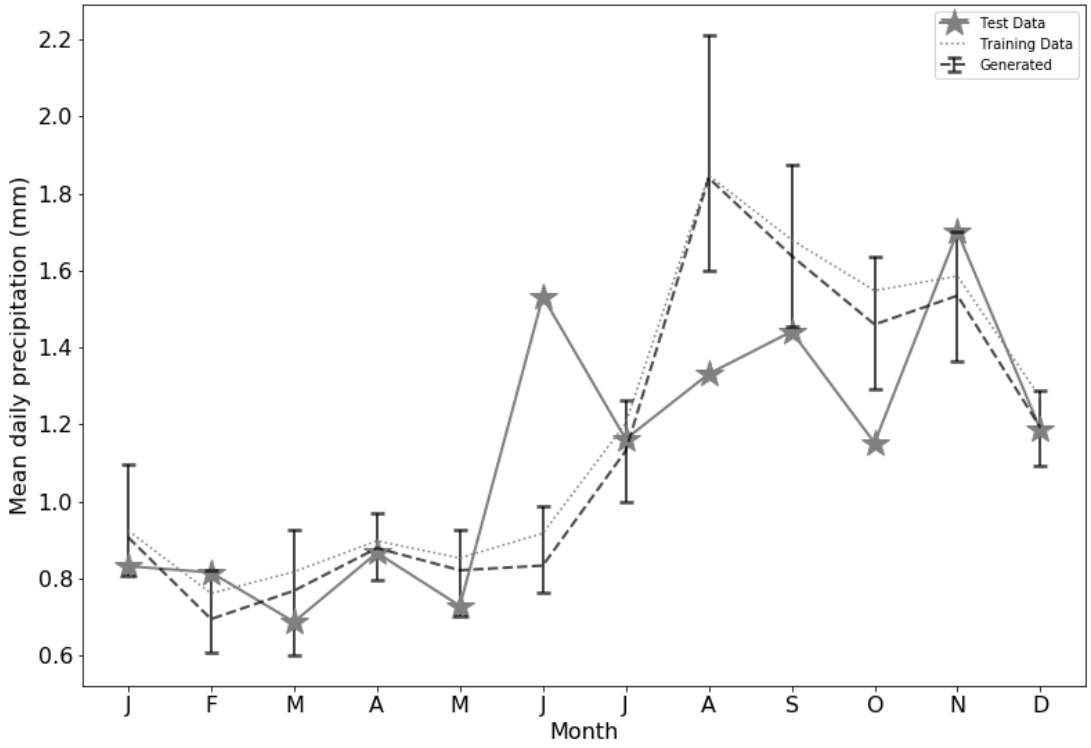


Figure 4.3 – Observed means in daily precipitation from the test dataset (stars) and the range of generated means using a first-order model (black dashed with error bars) at a) Key West, b) Phoenix, c) Reykjavik, d) Tallin and e) Aasiaat. Daily mean precipitation is calculated monthly, with error bars representing the largest mean and smallest mean from the 10 simulations. The dotted line represents the training dataset used to generate the 30-year simulations.

Annual cycles in mean daily precipitation are reproduced well at all locations (Figure 4.3). Where an observed mean does not fall within the range of generated values, the test and training datasets frequently show large differences in precipitation amounts. This is most noticeable in Phoenix (August through May), in Key West (July), in Tallinn (October), and in Aasiaat (June and August). Increasing the number of generated realisations from ten may result in a range of values that may better represent the true variability of the climate. Observed means generally fall within the range of generated means during drier months, with less agreement in the wetter months. This may once more be a result of better performance of the Markov-chain model than the gamma distribution at approximating daily precipitation amounts.

Reykjavik and Tallinn show the closest matches between observed and generated means (Table 4.3). A potential cause of the larger disagreement in Key West, Phoenix and Aasiaat can be explained through the weaker correlations in mean wet-day precipitation (Table 4.2). While all locations show strong correlations between the number of wet days in the observed and generated datasets, Key West, Phoenix and Aasiaat show slightly weaker correlations in mean wet-day precipitation amounts than at Reykjavik and Tallinn. In each of these locations, there are also fewer observed wet days per month (Figure 4.2) than at Reykjavik and Tallinn, in turn resulting in smaller sample sizes available to calculate wet-day shape and scale parameters. This reiterates the previously discussed issue with smaller sample sizes available to fit monthly gamma-distributions in daily precipitation, therefore perhaps requiring longer observed records to effectively capture the true distribution and resulting in the lower agreement shown in Figure 4.3.

In both methods of analysis presented in this chapter, approximating daily precipitation with the two-parameter gamma distribution is likely the prevalent cause of differences between observed and generated time series. This is unsurprising; although the gamma distribution is widely used to approximate daily precipitation, Vlček and Huth (2009) note that the Kolmogorov-Smirnov test tends to be widely misused in place of the Lilliefors test in climate literature, thus exaggerating the acceptance of the gamma distribution as a suitable fit. Nevertheless, the gamma distribution remains appropriate in over half of the weather stations across Europe examined by Vlček and Huth (2009), who also found that there tends to be little geographic coherence in the acceptance of the gamma distribution for daily precipitation amounts. The gamma distribution will however still be used in the GCWG while recognising its limitations, due to its important strengths: it is

positively skewed, bounded at zero, and has the flexibility to represent a range of rainfall distributions and shapes while utilising only two parameters (Husak et al., 2007; Wójcik et al., 2014). It is this simplicity of using only two parameters that will be exploited in Chapters 6, wherein these parameters will be scaled with increasing global mean surface temperature.

4.1.2 STOCHASTIC WEATHER GENERATOR PERFORMANCE AT REPRODUCING OBSERVED EXTREME PRECIPITATION RETURN PERIODS

Section 4.1.1 assessed the GCWG's capacity to reproduce the annual cycles of precipitation. Whilst it is important to accurately reproduce the observed mean behaviour of precipitation, the study of extremes and how they might change in the future holds great value. Extreme value analysis (EVA) will be used here to assess how accurately the GCWG can reproduce return periods in extreme precipitation. Katz (1999) outlines EVA for precipitation generated by a Markov-chain gamma-distribution model. The first-order Markov chain model outlined in Chapter 3 can be reparametrized in terms of the probability of a wet day (here defined as π), and the first-order autocorrelation coefficient, d, where

$$\pi = \frac{p_{01}}{p_{01} + p_{10}} \tag{4.1}$$

and

$$d = p_{11} - p_{01}. \tag{4.2}$$

In a specified period of T days, the number of occurrences of precipitation with intensity greater than 0.1mm is denoted by $N(T)$. Over this period, the maximum amount of precipitation is defined as

$$M_T = \max\{Z_1, Z_2, \ldots Z_{N(T)}\} \tag{4.3}$$

where $Z_k$is the amount of precipitation on the k-th day. In classical EVA, the following result

$$\Pr\{a_T(M_t - b_t) \leq x\} \rightarrow G(x) \equiv \exp(-e^{-x}), \tag{4.4}$$

$$-\infty < \mathrm{x} < \infty, \text{as T} \rightarrow \infty$$

can be used for a non-random, fixed number of independent, identically distributed variables (such as precipitation with the gamma distribution), where G is the limiting cumulative density function called the Type 1 extreme value distribution with zero location parameter and unit scale parameter (Katz, 1999). Following Katz's methodology, if the parent cumulative density function is a gamma distribution, one choice of normalising coefficients a_T, and b_T is defined as

$$a_T = \frac{1}{\beta},\ \frac{b_T}{\beta} = c_T + (\alpha - 1)\ln(c_T),\ \text{ where } c_T = \ln\left[\frac{T\pi}{\Gamma(\alpha)}\right] \qquad (4.5)$$

where $a_T > 0$, α and β are the shape and scale parameters of the gamma distribution respectively (as defined in Chapter 2, equations 2.3 and 2.4), and Γ denotes the gamma function. Note that $T\pi$ in equation 4.5 is the expected number of wet days. A design value, $x(r)$, is commonly introduced in engineered systems to allow for extreme events such that

$$r = \Pr\{M_T > x(r)\} \qquad (4.6)$$

corresponding to the return period of $\frac{1}{r}$ years (such that a probability of $r = 0.01$ refers to a 100-year return period). Though the autocorrelation coefficient, d, has been defined as part of the reparameterization process, the design value is only dependent on the parameters of the chain-dependent process (here, π, α and β). In this special case where the parent distribution is the gamma distribution, the design value can be expressed as

$$x(r) = b_T - \frac{\ln[-\ln(1-r)]}{a_T}. \qquad (4.7)$$

Equation 4.7 has been used to calculate design values for 10- and 50-year return periods from an observed record at each site. Here, the full observed record will be utilised (instead of differing training and test datasets as in Section 4.1.1) to provide the longest record possible to best calculate the design values. Design values have been calculated for the month with the *highest* mean daily precipitation (averaged over all days, wet *and* dry) from the observed record. This corresponds to a T of 28, 30 or 31 dependent month.

The ability of the first-order Markov-chain gamma-distribution model to reproduce the return periods for specified design values is tested. Instead of calculating the design value, the return period, *r*, is calculated using the *observed* design value corresponding to either a 10- or 50-year return period. Rearranging equation 4.7 therefore gives

$$r = 1 - e^{-e^{a_t\left(b_t - x(r)\right)}}. \quad (4.8)$$

The calculated design values (for 10- and 50-year return periods) from the observed record are substituted into equation 4.8 to calculate return periods from 300-years of generated data.

Table 4.4 – The generated return periods for design values with 10- and 50-year observed return periods during the month with the highest mean daily precipitation (calculated from observational data). Design values are calculated from the observed record, with return periods from the generated data corresponding to the relevant design value. Factor refers to how many times larger the generated return period is from the specified 10 or 50 years.

		10-YEAR RETURN PERIOD			**50-YEAR RETURN PERIOD**		
Location	**Month**	**Design Value (mm)**	**Return Period (years)**	**Factor**	**Design Value (mm)**	**Return Period (years)**	**Factor**
Key West	September	75.2	13.7	1.4	104.0	121.2	2.4
Phoenix	August	25.0	18.3	1.8	36.9	121.8	2.5
Reykjavik	February	26.5	20.0	2.0	35.6	140	2.8
Tallinn	August	34.8	14.4	1.4	47.5	77.5	1.6
Aasiaat	September	25.6	21.3	2.1	35.8	219.9	4.4

The GCWG simulates longer return periods than are present in the observed data, and in all instances, simulated 10-year return periods more accurately than 50-years. This is perhaps a reflection of the capabilities of the gamma distribution poorly representing extreme tail ends of daily precipitation; it is known that the extreme tail of gamma distributions is not always heavy enough (Furrer and Katz, 2008) leading to underestimations in the likelihoods of high precipitation intensities. Despite first-order models typically reproducing the full wet-spell length distribution more accurately than other orders, a known downfall of Markov-chain models is also in their accurate representation of extreme events (Wilson Kemsley et al., 2021). Locations with smaller design values are also represented more poorly than locations with larger design values.

Table 4.4 shows better reproductions of return periods in Key West and Tallinn than all other

locations. These two locations also have the highest design values for 10- and 50- year return periods.

As a primary purpose of constructing this GCWG is to use it in a tool to study changes to the climate and hence the frequency and intensity of extremes using the pattern scaling technique, the simplicity of the two-parameter gamma distribution is a major advantage. Furthermore, Osborn et al. (2016) have demonstrated the use of the pattern scaling technique for diagnosing changes in *total monthly* precipitation gamma shape parameters with GMST increase, showcasing the merits of the techniques. Similar methodology will analogously be applied to the *daily* shape and scale parameters in this thesis. Therefore, despite the significant overestimation of return periods, the gamma distribution will still be used in this study. It is perhaps advisable that, when studying changes to return periods using the perturbed weather generator, changes are considered relative to a stochastic simulation of the reference period, rather than relative to the raw observed data.

4.2 STOCHASTIC GENERATION OF THE SECONDARY VARIABLES

The strengths and weaknesses of a Markov-chain gamma-parameter model to reproduce observed precipitation series have been discussed in Chapter 3 and Section 4.1. While precipitation is considered a primary weather variable and is arguably one of the most important variables to accurately reproduce, there is also great importance in accurately reproducing daily temperature values. Similarly to precipitation, daily temperature time series are used in applications such as agricultural and ecological modelling.

4.2.1 GENERATING DAILY MAXIMUM AND MINIMUM TEMPERATURES

There are a range of methods that can be used to stochastically generate daily temperature time series, some of which have been briefly discussed in Chapter 2. Here, a selection of these methods will be examined in greater detail. While it is not uncommon to generate daily mean temperature and diurnal range, daily maximum (T_x) and minimum (T_n) temperatures will be simulated by the GCWG. This is in part due to much of the available observed weather station data containing daily maximums and minimums only, and also to allow the study of changes to extremes under different degrees of warming in Chapter 7,

alongside assessing the GCWG's capabilities at reproducing *observed* extreme temperature statistics.

WGEN, an early Markov-chain gamma-model for precipitation, produces daily maximum and minimum temperature time series using a multivariate weakly stationary generating processes, a process given by Matalas (1967), with the inclusion of daily solar radiation (Matalas, 1967; Richardson and Wright, 1984). While this method maintains correlations between temperature and solar radiation, parameters in this process are determined seasonally, and are not conditioned on precipitation status. Issues in reproducing observed means in T_x and T_n arose in locations where observations did not follow simple sinusoidal shapes, as assumed by the model. Extreme statistics were also reproduced less accurately than means. Solar radiation is also a required variable in this process, while many weather stations *only* record daily precipitation and temperature. To ensure the widest global applicability, it is more desirable to use the fewest variables possible that presently have the widest availability.

An alternative approach is used in LARS-WG (Semenov and Barrow, 2002). A Gaussian (normal) distribution is used to model T_x and T_n . The means and standard deviations of T_x and T_n are calculated *daily* from observations. A Fourier series is hence used to smooth the daily means and standard deviations to represent an annual cycle, with separate series fitted on wet and dry days (Qian et al., 2008). Unlike WGEN, this therefore incorporates conditioning on the precipitation status of the day. A lag-1 autocorrelation process is used between T_x and T_n, with different parameters depending on the precipitation status of the day. An issue that may arise with using daily means and standard deviations conditioned on the precipitation status is potentially small data samples. The maximum sample size for a given day of the year is equal to the number of years of observed data. In practice, some of these days will be wet, and some will be dry, meaning even smaller sample sizes. While fitting Fourier series will reduce noise arising from small sample sizes, in some locations where there are very few (or no) wet or dry days, the processes may not accurately represent actual temperature values.

AAFC-WG uses a slightly different approach that tends to reproduce observed temperatures better than LARS-WG, though with more complexity (Qian et al., 2005). An empirical distribution is fit to T_x and T_n residual series. Daily means and standard deviations are estimated using a spline interpolation procedure on monthly statistics. This is done separately for wet and dry days, once again maintaining correlations between the variables.

Much like WGEN, correlation matrices are calculated from T_x and T_n, and daily solar radiation; however, lag-0 and lag-1 matrices are also used in AAFC-WG. This builds upon WGEN in that generation of temperature is conditioned on the precipitation status, alongside maintaining correlations between maxima and minima in temperature and solar radiation (Qian et al., 2008). Due to the additional complexity of the method, the parameters are not suitable to demonstrate the applicability of the pattern scaling technique to a stochastic weather generator for generating daily temperature values (Chapter 7).

Here, the daily maximum and minimum temperatures will be generated using a conditional, first-order multivariate autoregressive approach, generally following methodology used in the Climatic Research Unit Weather Generator (CRUWG) (Jones et al., 2016) with some additions motivated by techniques present in other weather generators. The distributions of maximum and minimum temperature are much less skewed than precipitation (Richardson, 1981), and a commonality between many methods of stochastic temperature generation is the use of a normal distribution to model daily temperature values (Qian et al., 2008; Semenov and Barrow, 1997). As such, a Gaussian distribution will be used here.

An autoregressive approach will also be used due to the serial correlation of daily temperature values from one day to the next, caused by heat storage (Richardson, 1981). Autoregression is a common feature of many stochastic temperature models. Aside from strong daily correlations, temperature values are also generally linked to the precipitation status. On days with heavy precipitation, temperatures are more likely to fall below normal. The process also shows strong oscillations seasonally. Therefore, the proposed model for generating daily maximum and minimum temperatures uses autoregressive relationships calculated *at least* monthly, with previous day's temperature and either the previous or present day's precipitation (which is used is dependent on the precipitation status) as the predictors.

First, a time series of daily maximum and minimum temperature (e.g., observed or reanalysis) is converted to a residual series. To convert the observed time series to residuals, bimonthly means and standard deviations in T_x and T_n are calculated (where bimonthly refers to twice per month). Means and standard deviations are calculated for each precipitation transition state defined in Chapter 3 (WW, WD, DD and DW where, for example, DW refers to a wet day preceded by a dry day) for the first (days 1-15 inclusive) and second (days 16 onwards, regardless of month for simplicity) halves of the month. The

observed temperature record is hence standardised by subtracting the mean and dividing by the standard deviation corresponding to the appropriate half-month and precipitation state. This results in the calculation of 8 means and standard deviations per calendar month.

This residual series is then modelled as a first-order autoregressive process with a different model structure for each precipitation transition state. The generated precipitation time series is used to establish the transition state of a day, informing which of equations 4.9 – 4.12 to use to simulate the daily temperature residual (T_i). For the DD transition state (current dry day preceded by a dry day):

$$T_i = a_1 T_{i-1} + b_1 + \varepsilon_1 \tag{4.9}$$

for the WW transition state (current wet day preceded by a wet day):

$$T_i = a_2 T_{i-1} + b_2 + \varepsilon_2 \tag{4.10}$$

for the DW transition state (current wet day preceded by a dry day):

$$T_i = a_3 T_{i-1} + c_3 P_i + b_3 + \varepsilon_3 \tag{4.11}$$

and finally, the WD transition state (current dry day preceded by a wet day):

$$T_i = a_4 T_{i-1} + c_4 P_{i-1} + b_4 + \varepsilon_4 \tag{4.12}$$

Coefficients a_N, b_N and c_N are fitted using multiple linear regression of the standardised observed series of temperature and precipitation (also standardised). The suffixes i and $i-1$ indicate the present or previous day, respectively, and ε are independent, Gaussian error terms. While means and standard deviations are calculated bimonthly, regression equation coefficients are instead determined monthly. This is to reduce the risk of overfitting the model, alongside ensuring an appropriate number of samples are included in determining the regression coefficients. Linear autoregressive models of this type (conditioned on precipitation transition state) have been used to generate daily temperatures with good accuracy at a range of sites, including the Caribbean (Jones et al., 2016), the United Kingdom (Jones et al., 2009; Kilsby et al., 2007) and The Upper Indus Basin (Forsythe et al., 2014).

In instances where there are few (or no) temperature values corresponding to a specified precipitation state, the following process is used to determine the regression equation coefficients:

1. The preceding *and* following month will also be used to calculate the regression equations for *only* the precipitation status that does not have a large enough sample size. For example, if there are limited WW days in January, all WW days in December, January and February will instead be used.
2. If the sample remains too small, the preceding and following two months will be used. For example, if there are limited WW days in January, WW days in November, December, January, February and March will instead be used.

The generated residual series is converted back into temperature values by reversing the process used to convert the observed series to residuals. The generated residual is multiplied by the observed standard deviation and added to the observed mean corresponding to the correct half-month and precipitation transition state of the day. In some instances, generated T_n may exceed T_x . In these instances, a correction process described by Dabhi et al. (2021) is followed, where the mean temperature is calculated ($T_{mean} = \frac{T_x+T_n}{2}$) and a small positive number, δ, is randomly (where $\delta < 0.2$°C) added to and subtracted from T_{mean} to determine T_x and T_n respectively.

Where possible, means and standard deviations (conditioned on precipitation status) will be calculated bimonthly. These will be used to convert the observed and generated series to and from residuals. Bimonthly parameters will be used, where possible, to ensure the most globally representative model and in line with the methods outlined by Kilsby et al. (2007). The benefit of using bimonthly parameters is shown in Figure 4.4 for Reykjavik, Iceland, and Tallinn, Estonia. In some instances, the difference between the temperature in the first half of a month and the second half exceeds differences between monthly values themselves. This is most noticeable in the transition seasons, i.e., in the Spring (MAM) and Autumn (SON) months. In DJF and JJA, differences between half-months (and months themselves) are relatively small. In MAM and SON however, differences up to 4°C are noted between the first and second half of November and March in Tallinn, with differences between the monthly and bimonthly parameters up to 2°C in some instances in both locations.

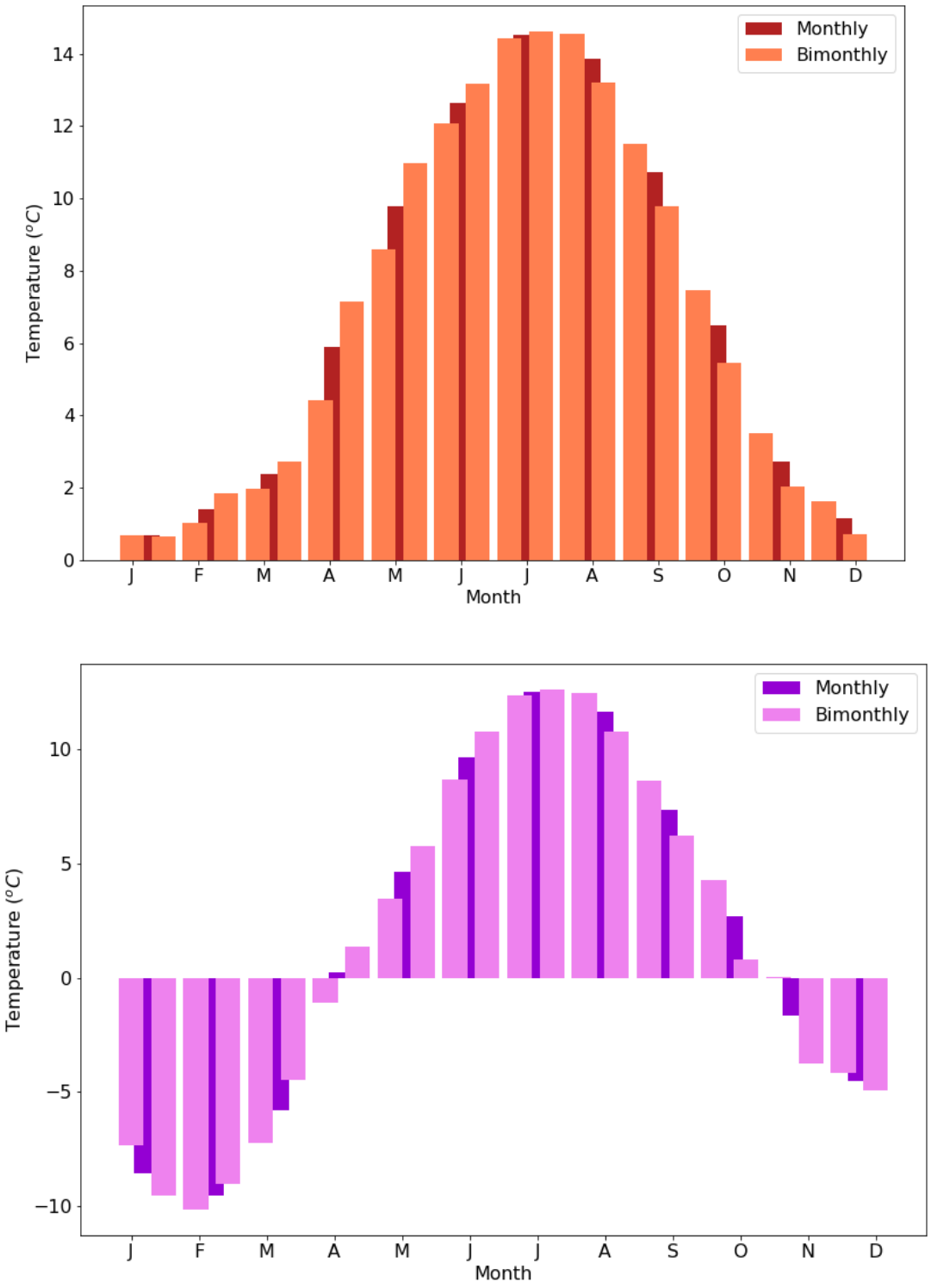


Figure 4.4 – Mean daily maximum temperatures calculated bimonthly and monthly in Reykjavik on days with WD transition states and mean daily minimum temperatures calculated bimonthly and monthly in Tallinn on days with DD transition states.

While in the displayed locations, each transition state contains large samples of temperatures on days with each transition sate, an issue may arise where there are few (or no) observations of a given state in a half-month. To address this issue, a user specified

minimum sample size will be required to estimate the bimonthly conditional means and standard deviations. In instances where the size of sample does not meet the specification due to limited or no observations, a best estimate is taken instead using the following process (hereafter referred to as the "temperature sample size correction process"):

1. If any sample size for a particular half-month and precipitation transition state is less than the specified minimum, the mean and standard deviation for that precipitation transition state is calculated using days with the correct transition state but for the *full* month instead of only a half-month period.
2. If the sample size remains too small, the means and standard deviations on wet or dry days for the corresponding *half*-month will be used instead. For WD and DD transition states, dry days will be used. For DW and WW, wet days will be used.
3. If the sample size is still too small, *three months* of temperature observations for days corresponding to the correct transition state (DW, WD, DD or WW) will be used instead, using the months preceding, including, and following the original half-month.
4. If the sample size is still too small, wet and dry means and standard deviations are instead calculated over three months. As in step 2, wet days will be used for the WW and DW transition states and dry days for DD and WD.
5. Finally, if the sample size remains too small, but not zero, Fourier series will be fit to the bimonthly means and standard deviations to smooth the data. The smoothed values will be used in place of the observational values. If the sample size is zero, the monthly mean (with no conditioning on precipitation status) will be used instead.

The outlined procedure aims to give a best estimate of the mean and standard deviation of the temperature in any half-month, conditioned where possible on precipitation status, whilst also capturing the annual cycle of temperature accurately. These steps are shown in Figure 4.5. Although the LARS-WG approach has not been used in the development of the GCWG, Fourier series will be used to smooth data in instances where sample sizes are not large enough (step 5).

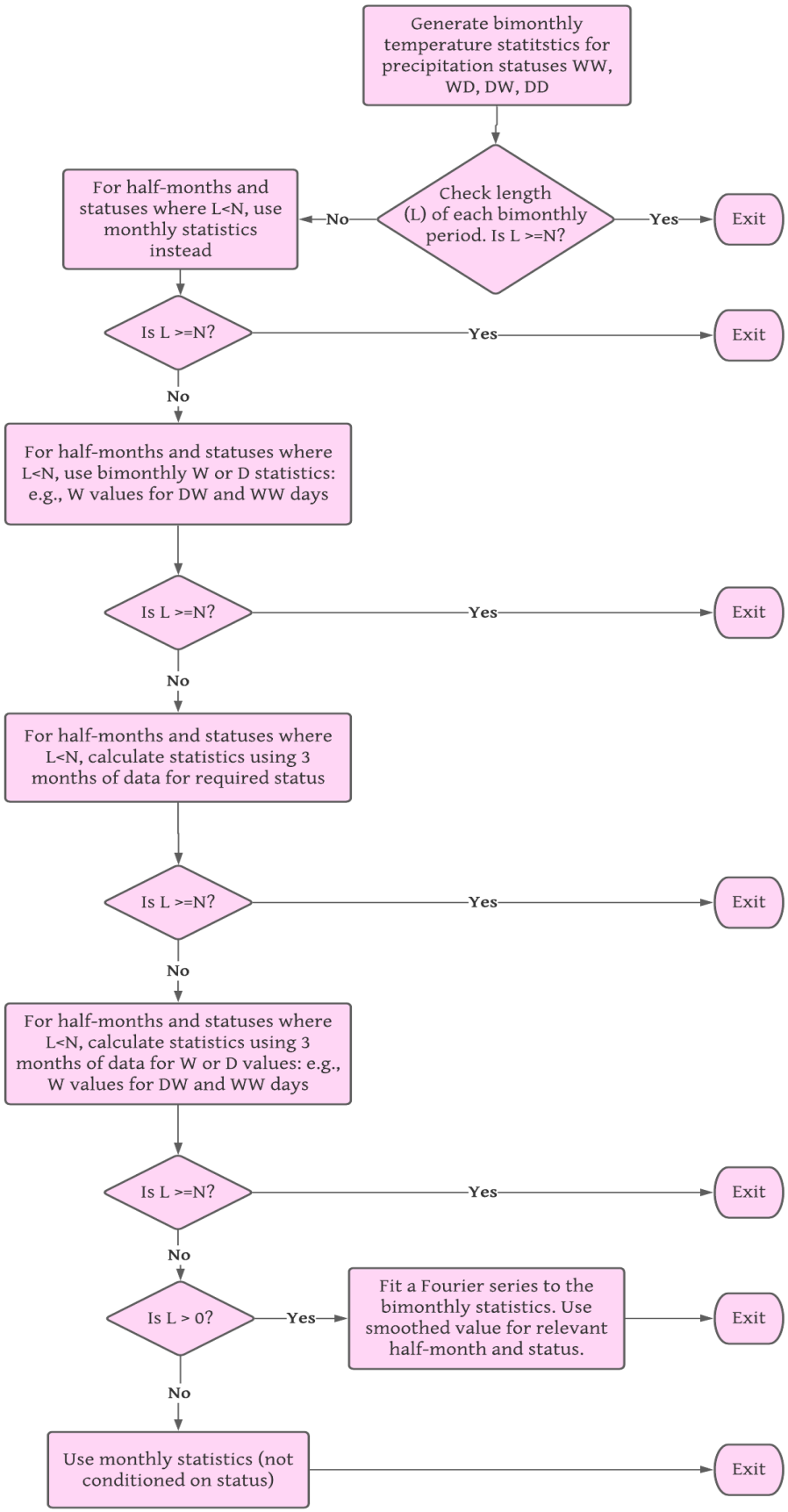


Figure 4.5 – Flowchart showing the steps in the temperature sample size correction process. N is the specified minimum sample size.

An example of this process is shown for Fort Huachuca, Arizona with a specified minimum sample size of N = 25 days (Figure 4.6). Before correcting the temperature sample size, several half-months contain little to no data for WW and DW days – a common occurrence for a dry location. Different half-months have been corrected using different stages of the temperature sample size correction process. From approximately 50 years of observed weather data, there are only 7 days with a WW precipitation status in the second half of January. After following the temperature sample size correction process, daily temperature statistics on WW days have instead been calculated using a three-month period (December, January, and February), resulting in 65 observations. In the second half of March, there are 18 DW days in the observed record. The mean temperatures have instead been calculated using DW days in all of March, resulting in 44 observations. Although temperature statistics for different half-months and precipitation statuses have been calculated using varying steps of the correction process, the annual cycle in Fort Huachuca has been preserved accurately. The standard error in the mean has also been reduced in several cases. This is reflected by the much smaller error bars for several half-months following the temperature sample size correction process. For example, in the second half of May, the standard error in the WW mean is 3.73°C. After increasing the sample size to >25 observations, the error has decreased by 81% to only 0.73°C. Standard errors in all half-months and precipitation statuses do not exceed 1.15°C after following the temperature sample size correction process. This increases the confidence in estimating an accurate value of the true temperature for a given transition state.

Figure 4.6 also highlights the importance of using four transition states, instead of simply wet or dry (as several of the previously published methods use), and that the results are physically plausible. Higher daily maximum temperatures are observed throughout the year during dry spells (DD), as might be expected in the desert climate of Fort Huachuca. However, the temperatures on dry days differ substantially according to whether the preceding day was dry (DD) or wet (WD). Daytime temperatures are further suppressed during a sequence of wet days, especially in February and March but also in many other months (WW means are cooler than DW means).

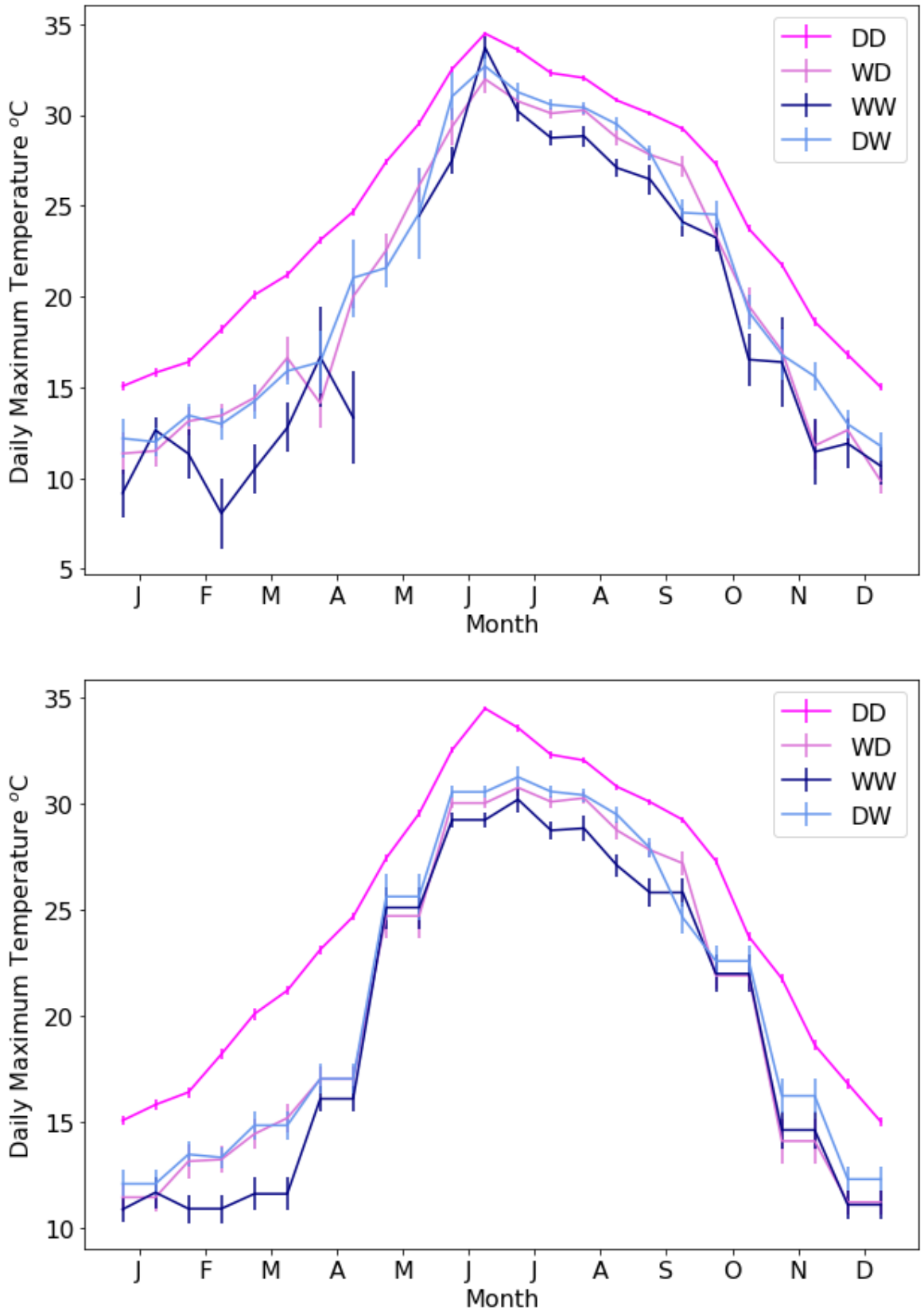


Figure 4.6 – Bimonthly means in daily maximum temperature at Fort Huachuca before (top) and after (bottom) the temperature sample size correction process. Error bars show the standard error in the mean.

Some locations may also not require means and standard deviations calculated on a bimonthly basis due to smaller increases or decreases in temperature over the course of a month. However, to ensure locations such as Tallinn (shown in Figure 4.4) and potentially many other locations globally are best represented, bimonthly parameters will be used as default in the GCWG. Locations with limited sample sizes for specific precipitation statuses have been tackled using the temperature sample size correction process (Figure 4.5).

4.2.2 STOCHASTIC WEATHER GENERATOR PERFORMANCE AT REPRODUCING OBSERVED MAXIMUM AND MINIMUM TEMPERATURE

Means and standard deviations in maximum and minimum temperature, conditioned on the precipitation status, have been calculated bimonthly with monthly regression equations. As in Section 4.1, training and test datasets differ, where the training dataset comprises two thirds of the full observed record, and the test dataset the remaining third (split every three years). As in section 4.1.1, ten 30-year periods of daily P, T_x and T_n have been generated using approximately 38 – 42 years of data. For the rest of this section, "observed" refers to the test dataset (one third of the observed data), while "generated" (or "simulated") has been produced using parameters derived from the training dataset (two thirds of the observed data). It can once again be expected that, should the observed dataset and the ten generated datasets be realisations of the same climate system, the observed means will generally lie within the range of generated means.

Table 4.5 – The number of months where the observed mean daily T_x or T_n falls within the range of generated means from ten 30-year simulations for both maximum and minimum temperature (column labelled "NO. MEANS"). The root-mean-squared-error (RMSE) has been calculated between the observed and generated means, monthly. RMSE shown here is the mean RMSE averaged over all 12 months.

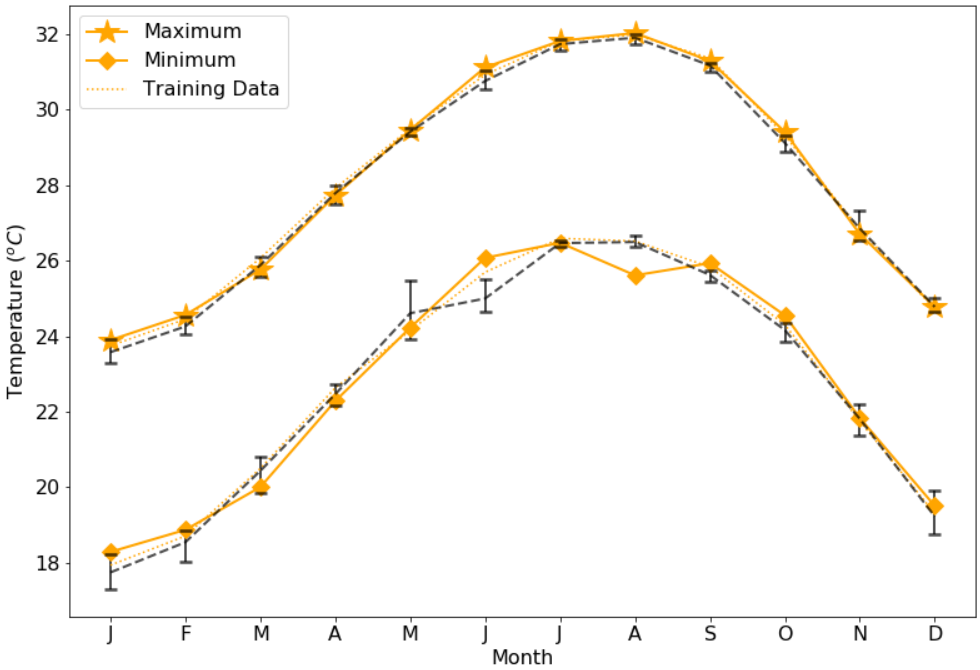
	NO. MEANS		RMSE	
SITE	T_x	T_n	T_x	T_n
Key West	10	7	0.10	0.13
Fort Huachuca	9	7	0.24	0.21
Reykjavik	11	10	0.24	0.26
Tallinn	9	8	0.37	0.42
Aasiaat	7	7	0.19	0.18

In each location, the observed mean lies within the range of generated means from the ten 30-year simulations in over half of the months (shown in Table 4.5 and Figure 4.7). In many cases where the range of generated means does not encompass the observed value from the testing period, the differences are nevertheless quite small. This is reflected by low mean RMSE at all sites (<0.5), including those where fewer observed means lies within the range of generated means. In all locations, the full range of generated means are captured within one standard deviation of the observed record for daily maximum and daily minimum temperature (standard deviation not shown here) even if the observed mean does not lie within the range of generated means. This suggests that the GCWG is generally producing good realisations of the true climate.

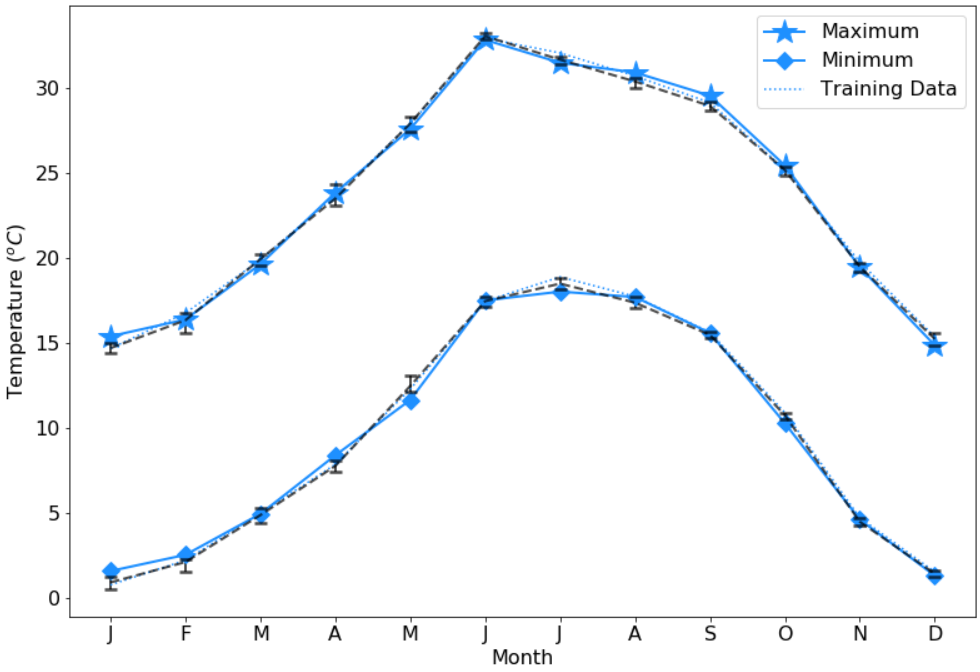
Performance is slightly better regarding maximum daily temperatures than minima. While daily temperatures are widely assumed to follow a normal distribution, Qian et al. (2008) notes that this is not always the case at some locations or for every month; this could be a cause for the different levels of agreement between observed and generated means. It is also known that the distribution of daily minimum temperature is slightly more skewed than maximums (Dabhi et al., 2021). This could result in slightly less accurate reproduction of observed minimum temperature, reflected by fewer generated means capturing the observed in Table 4.5.

Unlike precipitation, there tends to be little temporal structure regarding the months where the observed mean is not captured by the range of generated means. Instead, the most notable cause appears to be due to larger differences between the training and test datasets themselves. While generated means lie very close to the training data, it is important that the GCWG also captures the variability accurately. Like precipitation, the range of generated values may better capture true climate variability by increasing the number of simulations from ten. This way, the chances of generating an unusually warm or cold month (for example, test data for February in Tallinn or August in Key West respectively) or period (test data in Aasiaat) are increased.

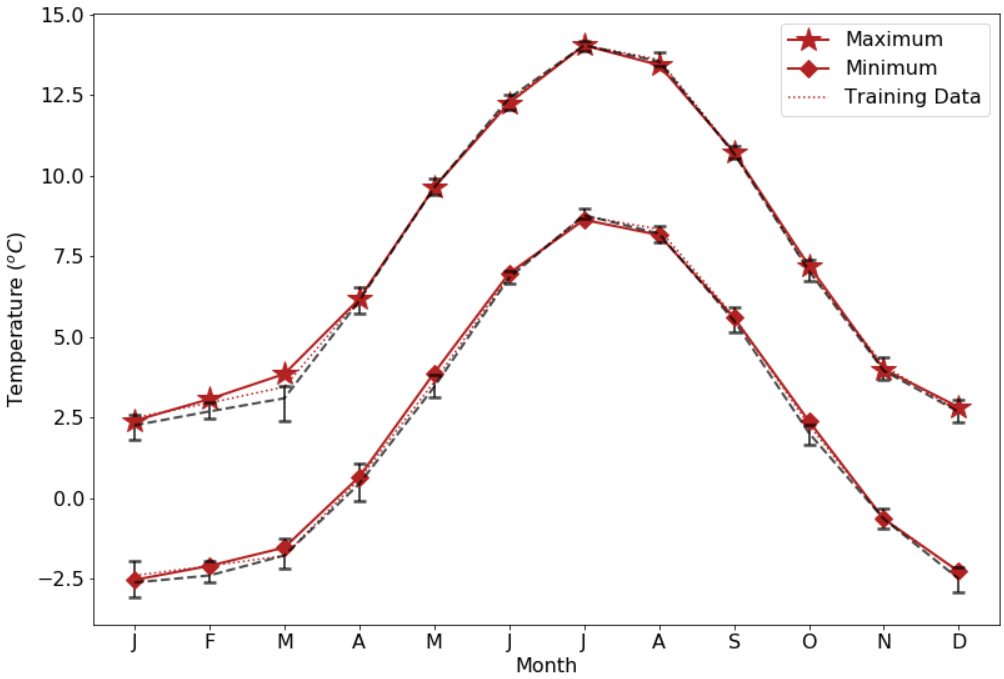
a)


Maximum
Minimum
Training Data
Temperature (°C)
32
30
28
26
24
22
20
18
J
F
M
A
M
J
J
A
S
O
N
D
Month

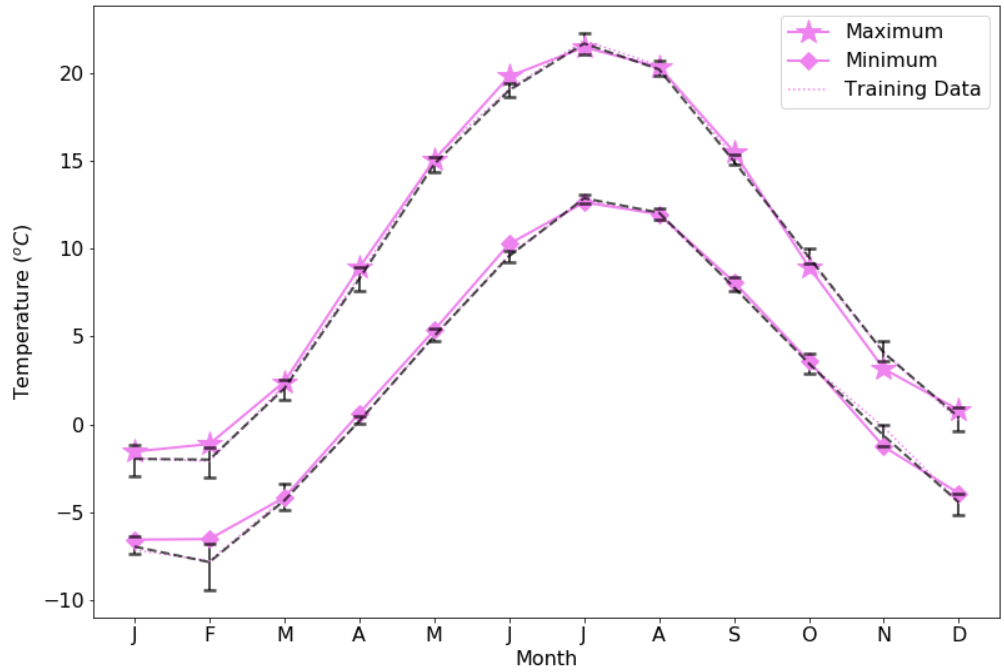
b)


Maximum
Minimum
Training Data
Temperature (°C)
30
25
20
15
10
5
0
J
F
M
A
M
J
J
A
S
O
N
D
Month

c)


Maximum
Minimum
Training Data
Temperature (°C)
15.0
12.5
10.0
7.5
5.0
2.5
0.0
−2.5
J
F
M
A
M
J
J
A
S
O
N
D
Month

d)


Maximum
Minimum
Training Data
Temperature (°C)
20
15
10
5
0
−5
−10
J
F
M
A
M
J
J
A
S
O
N
D
Month

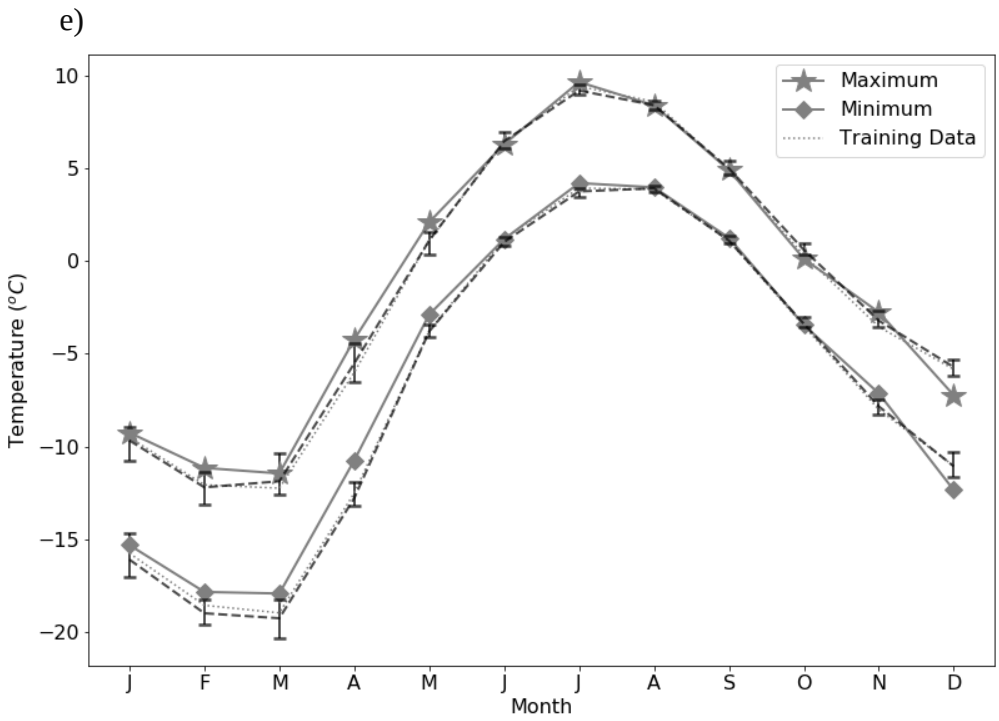


Figure 4.7 – Observed mean (stars and diamonds) and the generated means (black with error bars) in daily T_x and T_n (top and bottom curves, respectively) at a) Key West, b) Fort Huachuca, c) Reykjavik, d) Tallin and e) Aasiaat. Mean daily maxima and minima are calculated monthly, with error bars representing the largest mean and smallest mean from the 10 simulations. The dotted line represents the training dataset used to generate the 30-year simulations.

Figure 4.7 shows that the methods outlined in Section 4.2.1 accurately reproduce the observed annual cycles in T_x and T_n. In comparison with precipitation, the generated mean ranges in daily maximum and minimum temperature (error bars shown in Figure 4.7) are much smaller, though despite this, the number of months with observed means falling within the range of generated means is very similar. This is likely a reflection of the greater variability in daily precipitation than in daily temperature.

While it is important to accurately reproduce observed means in daily maximum and minimum temperature, the ability to capture the variability is arguably of greater importance in the study of extremes (Katz and Brown, 1992; Seneviratne et al., 2021). To assess the GCWG's ability to reproduce observed temperature variability, the observed and generated monthly variances have been compared using Bartlett's test for the ten 30-year realisations. This results in ten statistical tests between observed and generated variances per month, using the same training and test datasets as before. This results in 12 tests per simulation.

The mean number of simulations that show no statistical differences are taken over the 12 months and presented in Table 4.6 (out of a maximum 10).

Table 4.6 – The mean number of simulations showing no statistical difference between observed and generated variances per month. For example, a value of 7.3 represents on average 7.3 simulations show no statistically significant differences between the observed and generated variances per month.

VARIABLE	T_x	T_n	**MEAN**
Key West	6.4	4.0	5.2
Fort Huachuca	5.2	6.2	5.7
Reykjavik	6.8	7.3	7.1
Tallinn	5.9	6.3	6.1
Aasiaat	6.7	4.8	5.8
MEAN	6.2	5.7	

In most locations, at least half of the ten simulations show no statistical differences between the observed and generated variances per month on average with, no temporal patterns recognised (i.e., performance based on the annual cycle). Greater performance is shown in reproducing the variance in daily maximum temperatures than minima. This appears to be a common feature of this GCWG; the model tends to reproduce maxima with greater accuracy than minima. This is likely due to the aforementioned assumption of normality holding stronger for maximum temperatures than minima.

The Expert Team on Climate Change Detection and Indices (ETCCDI) established a suite of indices to represent the extreme ends of precipitation and temperature probability distributions. The ability to reproduce observed extreme indices in daily temperature will be tested using monthly and bimonthly temperature parameters. The warm spell duration indicator (WSDI) and cold spell duration indicator (CSDI) are two indices defined by the ETCCDI that have used to assess the annual count of warm and cold periods respectively in a record, relative to a baseline period (e.g., 1960 – 1990) (Kim et al., 2020). Here, two metrics have been introduced *based* on the WSDI and CSDI indices. The 90th percentile in daily maximum temperature and 10th percentile in daily minimum temperature has been calculated annually from the test dataset. Warm spells and cold spells occur when at least 6 consecutive days experience maxima or minima that exceed or fall below the 90th and 10th percentiles respectively. The annual count of warm and cold spells will be defined as the

WS90 and CS10. It is important to note that a warm (or cold) spell must last at *least* 6 days, though in some instances, the period may be longer. Regardless of length, the spell will only add one to the total annual count.

Table 4.7 – The annual number of observed and generated occurrences where the maximum temperature exceeds the 90th percentile for at least 6 consecutive days (WS90) and where the minimum temperature falls below the 10th percentile for at least 6 consecutive days (CS90).

	WS90		CS10	
LOCATION	**Observed**	**Generated**	**Observed**	**Generated**
Key West	2.2	0.7	1.9	0.2
Fort Huachuca	2.0	0.8	1.6	0.8
Reykjavik	1.8	1.9	1.3	1.1
Tallinn	1.8	1.7	1.8	2.1
Aasiaat	1.5	1.1	1.6	0.3

The GCWG reproduces the annual number of warm spells accurately at Tallin, Reykjavik and Aasiaat. Poorer performance in reproducing CS10 is present in most locations, with sizeable underestimates in Key West, Fort Huachuca and Aasiaat. Dabhi et al. (2021) found that a stochastic weather generator following an autoregressive lag-1 structure tended to underestimate cold spell extremes while hot spell extremes were generally well reproduced. This is likely due to the previously discussed more skewed minimum temperature distribution (Dabhi et al., 2021). This assumption of normality may hence result in the GCWG not capturing the potentially non-Gaussian shape of the more skewed minimum temperature distributions, causing poorer performance in reproducing cold spell extremes.

Poorer ability at reproducing extreme indices and (though to a lesser degree) observed variances in Key West and Fort Huachuca than other locations may be explained by a combination of model-order choice and the temperature sample size correction process. In Chapter 3, it was found that the distributions of wet-spell length, dry-spell length, *and* the interannual variability of precipitation occurrence are all best represented by third-order Markov chains in tropical regimes, of which Key West is categorised into. Weaker performance may therefore be due to inaccuracies in reproducing observed precipitation transition *states*. While Figure 4.2d shows strong correlations between the mean number of

wet days per month for observed and generated data, the exact precipitation state may not accurately be reproduced. For example, transition states DD and WD both result in a dry day but would correspond to a different regression equation, mean and standard deviation, used to calculate the temperature on that day.

A similar issue may arise in Fort Huachuca, which is in a dry regime. A dry regime may be more susceptible to longer spells with no precipitation than other regimes. A known limitation of the first-order Markov chain (that has been used here) is the ability to reproduce long dry spells (Lennartsson et al., 2008), and Chapter 3 concluded that distributions of dry-spell length are better captured by a third-order model in all locations. This means there may be a smaller proportion of DD days in the generated record than the observed record at Fort Huachuca. It is clear from Figure 4.6 that the temperature on DD days is much larger than on WD days, despite both corresponding to a dry day. This once again may affect the choice of regression equation, mean and standard deviation used to calculate the temperature.

The temperature sample size correction process (Figure 4.5) may also affect the ability to reproduce observed variances and hence the extreme indices. Some half-months may use temperature values from a range of different months depending on the length of the sample and which step of the process provides a long enough record. This may in turn result in a wider range of values used to calculate the temperature parameters for a given half-month, thus increasing the standard deviation used to convert the residual series into temperature values and therefore potentially overestimating the variance.

4.2.3 STOCHASTIC WEATHER GENERATOR PERFORMANCE AT REPRODUCING OBSERVED MULTIVARIATE EXTREMES

Whilst it is important to accurately reproduce precipitation and temperature independently, it is also desirable to correctly reproduce the correlations between the two. Relationships between the two variables are particularly important to accurately model in the study of concurrent extreme events. Strong impacts are associated with these multivariate extremes, and processes that cause extreme weather events tend to be spatially or temporally dependent. Despite this, relatively little attention has been given to such events and, traditionally, risk assessments tend to consider only one driver of extreme weather, potentially leading to underestimations of risk (Brunner et al., 2021; Zscheischler et al., 2018).

To assess the skill of the conditional, multiple linear regression method outlined in Section 4.2.1 (equations 4.9 – 4.12), the capabilities of the GCWG at reproducing observed correlations between the two variables have been assessed. The maximum and minimum temperatures on very wet days (exceeding the 90th percentile) and the mean maximum temperature over dry spells (10 days or more with no precipitation) have both been studied.

The maximum and minimum temperatures on days where the precipitation exceeds the 90th percentile (referred to here as very wet days) have been compared using observed and 300 years of generated data. As before, training and test datasets differ, though here the observed record has been split into a 1:1 ratio to provide longer test datasets for comparison with percentiles calculated from the training dataset (with test and training datasets comprising odd and even years of the record respectively). The reference 90th percentiles in daily precipitation for Key West, Fort Huachuca, Reykjavik, Tallinn and Aasiaat are 6.6, 1.5, 7.1, 5.9 and 3.3 mm per day respectively, calculated from the observed (test) dataset.

Boxplots have been produced for T_x and T_n on all days where precipitation is equal to or exceeds the reference 90th percentile for both observed and generated data (Figure 4.8). Temperatures on very wet days tend to be reproduced well by the GCWG. This is reassuring for the GCWG's applications in studying behaviour of concurrent wet and hot, or wet and cold, extremes. The GCWG generally reproduces the median, mean, and interquartile ranges of very wet day T_x and T_n well, though medians and interquartile ranges in T_x tend to be slightly better represented than in T_n. This may once more be due to inaccuracies previously discussed in reproducing the observed distribution of minimum temperature. While interquartile ranges are largely well reproduced, generated data tends to overestimate the ranges of temperature on very wet days slightly; this is most prevalent in Key West for minimum temperature, though also true for other locations. Outliers are also much more prevalent in generated datasets. Generally, these outliers fall *below* the range of temperatures; though a large number are present both above and below at Key West, T_n.

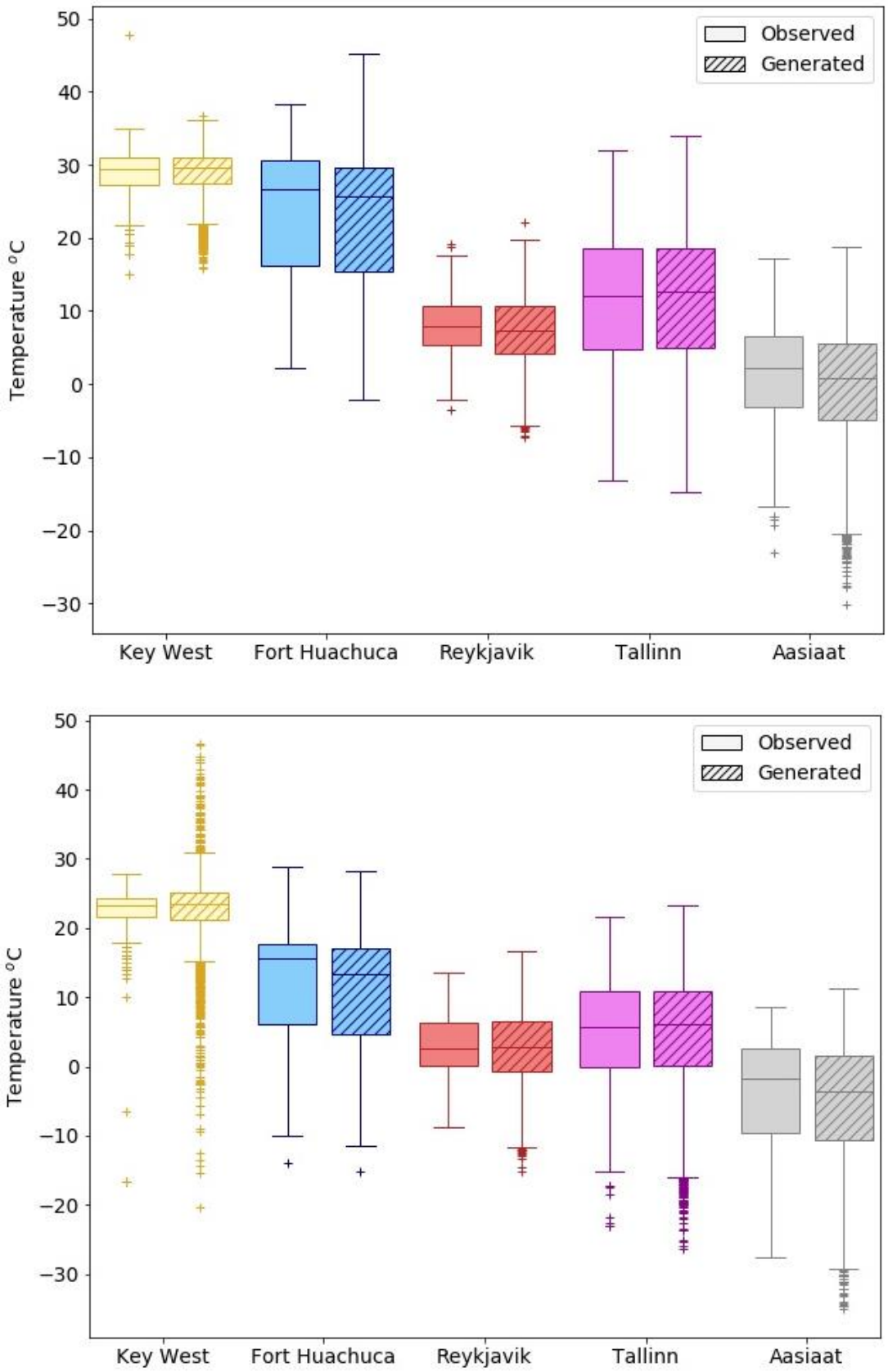


Figure 4.8 – Boxplot showing the observed (solid) and generated (dashed) maximum (top) and minimum (bottom) daily temperatures for very wet (daily precipitation exceeding the 90th percentile) days.

Extreme hot and dry conditions have the potential to intensify summer heatwaves and associated hazards, including drought and fire risk (Flannigan et al., 2000; Kilpeläinen et al., 2010). Vogel et al. (2020) studied changes in frequency of clustered extreme hot and dry events at different degrees of warming, projected from CMIP6 simulations. Vogel et al. (2020)'s method of studying concurrent extremes will be loosely used here. Here, a hot extreme is defined as a day where the maximum temperature exceeds the 99th percentile. Precipitation is aggregated over 90 days, with the date of the sum defined as the last day of the 90-day period. A dry extreme is defined as a day where the 90-day total precipitation falls below the 1st percentile in 90-day total precipitation. The number of times an extreme hot and extreme dry day fall within the *same month* is recorded over each 10-year period in either the observed data or in 300 years of generated data (i.e., in months per 10-years). As an example, a recording of 1 would correspond to 1 occurrence of an extremely hot and extremely dry day within the same month per 10 years on average. Here, the full observed record is used as both the training and test datasets. This is to ensure there are enough years in the observed record to fully capture the occurrence of such extremes. Vogel et al. (2020) instead use the 99.9th and 0.1st percentiles to define extreme hot and dry conditions respectively, however, their findings indicate very few to no regions show clustered hot and dry extremes under present day conditions using this definition, therefore slightly less extreme percentiles have been used here.

Table 4.8 – The number of occurrences of extreme hot and extreme dry days per month per 10 years, and concurrent extreme hot and dry days occurring in the same month per 10 years. "Obs." columns refer to the observed dataset, and "Gen." to the generated.

	HOT		DRY		HOT AND DRY	
LOCATION	**Obs.**	**Gen.**	**Obs.**	**Gen.**	**Obs.**	**Gen.**
Key West	11.6	11.6	3.3	0.3	0.4	0.1
Fort Huachuca	18.0	20.6	4.8	4.0	2.3	1.4
Reykjavik	12.0	13.7	3.7	0.7	0.4	0.1
Tallinn	14.1	12.1	3.6	0.7	0.4	0.1
Aasiaat	15.9	19.5	2.7	0.0	0.0	0.0

Included in Table 4.8 are the mean months per 10-year period where at least one day exceeds the 99th percentile in maximum temperature and falls below the 1st percentile in total 90-day precipitation. For concurrent extremes, it is unsurprising that values are close to (or

even) zero (see Figure 2 from Vogel et al. (2020)) except for Fort Huachuca, where 2.3 months per 10 years experience clustered hot and dry extremes. The occurrence of extremely hot days is reproduced well at each site. However, the number of months with extremely dry days is underestimated at most locations. This ultimately results in the occurrence of concurrent hot and dry extremes to be underestimated by the GCWG. While the frequency of these concurrent hot and dry extremes tends to be underestimated in all locations, values are very small to begin with, and the representation of univariate hot extremes are reproduced well. This is a limitation of this model (accurately reproducing extreme compound metrics) that should be assessed prior to each individual application. It is therefore perhaps advisable that when studying changes to the frequency of these clustered extremes, a reference period should be generated by the GCWG first (spanning the years of the observed record) and the frequency of such events determined from this synthetic time series. The frequency of concurrent extremes from a generated time series under different levels of warming can then be compared to the reference frequency multiplicatively to estimate the change in risk of these concurrent events. For example, if the occurrence of the concurrent extreme is 0.1 times every 10 years under present day climate as simulated by the GCWG but 0.5 under a future simulation, it could be estimated that the probability of such events co-occurring in the future is 5 times greater. This could then be applied to the observed occurrence to estimate future risk.

4.3 CHAPTER 4 SUMMARY

A precipitation and temperature weather generator using a Markov-chain gamma-model and a multivariate linear regression model, conditioned on precipitation transition state, has been introduced (GCWG) and its performance at reproducing observed weather at a selection of locations assessed. In wider literature, training and test datasets tend to be the same (i.e., the full observed record (Schoof and Pryor, 2008; Semenov et al., 1998; Vlček and Huth, 2009)), reducing the robustness of statistical testing. Where possible, training and test datasets should be independent of each other, to reduce bias and the risk of overestimating the skill of the model. While this may have resulted in more failures of the statistical tests, performance is generally very good, and the practical significance of the results is undeniable.

Correlations between observed and generated mean precipitation statistics are generally very high, particularly in metrics that are primarily testing the *occurrence* of precipitation (Table 4.2). This is reflected by high correlations between the observed and

generated mean number of wet days per month. Metrics that consider correlations between the *amounts* of precipitation tend to be weaker, though in most locations still remain strong. This is likely due to the use of a gamma distribution to approximate daily precipitation amounts. Key West, in the tropical regime (Table 4.1), shows weaker correlations in reproducing observed mean precipitation, and this is perhaps caused by a smaller sample size of wet days, with large amounts of precipitation, present in the training dataset used to calculate the shape and scale parameters of the gamma distribution.

The annual cycle of precipitation has been preserved with great accuracy in all locations, with the GCWG producing sensible realisations of the true climate from ten 30-year simulations. Conclusions from Chapter 3 have been reiterated; at all sites first-order models outperform third orders (except Key West, where first and third orders perform similarly). The performance of the model tends to be greater during drier months, and most of the differences between observed and generated precipitation is likely caused by larger differences in the training and test datasets. However, the GCWG must also represent the climate variability accurately; running more than ten simulations may increase the chances of producing exceptionally wet or dry months, which may have been the case for the test datasets in locations such as Aasiaat, where agreement between observed and generated data is weaker than other locations.

Annual cycles in maximum and minimum temperatures are well reproduced in all locations studied, with generated means falling within half of a standard deviation of the observed mean in all locations and all months studied and many observed means within the range of ten generated 30-year mean values. Annual cycles in both maximum and minimum daily temperature are clearly reproduced accurately (Figure 4.7). Variances also tend to be reproduced well, though with more noticeable differences in Key West and Fort Huachuca. This is perhaps caused by the temperature sample size correction process outlined in Figure 4.5, though it is of course necessary to ensure a sample size large enough to reduce further inaccuracies in calculating temperature statistics.

Reproducing extremes in both precipitation and temperature shows more inaccuracies than mean behaviour, though this is a known issue with Markov-type weather generators. Return periods for observed extreme precipitation are largely overestimated, by up to 2.6 and 5 times for 10 and 50-year return periods respectively. This is likely due to the gamma distribution's poor representation of the tail. Frequencies of WS90 are reproduced well in all locations, though the frequency of CS90 is widely underestimated. Maximum and

minimum temperatures are well reproduced on very wet days, though the GCWG tends to produce more extreme temperature values than are present in the observed record.

Despite underestimates in reproducing some extreme behaviour in precipitation and minimum temperature, the behaviour of the GCWG is generally very good, and has been calibrated using weather stations across all five climate regimes whilst maintaining relatively simple relationships between variables. Maintaining the simplicity of the input parameters is of particular importance when diagnosing relationships between the parameter and its response to increasing GMST (Chapter 7). Reiterating the objectives outlined in Chapter 1, these relationships between increasing GMST and each parameter will ultimately be used to perturb the GCWG's inputs to study a range of future climates at different global warming levels. To work around the underestimates in extreme behaviour, studying multiplicative changes (rather than absolute changes) is worth investigating. For example, when studying changes in return periods for a 50-year flood, future comparisons should be made to a GCWG generated baseline and then applied to the observed data.

5 GENERATION AND VALIDATION OF MODELS FOR THE TERTIARY VARIABLES

Wind speed, radiation and humidity are defined as “tertiary” variables due to their generation following, and their widely modelled dependency on, the primary (precipitation) and secondary (maximum and minimum temperatures) variables. Here, different models will be proposed for mean wind speed, solar radiation, and relative humidity. The dependency on the primary and secondary variables has been included in each model to ultimately allow the study of a full suite of stochastically produced climatic variables under different global warming levels (GWLs). Though this has not been shown in this thesis, a description of future work that may incorporate changes in the tertiary variables has been discussed in Chapter 8. The combined generation of the primary, secondary, and tertiary variables together constitute the full **G**lobally **C**alibrated stochastic **W**eather **G**enerator (GCWG) developed as part of this research.

The tertiary variables have been chosen partially following the Climatic Research Uni Weather Generator (CRUWG) (Jones et al., 2016) and are the variables required to calculate daily potential evapotranspiration (PET) using the Penman-Monteith equation in place of direct measurement. The Penman-Monteith equation is the Food and Agriculture Organization (FAO) recommended method used for calculating PET, widely used in literature, requiring a site’s latitude, air temperature, wind speed, radiation, and humidity. Of these parameters, it is solar radiation, temperature and wind speed that have the greatest effect on daily PET (Allen et al., 1998).

A sample of weather stations across the five Köppen climatic regimes have been used to validate models for the tertiary variables for use in the GCWG; the climatology and annual cycles for precipitation and temperature at these weather stations are outlined in greater detail in Chapter 4 (see Figure 4.1and Table 4.1). Observational data from weather stations will be used to determine a model for generating daily mean wind speed, with NASA POWER data (accessible from https://power.larc.nasa.gov/) used for solar radiation and relative humidity where observed records for the full suite of variables are absent.

5.1 DAILY MEAN WIND SPEED

5.1.1 A GENERALISED LINEAR MODEL FOR GENERATING MEAN DAILY WIND SPEED

Daily wind speed time series are necessary in several fields, including engineering sensitivity analyses, construction of wind speed sensitive structures and the operation and maintenance of wind farms (Lydia et al., 2016; Sahin and Sen, 2001). Wind speed is also one of the main meteorological variables that affects daily evapotranspiration (Sentelhas et al., 2010). This section introduces and validates a model for generating daily mean wind speeds.

A Markov-chain gamma-model and a multiple linear regression model have been used to produce synthetic time series for precipitation and temperature, respectively. Neither of these methods will be used to simulate mean daily wind speeds. This is due to the strongly, positively skewed distributions of daily mean wind speed which are not as accurately reproduced using methods such as linear regression. In comparison to precipitation and temperature, there is less literature on the stochastic simulation of mean daily wind speed. Daily maximums and means in hourly (or in some cases, sub-hourly) wind speeds have both received more attention in literature (Aksoy et al., 2004; Nfaoui et al., 1996; Sahin and Sen, 2001; Yan et al., 2002).

Though precipitation is non-normally distributed like wind speeds, a Markov-chain approach will not be used here. Wind speeds do not have the high proportion of zero values to account for, and while there are examples of the use of a Markov-chain model to generate wind speeds in literature (e.g., Aksoy et al., 2004; Jones and Lorenz, 1986; Sahin and Sen, 2001), a required feature is the definition of a number of states with specified wind speed values, as opposed to the simple (and intuitive) two-state (wet or dry), precipitation statuses outlined in Chapter 3. Wind speed states are defined by an upper and lower limit (e.g., 2 to 4m/s). Following the generation of a daily wind speed state, a random wind speed is chosen between the lower and upper limits of the state. Aksoy et al. (2004) suggested a ten-state Markov-chain model to generate hourly wind speeds, though they state that the upper and lower limits of a given state is "highly subjective" and Sahin and Sen (2001) express that choice of categorisation is "rather arbitrary". This subjectivity is also extended into the choice of the *number* of states. Herein lies the main issue with a multiple state Markov-chain wind speed generator for global use – the definition of the transition states. For example, a

location with small variability in daily wind speeds may need fewer states than a site with higher variability. This issue could also arise *within* an individual site, where it may not be suitable to define the same states for different months or seasons of the year. Furthermore, these Markov-chain models tend to be univariate in nature. An important feature of the GCWG is the incorporation of some dependence on the primary or secondary variables in the generation of each tertiary variable. This is to provide the opportunity to potentially study changes in the tertiary variables as a function of increasing global mean surface temperature (GMST) indirectly through the pattern scaling of the primary and secondary variables.

Generalised linear models (GLM) are an alternative, widely used technique discussed in literature to model daily wind speeds (Yan et al., 2002). GLMs consider a daily wind speed as a sample from its own distribution, where all observations are drawn from the same family of distributions. The Weibull, lognormal and Rayleigh distributions have all been used to model daily wind speeds (Sahin and Sen, 2001) though the gamma distribution has been used here. While Weibull is generally considered the best approximation to observed daily wind speed distributions, the differences between Weibull and gamma are deemed insignificant when studying geographical and seasonal patterns of change. As a primary objective of this GCWG is ultimately in the construction of future climate scenarios, the gamma distribution is deemed appropriate here. Furthermore, gamma distribution GLMs are simple to implement in Python. Only predictors, a specified GLM transformation (here, the log link), distribution (gamma) and daily wind speed shape parameters (calculated monthly from the observed record) are required. This is a significantly reduced number of parameters to the alternative Markov-chain approach, with much less subjectivity that arises from the choice of wind-speed states. While also possible to generate wind speeds using a bounded multiple linear regression (MLR) model (e.g., Jones et al. (2016)), GLM gamma-regressions may be preferential where the distribution of the predictand is highly skewed and, unlike alternative MLR approaches, will not need any additional constraints to ensure the generated variable is positive.

The mean of a gamma distribution, μ, will be estimated daily using a linear combination of parameters, where

$$\ln \mu = \beta_0 + \sum_j B_j x_j \tag{5.1}$$

where β_0 is a constant and all other βs are coefficients independently measuring the effect of the predictors, x_j. A log link function will be used to calculate the mean itself, such that

$\mu = \exp(\beta_0 + \sum_j B_j x_j)$. Model parameters are determined monthly, resulting in 12 sets of parameters.

Though following a multiple linear regression model, Jones et al. (2016) suggest the predictors for tertiary variables on day i are precipitation (P_i), mean temperature (T_{mean_i}), temperature range (T_{range_i}), and previous days' value (lag-1 autocorrelation coefficient). In the CRUWG, correlations between the tertiary variables generated should be indirectly maintained through their common dependencies on P_i, T_{mean_i} and T_{range_i}. However, as previously ascertained, a linear model is neither the most appropriate nor widely used in the simulation of daily wind speeds, and while predictors will be loosely based on the methodology of Jones et al. (2016), they will instead be in used in a GLM. Other predictors are discussed in literature, including geographical effects such as altitude, latitude, longitude, and seasonal cycles (Yan et al., 2002)

To determine which of the abovementioned parameters will be used to simulate daily wind speeds in the GCWG, three GLMs with differing predictors will be assessed (where W_{i-1} and P_i are previous day's wind speed and present day's precipitation respectively):

Model 1: W_{i-1},
Model 2: W_{i-1}, T_{x_i} and T_{n_i},
Model 3: W_{i-1}, T_{x_i}, T_{n_i} and P_i.

Though Jones et al. (2016) use T_{mean} and T_{range} as predictors, present day maximum (T_{x_i}) and minimum (T_{n_i}) temperature will be used instead. The GCWG generates maxima and minima instead of diurnal range and mean due to more widely available observations and to aid in a more accurate study of extreme weather. The following transformations

$$T_{mean} = \frac{T_x + T_n}{2} \tag{5.2}$$

and

$$T_{range} = T_x - T_n \tag{5.3}$$

link T_{mean} and T_{range} to T_x and T_n (noting that the former is only an approximation to the true 24-hour mean but is a commonly used approximation). All models discussed in literature, regardless of type (e.g., GLM or Markov chain), incorporate an element of autocorrelation. Here, each proposed model uses a lag-1 autocorrelation coefficient.

Autocorrelation is responsible for much of the variation in daily wind speeds, with some models using several days prior (Aksoy et al., 2004; Nfaoui et al., 1996; Yan et al., 2002). However, only lag-1 autocorrelation will be considered here; it has been found that wind speed autocorrelation coefficients show a negative trend with increasing lag (Khanmohammadi et al., 2021). Daily temperatures and precipitation amounts are the only other parameters considered. Reiterating the goal of the GCWG for use in the construction of climate scenarios under different GWLs, it is important that the predictors in the wind speed model are the ones that will be scaled with increasing GMST.

Maxima and minima in daily temperature have been included as predictors in both Model 2 and 3. Maxima and minima are considered as they are direct outputs from the GCWG in Chapter 4 and are closely related to the diurnal range and mean following equations 5.2 and 5.3. It is known that temperature gradients result in pressure differences, which in turn can produce changes to wind speed. For example, in a statistical analysis of the relationship between wind speed, pressure and temperature, Wooten (2011) found that wind speed is correlated to surface temperature.

Precipitation will be considered in Model 3. Higher wind speeds will promote greater evaporation, which may destabilise the boundary layer and trigger deep convection. Back and Bretherton (2005) found surface wind speeds are responsible for only small fraction of variability in precipitation in the Pacific ITCZ region, but there are significant correlations between the two variables at high-column relative humidities (Back and Bretherton, 2005). In a study of correlations between solar radiation, wind speed, and precipitation in California, Mohammadi and Goudarzi (2018) note a strong possibility of anti-correlation between wind speed and precipitation during very strong El Niño events (Mohammadi and Goudarzi, 2018). In a study in Sweden, Johansson and Chen (2003) found that in locations with steep orographic changes, wind speed is an indicator of strengths of frontal systems and therefore precipitation. Topography is another factor not considered in the Köppen classification system, and therefore the consideration of precipitation as a predictor may benefit the model in some locations with steep orographic changes.

To assess which predictors to use, log-likelihood ratio tests were performed at the 95% confidence level, comparing Model 2 with 1, and Model 3 with 2, to establish whether the additional parameters are necessary.

Table 5.1 – The number of months where a statistically significant improvement is found on the previous model using the log-likelihood ratio test and a confidence level of 95%. Model 2/1 represents the number of months where a statistically significant improvement is found when using Model 2 instead of Model 1 for generating daily mean wind speed.

	NO. OF MONTHS WITH STATISTICALLY SIGNIFICANT LOG-LIKELIHOOD RATIOS	
LOCATION	**Model 2/1**	**Model 3/2**
Key West	7	0
Fort Huachuca	11	2
Reykjavik	9	2
Tallinn	12	3
Aasiaat	12	9

In all instances, a statistically significant improvement in model fit was seen using daily maximum and minimum temperatures (Model 2) as well as previous days' wind speed (Model 1) in over half of the months. In Tallinn and Aasiaat, improvements were seen in all months. With the addition of precipitation (Model 3), statistically significant improvements are less common, and in Key West, there are none at all. Aasiaat is the only site where Model 3 shows statistically significant log-likelihood ratios in a majority of months. This suggests that in many instances, minimal benefit is added through the inclusion of precipitation as well as temperature.

The significance of each predictor can be further assessed using the p-values produced by the GLM for each month (Table 5.2). Previous day's wind speed is a statistically significant predictor in all months at all sites. This is an expected result, due to autocorrelation being responsible for much of the variation in daily wind speeds (Aksoy et al., 2004; Nfaoui et al., 1996; Yan et al., 2002). In at least half of the months (and at least 10 in two of the five sites), maximum and minimum temperatures are statistically significant predictors (except for T_{x_i} in Fort Huachuca, where it is statistically significant in only 5

months). Precipitation, however, is not commonly a statistically significant predictor, except for at Aasiaat. This supports the findings for Model 3 in the log-likelihood ratio tests shown in Table 5.1.

Table 5.2 – The number of months where a predictor is statistically significant (i.e., with p-values < 0.05) in the GLM for generating daily mean wind speed.

	NO. MONTHS WITH STATISTICALLY SIGNIFICANT PREDICTORS			
LOCATION	W_{i-1},	T_{x_i}	T_{n_i}	P_i
Key West	12	6	7	3
Fort Huachuca	12	5	6	2
Reykjavik	12	10	8	2
Tallinn	12	10	10	4
Aasiaat	12	11	12	10

Following the results presented in Table 5.1 and Table 5.2, daily mean wind speed will be generated using the previous day's value and the present day's maximum and minimum temperature (predictors in Model 2). Precipitation has not been included due to its limited improvement of the model at most sites, to minimise the risk of overfitting the model, and to reduce computational power required.

To validate the use of a gamma distribution in generating daily mean wind speed, time series have been generated for the observed mean daily wind speeds using the *observed* predictors for a range of weather stations across the range of Köppen climate regimes. The same weather stations introduced in Chapter 4, Table 4.1 will be used. Daily mean wind speed has been generated using time series of observed daily maximum and minimum temperatures, and previous days' wind speed. The Anscombe residuals,

$$A_{ij} = \sqrt[3]{\frac{y_{ij}}{\mu_{ij}}} \tag{5.4}$$

have been calculated, where y is the observed wind speed and μ the estimated mean of the gamma distribution on day i in location j (Yan et al., 2002). If the distribution of the Anscombe residuals is Gaussian, it asserts confidence that the data is well described by a

gamma distribution and can therefore be used to model daily mean wind speed. Figure 5.1 shows that the residuals are mostly normally distributed, with slight deviations in the upper and lower tails in Reykjavik and Aasiaat respectively.

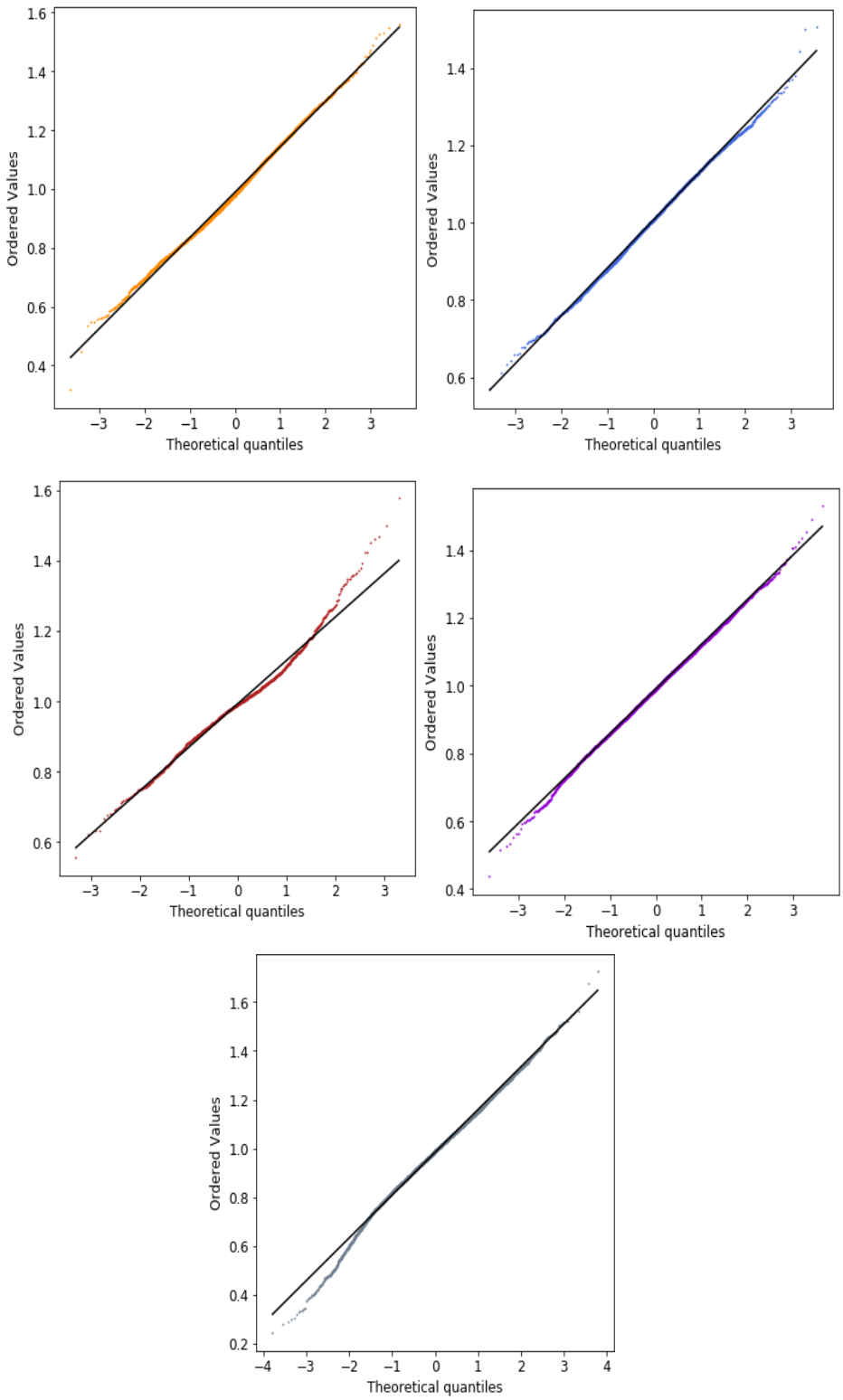


Figure 5.1 – Anscombe residual QQ plots for Key West, Fort Huachuca, Reykjavik, Tallinn and Aasiaat (top left to bottom middle) for daily mean wind speed.

After the mean of the distribution for a given day has been estimated, a wind speed is generated randomly from a gamma distribution with a constant shape parameter, ν, (calculated monthly from the observed record), but with varying scale, λ, based on the estimated value of μ, where

$$\lambda = \frac{\mu}{\nu}. \tag{5.5}$$

5.1.2 ASSESSING GENERALISED LINEAR MODEL PERFORMANCE AT REPRODUCING OBSERVED MEAN DAILY WIND SPEED

Generated precipitation and temperature values have been produced using the methods outlined in Chapters 3 and 4. Training and test datasets will differ using the same method introduced in Chapter 4, where the test dataset consists of every third year, and the training dataset all remaining years (a 2:1 training to test dataset ratio). Again, "observed" data refers to the test dataset and "generated" (or "simulated") means data that has been produced with the model fit to the training dataset. All references of wind speed in this section, unless explicitly stated otherwise, describe daily means.

Fifteen 30-year simulations have been produced for daily precipitation, maximum and minimum temperatures, and wind speed. Fifteen simulations, as opposed to ten, have been used here due to capture the greater variability in daily mean wind speeds than in the primary and secondary variables. The range of means produced by the fifteen simulations is calculated and compared to the observed means for each month. The same logic presented in Chapter 4 is followed here; if the observed dataset and the fifteen generated datasets are considered realisations of the same climate system, it can be expected that the observed means will generally lie within the range of generated means.

The Mann-Whitney U and Levene tests have been performed on the observed and generated wind speeds for each month over the fifteen 30-year simulations. This results in fifteen sets of statistical tests per month. The mean number of simulations per month showing no statistical differences between observed and generated data is shown in Table 5.3 (out of a maximum of fifteen).

Table 5.3 – The number of observed means that lie within the range of 15 generated means and the mean number of simulations (averaged over 12 months) that show no statistical differences out of a maximum of 15 using the Mann-Whitney U test to compare means, and Levene's test to compare variance (at the 95% significance level).

		MEAN SIMULATIONS PER MONTH SHOWING NO STATISTICAL DIFFERENCES	
LOCATION	**No. of observed means within range of generated**	**Mann-Whitney U**	**Levene's Test**
Key West	7	7.6	12.0
Fort Huachuca	6	8.0	12.1
Reykjavik	11	8.6	12.4
Tallinn	9	7.5	8.5
Aasiaat	9	6.3	9.1

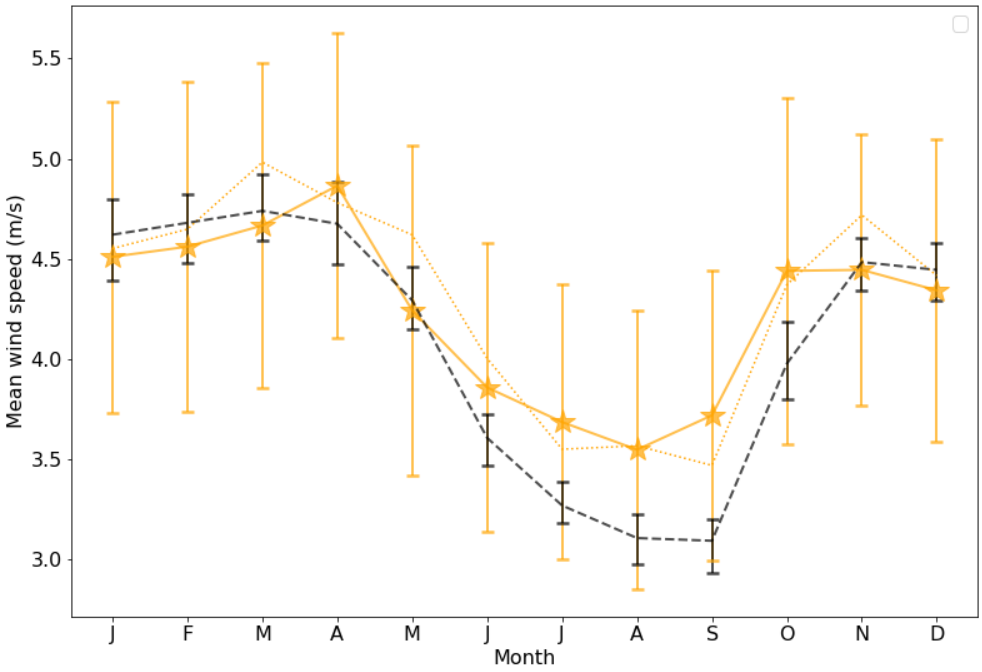
Performance is more dependent on site here than with temperature and precipitation though observed means lie within the range of generated means in at least half of the months at each site. All instances where the observed mean does not fall within the range of generated means, the full range of generated means lies within half a standard deviation of the observed. The performance of the model does not appear to be affected by the season or magnitude of wind speed (shown in Figure 5.2). Generated means tend to slightly underestimate mean daily wind speed, with observed means lying towards (or sometimes above) the upper end of the generated range.

In Key West, Fort Huachuca, and Reykjavik, over half of the simulations show, on average, no statistical differences between the observed and generated means. At all sites, the variance of the data is reproduced more accurately than the mean. In Key West, none of the observed and generated means are statistically similar from June to October. It is between these months that GCWG simulated temperature shows poorer agreement with the

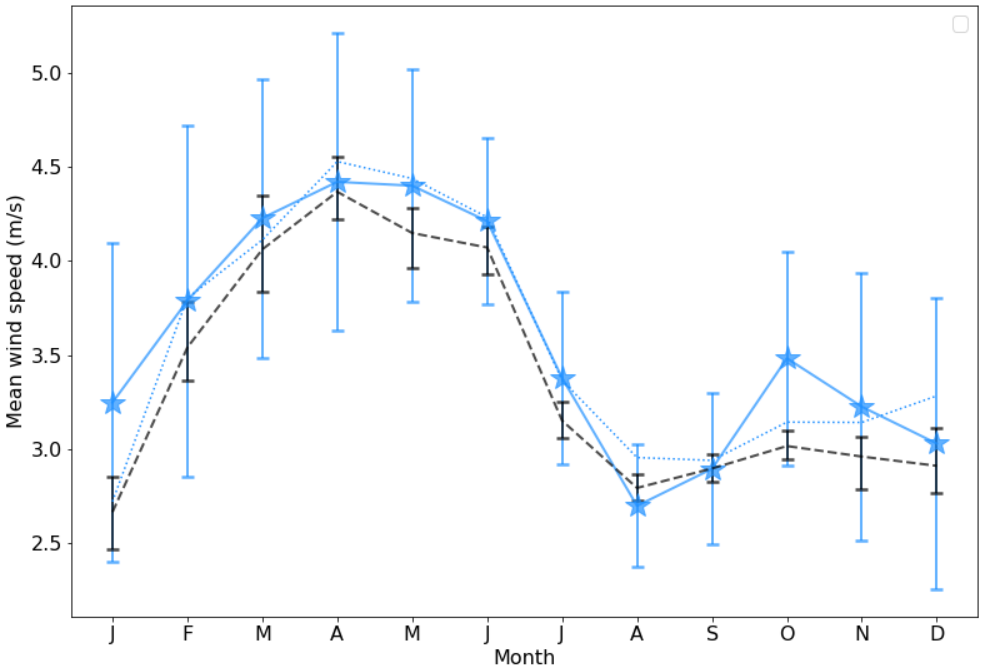
observed data, with observed maximum temperature in June and minimum temperatures in June, August, September and October, lying outside the range of generated values (see Chapter 4, Figure 4.7). In all other months except April, differences between observed and generated means were insignificant in all simulations. Similarly, all simulations in November, December, January and February show statistical differences in the mean at Aasiaat. Once again, these are the months for which the generated temperature series were less realistic than the summer months in Aasiaat, where observed means in maximum temperature lie outside the range of generated values in December and February, and minimum temperatures in November, December and February. This may demonstrate the impact inaccuracies in primary and secondary variables might have on the generation of tertiary variables.

A further reason for poorer performance in the statistical tests in Aasiaat may be due to the elimination of precipitation as a predictor. Table 5.1 and Table 5.2 show that most months at this site would benefit from the inclusion of precipitation as a predictor. This, in combination with the use of a gamma distribution (where a Weibull may perform better), to approximate daily values may also be responsible for some of the greater statistical differences between the observed and generated values. Figure 5.1 shows that the Anscombe residuals deviate from a normal distribution here more than at any other site. Although the residuals at Reykjavik also show deviation from a normal distribution in the upper tail, the number of values is less than in the lower tail at Aasiaat.

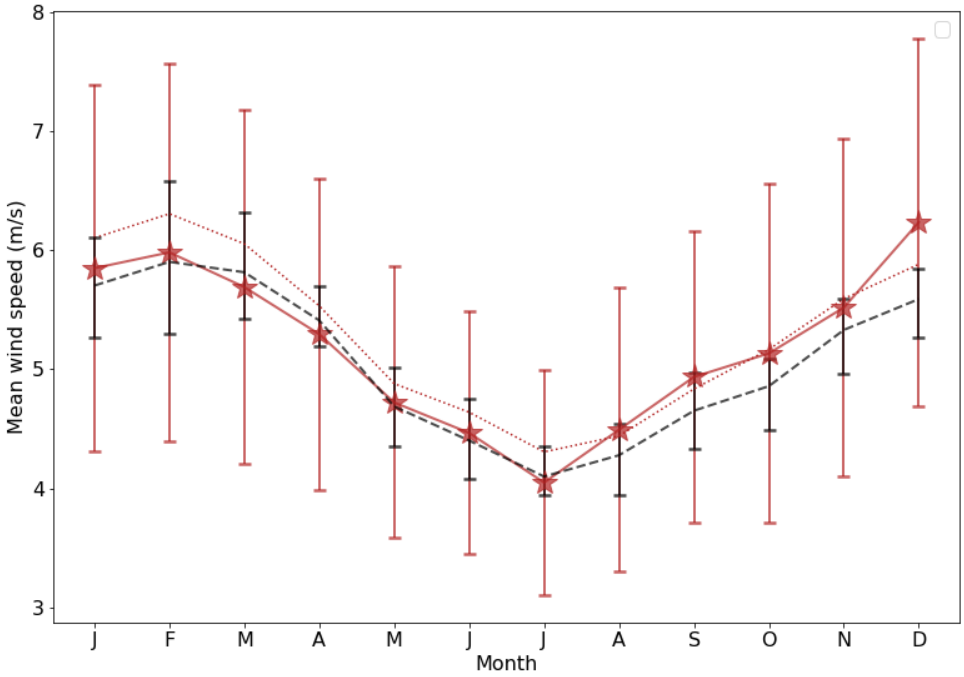
a)


Mean wind speed (m/s)
5.5
5.0
4.5
4.0
3.5
3.0
J
F
M
A
M
J
J
A
S
O
N
D
Month

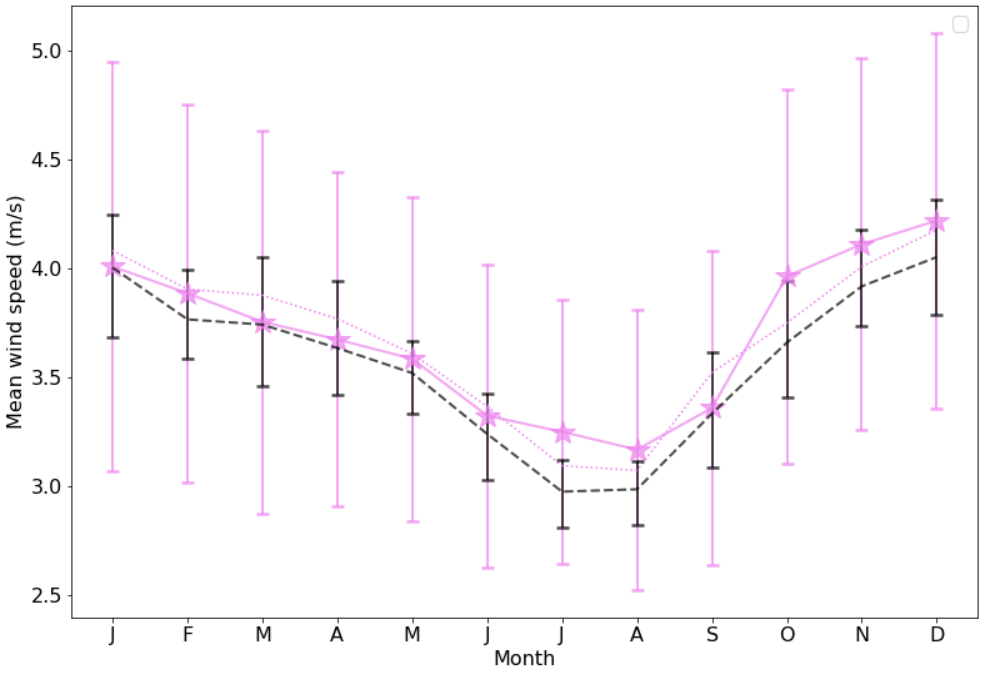
b)


Mean wind speed (m/s)
5.0
4.5
4.0
3.5
3.0
2.5
J
F
M
A
M
J
J
A
S
O
N
D
Month

c)


Mean wind speed (m/s)
8
7
6
5
4
3
J
F
M
A
M
J
J
A
S
O
N
D
Month

d)


Mean wind speed (m/s)
5.0
4.5
4.0
3.5
3.0
2.5
J
F
M
A
M
J
J
A
S
O
N
D
Month

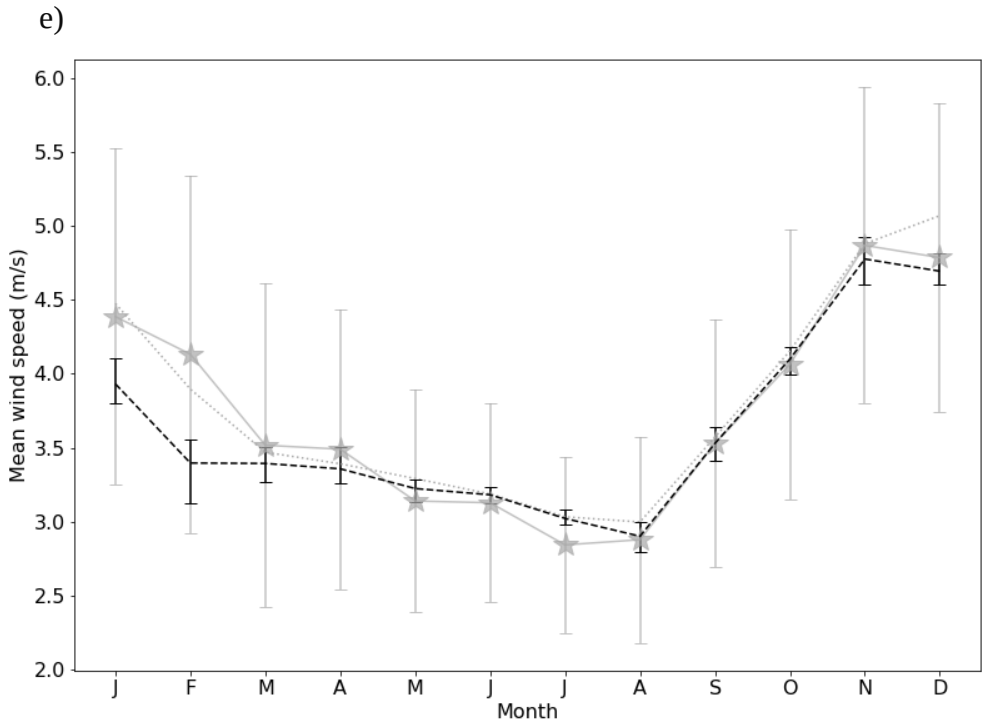


Figure 5.2 – Observed means in daily mean wind speed (coloured, stars with error bars showing one half of a standard deviation), wind speed from the training dataset (coloured, dotted, no error bars) and the generated means (black with error bars) calculated monthly, with error bars representing the largest mean from all fifteen simulations and the smallest mean (black, dashed) for a) Key West, b) Fort Huachuca, c) Reykjavik, d) Tallinn, and e) Aasiaat.

As with precipitation and temperature, the annual cycle in mean daily wind speed has been preserved accurately by the model. Despite some statistical differences between the observed data and the simulations, all locations show few practical differences between the datasets. For example, in Tallinn where, on average, 7.5 simulations per month show statistical differences in the mean, it is clear from Figure 5.2 that the differences between the observed and simulated values are not large. This supports the use of the GCWG in its ability to produce an alternative realisation of the climate that is similar in nature to the observed. It is also important to note that statistical significance is directly related to sample size. Generated sample sizes are large, with a minimum of 840 values. This will increase the likelihood of a statistical test deeming the samples statistically different, despite the closeness of the generated result to the observed.

Most sites show that months with greater observed standard deviations in mean daily wind speed generally tend to simulate a wider spread of values. This is most noticeable in Fort Huachuca, where the observed standard deviation in January is larger than August. Similarly, the range of generated means is larger in January than August. This indicates that the model is reproducing the variability between months successfully.

An important feature to reproduce is the correlations between wind speed and the primary and secondary variables. With T_x and T_n as predictors, some conditioning on precipitation status is also included because their generation depends on the wet/dry state of the current (and sometimes the previous) day. In this thesis, pattern scaling of wind speed input parameters has not been studied, but changes in future wind speed could be incorporated through the future changes in temperature, which will then influence the generated wind speeds. 100 years of precipitation, maximum and minimum temperature, and daily mean wind speed have been generated using the training dataset. Pearson's correlation coefficients have been calculated for wind speed-temperature and wind speed-precipitation relationships.

Table 5.4 – Pearson correlation coefficients between observed and generated wind speed with variables T_x, T_n and P. O and G represent observed and generated correlations respectively.

	T_x		T_n		P	
LOCATION	**O**	**G**	**O**	**G**	**O**	**G**
Key West	-0.25	-0.32	-0.10	-0.16	0.01	-0.06
Fort Huachuca	-0.06	-0.03	-0.01	0.01	0.01	-0.03
Reykjavik	-0.12	-0.11	-0.10	-0.17	0.23	0.05
Tallinn	-0.16	-0.19	-0.05	-0.08	0.04	-0.01
Aasiaat	-0.05	0.03	-0.10	-0.14	0.19	0.01

It is expected that correlations between wind speed and maximum and minimum temperatures should be reproduced well due to their use as predictors in the GLM. Correlations in most instances are reproduced with the correct sign and magnitude. Where the sign is incorrect, observed correlations between wind speed and temperature are extremely close to zero (no correlation). For example, in Fort Huachuca, the correlation between observed and generated wind speeds and minimum temperature is -0.01 and 0.01

respectively. Both values are essentially showing no correlation, despite the change of sign. Key West, Reykjavik and Tallin show the strongest (though still weakly negative) observed correlations between maximum temperature and wind speed. These are well reproduced by the GCWG. Observed near-zero correlations between wind speed and precipitation are also reproduced well at Key West, Fort Huachuca, and Tallinn. Aasiaat and Reykjavik show slightly stronger (though still weak) positive correlations between wind speed and precipitation in the observed dataset. In both locations, the GLM produces negligible correlations between the two variables, though in the same direction (positive).

Weak correlations between wind speed, temperature, and precipitation are not surprising. This is because at most locations, wind speeds are not usually thought to be strongly determined by these variables (Parlange and Katz, 2000). In all locations, observed and generated wind speeds show no correlations with precipitation, except Reykjavik and Aasiaat, where weakly positive correlations are present in the observed data. However, Table 5.2 showed that in most locations, temperatures do have significant effects on the regression. Despite weak correlations, it is important to incorporate a primary or secondary variable in this model; this is to allow changes in daily mean wind speed to be produced indirectly through the scaled future changes in the primary and secondary variables.

While the results presented here suggest previous days' wind speed and daily maximum and minimum temperatures are sufficient predictors in the GCWG, it is important to note that, whilst a site in each weather regime has been considered, the Köppen climate classification system does not consider wind speeds in its categorisations. Therefore, the use of the five regimes which are categorised on temperature and precipitation alone, may not be representative of the range of global wind speed behaviour. It is also important to note that in all locations except Key West, at least 2 months of the year were better represented by a model with the inclusion of precipitation (Table 5.1).

5.2 DAILY SOLAR RADIATION

Daily solar radiation (insolation) data is an important climatic variable for many applications. Such sectors include the application and development of renewable energy, in the study of environmental pollution, crop production and in hydrological and agricultural models (Shrestha et al., 2019). Solar radiation is also another quantity required to accurately calculate reference PET using the Penman-Monteith equation. One of the most important variables in the calculation of PET, solar radiation provides the energy that is required to change the state of water molecules from liquid to vapour.

There is greater discussion in literature regarding the stochastic generation of solar radiation than wind speeds, however it is widely known that there are several issues with observational solar radiation data. In comparison to the primary and secondary variables, solar radiation is rarely measured, and, where it is measured, poor maintenance of equipment and high costs of replacement results in several months of data missing or inadequate quality of data (Ayodele and Ogunjuyigbe, 2015). To counter this issue, NASA POWER datasets will be used instead of weather station records.

5.2.1 NASA POWER PROJECT DATA

Long, temporally consistent daily time series of precipitation, maximum and minimum temperature *with* solar radiation are not widely globally available or are lacking in quality and coherence. Although the European Climate Assessment & Dataset (ECAD) (Klein Tank et al., 2002) time series contain several stations with a full suite of climatic variables, locations are spatially close (i.e., Europe and North Africa) and do not capture a range of climatic regimes. Instead, data has been sourced from the NASA POWER Project (accessed from https://power.larc.nasa.gov/). Solar radiation data is available globally with records starting in 1981, a 1° x 1° latitude/longitude grid and meteorological variables on a ½° x ½° grid. Solar and meteorological parameters have been derived using a combination of satellite and model-based methods. In a comparison between NASA POWER data and weather station observations carried out at Fort Huachuca, Tallinn and Aasiaat (for example), annual cycles of the variables are very similar and as such, the NASA POWER data is therefore deemed an appropriate alternative for the validation of this model.

Table 5.5 – Site names, regimes and coordinates of NASA POWER data used in place of observational data.

LOCATION	CLIMATIC REGIME	LATITUDE	LONGITUDE
Brasília, Brazil	Tropical savanna, Aw	-15.8	-47.9
Fort Huachuca, USA	Hot desert, BWh	31.6	-110.3
Punta Arenas, Chile	Subpolar oceanic, Cfc	-53.1	-70.6
Tallinn, Estonia	Humid continental, Dfb	59.4	24.8
Aasiaat, Greenland	Polar, ET	68.7	-52.9

Brasília, Brazil has been used in place of Key West, USA, and Punta Arenas, Chile in place of Reykjavik, Iceland. Replaced locations are in the same sub-classification as previous sites and have been chosen to diversify the locations studied away from North America and Europe. While it is more desirable to use observational data in calibrating the model, it is also of importance to use long, temporally cohesive time series, that span a range of climatic regimes.

5.2.2 GENERATING SOLAR RADIATION BY TRANSFORMING TO A CLEARNESS INDEX AND USING A MULTIPLE LINEAR REGRESSION MODEL

Several methods have been discussed in literature to generate daily solar radiation time series. As with many other variables, most methods incorporate some form of autocorrelation. Nicks and Harp (1980) proposed a model conditioned on the precipitation status of the day (as generated by a Markov-chain model, utilising the same four states defined in Chapters 3 and 4), *independently* of temperature values. Monthly means, standard deviations and lag-1 correlation coefficients are the only parameters required in this method of solar radiation generation. While Nicks and Harp (1980) found that solar radiation was accurately reproduced in 11 out of 12 months, their study was carried out at only one site (Chickasha, Oklahoma) with only 11 years of training data, and non-independent training and test datasets. While this method does incorporate dependence on precipitation status,

most other methods outlined in the literature also contain an element of temperature dependence. This is due to a direct correlation between temperature and solar radiation.

Richardson (1981, 1982) suggested a multivariate, weakly stationary generating process, comprising of maximum temperature, minimum temperature, and solar radiation, with different parameters on wet and dry days. This method has briefly been discussed in Chapter 4, Section 4.1 in reference to temperature generation. Fourier series are used to smooth daily means and standard deviations in temperature and solar radiation, conditioned on the wet or dry status of the day. The observed records are hence converted to residuals, and matrices of lag-0 and lag-1 cross-correlation coefficients used to generate a residual series. For the application of this method, several parameters must be calculated for each day of the year. While Richardson (1981) suggests that the parameters can be reduced due to few differences in the inter-variable correlations between sites studied, only 3 sites were studied, all located in the contiguous USA. It is therefore not with confidence that these inter-variable correlations will remain the same across the globe, and, even after reducing the number of parameters, several still remain. This method will therefore not be used to generate solar radiation in the GCWG.

A first-order autoregressive, multivariate linear model will instead be used to simulate daily solar radiation time series following a transformation of the entire radiation time series to a clearness index (a measure of the atmospheric clarity). While stochastically modelling clearness index in this way is novel, such a transformation has previously been used by Ayodele and Ogunjuyigbe (2015) to predict monthly averages in solar radiation. A time series of daily clearness index will hence be produced and transformed back into a daily solar radiation value following generation of a full series. The rationale for targeting a clearness index rather than solar radiation directly is that physical features that determine temporal variations in atmospheric clarity (principally the presence and optical thickness of clouds) are linked to the occurrence of precipitation and to the minimum and maximum temperature, whereas the other contributors to solar radiation (day of year and latitude) are static. This is therefore a simple transformation that requires little computational power and the only additional variable required for the calculation is the readily available latitude of the site.

The maximum extraterrestrial radiation (DE) for a given day of the year (1 for 1st of January, 365 for 31st December) i, and specified latitude ϕ, can be calculated using

(5.6)

$$DE(i,\phi) = \frac{24}{\pi} I_{sc} E_0(d) \begin{bmatrix} \omega_s(i,\phi)\sin\delta(i)\sin\phi \\ +\cos\delta(i)\cos\phi\sin\omega_s(i,\phi) \end{bmatrix}$$

where the solar constant, $I_{sc} \sim 4.921 \frac{MJ}{m^2 hr}$, E_0 is the eccentricity factor, ω_s is the sunrise hour angle in radians, and δ the declination. For leap years, the final day of the year ($i = 366$) is replaced by $i = 365$ instead. The eccentricity factor varies with day, where

$$E_0(i) = 1 + 0.033 \cos\left(\frac{2\pi i}{365}\right)$$

with sunrise angle defined as

$$\omega_s(i,\phi) = \cos^{-1}(-\tan\phi \tan\delta(i))$$

with the constraints

$$\omega_s(i,\phi) = \pi \quad \text{if } -\tan\phi \tan\delta(i) > 1,$$

and

$$\omega_s(i,\phi) = 0 \quad \text{if } -\tan\phi \tan\delta(i) < 1.$$

Finally, the declination is defined as

$$\delta(i) = \delta_M \sin\left(\frac{2\pi(i+284)}{365}\right)$$

where

$$\delta_M = \frac{23.45\pi}{180}.$$

The observed insolation can be transformed to a clearness index using

$$C_i = \frac{R_i}{DE_i} \tag{5.7}$$

where C is clearness index, R is the observed solar radiation (in MJ) and DE is the maximum extraterrestrial radiation (in MJ) reaching the latitude of observation, calculated using equation 5.6 for Julian day, d. This in turn provides a time series of daily clearness indexes instead of radiation values.

To generate clearness index on day i (C_i), precipitation, maximum and minimum temperature, and previous day's clearness index will be used as predictors, such that

$$C_i = c + dP_i + eT_{x_i} + fT_{n_i} + gC_{i-1} + \varepsilon \quad (5.8)$$

where $d, e, f,$ and g are coefficients of regression, c is a constant, and ε is an independent, standard, Gaussian error term. Clearness index is a constrained variable; it cannot exceed or fall below 1 or 0 respectively. In instances where the model generates a value ≥ 0.99, the GCWG will set the value to 0.99. Similarly, for a day with a generated value of ≤ 0.01, the GCWG will set the value to 0.01. Finally, equation 5.7 can be rearranged such that generated clearness index is multiplied by the maximum potential DE on Julian day d, calculated from equation 5.6, to estimate a given day's total solar radiation. Prior to the regression, observed clearness index must be converted to a residual series using means and standard deviations relevant to the month of the observation. Regressions are fit monthly, resulting in 12 sets of coefficients that vary by site. Following the simulation of a residual time series, the values are converted back to an insolation using the corresponding month's observed mean and standard deviation.

Predictors have been chosen based on each variable's relationship with solar radiation. There is of course a direct, positive correlation between temperature and solar radiation. Simply, a fraction of the radiation reaching the Earth will be absorbed and converted into heat, therefore responsible for the temperature. The total radiation reaching and being absorbed by the Earth's surface from the sun is dependent on several factors, including geographic location, surface albedo, season, and weather conditions. It is intuitive that on a cloudy day where there is less insolation, temperature will likely be lower. Precipitation has also been chosen as a predictor because it is generally positively correlated to cloud coverage, which is in turn negatively correlated with insolation. The Pearson correlation coefficients between the variables at each site introduced in Table 5.5 between solar radiation and daily maximum temperature, minimum temperature and previous days' solar radiation are all statistically significant, and positive (strongly, at most sites). Conversely, correlations between solar radiation and precipitation are negative.

The same process (equation 5.8) can be used to generate daily insolation values directly without any transformation, simply substituting daily solar radiation values (measured in MJ), R, in place of clearness index, C. However, it is advised that solar radiation values are converted into clearness indexes in most cases, following a comparison of the residuals of regression. Time series of daily solar radiation and clearness index have been produced using the observed precipitation, maximum and minimum temperature, and previous day's solar radiation or clearness index respectively. The residuals (the difference

between the predicted and actual observed values) from each method have been compared at each site. If the residuals are normally distributed, a QQ-plot should show a straight line. It is clear from Figure 5.3 that the residuals are much closer to a normal distribution following the clearness index transformation in Tallinn and Punta Arenas. At some sites, the transformation from solar radiation to clearness index does little to change the distribution of the residuals (i.e., Brasília). In this instance, however, the original, untransformed distribution of solar radiation is much closer to a normal distribution than in Tallinn and Punta Arenas.

Based on the sites studied here, the transformation will be recommended prior to the regression. Transforming the variable does not seem to cause any ill effects in generating solar radiation times series in locations where the transformation does not improve the distribution of the residuals, like Brasília. Therefore, the transformation may ensure the most globally applicable solar radiation model. Furthermore, transforming the variable does not change the risk of overfitting, nor does it add any extra parameters to the model and thus any increase in computational time is negligible.

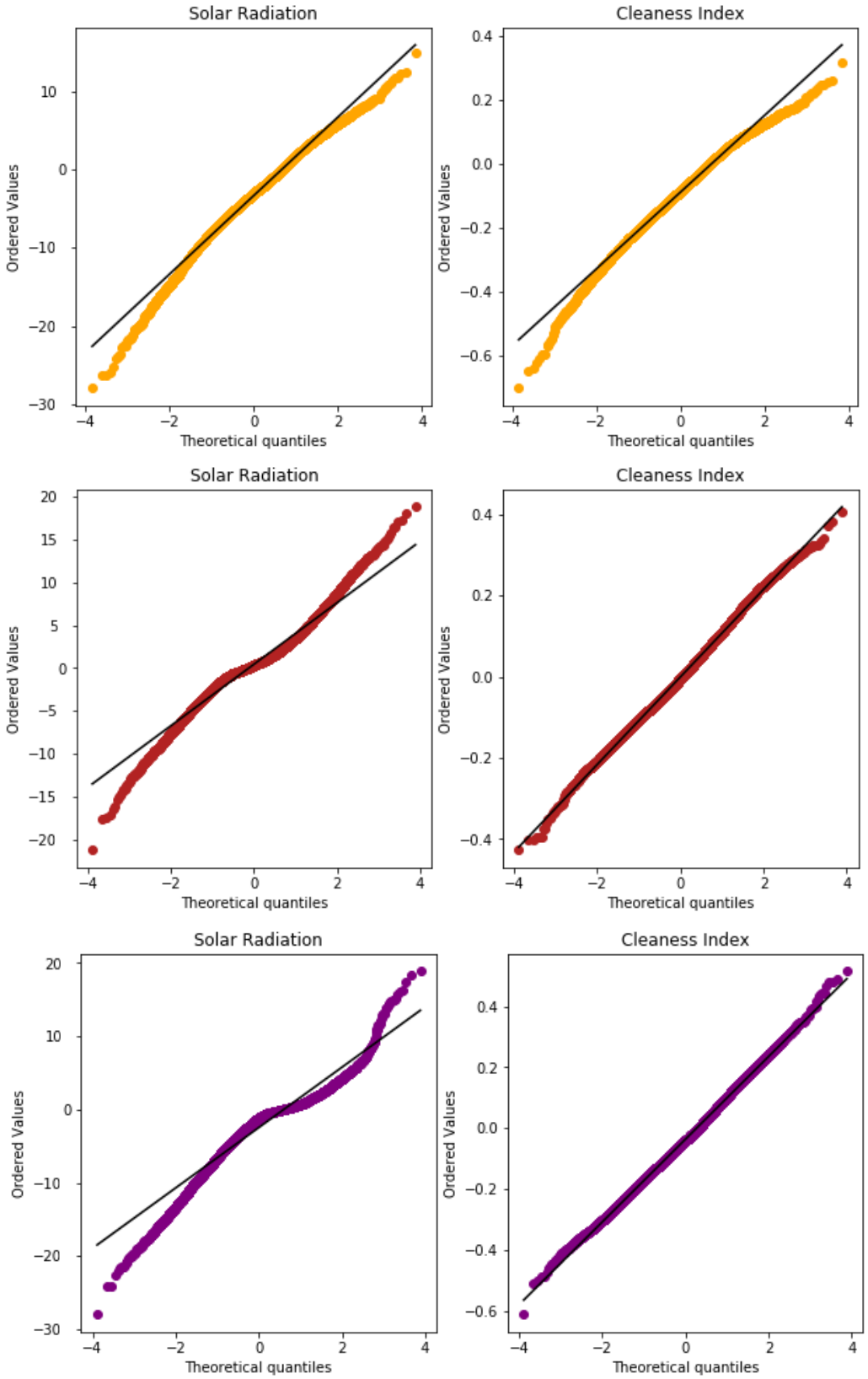


Figure 5.3 – QQ plots showing regression residuals for left) insolation and right) clearness index at top) Brasília, middle) Punta Arenas, and bottom) Tallinn.

5.2.3 ASSESSING THE MULTIPLE LINEAR REGRESSION MODEL PERFORMANCE AT REPRODUCING OBSERVED DAILY SOLAR RADIATION

The full NASA POWER series for each site is split into training and test datasets in the same 2:1 ratio as in Section 5.1.2. However, the NASA POWER datasets are significantly shorter in length than the previously used observed records, spanning approximately 30 to 40 years (with only 30 years of solar radiation data regardless of full record length). Splitting the data in the same 2:1 ratio as before (training to test datasets) leaves at least 20 years to calculate parameters in the training dataset, but only 10 years in the test dataset. For simplicity and consistency with other sections, test data will be referred to as "observed", despite originating from satellite/model outputs. "Generated" (or "simulated") will once again refer to the data simulated using the parameters calculated from the training dataset. The ability to reproduce the observed daily solar radiation, *not* clearness index, will be assessed here. NASA POWER solar radiation time series is provided in units of kWh/m^2. This has been converted into units of MJ/m^2.

Thirty 10-year simulations have been produced at each site. This is to reflect the smaller size of the test dataset (only 10 years) than for other variables. The range of means in daily insolation produced by the each of the thirty simulations is calculated and compared to the observed data by month. Once again, if the observed dataset and the thirty generated datasets are realisations of the same climate system, it can be expected that the observed means will generally lie within the range of generated means. The Mann-Whitney U test and Levene's test have been used to compare the mean and variance respectively of each of the thirty generated datasets.

Table 5.6 – The number of observed means that lie within the range of 30 generated means out of a maximum 12. The mean numbers of simulations averaged over 12 months that show no statistical differences out of a maximum of 30 using the Mann-Whitney test to compare means, and Levene's test to compare variance.

		MEAN SIMULATIONS PER MONTH SHOWING NO STATISTICAL DIFFERENCES	
LOCATION	**No. of observed means within range of generated**	**Mann-Whitney**	**Levene's Test**
Brasília	10	12.5	8.8
Fort Huachuca	11	16.8	14.5
Punta Arenas	8	20.1	24.0
Tallinn	11	21.4	19.8
Aasiaat	8	15.7	14.4

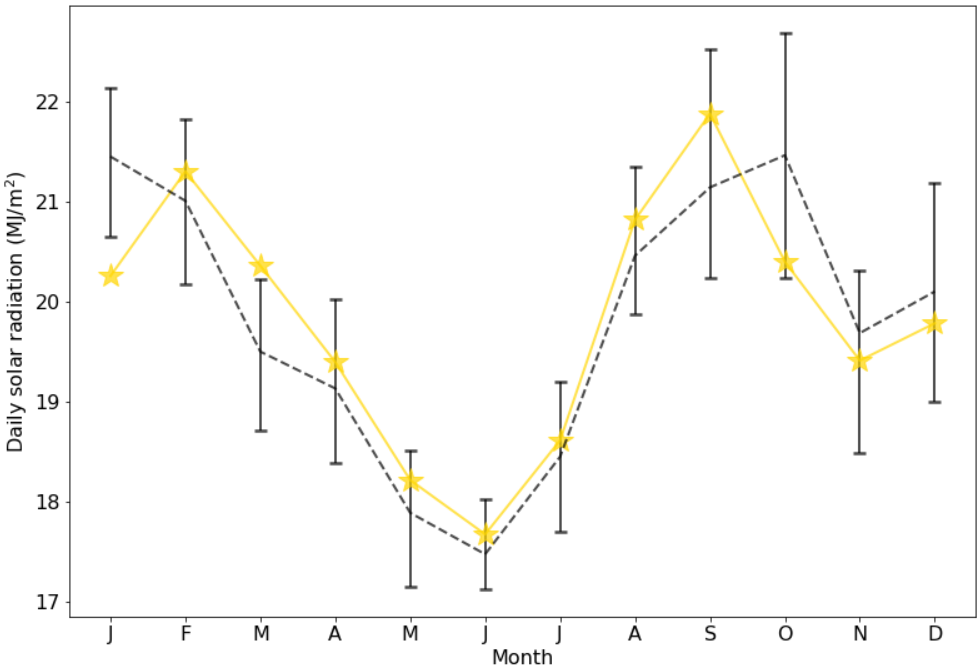
Observed monthly means in daily solar radiation lie within the range of generated means in at least 8 months at all sites. In any instance where the observed mean lies outside the range of generated means, the observed value lies within half a standard deviation of the full range. It is important to note that training and test datasets are much smaller here than in the validation of the GCWG's ability to reproduce other variables, resulting in larger variance in both test and training datasets. At all sites but Brasília, there are no significant differences between observed and generated means in at least half of the thirty simulations. Similarly, at least fourteen of the thirty simulations show no significant differences between the observed and generated variances at all sites but Brasília. Model performance is weaker in Brasília, where a mean of 12.5 simulations per month show statistically similar means, and only 8.8 for variance. A potential cause of the lower performance at Brasília may be the temperature correction sample size process outlined in Chapter 4, Section 4.2. Although Chapter 4 uses observational records as inputs to the model, issues present with the stochastic generation of precipitation and temperature discussed in Chapter 4 are still present here. Firstly, due to the even shorter training datasets here (in all variables), several half-months contain very few recordings for WD and DW days in Brasília. To provide a large enough sample size, the temperature for several half-months has instead been calculated

using several months of data (i.e., if the first half of January does not have enough a large enough sample of WD days, so WD days in December, January and February would be used instead). Temperatures make up half of the predictors to the solar radiation model and much like generating daily wind speed, any inaccuracies in reproducing the observed temperature accurately will also affect the resultant solar radiation model.

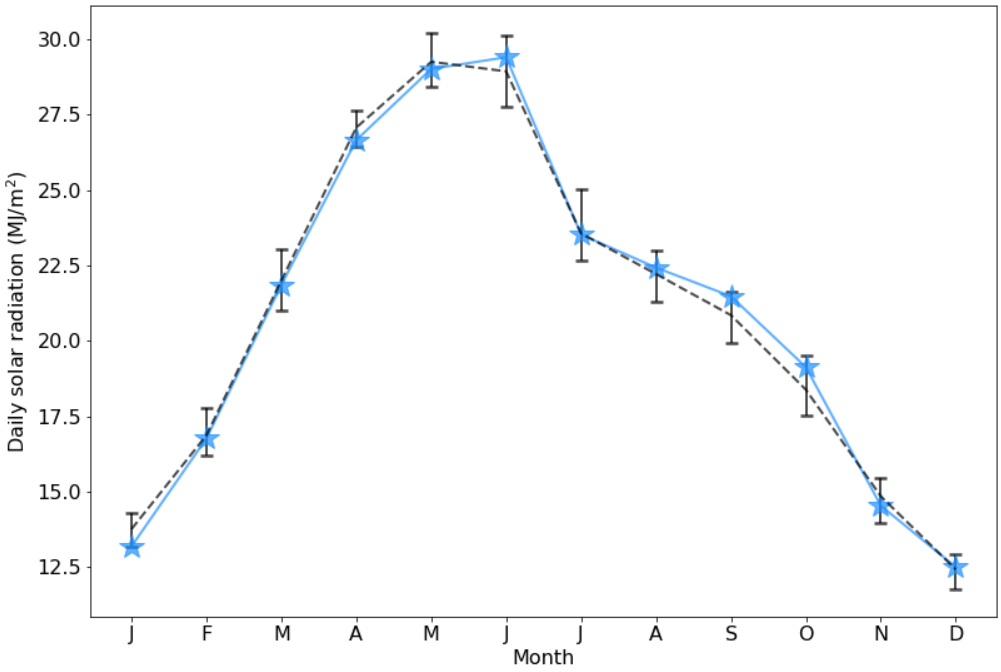
Another cause of the lesser performance may be due to the distribution of the regression residuals themselves at Brasília. The distribution of the solar radiation residuals is more normally distributed than other locations *prior* to the transformation into a clearness index. While the Gaussian-likeness has improved following the transformation, the improvement is much smaller than at other sites. Conversely, at Punta Arenas and Tallinn, residuals prior to the transformation are less normally distributed than Brasília, though following the transformation the residuals approximate a normal distribution extremely closely with much better agreement than Brasília (Figure 5.3). With residuals that are less normally distributed, the applicability of the regression is reduced.

Brasília has also been used instead of Key West to generate solar radiation following a first-order Markov-chain model to generate daily precipitation. In Chapter 3, it was determined that a third-order model most commonly represents precipitation in tropical regimes best for a range of metrics. Although both sites are sub-categorised as Aw (tropical savanna climate), Figures 3.9 and 3.12 show differences in the model order that best reproduces the observed interannual variability of precipitation occurrence (IVO) and amount (IVA) between the two locations, with different model orders performing better in different seasons at Brasília. This may result in slightly weaker precipitation performance in Brasília than at other sites, resulting in poorer performance in reproducing solar radiation.

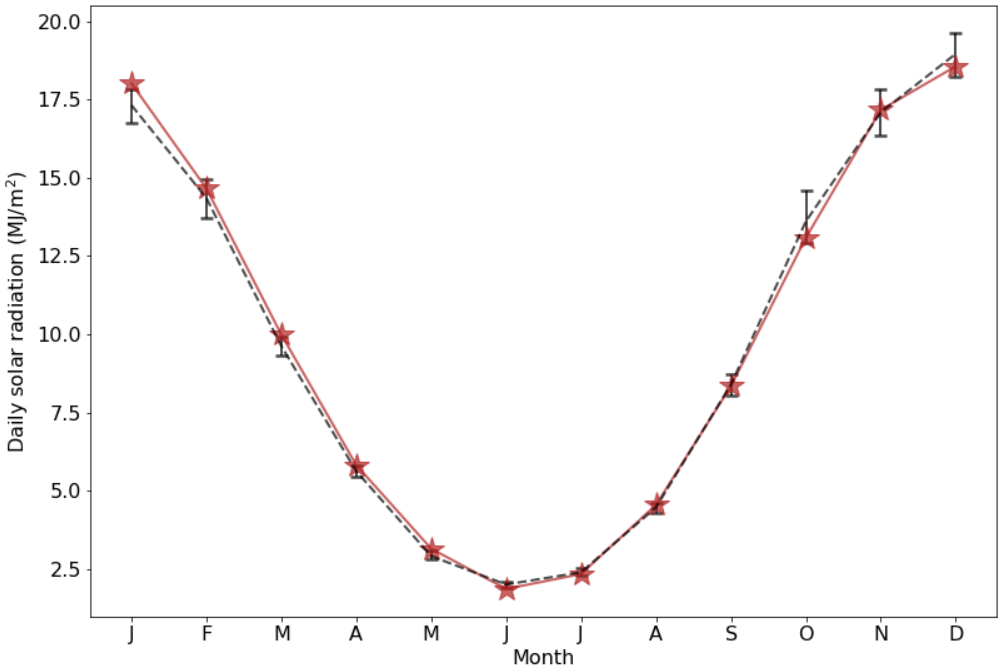
a)


Daily solar radiation (MJ/m²)
22
21
20
19
18
17
J
F
M
A
M
J
J
A
S
O
N
D
Month

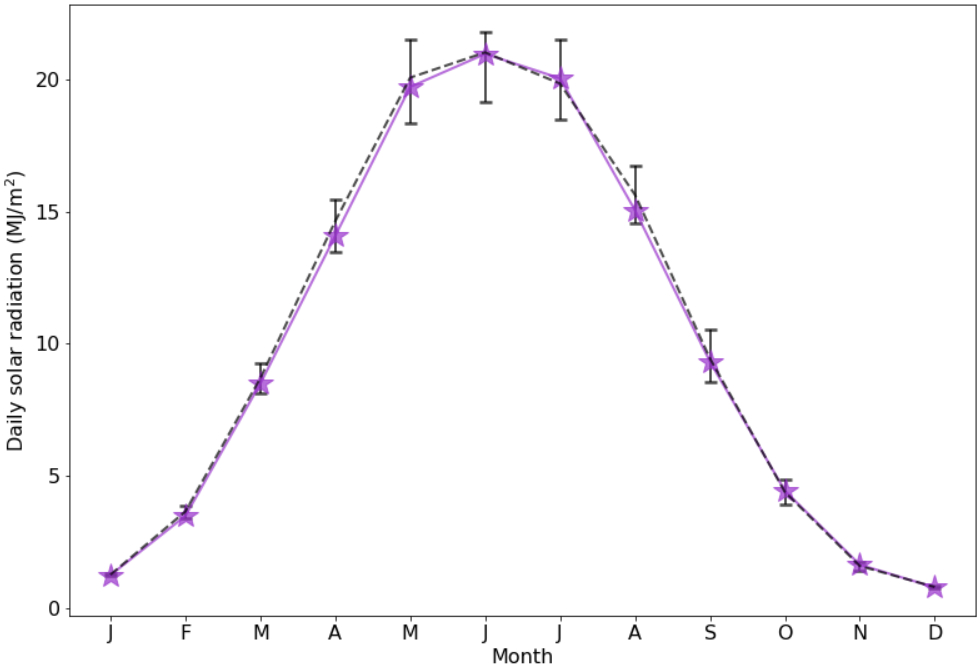
b)


Daily solar radiation (MJ/m²)
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
J
F
M
A
M
J
J
A
S
O
N
D
Month

c)


Daily solar radiation (MJ/m²)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
J F M A M J J A S O N D
Month

d)


Daily solar radiation (MJ/m²)
20
15
10
5
0
J F M A M J J A S O N D
Month

e)

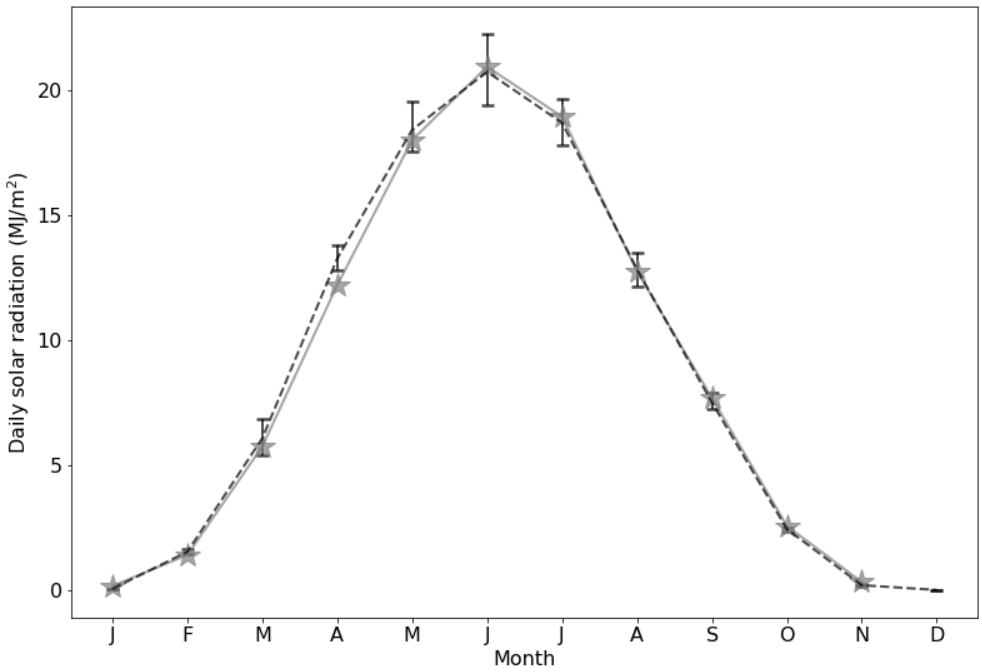


Figure 5.4 – Observed means in daily solar radiation (coloured stars) and the generated means in daily solar radiation (black with error bars), calculated monthly, with error bars representing the largest mean from all 30 simulations and the smallest mean (black, dashed) at a) Brasília, b) Fort Huachuca, c) Reykjavik, d) Punta Arenas, and e) Aasiaat.

Annual cycles in solar radiation have been preserved by the model accurately at all sites. As with wind speed, despite statistical differences between observed and generated means and variances, all sites show excellent practical significance, and once again supports the use of the GCWG as a tool to produce alternative realisations that are similar to the observed climate. This practical significance is particularly important at Brasília, Fort Huachuca and Aasiaat, where up to two thirds of the simulations show statistical differences. It is clear from Figure 5.4 that even in these locations, the annual cycles are captured exceptionally well at Fort Huachuca and Aasiaat. It is also reassuring to note the different shapes that the GCWG is able to capture; Brasília for example has a very weak annual cycle, whereas it is very strong in Aasiaat.

5.3 DAILY RELATIVE HUMIDITY

5.3.1 GENERATING DAILY RELATIVE HUMIDITY USING A MULTIPLE LINEAR REGRESSION MODEL

Relative humidity describes the ratio of how much water vapour is in the air to the maximum amount of water vapour the air could hold at any given temperature, expressed as a percentage. The primary reason for the inclusion of relative humidity in this GCWG is due to the role the variable plays in daily changes in reference evapotranspiration. With less available moisture in the air (therefore a lower relative humidity), evaporation is enhanced from open water sources and soils. Conversely, at higher relative humidities, evaporation rate is decreased. However, relative humidity is also an important variable in the study of wildfires, virology and human physiology (Williams et al., 2019). Changes to relative humidity can have direct and substantial impacts on physiological processes, including eye irritation and asthma, severe ill-effects on human comfort, the alteration of virus spread, alongside combinations of increased relative humidity and temperature reducing evaporative cooling of the body which can result in dangerous core body temperature levels (Gunawardhana et al., 2017; Wang et al., 2018).

Despite the significance of relative humidity in several fields, stochastically producing time series of daily relative humidity is largely undiscussed in literature in comparison to the primary, secondary, and other tertiary variables considered in this GCWG (Gunawardhana et al., 2017). Dew point is more frequently discussed, which can ultimately be converted into a relative humidity using daily temperature. Wallis and Griffiths (1997) proposed a model for generating dew point (alongside other variables) using a matrix of correlation coefficients between maximum and minimum temperatures, solar radiation, horizontal and vertical wind components at 1200 and 2400 LST, and dew point. This is similar to the process described by Matalas (1967) and Richardson (1981), but further developed to include more tertiary variables, assuming each variable is normally distributed and weakly-stationary.

However, to reduce errors that may arise through converting dew point temperature to relative humidity (using generated temperature values), a multivariate linear regression model will be used to *directly* generate relative humidity daily time series. Unlike the solar radiation to clearness index transformation, where only latitude is required, converting a generated time series of dew point temperature to relative humidity would also require the

generated daily temperature values. However, synthetic temperature values would also have been used to generate the residual dew point temperature series. By using the synthetic temperatures to generate a dew point temperature, and then in the conversion to a relative humidity, any errors in simulating temperature values will impact the resultant relative humidity twice.

Daily relative humidity (RH) values on day i are transformed into a residual series using monthly means and standard deviations. An autoregressive multivariate linear regression model,

$$RH_i = v + wP_i + xT_{x_i} + yT_{n_i} + zRH_{i-1} + \varepsilon \quad (5.9)$$

where w, x, y and z are coefficients of regression, v is a constant, and ε is an independent, standard, Gaussian error term, is used to generate a residual series. The constant and regression coefficients are fit monthly, resulting in 12 regressions. Following the simulation of a synthetic residual series, values are transformed back into relative humidities using the corresponding observed mean and standard deviation, calculated each month. Like clearness index, $RH\ is$ a constrained variable, though here with lower and upper limits of 0 and 100% respectively. If the model generates a value ≤1 or ≥ 99, values of 1 and 99% will be infilled respectively.

Daily temperatures, precipitation, and previous days' relative humidity have been chosen as predictors due to the relationships and significant correlations between relative humidity and these variables. Increased temperature results in the increased water-holding capacity of air, therefore increasing the maximum amount of water vapour that air can hold. Consequently, as temperature increases (if all other factors influencing relative humidity remain constant), relative humidity decreases. Conversely, an increase of precipitation will result in an increase in humidity due to more water available for evaporation. Similarly, higher humidities increase the likelihood of precipitation; if relative humidity reaches 100%, condensation and precipitation may occur. These known relationships are reflected by the Pearson correlations at each studied site between observed maximum/minimum temperature and relative humidity ranging from -0.20 to -0.80, precipitation and relative humidity between 0.20 and 0.40, and the lag-1 autocorrelation coefficient in all cases exceeding 0.60. All correlations are statistically significant, reassuring their use as predictors in the model.

5.3.2 ASSESSING THE MULTIPLE LINEAR REGRESSION MODEL PERFORMANCE AT REPRODUCING OBSERVED DAILY RELATIVE HUMIDITY

Relative humidity records are not widely available (Gunawardhana et al., 2017; Nicks and Harp, 1980). To address this issue, NASA POWER data at sites Brasília, Fort Huachuca, Punta Arenas, Tallinn and Aasiaat, introduced in Table 5.5, will be used once more. Relative humidities from the NASA POWER datasets are provided at a height of 2m. The same training and test data as in Section 5.2 has been used here (a 2:1 ratio of training to test datasets). For consistency with other sections, test data will be referred to as "observed", despite originating from model outputs. "Generated" (or "simulated") will once again refer to the data simulated using the parameters calculated from the training dataset.

Thirty 10-year simulations of daily relative humidity have been produced at each site, using generated maximum and minimum temperatures, precipitation, and previous days' humidity as predictors. Again, 10-year simulations are more representative of the length of test period (approximately 13 years here). The range of means in daily relative humidity produced by the thirty simulations is calculated and compared to the observed data for each month. Applying the same logic as before, if the observed dataset and the thirty generated datasets are realisations of the same climate system, it can be expected that the observed means will generally lie within the range of generated means.

The non-parametric Mann-Whitney U and Levene tests have been used to compare the observed and (thirty) generated means and variances respectively. Although some sites show approximate normal distributions in the observed and generated humidities, non-parametric tests will be used to ensure robustness where the distribution is non-normal, and to account for the constrained nature of the variable.

Table 5.7 – The number of observed daily means calculated monthly in relative humidity that lie within the range of 30 generated means out of a maximum 12. The mean numbers of simulations averaged over 12 months that show no statistical differences out of a maximum of 30 using the Mann-Whitney test to compare means, and Levene's test to compare variance.

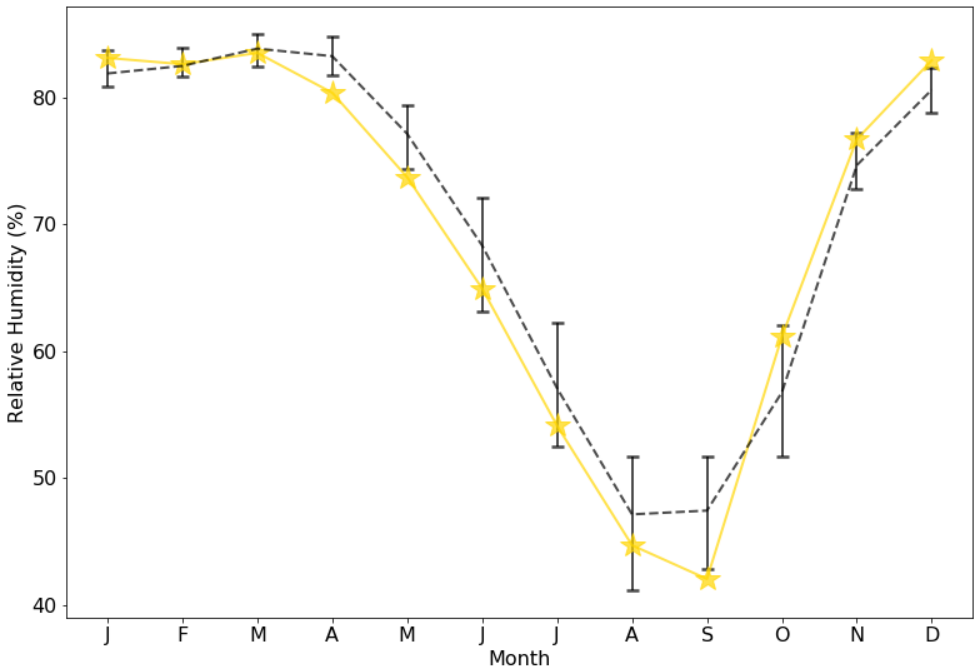
		MEAN SIMULATIONS PER MONTH SHOWING NO STATISTICAL DIFFERENCES	
LOCATION	**No. of observed means within range of generated**	**Mann Whitney**	**Levene's Test**
Brasília	8	20.7	22.3
Fort Huachuca	8	9.9	10.8
Punta Arenas	10	24.2	23.2
Tallinn	11	17.3	19.0
Aasiaat	10	19.1	15.2

At all sites, at least three quarters of observed means calculated monthly in daily relative humidity lie within the range of thirty generated means. At Brasília, Punta Arenas, Tallinn and Aasiaat, at last half of the simulations (averaged over all 12 months) show statistically similar means and variances. Here, Fort Huachuca shows the least accurate generation of relative humidities. An issue that has already been discussed in Section 5.2.3, it is important to note that the temperature sample size correction process introduced in Chapter 4, Section 4.2, may impact the resultant generated temperature and thus the tertiary variables. With shorter records used here than in Chapter 4, there are even more occurrences where the minimum sample size required to calculate bimonthly means and standard deviations in maximum and minimum temperature is not met. More half-months will therefore have the temperature parameters calculated using data not necessarily corresponding to the half-month (i.e., using three monthly means and standard deviations instead). Much like solar radiation, relative humidity is typically strongly correlated to (particularly maximum) temperature. Any inaccuracies in generating the temperature will therefore carry forward into the relative humidity.

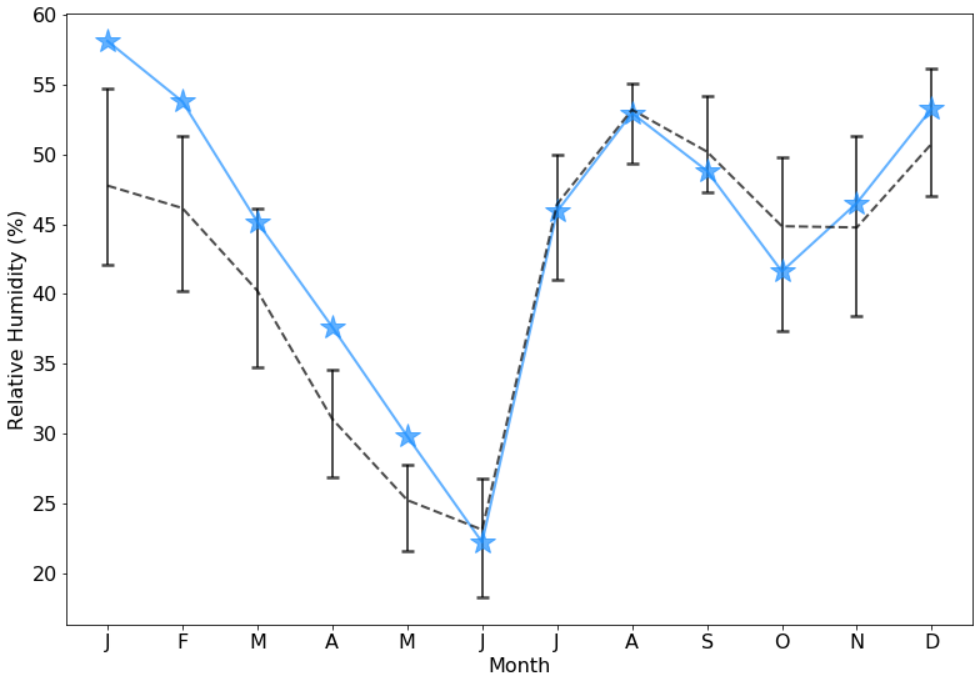
In Section 5.2.3 it was instead Brasília that showed the weakest agreement between observed and generated solar radiation, with the temperature sample size correction process

a suggested reason for the lesser performance. It could therefore be expected that Brasília would once again show the weakest performance at reproducing relative humidity, and Fort Huachuca at reproducing the solar radiation. However, performance may be dependent on the strength of the correlations at each site between the tertiary variable and temperature. For example, solar radiation is much more strongly negatively correlated to maximum temperature at Brasília (-0.8) than at Fort Huachuca (-0.4). Conversely, relative humidity is more strongly correlated with maximum and minimum temperature at Fort Huachuca (0.7 and 0.4 respectively) than at Brasília (0.4 and 0.0 respectively). This may result in inaccuracies in temperature generation affecting the variables differently depending on location.

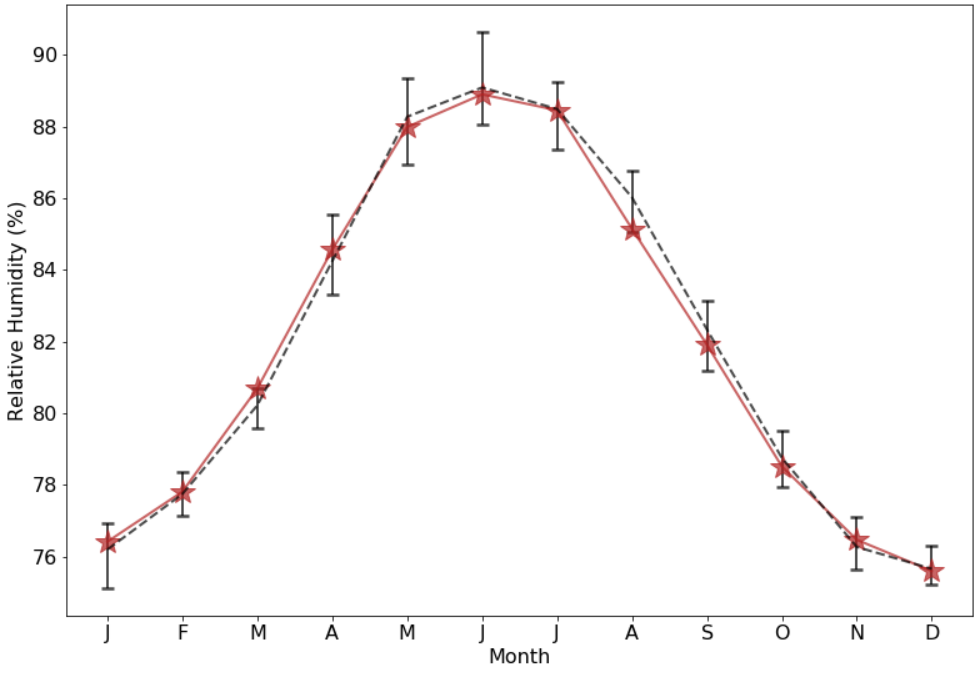
a)



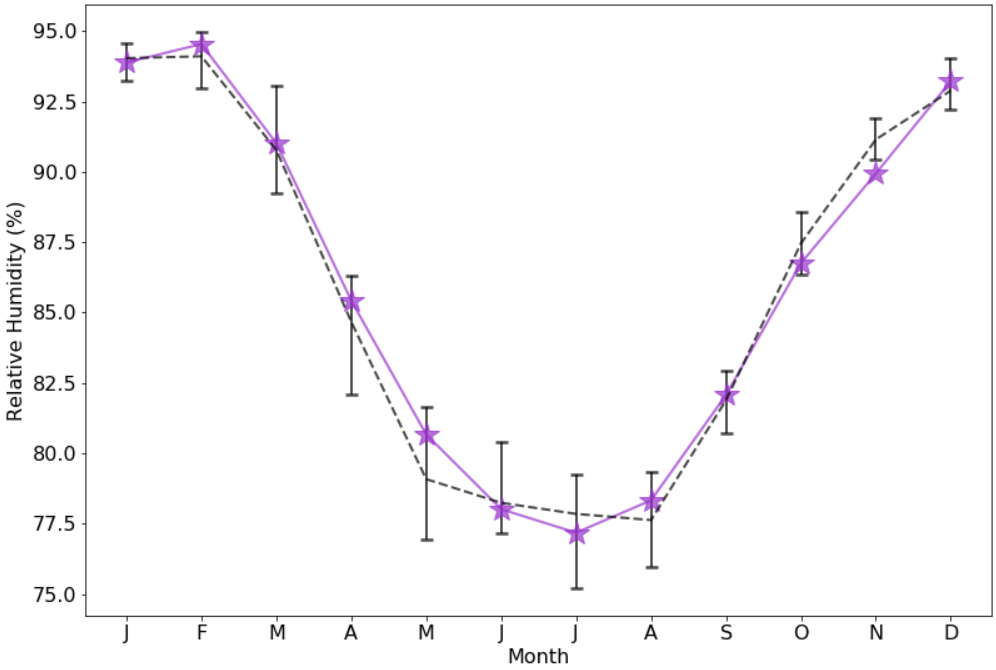
b)


Relative Humidity (%)
60
55
50
45
40
35
30
25
20
J
F
M
A
M
J
J
A
S
O
N
D
Month

c)


Relative Humidity (%)
90
88
86
84
82
80
78
76
J
F
M
A
M
J
J
A
S
O
N
D
Month

d)



e)

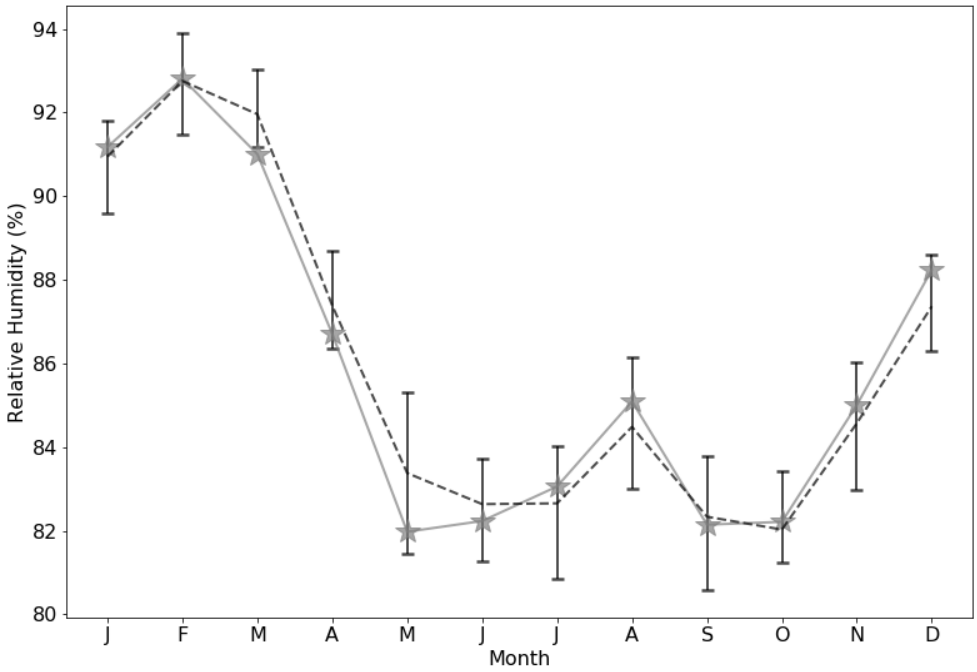


Figure 5.5 – Observed means in daily relative humidity (coloured stars) and the generated means in relative humidity (black with error bars) calculated monthly, with error bars representing the largest mean from all thirty simulations and the smallest mean, at from a) Brasília, b) Fort Huachuca, c) Reykjavik, d) Punta Arenas, and e) Aasiaat.

Annual cycles, shown in Figure 5.5, are reproduced accurately, with the range of generated means always lying within one standard deviation of the observed (not shown here) even in instances where there are a large number of statistical differences (e.g., Fort Huachuca). A range of shapes are present across the five sites. For example, Punta Arenas and Tallinn both have sinusoidal shapes but with peaks in June and February respectively. Aasiaat has a large peak in February, followed by a smaller peak in August. The shape of each distribution is reproduced with high accuracy. This is reassuring of the use of the multiple linear regression model to reproduce observed relative humidity, despite some of the statistical differences shown in Table 5.7. Despite statistically significant differences between the observed and simulated datasets, generated relative humidity remains meaningful for practical applications at Brasília, Punta Arenas, Tallinn, Aasiaat, and from July to December in Fort Huachuca.

Despite as many as 8 observed means lying within the range of generated means, only approximately one third of the simulations have statistically similar means and variances at Fort Huachuca. Although the temperature sample size correction process may contribute to this, a further cause of the inaccuracies in Fort Huachuca may be caused by large differences between the training and test datasets, demonstrated in Figure 5.6 (alongside Punta Arenas for comparison, where statistical differences between observed and generated simulations are far less prevalent). Differences of up to 10% are present between daily means (calculated monthly) in the training and test datasets at Fort Huachuca, conversely to a maximum difference of 1% in Punta Arenas. In Fort Huachuca, all five of the months where the observed means do not lie within the range of generated values are in the first five months of the year, where training and test datasets differ most strongly. Furthermore, the standard deviations in the test and training set are much larger in Fort Huachuca than at any other site. This potentially highlights an issue with using smaller datasets to determine GCWG input parameters.

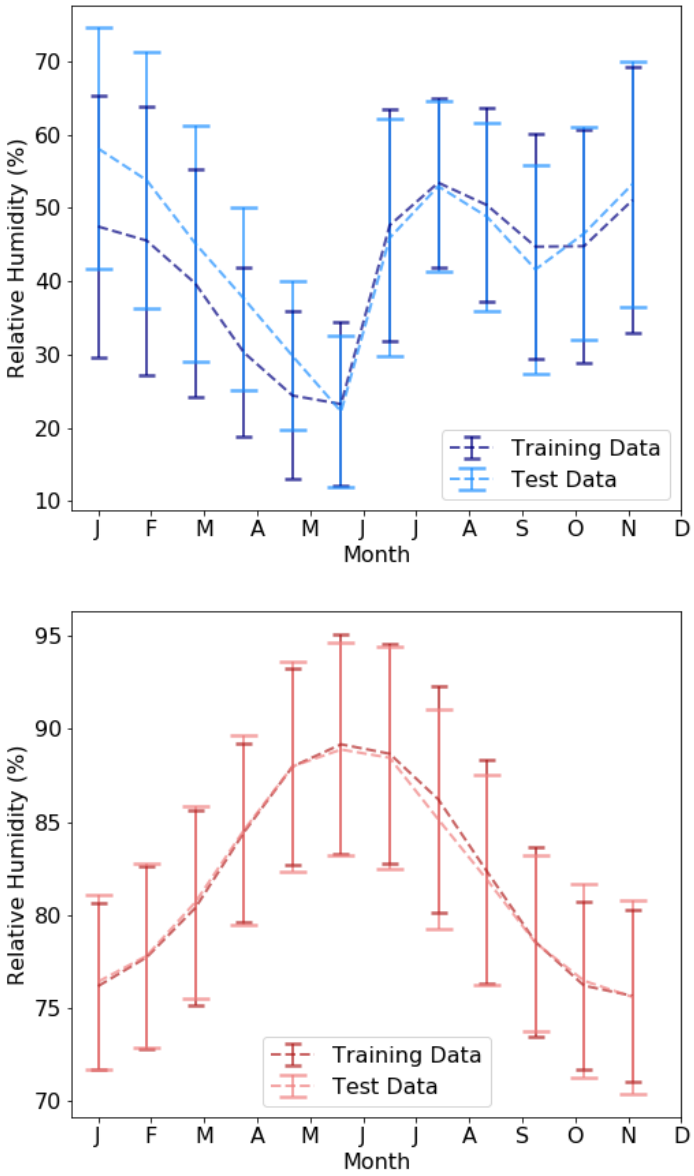


Figure 5.6 – Mean daily relative humidity in the training and test datasets for top) Fort Huachuca and bottom) Punta Arenas, calculated monthly.

Another potential cause for the weaker performance of the model may lie with using the linear regression model itself. Using the observed records for precipitation, maximum and minimum temperatures, and previous day's relative humidity, values for relative humidity have been produced using equation 5.7. The residuals (actual observed values minus predicted values) are shown in QQ-plots in Figure 5.7. If the residuals are normally distributed, they should show a straight line, and it is with confidence that the suggested multivariate linear regression is a suitable technique for modelling the variable. Residuals at Punta Arenas and Aasiaat resemble normal distributions the closest. These sites also have the fewest statistical differences between observed and generated relative humidities. Tallinn and Fort Huachuca show similar approximations to normal distributions, though to

a lesser degree than at Punta Arenas and Aasiaat. The QQ-plot for Brasília suggests the residuals of regression are not normally distributed; this may be a contributor towards the poorest statistical agreement with the observed values.

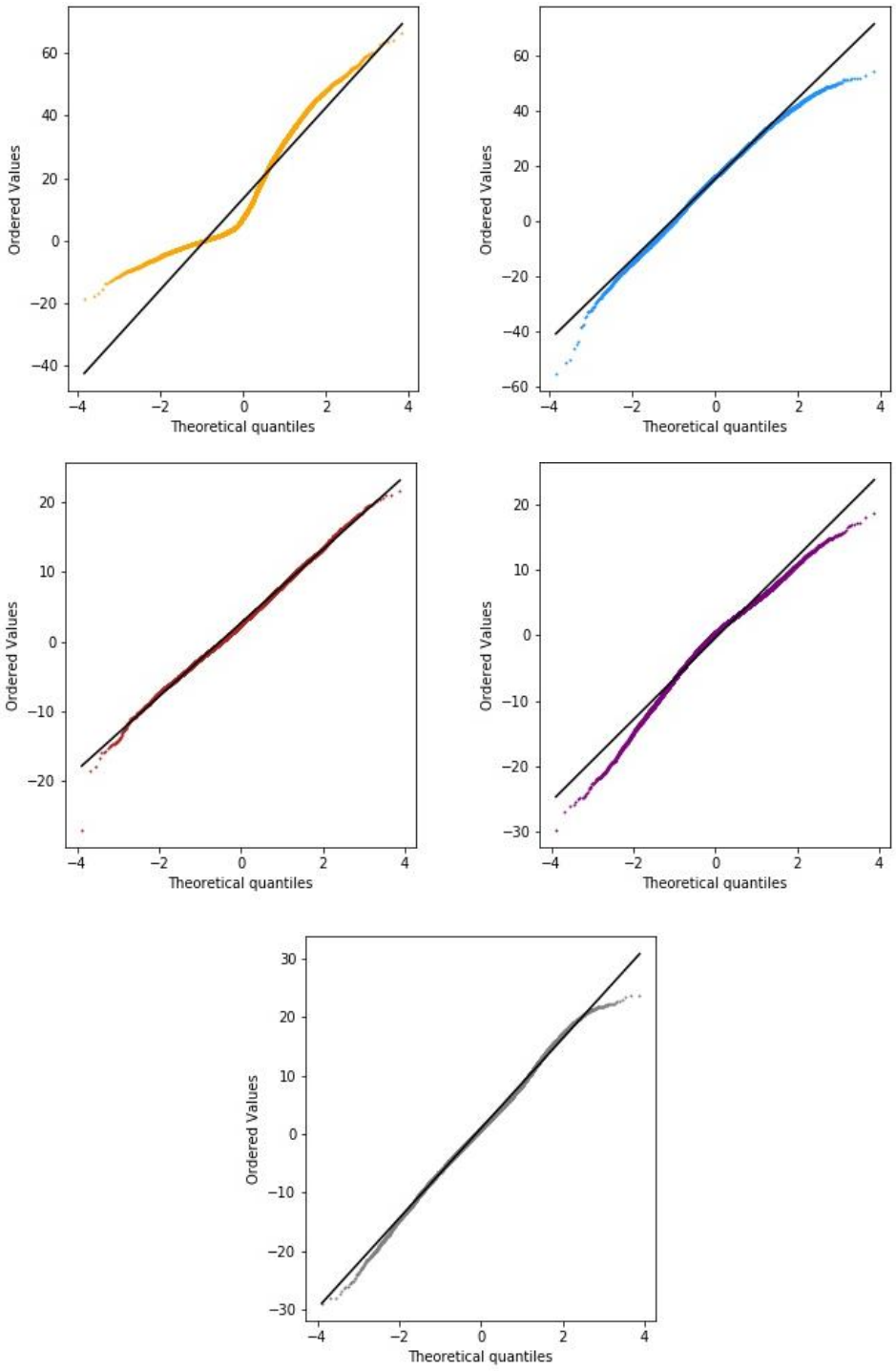


Figure 5.7 – QQ-plots showing the residuals of regression for top) Brasília and Fort Huachuca, middle) Reykjavik and Tallinn, and bottom) Aasiaat.

5.4 REPRODUCING OBSERVED INTER-TERTIARY VARIABLE CORRELATIONS

Models for simulating daily time series for each of the three tertiary variables have been introduced: a generalised linear model for wind speed which allows a non-normal distribution, a transformation to clearness index and multiple linear regression model for solar radiation, and a multiple linear regression model for relative humidity. 100 years of daily data has been generated for precipitation, maximum and minimum temperature, wind speed, solar radiation and relative humidity using the NASA POWER data split into a 2:1 ratio (training to test dataset). To determine whether correlations between the tertiary variables have been preserved through their individual dependencies on the primary and/or secondary variables, Pearson's correlation coefficient has been calculated for the observed (test) with the generated datasets between the tertiary variables.

Table 5.8 – Pearson's correlation coefficients from the observed and generated daily time series for relative humidity-wind speed and relative humidity-solar radiation

		Brasília	Fort Huachuca	Punta Arenas	Tallinn	Aasiaat
WIND SPEED	**Observed**	-0.33	0.07	-0.34	0.28	0.14
	Generated	-0.23	0.00	-0.06	0.26	0.10
SOLAR RADIATION	**Observed**	-0.30	-0.66	-0.71	-0.75	-0.40
	Generated	-0.35	-0.56	-0.65	-0.67	-0.35

The GCWG accurately reproduces correlations between relative humidity and wind speed, and relative humidity and solar radiation. Generated and observed correlations (not shown in Table 5.8) between solar radiation and wind speed are negligible (between -0.1 and 0.1) at all locations except Tallinn and Aasiaat, where observed and generated correlations are both weakly negative. Although the sites in this study show little correlation between the two variables, it is reassuring that the observed relationships between wind speed and solar radiation are reproduced accurately due to practical impacts relating to renewable energy sources. For example, Mohammadi and Goudarzi (2018) note potential anti-correlation between wind speed and solar radiation during very strong El Niño events across California. This relationship could be utilised in the planning of renewable energy

sources; where there is less solar energy potential, higher wind energy potential may complement the decrease.

A range of correlations between relative humidity and wind speed are present at the different sites. While all correlations are relatively weak, in some locations the correlation is positive (e.g., Tallinn and Aasiaat) and others negative (Brasília and Punta Arenas). In most cases, the GCWG reproduces the observed correlations with a good degree of accuracy. The largest differences between observed and generated correlations are present at Punta Arenas. Despite this, the observed correlation remains weakly negative (-0.34), with an extremely weak negative correlation in the generated dataset (-0.06).

Knowing that solar radiation is strongly correlated to temperature and relative humidity negatively correlated to temperature, the observed negative correlations shown in Table 5.8 between solar radiation and relative humidity are expected. It is reassuring the GCWG reproduces these correlations between relative humidity and solar radiation with the correct sign and magnitude at each site. Correlations vary from weakly to strongly negatively correlated, with the largest difference between any observed and generated correlation only 0.09. This demonstrates the GCWG's good ability at reproducing correlations between the tertiary variables accurately. Of the three inter-variable relationships, this is arguably the most important to accurately reproduce. Sunny weather (i.e., high solar radiation) in combination with low relative humidity is important for agricultural processes (such as cereal crop harvesting) and can increase forest-fire ignition risk (Ruosteenoja and Räisänen, 2013). While the two variables tend to be negatively correlated, in instances where both are higher than usual (alongside temperature) there is an increased risk to human mortality. Furthermore, in the absence of observational data, several methods rely on relative humidity (alongside other variables) to calculate daily solar radiation.

Punta Arenas and Brasília show strong (>-0.5) and weak (<-0.5) negative correlations, respectively, between relative humidity and solar radiation. Using the observed record and 100 years of generated data, boxplots have been produced for the relative humidities on all days where the solar radiation falls below the 10^{th} percentile and exceeds the 90^{th} percentile (with percentiles calculated from the observed data). The inverse relationship between the two variables is clearly shown, with higher average relative humidities on the days falling below the 10^{th} percentile in solar radiation, and lower average relative humidities on days exceeding the 90^{th} percentile in solar radiation. The range of relative humidities on days where the solar radiation falls below the 10^{th} percentile is also

accurately reproduced in both locations. While the mean relative humidity on days with solar radiation exceeding the 90th percentile is well reproduced, Figure 5.8 shows much larger ranges in Punta Arenas and (though to a lesser degree) Brasília in the generated datasets. It is important to reiterate that the sample size of the observed data is much smaller than the generated; hence many more outliers in the generated data shown in Figure 5.8. Despite this, the upper tail is generally reproduced well, with differences between the observed and generated values of around 4%. Due to the greater negative implications of simultaneous high solar radiation (and temperature) alongside high relative humidity, it is arguably more important to reproduce the upper tail than the lower tail effectively.

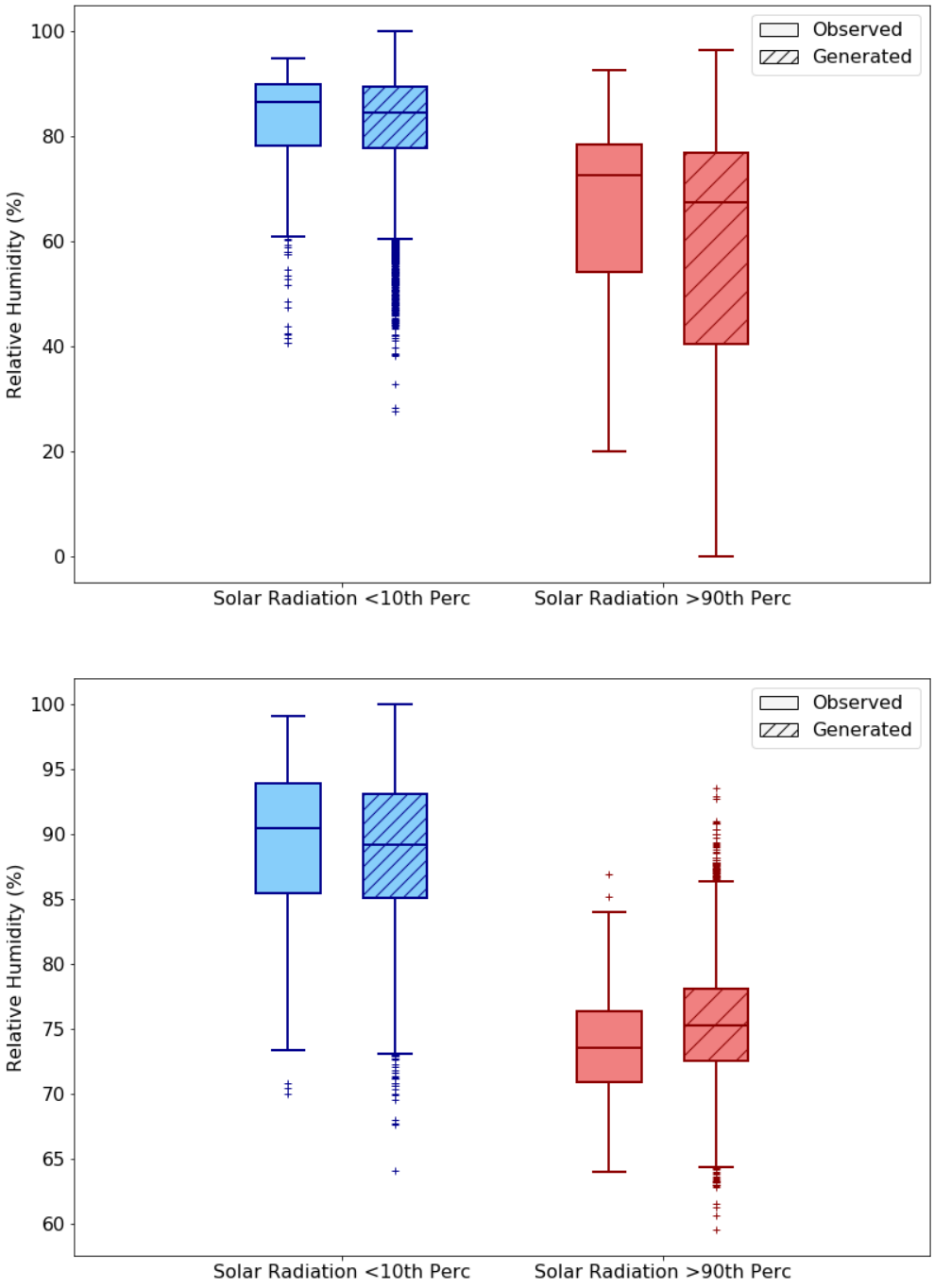


Figure 5.8 – Relative humidities on days where the solar radiation falls below the 10th percentile (calculated from the observed data) and exceeds the 90th percentile (calculated from the observed data) for top) Brasília and bottom) Punta Arenas.

5.5 CALCULATION OF REFERENCE EVAPOTRANSPIRATION USING OBSERVED AND GENERATED WEATHER VARIABLES

Evapotranspiration is an important agrometeorological variable in the hydrological cycle, describing evaporation and transpiration. A widely used variable in hydrological and agricultural models, evapotranspiration is used to understand local water budgets and cycles and is often used to assist in the development of infrastructure and irrigation planning and management. Furthermore, evapotranspiration is also an important variable in distinguishing the climatology of a region. Understanding precipitation at a site alone may not be enough to distinguish between a dry or moist climate. However, evapotranspiration is not a widely observed variable due to a wide range of complexities in taking measurements. These include flaws in experimental design, interpretation of results and even the measurement equipment itself (Allen et al., 2011).

Reference evapotranspiration describes the rate of evaporation and transpiration from a reference surface (that is not short of water), and was initially developed to study the atmospheric evaporative demand independently of crop type, development, and management practices (Allen et al., 2006). Reference evapotranspiration is a purely climatological variable and can be calculated using meteorological variables only. The main meteorological variables that affect reference evapotranspiration are solar radiation, air temperature, humidity, and wind speed. Using the outlined models for daily wind speed, solar radiation and relative humidity, alongside the models described in Chapter 4 for generating maximum and minimum temperature, daily reference evapotranspiration can be calculated.

Reference evapotranspiration has been calculated with the ETo Python package (documentation accessible from https://eto.readthedocs.io/en/latest/intro.html) using the FAO Penman-Monteith method. This package is also able to estimate reference evapotranspiration in the absence of several of the tertiary variables, making it applicable at sites that have not recorded variables such as solar radiation or humidity. Reference evapotranspiration will be calculated at Punta Arenas and Brasília. These sites have been chosen due to the excellent performance of the GCWG in temperate climates and the weaker (though still good) performance in tropical climates. NASA POWER data for Punta Arenas and Brasília has been split into the same 2:1 ratio (training to test datasets) as before.

Reference evapotranspiration is first calculated from the observed (test) dataset. References to “observed” evapotranspiration describe reference evapotranspiration calculated from the observed (test) dataset. Similarly, “generated” (or “simulated”) reference evapotranspiration refers to reference evapotranspiration calculated from the generated variables, using parameters calculated from the training dataset. Thirty 10-year simulations of maximum and minimum temperature, mean wind speed, solar radiation and relative humidity have been produced using the training dataset and hence daily reference evapotranspiration calculated. Thirty 10-year simulations (as opposed to ten 30-year simulations as in Chapter 4 and Section 5.1) have been used to once again better reflect the shorter length of the test dataset.

Figure 5.9 shows excellent practical agreement between generated and observed monthly means in daily reference evapotranspiration at Punta Arenas. In most months, generated evapotranspiration at Brasília shows good practical agreement with the observed. Weaker performance at Brasília than Punta Arenas is expected, due to weather stations in tropical regimes consistently showing less accuracy in reproducing (in particular, tertiary) weather variables than those in temperate locations (such as Punta Arenas). It is important to reiterate that the test dataset only spans 10 years. This is not a particularly long record, especially when considering derived quantities such as reference evapotranspiration that depend upon several climatic variables. This may be responsible for some of the differences between the observed and generated datasets. However, the annual cycle at both locations is captured accurately, with the shape, timing and amplitudes of the generated reference evapotranspiration closely aligning with the observed. This is reassuring of the proficiency of the GCWG.

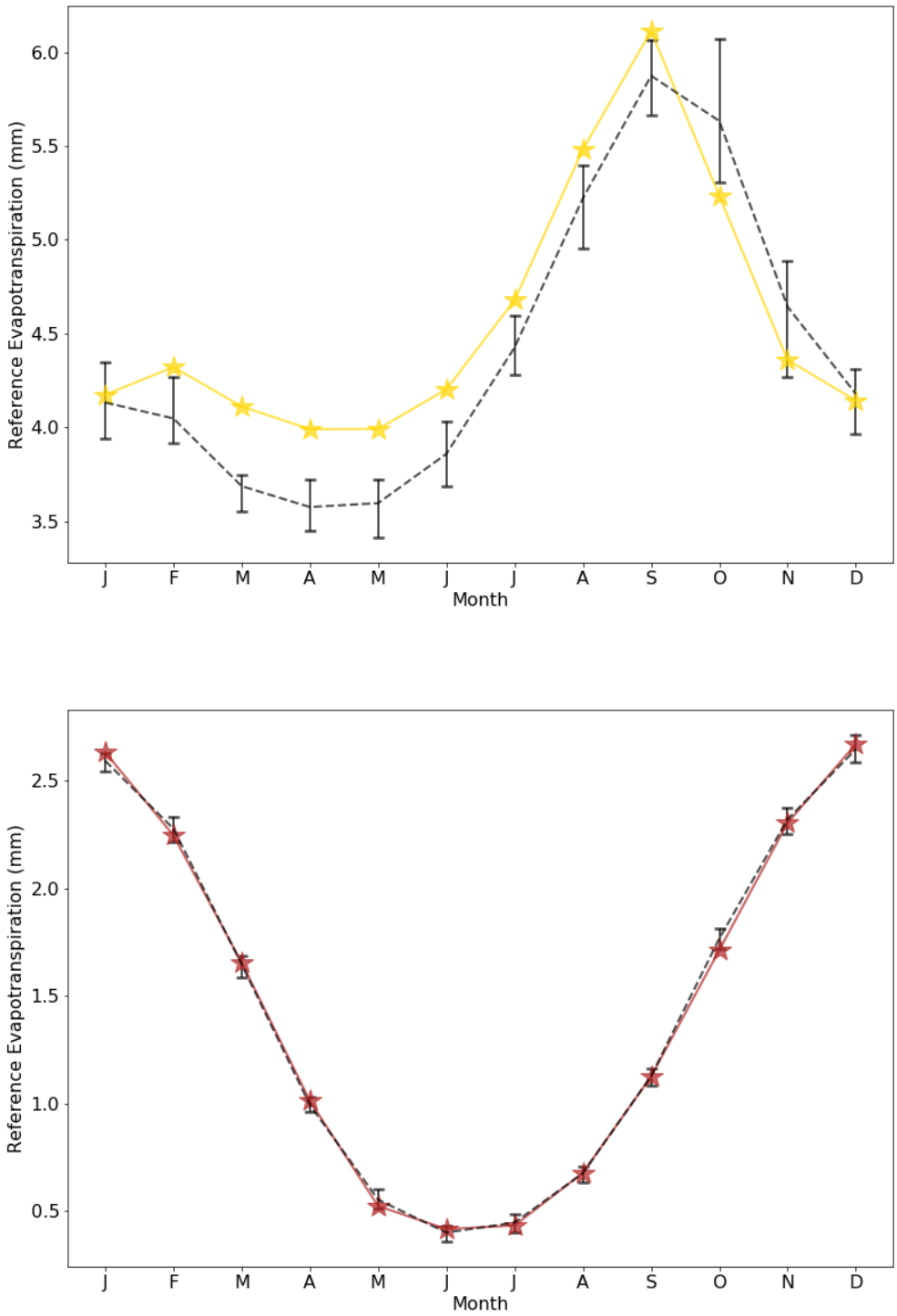


Figure 5.9 – Observed mean daily evapotranspiration (coloured stars) and generated mean daily evapotranspiration calculated from generated variables (black with error bars), with error bars representing the largest mean from all 30 simulations and the smallest mean (black, dashed) at top) Brasília and bottom) Punta Arenas.

At Punta Arenas, the observed reference evapotranspiration lies within the range of generated values in all months, with at least 60% of the simulations showing statistically similar means and variances. The GCWG does, however, produce several months where the range of generated monthly means in daily reference evapotranspiration does not capture the observed at Brasília. This is reflected by only a small proportion of statistically similar means (approximately 17%).

Table 5.9 - The number of means in daily reference evapotranspiration calculated from the observed (test) dataset that lie within the range of 30 means calculated from generated variables out of a maximum 12. The mean numbers of simulations averaged over 12 months that show no statistical differences out of a maximum of 30 using the Mann-Whitney test to compare means, and Levene's test to compare variance.

		MEAN SIMULATIONS PER MONTH SHOWING NO STATISTICAL DIFFERENCES	
Location	**No. observed means within range of generated**	**Mann Whitney**	**Levene's Test**
Brasília	3	5.1	10.9
Punta Arenas	12	18.1	19.8

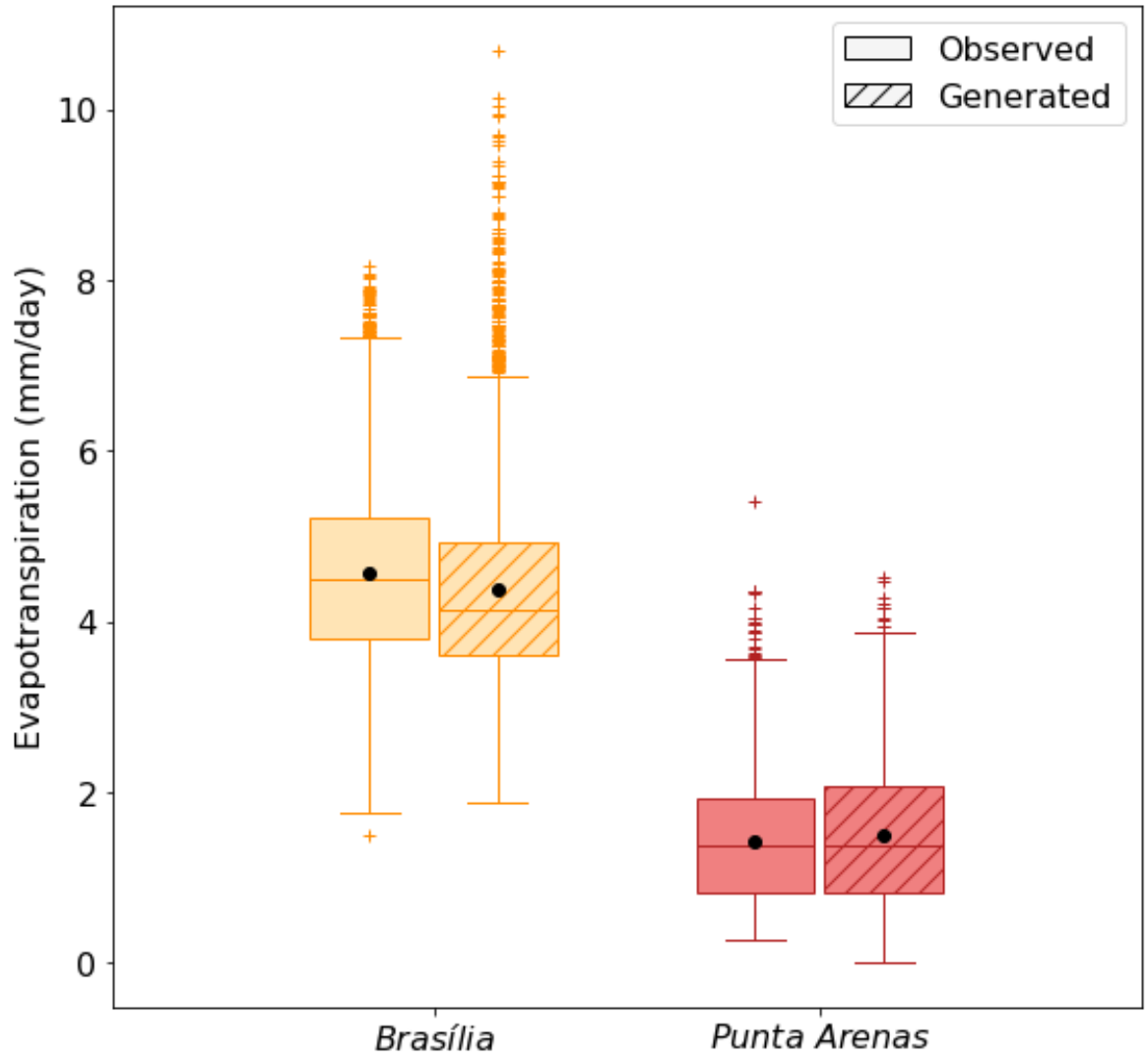


Figure 5.10 – Evapotranspiration calculated using observed and generated variables for Brasília and Punta Arenas. One of the thirty simulations is shown here.

The overall distributions of all daily values show few differences between the two datasets at both sites (Figure 5.10). Observed reference evapotranspiration at Brasília shows a larger median, mean, and range than Punta Arenas, which is accurately captured by the GCWG. There are many more outliers in the upper tail at Brasília, indicating a more skewed distribution, once again accurately captured by the generated data. This demonstrates the capability of the GCWG to capture a range of distributions in daily reference evapotranspiration. Figure 5.9 and Figure 5.10 highlight the merits of not only statistically testing the data, but to also visualise it. Although there are several statistical differences, the generated time series is generally showing a realistic realisation of the observed climate. This reiterates once more that with larger sample sizes, statistical tests are more likely to pick up on small differences between the datasets. However, it is clear that for many

applications, the GCWG will sufficiently simulate an accurate time series of weather variables that can be used to calculate a realistic realisation of the reference evapotranspiration.

5.6 CHAPTER 5 SUMMARY

Models for generating daily time series for each of the tertiary variables (mean wind speed, solar radiation and relative humidity) have been introduced. A requirement of each model was to incorporate some dependency on a primary (precipitation) or secondary (maximum and minimum temperature) variable. This is because only the primary and secondary variables will be scaled with increasing global mean surface temperature in Chapters 6 and 7. Changes to the tertiary variables as a function of GMST could therefore be incorporated through the changes in precipitation and temperature (though not investigated as part of this research).

A generalised linear model (GLM) to simulate daily mean wind speeds has been demonstrated, using lag-1 autocorrelation, maximum and minimum temperatures as predictors with a gamma distribution and a log-link function. The model shows good statistical agreement and excellent practical agreement with observed data at each of five sites (Key West, Fort Huachuca, Reykjavik, Tallinn, and Aasiaat). Inter-variable correlations are generally well reproduced, though at most sites correlations are weak. A multivariate linear regression model has been used to produce solar radiation and relative humidity time series using lag-1 autocorrelation coefficients, maximum and minimum temperatures, and precipitation as predictors. NASA POWER data (which combines satellite and model data) has been used in place of weather station observations due to a lack of high quality, multivariate records, using Brasília in place of Fort Huachuca and Punta Arenas in place of Reykjavik. A transformation to and from a clearness index is recommended for the solar radiation model. This is to ensure that regression residuals are approximately normally distributed. Generated solar radiations and relative humidities show good statistical agreement with observed data at most sites, with Brasília and Fort Huachuca showing the least accuracy. Despite this, annual cycles are accurately reproduced at each site, and generated data shows excellent practical agreement with the observed at most sites.

Temperatures and tertiary variables have been used to calculate daily reference evapotranspiration with the ETo Python package, using the FAO Penman-Monteith equation. Generated and observed reference evapotranspiration time series show excellent

agreement at Punta Arenas, and to a lesser degree, at Brasília. This is an expected result; in all comparisons between generated and observed data, the GCWG tends to perform better in temperate locations (i.e., Reykjavik and Punta Arenas) than at tropical sites (i.e., Key West and Brasília). However, at both sites, the generated time series continues to demonstrate the GCWGs ability to reproduce sensible realisations of the climate, with good practical agreement between generated and observed datasets.

This chapter has demonstrated the strengths of this simple GCWG, utilising a limited number of parameters for each tertiary variable, whilst maintaining some dependency on the primary and/or secondary variables. Furthermore, this dependency has accurately captured the correlations between the tertiary variables without the inclusion of additional predictors. Using weather stations that cover a range of climatic regimes provides confidence of the globally representative nature of the GCWG. Generated time series for each tertiary variable shows good to excellent practical agreement with observations, supporting the skill of the GCWG at reproducing realistic realisations of a location's actual climate.

6 PATTERN SCALING THE PRIMARY VARIABLE WEATHER GENREATOR PARAMETERS

Chapters 3 to 5 saw the development of the **G**lobally **C**alibrated stochastic **W**eather **G**enerator (GCWG), validated using a global network of gridded weather station observations, and a selection of observations representative of the five Köppen climate regimes (tropical, dry, temperate, and polar) (Köppen, 1900). In this chapter, precipitation parameter response to external climate forcings will be diagnosed from two General Circulation Models (GCMs) as a function of increasing global mean surface temperature (GMST). To reiterate, these parameters include monthly P_{00}, P_{11} (the probability of a dry day given that the previous day is dry, and the probability of a wet day given that the previous day is wet, respectively), α and β (wet-day shape and scale, respectively).

Responses will be diagnosed from the IPSL-CM6A-LR and ACCESS-ESM1.5 GCMs and have hence been compared to each other and wider literature. Following the diagnosis of primary parameter response, parameters calculated directly from weather station observations will be perturbed to generate precipitation time series that may be representative of a range of global warming levels (GWLs) whilst reducing GCM-induced errors at representing local-scale climate arising from their course resolution and parameterisation of several sub-GCM grid-scale processes (such as cloud microphysics and local forcings).

Section 6.1 recaps the Markov-chain gamma-distribution model discussed in Chapters 2, 3 and 4, and reiterates some of the literature discussed in Chapter 2 pertaining to pattern scaling. This approach has been taken to facilitate the preparation and submission of a manuscript titled “Pattern Scaling the Parameters of a Stochastic Weather Generator to Represent Changes in Climate: A Demonstration for Primary Variables” to *Climatic Change* in due course. The GCWG is therefore not referred to explicitly in the manuscript, though the parameters and models described are those used in the GCWG.

6.1 PATTERN SCALING THE PARAMETERS OF A STOCHASTIC WEATHER GENERATOR TO REPRESENT CHANGES IN CLIMATE: A DEMONSTRATION FOR PRIMARY VARIABLES

6.1.1 INTRODUCTION

General Circulation Models (GCMs) are the most sophisticated and widely utilized tools used to study the climate's response to external forcings. However, GCM outputs may not be directly suitable for impact assessments due to their coarse resolution (Mitchell, 2003) or, in cases where the resolution *is* suitable (e.g., the recent High Res MIP (Haarsma et al., 2016)), limited numbers of simulations representing only a few scenarios have been run, therefore not fully capturing unforced variability. Several methods have been discussed in the literature to address the mismatch between GCM area-averaged outputs and the local-scale resolutions required for hydrological, agricultural, and ecological assessments, including statistical and dynamical downscaling techniques (Maraun et al., 2010). Stochastic weather generators are a tool that have been used to statistically downscale GCM outputs and produce long time-series for a range of weather variables at a local-scale (Jones et al., 2016; Wilks, 2010, 1999a). The weather generator parameters are typically estimated using a local observed record, and then projected changes in the parameters (to represent a possible future climate) are diagnosed from GCM simulations for a specific scenario and time period. While this addresses the resolution mismatch, it can only be applied to scenarios or time periods simulated by the GCM. Although a defined range of climate scenarios have been considered by the sixth Coupled Model Intercomparison Project (CMIP6) GCMs covering mitigated, weak, and high forcing scenarios, there is no quantification of each scenario's likelihood (Tebaldi et al., 2021; van Vuuren et al., 2011) and therefore the study of a handful of scenarios alone does not capture the range of uncertainty.

Pattern scaling (PS) addresses the issue of limited simulations. Originally developed to create transient projections from equilibrium responses of a GCM to doubled CO_2 concentration, PS has more recently been used to construct projections for scenarios and time periods where fully coupled projections are not available, and to better understand uncertainties associated with inter-model variability (Lynch et al., 2017; Osborn et al., 2016; Santer et al., 1990; Tebaldi and Arblaster, 2014). It is an approach that has been applied within the IPCC Sixth Assessment Report (IPCC, 2021a). A computationally inexpensive technique, PS assumes a linear relationship between local climate change (e.g., within a grid

cell), and the global mean surface temperature (GMST) change. It uses this linearisation to approximately emulate the projections from a GCM. Osborn et al. (2016, 2018) argued that this linear approximation does not need to be perfect (it is an approximation after all) but that, to be useful, the errors arising from this approximation should be small relative to the other sources of uncertainty (e.g., small relative to the spread of results in an ensemble of different GCMs). A spatial pattern of this relationship is derived from transient GCM data expressed in a normalised way, for instance, as change per degree Celsius of GMST change. The resulting pattern can hence be scaled by a specific GMST change (the 'scaling factor'), or a time series of GMSTs, to emulate the climate at a specified global warming level (GWL).

The raw GCM data used to define the spatial patterns are usually composed of monthly, or seasonal, means and are typically time filtered (e.g., a multi-year running average) to clarify the external forcing signal. As a result, further treatment of the PS output is needed to investigate extreme events at different time scales (e.g., daily events), and variability responses (e.g., inter-annual variability) to forcings are not routinely accounted for despite their importance (Katz and Brown, 1992; Seneviratne et al., 2021) – although variability parameters can be scaled, independently, alongside the mean of the climate variable itself, e.g., Osborn et al. (2016).

Stochastic weather generators (SWGs) can be used to downscale GCM outputs and are frequently used to produce time series for a range of weather variables in the absence of high quality, consistent records, for use in hydrological and agricultural risk assessments (Semenov et al., 1998). Long, synthetic time series also allow for more accurate estimation of the probabilities of extreme events, such as long wet or dry spells, where observed records may be too short. SWGs are computationally inexpensive and can produce several realisations of the climate that they have been calibrated by. However, unlike GCMs, SWGs do not represent the physical processes that determine the climate change induced by external forcing. This means that they cannot, on their own, be used to produce realisations of the future climate without some perturbation. Instead, changes in their parameters need to be provided by an external source (such as diagnosing changes from GCM or Regional Climate Model (RCM) simulations). Note that other approaches to statistical downscaling are conditioned on large-scale variables such as atmospheric temperature, humidity and circulation patterns, and these predictor variables also need to be diagnosed from GCM or RCM simulations.

This study, therefore, applies PS to the first-order transition probabilities and wet-day gamma parameters of a SWG (defined in Section 6.1.2) with the aim of generating long time-series emulating presently un-simulated scenarios (i.e., representative of presently unmodelled GCM - scenario combinations), whilst also addressing the issues regarding coarse GCM resolution and reducing the impact of GCM-induced errors in the projection of local scale climates. To demonstrate this approach, global gridded transition probability and gamma parameter responses to external forcings such as increased greenhouse gas emissions will be diagnosed using the IPSL-CM6A-LR GCM. GMST is used as a metric to quantify the changes to regional climates and therefore, responses will be discussed relative to changing GMST (i.e., how do these parameters scale with GMST?). The implications of the response patterns to resultant precipitation will be interpreted and hence compared to observed and projected trends in precipitation with changing GMST. Finally, the response patterns will be used to perturb transition probabilities calculated from weather station observations at Santarém, Brazil and Reykjavik, Iceland, and generate precipitation occurrence under three global warming levels (GWLs); 1.5, 2.0 and 4.0°C.

6.1.2 DATA AND METHODS

6.1.2.1 SELECTION OF CLIMATE SCENARIOS AND GCMS

A wide range of future forcing pathways are covered by the CMIP6 ScenarioMIP "Tier 1" experiments, from strongly mitigated (i.e., SSP1-2.6) to high end forcing (i.e., SSP5-8.5) scenarios (O'Neill et al., 2016; Riahi et al., 2017; Tebaldi et al., 2021). Three of the four experiments (SSP1-2.6, SSP2-4.5 and SSP5-8.5) are analogous to CMIP5 experiments, studying similar levels of radiative forcing by 2100 (corresponding to RCPs 2.6, 4.5 and 8.5). SSP3-7.0 has been constructed to fill a gap between medium (SSP2-4.5) and high (SSP5-8.5) pathways (Tebaldi et al., 2021). Weaker forcing or strongly mitigated scenarios are known to produce less well-defined response patterns, which, when extrapolated, lack characteristics present in higher forcing responses. Strongly mitigated scenarios are also thought to show non-linear precipitation changes (Wu et al., 2010) with the spatial characteristics of warming changing as the deep ocean temperature reaches equilibrium (May, 2012). Whilst it is possible to extrapolate patterns from low-end warming scenarios, it has been recommended that patterns should instead be diagnosed from strong forcing scenarios (Mitchell, 2003) and it has been shown that response patterns diagnosed excluding high-end scenarios may perform poorly beyond certain GWLs (Osborn et al.,

2018). It is because of this that, of the four widely available Tier 1 ScenarioMIP scenarios, only the high-end forcing scenarios, SSP3-7.0 and SSP5-8.5, will be used to produce climate response patterns. SSPs 3-7.0 and 5-8.5 both result in higher GMST increases by 2100 than other SSPs. This will reduce inaccuracies that may arise when scaling SWG input parameters changes from low-end warming scenarios. Furthermore, the pooling of two SSPs provides a larger sample size, thus improving the signal-to-noise ratio of the forced response against internal climate variability (Osborn et al., 2018).

SSP5 represents the high-end of future pathways and describes a future with high mitigation challenges with fossil-fuel dominated development. Investments in human and social capital are increased, with a greater emphasis on resource and energy intensive lifestyles globally. In turn, there are higher energy demands with high exploitation of fossil-fuel resources and delays in uptake of climate mitigation policies (Riahi et al., 2017) resulting in an external forcing of $8.5 Wm^{-2}$ by 2100. Alternatively, SSP3 describes a future of resurgent nationalism with high societal vulnerability. SSP3-7.0 is an update on RCP7.0 and represents a medium-to-high pathway. This scenario combines substantial land use changes (including in global forest cover), high near-term climate-forcing emissions, and the largest population growth of all scenarios (12.6 billion by 2100) (Riahi et al., 2017).

While any GCM could have been chosen because this is a proof-of-concept rather than putting forward a new climate change projection, there are nevertheless some reasons why IPSL-CM6A-LR was chosen. First, it has an initial condition ensemble of runs for each of the SSPs considered here (rather than a single run) which allows more accurate climate change patterns to be diagnosed when the signal-to-noise ratio is small (Mitchell, 2003). Here, four realisations of both SSPs have been used to diagnose the responses. Second, it has a moderate-to-high climate response: the GMST change by 2100 slightly exceeds the "very likely" ranges assessed by AR6 (IPCC, 2021) and its equilibrium climate sensitivity is 4.52°C, which lies towards the upper end of AR6 assessed range. These slightly stronger GMST changes will also improve the signal-to-noise ratio. These advantages will help to make the diagnosed change patterns more accurate, so that differences between PS emulated and GCM projections at a specified GWL can be investigated. The inter-model spread of responses is likely to be larger (Osborn et al., 2016; Tebaldi and Arblaster, 2014) but quantifying that is not the focus of this study.

Four initial-condition ensemble members with differing realisations of SSP5-8.5 and SSP3-7.0 ScenarioMIP simulations *and* historical CMIP6 simulations will be used to

diagnose the spatial response of the first-order Markov probabilities. Historical simulations run from 1850 – 2015, and ScenarioMIP simulations from 2015 – 2100. Each realisation has unique, internally driven components that are independent of the forced external climate signal which remains constant between realisations. The components of unforced variability in the ensemble members are independent of each other (Jones et al., 2011), thus it can be expected that pooling four ensemble members will reduce the size of the internal, unforced variability by half relative to using a single member (Osborn et al., 2018). Increasing the number of ensemble members beyond four is not computationally efficient for the improvement in the error; to further reduce the error by another half would require sixteen ensemble members (with many GCMs not simulating this number of realisations).

Historical simulations have been concatenated with the corresponding ScenarioMIP realisation to produce a temporally consistent series from 1850 to 2100. For example, the first realisations of SSP3-7.0 and SSP5-85 (r1i1p1f1) have been concatenated with the first realisation of the historical simulation (r1i1p1f1). The concatenated time series will hereafter be referred to as historical+SSP3-7.0 and historical+SSP5-8.5. The four historical+SSP3-7.0 and four historical+SSP5-8.5 ensemble members will be pooled, resulting in eight temporally consistent time series from 1850-2100.

6.1.2.2 DETERMINING THE SPATIAL RESPONSE PATTERNS

There are several different methods of generating daily precipitation time series, including Markov-chain gamma-distribution models, series-type empirical distributions (Semenov and Barrow, 2002), Markov renewal processes (Foufoula-Georgiou and Lettenmaier, 1987) and the Neyman–Scott Rectangular Pulse method (Cowpertwait, 2004). A summary of the approaches is provided in Chapter 2, Section 2.3. Here, due to the simplicity of the parameters lending themselves suitably to pattern scaling, a Markov-chain gamma-distribution model will be used here.

First-order, 2-state Markov chains are frequently used to stochastically simulate daily precipitation occurrence (Richardson and Wright, 1984), where two-state refers to the precipitation status of a day i.e., either wet or dry, with “wet” here defined as a precipitation amount of at least 0.1mm. Following the generation of a binary number sequence representing wet or dry days, Markov-chain weather generators typically attribute an amount of precipitation to a wet day independently from a two-parameter gamma distribution, fitted monthly. While higher-order models have merits first order Markov chains are the most

commonly used in literature (Jimoh and Webster, 1996; Schoof and Pryor, 2008), and in a global comparison of zeroth, first, second, and third model-order performance, Wilson Kemsley et al. (2021) showed that first-order models are the most accurate at reproducing observed distributions of wet-spell length and outperformed all others using the Bayesian Information Criterion (BIC) globally. For this reason, first-order Markov chains only will be considered here.

The response of first order transition probabilities and wet-day gamma parameters to changing GMST will be diagnosed from IPSL-CM6A-LR. First order transition probabilities are defined as

$$P_{ij} = \frac{n_{ij}}{\sum_{m=0}^{1} n_{im}} \tag{6.1}$$

where i and j represent wet (1) or dry (0) days and n_{ij} the number of days in the observed record corresponding to the precipitation state defined by i and j on the preceding and current day respectively. Transition probabilities are calculated for each of the twelve calendar months, resulting in four probabilities per month: P_{11} (a wet day preceding a wet day), P_{00} (a dry day preceding a dry day), P_{01} (a dry day preceding a wet day) and P_{10} (a wet day preceding a dry day), though the number of independent transition probabilities is only two per month, where $P_{11} = 1 - P_{10}$. Patterns for P_{00} and P_{11} will be presented here.

For a given month, the change in a transition probability, P, at a GCM grid-cell with latitude and longitude coordinates x and y will be regressed using least squares regression against the GMST change relative to 1850-1900, such that

$$P_{xyn} - P_{xy,n=0} = a_{P_{xy}}(T_n - T_{n=0}) + c_{P_{xy}} \tag{6.2}$$

where n is any 30-year window and with $n = 0$ referring to the first 30-year window used to calculate parameters. The gradient of the regression, $a_{P_{xy}}$ and the intercept, $c_{P_{xy}}$ will differ for each cell. The gradients plotted globally form the spatial response pattern for each transition probability. The full historical+SSP3-7.0 time series will be used in the regression, but only the responses diagnosed from 2000-2100 of the historical+SSP5-8.5 time series will be considered. This is to avoid duplication of responses diagnosed from the historical simulation.

Linear regression has been chosen over a simple time-slice (or "delta-change") method. The time-slice approach is more sensitive to random noise caused by internal

variability which may then affect the size of the scaled responses. Unlike the time-slice approach, regression effectively uses the full simulated time-series (i.e., 1850-2100) further reducing the influence of internal variability (Ruosteenoja et al., 2007) and eliminates differences that may arise in choosing the epoch of the slices (Lynch et al., 2017).

The response of wet-day gamma-distribution parameters to changing GMST is determined as a fractional change, as opposed to a least-squares regression (Osborn, 1997; Osborn et al., 2016). The shape (α) and scale (β) parameters for days where the precipitation exceeds 0.1mm are similarly calculated from 30-year windows every 5 years from 1850 to 2100 using the historical+SSP3-7.0 and historical+SSP5-8.5 time series for each month. Shape and scale parameters are calculated using the Thom estimators

$$\alpha = \frac{1 + \left(1 + \frac{4D}{3}\right)^{\frac{1}{2}}}{4D} \tag{6.3}$$

$$\beta = \frac{\bar{x}}{\alpha} \tag{6.4}$$

with sample statistic

$$D = \ln(\bar{x}) - \frac{1}{n}\sum_{i=1}^{n} \ln(x_i) \tag{6.5}$$

where x is the precipitation on wet day i, with a total of n wet days. Thom estimators make better use of the information in a dataset and are considered more efficient than moment estimators (Wilks, 2011).

Although PS is usually applied to mean climate, recently Osborn et al. (2016) applied the pattern scaling technique to *monthly* precipitation shape parameters – that, together with the mean precipitation change, determines the distribution of monthly total precipitation. Similar methodology will be applied here, though to *daily* wet-day shape and scale parameters. Fractional changes to these parameters, referred to collectively as W, are regressed against the change in GMST anomaly (relative to 1850-1900) for each grid cell in a GCM, with coordinates x and y, such that

$$\frac{W_{xyn}}{W_{xy,n=0}} = a_{W_{xy}}(T_n - T_{n=0}) + c_{W_{xy}}. \tag{6.6}$$

where n corresponds to the n^{th} 30-year period and $n = 0$ once again refers to the first 30-year period with a calculated parameter. $W_{xy,n=0}$ is averaged across all ensemble members in the first recorded 30-year window (usually 1850-1880) where possible. However, some grid cells may have limited or no wet days during a given month over the first (or any) 30-year period making an estimation of the wet-day gamma distribution challenging or impossible. If, in a given month for any 30-year period, there are less than 30 wet days (i.e., one per month), no gamma distribution will be recorded. If this happens to be in the first 30-year window (i.e., 1850-1880), $W_{xy,n=0}$ in equation 6.6 is instead the *first* 30-year period that contains a calculated gamma parameter. This may be averaged across all ensemble members (if all of them contain recordings in the same 30-year window) or just one ensemble member. If $W_{n=0}$ is *not* calculated from the first 30-year period, the corresponding $T_{n=0}$ in equation 6.6 will also be calculated over the same 30-year period.

Unlike the transition probabilities, this may result in some locations and months containing fewer data points for regression. If there are less than ten recordings of W_{xyn} across all simulations (i.e., at least ten points in the regression), it is determined that there is not enough data to robustly identify a change in the parameter with changing GMST, and thus the slope $a_{W_{xy,}}$, for a given month and grid cell, is set equal to zero. This is a conservative choice: in the absence of robust information about the projected change, *no* change is assumed.

6.1.2.3 APPLICATION OF THE SPATIAL RESPONSE PATTERNS

The GCM-diagnosed response of a transition probability to changing GMST can be utilised by a SWG to produce local scale, daily time series, under a range of different climate scenarios and time periods. The transition probabilities calculated directly from observations can be perturbed for a range of GWLs using the gradient from equation 6.2 for the corresponding grid-cell. Data from two weather stations are used here (Santarém, Brazil, and Reykjavik, Iceland), which have been chosen based on large and differing parameter responses at their locations (presented in Section 6.1.3). Furthermore, these locations are classified as having tropical (Santarém) and temperate (Reykjavik) climates. Chapters 3 through 5 showed the strongest performance of the Markov-chain gamma-distribution weather generator at reproducing observed precipitation in temperate regimes, alongside the weakest performance in tropical locations.

First-order transition probabilities will be scaled additively;

$$P_T = P_0 + a_{P_{xy}} \Delta T \quad (6.7)$$

where P_T is the new transition probability following a specified global temperature change ΔT relative to the observed period (where the observed transition probability is P_0). Coordinates i and j refer to the grid-cell that encloses the coordinates of the location of the weather station used to calculate P_0. Probabilities have upper (1) and lower (0) limits. If after scaling $P_T > 1$ or $P_T < 0$, the transition probability is simply truncated just below the limits at 0.999 and 0.001 respectively.

Osborn et al. (2016) suggested that a multiplicative approach instead of an additive approach was appropriate for pattern scaling shape and scale parameters, because they are bounded at zero. Wet-day gamma distribution parameters are therefore scaled multiplicatively;

$$W_T = W_0(1 + a_W \Delta T) \quad (6.8)$$

where W_T is the new gamma parameter following the global temperature change ΔT relative to the observed period (for which the observation-based gamma parameters are W_0). Once again, the scaled parameters must be constrained to the physically allowed range ($W > 0$). If a parameter falls below 0 following scaling, it will instead be set to an arbitrary small number (i.e., 0.01). Though using a different method, Wilby and Wigley (2002) similarly compared changes in daily gamma distributions to assess precipitation under different scenarios.

The perturbed transition probabilities, P_T, and wet-day parameters, W_T, can hence be used to generate time series of precipitation under the specified GWLs. This method of synthetic time series generation avoids GCM-induced bias because the starting point is the parameter set fitted to observations and only changes in parameters are taken from the GCM. For a given site, the area-averaged data can vary significantly from the local, weather station scale data. Using local-scale data perturbed by GCM predicted trends will reduce errors in estimating the local climate. An example of this will be given in the following Section.

6.1.3 RESULTS AND DISCUSSION

6.1.3.1 TRANSITION PROBABILITY RESPONSE TO CHANGING GMST

Spatial responses for dry-day to dry-day and wet-day to wet-day probabilities, hereafter referred to as P_{00} and P_{11}, have been produced using IPSL-CM6A-LR and are shown in Figure 6.1 (see Appendix for example linear regression scatter plots). Deductions about the transition probabilities, P_{01} and P_{10}, can be drawn from P_{00} and P_{11} respectively. For visualisation purposes, the maximum and minimum transition probability responses (gradient in equation 6.2) are capped at 6 and -6 respectively, where a value of 6 (-6) indicates an increase (decrease) of 6 in a transition probability per degree GMST increase. Note that the colour bars have been reversed for P_{11} and P_{00} to indicate increasing wet or dry conditions with consistent colours. While spatial distributions have been produced over land and ocean, attention will be focussed over land-surface areas due to the increased relevance of these locations in impact assessments. Seasons may be referred to by their months, where DJF refers to December, January, February, MAM March, April, May, JJA June, July, August and SON September, October and November.

Several regions, now discussed in turn, show strong transition probability responses to increasing GMST. Firstly, within South America, strong responses are present with large magnitudes in the north, east, and monsoon regions, including across the Amazon rainforest. IPSL-CM6A-LR generally shows a decrease in the number of wet days alongside increased dry day frequency and dry-spell length across these regions., with decreasing P_{11} (therefore increasing P_{10}) and increasing P_{00} (hence decreasing P_{01}) in several months (note that the mean length of dry spells is $\frac{1}{1-P_{00}}$). During the wettest months over the Amazon region (January to March), P_{11} remains mostly constant with increasing GMST. North-eastern Brazil shows strong decreases in P_{11} in April and May, around the start of the region's dry season. This, combined with the strongest P_{00} responses during the dry season (May to October), the length and intensity of the dry season in these regions may increase, heightening the risk of meteorological droughts. This, alongside homogenous decreases in P_{11} from July to October over Brazil, reiterates literature findings that the maximum number of consecutive dry days per year (CDD) in the Amazon basin and north-eastern Brazil, alongside the length of the dry season, is projected to increase with climate change (Sörensson et al., 2010).

Response patterns are less pronounced in North America than South America. Increases in P_{11} are present across Canada and Alaska throughout the winter (DJF) with slight decreases in all other seasons. Conversely, decreases in P_{00} are present in the winter with increases in other months. This suggests an overall increase in wet-day frequency and wet-spell length during winter. Responses are weaker in magnitude across the contiguous USA, closer to zero for large spatial areas. Where there are stronger responses, mainly in western USA, some months show increases in P_{00} with only slight decreases in P_{11}. This suggests only a slight increase to dry-day frequency, with longer dry-spells. This differs over the east coast where there are increases in P_{11} from May to October.

In Central America, P_{11} remains relatively constant with increasing GMST. Exceptions are present primarily over coastal grid cells from May to August, where large negative trends are noted. Central America shows much stronger responses than North America for P_{00} from June to September, where large increases can be seen, with responses bounded by large decreases over the Pacific Ocean. While the wet and dry seasons vary by country and are dependent on temperature anomalies in the east-equatorial Pacific and Tropical North Atlantic Oceans (Hidalgo et al., 2013), the strongest probability responses tend to fall in an individual country's wet season. The responses suggest increased CDD alongside a reduced occurrence of wet days during these wet seasons. This agrees with the findings of Ortega et al. (2021) that decreases in overall precipitation are projected by the end of the 21st century in this region.

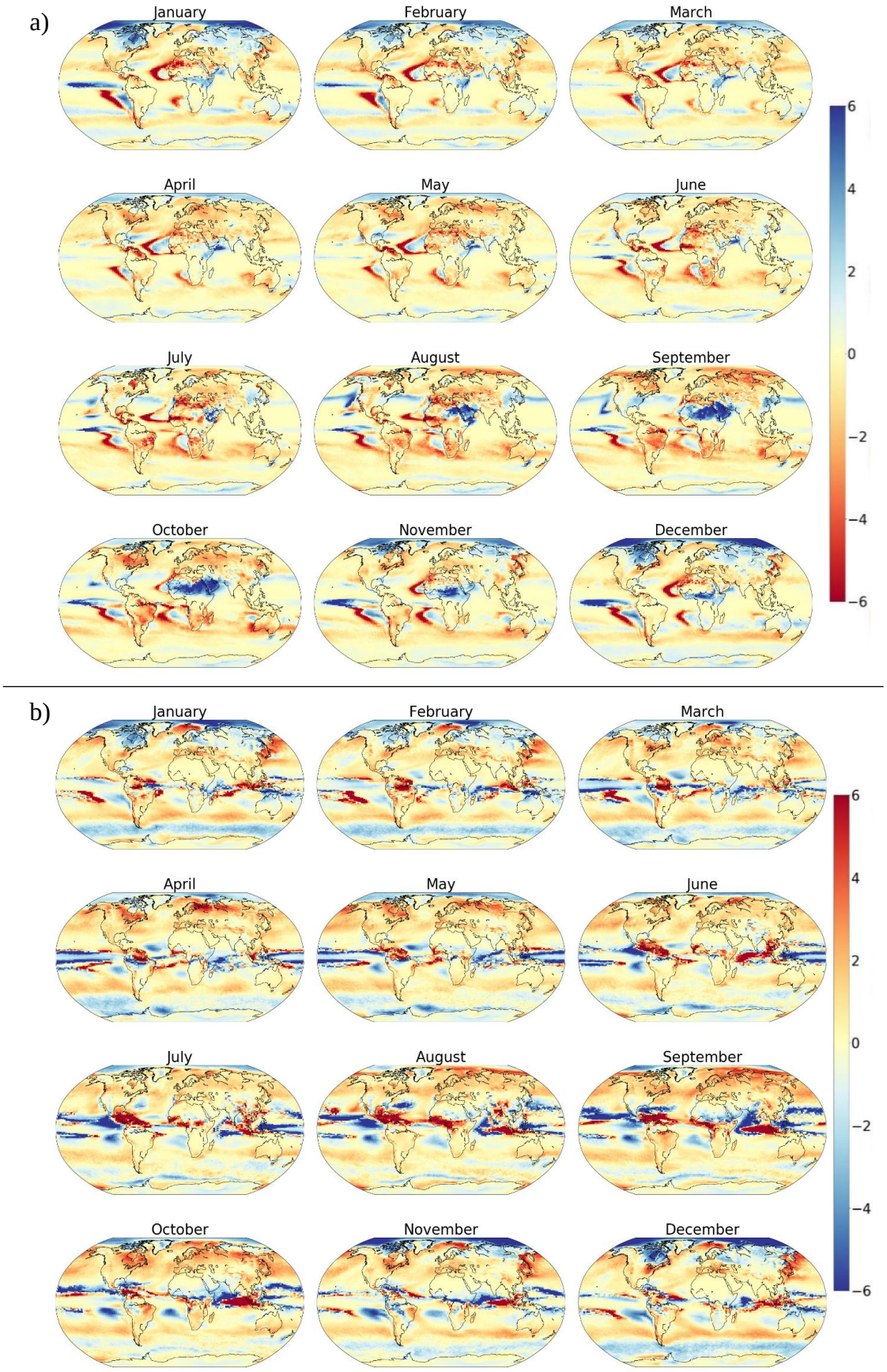


Figure 6.1 – Monthly change in P_{11} (a) and P_{00} (b) per degree GMST increase, diagnosed using pooled IPSL-CM6A-LR historical, SSP3-7.0 and SSP5-8.5.

Spatial responses of P_{11} to increasing GMST vary across Asia. Across Saudi Arabia, Yemen and Oman, strong seasonal changes dominate with large increases in P_{11} from May to November. This indicates decreased CDD in what is generally already considered a dry season in this region (Almazroui et al., 2012). Decreases in P_{11} are present during the wet season, though with much smaller magnitude than the increases during the dry season. P_{00} responses to increasing GMST across this region are smaller in magnitude. Some decreases are noted in the dry months, though with much less spatial coverage and magnitude than changes in P_{11}. Still, this remains an annual increase in wet-day frequency. During monsoon seasons (for example, JJA in India and July and August over East Asia), the response of the transition probabilities show little change. This is likely due to monsoon dynamics governing the precipitation patterns, resulting in precipitation occurrence on most days. Across south and central Asia, the magnitude of P_{11} response is close to zero, though some increases in east India can be seen. Increases in P_{00} are present across the Himalayan Mountain range, while regions just south see little change. While not as strong, this is also notable with P_{11} responses, where across the Himalayas there are decreases in P_{11} and increases further south. These responses suggest an increased dry-day frequency alongside increased CDD. Responses across north and west China and east Russia change seasonally and reiterate patterns across similar latitudes in North America. P_{11} increases with GMST in winter (generally November to March) and decreases in the summer months (April to October). As with P_{11}, P_{00} patterns tend to reiterate changes at similar latitudes in North America. However, patterns in north-east China and Korea differ, with P_{11} decreasing in winter and increasing in summer (though with less magnitude and sparser spatial coverage), alongside increases in P_{00} large in magnitude during the winter months. Altogether, this suggests less frequent precipitation, with longer dry spells, during the winter season.

Magnitudes of the probability responses are large across much of Africa. Decreasing P_{11} responses, large in magnitude, are present across much of west Africa from May to August. Contrasting the decreases in P_{11}, large increases in P_{00} are present from July to September. This strongly supports the findings of Klutse et al. (2018), wherein 80% of studied RCMs agreed with increasing CDD and decreasing maximum consecutive wet days per year (CWD) with rising GMST over the Guinea Coast region (Klutse et al., 2018). Decreasing P_{11} probability is also seen in southern Africa (June through October) though with smaller magnitude, with increases in P_{00} present during several months. This agrees with studies into increased drought and CDD risk in the southern African region (Almazroui et al., 2021). Conversely, strong increases in P_{11} are present across central and east Africa

from August to January, covering large spatial areas in SON. The combination of changes across central Africa and the Sahara suggests an increased wet-day frequency. In the east African region, the response of P_{00} to GMST increase is weaker, though decreases are present in SON. This indicates an increased number of wet days, implying wetter weather (in the absence of the wet-day parameters). These responses reiterate the general consensus of a wetter east Africa as a function of climate change (Cooper et al., 2008).

In general, as GMST increases, IPSL-CM6A-LR projects a slightly higher frequency of wet days, with longer wet spells in the winter (DJF) over much of northern Europe. The opposite is true in the summer (JJA); a higher frequency of dry days, with longer dry spells, is indicated. Central Europe shows approximately constant P_{11} responses, with spatial responses across Europe showing some similarities to North American responses at similar latitudes. During the winter months, northern Eurasia shows widespread increases, though small in magnitude, in P_{11}. As with north Canada, these responses reduce in magnitude and coverage until summer, where most of Europe and Eurasia experiences decreases in P_{11}. P_{00} responses also show strong seasonal variations. Winter months show spatially widespread decreases in P_{00}, while the rest of the year shows increasing P_{00}, with some very strong, localised responses in April, May, September and October over the Eurasian subarctic and wider, but more modest increases over the rest of the continent.

Responses across Oceania are smaller in magnitude than other regions. While much of Australia shows weak patterns, increasing P_{11} is noted in the east (New South Wales and Queensland) from November to March, with greatest magnitude in December and January. Decreases in P_{11} are present in several other months, with decreases found in south Australia, evident from July onwards, with very strong southern coastal decreases by October. P_{00} is less sensitive to increasing GMST with small magnitudes of change. The largest changes (decreases) are present in north Australia in December, January, and February. Slight increases are shown in south Australia from August to November, though with lesser magnitude. Responses over islands in Oceania are less defined.

6.1.3.2 WET-DAY GAMMA DISTRIBUTION PARAMETER RESPONSE TO CHANGING GMST

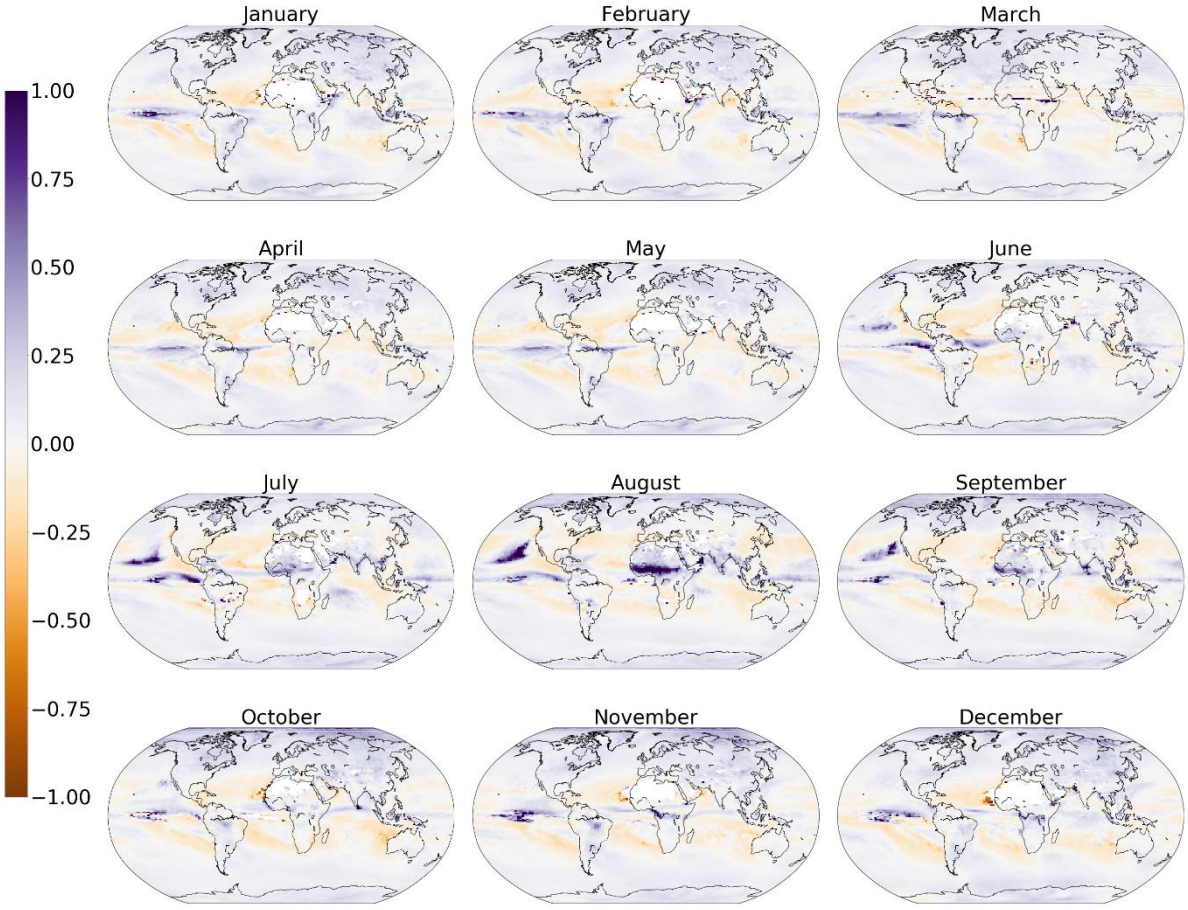


Figure 6.2 – Fractional change in wet-day shape parameter per degree GMST increase, diagnosed from IPSL-CM6A-LR historical, SSP3-7.0 and SSP5-8.5.

The wet-day gamma parameter responses are shown as fractional changes per degree warming, where a value of +0.5 indicates an increase of 50% of the original value per degree warming, i.e., for a shape parameter of 3 and a response of 0.5, the parameter following 1 degree of warming becomes 4.5 (a 50% increase). Shape and scale parameter changes (per °C) have been capped at +/-1, and +/-0.5, respectively, for visualisation purposes (Figure 6.2 and Figure 6.3). The first finding from these maps is that these patterns are spatially coherent. This indicates that the dependence of these parameters on GMST change can be diagnosed without being dominated by local scale noise and hence that the idea to pattern scale the parameters of the weather generator has merit.

While interpreting transition probability response to changing GMST is intuitive, inferring changes to the wet-day parameters is less straightforward. It is difficult to reach

conclusions on changes to resultant wet-day gamma distributions through studying changes to one parameter alone, since unlike P_{11} and P_{00}, shape and scale parameters are not independent of each other, each affecting the wet-day distribution's properties. A change in the shape parameter will, instinctively, result in a distribution with a different shape; for example, a large increase in shape from an initial value of 1 will adjust the distribution from an exponential-type character towards a more bell-shaped function, with the breadth of all distributions (irrespective of shape) also determined by the scale parameter. A decrease in shape parameter indicates a more positive skew. Here, shape responses will be discussed in absence of any changes in the scale parameter.

Changes to the shape parameter (shown in Figure 6.2) are seemingly less affected by the seasons than the transition probabilities, with fewer spatial patterns of note. All months show slight fractional increases to the shape parameter over most land surface areas. Most regions do not show strong or seasonal changes to shape parameter response, with noteworthy exceptions in Africa and South America.

The shape parameter shows strong, seasonally changing responses in West and Central Africa, with spatially widespread responses from July to September. These large increases may result in a less skewed distribution in a warmer climate (e.g., a present-day climate with many low-precipitation days but occasional very wet days could move towards one with more consistent precipitation amounts), with the strongest responses coinciding with strong increasing (decreasing) P_{00} (P_{11}) probability. This suggests large changes to daily precipitation in this region. However, the actual chance of very wet days also depends on the scale parameter and on the probability of wet days, both of which are also changing. Much of Saharan Africa shows no response to increasing GMST simply based on a limited sample of wet days per 30-year period available to calculate the parameters. Unlike transition probability response, South Africa shows seasonally consistent responses, small in magnitude.

While much of South America shows patterns that reflect the global, slight increases in wet-day shape parameter, stronger responses are present over the Amazon region from February to May, and southeast Brazil throughout the year. These tend to be the locations that show the strongest transition probability responses. While several land areas show slight increases, Brazil sees slight decreases in shape parameter along the east coast in August, September and October. This indicates a move towards a more positively skewed, exponential-type distribution.

Responses across North and Central America are approximately constant annually, with slight increases shown in most months and greatest magnitudes present in the high latitudes. Exceptions are noted across east USA from June to September, where slight decreases are present. Unlike the responses across Africa and South America, coherence with transition probability response is not noted. Much like the transition probability responses, patterns over Europe and most of Eurasia tend to reflect those in similar latitudes over North America.

For most months, Southeast Asia and India also show responses that reflect the patterns over most land surface areas. However, some localised, positive responses are present over Southeast Asia and India in several months, for example, August. While P_{11} is mostly constant with increasing GMST in these regions, P_{00} shows strong increases (over Burma, Thailand and South China). The combination of an increased chance of consecutive dry days, alongside a less skewed precipitation distribution, suggests changes to the resultant precipitation character of these regions, depending, of course, upon scale changes too. Responses over central and northern China usually follow patterns present in similar latitudes across North America and Europe, with increases to the shape parameter in most months. However, in July, August and September, large areas in the east show slight decreases to the shape parameter with increasing GMST. These months fall within China's wet season (May to September), indicating large implications on the overall mean daily precipitation during these months.

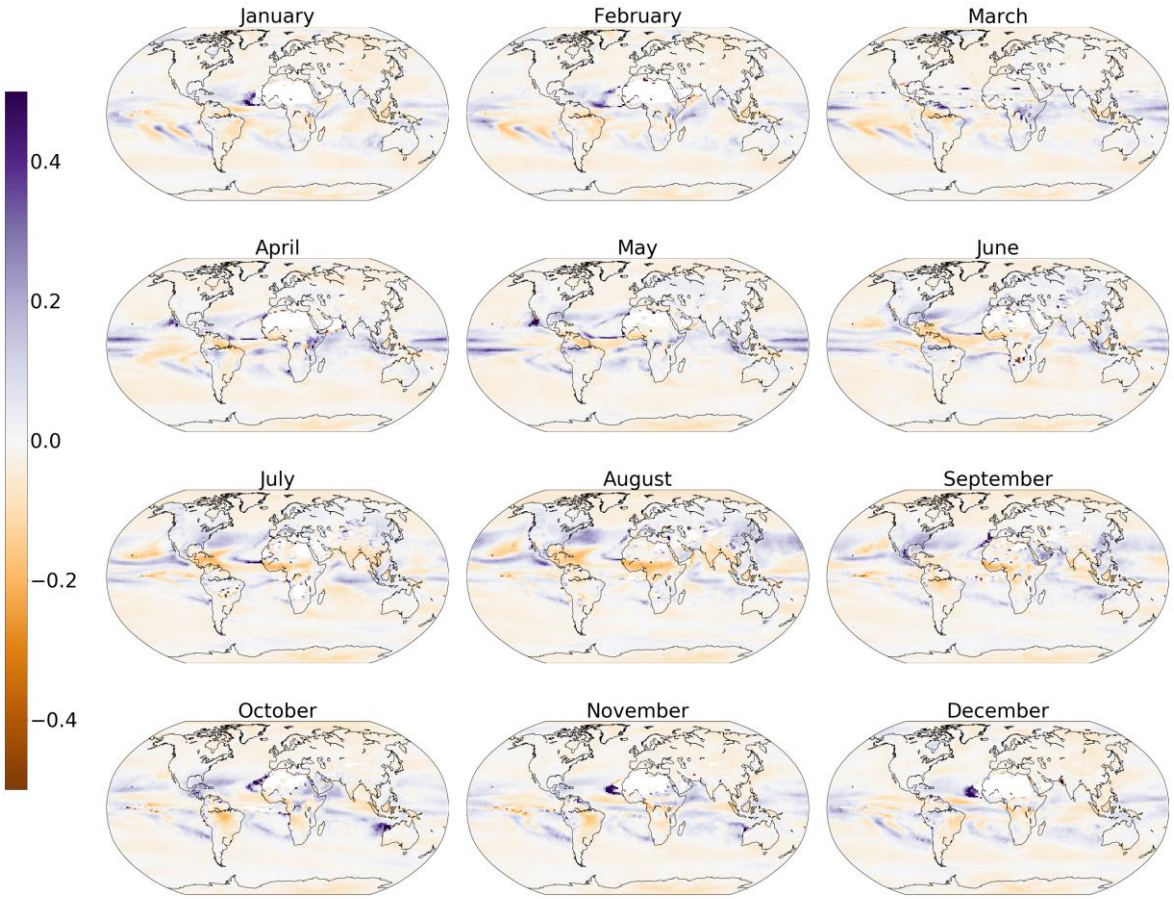


Figure 6.3 – Fractional change in wet-day scale parameter per degree GMST increase, diagnosed from IPSL-CM6A-LR historical, SSP3-7.0 and SSP5-8.5.

Scale parameters show weaker responses to GMST increase than the shape parameters, this time with widespread, slight decreases. Regions that show the largest changes in shape parameter tend to also show the strongest scale parameter responses to GMST change, with more global seasonal changes (Figure 6.3). The scale parameter defines the horizontal and vertical scaling of the distribution, where in the absence of changes to the shape distribution, an increase indicates a stretching of the distribution, a shift towards larger precipitation amounts and greater variability. Conversely, a decrease indicates a shift towards smaller precipitation amounts, less variability and loads the distribution towards smaller precipitation amounts.

Parts of Africa again show the strongest seasonal changes to parameter response, especially in and to the south of the Sahel. From June to September, decreases in scale with increasing GMST are present across West and Central Africa. This coincides with the strongest increases of the shape parameter with increasing GMST. While the increasing shape parameter suggests a distribution with less exponential features (less positive

skewness) but also an increased mean and variance, it is accompanied by a decrease in scale which reduces the mean and variance (without changing the skewness) so that the net effect is hard to determine from the individual parameter changes. The varying wet-day parameter responses, alongside strong transition probability response, result in changes to the region that are not intuitive. This is unsurprising, as Cooper et al. (2008) note particular uncertainties with regards to precipitation projections in the West African region.

South America shows decreases in the scale parameter with increasing GMST in several months, present over Amazon and central Brazil from September to November. This coincides with the dry season in these regions, indicating a distribution that is loaded towards smaller precipitation amounts on the few wet days that do occur. Such a negative trend is additionally worrisome when considering these regions show strong P_{00} and P_{11} responses that indicate increasing dry day frequency. Conversely, slight increases in Chile and Peru are present in several months, indicating a shift of the distribution towards higher wet-day precipitation amounts.

While shape parameter response remained widely constant with increasing GMST and with positive magnitude across Central and North America, seasonally varying patterns are present with the scale response. Neutral responses are present from November to April with increases present in all other months, focussed on the east coast of the USA and appearing to peak in September. Unlike the patterns across Africa, coherence with transition probability response is not noted. Increases are present over Central America from June to August. In the absence of changes to the shape parameter, increases to the scale indicate increases to the mean and variance.

Shape parameter response to GMST over India and Southeast Asia shows slight increases, notably in August, with decreases in the scale parameter during July to September, indicating a shift of the wet-day precipitation distribution towards lighter precipitation amounts. Eastern China, once more, shows patterns of a different sign to elsewhere in Southeast Asia and India, with moderate increases visible predominantly from June to September and within China's wet season. On account of the scale response, at least, a precipitation regime with a wider spread of values may be expected.

6.1.3.3 APPLICATION OF RESPONSES TO A STOCHASTIC WEATHER GENERATOR

To demonstrate the applicability of the response patterns to a stochastic weather generator, synthetic precipitation time series at two weather stations under different GWLs have been produced using perturbed parameters (following equations 6.7 and 6.8). Santarém, Brazil, and Reykjavik, Iceland have been chosen as case study locations for reasons provided in Section 6.1.2. Santarém sits at the confluence of the Tapajós and Amazon Rivers. A tropical monsoon climate, Santarém experiences similar temperatures throughout the year, with strongly defined wet (November through July) and dry (August through October) seasons. GCM-diagnosed responses at Santarém show generally increasing P_{00} and decreases in P_{11}, the latter particularly pronounced in September and October. Reykjavik is located on the southwest coast of Iceland. A subpolar oceanic climate, Reykjavik too has defined wet and dry seasons, coinciding with the winter (DJF) and summer (JJA) months respectively. Transition probability responses at Reykjavik are weaker than Santarém, alongside year-round, slight increases to the shape parameter. Global Summary of the Day data, provided by the National Centers for Environmental Information (NCEI), have been used for Santarém (dataset identifier: NCEI DSI 3505, accessible from https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00516) and the European Climate Assessment & Dataset (ECAD) for Reykjavik (Klein Tank et al., 2002). Note that this section aims to demonstrate the applicability of the PS method to perturbing the parameters of a SWG and not in-depth analysis of the future climate at the two case study locations.

Observed parameters, P_O and W_O, for each month have been calculated from the weather station records. A synthetic 100-year precipitation time series is first generated using the SWG configured with observed parameters directly from weather station observations. This is hereafter referred to as the "reference" period. The GCM grid cell that encompasses the geographical coordinates of each weather station has been used to determine $a_{P_{xy}}$ and $a_{W_{xy}}$. Parameters have been scaled using GMST increases (relative to the 1850-1900 mean) representing GWLs of 1.5, 2.0, and 4.0°C. The time periods covered by the observed data records used to estimate the reference SWG parameters differ between the two sites. The mean observed GMST anomaly using the HadCRUT5 dataset (Morice et al., 2021) relative to 1850-1900, are 0.78°C for the time period with observations at Santarém (1990 – 2020) and 0.56°C for the period with observations at Reykjavik (1970 –

2020). Observed GMST anomalies have been calculated using the HadCRUT5 dataset (Morice et al., 2021). The difference between the observed GMST and each GWL is calculated and substituted into ΔT in equations 6.7 and 6.8 to determine P_T and W_T respectively. Using the scaled parameters, further 100-year time series are produced for both stations.

Figure 6.4 demonstrates the resultant transition probabilities for all months and the wet-day gamma distributions for April at Santarém following perturbation with the IPSL-CM6A-LR response patterns. April has been provided as an example due to large changes in the scaled wet-day distribution. Perturbed transition probabilities show large increases to P_{00} from January to April, indicating increases to mean dry-spell length during these months. There are also large changes to distribution of precipitation amount at the GWLs in April. The shape parameter has increased from the reference period to a GWL (from 0.7 to 1.9) and is accompanied by a decrease in the scale (from 20.7 to 8.5). Such changes are reflected in the mean and variance. During the reference period, the mean wet-day precipitation is 15.1mm/day with a standard deviation of 17.6mm/day. At GWL of 4.0°C, the mean has increased to 16.4mm/day with a standard deviation of 11.8mm/day.

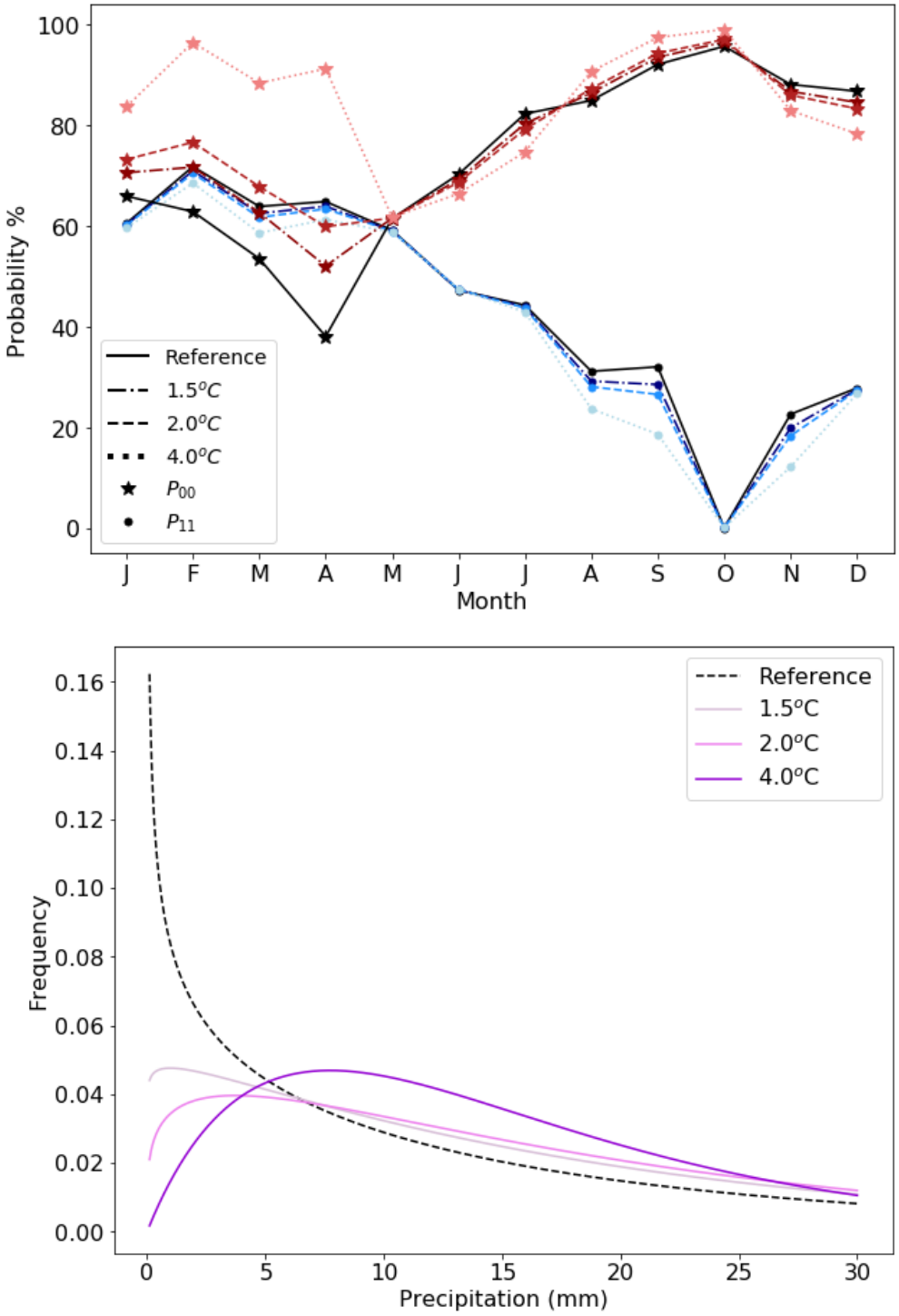


Figure 6.4 – Transition probabilities (top panel: blue lines scaled P_{11} orange lines scaled P_{00}) and wet-day gamma distribution for April (bottom) at Santarém at different GWLs with shapes 0.7, 1.0, 1.2, and 1.9 and scales 20.7, 18.0, 16.1 and 8.5 during the reference period and GWLs of 1.5, 2.0 and 4.0°C respectively.

Santarém shows large decreases in mean precipitation from January to April with increasing GMST (Figure 6.5), particularly beyond 2°C. In all other months, changes are minimal. Strong transition probability responses drive this decreased precipitation; the mean number of wet days per year decreases from 108.8 during the reference period to 75.0 by

4°C GWL. February, March and April show exceptionally strong increases in P_{00} (+10.3% and +15.7% per °C GMST increase respectively for March and April), relating to increases of 35.8% and 54.6% with respect to the reference period (see Figure 6.4). Large decreases in the mean number of wet days in April further reflect this increase; the reference period experiences an average of 20.8 wet days per month, whereas at a GWL of 4°C, the mean number of wet days is only 11.3. While the decreasing number of wet days in March and April are a primary driver for the decreases in mean precipitation by 4°C, at lower GWLs there are slight increases in mean wet-day precipitation during these months, reflected in the perturbed wet-day gamma distribution. An example is shown for April in Figure 6.4, where the mean has only increased by 1.3mm/day from the reference period to GWL 4.0°C. As the GMST increases, the distribution shows greater variability with higher values and less skew. At a GWL of 2°C (approximately 1.5°C warmer than the reference period), there are still 19.0 wet days per April. This is only a slight decrease on the reference period's 20.8 days. However, at 2°C, the mean wet-day precipitation has increased by 20%. This reflects the changes to the wet-day distribution and further, more pronounced changes occur under the 4°C GWL. This results in small increases in monthly mean daily precipitation between the lower GWLs. At the 2°C GWL the distribution shows greater variability a higher mode (around 3mm/day). More pronounced changes in the distribution occur under the 4°C GWL, where the April mode rises to about 10mm/day. However, at this GWL in April, there are only 11.3 wet days per month, thus resulting in the large reductions in mean precipitation despite the changes to the wet-day distribution.

Changes to precipitation in March and April are in stark contrast with May (and indeed the remainder of most of the year), where the responses of transition probabilities and wet-day gamma distributions to changing GMST are close to zero, thus resulting in little change to the number of and precipitation on wet days. The reduction in precipitation at a GWL of 4°C for the first quarter of the year is shown to be distinct and agrees with those changes discussed in literature. For example, Sörensson et al. (2010) found increases to CDD, alongside significant decreases in winter (DJF) and spring (MAM) total precipitation in the Amazon basin. Averaging the generated precipitation over these seasons agrees with these findings. Ortega et al. (2021) note that, in this region, agreement amongst CMIP6 models is only present during SON. This happens to coincide with the season showing the smallest changes at each GWL.

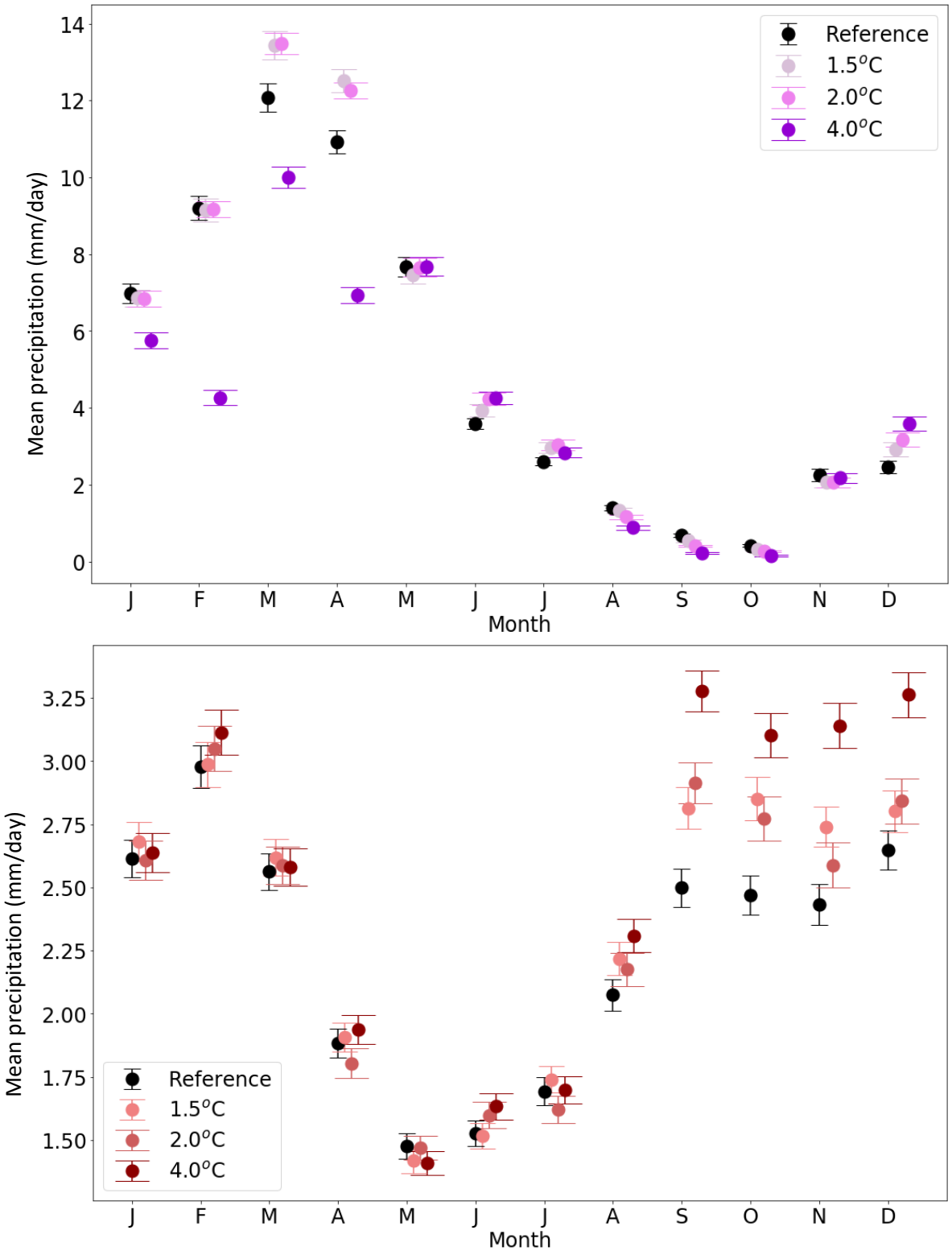


Figure 6.5 – Mean precipitation (mm/day) at Santarém (top) and Reykjavik (bottom), at different GWLs. Error bars show the standard error in the mean.

Conversely, mean daily precipitation at Reykjavik is projected to increase sizeably in several months (especially September to December). The largest increases are present during the winter months, DJF (already the wettest season). In these months, increases of at least 14% in mean wet-day precipitation are present, with up to 21% in December. Alongside increases in mean precipitation, there are large changes to the variability over these months. Increases in mean precipitation are accompanied by an increase of 11.9% (averaged from September to December) in the standard deviation of daily precipitation, with the largest increase in December (15.5%). The total number of wet days per month is relatively unchanged (i.e., from 18.1 during the reference period to 18.4 at GWL 4°C) during these months. This change in variability is therefore solely from the changes to the wet-day precipitation distributions, and not the change in wet or dry day frequency. Changes in variability may therefore result in the number of very wet days, resulting in a potentially increased risk of extreme precipitation. This is a known result of warming GMST; as the atmosphere warms, the water holding capacity increases (Trenberth, 2011) resulting in larger rainfall events.

Changes to the mean precipitation and the number of wet days per month from May to July (the driest months) are much smaller. Figure 6.6 shows that the transition probabilities during these months are approximately constant. Changes to the distribution of wet-day precipitation are not large enough to affect the mean precipitation over these months. However, from January to April the probability of a dry day occurring is slightly increasing, with increases in P_{00} and decreases in P_{11}. Despite this, the precipitation in January, March, and April remains fairly unchanged between the GWLs. This is due to changes in the wet-day gamma parameters. From January to May, the mean wet-day precipitation is projected to increase by an average of 11.9%. Therefore, despite decreases in the number of wet days per month, the precipitation that may be expected on any wet day during this period is larger. The percentage change is largest in February (15.7%) where the total precipitation increases at each GWL, despite the decreases to the wet-day frequency.

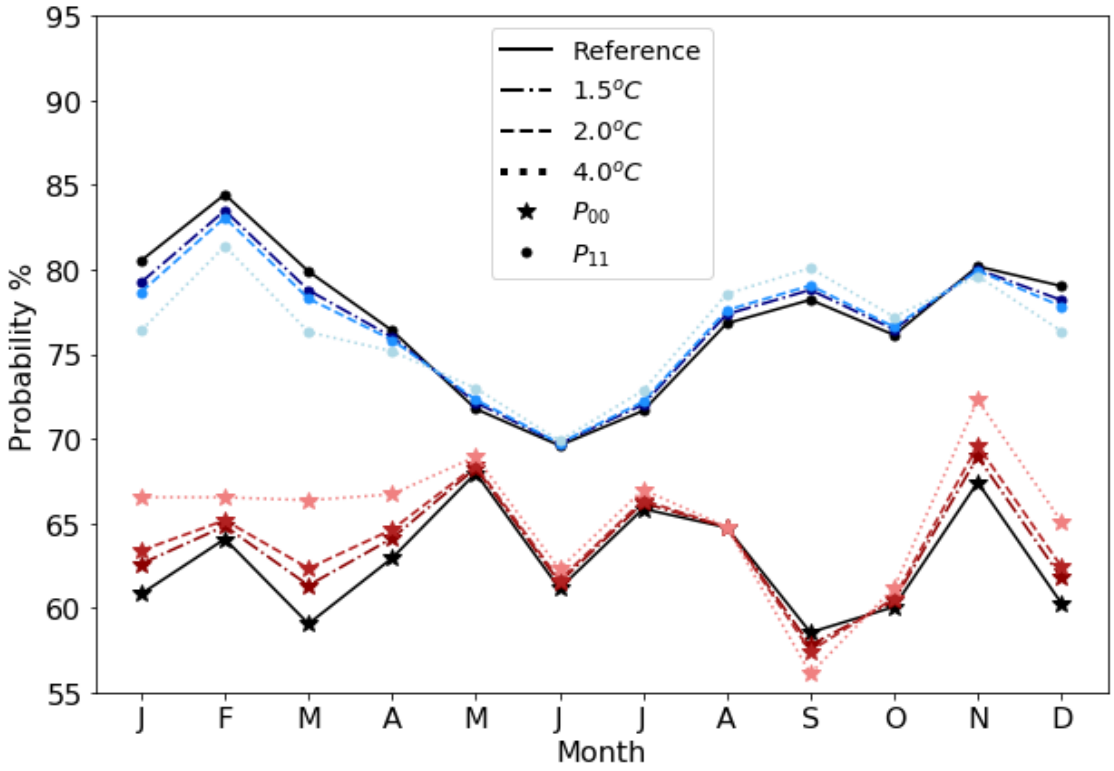


Figure 6.6 – Transition probabilities (blue lines scaled P_{11}, orange lines scaled P_{00}) at Reykjavik under different GWLs.

To further investigate the advantages of using a SWG when configured with parameters that are derived from observations and then scaled, as opposed to configuration with parameters that are diagnosed *directly* from GCM daily precipitation data (e.g., for reference period and then again for some future climate period), the existing Santarém data is compared with a new synthetic SWG series representing the latter approach. For this comparison, GCM daily precipitation data for the cell containing Santarem is extracted for the four IPSL-CM6A-LR SSP3-7.0 ensemble members. From these series, any simulated year where the GMST change (relative to the 1850 – 1900 mean) corresponds to 0.78 ∓ 10% (the reference GWL) or 4°C ∓ 10% is identified and used to define two sets of SWG transition probabilities and wet-day gamma parameters (a reference set, and a GWL 4°C set). Using these parameters, two new batches of 100-years precipitation are generated via the SWG for the reference *and* warming period. These data are referred to as the "GCM-driven" data and are compared to the existing stochastically simulated Santarém data (shown in Figure 6.5, top panel) generated using observed parameters and the scaled-observed parameters (referred to here as "observation-driven").

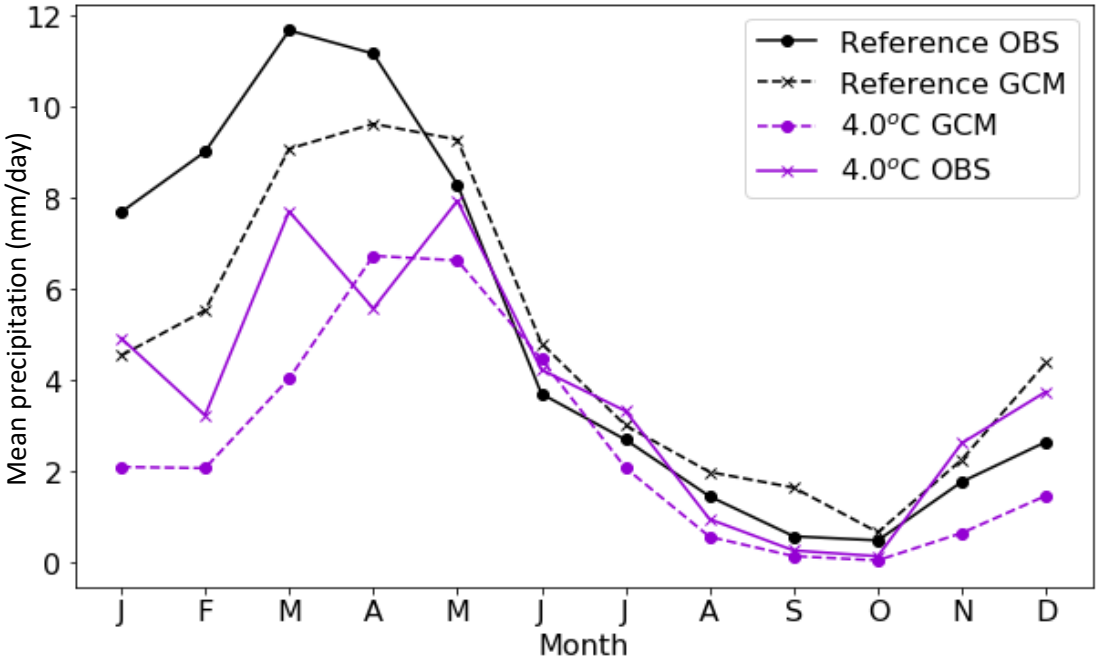


Figure 6.7 – Mean daily precipitation from 100-year GCWG simulations using GCM-driven and observation-driven parameters for the reference period and a GWL of 4 ℃.

Through comparison of the reference periods, it is evident from Figure 6.7 that there are errors in modelling local climate using the GCM-driven precipitation. Observation-driven precipitation peaks in March, with GCM-driven precipitation peaking later in April. GCM-driven precipitation underestimates the observed mean precipitation in March by 21%, and, from August to November (the driest months), observed precipitation is overestimated by 35%. Furthermore, a shorter dry season and a wetter wet season are simulated by GCM-driven precipitation than observations suggest. Accurately depicting the wet and dry seasons, alongside the precipitation amounts, is crucial for water resourcing and management (for example). Comparing precipitation at 4°C demonstrates that general trends are reflected by both the GCM- and observation-driven precipitation time series. Wet seasons both show decreases in precipitation, with smaller changes during the dry season. The months with the largest decreases in the GCM-driven precipitation at the 4°C GWL also show the largest decreases in the observation-driven data. While most months of observation-driven data show changes in precipitation in the same direction as the GCM-driven data, on occasion the direction of the change differs (e.g., during the dry season). However, some differences between the precipitation in the GCM-driven and observation-driven data are not surprising, as scaling responses have been constructed using SSP5-8.5 data alongside the SSP3-7.0 simulations that are shown here. Additionally, constraints associated with using a linear relationship to diagnose the scaling responses will, by

definition, not always precisely capture actual GCM-projected changes in precipitation character for some grid cells and months.

Time series stochastically generated using local-scale, observed parameters perturbed by area-averaged responses to GMST increase can be used to assess a range of impacts on precipitation character, including monthly totals, interannual variability, the distribution of wet- and dry-spell lengths and extreme indices, while reducing GCM-induced biases. It is important to note that different GCMs will likely produce differing responses of the parameters to increasing GMST, and the demonstrative analysis in this study uses just one GCM as an example. Through diagnosing similar response patterns from alternative GCMs, a wider range of realisations for specific global warming levels can be constructed to better explore and account for a range of uncertainties for consideration in robust adaptation plans.

6.1.4 CONCLUSION

The applicability of the pattern scaling technique to the transition probabilities and wet-day gamma parameters of a stochastic weather generator have been demonstrated. The response of two transition probabilities, P_{11} and P_{00}, and wet-day gamma distribution parameters to changing GMST have been diagnosed using four ensemble members each of historical, SSP3-7.0 and SSP5-8.5 simulations produced by IPSL-CM6A-LR. These scaling patterns have been diagnosed globally and have then been applied to two specific locations to illustrate the approach in full. Scaled responses have been used to perturb parameters calculated from weather station observations and hence generate 100 years of precipitation at Santarém, Brazil, and Reykjavik, Iceland, for a selection of GWLs.

Transition probability response generally agrees with wider studies, CMIP5, CMIP6, and downscaled projections. The response of the transition probabilities to changing GMST is seasonally varying, with the strongest land responses generally across Africa, South America and Southeast Asia. In the northern hemisphere, similar latitudes tend to show similar responses in magnitude, direction, and spatial coverage. This is most noticeable with P_{11} where winter months (DJF) show widespread increases with GMST, and summer months show prevalent decreases. The response of the wet-day gamma parameters to changing GMST are less intuitive to interpret, and in most locations, seasonally varying patterns are not present. However, once again, Africa and South America show the largest, seasonally

varying responses, indicating changes to the precipitation nature in these regions, in combination with strong transition probability response.

The effect of transition probability and wet-day gamma parameter responses to GMST increase have been demonstrated for two locations. Observed parameters at Santarém, Brazil and Reykjavik, Iceland, have been perturbed using local responses to increasing GMST to produce time series at three GWLs; 1.5, 2.0 and 4.0°C. Both locations show different responses to GMST increase. In Santarém, several months have very strong transition probability responses, associated with large decreases in wet-day frequency and increasing P_{00} in several months. Wet-day gamma parameter responses result in distributions with increased shape, and slight increases to mean daily precipitation in some months. In these same months (February, March and April for example), by 4°C, however, increases to P_{00} are so great that the number of wet days per month drops substantially, with simultaneous and large decreases in mean daily precipitation. Conversely, responses at Reykjavik show sizeable increases in mean daily precipitation during the wet season, though predominantly caused by changes to the wet-day gamma distributions. Transition probability response is weaker than at Santarém, with the largest differences in precipitation occurrence between the GWLs occurring during January to March. Despite decreases in the number of wet days, mean daily precipitation tends to remain unchanged, once again due to changes in the wet-day parameters resulting in greater precipitation amounts on wet days.

Applying PS to a SWG utilises the advantages of both techniques and unifies them into a single methodology. PS provides a method for the construction of a wide range of climate scenarios to capture a range of uncertainties with computational efficiency. SWGs can produce several realisations of long, local-scale time series, that are suitable for impact assessments, and which can be statistically representative of the weather at the site used for calibration. This study has utilised parameter responses to GMST change to perturb observed parameters at Santarém and Reykjavik, generating synthetic data with reduced GCM-induced bias, at a local-scale more appropriate for impact assessors and as inputs to hydrological, agricultural, and ecological models, and with the potential to emulate scenarios, time periods, and GMST increases not simulated by GCMs.

6.2 RESPONSES FROM AN ADDITIONAL GCM – ACCESS-ESM1.5

The above manuscript (Section 6.1) outlines the methods for constructing transition probability and wet-day gamma parameter responses from increasing GMST using the IPSL-CM6A-LR GCM. To better appreciate inter-model uncertainties between response patterns, diagnosis of global parameter response patterns from a further GCM, ACCESS-ESM1.5, will be performed. ACCESS-ESM1.5 has the same nominal resolution as IPSL-CM6A-LR but a lower sensitivity to climate forcing.

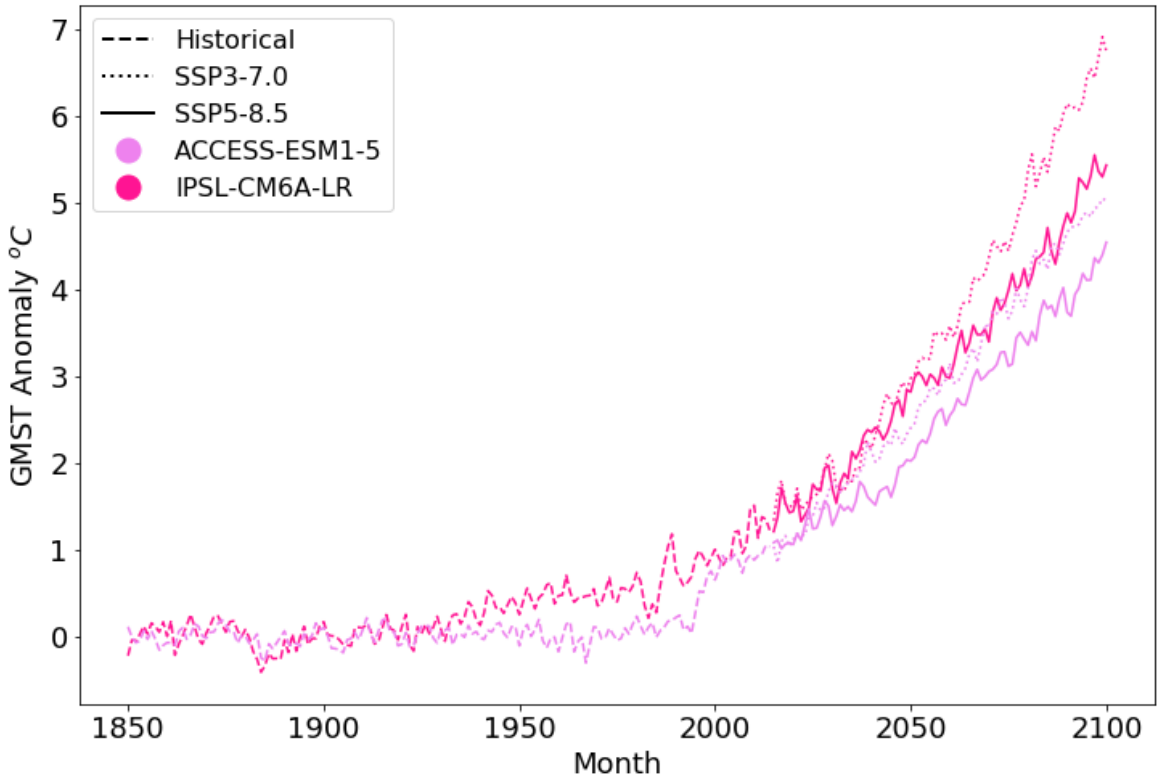


Figure 6.8 – The GMST anomalies relative to 1850-1900 of the two GCMs (ACCESS-ESM1.5 and IPSL-CM6A-LR) used to diagnose GCWG precipitation parameter response to GMST change.

A range of GMST anomalies by 2100 is projected by the two GCMs, shown in Figure 6.8. CMIP6 models show stronger future global warming trends than CMIP5 projections due to increased climate sensitivity values. The GMST increases projected by ACCESS-ESM1.5 fall just above the best estimates for SSP3-7.0 and SSP5-8.5 (3.6 and 4.4°C respectively) (IPCC, 2021a). Osborn et al. (2018) found that global temperature increases greater than 3°C have very small errors in diagnosing response patterns across most global land areas (with exceptions in Northern Hemisphere summers). This is due to the externally forced signal becoming much stronger compared to the unforced internal variability. Both

ACCESS-ESM1.5 and IPSL-CM6A-LR GCMs project GMST increases greater than 3°C for each scenario used in the construction of response patterns.

Responses will be diagnosed from ACCESS-ESM1.5 precipitation projections analogously to IPSL-CM6A-LR using equations 6.2 and 6.6. Four realisations each of the concatenated historical+SSP3-7.0 and historical+SSP5-8.5 daily precipitation time series will be pooled, resulting in eight temporally consistent time series from 1850-2100. As before, the full historical+SSP3-7.0 time series will be used in the regression, but only 2000-2100 from historical+SSP5-8.5 to avoid double counting the historical part. ACCESS-ESM1.5 results will be discussed and compared to the IPSL-CM6A-LR patterns presented earlier in Section 6.1.3.

6.2.1 TRANSITION PROBABILITY RESPONSE

Much like IPSL-CM6A-LR, ACCESS-ESM1.5 P_{11} shows strong, seasonally varying responses to GMST across north-eastern South America, South American monsoon regions, and Canada (Figure 6.9). The magnitude of P_{00} response to increasing GMST varies to IPSL-CM6A-LR's across Central America, the Arabian Peninsula, the USA and Europe, though generally agreeing with the direction of change, with larger differences over east Africa, Southeast Asia, and Australia.

South America once more shows some of the strongest transition probability responses to increasing GMST. Responses across the Amazon and monsoon regions of South America are generally similar to IPSL-CM6A-LR, and overall increases in dry-day frequency and dry-spell length are once more projected across north, east, and monsoon regions, showing decreasing P_{11} and increasing P_{00}. This further supports several studies pertaining to increased CDD (Alexander et al., 2006; Kim et al., 2020; Marengo et al., 2009) across these regions. A feature that is not present in IPSL-CM6A-LR diagnosed patterns but is present here, is a strong increase in P_{00} adjacent to and across the Andes Mountain range in most months. The greatest magnitudes are present in the Dry (mid) and Tropical (northwest) Andes. P_{11} changes are less prominent in this area, with slight negative trends in the Wet (south) Andes. Projected precipitation changes in the eastern slope of and equatorial Andes are generally poorly agreed upon between models (Ortega et al., 2021). This is reflected by IPSL-CM6A-LR showing decreases in P_{00} across this region, while, depending on month, ACCESS-ESM1.5 shows strong increases to no response at all. P_{11} remains approximately constant with increasing GMST from November to March. The strongest

decreasing responses of P_{11} to increasing GMST are present in July to October, during the dry season. Increases in P_{00} are found in this region all year round, with varying magnitudes. These combined responses will ultimately lead to an annual increase in dry days alongside increased CDD, particularly during the dry season.

While IPSL-CM6A-LR shows responses that vary in magnitude across North America, ACCESS-ESM1.5 shows more uniform, strong responses (though generally in the same direction as IPSL-CM6A-LR). During the spring (MAM) and summer (JJA), responses across central Canada show strong decreases (increases) in P_{11} (P_{00}). This suggests summers with more frequent and longer dry spells. In the winter months (DJF), P_{11} is constant with GMST, though some increasing trends are noted across north Canada and Alaska. Alongside increases in P_{11}, decreases in P_{00} are prevalent in these areas. As the seasons change from spring (MAM) to summer (JJA), the spatial area in north Canada with increasing (decreasing) P_{11} (P_{00}) begins to reduce until by summer, little area shows these increases (decreases). This suggests north Canada and Alaska may expect winter months with more frequent, longer wet spells. The contiguous USA shows widespread, though smaller in magnitude relative to much of Canada, decreases in P_{11} during most months. P_{00} response patterns mirror this; large regions (focussed on the centre and west) of the contiguous USA show increases in P_{00}, with smaller areas showing small decreases in the east.

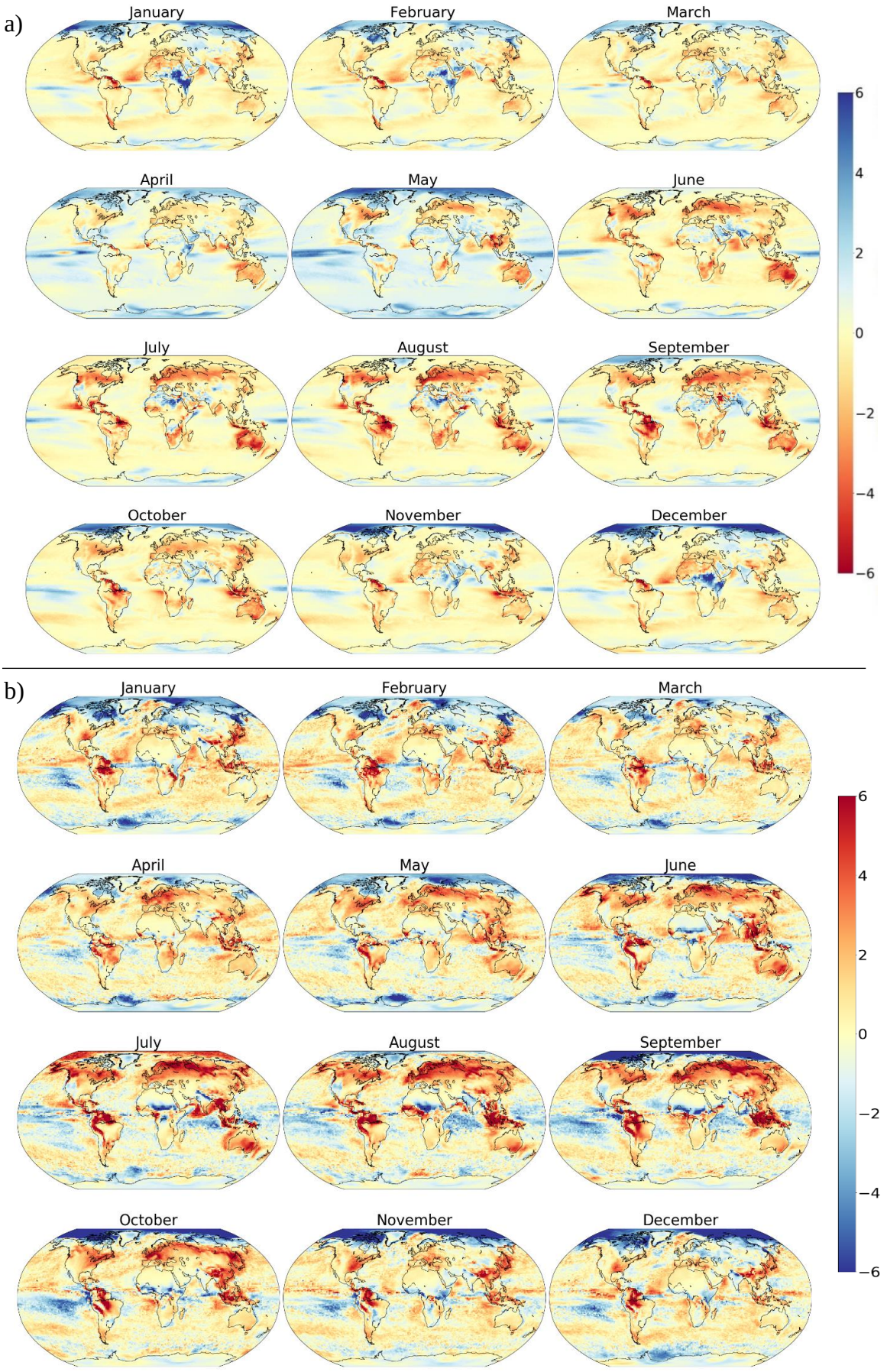


Figure 6.9 – Change in P_{11} (a) and P_{00} (b) per change degree GMST increase, diagnosed from ACCESS-ESM1.5historical, SSP3-7.0 and SSP5-8.5.

P_{11} response patterns produced by ACCESS-ESM1.5 over Africa are similar to IPSL-CM6A-LR, showing widespread increases in the east and decreases in the south. Differences between GCM responses exist mainly regarding the season of change; in SON, ACCESS-ESM1.5 shows little response over east Africa, with the largest increases in DJF. IPSL-CM6A-LR instead shows much larger spatial coverage, with increases present in DJF, and less so in SON. While IPSL-CM6A-LR shows decreases in P_{11} over north Africa, neither GCM indicates a sizeable increase in CDD over north and Saharan Africa. This is not altogether surprising, as it is known that CMIP6 models tend to underestimate CDD in this region (Kim et al., 2020). Seasonally varying decreases in P_{11} are present in west Africa, with similar spatial patterns to IPSL-CM6A-LR, though lesser in magnitude. However, in this region ACCESS-ESM1.5 diagnosed P_{00} patterns are generally very similar to IPSL-CM6A-LR. An overall increase in P_{00} and a decrease in P_{11} once more suggests more frequent, longer dry spells in this region, potentially leading to an intensification of CDD over west Africa, and therefore agreeing with IPSL-CM6A-LR responses.

Slight increases in P_{11} are present year-round across most of India. ACCESS-ESM1.5 shows strong patterns surrounding the Himalayas for P_{00}. While this location tends to show similar responses to IPSL-CM6A-LR, the increases east of the Himalayas and decreases just southwest are much stronger and are present most of the year. Like IPSL-CM6A-LR, patterns across north and west China and east Russia are consistent with similar latitudes in Europe and North America for both P_{11} and P_{00}. Conversely, large differences are noted across Southeast Asia, where much stronger increases in P_{00} than IPSL-CM6A-LR are present for much more of the year. Furthermore, in this region there are large decreases in P_{11} from June to November. Combined with year-round increases in P_{00} (also peaking in magnitude from June to November), large increases in the length and frequency of dry spells is suggested, agreeing closer with projections from a range of RCMs predicting increases in CDD and decreases in CWD over several regions in Southeast Asia than IPSL-CM6A-LR responses (Ge et al., 2019).

It is also clear from Figure 6.1 – Figure 6.9 that differences in the way the GCMs resolve large topographical features affects the resultant pattern, evidenced by patterns over the Andes and the Himalayas. In transition probability responses produced by ACCESS-ESM1.5, there are clear spatial patterns that represent the aforementioned mountain ranges. This is much less prevalent in IPSL-CM6A-LR, and in the responses of the wet-day parameters.

A further, major difference between the response patterns diagnosed from IPSL-CM6A-LR and ACCESS-ESM1.5 is present over Australia, where much stronger patterns are visible for the latter GCM. In all seasons, P_{11} decreases with increasing GMST, though with varying magnitudes. Patterns are strongest from May through October. The P_{00} response patterns, while not as strong in magnitude or spatial coverage, show increases concentrated between May and October. New South Wales in particular sees the strongest decreases and increases in P_{11} and P_{00} respectively. This supports CMIP5 projected drying trends in east Australia (Dai et al., 2018).

It is known that inter-model uncertainty is much greater than the uncertainty within each GCM. This is particularly true for modelling changes in precipitation, and, in a recent study, Kim et al. (2020) note that large inter-model spread exists regarding the simulation of *extreme* precipitation. It is therefore not expected that each GCM will produce responses that are identical. Inter-model spread is reflected by some of the differences presented in the response patterns shown here. While similarities exist between the GCMs in the responses of P_{11} and P_{00} to changing GMST over much of Africa, exact geographical location, season, and magnitude of change varies between each model. For example, in SON, little change in P_{11} is projected over east Africa by ACCESS-ESM1.5, with increases in DJF. Conversely, IPSL-CM6A-LR shows increasing P_{11} responses that are very large in magnitude and spatial coverage in SON, with little change in DJF. IPSL-CM6A-LR and ACCESS-ESM1.5 transition probability responses alone do not accurately reflect projected increased aridification and desertification in the arid and semiarid regions of east Africa (Haile et al., 2020). This is not an issue solely present in CMIP6 projections; Cooper et al. (2008) found that in a comparison of 21 CMIP5 GCMs, models do not necessarily agree on positive or negative changes in precipitation trends, with discrepancies across African regions.

Despite some differences between the GCM diagnosed responses, there are several regions where there is good agreement. This tends to be where the magnitude of the pattern is strong or covers large geographical areas. For example, North America and Europe show similar responses with seasonal differences in P_{11} and P_{00} reiterated by locations with similar latitudes. There is generally good agreement between the GCMs that South America will experience a widespread overall decrease in wet days, with increased dry spell length and frequency, though with varying amounts of spatial coverage. Both ACCESS-ESM1.5 and IPSL-CM6A-LR show similar patterns over South American monsoon regions, southern Africa, and Southeast Asia, where there are projected increases to the intensification of consecutive dry days. Both GCMs predicts large increases in P_{00} across South American

monsoon regions and southern Africa, though with largest magnitude shown in ACCESS-ESM1.5 over Southeast Asia. However, Almazroui et al. (2021) state that "CMIP6 models tend to overestimate the CDD over some regions" (page 493), including several locations that the two GCMs produce consistent increasing P_{00} responses, including north and central east Africa and Amazonia, reiterated by the findings of Kim et al. (2020).

6.2.2 WET-DAY GAMMA PARAMETER RESPONSE

Although transition probability responses diagnosed by ACCESS-ESM1.5 and IPSL-CM6A-LR are relatively similar, the same is not true regarding the parameters of the daily wet-day gamma distributions. This highlights some of the large inter-model differences regarding the modelling of precipitation. It is not altogether surprising that there are significant differences in the wet-day parameter responses to GMST increase; it is known that there is much uncertainty in precipitation projections (Kim et al., 2020; Torres and Marengo, 2013). Section 6.1.3.2 demonstrated that it is challenging to make assumptions on the changes to the resultant daily precipitation distribution through the study of changing parameters alone, though a comparison between the modelling centres' response patterns will be discussed.

While IPSL-CM6A-LR shows generally increasing shape parameters over most land-surface areas, ACCESS-ESM1.5 instead shows slight decreases (Figure 6.10). Responses are also smaller in magnitude than IPSL-CM6A-LR. Note the maximum and minimum responses here have been capped at +0.5 and -0.5, respectively. Despite differences in sign, similarities are found regarding locations that show the strongest magnitudes to changing GMST. South America and Africa show seasonally changing responses with large magnitudes, while North America, Europe, Central and East Asia show generally seasonally consistent responses. Much like the transition probability patterns, locations with similar latitudes in the northern hemisphere show similar responses to GMST increase. Strongest decreases to the shape parameter are present during winter (DJF).

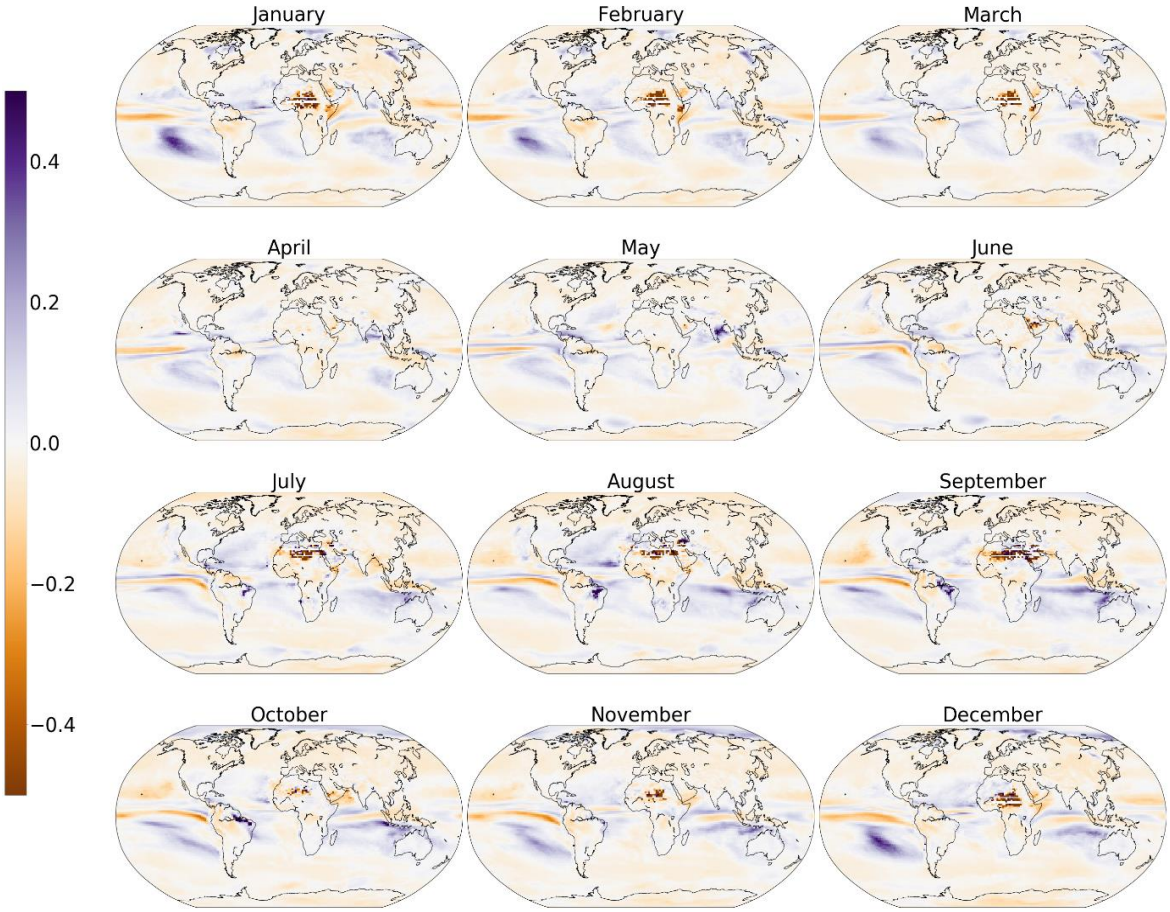


Figure 6.10 – Fractional change in wet-day shape parameter per degree GMST increase, diagnosed from ACCESS-ESM1.5 historical, SSP3-7.0 and SSP5-8.5.

From November to February, spatially dense decreases to the shape parameter, large in magnitude, are present across Central Africa. It is in these months that no responses are diagnosed from IPSL-CM6A-LR due to too few wet days in the simulations. This suggests ACCESS-ESM1.5 altogether projects more precipitation over Central Africa than IPSL-

CM6A-LR. Here, there is little coherence with the transition probability response (where there is little change to P_{00}, with slight increases to P_{11}). It is also these locations that in July to September, IPSL-CM6A-LR shows spatially widespread increases to the shape parameter, large in magnitude. In August and September, north-east Africa and the Arabian Peninsula also show strong decreases. These decreases, in absence of scale responses, indicate a potential decrease in variability and mean of wet-day precipitation amounts alongside greater skew.

Much of southern South America show responses that reflect trends shown over most other land-surface areas, with slight decreases to the shape parameter in several months. It is once again over the Amazon regions and central South America where responses deviate most from other land-surface areas. In these regions, large decreases in shape are present during most months. However, notable increases are also present from July to September in eastern Brazil. Though the spatial coverage is small, the magnitude is large. While the resolution of the topography of the Andes evidently affected transition probability response, there appears to be less influence on shape response.

Further locations with noteworthy differences to larger scale geographic trends include India, north-east Russia and, to a lesser degree, Southeast Asia. The monsoon and rainy seasons in India typically fall between June and September. In May and June, increases to the shape parameter are large in magnitude and cover much of east and central India, suggesting changes to the variability and a decrease in skew of wet-day precipitation at the start of the monsoon season. Increases are also present in north-east Russia in January to March. These regions also show strong transition probability response patterns in all months generally indicating a move towards more wet days (with increasing P_{11} and decreasing P_{00}). In combination with the increasing shape parameter, this may suggest wet days with less consistent precipitation could be expected. Finally, Southeast Asia tends to show decreases that are larger in magnitude than most other land-surface areas in Asia from July to September (and to a lesser degree, most other months), indicating a more skewed distribution. Conversely, increases are present in Indonesia and Papua New Guinea from March to May.

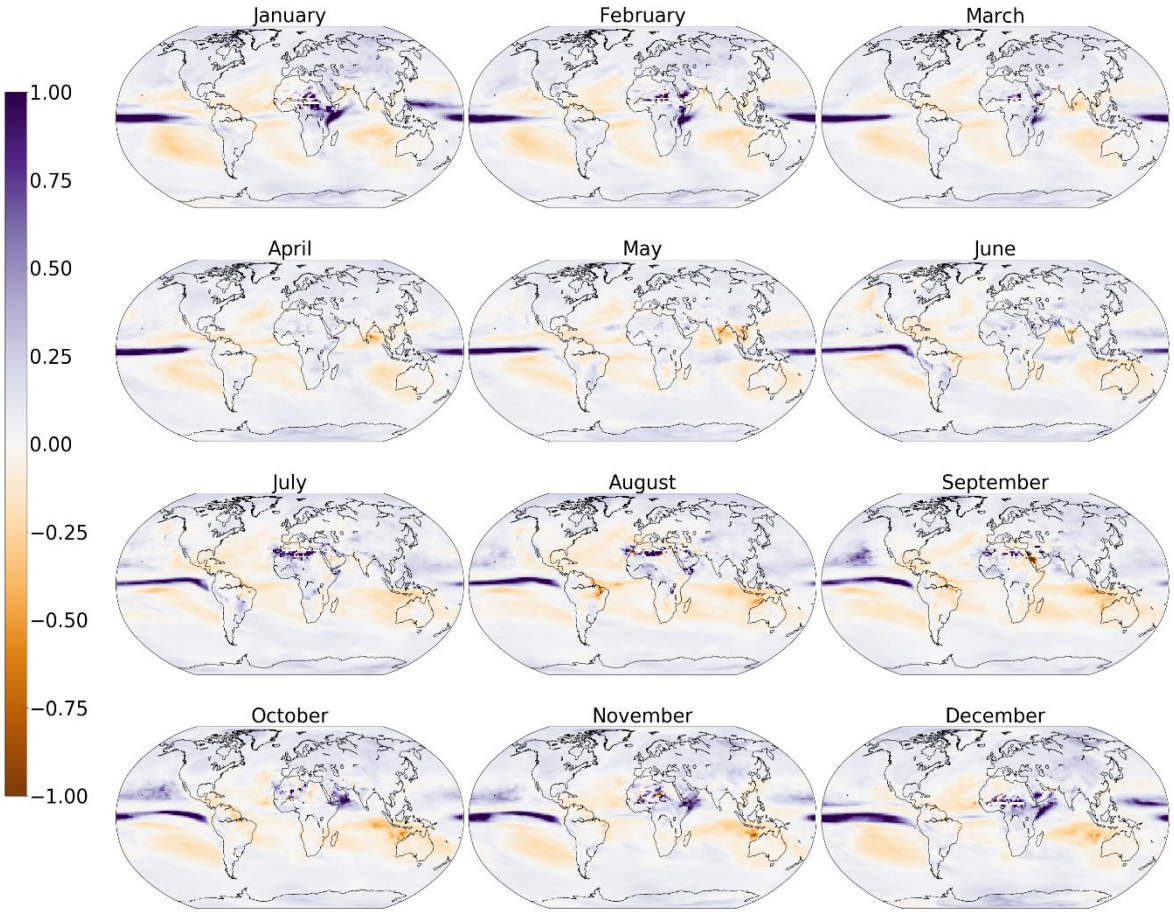


Figure 6.11 – Fractional change in wet-day scale parameter per degree GMST increase, diagnosed from ACCESS-ESM1.5 historical, SSP3-7.0 and SSP5-8.5.

ACCESS-ESM1.5 diagnosed wet-day scale parameter responses show greater magnitude than the shape parameter, and the analogous scale responses diagnosed from IPSL-CM6A-LR – note the colourbar has instead been capped at +1 and -1 for maximum and minimum responses, respectively. Seasonal changes are present in the scale response similar to the shape response. ACCESS-ESM1.5 generally shows a widening of the wet-day precipitation distribution and a move towards larger precipitation values, with increases in scale parameter over most land-surface areas, including North America, Europe, and Asia, where increases are present regardless of season, with magnitudes once again peaking during DJF. Africa and South America once more shown the strongest responses with the most seasonal variation.

Australia shows some spatial and seasonal variability in scale parameter response, though with generally small magnitude. These responses are not dissimilar from the scale responses diagnosed from IPSL-CM6A-LR, where changes small in magnitude are present

in several months. From February to May, an increase in the scale parameter with GMST is present in most regions, indicating a widening of the distribution (similar to several land-surface areas). In May, the east coast shows decreases, which cover an increasingly large area up until October. Unlike IPSL-CM6A-LR, ACCESS-ESM1.5 also shows large increases in dry day frequency and spell length, with P_{00} and P_{11} responses large in magnitude on the east coast from June to October. Changes in the shape are typically small in magnitude, therefore indicating that alongside general drying, when precipitation does occur during these months, smaller amounts with less variability could be expected, further exacerbating drying.

Africa shows much variation in scale parameter response that does not appear to be related to the shape parameter responses. From April to June, most of the continent shows slight increases in scale. This is in stark contrast to the IPSL-CM6A-LR diagnosed responses, where these months show the strongest, decreasing responses over much of Central Africa. ACCESS-ESM1.5 instead shows increases large in magnitude that are present in several other months. From December to March, Central and east Africa show large increases, indicating a shift of the distribution towards larger precipitation values and perhaps greater variability (when considered in absence of shape changes). Of course, in this region changes are present in both the shape *and* scale parameters, leading to uncertainty regarding the resultant changes to the wet-day precipitation distribution.

The scale parameter response varies less so than the shape parameter over South America. Western South America generally shows increases in all months, while the east typically shows decreases. The eastern decreases in scale parameter cover the largest spatial area in August and September. This coincides with the largest increases to the shape parameter. This may indicate a wet-day distribution with less skew alongside an increased load towards smaller precipitation values. However, it is difficult to ascertain changes to the variability, as increasing shape implies greater variability, while decreasing scale results in the opposite. Transition probability responses indicate a large decrease in the total number of wet days (see Figure 6.9), altogether suggesting large changes in the monthly distributions of daily precipitation in this region.

North America, Europe and most of Asia show similar spatial and seasonal trends in the scale response as the shape response, though here with widespread increases (as opposed to decreases in shape). Scale response over India once more show differences to overall trends in the region, with concentrated decreases in the parameter in May and June in the

east. This coincides with large increases in the scale parameter, once again indicating sizable changes in daily precipitation around the start of the monsoon season. Changes in precipitation during the Indian monsoon season have been established by Loo et al. (2015), wherein widespread decreases are noted across central and southern India. Conversely, increased in precipitation is noted in the northeast. The patterns in Figure 6.11 also make a distinction between these regions, with the scale responses differing in sign between these areas.

6.3 APPLICATION OF IPSL-CM6A-LR AND ACCESS-ESM1.5 DIAGNOSED RESPONSES TO OBSERVED DATA

Section 6.1.3.3 demonstrated the application of the scaled response patterns to observed parameters at Santarém, Brazil, and Reykjavik, Iceland. To further validate the application of the technique, compare results from different GCMs, and thus better demonstrate inter-model uncertainties, the responses shown in Figure 6.2 and Figure 6.9 will also be applied to observational data at Santarém and Reykjavik. The precipitation produced by parameters scaled by IPSL-CM6A-LR diagnosed responses will be the same as in Section 6.1.3.3 for Santarém and Reykjavik, though here ACCESS-ESM1.5 diagnosed responses will also be presented. Alongside the reasons outlined in Section 6.1.3.3 for the choice of Santarém and Reykjavik as case studies, the sites are categorised as tropical and temperate, respectively. Chapter 3 determined that the GCWG shows weakest performance in tropical regimes and strongest performance in temperate and continental classifications. Therefore, Santarém and Reykjavik have been chosen to also demonstrate the technique in regions with different GCWG skill. Figure 6.12 demonstrates the locations of both sites.

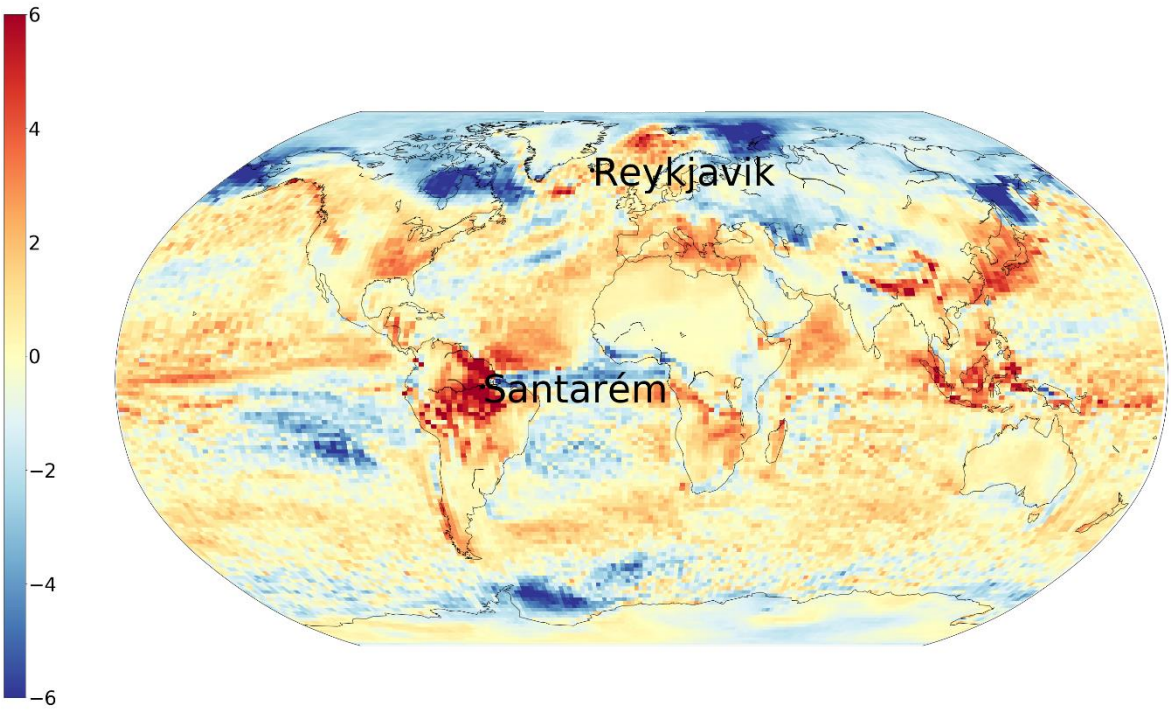


Figure 6.12 – ACCESS-ESM1.5 diagnosed P00 response to increasing GMST in February, showing the locations of Santarém and Reykjavik. February has been provided as an example based on strong responses.

100 years of daily precipitation has been generated at each site using parameters calculated from the observed records. Parameters are then perturbed using the transition probability and wet-day gamma parameter responses diagnosed from IPSL-CM6A-LR and ACCESS-ESM1.5 (presented in Figures 6.2, 6.3, 6.9 and 6.10) following equations 6.7 and 6.8. ΔT has been substituted such that the altered parameters are representative of GWLs of 2.0 and 4.0°C. Precipitation at GWLs of only 2.0 and 4.0°C (excluding 1.5°C) have been produced here for ease of comparison in Figure 6.13 and Figure 6.14. Once again, it is important to note that the observed precipitation records for Santarém and Reykjavik do *not* correspond to the same GMST anomalies relative to the 1850-1900 mean. The mean observed GMST anomaly has been averaged over years in the observed record using the HadCRUT5 dataset (Morice et al., 2021). To reiterate, at Santarém, this is 0.78°C (from 1990 – 2020) and at Reykjavik, the averaged observed GMST anomaly for its reference period is 0.56°C (from 1970 – 2020). 100 years of daily precipitation time series at Santarém and Reykjavik have already been produced using parameters scaled by the IPSL-CM6A-LR responses. Figure 6.13 and 6.14 show the previously presented results alongside 100 years of precipitation generated by parameters scaled by ACCESS-ESM1.5. Precipitation

produced using parameters perturbed by IPSL-CM6A-LR and ACCESS-ESM1.5 diagnosed responses will be referred to as “IPSL-scaled” and “ACCESS-scaled” for simplicity.

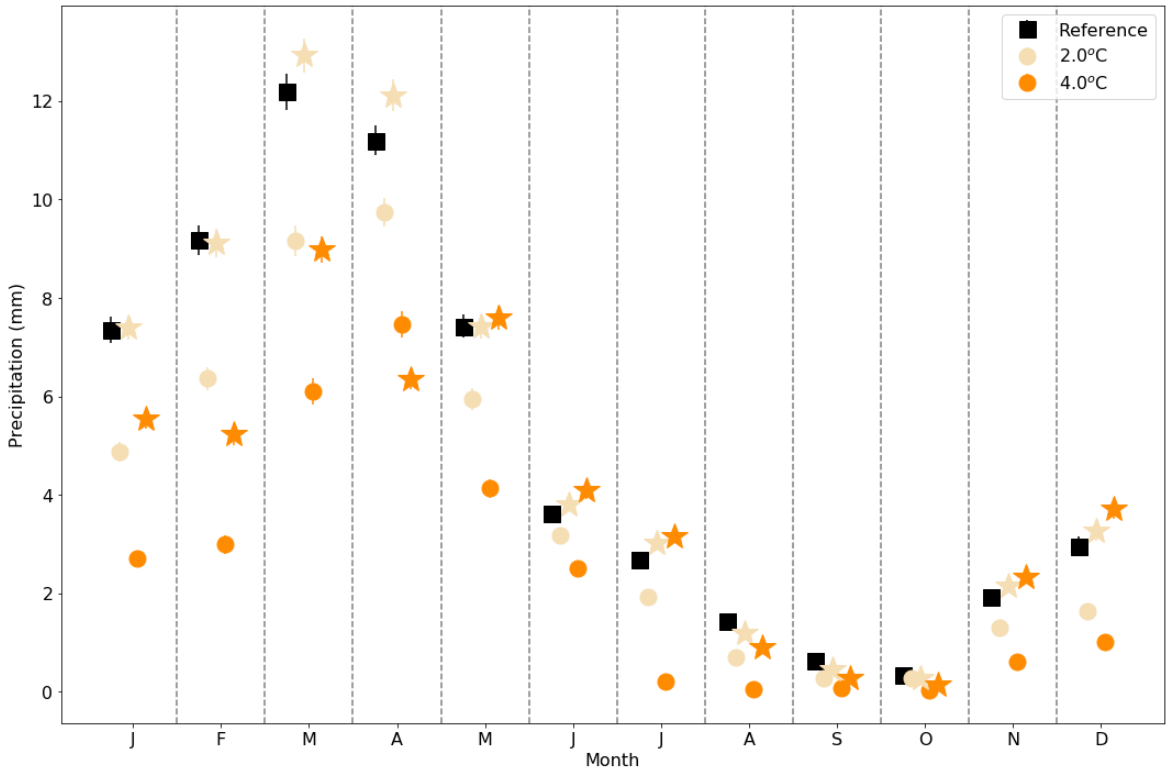


Figure 6.13 – Mean daily precipitation in Santarém during the reference period (black squares), and at the 2.0 and 4.0°C GWLs produced using ACCESS-ESM1.5 (coloured circles) and IPSL-CM6A-LR (coloured stars) precipitation parameter responses. Error bars are the standard error in the mean, hidden by the symbols in some cases where standard error is small.

At GWL 4.0°C, IPSL-scaled and ACCESS-scaled precipitation both show large decreases relative to the reference period during the wet season at Santarém (January to May). ACCESS-scaled precipitation shows an approximately linear decrease in precipitation between the GWLs in all months, and in all but one month (April), decreases exceed those in IPSL-scaled precipitation. IPSL-scaled precipitation, however, shows small increases in March and April at the 2.0°C GWL followed by decreases at 4.0°C The differences in their respective wet-day gamma parameter responses are primarily responsible for the conflicting changes at 2.0°C. This is unsurprising; ACCESS-ESM1.5 shows approximately constant shape response with decreasing scale, resulting in decreasing precipitation. Conversely, the interactions between shape, scale, and precipitation occurrence result in slight increases to the IPSL-scaled precipitation (discussed in more detail in Section 6.1.3.3). Further differences are noted during the dry season. ACCESS-

scaled precipitation clearly depicts a dry season that is at least one month longer, starting in July (instead of August) where, at 4.0°C, there is less precipitation than in any dry-season month during the reference period. Instead, IPSL-scaled precipitation is approximately constant (showing marginal increases) during November and December. ACCESS-scaled precipitation continues to show large decreases, arguably extending the dry season beyond October and into December.

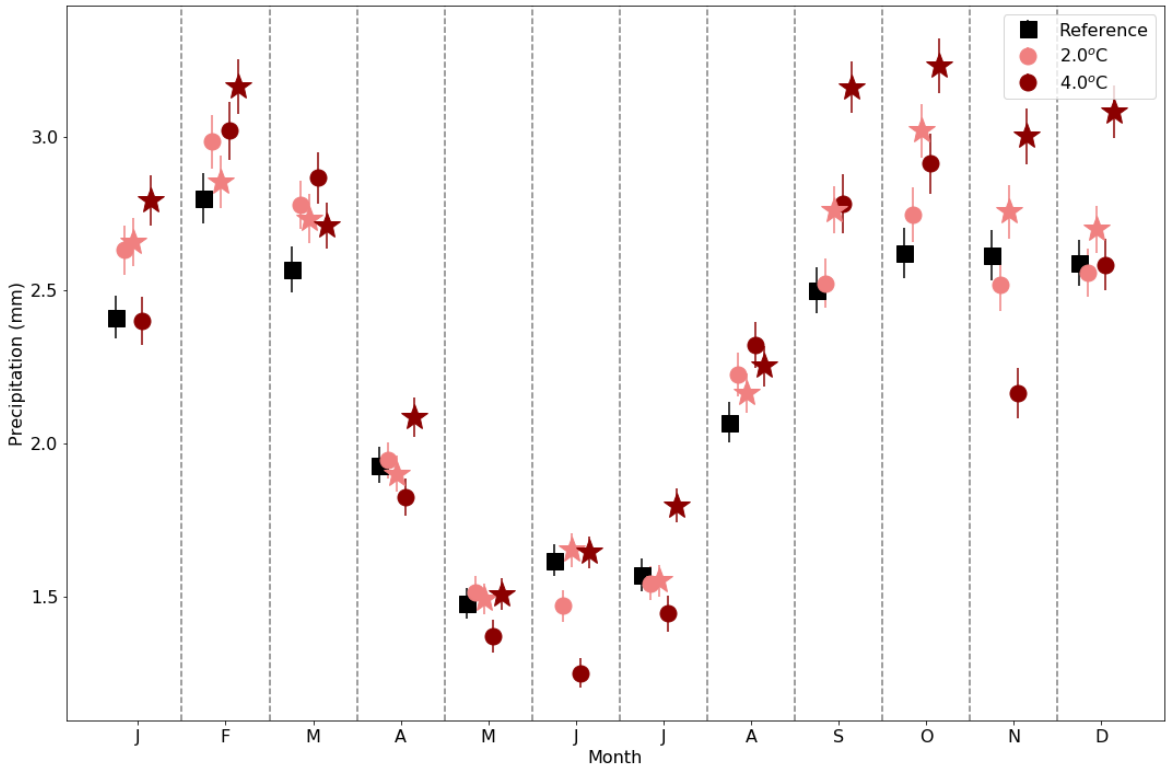


Figure 6.14 – Mean daily precipitation in Reykjavik during the reference period (black squares), and at the 2.0 and 4.0°C GWLs produced using ACCESS-ESM1.5 (circles) and IPSL-CM6A-LR (stars) precipitation parameter responses. Error bars are the standard error in the mean, hidden by the symbols in some cases where standard error is small.

Precipitation changes are smaller in magnitude (and hence severity) at Reykjavik than at Santarém. ACCESS- and IPSL-scaled precipitation both agree with the direction of precipitation change in several months. Both GCM-scaled time series project the largest changes in precipitation during Reykjavik's wettest months (September through March). There are several instances (in both ACCESS- and IPSL-scaled precipitation) where increases (or decreases) are present at 2.0°C followed by a reversal at 4.0°C. ACCESS-scaled examples include January April, May, and December. IPSL-scaled examples include April and July. This is likely caused by interactions between the changing number of wet days per month alongside sizeable changes in the wet-day precipitation distributions,

wherein at smaller GMST increases, either changes to the frequency of wet-day occurrence or changes to the wet-day distribution dominate the resultant changes in precipitation. Precipitation in November exemplifies this. While transition probability response is generally similar between the two GCMs, differences in their wet-day gamma-parameter responses are present. For example, IPSL-CM6A-LR shows a 7% increase in shape per degree GMST increase with approximately constant scale. ACCESS-ESM1.5 instead shows a decrease of 3% in the shape parameter, with an increase of 6% in the scale. These responses ultimately result in an ACCESS-scaled wet-day distribution that produces slightly smaller precipitation values, while the IPSL-scaled wet-day distribution shows much more variability with higher values.

To gain some insight into the differences between ACCESS- and IPSL-scaled precipitation extremes, changes to the 99th percentile of dry-spell length and "50-year" precipitation events will be assessed (changes to the 99th percentile of wet-spell length are much smaller than dry-spell length under each GWL and are therefore not shown here).

Table 6.1 – 99th percentile in dry-spell length distribution under different GWLs. GWLs 0.78°C and 0.56°C correspond to the reference periods at Santarém and Reykjavik, respectively.

		DRY SPELL 99TH PERCENTILE (DAYS)	
	GWL (°C)	**IPSL-CM6A-LR**	**ACCESS**
Santarém	0.78	33.0	33.0
	2.00	40.8	64.0
	4.00	63.4	137.0
Reykjavik	0.56	11.0	11.0
	2.00	12.0	12.0
	4.00	12.0	14.0

At Santarém, each GCM projects large increases in the 99th percentile of dry-spell length. This is unsurprising due to the strong decreasing and increasing P_{11} and P_{00} trends respectively in both GCMs. ACCESS-scaled precipitation shows an increase in the 99th percentile of dry spell corresponding to an increase of approximately four times the reference period. Large increases in the 99th percentile of dry-spell length in the ACCESS-scaled precipitation are also unsurprising due to the addition of 3 months to the dry season

(Figure 6.13). While this is an exceptionally large increase, it is known that CMIP6 models overestimate consecutive dry days over much of South America, including Santarém's location (Almazroui et al., 2021; Kim et al., 2020), and that perhaps in this instance, the precipitation generated by IPSL-CM6A-LR may be a better reflection of future trends. Reykjavik shows smaller changes to the 99th percentile of dry-spell length, though this is expected. Both GCMs project changes over Iceland that are much smaller in magnitude than Brazil, with relatively small changes to the daily precipitation.

In Chapter 4, Section 4.1.2, the return periods of extreme events were studied at five weather stations (including Reykjavik). It was noted that the GCWG (much like many stochastic weather generators) tends to underestimate the frequency of extreme precipitation, wherein return periods for 10-year and 50-year extreme events calculated directly from the observed data are largely underestimated (shown in Chapter 4, Section 4.1.2). Instead, a comparison between the *generated* return periods at different GWLs relative to the reference period was advised to reduce GCWG errors in simulating such extremes. To demonstrate this, design values for 50-year return periods are calculated directly from the observed weather station data at both Santarém and Reykjavik (following the same equations 4.5 – 4.7). Using 500 years of generated data (with unscaled parameters, calculated from the observed record), the generated return period corresponding to the *observed* 50-year design value is calculated from the reference run. The return periods for the *observed* 50-year design value are then calculated for GWLs 2.0 and 4.0°C during the wettest month of the observed record. Despite significant underestimates in the generated return periods (i.e., the reference period), events will still be referred to as a "50-year" event. The return periods at different GWLs can then be compared to the generated return period for the reference period to establish a multiplication factor (i.e., the return period at each GWL divided by the generated reference return period). This multiplication factor can then be used to estimate the future frequency of such an event (i.e., multiplication factor multiplied by 50-years).

Table 6.2 – Return periods for "50-year" extreme precipitation generated using observed parameters (reference period) and parameters scaled for GWLs of 2.0 and 4.0° C during the wettest month of the year. Design values are calculated from the observed record. Return periods are calculated from the generated data corresponding to the relevant design value. GCWG "50-year" Return Period is the "50-year" return period calculated from the stochastically simulated reference run using the observed design value.

				RETURN PERIOD (YEARS)	
Location	**Design Value (mm)**	**GCWG "50-year" Return Period (yrs)**	**GCM**	**GWL 2.0°C**	**GWL 4.0°C**
Santarém	158.2	111	IPSL-CM6A-LR	200	1000+
			ACCESS-ESM1.5	250	1000+
Reykjavik	35.6	140	IPSL-CM6A-LR	101	74
			ACCESS-ESM1.5	67	45

Santarém sees a large increase in the return periods of observed "50-year" precipitation, resulting in a decreased frequency of such occurrences. This is unsurprising due to the large increases in dry-spell length, caused by large increases in P_{00}. While both GCMs predict at least a tenfold increase to the return period of "50-year" extreme precipitation by 4.0°C (and thus a decrease in the frequency of occurrence), this varies at 2.0°C, wherein IPSL-scaled precipitation shows an increased return period of 1.8 times the reference period, and ACCESS-scaled 2.25. While P_{00} response is similar between the two GCMs, the magnitude of the P_{11} response is more varied. It is evident from Figure 6.1 that IPSL-CM6A-LR projects weaker and less spatially widespread P_{11} responses than ACCESS-ESM1.5 over much of South America. This results in a P_{11} response of only -0.96%/°C under IPSL-CM6A-LR (in February) where ACCESS-ESM1.5 has a response of -3.2%/°C.

Conversely, in Reykjavik, the return period for extreme precipitation decreases under both GCMs, resulting in greater frequency of "50-year" precipitation, though with large deviation between the GCMs. A decrease by a factor of 3.1 has occurred from the reference period to a GWL of 4.0°C under ACCESS-ESM1.5. This multiplication factor, when applied to the *observed* data (i.e., not to the generated GCWG-reference data) would suggest a 50-year flood occurring instead every 16 years. It is interesting that the return

period under ACCESS-ESM1.5 shows a greater decrease in return period than IPSL-CM6A-LR. From Figure 6.14 alone, it is clear that IPSL-scaled precipitation generally shows larger increases than ACCESS-scaled. Despite this, larger decreases in P_{11} are diagnosed by IPSL-CM6A-LR than ACCESS-ESM1.5 in February. This suggests that there will likely be longer wet-spells in the ACCESS-scaled precipitation than IPSL-scaled, perhaps contributing to the larger increases to the occurrence of "50-year" precipitation.

6.4 CHAPTER 6 SUMMARY

This chapter introduces the pattern scaling technique to the input parameters of the GCWG, utilising the advantages of both PS and SWGs. SWGs are an ideal tool for producing long, temporally consistent time series for a suite of weather variables. They are commonly used to produce high-quality data as an input for impact assessments in several fields, including hydrology, agriculture, and ecology. Capable of producing very long time series with computational efficacy, SWGs are also ideal for robustly assessing extremes at a specified GWL. Although a widely used statistical downscaling technique, SWGs can only produce time series corresponding to climate scenarios and time periods that GCMs have simulated and from which changes in the key parameter values can be determined. The few scenarios GCMs have simulated alone cannot be used to fully address climate uncertainty. Herein lies the gap that this chapter (and Chapter 7) has addressed.

PS is a computationally efficient technique that can be used to emulate a range of climate scenarios and time periods, though typically only changes in mean climate have been studied, generally over monthly, seasonal, or annual temporal scales. This chapter instead applies PS to daily precipitation statistics. The principles of PS have been applied to the input parameters of a SWG (here, the GCWG), including the transition probabilities (P_{00} and P_{11}) and wet-day shape and scale parameters. This provides a technique that not only emulates changes in mean local-scale climate at specified GWLs, but also better represents changes to local-scale variability, therefore allowing more robust estimates regarding the changing risk of extremes. A proposed manuscript for publication in *Climatic Change* detailing the responses of the GCWG parameters (transition probabilities and wet-day gamma parameters) diagnosed from the IPSL-CM6A-LR GCM and subsequent application to weather station data at Santarém, Brazil, and Reykjavik, Iceland has been included. The responses of the GCWG parameters have additionally been diagnosed from the ACCESS-ESM1.5 GCM to gain insight into the inter-model uncertainties in stochastic weather generator response patterns.

There are several similarities between the GCM-diagnosed transition probability response patterns. Similarities primarily exist in locations with responses large in magnitude or spatial area, resulting in agreement over much of North America, Europe (where responses are large in geographic coverage) and South America (where responses are strong in magnitude). In locations where similar transition probability responses are present, trends generally agree with projections in literature. Several similarities exist regarding the geographic location of strong shape and scale responses between the two GCMs, though oftentimes with opposing magnitudes. It is known that differences between the GCMs themselves are the largest cause of uncertainties in the diagnosis of response patterns. This is evidenced by discrepancies in the wet-day parameter responses, and transition probability response over Australia, southeast Asia, and east Africa.

Daily precipitation time series have been produced for Santarém, Brazil and Reykjavik, Iceland, using responses diagnosed from the two GCMs for GWLs of 2.0 and 4.0°C. This has demonstrated the application of the pattern scaling technique through perturbing local-scale parameters (calculated directly from observational records) by the area-averaged relationships to increasing GMST diagnosed from the GCMs. Stochastic weather generators can produce long records at a specified GWL, allowing robust calculation of the risk of extreme events. This has been demonstrated here; the 99th percentile of dry-spell length, and changes to the return period of a 50-year extreme precipitation have both been calculated from a reference period and at the two GWLs.

It is of course exceptionally important that there are robust methods for simulating future precipitation for a range of climate scenarios to better understand uncertainties, especially considering the known inter-model differences in GCM precipitation projections. To reiterate, the advantages of both stochastic weather generators and pattern scaling have been utilised here. Long, daily time series can be produced at a specified GWL, incorporating changes to both the mean and variability at a local scale, suitable for use in impact assessments or as an input to hydrological models. This has avoided GCM-induced bias in local-scale simulations whilst providing a method of robustly assessing the risk of extreme events at different GWLs. This method of producing local scale time series has advantages over several traditional pattern scaling techniques due to the inclusion of changes in the variance through the scaling of the wet-day gamma parameters on daily temporal scales.

7 PATTERN SCALING THE SECONDARY VARIABLE WEATHER GENERATOR PARAMETERS

Chapter 6 has demonstrated the application of the pattern scaling (PS) technique to the precipitation parameters (transition probabilities and wet-day gamma parameters) of the **G**lobally **C**alibrated stochastic **W**eather Generator (GCWG) developed in Chapters 3 to 5. While future changes to precipitation are exceptionally important to consider with regards to the assessment of impacts which a changing climate may have on the hydrological cycle, water resource management, and to the frequency and severity of extreme precipitation, considering changes in the distribution of maximum and minimum daily temperatures is also of great value. Changes to the maximum or minimum temperature may have sizeable impacts on human comfort and mortality, crop yields, and the exacerbation of the conditions required for drought or wildfire ignition through evapotranspiration (Li et al., 2009; Luber and McGeehin, 2008; Seneviratne et al., 2021; Stott et al., 2004). This chapter is analogous to Chapter 6, though instead applying PS to a selection of GCWG temperature parameters (introduced in Chapter 4).

In the GCWG, temperature is considered a secondary variable and is generated following the simulation of a synthetic precipitation time series. To recap, observed minimum and maximum temperatures (T_x and T_n, respectively) are transformed into a residual series using means and standard deviations that have been calculated bimonthly (i.e., from days 1-15 for the first half of the month, and days 16 onwards for the second half of the month) and conditioned on precipitation transition state. Precipitation transition state (or precipitation status) refers to the precipitation on the day of the temperature measurement *and* the preceding day, with one of four states possible: WW, WD, DW, or DD (where, for example, WD refers to a dry day preceded by a wet day). This results in the calculation of 8 means and standard deviations per month. A linear regression model with equations that are dependent on the precipitation transition state on a given day is used to generate a residual series of T_x and T_n. Following the transformation, regression coefficients are fitted monthly for each precipitation state, resulting in a total of 4 regression equations per month. Predictors in the linear model are also dependent on the precipitation state (see equations 4.9 – 4.12, Chapter 4). The residual series is then generated using the synthetic precipitation time series also produced by the GCWG. Residuals are converted back into temperature

values using the corresponding observed mean and standard deviation for the precipitation status and day of the month.

Of these various parameters (means and standard deviations of T_x and T_n, together with the regression coefficients that link the temperature residual series to the predictors), it is the conditional, bimonthly means and standard deviations in T_x and T_n that will be scaled with increasing global mean surface temperature (GMST). Regression coefficients will not be scaled as a function of GMST change and will therefore be assumed to remain constant regardless of GWL. There is little literature assessing whether the intervariable correlations between temperature and precipitation will scale linearly (or change at all) as a function of GMST, and therefore is beyond the scope of this research. However, it is known that temperature changes scale more linearly with GMST change than precipitation patterns (Tebaldi and Arblaster, 2014), partly due to a smaller component of natural variability that affects the estimates of the forced signal of temperature change compared to precipitation. However, as with precipitation scaling, temperature is usually scaled over larger temporal scales, for example monthly, seasonally, or annually. Here, daily statistics are scaled whilst additionally incorporating conditioning on precipitation status.

The changes in local-scale temperature parameters are assumed to be reasonably represented by changes in General Circulation Model (GCM) grid-scale temperature. GCM grid-scale responses of the bimonthly mean and standard deviations in daily T_x and T_n, conditioned on precipitation status, to changing GMST will be diagnosed analogously to the transition probability response. The changes to a bimonthly mean or standard deviation, S, in T_x and T_n at a GCM grid-cell with latitude and longitude coordinates x and y will be regressed using linear least squares regression against the change in GMST anomaly (relative to 1850-1900), such that

$$S_{xyn} - S_{xy,n=0} = a_{S_{xy}}(T_n - T_{n=0}) + c_{S_{xy}} \quad (7.1)$$

where n is any 30-year window and with $n = 0$ referring to the first 30-year window used to calculate parameters (1850-1880). Advantages of using linear least squares regression instead of alternative methods (such as the time-slice method) are given in Section 6.1.2, though in short, regression methods reduce the impact of noise caused by internal variability relative to the alternative time-slice approach and effectively utilise the full simulated time series. The gradients of the regressions, $a_{S_{xy}}$ and the intercept, $c_{S_{xy}}$ will differ for each grid-

cell. The gradients will be plotted on a map to create the resultant spatial response pattern for each mean and standard deviation (i.e., 8 per month).

A similar issue to calculating the gamma parameters is present here. In some half-months, there may be no recorded temperatures for a given transition state. For example, in a given half-month there may be no wet days following dry days, resulting in no data for the DW transition state. While no shape or scale parameters were estimated for similar occurrences in Chapter 6, here the temperature sample size correction process (defined in Chapter 4, Figure 4.5) will be used to estimate a temperature corresponding to the half-month and transition state using a minimum sample size of 25 daily values. This increases the computational time required significantly, and as such only responses for GCM grid-cells with a land-surface area equalling 25% or more will be determined. Due to the added computational complexity required in determining bimonthly temperature means and standard deviations that are conditioned on the precipitation status, only one GCM has been used to construct temperature parameter response patterns to provide a prototype to demonstrate the approach. Furthermore, instead of calculating temperature parameters over 30-year periods shifted in steps of 5 years, steps of 10 years were used. This is also to compromise between reducing the computational time required to diagnose the response and ensuring a suitable signal-to-noise ratio.

Only ACCESS-ESM1.5 will be used to diagnose the response of means and standard deviations in T_x and T_n to GMST change (Ziehn et al., 2020). While this will not provide an insight into inter-model uncertainties associated with temperature projections, there is generally greater agreement regarding the resultant temperature response patterns than precipitation responses (Tebaldi and Arblaster, 2014). As in Chapter 6, ScenarioMIP SSP5-8.5 and SSP3-7.0 and CMIP6 historical projections have been used to diagnose parameter response, with corresponding realisations of historical and ScenarioMIP simulations concatenated. Concatenated historical and SSP3-7.0 and SSP5-8.5 simulations will be referred to as historical+SSP3-7.0 and historical+SSP5-8.5 respectively. Due to time constraints, only two realisations each of historical+SSP5-8.5 (r1 and r2) and historical+SSP3-7.0 (r3 and r4) have been used to diagnose response patterns, resulting in *four* (as opposed to eight) pooled time series. Corresponding historical+SSP5-8.5 and historical+SSP3-7.0 ensemble members for precipitation must be used to determine the precipitation status of a given day. Strong forcing scenarios have been used once again to ensure a robust signal-to-noise ratio and to increase accuracy though interpolating scenarios as opposed to extrapolating. While the use of only four pooled scenarios (as opposed to the

eight used in Chapter 6) may reduce the signal-to-noise ratio and potentially obscure some of the climate change signal from internal variability, these errors are less pronounced in comparison to precipitation projections.

Regression coefficients will be assumed to remain constant with changing GMST, and only the bimonthly means in maximum and minimum temperature will be scaled with GMST. Responses will once again be used to perturb parameters calculated directly from observations, this time additively, where the new mean or standard deviation, S_T, following a GMST of ΔT relative to the *observed* period is equal to

$$S_T = S_O + a_{S_{xy}} \Delta T \tag{7.2}$$

where S_O is the parameter calculated from the observed (local-scale) record. Perturbed bimonthly means and standard deviations can hence be used to produce synthetic time series at different global warming levels (GWLs) using the corresponding precipitation time series produced by the shape, scale, and transition probabilities that have been analogously scaled by the ACCESS-ESM1.5 diagnosed precipitation parameter responses. This will provide local-scale time series that maintain intervariable correlations while avoiding GCM induced errors in simulating climate at a high resolution.

In this chapter, GCM grid-scale bimonthly mean and standard deviation T_x and T_n global responses to increasing GMST will be presented, conditioned on the precipitation status of the day. ACCESS-ESM15 diagnosed precipitation (see Chapter 6, Section 6.3) and temperature parameter responses will hence be used to perturb observed parameters at Santarém, Brazil, and Reykjavik, Iceland to produce daily T_x and T_n time series at GWLs of 1.5, 2.0 and 4.0°C.

7.1 SECONDARY VARIABLE RESPONSES

The response of bimonthly temperature means and standard deviations used to convert observed temperatures to a residual series (and hence from a generated residual series to temperature values) have been diagnosed using the ACCESS-ESM1.5 GCM. Means and standard deviations in T_x and T_n have been calculated for the first and second half of each calendar month, considering four precipitation transition states: DD, WD, DW, and WW. Seasons are referred to by their months, where SON refers to September, October and November, DJF December, January, February, MAM March, April, May, and JJA June, July, August. Though responses have been determined bimonthly, for concision and to better

demonstrate seasonal changes, the responses have been averaged over SON, DJF, MAM and JJA. This corresponds to an average of 6 responses (first and second half of the month for each of the three months) averaged per season and per precipitation transition state.

Table 7.1 – Global land and annual average grid-cell mean change (°C) per degree GMST increase (°C) (in units of °C /°C) on days with precipitation transition states DD, DW, WD and WW (here the response has been averaged using the responses from both the first and second half of a month).

	MEAN	
Precipitation Status	T_x	T_n
DD	1.42	3.71
WD	1.33	3.56
DW	1.30	3.31
WW	1.22	2.94

Mean T_x and T_n responses over all land-surface areas have been averaged for each precipitation status, with the averaged responses shown in Table 7.1.When averaged over all land-surface areas, grid cell increases in T_x and T_n are larger than the rate of GMST increase. It is virtually certain that warming over land exceeds ocean areas, resulting in warming rates that exceed the GMST increase (Lee et al., 2021). It is also evident that grid cell temperature increase on dry days exceeds increases on wet days and that the daily minimum temperatures are rising at a quicker rate than the daily maximums. This is an expected result; it is widely known that night-time temperatures (where daily minima typically occur) have been increasing at a greater rate (approximately double) than day time maximums over land-surface areas (Alexander et al., 2006; Salinger, 2005). This is thought to be caused by increased levels of mean daily cloud cover that dampen daytime warming relative to night-time (Cox et al., 2020). ACCESS-ESM1.5 diagnosed responses approximately reiterate this, with minimum daily temperatures warming at a rate between 2.3 and 2.7 times quicker than maxima (depending on precipitation transition state).

7.1.1 DAILY MAXIMUM MEAN TEMPERATURE RESPONSE

The response of GCM grid-scale mean daily maximum temperature to increasing GMST has been diagnosed from the ACCESS-ESM1.5 GCM. In Figure 7.1 and Figure 7.2

responses have been capped at -1°C and 5°C per degree GMST (units hereafter referred to as °C/°C) increase for ease of viewing, though in some locations the responses may be greater. Mean T_x responses vary substantially across the globe, though perhaps with less spatial variability than precipitation. Exceptionally strong responses are present in South America, and over high-latitudes in North America and Eurasia, with several other land-surface areas showing regional mean T_x increases that are larger than 1°C/°C, including Australia, northern and southern Africa, southern Europe, the Arabian Peninsula, and India. While not specifically considering daily maximum temperatures, a similar global pattern is shown in Chapter 4 of the IPCC WGI, Figure 4.31, at different levels of global warming, where North American and Eurasian high-latitudes, Amazonian South America, central Eurasia and northern and southern Africa show the strongest warming (Lee et al., 2021). Patterns are generally similar between the precipitation states, with similarities predominantly present on dry days (DD and WD, Figure 7.1) or wet days (DW and WW, Figure 7.2). Areas that show strong patterns will be discussed in turn.

In northern Eurasia, increases in mean T_x that far exceed 1°C/°C are geographically widespread for each precipitation status, with the strongest responses in the highest latitudes. Differences in response are small between precipitation transition states, with the strongest responses for all states occurring in the north during the autumn months (SON). Though slightly smaller in magnitude, spatially widespread increases that far exceed 1°C/°C are present in north and northeast Eurasia throughout all other seasons. Responses over central Asia are geographically extensive during summer (JJA) for all precipitation statuses. DD days show smaller increases during winter, while all other precipitation statuses show slightly stronger patterns in all months except December and January. In central and southern Europe, mean T_x responses are closer to 1°C/°C for many months of the year. However, during the summer (JJA), all precipitation statuses show strong positive responses, indicating a large rise in mean T_x that is generally independent of precipitation status. Therefore, regardless of any projected changes in precipitation, it is likely that the greatest warming over the Mediterranean region may be experienced in the summer. This is in agreement with the findings of Lionello and Scarascia (2018), who note larger increases in T_x during the summer months than the winter months, when more surplus energy is partitioned into sensible heat transfer.

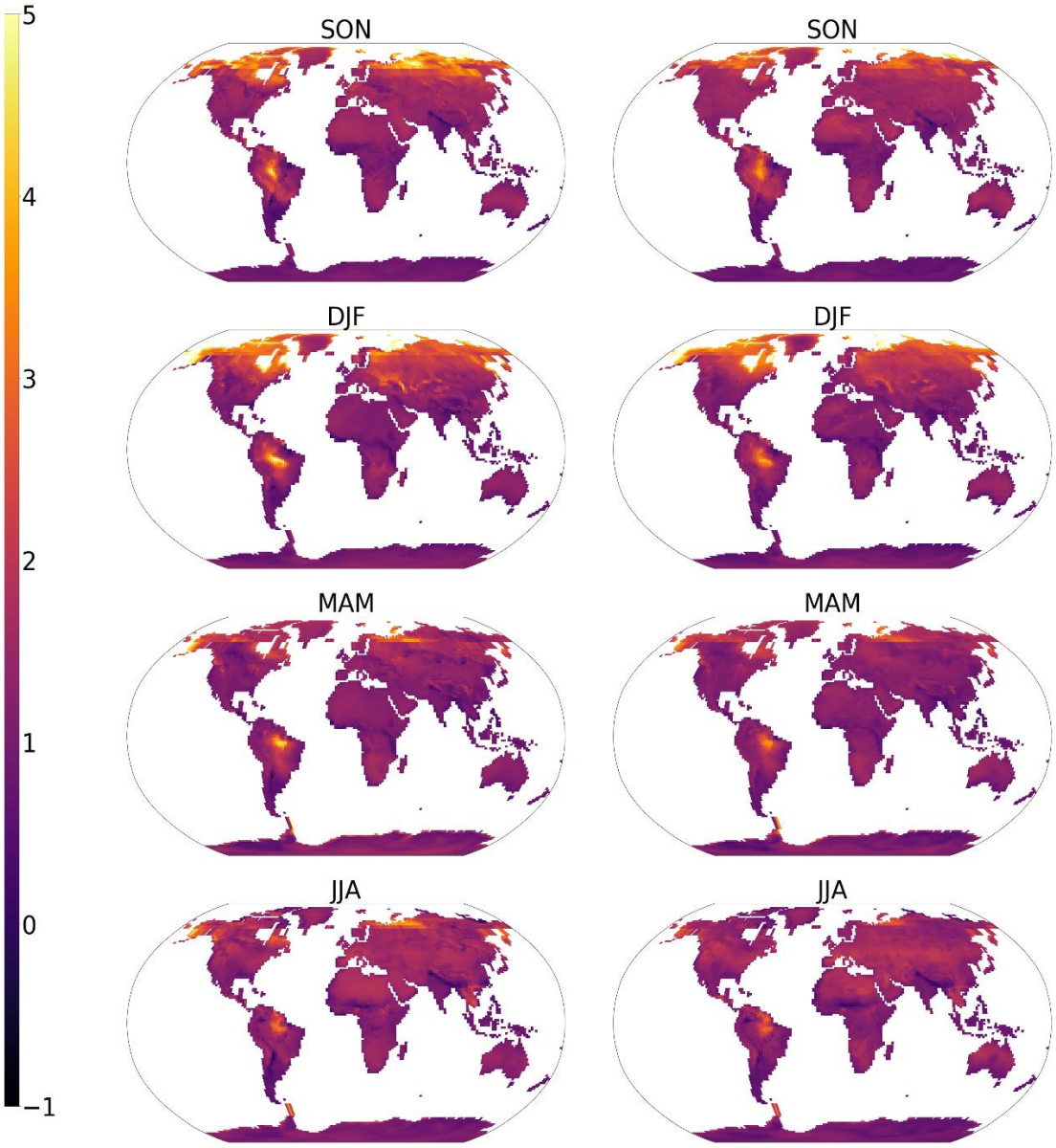


Figure 7.1 – Change in mean maximum daily temperature (°C) per degree GMST increase(°C) seasonally averaged using all bimonthly means within each season on DD days (left) and WD days (right). Colour bar is in units of °C /°C.

Subarctic land surface areas show strong responses to GMST increase regardless of precipitation occurrence or continent (i.e., both Eurasia and North America), most prominent in SON and DJF (with strongest magnitude and spatial coverage in January and February). These responses are unsurprising due to polar amplification, a known phenomenon of anthropogenic climate change that results in larger changes to the surface temperature of the high latitudes than the global average (Lee et al., 2021). Several feedback mechanisms are present in the high latitudes, including the surface albedo, lapse rate, and Planck feedbacks

(Smith et al., 2019). While these mechanisms are present at other latitudes, the mechanisms in combination in relation to other latitudes is largely responsible for polar amplification. Lapse rate is responsible for negative feedback in lower latitudes and positive feedbacks over the high latitudes (during SON and DJF months), thus intensifying polar amplification. Positive lapse rate feedback arises when the surface warms at a greater rate than the atmosphere, resulting in decreased cooling from long-wave radiation emitted to space (Boeke et al., 2021). Planck feedback is the rate at which the loss of infrared energy per unit of vertically uniform warming of the surface and troposphere is increasing (Cronin, 2020). While Planck feedback is negative everywhere, the effect is smaller at the poles, thus reducing the cooling experienced at these high latitudes relative to other latitudes. Finally, with decreasing sea ice coverage as the planet warms, surface albedo at the poles contributes to positive feedback. The strongest mean T_x responses over the subarctic regions also typically coincide with the lowest sea ice cover (usually mid-September) (Belchansky et al., 2004; Serreze and Meier, 2019). This may accelerate the rate of sea ice melt further or increase the length of the melt season. These mechanisms are clearly represented by a strong polar response that is presented in these responses, and features relating to polar amplification will continue to be present throughout the global patterns that are presented in this chapter.

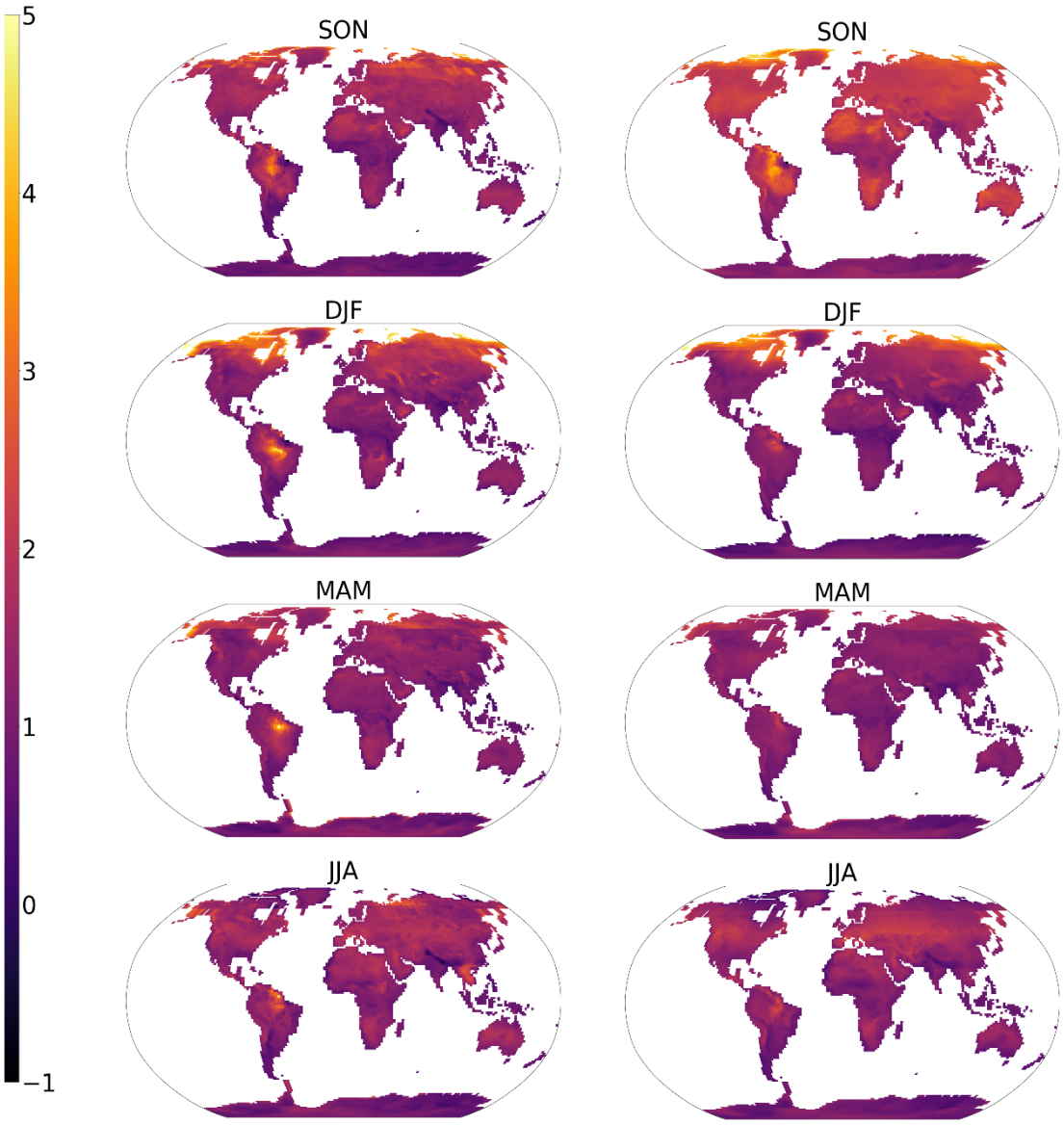


Figure 7.2 – Change in mean maximum daily temperature (°C) per degree GMST increase(°C) seasonally averaged using all bimonthly means within each season on DW days (left) and WW days (right). Colour bar is in units of °C /°C.

South America shows strong, spatially dense increases in mean T_x per degree GMST increase regardless of precipitation status. The strongest responses tend to fall on any dry day, where mean T_x increase exceeds the GMST increase substantially in several months over the Amazon and South American monsoon regions. Increases in DD temperature are particularly alarming when considered alongside the increasing responses of P_{00} in Section 6 and align with the studies of Vogel et al. (2020), who demonstrated large increases in the frequency of concurrent hot and dry extremes in the Northeast Brazil and Amazon rainforest regions. This also aligns with several other studies, projecting large increases in the

occurrence of extremely hot days and heatwaves (Marengo et al., 2009; Vincent et al., 2005). Though large increases are present on DD days in all months, the temperature on WD days shows greater increases during the dry season (August to November). The mean maximum temperature on WW days shows a weaker response to GMST increase, with changes close to 1°C/°C in DJF and MAM, (though in particular, December, January, May and June). This is during the wet season over the Amazon rainforest. In combination with the increasing frequency of dry days and length of dry spells, it is likely that increases in mean T_x may be dominated by the WD and DD day responses.

While transition probability response across Africa indicated localised, seasonal changes to precipitation occurrence, mean T_x responses are more geographically consistent across the continent. Regardless of precipitation status, responses *below* 1°C/°C frequently occur in several seasons, predominantly in December and January, with DD days showing the smallest increases in T_x over the most months (October to January) across equatorial Africa. Regardless of strength of response or precipitation status, equatorial Africa shows weaker responses than the north and south. The spatial changes approximately resemble a distinction between African drylands (in the north and south) and non-drylands. The responses imply increased warming over the dryland regions. Responses in southern Africa are unsurprising as it is thought that warming is exacerbated due to strong soil-moisture-temperature coupling (Gevaert et al., 2018; Lee et al., 2021). ACCESS-ESM1.5 diagnosed responses show large increases in P_{00} from June to September over southern Africa, suggesting increased drying with GMST increase. These conditions may lead to an increased risk of drought and may have negative implications on human health and agriculture.

Further hotspots of large mean T_x increase (that have not already been discussed) include Vietnam, Cambodia and Thailand. On DD and WD days, increases are strongest in August, with WW days showing the strongest responses in June, and DW days in June and July. These strong responses coincide with the start of the southwest summer monsoon season, where the probability of a wet days (WW or DW) will be higher than dry days. This suggests future warming will most likely follow the responses on wet days more closely than dry days.

7.1.2 DAILY MAXIMUM TEMPERATURE STANDARD DEVIATION RESPONSE

Regions that show the largest changes to mean T_x typically also show strong daily standard deviation responses. Most notably, these areas include the Amazon and monsoon regions of South America and the high latitudes of North America and Eurasia. These regions also show seasonal fluctuations regarding spatial coverage, magnitude, and direction of the response. Strong responses are also noted across Africa and Australia. Responses shown have been capped at -0.5°C/°C and 0.5°C/°C change in standard deviation. Dry day (DD and WD) and wet day (DW and WW) responses show several similarities, usually only differing slightly regarding spatial coverage. For this reason, only the DD and WW responses have been included here, averaged seasonally (for similar reasoning as Section 7.1.1).

T_x standard deviation responses across Eurasia show large seasonal and spatial variations. There are sizeable seasonal differences in the responses across Europe, with large differences present on days with differing precipitation statuses. In southern and eastern Europe, the strongest responses are on DD days in SON and MAM (transition seasons), where widespread decreases in standard deviation are present. Conversely, there is little response on WW days during these months. Instead, increases on WW days are present in several other months, though with less spatial coverage and magnitude than DD responses, peaking in geographic coverage during the summer (JJA). On DD days, northern Europe and much of Russia show widespread decreases in several months, though strong increases are also noted in JJA. Summer changes coincide with strong increases to P_{00} and decreases in P_{11} (Figure 6.9) resulting in a drying of the region. This drying may therefore be responsible for the increases to the standard deviation; drier soils can reduce the buffering effect of heat going into evaporation rather than increasing the temperature, therefore resulting in a higher standard deviation than wetter soils. Responses on WW days reiterate DD decreases in December to March, followed by almost exclusively increasing responses present from June to September.

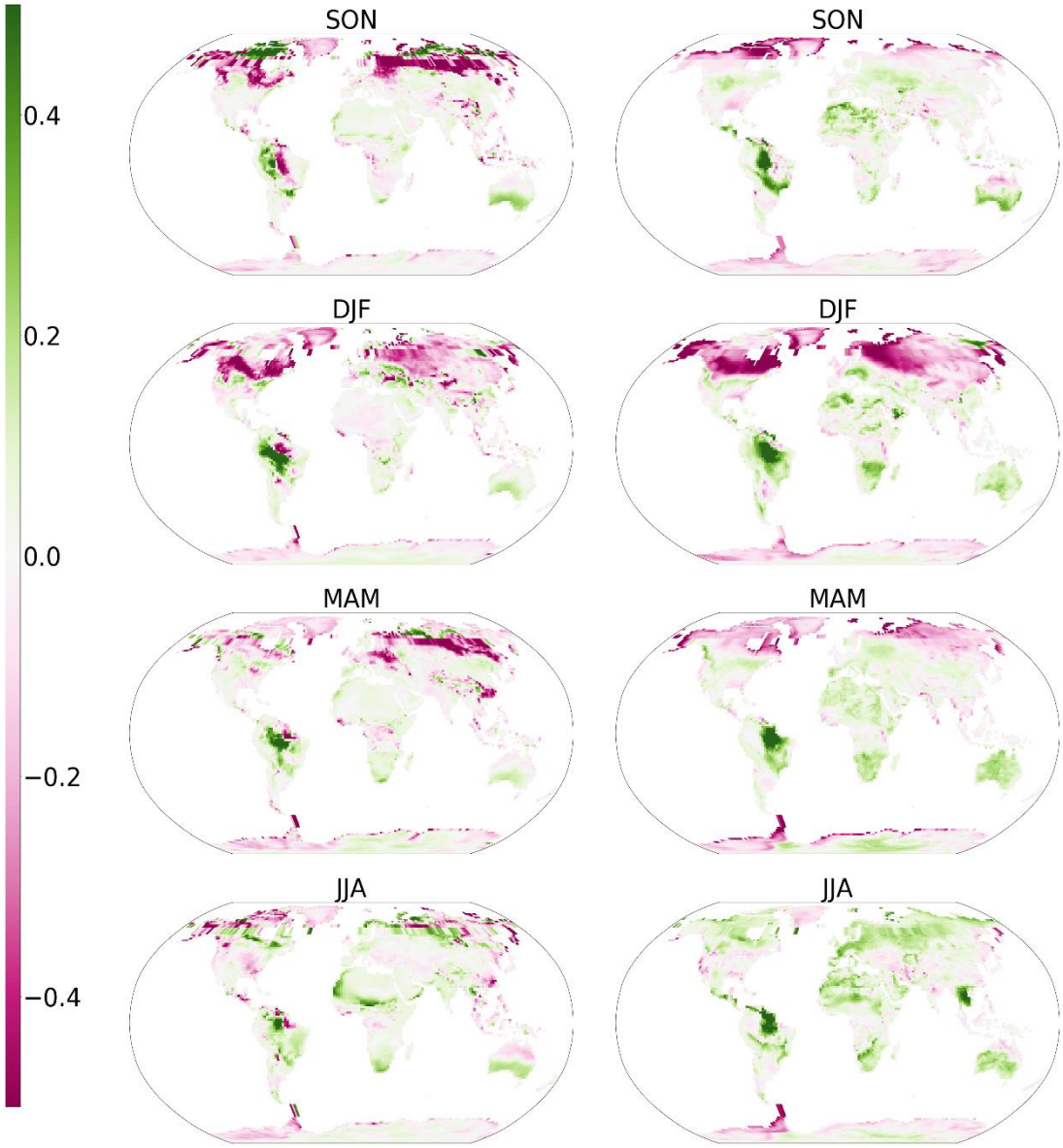


Figure 7.3 – Change in standard deviation of daily maximum temperatures per degree GMST increase on (left) DD and (right) WW averaged seasonally. Colour bar units are ℃/℃.

The largest standard deviation responses (decreases) coincide with the greatest *increases* to mean T_x as a function of increasing GMST over most northern and high latitudes. Such responses indicate rapid warming in these latitudes with less daily variability. It should be noted that decreases in the standard deviation of these subarctic latitudes may be more pronounced due to the loss of sea-ice as a consequence of increased surface temperature, resulting in exposed ocean which in turn has a higher heat capacity than sea-ice, therefore resulting in fewer fluctuations in daily temperature (Bathiany et al., 2018). These regions do however tend to show the greatest inter-model and inter-scenario

differences in regional temperature response (Herger et al., 2015; Tebaldi et al., 2021). Inter-model differences are primarily caused by disagreements regarding the edges of sea ice retreat and inter-scenario differences due to differences in timings regarding persistent ice melt.

Though responses are shown seasonally, hotspots of strong standard deviation response are noted over Cambodia, Thailand and Vietnam for only a few months of the year (much like the mean T_x response). On DD days, decreases in the standard deviation are present in June and July, followed by large increases in August (resulting in the net increase shown in Figure 7.3). On WW days, increases are present in June and July. These strong changes are once again during the monsoon season. Due to the responses falling during the monsoon season, the region tends to have many more wet days than dry days. Therefore, despite decreases to the standard deviation on DD days, it is likely that the standard deviation in T_x will primarily show increases. This, in conjunction with the sizable mean T_x increases, may result in a large increase in the number of extremely hot days and heatwaves due to greater variability. It is, however, important to note the transition probability responses in this region will also impact the future standard deviation in T_x. ACCESS-ESM1.5 diagnosed transition probability responses indicate increases in dry-day frequency from June to August. The resultant changes to T_x standard deviation will therefore be dependent on the total number of wet or dry days.

A range of responses are present over North America. In the high latitudes, decreases in standard deviation are widespread and large in magnitude on DD and WW days during the winter (DJF) and early spring (MAM) months. By summer (JJA), the wet-day responses across Canada almost exclusively show slight increases (with some exceptions in the northeast on DW days in August). Conversely, dry-day responses show greater variations regarding the direction of change, with both increases and decreases present across the high-latitudes (most visible in SON). Transition probability responses in summer across these regions generally show a drying trend, with an increased frequency of dry days and increases in dry-spell length. Based on the variability of dry-day standard deviation responses, it is difficult to deduce the resultant risk of temperature extremes in this region from the response patterns alone. In comparison to Canada and the subarctic grid-cells, responses over the contiguous USA are relatively weak. The strongest changes are present in the north and the east coast, with occasional localised hotspots of strong response and few seasonal variations.

Bathiany et al. (2018) note that in a comparison of climate models, Amazonia showed exceptionally large increases to the annual T_x standard deviation due to changes occurring in *all* seasons. This is loosely true for the ACCESS-ESM1.5 diagnosed responses. On DD days, increases to the standard deviation of T_x are present over western Brazil (the Amazon region) bounded by the Andes Mountain range in Peru and Bolivia in all seasons. However, a region of decreasing standard deviation is present just to the east of the increasing response. Conversely, on WW days, these regions almost exclusively show *increases* to the standard deviation, with some exceptions to the east (that are slightly more present on DW days, not shown here). From December to May, the Amazon regions experience a wet season, meaning changes to the resultant distribution of T_x may be impacted more by wet-day responses than those on dry days. However, it has been established that there will be an increase to the frequency of dry days and length of dry spells in this region, coinciding with some decreasing dry-day standard deviation responses. However, there are several regions where WW and DD responses show agreement in sign (increasing), and it can thus be expected that there will be sizeable increases to the variability of T_x which, alongside large increases to mean T_x, may increase the risk of hot extremes. In regions where there are opposite signs of change on wet and dry days, changes to precipitation occurrence must also be considered.

While the response of mean daily T_x was weaker over Africa in comparison to other regions, concentrated geographical regions experience comparatively strong increases in the standard deviation. On DD days, responses are weak from January through to May, however in JJA, equatorial Africa shows a belt of increasing standard deviation, alongside increases along the north-western coast. This is also present on WD days, though perhaps with wider spatial coverage and more noise. Similar patterns are present on WW days, though increases continue into SON, also moving northward. Increases large in magnitude are present on the north coast from January to March (this is also the case for DW and, to a lesser extent, on WD days). Year-round increases are present in southern Africa on WW days, with increases smaller in magnitude present for all other precipitation statuses. Although mean T_x responses were close to 1°C/°C GMST increase, increases in the standard deviation may result in a greater frequency of extremely hot days. Decreases are also present, with concentrated, localised spatial coverage over Liberia, Cote d'Ivoire and Ghana all year (mainly on DD days).

Mean T_x response was fairly consistent over Australia (with magnitude consistently slightly larger than 1°C/°C), however, here there are seasonal and spatial changes to the standard deviation response. Most of the year, southern Australia shows increases in the standard deviation for each precipitation status resulting in an increased daily variability. Similarly, on WW days, increases cover most of the country from February to April. On DD days, it is primarily only the south that shows increases. The north shows decreases in SON on both DD and WW days, coinciding with the dry season. Although T_x response shows fewer spatial patterns than other regions, on dry days northern Australia shows the strongest warming during these months. This suggests an overall hotter northern Australia, with less variation between daily T_x values.

7.1.3 DAILY MINIMUM MEAN TEMPERATURE RESPONSE

The responses of the mean T_n to GMST increase are presented here. Note a different scale for the changes to the means; regional responses have been capped at 1°C/°C to 6°C/°C. As before, some locations may have responses that exceed 6°C, though for visualisation purposes these have been capped. The responses are generally larger in magnitude than those for mean $T_x$with less geographically concentrated patterns and greater spatial variation. Regions that experience the strongest T_x response to GMST increase also tend to show the strongest T_n response to GMST increase. These include Northern Canada, subarctic grid-cells, and much of South America. Responses over Europe and much of Asia also show seasonally varying strong responses. This is reflected by high correlation coefficients between the responses of mean T_x and T_n for each precipitation status, shown in Table 7.2 (with a mean correlation of 0.71). Strong responses will once again be discussed in turn.

Table 7.2 – Pearson's correlation coefficient between mean T_x and T_n response. Correlations are calculated using the response of T_x and T_n means from the same grid cell.

PRECIPITATION STATUS	MEAN $T_x - T_n$ CORRELATION
DD	0.68
WD	0.64
DW	0.73
WW	0.82
Mean	0.71

While much of central Eurasia showed mean T_x responses between 1°C/°C and 3°C/°C, T_n means show stronger responses of up to at least 6°C with large differences seasonally. This strong mean T_n response is unsurprising; Alexander et al. (2006) found in a global comparison of observed changes to extreme indices, the largest annual change in extreme events corresponding to changes in minimum temperature is present over Eurasia. During spring (MAM) and summer (JJA) months, vast geographical regions show increases in T_n, peaking in magnitude during DJF (with slightly stronger responses on dry days than wet days). Based on the transition probability responses (Chapter 6, Figure 6.9), it is likely that the proportion of WW days in the future will decrease over central Eurasia, wherein overall wet day probability over much of this region is decreasing, particularly during MAM and JJA. This would lead to a higher proportion of DD and WD days, resulting in increases in mean T_n that are more aligned with dry-day temperature responses. During the autumn (SON) and extending into the winter (DJF) months, much of Europe shows increases closer 1°C/°C on dry days, while northeast Russia, east China and India all show stronger increases, with Russia showing the strongest increases in winter with increases of up to 6°C/°C regardless of precipitation state. Increases are present in all months and for each precipitation transition state in subarctic cells, demonstrating polar amplification once more. In central and Mediterranean Europe, the strongest responses are most prominent in autumn and summer months, with larger magnitudes than T_x. This once again implies greater warming during the summer months, however, in a comparison of CMIP5 projections, Lionello and Scarascia (2018) found greater increases in $T_x$than T_n, resulting in an *increased* diurnal temperature range over these regions. Responses diagnosed from CMIP6 ACCESS-ESM1.5

imply the opposite, perhaps highlighting the range of uncertainties associated within climate models themselves.

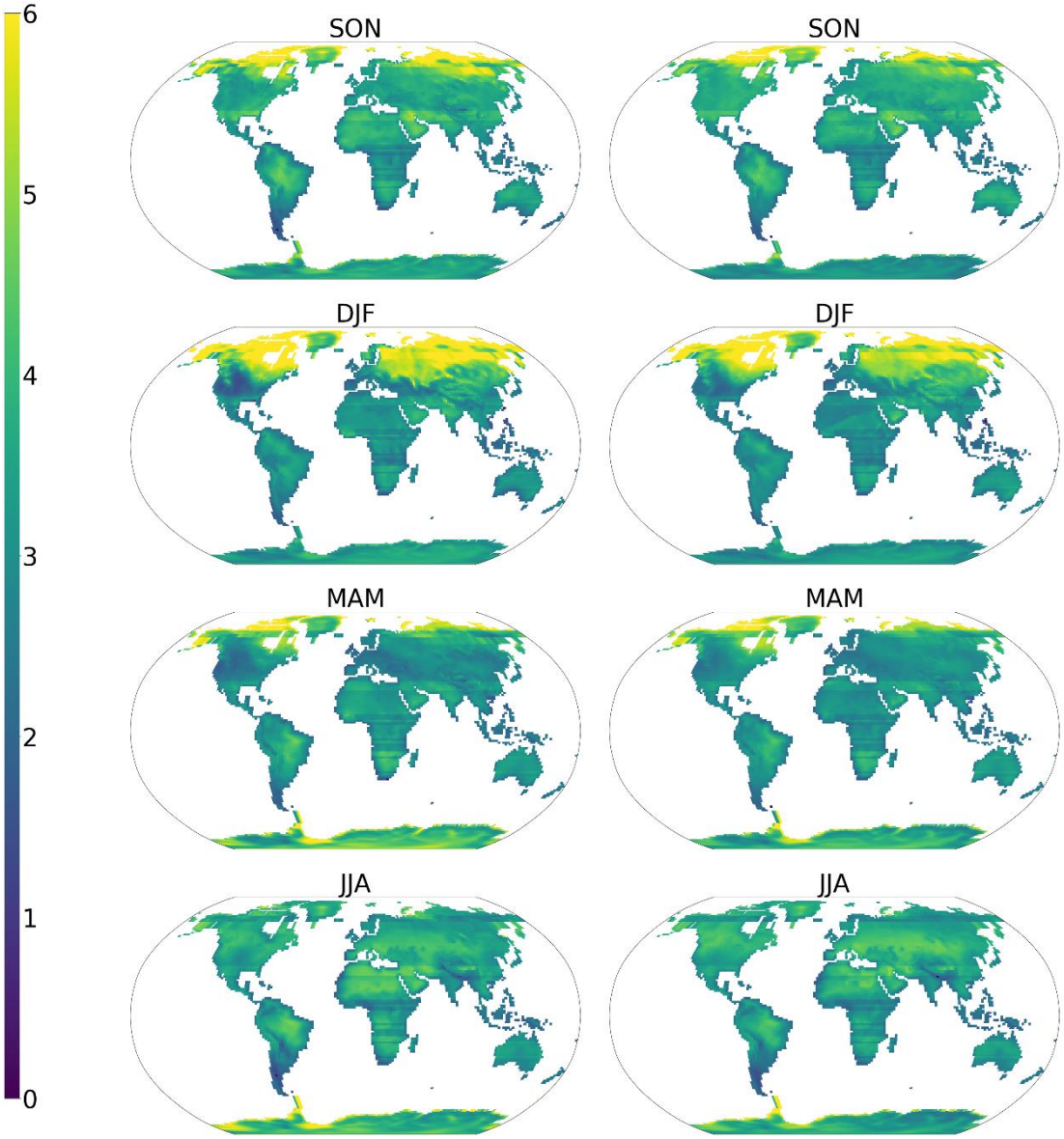


Figure 7.4 – Change in mean minimum daily temperature (°C) per degree GMST increase(°C) seasonally averaged using all bimonthly means within each season on DD days (left) and WD days (right). Colour bar is in units of °C /°C.

North America shows a range of T_n responses. In winter (DJF) the contiguous USA shows responses closer to 1°C/°C for all precipitation statuses. WW days perhaps show the strongest seasonal changes, with widespread strong increases from summer (JJA) through to autumn (SON). Other precipitation statuses also show an intensification of response during these seasons, though with less magnitude than on WW days. Responses are stronger over Canada and once again, in the subarctic grid cells, than the contiguous USA, again

demonstrating polar amplification, where increases reach up to 6°C/°C on DD, WD and DW days in all months except JJA. Responses are slightly weaker on WW days in this region. Northern Canada and subarctic grid-cells show seasonally varying transition probability responses, generally projecting increases to the frequency and length of wet spells during the winter (DJF) and conversely increases to the frequency and length of dry spells during the summer. Despite the strongest summer T_n patterns occurring on WW days, it is increasingly likely that the DD and WD responses will have a greater effect on the resultant temperature distribution, resulting in lower increases.

While the strongest response of mean T_x to GMST increase on DD, WD, DW and WW days is concentrated over small spatial regions with large magnitude over South America, the strongest increases in mean T_n tend to be smaller in magnitude but with greater geographic coverage. Responses are strong on any dry day (DD or WD) and most notable over Brazil from March to October, where there are increases in mean T_n of approximately 3 – 4°C/°C. Responses on WW days are widespread and strong from June to September, covering almost the entirety of the continent (except for southern Argentina and Chile). It is important to note that these strong responses occur during the region's dry season, where there are usually few wet days. This may result in several months having calculated mean temperatures using various steps of the temperature correction sample size process, resulting in patterns that are closer in nature to previous months' responses or the response on days with different precipitation status. The changes shown in Figure 7.4 and Figure 7.5 support the findings of Vincent et al. (2005), noting decreases in the diurnal temperature range from a selection of weather station observations in this region, caused by increases in T_n that exceed increases in T_x.

Across Africa, mean T_n responses show similarities to mean T_x, with widely geographically coherent patterns. DD days once again show the weakest responses, though this time with greater magnitude than the patterns mean T_x. All precipitation statuses show the weakest responses in December, January and February, and once more there are stronger responses in the northern and southern regions (drylands) of the continent than equatorial Africa.

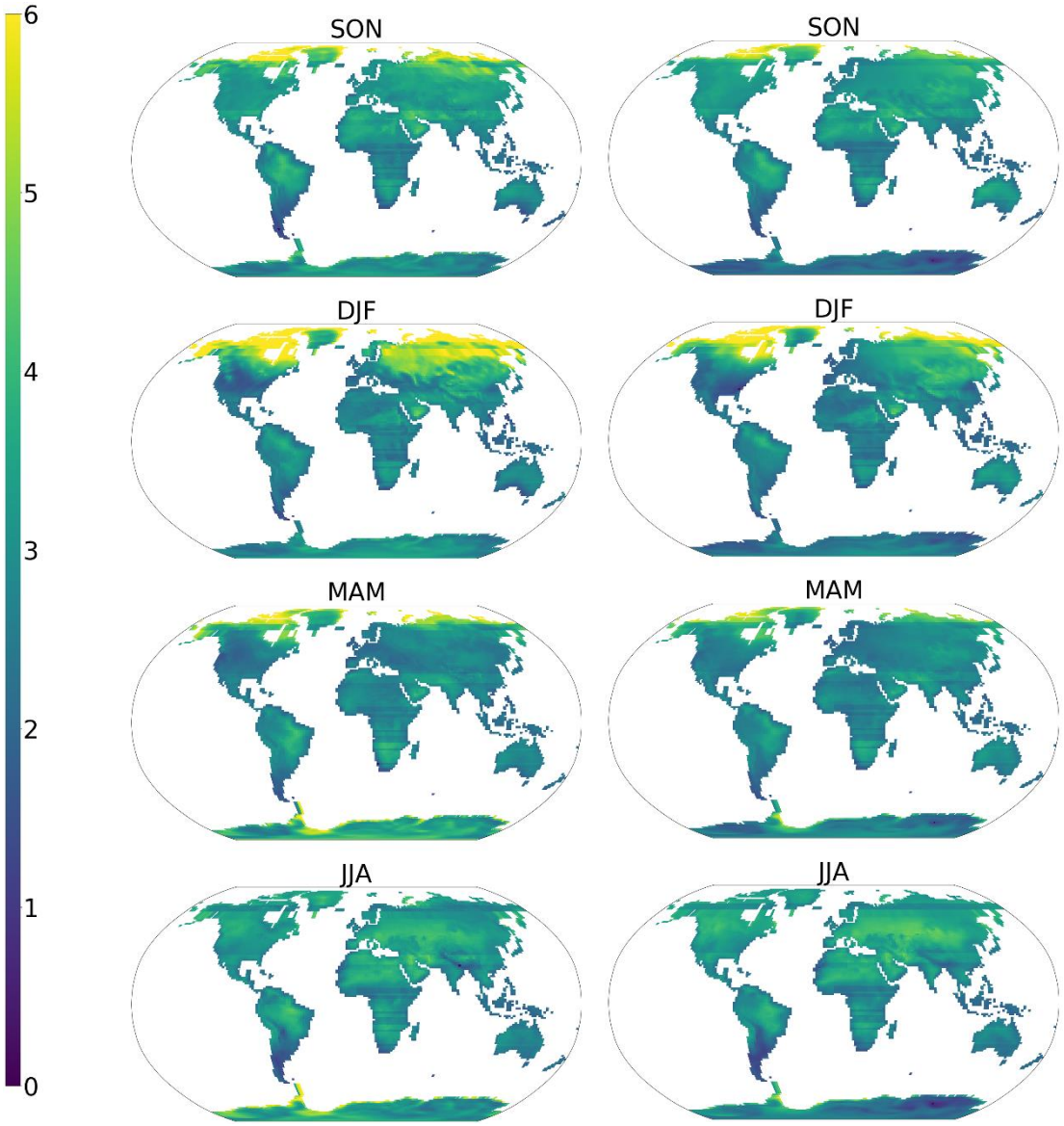


Figure 7.5 – Change in mean minimum daily temperature (°C) per degree GMST increase(°C) seasonally averaged using all bimonthly means within each season on DW days (left) and WW days (right). Colour bar is in units of °C /°C.

7.1.4 DAILY MINIMUM TEMPERATURE STANDARD DEVIATION RESPONSE

Much like the stronger mean T_n responses, T_n daily standard deviations also show stronger responses to GMST increase than T_x (shown in Figure 7.4 and Figure 7.5). Note the colour bar here has been capped at -0.75°C/°C and +0.75 °C/°C to capture the stronger responses (relative to T_x standard deviation). However, much like the similarities between spatial patterns in mean $T_x$and T_n, there are also several similarities in the response of the T_n standard deviation to the standard deviation of T_x. These include strong decreasing

responses over Canada and north Eurasia, alongside increases over South America. DD and WD and WW and DW also show similar responses to each other, so once more only the DD and WW responses will be shown here, averaged seasonally.

Daily T_n standard deviations across Eurasia again show exceptionally strong responses to GMST increase, varying seasonally. On wet days, decreases cover much of north and eastern Eurasia from autumn (SON) through to spring (MAM), peaking in coverage and magnitude in winter (DJF). Central Asia shows slight increases from March to September (though smaller in magnitude, and not shown in Figure 7.6 due averaging seasonally) while the geographical coverage of decreasing responses in the high latitudes diminishes in size. South China particularly shows strong increases in standard deviation from DJF, and Vietnam, Cambodia and Thailand show similarly strong increases in JJA and early SON. The implications of an increasing standard deviation may result in a greater frequency of extremely high (and low) T_n, resulting in, perhaps, extremely hot (or cold) nights. However, due to the increasing mean T_n, it is unlikely that there will be a sizable increase in extremely cold nights, though considering the increases to the mean and standard deviation of T_n in this region, alongside T_x responses, may result in an increased occurrence of concurrent hot days *and* hot nights. The consecutive occurrence of extremely hot days and hot nights is widely used as a measure of extremely hot spells that are considered exceptionally dangerous for human health and have been used to explain heatwave mortality rates (Zhu et al., 2021). In a study of consecutive occurrence of hot days and hot nights (COH), Zhu et al. (2021) found that tropical regions (including Southeast Asia) can expect an increased frequency of COH under RCP8.5, supporting the conclusions made here. Across Europe, DD days show a generally similar pattern to north Eurasia, though with less geographically widespread decreases in the standard deviation from November to April. Instead, a concentrated region of large decrease is noted across central and eastern Europe from November to March. Both wet and dry-day responses are weaker over summer (JJA).

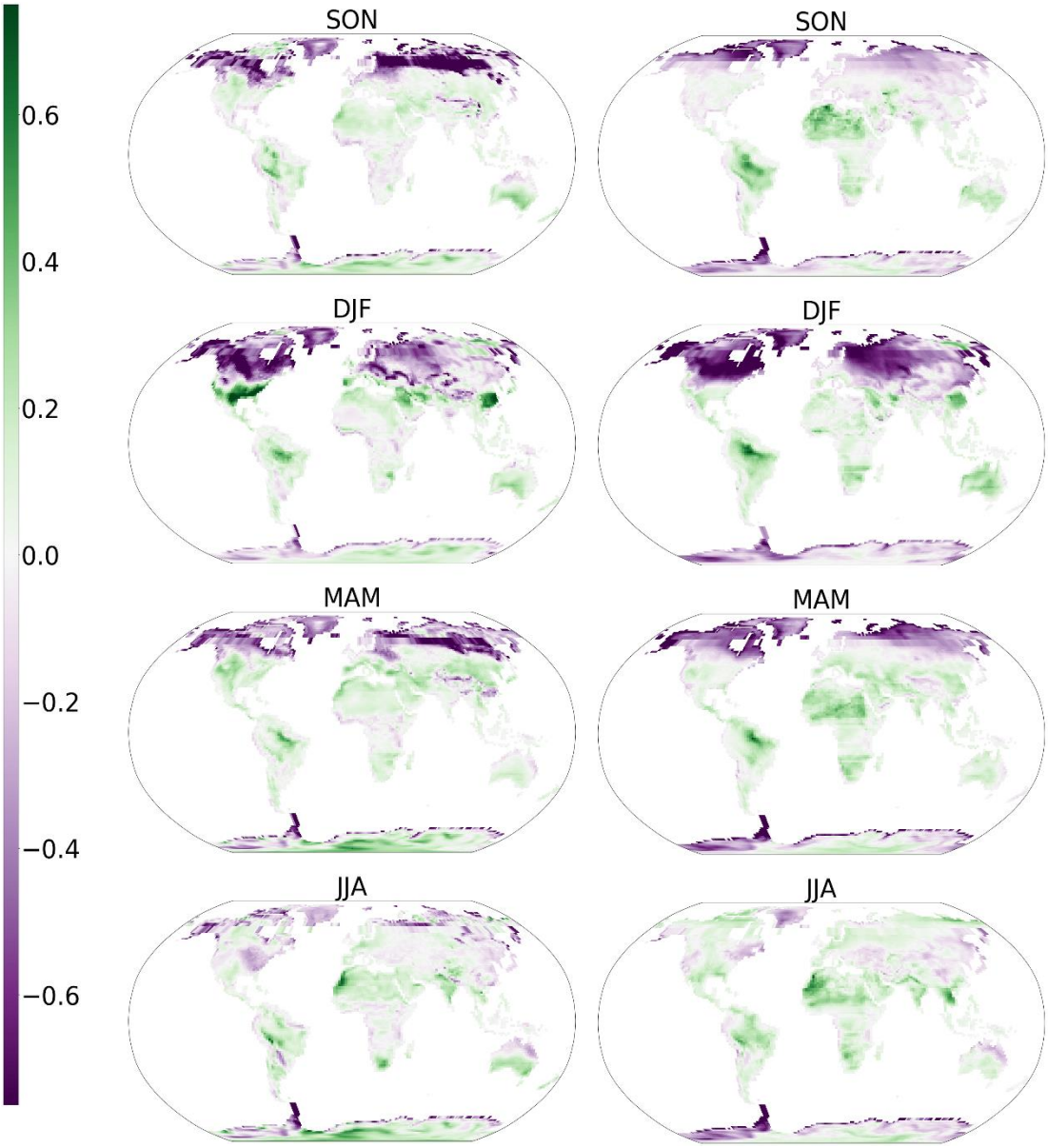


Figure 7.6 – Change in standard deviation of daily minimum temperatures per degree GMST increase on (left) DD and (right) WW averaged seasonally. Colour bar units are °C /°C.

North America shows two distinct patterns. Across the high latitudes, widespread decreases, large in magnitude, are present, with the greatest responses in winter (DJF) and the first few months of spring (MAM) on both wet and dry days. The responses here show decreased standard deviation in high-latitude grid-cells, perhaps once again resulting from loss of sea-ice (Bathiany et al., 2018). Conversely, across the contiguous USA (and in some months, southern Canada), responses indicating an increased standard deviation on dry days are present, primarily from November through to spring. This is also true for wet days, though with less magnitude. These months are typically the coldest in the USA and coincide

with only small changes to mean T_n, therefore there perhaps an increased risk of both extremely hot *and*, unlike previously where mean T_n increases rapidly (e.g., southeast Asia), *cold* nights.

Once again there is a distinct pattern across Africa, wherein northern and southern African drylands show increases in several months, with equatorial and sub-Saharan Africa generally show little change. This is most prominent in the WW response patterns, where all seasons (except DJF) show widespread increases to the standard deviation. This may also indicate an increased frequency of exceptionally hot nights which, in conjunction with the large increases in T_x standard deviation, may contribute to an increased frequency of COH. This agrees with the findings of Zhu et al. (2021), finding large increases in COH frequency over western Africa. An increased frequency of COH was also found over eastern Africa, though T_n responses show smaller magnitudes than in T_x here.

While the standard deviation of T_x responses in Amazonian South America showed both increases *and* decreases, here the T_n standard deviation responses only show increases over these regions, alongside most of Brazil. Increases are present in all months and for all precipitation statuses. Responses are less geographically concentrated than the analogous T_x standard deviation responses and seemingly have greater spatial coverage on wet days. Despite increases to mean T_n, large increases in the standard deviation may cause a greater number of cold nights. Alongside a potentially decreasing diurnal range (through the faster warming of T_n than T_x) the Amazon Basin region may also experience a similar increase in COH as Southeast Asia. Once again, with increasing T_n mean and standard deviation, it is unlikely that there will be a large increase in extremely cold nights however there may be an increased frequency of extremely hot nights. In combination with the large increases to the T_x mean and standard deviation, the region is at risk of greater COH frequency. This is once again supported by Zhu et al. (2021), who note strong increases to COH frequency also occurring in the Amazon basin.

7.2 APPLICATION OF THE RESPONSES TO A STOCHASTIC WEATHER GENERATOR

For consistency with Chapter 6, Section 6.3, observed weather at Santarém, Brazil and Reykjavik, Iceland, will be used to demonstrate the application of the temperature response patterns to the stochastic generation of data using the GCWG. Santarém is a tropical location (like Key West, Florida and Brasília, Brazil, used in previous chapters in

the validation of the GCWG) with exceptionally strong regional temperature responses to GMST increase. Reykjavik has a temperate climate, with temperature responses that are somewhat affected by polar amplification. The GCWG has shown consistently strong performance in temperate regimes, with slightly weaker performance for tropical climates. Sites in each of these regimes have been chosen to demonstrate the applicability of PS at sites where the GCWG typically shows stronger and slightly weaker (though still with good accuracy) performance.

Bimonthly temperature parameters must first be calculated from the observed records. Parameters can then be scaled using the responses presented in Section 7.1 following equation 7.2. ΔT is substituted into equation 7.2 such that it is equal to the difference between three GWLs – 1.5, 2.0 and 4.0°C – and the GMST anomaly that is averaged over the period covered by the observed record. As in Chapter 6, the GMST anomaly is calculated from the HadCRUT5 observed dataset (Morice et al., 2021), relative to the 1850 – 1900 mean. The GMST anomalies averaged over the observed records for Santarém and Reykjavik correspond to 0.78°C and 0.56°C respectively. Based on the similarities between the responses during the first half and second half of the month, and, for standard deviation, the similarities between responses on wet days (WW and DW) and dry (DD and WD) days, it is perhaps unnecessary to diagnose and hence apply responses bimonthly or based on all *four* precipitation statuses (i.e., distinguishing between wet or dry may be adequate) or alternatively, monthly responses instead of bimonthly may be sufficient. However, as this was not known until they had already been diagnosed, bimonthly responses considering all four precipitation statuses will still be used here. Following the scaling of the parameters, 100 years of daily precipitation and temperature data will be generated using the observed and the three sets of perturbed parameters representative of the different GWLs. Changes to the distributions of daily maximum and minimum temperature at the different GWLs will be compared to the data generated using observed parameters (referred to as the reference period).

At Santarém, large increases in T_x and T_n are present at each GWL that far exceed the increases in GMST (Figure 7.7). For example, when the GMST anomaly is increased by 3.2°C from the reference period to a GWL of 4.0°C, the mean maximum temperature in Santarém is projected to increase by 7.5°C. This corresponds to a warming rate of over double the GMST increase. The rate of warming is even faster for minimum temperatures, where the rate is almost four times the rate of GMST increase. The largest increases relative

to the reference period in T_x and T_n are present from March to August. Such large increases in both T_x and T_n coincide with the region's dry season. When considered with the precipitation changes presented in Chapter 6, increases in maximum and minimum temperature during the dry season may exacerbate the frequency of compound hot and dry extremes in the future. Aside from the obvious changes to the mean daily maximum and minimum temperatures, such increases in T_n play an important role in the decreasing of the diurnal temperature range.

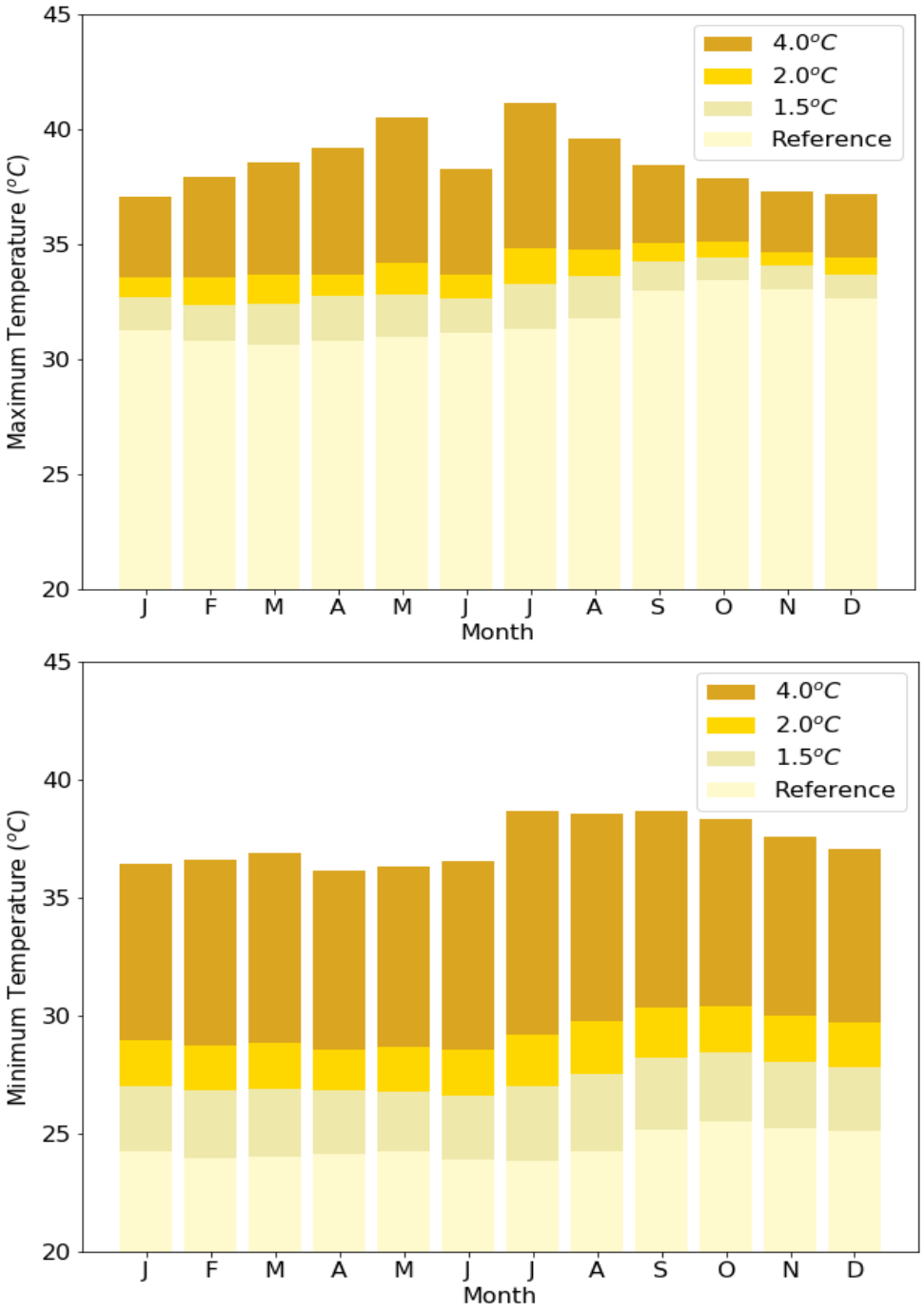


Figure 7.7 - Monthly mean daily maximum (top) and minimum (bottom) temperature at different GWLs at Santarém, Brazil.

It is known that the variability of a distribution is more relevant in the causation of extreme events (Katz and Brown, 1992; Seneviratne et al., 2021) than changes to the mean. The temperature distributions shown in Figure 7.8 are much wider at the higher GWL, showing increased variability at higher GMSTs. While the changes to mean T_x are smaller than the changes to mean T_n at Santarém, the changes to the T_x standard deviation are much larger, resulting in a standard deviation at 4.0°C GWL of 3.2°C in comparison to 1.5°C at GWL 1.5°C. Conversely, the standard deviation of T_n is 1.6°C at GWL 4.0°C compared to 1.3° at 1.5°C. Changes to the standard deviation of T_x in particular will likely result in an increased occurrence of extremely hot events at the higher GWLs. These changes support several studies, indicating substantial increases in hot extremes over South American Amazon regions (Marengo et al., 2009; Vogel et al., 2020; Zhu et al., 2021).

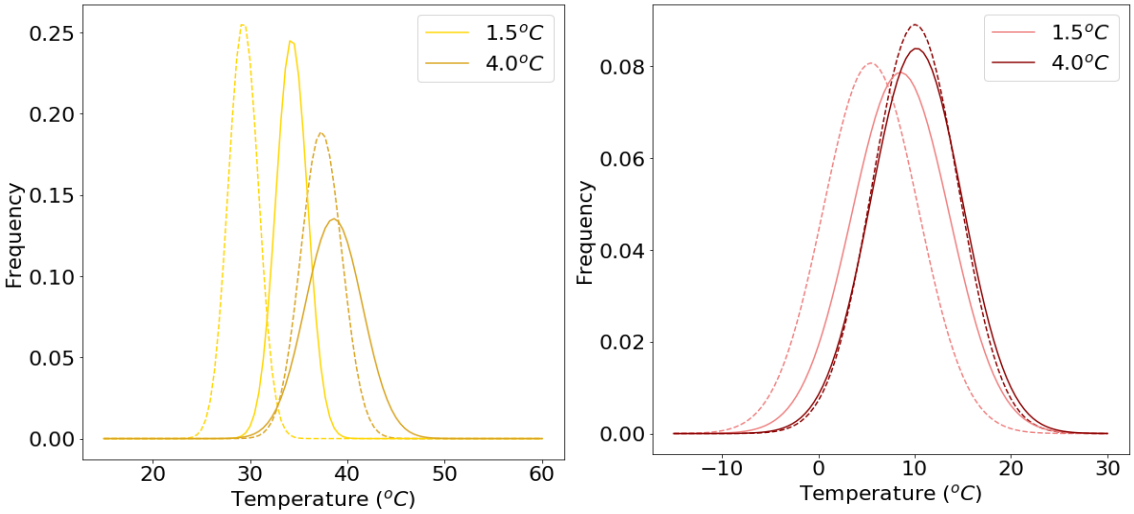


Figure 7.8 – The distribution of daily maximum (solid line) and minimum (dashed line) temperatures at GWLs of 1.5 and 4 ℃ calculated using daily temperatures from all months for Santarém (left) and Reykjavik (right).

It is important to note that the responses with the largest magnitude are usually on DD days. It is also known that the CMIP6 models tend to overestimate the frequency of consecutive dry days (CDD). In Chapter 6, Section 6.2.1, it was shown that precipitation parameters perturbed by the ACCESS-ESM1.5 GCM show large increases in the 99th percentile of dry-spell length distribution, indicating large increases to the number of DD days. The occurrence of DD days therefore may be overestimated in the stochastic simulation of precipitation here, in turn resulting in the DD-scaled temperature means and standard deviations being used more frequently.

At Reykjavik, increases in T_x and $T_n$are smaller than at Santarém. Between the reference period and GWL of 4°C, T_x is projected to increase by 3.3°C; approximately the same as the GMST increase between the two GWLs (of 3.4°C). T_n however, increases at approximately twice the rate of T_x (and also GMST). The largest warming in T_x and T_n tends to occur during the winter months (DJF) and March (early spring, MAM). Unlike Santarém, *decreases* to the standard deviation occur at each GWL (Figure 7.8). By 4.0°C GWL, the standard deviation has decreased by 11% for T_x and 20% for T_n relative to the reference period. Without the increased mean temperature, this would imply a lesser frequency of extreme events. However, with the addition of increased mean T_x and T_n, the frequency of warm extreme events (if defined relative to the reference period) will likely still increase. The reduction in standard deviation will however reduce some potential increases to frequency and/or severity of such events, and when coupled with the warmer mean temperature the frequency and/or severity of cold extremes will be much reduced.

An issue with the GCWG pertaining to the faster warming of T_n arises in Reykjavik (though of course not limited just to Reykjavik). More days are generated by the GCWG where T_n exceeds T_x. Wherever the maximum and minimum distributions overlap, it is possible for the GCWG to generate a larger T_n than T_x. This is demonstrated in Figure 7.8, wherein at GWL of 4.0°C, much of the T_n than T_x distributions at Reykjavik overlap (and, at the tail-ends of the distribution, at Santarém). In Chapter 4, a correction process was introduced to account for instances where this may occur, following methods defined by Dabhi et al. (2021). In instances where T_n exceeds T_x, the mean temperature is calculated ($T_{mean} = \frac{T_x+T_n}{2}$) and a small positive number, δ , is randomly chosen (where $\delta < 0.2$°C). This number is added to and subtracted from T_{mean} to determine T_x and T_n respectively.

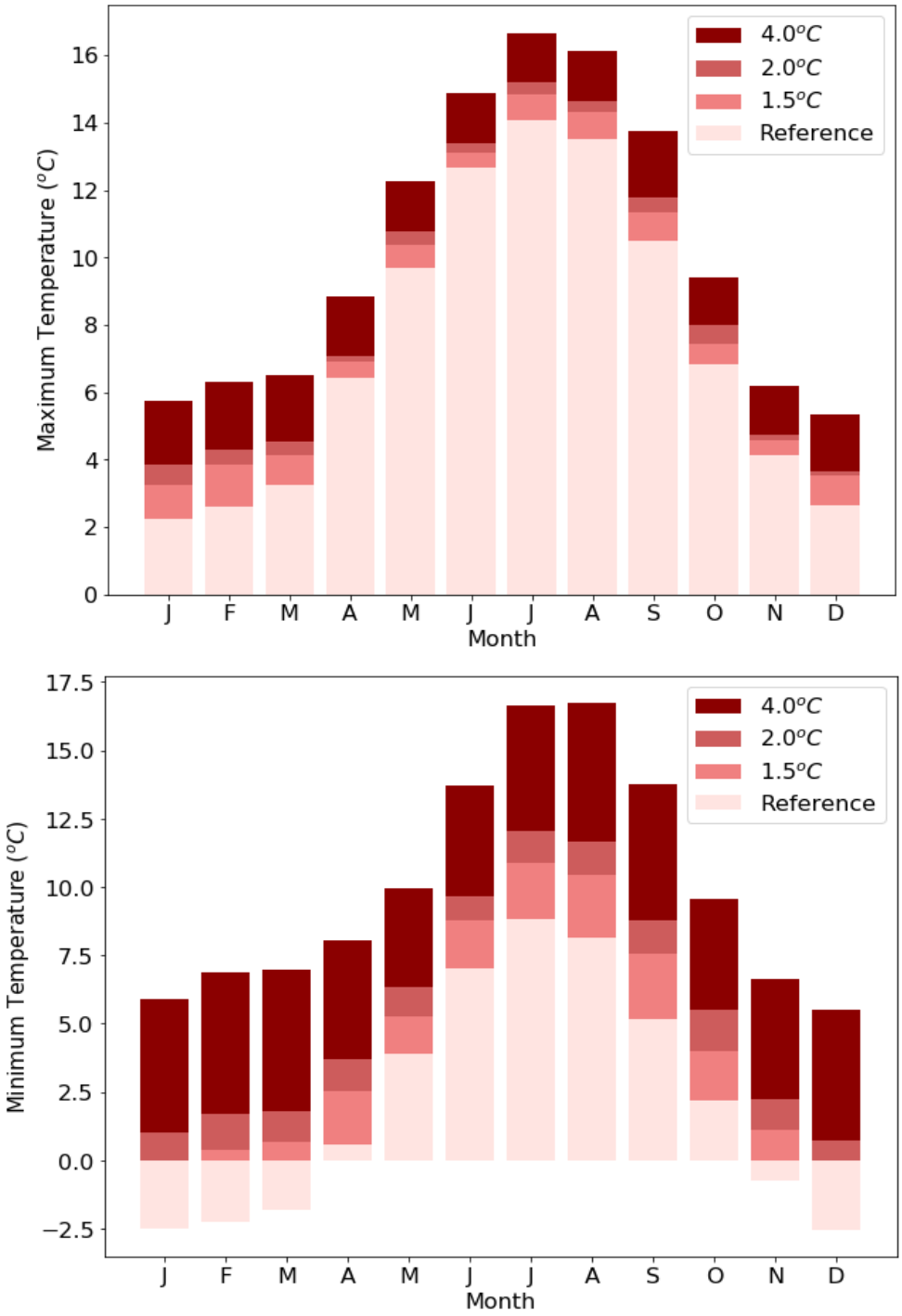


Figure 7.9 – Monthly mean daily maximum (top) and minimum (bottom) temperature at different GWLs at Reykjavik, Iceland.

7.2.1 EXTREME EVENTS AT HIGHER GLOBAL WARMING LEVELS

To better assess some of the associated risks of increasing GMST pertaining to an increased (or decreased) frequency of extreme events at different GWLs, the WS90 and

CS10 indices introduced in Chapter 4 will be used to study the generated time series at Santarém and Reykjavik. To reiterate, WS90 is a measure of warm spells where a warm spell is defined as a period of at least 6 days where the maximum temperature exceeds the 90th percentile. CS90 is a measure of cold spells, where a cold spell is similarly defined as a period of at least 6 days where the minimum temperature falls below the 10th percentile. The annual count includes spells that are 6 days or more (i.e., a cold spell of 10 days is counted as 1 cold spell, the same as a cold spell of 6 days). Percentiles are calculated from the reference period; this is to give an indication of the increased or decreased frequency of events relative to the observed period (note 1.5°C projections have been omitted from Table 7.3 for brevity).

Table 7.3 – Annual WS90 and CS10 count and length in Santarém and Reykjavik during the reference period and at GWLs 2.0 and 4.0°C. Percentiles have been calculated from the reference period.

		WS90			CS10		
LOCATION		**Ref**	**2.0°C**	**4.0°C**	**Ref**	**2.0°C**	**4.0°C**
Santarém	Count	1.6	12.1	5.9	0.1	0.0	0.0
	Length	8.9	21.0	58.5	6.0	0.0	0.0
Reykjavik	Count	1.7	3.4	4.4	1.1	0.0	0.0
	Length	8.3	10.8	17.8	8.0	6.0	0.0

As is expected under a warming climate, the number of warm spells each year increases from the reference period to GWL 4°C at both sites, alongside increases to the mean length of warm spells. An exceptionally large increase of 557% in warm spell length is present in Santarém, wherein at a GWL of 4°C, the total number of individual warm spells per year has *decreased* in comparison with a GWL of 2.0°C. However, this is far from an indicator of a reduction in days that can be categorised as belonging to a warm spell; instead, the number of discrete spells reduces as individual spells join together into fewer but longer spells. For example, an average of 5.9 warm spells per year with average length 58.8 corresponds to a total of 345.2 days that belong to a warm spell. This is a huge increase on the 15.8 mean days belonging to a warm spell per year during the reference period. At Reykjavik, there is also an increase in the annual count of warm spells, though to a much smaller degree than Santarém. The mean number of warm spells increases by a factor of 2.5, with an increase in length of 114%. Both sites see decreases in CS10 resulting in 0 cold

spells by GWL 4.0°C. This is expected; at both sites, mean T_n has increased at a rate that exceeds the GMST increase by at least double. This is accompanied by a small increase in the standard deviation at Santarém and a decrease at Reykjavik. The small increase in standard deviation at Santarém is not enough to produce exceptionally low T_n values considering the large increases to the mean. Although many impacts associated with a warming climate are negative, for some factors, a reduction in the number of cold nights may be beneficial, including reducing animal mortality caused by cold stress (Jones, 2001).

This next method of analysis is analogous to Chapter 4, Section 4.2.3, where the skill of the GCWG at reproducing observed multivariate extremes was assessed. The occurrence of extremes in temperature and precipitation in quick succession (or simultaneously) are considered more damaging to society, ecosystems, agriculture and human health and comfort in comparison to their isolated occurrence (Dash and Maity, 2021; Zscheischler et al., 2018). It is therefore desirable that tools are able to study changes to the frequency or severity of multivariate extremes for successful adaptation, though only recently has more attention been given to such events, including in Chapter 11 of the recent IPCC report (Seneviratne et al., 2021). An advantage of constructing synthetic precipitation and temperature time series using a pattern scaled GCWG is that correlations between the variables are maintained. Alongside the advantages pertaining to scaling variability in both variables, an opportunity to assess the changing risk of consecutive extremes in temperature and precipitation is provided.

The warmest temperatures typically coincide with the dry seasons over most land regions. Such a negative correlation is typically driven by land surface feedbacks that are associated with the impacts of soil moisture limitation on surface temperature (Zscheischler and Seneviratne, 2017). Therefore, concurrent hot and wet extremes are unusual in many regions. Still, there are several studies which have assessed the changing risk of such extremes, including a study into the changing frequency of joint hot and wet extremes across India (Dash and Maity, 2021) and a study into the global co-occurrence of extreme hot and wet events over short time periods at different GWLs (Vogel et al., 2020). Maximum temperature on days where the precipitation exceeds the 95th percentile calculated from the reference period (referred to as “very rainy” days) at each site and GWL is determined.

At both sites, the maximum temperature on very rainy days increases at each GWL (Figure 7.10). This is of course an expected outcome considering the warming at each site, presented in Figures 7.7 and 7.9. At each GWL, the maximum temperature on very rainy days shows negative skew. At Reykjavik, there are few changes to the distribution of maximum temperature on very rainy days at each GWL, though the skew becomes more apparent at larger GWLs. This is reflected by similar increases in temperature on very rainy days to the temperature on all days from the reference period to GWL 4.0°C (of 2.9°C and 2.8°C, respectively). The number of very rainy days has also increased, though this is expected based on the increasing precipitation in several months (Chapter 6, Section 6.3).

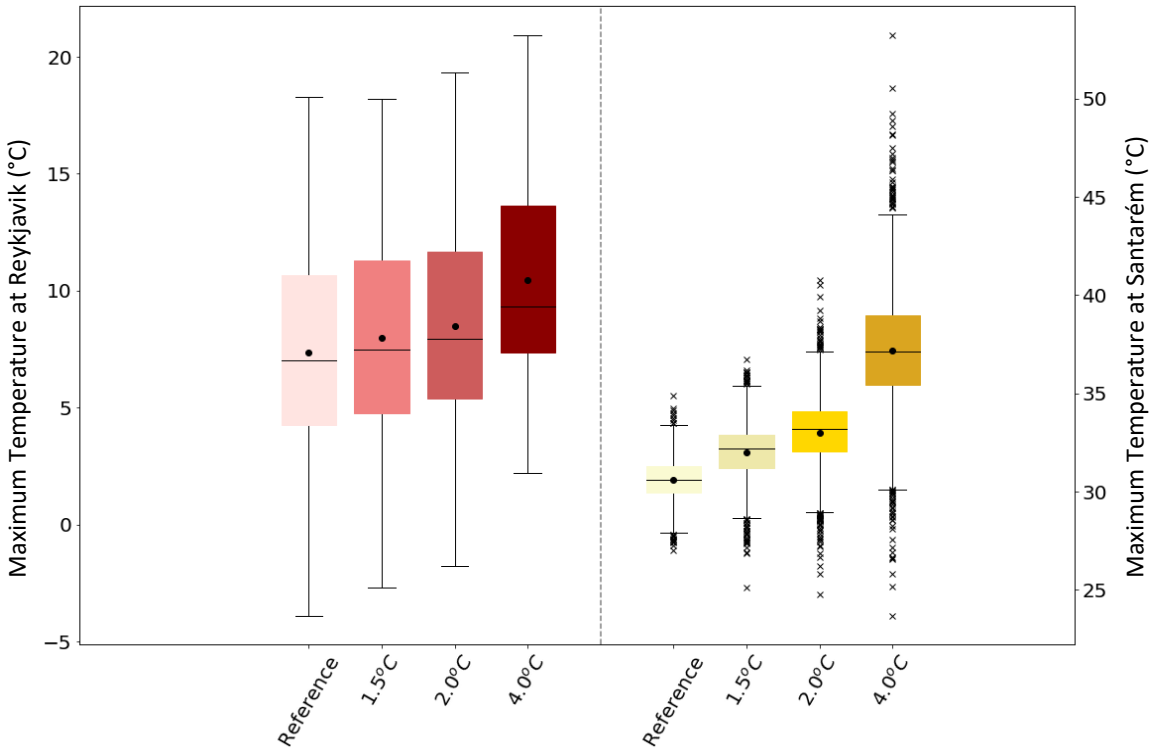


Figure 7.10 - Daily maximum temperatures on days with precipitation exceeding the 95th percentile from the reference period

In Santarém, the number of days where the precipitation exceeds the 95th percentile (calculated from the reference period) decreases at each GWL (corresponding to a decrease of approximately 50% from the reference period to GWL of 4.0°C). This is also unsurprising considering the exceptionally large increases to the dry-spell length shown in Chapter 6, Section 6.3. On days where the precipitation exceeds the 95th percentile, there has been a mean increase of 6.8°C in the maximum temperature from the reference period to GWL of 4.0°C. This is similar to (though slightly smaller than) than mean increases to maximum temperature irrespective of precipitation occurrence (7.5°C). Unlike Reykjavik, the temperature on very wet days does not show skew. Instead, the interquartile range is small

(increasing at higher GWLs) though several outliers are present at each GWL. Despite the number of very rainy days decreasing, the number of outliers shown in Figure 7.10 has increased by 4.0°C GWL. The upper outliers therefore indicate a higher likelihood of extremely hot and very rainy days, which aligns with the large increases to both the mean and standard deviation of T_x at Santarém.

Finally, the occurrence of extremely hot *and* extremely dry conditions at different GWLs will be assessed analogously to Chapter 4, Section 4.2.3. Concurrent hot and extreme events are associated with several hazards, including increased wildfire risk (Zscheischler et al., 2018), exacerbation of summer heatwaves, negative impacts on crop yield, disease spread, and human mortality (Luber and McGeehin, 2008). It is therefore imperative that there are robust methods for assessing changes to the frequency and/or severity of such events. The co-occurrence of a hot and dry extreme is defined in the same way as in Chapter 4, Section 4.2.3, loosely based on the methodology of Vogel et al. (2020). Here, an extremely hot day is one in which the maximum temperature exceeds the 99th percentile calculated from the reference period. Total precipitation is summed over a rolling 90-day period, with the date of the period defined as the last calendar day of the 90-day window. A dry extreme is defined as a day where the 90-day precipitation falls below the 1st percentile calculated from the reference period. The mean number of times where an extremely hot day *and* extremely dry day fall within the same month are recorded per every 10-years in Table 7.4.

Table 7.4 – The number of occurrences of extreme hot and extreme dry days per month per 10 years.

	HOT AND DRY EVENTS			
GWL	**Ref**	**1.5°C**	**2.0°C**	**4.0°C**
Santarém	0.4	2.5	2.6	9.9
Reykjavik	0.8	1.1	1.0	0.9

Co-occurrences of extremely hot and extremely dry days only fall within the same month on average 0.4 times during the reference period at Santarém. This is projected to increase by a factor of almost 25 by GWL 4.0°C, resulting in almost 10 months per 10-years (i.e., 1 month per year) that experience at least one extremely hot and extremely dry day. This reflects the findings of Vogel et al. (2020) wherein Amazonian South America showed increases in the co-occurrence of extreme hot and dry events during the same year, month,

and week that exceeded most other land-surface areas. This is certainly reassuring of the pattern scaled GCWG's capabilities at assessing the changing frequencies of such events at higher GWLs. Such increases in the co-occurrence of extreme hot and dry days per month per 10-years also suggests increases to the frequency of weekly clustered extremes. To elaborate, during the reference period, *no* weeks contain an extremely hot *and* extremely dry day per 10-year period. By a GWL of 4.0°C however, an average of 0.4 weeks per 10-year period experiences at least one extremely hot and extremely dry day.

During the reference period, Reykjavik experiences just over double the number of consecutive hot and dry extremes than Santarém (recall that "hot" and "dry" are defined relative to local conditions, i.e., the thresholds are different between the sites). The frequency of consecutive hot and dry extremes however remains relatively unchanged at the different GWLs. This is likely due to increasing precipitation, thus decreasing the frequency of isolated extreme dry event, alongside increases to the maximum temperature, thus increasing the frequency of isolated extreme hot events. Larger increases to these events at Santarém than Reykjavik are also likely for two reasons. The first is that regional mean maximum daily temperature responses at Santarém exceed those at Reykjavik for all transition states, therefore greater warming is expected. Furthermore, the standard deviation in T_x increases much more at Santarém than Reykjavik (where the standard deviation decreases). This will ultimately result in a greater spread of T_x values, with greater frequency of values that are far from mean T_x.

7.3 CHAPTER 7 SUMMARY

The GCWG presented in Chapters 3 – 5 can simulate daily time series for maximum and minimum temperature, conditioned on the generated precipitation status of the day, using monthly regression equations and bimonthly means and standard deviations. GCM grid cell-scale (sometimes referred to as regional) responses of the conditional bimonthly means and standard deviations in maximum and minimum temperature to changing GMST over land-surface areas have been diagnosed from four ensemble members of the ACCESS-ESM1.5 GCM using linear least square regression. Monthly linear regression coefficients (relating precipitation and temperature) have been assumed to remain constant as a function of GMST increase.

The responses presented in Figures 7.1 to 7.6 show several similarities with trends present in historical temperature observations and studies on future warming patterns based

on CMIP5 and CMIP6 projections. The most notable features are the stronger warming signals over the high latitudes and subarctic regions, with temperatures increasing more in the northern hemisphere than the southern (Tebaldi and Arblaster, 2014). Hotspots in warming are also present over South America, Southeast Asia and southern Europe. Mean minimum temperature responses are usually less spatially concentrated than maximum temperatures, with wider regions showing large increases. The rate of warming is greater with regards to daily minimum temperature than maximum temperature, indicating a faster warming of night-time temperatures than daytimes. Responses are typically similar on wet days (WW and DW) and dry days (DD and WD), though with some differences in magnitude and spatial coverage. The largest differences shown in response are typically between WW and DD days. Changes to the distribution of temperature are therefore also dependent on changes to the precipitation occurrence.

Daily standard deviations tend to show the strongest response to GMST change in regions that also show strong mean responses, typically including the high-latitudes, North America, and much of Europe and Eurasia. Strong standard deviation signals in the high latitudes are unsurprising. It is known that due to sea-ice loss, the variability in daily temperatures will reduce due to the increased heat capacity of the exposed ocean. However, it is also these regions that show the greatest inter-model differences due to the disagreements in the rate of sea-ice loss. Strong responses are also present over the African drylands, Australia and Southeast Asia. There is generally a spatial separation in the response patterns between the decreases over the high latitudes and subarctic grid cells, and the increases that are present over most other locations. This pattern is particularly pronounced in the T_n standard deviation responses. There are of course some exceptions; on dry days, regions in South America, southwest Africa, and Southeast Asia all show large decreases for at least three months of the year in the T_x standard deviation. On wet days from June to September, high latitudes show weaker T_n standard deviation responses, positive in magnitude.

The scaling of observed parameters and hence the construction of future time series at different GWLs have been demonstrated using weather station data at Santarém, Brazil, and Reykjavik, Iceland. Santarém shows warming in both T_x and T_n at a much faster rate than Reykjavik, with both sites showing sizable increases in the rate of daily minimum temperature warming than maximum temperature. While it is of course important to have accurate depictions of mean, regional-scale climate change, the diagnosed changes in the

daily standard deviations of T_x and T_n are invaluable in the construction of local-scale synthetic time series that are suitable are inputs for hydrological, agricultural and impact assessment models due to the incorporation of changes to the variability. This is a further advantage of the scaling of *daily* parameters; oftentimes monthly, seasonal, or annual statistics are not of the correct temporal scale for such assessments. The ability to produce long local-scale time series also provides a method for studying future changes to the frequency and/or severity of univariate and multivariate extremes. Such analysis of extremes has been demonstrated here; changes to warm and cold spell occurrence and length, the maximum temperature on very rainy days, and the frequency of extremely hot and dry day co-occurrence at the three GWLs at Santarém and Reykjavik.

Chapters 6 and 7 have demonstrated the diagnosis of the GCWG's primary and secondary variables' response to GMST increase and the subsequent perturbation of observed weather station parameters to produce synthetic time series at different GWLs. There are several strengths in this method of emulating future climate scenarios. Firstly, stochastically generated time series have fine temporal and spatial scale. The GCWG has been developed using weather station data, which of course has a much higher resolution than direct GCM-outputs. This makes the GCWG-simulated data more appropriate for local-scale impact assessments. On the same note, station-scale data may contain weather patterns that are not represented by GCM-scale outputs (for example caused by steep topographic changes or local forcings that are not resolved by the GCM). These may be better captured at future GWLs using the scaled GCWG opposed to direct GCM output. Secondly, pattern scaling has typically been used for constructing scenarios with larger temporal scales, usually monthly, seasonal, or annual statistics. Daily time series are more appropriate for impact assessments or as inputs to hydrological and agricultural models. Furthermore, the GCWG can generate very long time series at any (non-transient) GWL with computational efficiency. This provides temporally consistent, long records that are suitable for the robust assessment to the changes at a specific GWL. On a similar note, the PS technique provides a method of constructing several realisations of the climate for a range of GWLs, including those not presently simulated by GCMs, therefore providing a method to better represent and understand future climate uncertainty. Finally, another strength is in the better inclusion of variability that traditional pattern scaling methods. Daily wet-day precipitation shape *and* scale parameters, alongside temperature standard deviations (conditional on scaled precipitation status) all contribute to changes in precipitation and temperature variability.

8 THESIS SUMMARY AND CONCLUDING REMARKS

The development of a stochastic weather generator (SWG) that has been calibrated using a selection of weather stations representative of the five overarching Köppen climate classifications (tropical, dry, temperate, continental and polar (Ackerman, 1941)) has been presented in Chapters 3 through 5 (named the **G**lobally **C**alibrated stochastic **W**eather **G**enerator, GCWG). The GCWG produces synthetic daily time series for a suite of weather variables, including precipitation (a primary variable), maximum and minimum temperature (secondary variables), mean wind speed, solar radiation, and relative humidity (tertiary variables). The pattern scaling (PS) technique was then extended to the estimation of the key parameters of the GCWG under a warmer climate. This section summarises the methods and key findings presented in this thesis as part of the development of the GCWG and the subsequent application of PS to construct local-scale daily time series at different global warming levels (GWLs).

The models used to generate each of the variables and the parameters required are summarised in Table 8.1. This thesis has focussed on the stochastic generation and subsequent application of PS to precipitation and temperature only. Though the tertiary variables will likely be affected by climate change, there is more literature documenting the effects of climate change on the primary and secondary variables, including detrimental impacts on hydrological and agricultural planning, human health, comfort and mortality, and changes to extreme event severity and occurrence (Cook et al., 2018; Kilpeläinen et al., 2010; Seneviratne et al., 2021; Trenberth, 2011). Furthermore, many studies have assessed the applicability of PS to these variables and critiqued the assumptions underlying the technique (e.g., Tebaldi and Arblaster (2014), Osborn et al. (2016, 2018), Mitchell (2003), Cabré et al. (2010)). However, even for precipitation and temperature, PS has almost universally been applied only to changes in their monthly means. Here, the literature is advanced through the application of the PS technique to the parameters of a SWG, i.e., changes in parameters that represent daily timescale characteristics including variability, skewness, and sequencing of wet and dry days (which have knock on effects to the other variables via their dependence on precipitation state). The structure and parameters of the GCWG have therefore intentionally been kept simple to allow for pattern scaling.

A selection of weather stations representative of the five climate classifications have been used for validation of the GCWG. In the validation of GCWG's ability to reproduce observed time series, separate training and test (referred to as "observed") datasets have been used throughout. This is an improvement over some prior studies that used the same training and test datasets to evaluate a model. This has ensured the robustness of the statistical testing and subsequent analysis. It is important to note that using different test and training datasets will likely increase the number of generated datasets showing statistically significant differences from the observed than other studies where test and training datasets are the same. Furthermore, a wide range of metrics (including those that are not widely used and are challenging in nature) in combination have been used to validate the GCWG. A range of properties have therefore been evaluated, including the reproduction of observed intervariable relationships, extremes, and compound events.

Chapter 3 assessed the performance of four orders of Markov chain at accurately reproducing precipitation from a network of global weather station observations. While similar analysis has been presented elsewhere, studies have typically been constrained to smaller geographic regions such as the USA (Schoof and Pryor, 2008), Nigeria (Jimoh and Webster, 1996) or Sweden (Lennartsson et al., 2008), and typically only using one or two performance metrics. This thesis has expanded upon previous literature by using Global Historical Climatology Network Daily (GHCN-D) observations, aggregated into 5° latitude by 5° longitude cells with at least 100 cells present in each of the five Köppen climate regimes. Furthermore, six metrics were used to assess model-order performance across the five climatic regimes, including Bayesian Information Criteria, the ability to reproduce observed distributions of wet- and dry-spell lengths, the ability to reproduce observed interannual variability of precipitation occurrence and amount, and the ability to reproduce observed mean annual maximum 5-day precipitation. This thesis finds that performance is generally more sensitive to assessment metric rather than climate regime, with exceptions in tropical regions where third-order models typically perform best regardless of metric. Much of this work has been published in the *International Journal of Climatology* by Wilson Kemsley et al. (2021). Despite the good performance of third-order models, first orders have been used to generate precipitation occurrence throughout the remainder of this thesis to reduce the number of scalable parameters. Nevertheless, future work could apply these techniques to higher order models, provided that GCM ensembles are large enough to reliably estimate changes in the higher order transition probabilities.

Table 8.1 – Primary, secondary, and tertiary variables and the models used to simulate them. P is precipitation, T_x and T_n are maximum and minimum temperatures respectively, W is mean daily wind speed, R solar radiation and H relative humidity.

		MODEL	PARAMETERS
Primary	**Occurrence** **P**	Markov-chain model; first order for wet-spell length, third order for tropical regimes, dry-spell length. Probabilities calculated monthly.	P_{00} and P_{11} (first order) P_{0000}, P_{0010}, P_{0110}, P_{0100}, P_{1000}, P_{1010}, P_{1110}, P_{1110} (third order)
	Amount	"Wet" day defined as >0.1mm. Values taken independently from monthly fitted gamma distributions.	Wet-day shape (α) and scale (β) gamma
Secondary	**T_x and T_n**	Multiple linear regression models to generate residual series with coefficients fitted monthly with predictors determined by precipitation status. Residual series converted using bimonthly means and standard deviations conditioned on precipitation status.	Mean and standard deviation on WW, WD, DW and DD days Regression coefficients, e.g., a_1, b_1 and ε_1
Tertiary	**W**	Generalised linear model using log-link function and a gamma distribution, fitted monthly. Maximum and minimum temperature and lag-1 autocorrelation used as predictors.	Monthly shape, ν, scale, λ, and generalised linear model coefficients, β_n
	R	Following a transformation to a residual series of clearness index, multiple linear regression model fitted monthly using precipitation, maximum and minimum temperature, and lag-1 autocorrelation as predictors. Residual series calculated using monthly means and standard deviations.	Station latitude, monthly mean, standard deviation, and regression coefficients
	H	Converted to residual series using monthly means and standard deviations. Multiple linear regression model, fitted monthly using precipitation, maximum and minimum temperature, and lag-1 autocorrelation as predictors.	Monthly mean, standard deviation, and regression coefficients

At the selection of weather stations used in this thesis, the GCWG has reproduced the primary and secondary variables with excellent practical accuracy. Annual cycles are reproduced well, and the GCWG has synthesised credible realisations of the climate at each site for both the primary and secondary variables. Errors between observed and generated time series of maximum and minimum temperature are small. The ability of the GCWG to

reproduce observed extremes in both precipitation and temperature has also been assessed. The frequency of extreme event occurrence is typically underestimated. For example, return periods of extreme precipitation events are typically overestimated by a factor of between 1.4 and 4.4. This results in an underestimation of, for example, 50-year rainfall events. While this is a drawback of the model, it is a widely known issue with SWGs that is not unique to the GCWG (Dubrovský, 1997; Kilsby et al., 2007; Soltani and Hoogenboom, 2003a). To utilise the GCWG in the study of the changing risks to the frequency of extreme events, a comparison relative to a generated reference period (as opposed to observations) may be used instead. An example of this is provided in Chapter 7. Reproducing multivariate extreme events shows variable performance. The maximum and minimum temperature on very wet days is reproduced well at all sites. The number of extremely hot *and* extremely dry days per month is slightly underestimated by the model. Univariate extremes in maximum temperature have been well reproduced at several sites, and, in all sites, the mean number of occurrences of extremely hot days per month has been captured well by the GCWG.

Tertiary variables use the primary and/or secondary variables as predictors in each of their respective models. This may provide the potential to assess changes in tertiary variables as a function of the primary and secondary variables (see Section 8.1 for future work). Much like the primary and secondary variables, the tertiary variables have been well reproduced by the GCWG at all sites with inter-tertiary variable accurately simulated. The GCWG shows similar performance at reproducing solar radiation and relative humidity. Accuracy is weaker at sites where the calculation of bimonthly means and standard deviations in temperature may have been adjusted using the temperature sample size correction process, presented in Chapter 4. These variables are generally strongly correlated to temperature, therefore inaccuracies in the generation of temperature may carry forward to the generation of solar radiation and relative humidity. Observed variance in wind speed has been more accurately captured than its mean. However, annual cycles in each variable have been captured with a great degree of accuracy, with annual maxima, minima and ranges also reproduced accurately. This has been exemplified by good reproduction of reference evapotranspiration at sites in tropical and temperate regimes (where the GCWG performs worst and best respectively).

With the good performance of the GCWG demonstrated, PS has hence been used to diagnose a selection of primary and secondary variable parameters' responses to increasing GMST. Pattern scaling has typically been used to construct scenarios using scaled changes in means over larger temporal scales (e.g., annual, seasonal, monthly). Here, parameters

representing daily climate have been scaled, with a greater emphasis on capturing changes to climate variability than many previous studies.

This thesis has presented the spatial responses of the transition probabilities and wet-day gamma parameters to increasing GMST, diagnosed from two GCMs (IPSL-CM6A-LR and ACCESS-ESM1.5). The purpose was to illustrate how the approach could be applied in practice and to find out whether the response patterns of the parameters looked physically plausible in relation to prior work, rather than (e.g.) being dominated by local-scale noise. Responses show several similarities between the two GCMs, though an insight into inter-model uncertainties (one of the largest uncertainties associated with modelling the future climate) has been demonstrated where the spatial patterns differ. Regions where the responses are similar include South America, Canada, and Northern Eurasia. Africa, Australia and Southeast Asia show the largest differences in the response patterns between the two GCMs. However, differences are typically regarding magnitude of the signal and not the direction of the response. The wet-day shape parameters show larger differences between the two modelling centres, resulting in different changes to the variability of the resultant wet-day gamma distribution. The response of bimonthly means and standard deviations, conditioned on precipitation status (WW, WD, DW or WD), in maximum and minimum temperature to increasing GMST have also been diagnosed using the ACCESS-ESM1.5 GCM. Regions that show the strongest precipitation parameter response typically showed strong temperature parameter response. As expected, minimum temperatures showed warming at a greater rate than maxima, with large changes in variability across several regions in the Northern Hemisphere.

Following the diagnosis of the responses, parameters calculated directly from observations can be perturbed under a specified GMST increase (e.g., a transiently increasing scenario of GMST or, as here, a specific GWL). The perturbed parameter values can hence be used in the GCWG to generate time series at a range of GWLs. There are several strengths to this technique that will be briefly summarised:

- Scenarios (e.g., specific GWLs or scenarios intermediate to the commonly used Shared Socioeconomic Pathways SSPs) not presently simulated by GCMs can be emulated
- Several realisations of a climate scenario can be produced with computational efficiency
- Greater incorporation of changes in climate variability than other PS techniques

- Initial parameter values obtained by fitting to observations include local-scale climatology and as such reduce GCM errors
- Long time series can be produced for the robust assessment of extreme events and subsequent changes as a function of GMST increase.

PS has been widely used to emulate scenarios for time periods and scenarios that are not presently simulated by GCMs. The construction of precipitation and temperature scenarios at GWLs of 1.5, 2.0 and 4.0°C has been demonstrated in Chapters 6 and 7, though theoretically any GWL may be used. The use of a SWG in the construction of these scenarios also allows for the generation of several different realisations at a single GWL, which provides an opportunity to address climate uncertainty from unforced variability (noting, however, the smaller interannual variability often produced by this type of weather generator and evaluated in Chapter 3). While not demonstrated at the different GWLs in Chapters 6 and 7, the use of several realisations in the validation of the GCWG was shown in Chapters 4 and 5. This provides a method to study a range of uncertainties in the future climate at a single GWL.

Several PS studies have typically only assessed changes to mean climate, usually diagnosing the spatial response of monthly, seasonal or annual statistics in temperature and/or precipitation to increasing GMST (Doblas-Reyes et al., 2021). Although there has more recently been discussion into better incorporation of climate variability, this has been limited to the shape parameter of *monthly* total precipitation gamma distributions, with daily temporal scales not addressed (Osborn et al., 2016). Changes in climate variability are particularly important in studying the changes to the frequency and severity of extreme events. This thesis has scaled multiple parameters that contribute to changes in variability on a daily temporal scale. For precipitation, the response of both the wet-day gamma parameters *and* transition probabilities to increasing GMST contribute to changes in the variability. Shape and scale parameters determine the mean, variance, and skew of wet-day precipitation amounts. This directly impacts the variability of daily precipitation. Changes to precipitation occurrence also impact the variability. For example, if the total monthly precipitation increases, but the number of wet days per month decreases, there will be an increase to the variability for that month. Standard deviations in daily temperature that are conditioned on the precipitation statuses (which, at several locations are also changing) have also been scaled with GMST.

GCM scale data is typically referred to as "area-averaged" (e.g., representing the average over a GCM grid cell or larger), resulting in errors if used to represent local-scale present-day climate. The errors arise due to parameterisation of several sub-GCM grid-scale processes and steep topographic effects, alongside unresolved local forcings such as black-carbon or sulphate aerosols (Murphy, 1999; Tebaldi and Arblaster, 2014). This may result in local, weather-station scale data being misrepresented by the area-averaged climate. Applying GCM-derived responses to parameters calculated directly from observations will therefore reduce inaccuracies from GCMs, and therefore better represent the local-scale climate. An example of this has been provided in Chapter 6 for Santarém, Brazil.

Finally, scaling of the GCWG parameters provides an opportunity to produce synthetic precipitation and temperature daily time series of any length at a specified GWL. Several impact assessments (e.g., hydrological, ecological, agricultural) require very long time series to robustly determine risks. This method will provide impact assessors a tool that can produce long time series at a range of GWLs with computational efficiency, and the changing risk of extreme events may therefore be determined. This has been demonstrated in Chapters 6 and 7, including an assessment into the changing risks of compound extremes.

This thesis has combined a stochastic weather generator that has been calibrated globally (GCWG) with the pattern scaling technique to aid in the construction of scenarios that address future climate uncertainty, evaluate the changing risk to climate extremes, and emulate scenarios and time periods that GCMs have not yet simulated. Chapters 3 and 6 present manuscripts that have been published in the *International Journal of Climatology* and are being prepared for submission to *Climatic Change*, respectively, demonstrating novel research into global Markov-chain model order performance and the application of the pattern scaling technique to the parameters of a Markov-chain gamma-distribution SWG. The work presented in this thesis is an exciting development into the stochastic generation of weather under different climate scenarios, paving the way for further research in this field.

8.1 FUTURE RESEARCH THAT COULD BUILD ON THIS THESIS

This thesis has demonstrated strengths of the pattern scaling technique applied to the parameters of a stochastic weather generator. This has opened the door for several interesting developments that may be considered in furthering this research. Firstly, this thesis aimed to demonstrate the application of the PS technique to the parameters of a SWG and briefly investigate inter-model uncertainties in precipitation parameter response. To further capture

inter-model uncertainties, parameter response could be diagnosed from further GCMs. It has been evidenced in Chapter 6 that large differences in the resultant time series at GWLs may arise due to sizeable differences in parameter response simulated by two GCMs. This is a relatively simple extension, though it is intensive regarding computational time required to diagnose parameter changes, and for this reason has not been included in this thesis. Once the parameter response patterns have been diagnosed, however, the use of the combined PS-GCWG system allows weather sequences under many possible future climates to be generated very quickly and with much less computational power than running additional GCM simulations.

In this thesis, the parameters of a Markov-chain gamma-distribution SWG have been scaled with GMST. The parameters of such SWGs are simple to calculate from observations *and* GCM projections. However, there are several other methods of generating daily precipitation time series that each have their own strengths (discussed in greater detail in Chapter 2, Section 2.3). For example, series-type weather generators (e.g., LARS-WG (Semenov and Barrow, 2002)) are thought to represent the lengths of wet and dry spells with greater accuracy than Markov-type weather generators, though with several more parameters. The tail-ends of wet-spell and dry-spell length distributions have implications on the occurrence of extreme events such as drought. It therefore may be interesting to diagnose the response of parameters to climate forcing from alternative SWGs. It is, however, important to reiterate that the parameters of a Markov-chain gamma-distribution SWG are simple to calculate from GCM projections, resulting in a relatively straightforward, intuitive regression. The perturbation of alternative stochastic precipitation models using the response to climate forcing diagnosed from GCM projections may be more complex, especially for models that rely on sampling from observed data rather than based on parameters.

In addition to the parameters detailed in Table 8.1, there are several parameters that have been coded implicitly to the GCWG. These include the threshold precipitation for a wet day (the GCWG uses 0.1mm), the definition of a season throughout the analyses (assumed to be the temperate four seasons) and, for example, the number of additional months searched to ensure that the sample size for calculation of parameters (and similarly, their response to forcings) is large “enough”. These implicit parameters have not been assessed as part of this work, and have been decided based on wider literature and scientific reasoning. However, these choices *may* affect the resultant performance of the

GCWG, and testing of the weather generator's sensitivity to changes in these choices may prove interesting.

The observational data used in this thesis has been assumed to be of good quality; it is also potentially the only observational data that many impact assessors have access to. It is known that observational datasets are potentially prone to a range of biases. For example, a disproportionate number of observed precipitation amounts tends to end in "5" or "10" (dubbed the 5/10 error). One speculated cause of the 5/10 error is thought to be due to the way in which measuring sticks are labelled or marked (amongst a range of other potential causes; see Daly et al. 2007 for comprehensive suggestions). Further sources of bias in precipitation observations include: fewer recordings on Sundays, changes in data capture method during within an observational period, poorly collected or archived data and changes in site location that result in changed topography (Wilby et al., 2017; Viney and Bates, 2004). Such biases have been found to underestimate, for example, wet- and dry-spell lengths. Therefore, assessment of the quality of the raw observational data may be beneficial in truly representing the local-scale climate.

Chapter 3 presented an analysis of Markov chain model-order performance using a global, gridded network of weather station observations. The value of this global analysis is clear. The gamma distribution shows good performance at reproducing the overall distribution of wet-day precipitation; however, weaknesses were highlighted in Chapter 4 at reproducing extreme behaviour. Such inaccuracies might be reduced by using different distributions in some climatic regimes where the gamma distribution is less suitable. The choice of distribution has received relatively little attention in comparison to the choice of model order in both this thesis and the wider literature. It may therefore be interesting to discern whether there are differences regarding the best distribution to represent daily precipitation *amounts*, particularly the extreme end of the distribution, that are dependent on climate regime, or, as with Markov chain model-order, assessment metric, following a similar study to Chapter 3.

A further area of interest lies in the correlations between the primary and secondary variables. In this thesis, the correlations between the primary and secondary variables are assumed unchanging as a function of GMST increase due to limited literature assessing the relationship between the variables with climate change, and due to the computational complexity and larger sample sizes (e.g., more GCM ensemble members) required to diagnose the response of several coefficients from 4 regression equations per month. The

changing relationship between the two variables, and whether it may scale linearly with GMST, may be interesting to investigate.

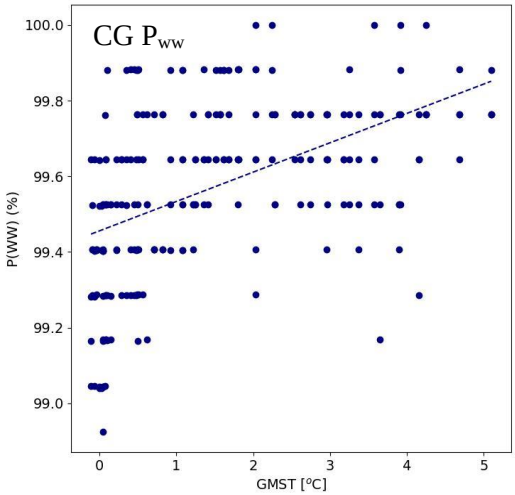
Finally, Chapter 5 introduced multiple linear regression models for both solar radiation and relative humidity. Tertiary variable parameters have not been scaled as a function of GMST and future work could consider how best to do this. It has been mentioned throughout the thesis that changes in the tertiary variables as a function of GMST may arise indirectly as a function of changes to the primary and secondary variables, which are perturbed by scaling their parameters with GMST. However, the equations provided in Chapter 5, Section 5.2.2 and 5.3.1 cannot be used to model future climates without some adjustment. Linear regression coefficients are calculated from the *residual* series of *observed* solar radiation, maximum temperature, minimum temperature, and precipitation. The *generated* series of precipitation, maximum and minimum temperature are also converted into residual series using *perturbed* (observed plus pattern-scaled changes according to the change in GMST) means and standard deviations. Therefore, the residual series do not themselves capture any changes to the mean temperature or precipitation, only the variability. Using the regression coefficients calculated from the observed residual series, a synthetic solar radiation residual series is generated. This of course results in the resultant solar radiation time series only incorporating changes to the variability as a function of change in the other variables, and not the mean. Future work could consider an alternative approach, such as scaling parameters such as monthly shape (wind speed) or means and standard deviations (solar radiation and relative humidity) with increasing GMST to produce changes to the tertiary variables.

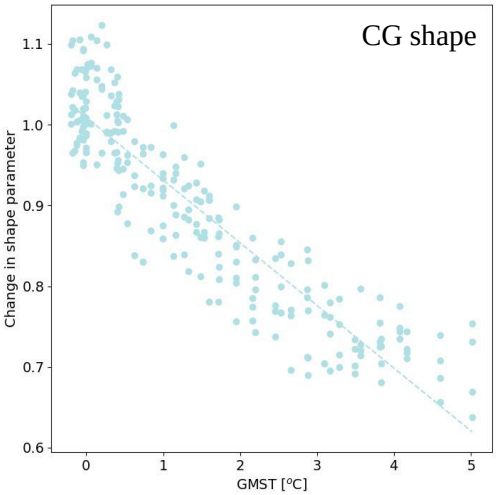
APPENDIX

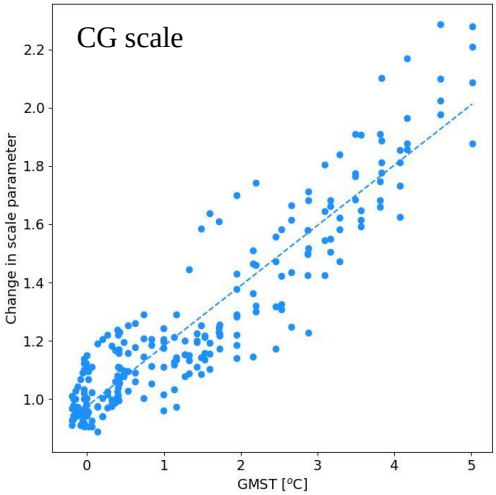
It's widely known that precipitation scales less linearly than temperature. To demonstrate the linear regressions for the Markov probabilities and wet-day parameters, a selection of grid-cells from the IPSL-CM6A-LR GCM representative of the five overarching Köppen climatic regimes have been chosen to demonstrate the varying strengths of regression (Figure A.1). R^2 values for the linear regressions have been calculated, and are shown in Table A.1. X-axis GMST shown in the corresponding figures is in reference to GMST change relative to the first 30-year period used to calculate the parameter. Note that linear regression performed on both shape and scale parameters are fractional changes relative to the first 30-year period used to calculate the parameter as discussed in Section 6.

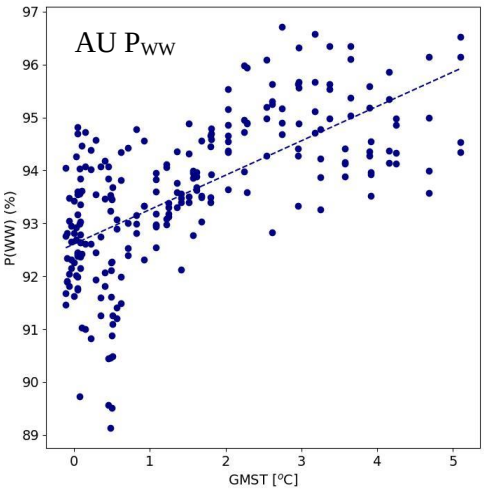
Table A.1 – Locations, month, Köppen zone, and R^2 scores for the linear regressions performed on a range of GCWG parameters. Corresponding locations are shown in Figure A.1 using the codes shown in the "Figure" column below. Location, latitude, and longitude, refer to a location within the GCM grid-cell.

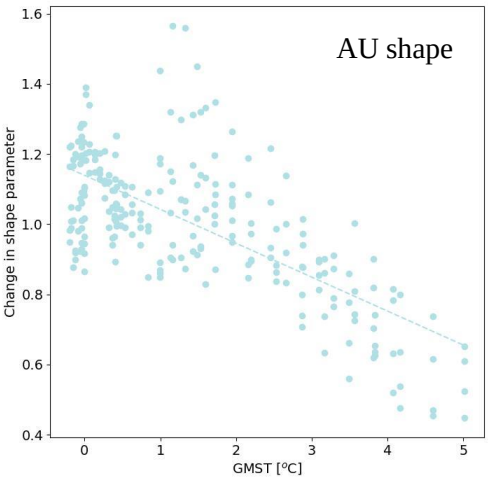
				Linear Regression R^2 for Parameter		
Figure	**Location (lat, lon)**	**Month**	**Koppen Zone**	**P_{ww}**	**shape**	**scale**
CG	Lobolo, Democratic Republic of the Congo, 7.4S, 30E	Feb	Aw	0.23	0.83	0.85
AU	Calvert, Australia, 16.2S, 137.5E	Jan	Am	0.43	0.48	0.11
CD	Bir Arakay, Chad, 15.2N, 20.0E	Nov	BWh	0.41	0.50	0.80
EC	Villa Seca, Ecuador, 3.8S, 80W	Jan	BSh	0.50	0.15	0.40
GR	Alma, Georgia, USA, 31.4N, 82W	Dec	Cfa	0.18	0.54	0.01
FR	Les Angles, France, 43.0N, 0E	May	Cfb	0.81	0.51	0.54
IN	St. Leon, Indiana, USA, 39.29N, 85W	July	Dfa	0.71	0.77	0.60
NW	Folldal, Norway 62.1N, 10E	Feb	Dfc	0.66	0.55	0.72
GL	Greenland, 72N, 35W	Nov	EF	0.54	0.68	0.85

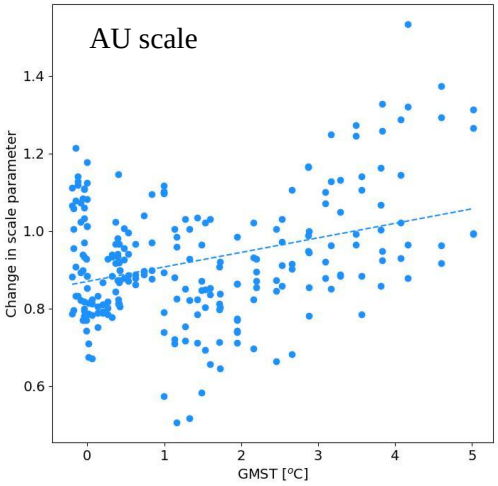

CG P_{ww}
P(WW) (%)
GMST [°C]

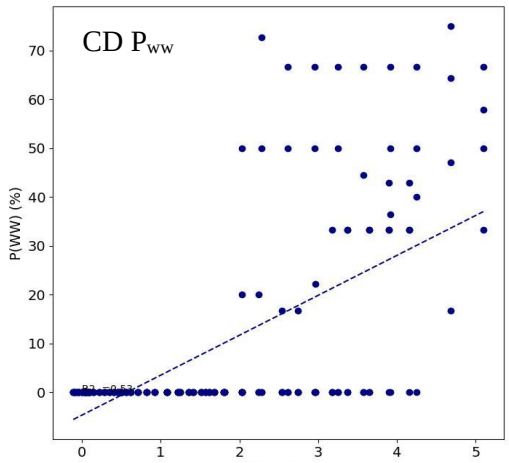

CG shape
Change in shape parameter
GMST [°C]

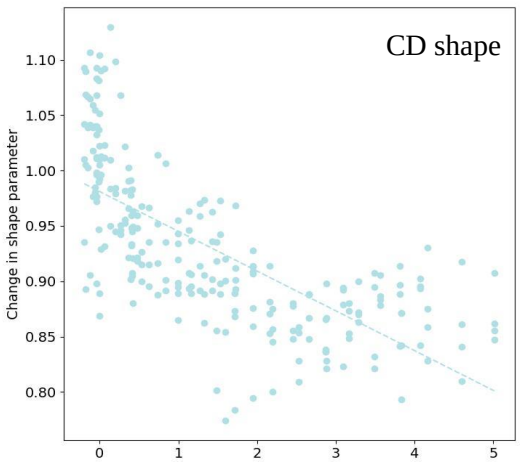

CG scale
Change in scale parameter
GMST [°C]

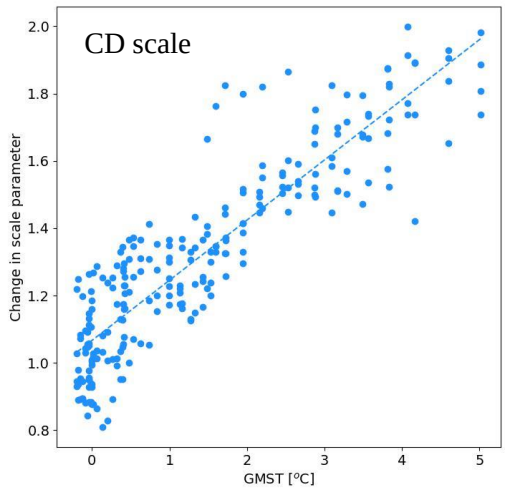

AU P_{ww}
P(WW) (%)
GMST [°C]

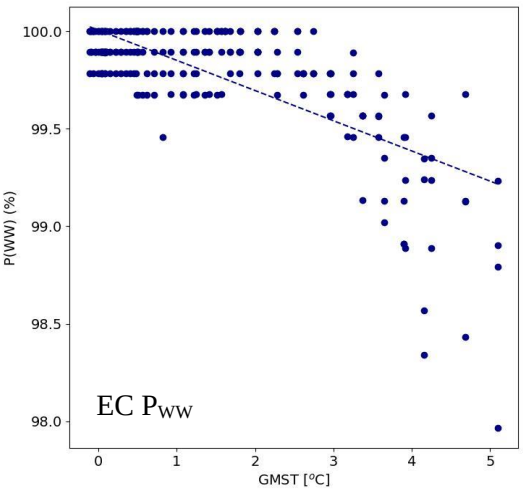

AU shape
Change in shape parameter
GMST [°C]

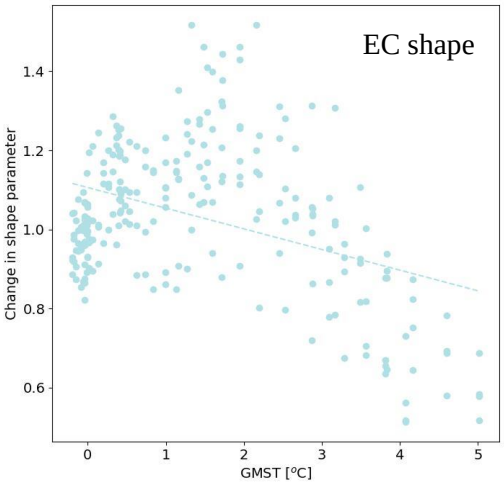

AU scale
Change in scale parameter
GMST [°C]

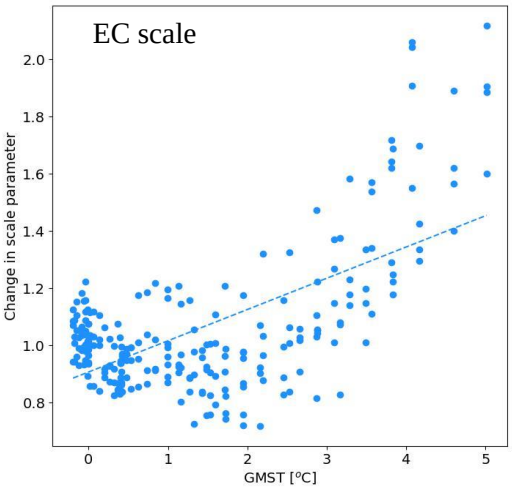

CD P_{ww}
P(WW) (%)

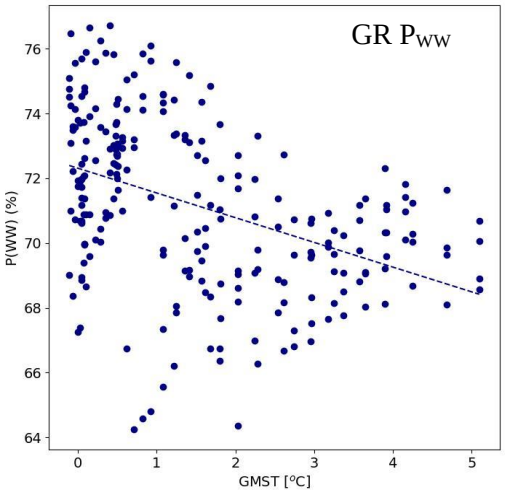

CD shape
Change in shape parameter

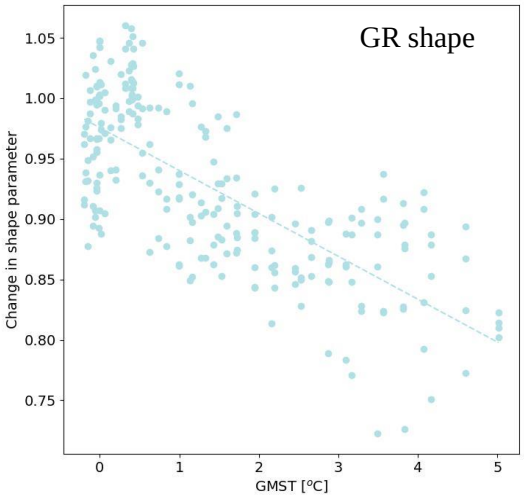

CD scale
Change in scale parameter
GMST [°C]

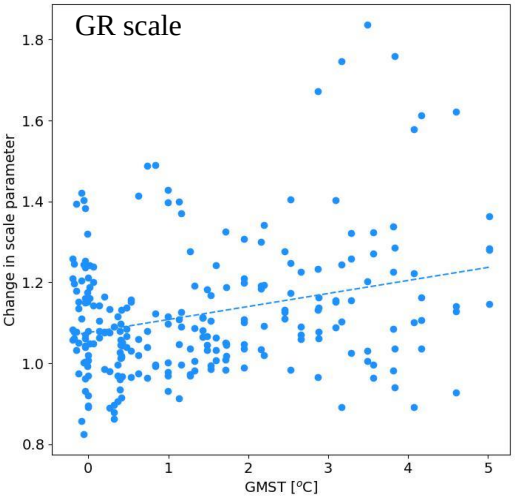

EC P_{ww}
P(WW) (%)
GMST [°C]

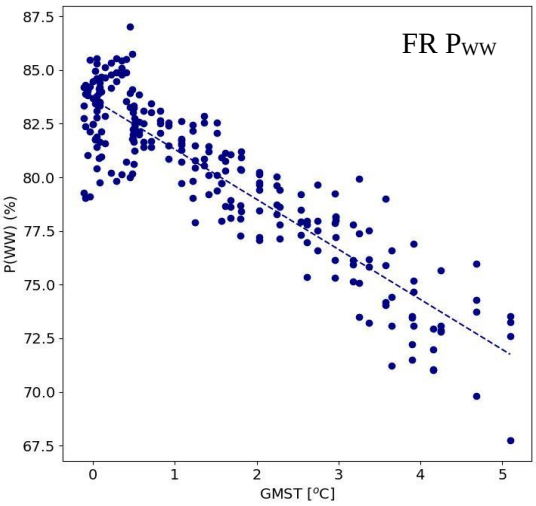

EC shape
Change in shape parameter
GMST [°C]

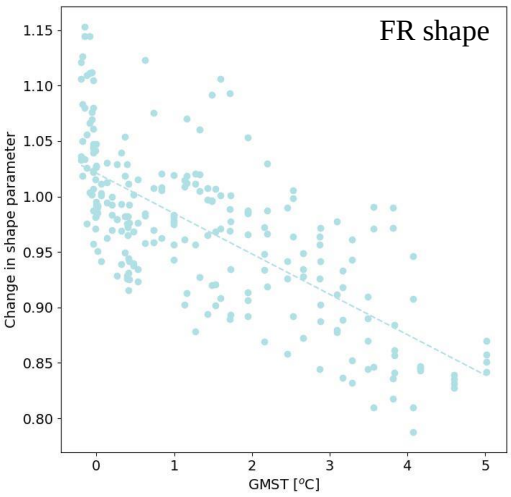

EC scale
Change in scale parameter
GMST [°C]

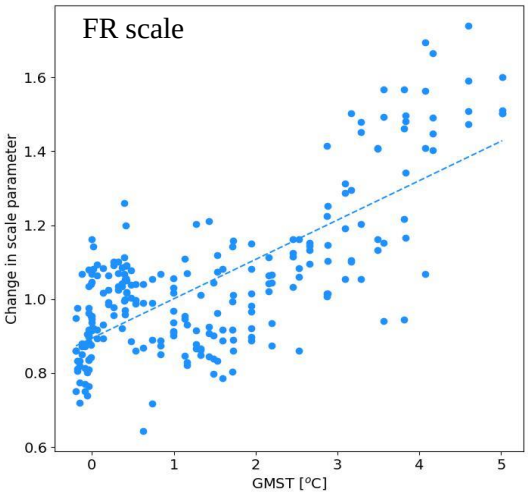

GR P_{WW}
P(WW) (%)
GMST [°C]

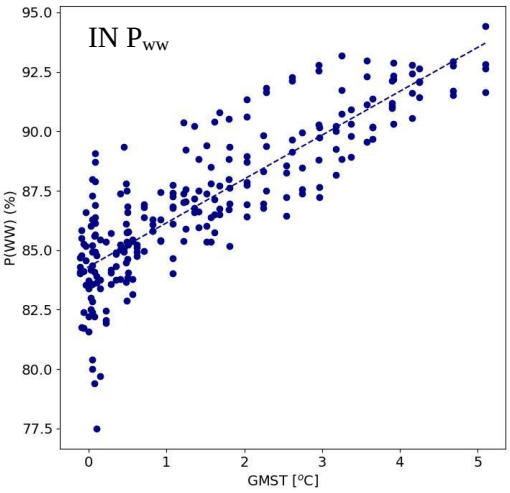

GR shape
Change in shape parameter
GMST [°C]

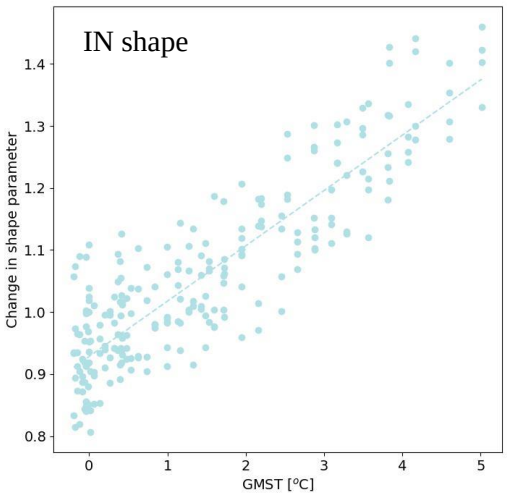

GR scale
Change in scale parameter
GMST [°C]

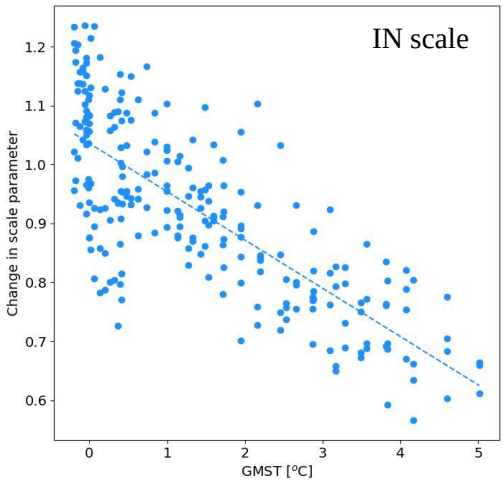

FR P_{WW}
P(WW) (%)
GMST [°C]

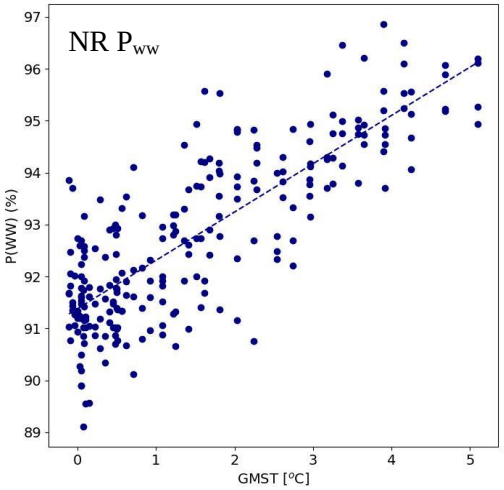

FR shape
Change in shape parameter
GMST [°C]

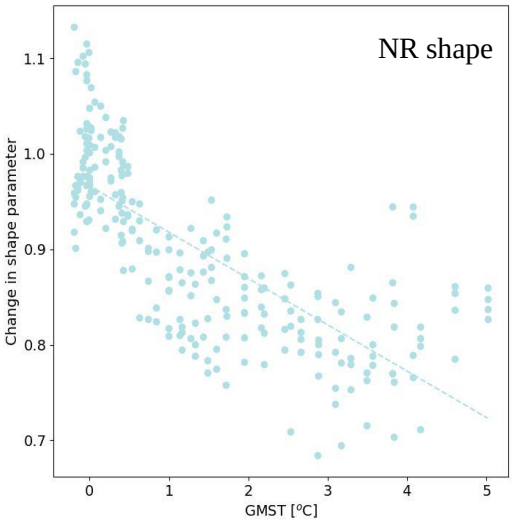

FR scale
Change in scale parameter
GMST [°C]

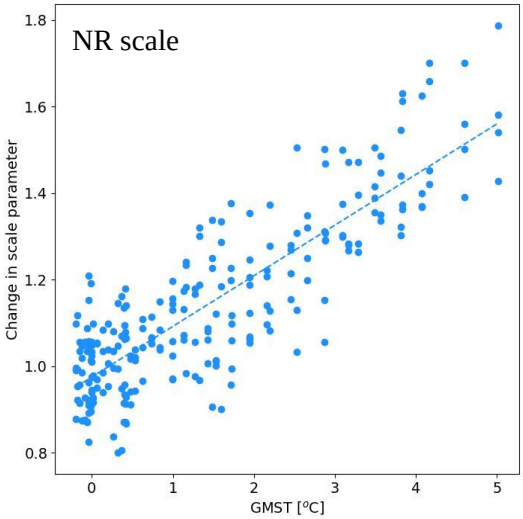

IN P_{ww}
P(WW) (%)
GMST [°C]


IN shape
Change in shape parameter
GMST [°C]


IN scale
Change in scale parameter
GMST [°C]


NR P_{ww}
P(WW) (%)
GMST [°C]


NR shape
Change in shape parameter
GMST [°C]


NR scale
Change in scale parameter
GMST [°C]

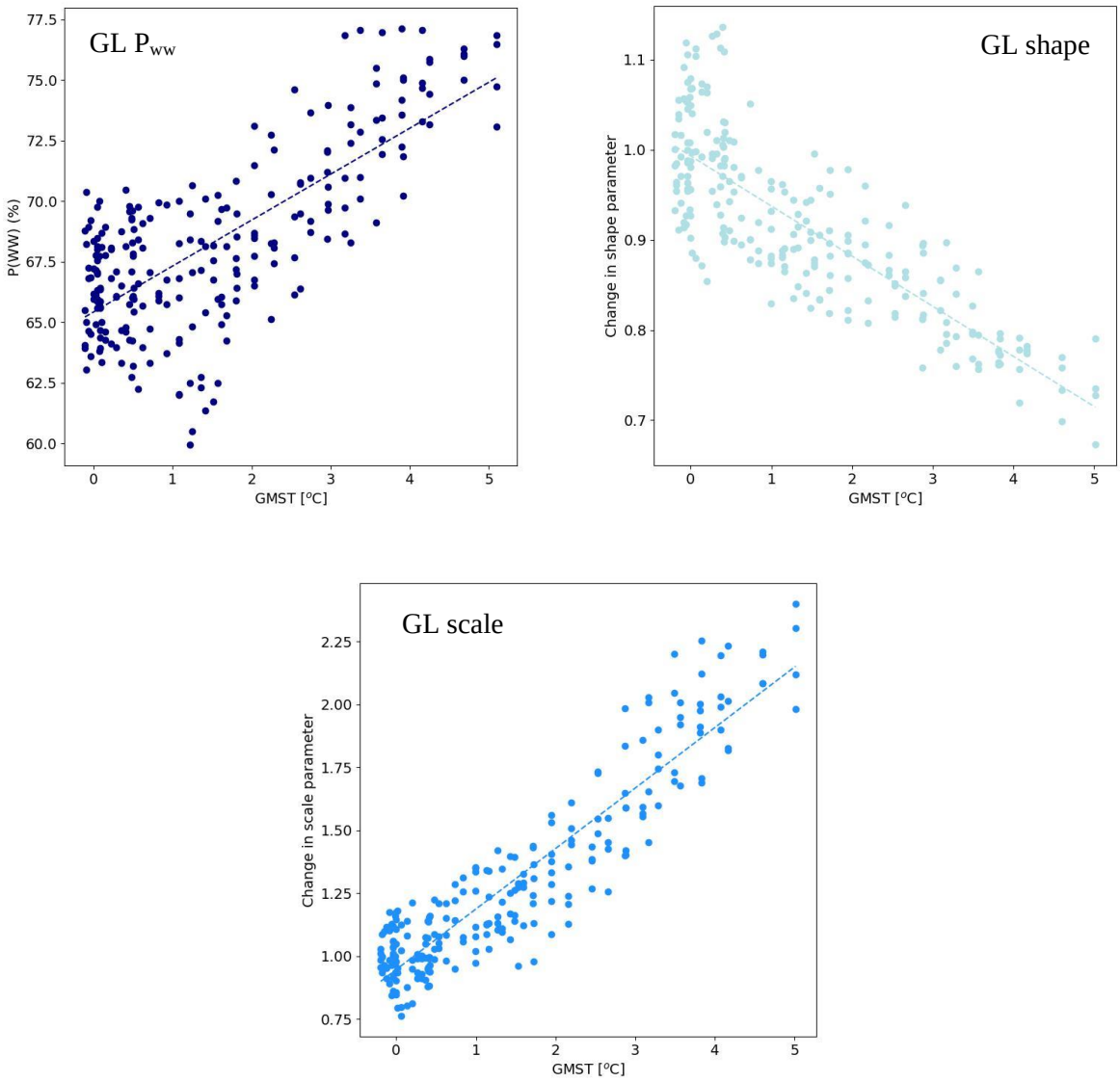


Figure A.1 – Scatter plots showing precipitation parameters against change in GMST. Location codes (e.g., GL) correspond to approximate GCM grid-cell location, with details given in Table A.1.

REFERENCES

Ackerman, E.A., 1941. The Köppen Classification of Climates in North America. Geogr. Rev. 31, 105–111.

Ailliot, P., Allard, D., Monbet, V., Naveau, P., 2015. Stochastic weather generators: an overview of weather type models. J. la Société Française Stat. 156, 101–113.

Aksoy, H., Toprak, Z.F., Aytek, A., Ünal, N.E., 2004. Stochastic generation of hourly mean wind speed data. Renew. Energy 29, 2111–2131. https://doi.org/10.1016/j.renene.2004.03.011

Alexander, L. V., Zhang, X., Peterson, T.C., Caesar, J., Gleason, B., Klein Tank, A.M.G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D.B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., Vazquez-Aguirre, J.L., 2006. Global observed changes in daily climate extremes of temperature and precipitation. J. Geophys. Res. Atmos. 111, 5109. https://doi.org/10.1029/2005JD006290

Alexeeff, S.E., Nychka, D., Sain, S.R., Tebaldi, C., 2018. Emulating mean patterns and variability of temperature across and within scenarios in anthropogenic climate change experiments. Clim. Change 146, 319–333. https://doi.org/10.1007/s10584-016-1809-8

Allen, R.G., Pereira, L.S., Howell, T.A., Jensen, M.E., 2011. Evapotranspiration information reporting: I. Factors governing measurement accuracy. Agric. Water Manag. https://doi.org/10.1016/j.agwat.2010.12.015

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 2006. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56, Remote Sensing of Environment.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. FAO Irrigation and Drainage Paper No. 56 - Crop Evapotranspiration, in: Food and Agriculture Organisation of the United Nations, Rome 300.

Almazroui, M., Islam, M.N., Jones, P.D., Athar, H., Rahman, M.A., 2012. Recent climate change in the Arabian Peninsula: Seasonal rainfall and temperature climatology of Saudi Arabia for 1979-2009. Atmos. Res. 111, 29–45.

https://doi.org/10.1016/j.atmosres.2012.02.013

Almazroui, M., Saeed, F., Saeed, S., Ismail, M., Ehsan, M.A., Islam, M.N., Abid, M.A., O'Brien, E., Kamil, S., Rashid, I.U., Nadeem, I., 2021. Projected Changes in Climate Extremes Using CMIP6 Simulations Over SREX Regions. Earth Syst. Environ. 5, 481–497. https://doi.org/10.1007/s41748-021-00250-5

Arnell, N.W., Gosling, S.N., 2016. The impacts of climate change on river flood risk at the global scale. Clim. Change 134, 387–401. https://doi.org/10.1007/s10584-014-1084-5

Arnell, N.W., Lowe, J.A., Brown, S., Gosling, S.N., Gottschalk, P., Hinkel, J., Lloyd-Hughes, B., Nicholls, R.J., Osborn, T.J., Osborne, T.M., Rose, G.A., Smith, P., Warren, R.F., 2013. A global assessment of the effects of climate policy on the impacts of climate change. Nat. Clim. Chang. 3, 512–519. https://doi.org/10.1038/nclimate1793

Ayodele, T.R., Ogunjuyigbe, A.S.O., 2015. Prediction of monthly average global solar radiation based on statistical distribution of clearness index. Energy 90, 1733–1742. https://doi.org/10.1016/j.energy.2015.06.137

Back, L.E., Bretherton, C.S., 2005. The relationship between wind speed and precipitation in the Pacific ITCZ. J. Clim. 18, 4317–4328. https://doi.org/10.1175/JCLI3519.1

Bannayan, M., Crout, N.M.J., 1999. A stochastic modelling approach for real-time forecasting of winter wheat yield. F. Crop. Res. 62, 85–95. https://doi.org/10.1016/S0378-4290(99)00008-8

Barrow, E., Hulme, M., Semenov, M., 1996. Effect of using different methods in the construction of climate change scenarios: Examples from Europe. Clim. Res. 7, 195–211. https://doi.org/10.3354/cr007195

Bathiany, S., Dakos, V., Scheffer, M., Lenton, T.M., 2018. Climate models predict increasing temperature variability in poor countries. Sci. Adv. 4. https://doi.org/10.1126/sciadv.aar5809

Belchansky, G.I., Douglas, D.C., Platonov, N.G., 2004. Duration of the Arctic sea ice melt season: Regional and interannual variability, 1979-2001. J. Clim. 17, 67–80. https://doi.org/10.1175/1520-0442(2004)017<0067:DOTASI>2.0.CO;2

Belda, M., Holtanová, E., Halenka, T., Kalvová, J., 2014. Climate classification revisited:

From Köppen to Trewartha. Clim. Res. 59, 1–13. https://doi.org/10.3354/cr01204

Bellprat, O., Doblas-Reyes, F., 2016. Attribution of extreme weather and climate events overestimated by unreliable climate simulations. Geophys. Res. Lett. 43, 2158–2164. https://doi.org/10.1002/2015GL067189

Boeke, R.C., Taylor, P.C., Sejas, S.A., 2021. On the Nature of the Arctic's Positive Lapse-Rate Feedback. Geophys. Res. Lett. 48, e2020GL091109. https://doi.org/10.1029/2020GL091109

Boucher, O., Servonnat, J., Albright, A.L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.Y., Guenet, B., Guez, Lionel, E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A.K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., Vuichard, N., 2020. Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. J. Adv. Model. Earth Syst. 12. https://doi.org/10.1029/2019MS002010

Brunner, M.I., Gilleland, E., Wood, A.W., 2021. Space-time dependence of compound hot-dry events in the United States: Assessment using a multi-site multi-variable weather generator. Earth Syst. Dyn. 12, 621–634. https://doi.org/10.5194/esd-12-621-2021

Cabré, M.F., Solman, S.A., Nuñez, M.N., 2010. Creating regional climate change scenarios over southern South America for the 2020's and 2050's using the pattern scaling technique: Validity and limitations. Clim. Change 98, 449–469. https://doi.org/10.1007/s10584-009-9737-5

Caya, D., Laprise, R., 1999. A semi-implicit semi-Lagrangian regional climate model: The Canadian RCM. Mon. Weather Rev. 127, 341–362. https://doi.org/10.1175/1520-

0493(1999)127<0341:asislr>2.0.co;2

Chen, D., Chen, H.W., 2013. Using the Köppen classification to quantify climate variation and change: An example for 1901–2010. Environ. Dev. 6, 69–79. https://doi.org/10.1016/J.ENVDEV.2013.03.007

Chen, D., Rojas, M., Samset, B.H., Cobb, K., Diongue Niang, A., Edwards, P., Emori, S., Faria, S.H., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., Mustafa, S.K., Plattner, G.-K., Tréguier, A.-M., 2021. Framing, Context, and Methods., in: Masson-Delmott, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Chen, J., Brissette, F.P., 2014. Comparison of five stochastic weather generators in simulating daily precipitation and temperature for the Loess Plateau of China. Int. J. Climatol. 34, 3089–3105. https://doi.org/10.1002/joc.3896

Chin, E.H., 1977. Modeling Daily Precipitation Occurrence Process With Markov Chain. Water Resour. Res. 13, 949–956. https://doi.org/10.1029/WR013i006p00949

Coe, R., Stern, R.D., 1982. Fitting Models to Daily Rainfall Data. J. Appl. Meteorol. 21, 1024–1031. https://doi.org/10.1175/1520-0450(1982)021<1024:fmtdrd>2.0.co;2

Cole, J.E., Overpeck, J.T., Cook, E.R., 2002. Multiyear La Niña events and persistent drought in the contiguous United States. Geophys. Res. Lett. 29, 25-1-25–4. https://doi.org/10.1029/2001GL013561

Cook, B.I., Mankin, J.S., Anchukaitis, K.J., 2018. Climate Change and Drought: From Past to Future. Curr. Clim. Chang. Reports. https://doi.org/10.1007/s40641-018-0093-2

Cooper, P.J.M., Dimes, J., Rao, K.P.C., Shapiro, B., Shiferaw, B., Twomlow, S., 2008. Coping better with current climatic variability in the rain-fed farming systems of sub-Saharan Africa: An essential first step in adapting to future climate change? Agric. Ecosyst. Environ. 126, 24–35. https://doi.org/10.1016/j.agee.2008.01.007

Coumou, D., Rahmstorf, S., 2012. A decade of weather extremes. Nat. Clim. Chang.

https://doi.org/10.1038/nclimate1452

Courty, L.G., Wilby, R.L., Hillier, J.K., Slater, L.J., 2019. Intensity-duration-frequency curves at the global scale. Environ. Res. Lett. 14, 084045. https://doi.org/10.1088/1748-9326/ab370a

Cowpertwait, P.S.P., 2004. Mixed rectangular pulses models of rainfall. Hydrol. Earth Syst. Sci. 8, 993–1000. https://doi.org/10.5194/hess-8-993-2004

Cowpertwait, P.S.P., 1994. Generalized point process model for rainfall. Proc. R. Soc. London, Ser. A Math. Phys. Sci. 447, 23–37. https://doi.org/10.1098/rspa.1994.0126

Cowpertwait, P.S.P., O'Connell, P.E., Metcalfe, A. V., Mawdsley, J.A., 1996. Stochastic point process modelling of rainfall. I. Single-site fitting and validation. J. Hydrol. 175, 17–46. https://doi.org/10.1016/S0022-1694(96)80004-7

Cox, D.T.C., Maclean, I.M.D., Gardner, A.S., Gaston, K.J., 2020. Global variation in diurnal asymmetry in temperature, cloud cover, specific humidity and precipitation and its association with leaf area index. Glob. Chang. Biol. 26, 7099–7111. https://doi.org/10.1111/gcb.15336

Cronin, T.W., 2020. How well do we understand the Planck feedback?, in: AGU Fall Meeting. AGU.

Dabhi, H., Rotach, M.W., Dubrovský, M., Oberguggenberger, M., 2021. Evaluation of a stochastic weather generator in simulating univariate and multivariate climate extremes in different climate zones across Europe. Meteorol. Zeitschrift 30, 127–151. https://doi.org/10.1127/metz/2020/1021

Dai, A., Zhao, T., Chen, J., 2018. Climate Change and Drought: a Precipitation and Evaporation Perspective. Curr. Clim. Chang. Reports. https://doi.org/10.1007/s40641-018-0101-6

Daly, C., Gibson, W.P., Taylor, G.H., Doggett, M.K., Smith, J.I., 2007. Observer Bias in Daily Precipitation Measurements at United States Cooperative Network Stations. Bull. Am. Meteorol. Soc. 88, 899-912. https://doi.org/10.1175/BAMS-88-6-899

Dash, S., Maity, R., 2021. Revealing alarming changes in spatial coverage of joint hot and wet extremes across India. Sci. Rep. 11, 1–15. https://doi.org/10.1038/s41598-021-97601-z

Dawkins, L.C., Osborne, J.M., Economou, T., Darch, G.J.C., Stoner, O.R., 2022. The Advanced Meteorology Explorer: a novel stochastic, gridded daily rainfall generator. J. Hydrol. 607, 127478. https://doi.org/10.1016/j.jhydrol.2022.127478

Doblas-Reyes, F.J., Sörensson, A.A., Almazroui, M., Dosio, A., Gutowski, W.J., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W.-T., Lamptey, B.L., Maraun, D., Stephenson, T.S., Takayabu, I., Terray, L., Turner, A., Zuo, Z., 2021. Linking Global to Regional Climate Change, in: Masson-Delmott, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. In Press.

Dubrovský, M., 1997. Creating Daily Weather Series with Use of the Weather Generator. Environmetrics 8, 409–424. https://doi.org/10.1002/(SICI)1099-095X(199709/10)8:5<409::AID-ENV261>3.0.CO;2-0

Dunn, R.J.H., Alexander, L. V., Donat, M.G., Zhang, X., Bador, M., Herold, N., Lippmann, T., Allan, R., Aguilar, E., Barry, A.A., Brunet, M., Caesar, J., Chagnaud, G., Cheng, V., Cinco, T., Durre, I., de Guzman, R., Htay, T.M., Wan Ibadullah, W.M., Bin Ibrahim, M.K.I., Khoshkam, M., Kruger, A., Kubota, H., Leng, T.W., Lim, G., Li-Sha, L., Marengo, J., Mbatha, S., McGree, S., Menne, M., de los Milagros Skansi, M., Ngwenya, S., Nkrumah, F., Oonariya, C., Pabon-Caicedo, J.D., Panthou, G., Pham, C., Rahimzadeh, F., Ramos, A., Salgado, E., Salinger, J., Sané, Y., Sopaheluwakan, A., Srivastava, A., Sun, Y., Timbal, B., Trachow, N., Trewin, B., van der Schrier, G., Vazquez-Aguirre, J., Vasquez, R., Villarroel, C., Vincent, L., Vischel, T., Vose, R., Bin Hj Yussof, M.N.A., 2020. Development of an Updated Global Land In Situ-Based Data Set of Temperature and Precipitation Extremes: HadEX3. J. Geophys. Res. Atmos. 125. https://doi.org/10.1029/2019JD032263

Eames, M., Kershaw, T., Coley, D., 2011. On the creation of future probabilistic design weather years from UKCP09. Build. Serv. Eng. Res. Technol. 32, 127–142. https://doi.org/10.1177/0143624410379934

Easterling, D.R., Meehl, G.A., Parmesan, C., Changnon, S.A., Karl, T.R., Mearns, L.O., 2000. Climate extremes: Observations, modeling, and impacts. Science (80-.). 289,

2068–2074. https://doi.org/10.1126/science.289.5487.2068

Eden, J.M., Widmann, M., 2014. Downscaling of GCM-simulated precipitation using model output statistics. J. Clim. 27, 312–324. https://doi.org/10.1175/JCLI-D-13-00063.1

Edwards, P.N., 2011. History of climate modeling. Wiley Interdiscip. Rev. Clim. Chang. 2, 128–139. https://doi.org/10.1002/wcc.95

Elsner, J.B., 2020. Continued increases in the intensity of strong tropical cyclones. Bull. Am. Meteorol. Soc. https://doi.org/10.1175/BAMS-D-19-0338.1

Fatichi, S., Ivanov, V.Y., Caporali, E., 2011. Simulation of future climate scenarios with a weather generator. Adv. Water Resour. 34, 448–467. https://doi.org/10.1016/j.advwatres.2010.12.013

Flannigan, M.., Stocks, B.., Wotton, B.., 2000. Climate change and forest fires. Sci. Total Environ. 262, 221–229. https://doi.org/10.1016/S0048-9697(00)00524-6

Forsythe, N., Fowler, H.J., Blenkinsop, S., Burton, A., Kilsby, C.G., Archer, D.R., Harpham, C., Hashmi, M.Z., 2014. Application of a stochastic weather generator to assess climate change impacts in a semi-arid climate: The Upper Indus Basin. J. Hydrol. 517, 1019–1034. https://doi.org/10.1016/j.jhydrol.2014.06.031

Foufoula-Georgiou, E., Lettenmaier, D.P., 1987. A Markov Renewal Model for rainfall occurrences. Water Resour. Res. 23, 875–884. https://doi.org/10.1029/WR023i005p00875

Frei, C., Christensen, J.H., Déqué, M., Jacob, D., Jones, R.G., Vidale, P.L., 2003. Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. J. Geophys. Res. Atmos. 108, 4124. https://doi.org/10.1029/2002jd002287

Frei, C., Schöll, R., Fukutome, S., Schmidli, J., Vidale, P.L., 2006. Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models. J. Geophys. Res. Atmos. 111, 6105. https://doi.org/10.1029/2005JD005965

Furrer, E.M., Katz, R.W., 2008. Improving the simulation of extreme precipitation events by stochastic weather generators. Water Resour. Res. 44. https://doi.org/10.1029/2008WR007316

Gabriel, K.R., Neumann, J., 1962. A Markov chain model for daily rainfall occurrence at Tel Aviv. Q. J. R. Meteorol. Soc. 88, 90–95. https://doi.org/10.1002/qj.49708837511

Gates, P., Tong, H., 1976. On Markov Chain Modeling to Some Weather Data. J. Appl. Meteorol. 15, 1145–1151. https://doi.org/10.1175/1520-0450(1976)015<1145:OMCMTS>2.0.CO;2

Ge, F., Zhu, S., Peng, T., Zhao, Y., Sielmann, F., Fraedrich, K., Zhi, X., Liu, X., Tang, W., Ji, L., 2019. Risks of precipitation extremes over Southeast Asia: Does 1.5°C or 2°C global warming make a difference? Environ. Res. Lett. 14, 044015. https://doi.org/10.1088/1748-9326/aaff7e

Geng, S., Penning de Vries, F.W.T., Supit, I., 1986. A simple method for generating daily rainfall data. Agric. For. Meteorol. 36, 363–376. https://doi.org/10.1016/0168-1923(86)90014-6

Gevaert, A.I., Miralles, D.G., de Jeu, R.A.M., Schellekens, J., Dolman, A.J., 2018. Soil Moisture-Temperature Coupling in a Set of Land Surface Models. J. Geophys. Res. Atmos. 123, 1481–1498. https://doi.org/10.1002/2017JD027346

Giorgi, F., 2008. A simple equation for regional climate change and associated uncertainty. J. Clim. 21, 1589–1604. https://doi.org/10.1175/2007JCLI1763.1

Giorgi, F., 2005. Interdecadal variability of regional climate change: Implications for the development of regional climate change scenarios. Meteorol. Atmos. Phys. 89, 1–15. https://doi.org/10.1007/s00703-005-0118-y

Gregory, J.M., Wigley, T.M.L., Jones, P.D., 1993. Application of Markov models to area-average daily precipitation series and interannual variability in seasonal totals. Clim. Dyn. 8, 299–310. https://doi.org/10.1007/BF00209669

Grübler, A., Nakicenovic, N., 2001. Identifying dangers in an uncertain climate. Nature 412, 15–15. https://doi.org/10.1038/35083752

Guidoum, A.C., 2015. Kernel Estimator and Bandwidth Selection for Density and its Derivatives: The kedd Package.

Gunawardhana, L.N., Al-Rawas, G.A., Kazama, S., 2017. An alternative method for predicting relative humidity for climate change studies. Meteorol. Appl. 24, 551–559. https://doi.org/10.1002/met.1641

Haarsma, R.J., Roberts, M.J., Vidale, P.L., Catherine, A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N.S., Guemas, V., Von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L.R., Lu, J., Luo, J.J., Mao, J., Mizielinski, M.S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., Von Storch, J.S., 2016. High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. Geosci. Model Dev. 9, 4185–4208. https://doi.org/10.5194/gmd-9-4185-2016

Haile, G.G., Tang, Q., Hosseini-Moghari, S.M., Liu, X., Gebremicael, T.G., Leng, G., Kebede, A., Xu, X., Yun, X., 2020. Projected Impacts of Climate Change on Drought Patterns Over East Africa. Earth's Futur. 8. https://doi.org/10.1029/2020EF001502

Harrison, M., Waylen, P., 2000. A note concerning the proper choice for Markov model order for daily precipitation in the humid tropics: A case study in Costa Rica. Int. J. Climatol. 20, 1861–1872. https://doi.org/10.1002/1097-0088(20001130)20:14<1861::AID-JOC577>3.0.CO;2-9

Hashmi, M.Z., Shamseldin, A.Y., Melville, B.W., 2011. Comparison of SDSM and LARS-WG for simulation and downscaling of extreme precipitation events in a watershed. Stoch. Environ. Res. Risk Assess. 25, 475–484. https://doi.org/10.1007/s00477-010-0416-x

Hassan, Z., Shamsudin, S., Harun, S., 2014. Application of SDSM and LARS-WG for simulating and downscaling of rainfall and temperature. Theor. Appl. Climatol. 116, 243–257. https://doi.org/10.1007/s00704-013-0951-8

He, Y., Manful, D., Warren, R., Forstenhäusler, N., Osborn, T.J., Price, J., Jenkins, R., Wallace, C., Yamazaki, D., 2022. Quantification of impacts between 1.5 and 4 °C of global warming on flooding risks in six countries. Clim. Change 170, 1–21. https://doi.org/10.1007/s10584-021-03289-5

Herger, N., Sanderson, B.M., Knutti, R., 2015. Improved pattern scaling approaches for the use in climate impact studies. Geophys. Res. Lett. 42, 3486–3494. https://doi.org/10.1002/2015GL063569

Hidalgo, H.G., Amador, J.A., Alfaro, E.J., Quesada, B., 2013. Hydrological climate change projections for Central America. J. Hydrol. 495, 94–112. https://doi.org/10.1016/j.jhydrol.2013.05.004

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S.,

Kim, H., Kanae, S., 2013. Global flood risk under climate change. Nat. Clim. Chang. 3, 816–821. https://doi.org/10.1038/nclimate1911

Hosseini, R., Le, N., Zidek, J., 2011. Selecting a binary Markov model for a precipitation process. Environ. Ecol. Stat. 18, 795–820. https://doi.org/10.1007/s10651-010-0169-1

Hulme, M., 1992. A 1951-80 global land precipitation climatology for the evaluation of general circulation models. Clim. Dyn. 7, 57–72. https://doi.org/10.1007/BF00209609

Husak, G.J., Michaelsen, J., Funk, C., 2007. Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. Int. J. Climatol. 27, 935–944. https://doi.org/10.1002/joc.1441

IPCC-TGICA, 1999. Guidelines on the Use of Scenario Data for Climate Impact and Adaptation Assessment. Version 1. Prepared by Carter, T.R., M. Hulme, and M. Lal, Intergovernmental Panel on Climate Change, Task Group on Scenarios for Climate Impact Assessment.

IPCC, 2022. Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. In Press.

IPCC, 2021a. The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, in: Masson-Delmott, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), Climate Change 2021. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC, 2021b. Annex II: Models [Gutiérrez, J M., A.-M. Tréguier (eds.)], in: Masson-Delmott, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 2087–2138. https://doi.org/10.1017/9781009157896.016.2087

Jimoh, O.D., Webster, P., 1996. The optimum order of a Markov chain model for daily rainfall in Nigeria. J. Hydrol. 185, 45–69. https://doi.org/10.1016/S0022-1694(96)03015-6

Jones, C.D., Hughes, J.K., Bellouin, N., Hardiman, S.C., Jones, G.S., Knight, J., Liddicoat, S., O'Connor, F.M., Andres, R.J., Bell, C., Boo, K.O., Bozzo, A., Butchart, N., Cadule, P., Corbin, K.D., Doutriaux-Boucher, M., Friedlingstein, P., Gornall, J., Gray, L., Halloran, P.R., Hurtt, G., Ingram, W.J., Lamarque, J.F., Law, R.M., Meinshausen, M., Osprey, S., Palin, E.J., Parsons Chini, L., Raddatz, T., Sanderson, M.G., Sellar, A.A., Schurer, A., Valdes, P., Wood, N., Woodward, S., Yoshioka, M., Zerroukat, M., 2011. The HadGEM2-ES implementation of CMIP5 centennial simulations. Geosci. Model Dev. 4, 543–570. https://doi.org/10.5194/gmd-4-543-2011

Jones, D.I., Lorenz, M.H., 1986. An application of a Markov chain noise model to wind generator simulation. Math. Comput. Simul. 28, 391–402. https://doi.org/10.1016/0378-4754(86)90074-1

Jones, P.D., Harpham, C., Burton, A., Goodess, C.M., 2016. Downscaling regional climate model outputs for the Caribbean using a weather generator. Int. J. Climatol. 36, 4141–4163. https://doi.org/10.1002/joc.4624

Jones, P.D., Harpham, C., Kilsby, C., Glenis, V., Burton, A., 2009. Projections of future daily climate for the UK from the Weather Generator, Weather.

Jones, P.G., Thornton, P.K., 1993. A rainfall generator for agricultural applications in the tropics. Agric. For. Meteorol. 63, 1–19. https://doi.org/10.1016/0168-1923(93)90019-E

Jones, R.N., 2001. An environmental risk assessment/management framework for climate change impact assessments. Nat. Hazards 23, 197–230. https://doi.org/10.1023/A:1011148019213

Katz, R.W., 1999. Extreme value theory for precipitation: Sensitivity analysis for climate change. Adv. Water Resour. 23, 133–139. https://doi.org/10.1016/S0309-1708(99)00017-2

Katz, R.W., 1981. On Some Criteria for Estimating the Order of a Markov Chain. Technometrics 23, 243–249. https://doi.org/10.1080/00401706.1981.10486293

Katz, R.W., Brown, B.G., 1992. Extreme events in a changing climate: Variability is more important than averages. Clim. Change 21, 289–302. https://doi.org/10.1007/BF00139728

Katz, R.W., Parlange, M.B., 1998. Overdispersion Phenomenon in Stochastic Modeling of Precipitation. J. Clim. 11, 591–601. https://doi.org/10.1175/1520-0442(1998)011<0591:OPISMO>2.0.CO;2

Kennett, E.J., Buonomo, E., 2006. Methodologies of pattern scaling across the full range of RT2A GCM ensemble members. Met Office Hadley Centre for Climate Prediction and Research: Exeter, UK.

Khanmohammadi, N., Rezaie, H., Behmanesh, J., 2021. The effect of autocorrelation on the diagnostic procedures. Meteorol. Atmos. Phys. 133, 565–577. https://doi.org/10.1080/03610926.2021.1971246

Kilpeläinen, A., Kellomäki, S., Strandman, H., Venäläinen, A., 2010. Climate change impacts on forest fire potential in boreal conditions in Finland. Clim. Change 103, 383–398. https://doi.org/10.1007/s10584-009-9788-7

Kilsby, C.G., Jones, P.D., Burton, A., Ford, A.C., Fowler, H.J., Harpham, C., James, P., Smith, A., Wilby, R.L., 2007. A daily weather generator for use in climate change studies. Environ. Model. Softw. 22, 1705–1719. https://doi.org/10.1016/j.envsoft.2007.02.005

Kim, Y.H., Min, S.K., Zhang, X., Sillmann, J., Sandstad, M., 2020. Evaluation of the CMIP6 multi-model ensemble for climate extreme indices. Weather Clim. Extrem. 29, 100269. https://doi.org/10.1016/j.wace.2020.100269

Kittel, T.G.F., Giorgi, F., Meehl, G.A., 1998. Intercomparsion of regional biases and doubled CO2-sensitivity of coupled atmosphere-ocean general circulation model experiments. Clim. Dyn. 14, 1–15. https://doi.org/10.1007/s003820050204

Klein Tank, A.M.G., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., Van Engelen, A.F.V., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Antonio López, J., Dahlström, B., Moberg, A., Kirchhofer,

W., Ceylan, A., Pachaliuk, O., Alexander, L. V., Petrovic, P., 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int. J. Climatol. 22, 1441–1453. https://doi.org/10.1002/joc.773

Klutse, N.A.B., Ajayi, V.O., Gbobaniyi, E.O., Egbebiyi, T.S., Kouadio, K., Nkrumah, F., Quagraine, K.A., Olusegun, C., Diasso, U., Abiodun, B.J., Lawal, K., Nikulin, G., Lennard, C., Dosio, A., 2018. Potential impact of 1.5 °C and 2 °C global warming on consecutive dry and wet days over West Africa. Environ. Res. Lett. 13, 055013. https://doi.org/10.1088/1748-9326/aab37b

Knutson, T.R., Camargo, S.J., Chan C.L.J., Emanuel. K., Ho, C-H., Kossin, J., Mohapatra, M., Satoh, M., Sugi, M., Walsh, K., Wu, L., 2019. Tropical Cyclones and Climate Change Assesment: Part I: Detection and Attribution. Bull. Am. Meteorol. Soc., 100, 1987-2007. https://doi.org/10.1175/BAMS-D-18-0189.1

Knutson, T.R., McBride, J.L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J.P., Srivastava, A.K., Sugi, M., 2010. Tropical cyclones and climate change. Nat. Geosci. https://doi.org/10.1038/ngeo779

Köppen, W., 1900. Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt. Geogr. Z. 6, 593–611.

Kreienkamp, F., Philip, S.Y., Tradowsky, J.S., Kew, S.F., Lorenz, P., Arrighi, J., Belleflamme, A., Bettmann, T., Caluwaerts, S., Chan, S.C., Ciavarella, A., De Cruz, L., de Vries, H., Demuth, N., Ferrone, A., Fischer, E.M., Fowler, H.J., Goergen, K., Heinrich, D., Henrichs, Y., Lenderink, G., Kaspar, F., Nilson, E., L Otto, F.E., 2021. Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021, Royal Netherlands Meteorological Institute (KNMI). Deutscher Wetterdienst.

Larsen, G.A., Pense, R.B., 1982. Stochastic Simulation of Daily Climatic Data for Agronomic Models. Agron. J. 74, 510–514. https://doi.org/10.2134/agronj1982.00021962007400030025x

Lee, J.-Y., Marotzke, J., Bala, G., L. Cao, S.C., Dunne, J.P., Engelbrecht, F., Fischer, E., Fyfe, J.C., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S., Zhou, T., 2021. Future global climate: scenario-based projections and near-term information, in: Masson-Delmott, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E.,

Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

Lennartsson, J., Baxevani, A., Chen, D., 2008. Modelling precipitation in Sweden using multiple step markov chains and a composite model. J. Hydrol. 363, 42–59. https://doi.org/10.1016/j.jhydrol.2008.10.003

Li, Y., Ye, W., Wang, M., Yan, X., 2009. Climate change and drought: a risk assessment of crop-yield impacts. Clim. Res. 39, 31–46. https://doi.org/10.3354/cr00797

Lionello, P., Scarascia, L., 2018. The relation between climate change in the Mediterranean region and global warming. Reg. Environ. Chang. 18, 1481–1493. https://doi.org/10.1007/s10113-018-1290-1

Loo, Y.Y., Billa, L., Singh, A., 2015. Effect of climate change on seasonal monsoon in Asia and its impact on the variability of monsoon rainfall in Southeast Asia. Geosci. Front. 6, 817–823. https://doi.org/10.1016/j.gsf.2014.02.009

Luber, G., McGeehin, M., 2008. Climate Change and Extreme Heat Events. Am. J. Prev. Med. https://doi.org/10.1016/j.amepre.2008.08.021

Lustenberger, A., Knutti, R., Fischer, E.M., 2014. The potential of pattern scaling for projecting temperature-related extreme indices. Int. J. Climatol. 34, 18–26. https://doi.org/10.1002/joc.3659

Lydia, M., Suresh Kumar, S., Immanuel Selvakumar, A., Edwin Prem Kumar, G., 2016. Linear and non-linear autoregressive models for short-term wind speed forecasting. Energy Convers. Manag. 112, 115–124. https://doi.org/10.1016/j.enconman.2016.01.007

Lynch, C., Hartin, C., Bond-Lamberty, B., Kravitz, B., 2017. An open-Access CMIP5 pattern library for temperature and precipitation: Description and methodology. Earth Syst. Sci. Data 9, 281–292. https://doi.org/10.5194/essd-9-281-2017

Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themel, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M., Thiele-Eich, I., 2010. Precipitation downscaling under climate change: Recent developments to bridge the gap between

dynamical models and the end user. Rev. Geophys. 48. https://doi.org/10.1029/2009RG000314

Marengo, J.A., Jones, R., Alves, L.M., Valverde, M.C., 2009. Future change of temperature and precipitation extremes in south america as derived from the PRECIS regional climate modeling system. Int. J. Climatol. 29, 2241–2255. https://doi.org/10.1002/joc.1863

Matalas, N.C., 1967. Mathematical assessment of synthetic hydrology. Water Resour. Res. 3, 937–945. https://doi.org/10.1029/WR003i004p00937

Matte, D., Larsen, M.A.D., Christensen, O.B., Christensen, J.H., 2019. Robustness and scalability of regional climate projections over Europe. Front. Environ. Sci. 7, 163. https://doi.org/10.3389/fenvs.2018.00163

May, W., 2012. Assessing the strength of regional changes in near-surface climate associated with a global warming of 2°C. Clim. Change 110, 619–644. https://doi.org/10.1007/s10584-011-0076-y

Mendelsohn, R., Emanuel, K., Chonabayashi, S., Bakkensen, L., 2012. The impact of climate change on global tropical cyclone damage. Nat. Clim. Chang. 2, 205–209. https://doi.org/10.1038/nclimate1357

Milly, P.C.D., Wetherald, R.T., Dunne, K.A., Delworth, T.L., 2002. Increasing risk of great floods in a changing climate. Nature 415, 514–517. https://doi.org/10.1038/415514a

Mitchell, J.F.B., Johns, T.C., Eagles, M., Ingram, W.J., Davis, R.A., 1999. Towards the construction of climate change scenarios. Clim. Change 41, 547–581. https://doi.org/10.1023/a:1005466909820

Mitchell, T.D., 2003. Pattern Scaling. An Examination of the Accuracy of the Technique for Describing Future Climates. Clim. Change 60, 217–242. https://doi.org/10.1023/A:1026035305597

Mohammadi, K., Goudarzi, N., 2018. Study of inter-correlations of solar radiation, wind speed and precipitation under the influence of El Niño Southern Oscillation (ENSO) in California. Renew. Energy 120, 190–200. https://doi.org/10.1016/j.renene.2017.12.069

Morice, C.P., Kennedy, J.J., Rayner, N.A., Winn, J.P., Hogan, E., Killick, R.E., Dunn, R.J.H., Osborn, T.J., Jones, P.D., Simpson, I.R., 2021. An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. J. Geophys. Res. Atmos. 126, e2019JD032361. https://doi.org/10.1029/2019JD032361

Mueller, B., Zhang, X., Zwiers, F.W., 2016. Historically hottest summers projected to be the norm for more than half of the world's population within 20 years. Environ. Res. Lett. 11, 044011. https://doi.org/10.1088/1748-9326/11/4/044011

Murphy, J., 1999. An evaluation of statistical and dynamical techniques for downscaling local climate. J. Clim. 12, 2256–2284. https://doi.org/10.1175/1520-0442(1999)012<2256:aeosad>2.0.co;2

Nfaoui, H., Buret, J., Sayigh, A.A.M., 1996. Stochastic simulation of hourly average wind speed sequences in Tangiers (Morocco). Sol. Energy 56, 301–314. https://doi.org/10.1016/0038-092X(95)00103-X

Nicks, A.D., Harp, J.F., 1980. Stochastic generation of temperature and solar radiation data. J. Hydrol. 48, 1–17. https://doi.org/10.1016/0022-1694(80)90062-1

O'Neill, B.C., Kriegler, E., Riahi, K., Ebi, K.L., Hallegatte, S., Carter, T.R., Mathur, R., van Vuuren, D.P., 2014. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. Clim. Change 122, 245–266. https://doi.org/10.1201/b20720-20

O'Neill, B.C., Tebaldi, C., Van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.F., Lowe, J., Meehl, G.A., Moss, R., Riahi, K., Sanderson, B.M., 2016. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. Geosci. Model Dev. 9, 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Ochola, W.O., Kerkides, P., 2003. A Markov chain simulation model for predicting critical wet and dry spells in Kenya: Analysing rainfall events in the kano plains. Irrig. Drain. 52, 327–342. https://doi.org/10.1002/ird.94

Ortega, G., Arias, P.A., Villegas, J.C., Marquet, P.A., Nobre, P., 2021. Present-day and future climate over central and South America according to CMIP5/CMIP6 models. Int. J. Climatol. 41, 6713–6735. https://doi.org/10.1002/joc.7221

Osborn, T.J., 1997. Areal and point precipitation intensity changes: Implications for the

application of climate models. Geophys. Res. Lett. 24, 2829–2832. https://doi.org/10.1029/97GL02976

Osborn, T.J., Wallace, C.J., Harris, I.C., Melvin, T.M., 2016. Pattern scaling using ClimGen: monthly-resolution future climate scenarios including changes in the variability of precipitation. Clim. Change 134, 353–369. https://doi.org/10.1007/s10584-015-1509-9

Osborn, T.J., Wallace, C.J., Lowea, J.A., Bernie, D., 2018. Performance of pattern-scaled climate projections under high-end warming. Part I: Surface air temperature over land. J. Clim. 31, 5667–5680. https://doi.org/10.1175/JCLI-D-17-0780.1

Parlange, M.B., Katz, R.W., 2000. An extended version of the Richardson model for simulating daily weather variables. J. Appl. Meteorol. 39, 610–622. https://doi.org/10.1175/1520-0450-39.5.610

Pastén-Zapata, E., Jones, J.M., Moggridge, H., Widmann, M., 2020. Evaluation of the performance of Euro-CORDEX Regional Climate Models for assessing hydrological climate change impacts in Great Britain: A comparison of different spatial resolutions and quantile mapping bias correction methods. J. Hydrol. 584, 124653. https://doi.org/10.1016/j.jhydrol.2020.124653

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Peleg, N., Fatichi, S., Paschalis, A., Molnar, P., Burlando, P., 2017. An advanced stochastic weather generator for simulating 2-D high-resolution climate variables. J. Adv. Model. Earth Syst. 9, 1595–1627. https://doi.org/10.1002/2016MS000854

Qian, B., Gameda, S., Hayhoe, H., 2008. Performance of stochastic weather generators LARS-WG and AAFC-WG for reproducing daily extremes of diverse Canadian climates. Clim. Res. 37, 17–33. https://doi.org/10.3354/cr00755

Qian, B., Hayhoe, H., Gameda, S., 2005. Evaluation of the stochastic weather generators LARS-WG and AAFC-WG for climate change impact studies. Clim. Res. 29, 3–21. https://doi.org/10.3354/cr029003

Racsko, P., Szeidl, L., Semenov, M., 1991. A serial approach to local stochastic weather models. Ecol. Modell. 57, 27–41. https://doi.org/10.1016/0304-3800(91)90053-4

Rajagopalan, B., Lall, U., Tarboton, D.G., 1997. Evaluation of kernel density estimation methods for daily precipitation resampling. Stoch. Hydrol. Hydraul. 11, 523–547. https://doi.org/10.1007/BF02428432

Riahi, K., van Vuuren, D.P., Kriegler, E., Edmonds, J., O'Neill, B.C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J.C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L.A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J.C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., Tavoni, M., 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Glob. Environ. Chang. 42, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Ribalaygua, J., Torres, L., Pórtoles, J., Monjo, R., Gaitán, E., Pino, M.R., 2013. Description and validation of a two-step analogue/regression downscaling method. Theor. Appl. Climatol. 114, 253–269. https://doi.org/10.1007/s00704-013-0836-x

Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. Water Resour. Res. 17, 182–190. https://doi.org/10.1029/WR017i001p00182

Richardson, C.W., Wright, D.A., 1984. WGEN: A Model for Generating Daily Weather Variables. United States Dep. Agric. Agric. Res. Serv. ARS-8 83.

Roberts, M.J., Camp, J., Seddon, J., Vidale, P.L., Hodges, K., Vannière, B., Mecking, J., Haarsma, R., Bellucci, A., Scoccimarro, E., Caron, L.P., Chauvin, F., Terray, L., Valcke, S., Moine, M.P., Putrasahan, D., Roberts, C.D., Senan, R., Zarzycki, C., Ullrich, P., Yamada, Y., Mizuta, R., Kodama, C., Fu, D., Zhang, Q., Danabasoglu, G., Rosenbloom, N., Wang, H., Wu, L., 2020. Projected Future Changes in Tropical Cyclones Using the CMIP6 HighResMIP Multimodel Ensemble. Geophys. Res. Lett. 47, e2020GL088662. https://doi.org/10.1029/2020GL088662

Roeckner, E., Bengtsson, L., Feichter, J., Lelieveld, J., Rodhe, H., 1999. Transient climate change simulations with a coupled atmosphere-ocean GCM including the tropospheric sulfur cycle. J. Clim. 12, 3004–3032. https://doi.org/10.1175/1520-0442(1999)012<3004:TCCSWA>2.0.CO;2

Rummukainen, M., 2016. Added value in regional climate modeling. Wiley Interdiscip. Rev. Clim. Chang. 7, 145–159. https://doi.org/10.1002/wcc.378

Rummukainen, M., 1997. Methods for statistical downscaling of GCM simulations. Reports Meteorololgy Climatol. 29.

Ruosteenoja, K., Räisänen, P., 2013. Seasonal changes in solar radiation and relative humidity in europe in response to global warming. J. Clim. 26, 2467–2481. https://doi.org/10.1175/JCLI-D-12-00007.1

Ruosteenoja, K., Tuomenvirta, H., Jylhä, K., 2007. GCM-based regional temperature and precipitation change estimates for Europe under four SRES scenarios applying a super-ensemble pattern-scaling method. Clim. Change 81, 193–208. https://doi.org/10.1007/s10584-006-9222-3

Sahin, A.D., Sen, Z., 2001. First-order Markov chain approach to wind speed modelling. J. Wind Eng. Ind. Aerodyn. 89, 263–269. https://doi.org/10.1016/S0167-6105(00)00081-7

Salinger, M.J., 2005. Climate variability and change: Past, present and future - An overview. Increasing Clim. Var. Chang. 9–29. https://doi.org/10.1007/1-4020-4166-7_3

Santer, B.D., Wigley, T.M.L., Schlesinger, M.E., Mitchell, J.F.B., 1990. Developing climate scenarios from equilibrium GCM results.

Savaresi, A., 2016. The Paris Agreement: A new beginning? J. Energy Nat. Resour. Law 34, 16–26. https://doi.org/10.1080/02646811.2016.1133983

Schoof, J.T., Pryor, S.C., 2008. On the proper order of Markov chain model for daily precipitation occurrence in the contiguous United States. J. Appl. Meteorol. Climatol. 47, 2477–2486. https://doi.org/10.1175/2008JAMC1840.1

Schwarz, G., 1978. Estimating the Dimension of a Model. Ann. Stat. 6, 461–464. https://doi.org/10.1214/aos/1176344136

Semenov, M.A., 2008. Simulation of extreme weather events by a stochastic weather generator. Clim. Res. 35, 203–212. https://doi.org/10.3354/cr00731

Semenov, M.A., Barrow, E.M., 2002. A Stochastic Weather Generator for Use in Climate Impact Studies.

Semenov, M.A., Barrow, E.M., 1997. Use of a stochastic weather generator in the development of climate change scenarios. Clim. Change 35, 397–414. https://doi.org/10.1023/A:1005342632279

Semenov, M.A., Brooks, R., Barrow, E., Richardson, C., 1998. Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. Clim. Res. 10, 95–107. https://doi.org/10.3354/cr010095

Seneviratne, S.I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Luca, A. Di, Ghosh, S., Iskandar, I., Kossin, J., S.Lewis, Otto, F., Pinto, I., M.Satoh, Vicente-Serrano, S.M., Wehner, M., Zhou, B., 2021. Weather and Climate Extreme Events in a Changing Climate, in: Masson-Delmott, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge and New York, NY, USA, pp. 1513–1766. https://doi.org/10.1029/2018GL080768

Sentelhas, P.C., Gillespie, T.J., Santos, E.A., 2010. Evaluation of FAO Penman-Monteith and alternative methods for estimating reference evapotranspiration with missing data in Southern Ontario, Canada. Agric. Water Manag. 97, 635–644. https://doi.org/10.1016/j.agwat.2009.12.001

Serenaldi, F., Kilsby, C.G., 2012. A modular class of multisite monthly rainfall generators for water resource management and impact studies. J. Hydrol. 464-465, 528-540. https://doi.org/10.1016/j.jhydrol.2012.07.043

Serreze, M.C., Meier, W.N., 2019. The Arctic's sea ice cover: trends, variability, predictability, and comparisons to the Antarctic. Ann. N. Y. Acad. Sci. 1436, 36–53. https://doi.org/10.1111/nyas.13856

Shiogama, H., Stone, D.A., Nagashima, T., Nozawa, T., Emori, S., 2013. On the linear additivity of climate forcing-response relationships at global and continental scales. Int. J. Climatol. 33, 2542–2550. https://doi.org/10.1002/joc.3607

Shrestha, A.K., Thapa, A., Gautam, H., 2019. Solar radiation, air temperature, relative humidity, and dew point study: Damak, Jhapa, Nepal. Int. J. Photoenergy 2019. https://doi.org/10.1155/2019/8369231

Skansi, M. de los M., Brunet, M., Sigró, J., Aguilar, E., Arevalo Groening, J.A., Bentancur, O.J., Castellón Geier, Y.R., Correa Amaya, R.L., Jácome, H., Malheiros Ramos, A., Oria Rojas, C., Pasten, A.M., Sallons Mitro, S., Villaroel Jiménez, C., Martínez, R., Alexander, L. V., Jones, P.D., 2013. Warming and wetting signals emerging from analysis of changes in climate extreme indices over South America. Glob. Planet. Change 100, 295–307. https://doi.org/10.1016/j.gloplacha.2012.11.004

Smith, D.M., Screen, J.A., Deser, C., Cohen, J., Fyfe, J.C., García-Serrano, J., Jung, T., Kattsov, V., Matei, D., Msadek, R., Peings, Y., Sigmond, M., Ukita, J., Zhang, X., 2019. The Polar Amplification Model Intercomparison Project (PAMIP) contribution to CMIP6: Investigating the causes and consequences of polar amplification. Geosci. Model Dev. 12, 1139–1164. https://doi.org/10.5194/gmd-12-1139-2019

Soltani, A., Hoogenboom, G., 2003a. A statistical comparison of the stochastic weather generators WGEN and SIMMETEO. Clim. Res. 24, 215–230. https://doi.org/10.3354/cr024215

Soltani, A., Hoogenboom, G., 2003b. Minimum data requirements for parameter estimation of stochastic weather generators. Clim. Res. 25, 109–119. https://doi.org/10.3354/cr025109

Sörensson, A.A., Menéndez, C.G., Ruscica, R., Alexander, P., Samuelsson, P., Willén, U., 2010. Projected precipitation changes in South America: A dynamical downscaling within CLARIS, in: Meteorologische Zeitschrift. pp. 347–355. https://doi.org/10.1127/0941-2948/2010/0467

Sørland, S.L., Schär, C., Lüthi, D., Kjellström, E., 2018. Bias patterns and climate change signals in GCM-RCM model chains. Environ. Res. Lett. 13, 074017. https://doi.org/10.1088/1748-9326/aacc77

Steinschneider, S., Brown, C., 2013. A semiparametric multivariate, multisite weather generator with low-frequency variability for use in climate risk assessments. Water Resour. Res. 49, 7205–7220. https://doi.org/10.1002/wrcr.20528

Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M.M.B., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., 2013. Climate change 2013 the physical science basis: Working Group I contribution to the fifth assessment report of the intergovernmental panel on climate change, Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the

Intergovernmental Panel on Climate Change. https://doi.org/10.1017/CBO9781107415324

Stott, P.A., Stone, D.A., Allen, M.R., 2004. Human contribution to the European heatwave of 2003. Nature 432, 610–614. https://doi.org/10.1038/nature03089

Sun, Q., Zhang, X., Zwiers, F., Westra, S., Alexander, L. V., 2021. A global, continental, and regional analysis of changes in extreme precipitation. J. Clim. 34, 243–258. https://doi.org/10.1175/JCLI-D-19-0892.1

Tebaldi, C., Arblaster, J.M., 2014. Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. Clim. Change 122, 459–471. https://doi.org/10.1007/s10584-013-1032-9

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., Van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K., Hurtt, G., Kriegler, E., Meehl, G., Moss, R., Bauer, S., Boucher, O., Brovkin, V., Yhb, Y., Dix, M., Gualdi, S., Guo, H., John, J., Kharin, S., Kim, Y.H., Koshiro, T., Ma, L., Olivié, D., Panickal, S., Qiao, F., Rong, X., Rosenbloom, N., Schupfner, M., Séférian, R., Sellar, A., Semmler, T., Shi, X., Song, Z., Steger, C., Stouffer, R., Swart, N., Tachiiri, K., Tang, Q., Tatebe, H., Voldoire, A., Volodin, E., Wyser, K., Xin, X., Yang, S., Yu, Y., Ziehn, T., 2021. Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6. Earth Syst. Dyn. 12, 253–293. https://doi.org/10.5194/esd-12-253-2021

Tebaldi, C., Knutti, R., 2018. Evaluating the accuracy of climate change pattern emulation for low warming targets. Environ. Res. Lett. 13, 055006. https://doi.org/10.1088/1748-9326/aabef2

Tebaldi, C., Wehner, M.F., 2018. Benefits of mitigation for future heat extremes under RCP4.5 compared to RCP8.5. Clim. Change 146, 349–361. https://doi.org/10.1007/S10584-016-1605-5/TABLES/1

Thom, H.C., 1958. A Note on the Gamma Distribution. Mon. Weather Rev. 86, 117–122.

Tong, H., 1975. Determination of the Order of a Markov Chain by Akaike's Information Criterion. J. Appl. Probab. 12, 488–497.

Torres, R.R., Marengo, J.A., 2013. Uncertainty assessments of climate change projections over South America. Theor. Appl. Climatol. 112, 253–272.

https://doi.org/10.1007/s00704-012-0718-7

Trenberth, K.E., 2011. Changes in precipitation with climate change. Clim. Res. 47, 123–138. https://doi.org/10.3354/cr00953

Trenberth, K.E., Dai, A., Van Der Schrier, G., Jones, P.D., Barichivich, J., Briffa, K.R., Sheffield, J., 2014. Global warming and changes in drought. Nat. Clim. Chang. https://doi.org/10.1038/nclimate2067

van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C., Kram, T., Krey, V., Lamarque, J.F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S.J., Rose, S.K., 2011. The representative concentration pathways: An overview. Clim. Change 109, 5–31. https://doi.org/10.1007/s10584-011-0148-z

Vautard, R., Gobiet, A., Sobolowski, S., Kjellström, E., Stegehuis, A., Watkiss, P., Mendlik, T., Landgren, O., Nikulin, G., Teichmann, C., Jacob, D., 2014. The European climate under a 2°C global warming. Environ. Res. Lett. 9, 034006. https://doi.org/10.1088/1748-9326/9/3/034006

Vincent, L.A., Peterson, T.C., Barros, V.R., Marino, M.B., Rusticucci, M., Carrasco, G., Ramirez, E., Alves, L.M., Ambrizzi, T., Berlato, M.A., Grimm, A.M., Marengo, J.A., Molion, L., Moncunill, D.F., Rebello, E., Anunciação, Y.M.T., Quintana, J., Santos, J.L., Baez, J., Coronel, G., Garcia, J., Trebejo, I., Bidegain, M., Haylock, M.R., Karoly, D., 2005. Observed trends in indices of daily temperature extremes in South America 1960-2000. J. Clim. 18, 5011–5023. https://doi.org/10.1175/JCLI3589.1

Viney, N.R., Bates B.C., 2004, It never rains on Sunday: the prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set. Int. J. Climatol. 24, 1171-1192. https://doi.org/10.1002/joc.1053

Vlček, O., Huth, R., 2009. Is daily precipitation Gamma-distributed? Adverse effects of an incorrect use of the Kolmogorov-Smirnov test. Atmos. Res. 93, 759–766. https://doi.org/10.1016/j.atmosres.2009.03.005

Vogel, M.M., Hauser, M., Seneviratne, S.I., 2020. Projected changes in hot, dry and wet extreme events' clusters in CMIP6 multi-model ensemble. Environ. Res. Lett. https://doi.org/10.1088/1748-9326/ab90a7

Vrac, M., Naveau, P., 2007. Stochastic downscaling of precipitation: From dry events to heavy rainfalls. Water Resour. Res. https://doi.org/10.1029/2006WR005308

Vrac, M., Stein, M., Hayhoe, K., 2007. Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing. Clim. Res. 34, 169–184. https://doi.org/10.3354/cr00696

Vu, T.M., Mishra, A.K., 2020. Performance of multisite stochastic precipitation models for a tropical monsoon region. Stoch. Environ. Res. Risk Assess. 34, 2159–2177. https://doi.org/10.1007/s00477-020-01871-4

Wallis, T.W.R., Griffiths, J.F., 1997. Simulated meteorological input for agricultural models. Agric. For. Meteorol. 88, 241–258. https://doi.org/10.1016/S0168-1923(97)00035-X

Wang, P., Goggins, W.B., Chan, E.Y.Y., 2018. A time-series study of the association of rainfall, relative humidity and ambient temperature with hospitalizations for rotavirus and norovirus infection among children in Hong Kong. Sci. Total Environ. 643, 414–422. https://doi.org/10.1016/j.scitotenv.2018.06.189

Wilby, R.L., Clifford, N.J., De Luca, P., Harrigan, S., Hillier, J.K., Hodgkins, R., Johnson, M.F., Matthews, T.K.R., Murphy, C., Noone, S.J., Parry, S., Prudhomme, C., Rice, S.P., Slater, L., Smith, K.A., Wood, P.J., 2017. The 'dirty dozen' of freshwater science: detecting then reconciling hydrological data biases and errors. WIREs Wat. 4, e1209. https://doi.org/10.1002/wat2.1209

Wilby, R.L., Conway, D., Jones, P.D., 2002a. Prospects for downscaling seasonal precipitation variability using conditioned weather generator parameters. Hydrol. Process. 16, 1215–1234. https://doi.org/10.1002/hyp.1058

Wilby, R.L., Dawson, C.W., Barrow, E.M., 2002b. SDSM - A decision support tool for the assessment of regional climate change impacts. Environ. Model. Softw. 17, 145–157. https://doi.org/10.1016/s1364-8152(01)00060-3

Wilby, R.L., Wigley, T.M.L., 2002, Future changes in the distribution of daily precipitation totals across North America. Geophys. Res. Lett. 29, 39-1. https://doi.org/10.1029/2001GL013048

Wilby, R.L., Wigley, T.M.L., 1997. Downscaling general circulation model output: A review of methods and limitations. Prog. Phys. Geogr. https://doi.org/10.1177/030913339702100403

Wilks, D.S., Wilby, R.L., 1999. The weather generation game: a review of stochastic

weather models. Prog. Phys. Geogr. 23, 329-357. https:/doi.org/10.1177/030913339902300302

Wilks, D.S., 2011. Statistical methods in the atmospheric sciences, Academic Press.

Wilks, D.S., 2010. Use of stochastic weather generators for precipitation downscaling. Wiley Interdiscip. Rev. Clim. Chang. 1, 898–907. https://doi.org/10.1002/wcc.85

Wilks, D.S., 1999a. Multisite downscaling of daily precipitation with a stochastic weather generator. Clim. Res. 11, 125–136. https://doi.org/10.3354/cr011125

Wilks, D.S., 1999b. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric. For. Meteorol. 93, 153–169. https://doi.org/10.1016/S0168-1923(98)00125-7

Williams, A.P., Abatzoglou, J.T., Gershunov, A., Guzman-Morales, J., Bishop, D.A., Balch, J.K., Lettenmaier, D.P., 2019. Observed Impacts of Anthropogenic Climate Change on Wildfire in California. Earth's Futur. 7, 892–910. https://doi.org/10.1029/2019EF001210

Wilson Kemsley, S., Osborn, T.J., Dorling, S.R., Wallace, C., Parker, J., 2021. Selecting Markov chain orders for generating daily precipitation series across different Köppen climate regimes. Int. J. Climatol. 41, 6223–6237. https://doi.org/10.1002/joc.7175

Wójcik, R., Pilarski, M., Miętus, M., 2014. Statistical downscaling of probability density function of daily precipitation on the Polish coast. Meteorol. Hydrol. Water Manag. 2, 27–36. https://doi.org/10.26491/mhwm/21590

Wood, A.W., Leung, L.R., Sridhar, V., Lettenmaier, D.P., 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Clim. Change 62, 189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9e

Wooten, R.D., 2011. Statistical Analysis of the Relationship Between Wind Speed, Pressure and Temperature. J. Appl. Sci. 11, 2712–2722.

Wu, J., Gao, X., 2020. Present day bias and future change signal of temperature over China in a series of multi-GCM driven RCM simulations. Clim. Dyn. 54, 1113–1130. https://doi.org/10.1007/s00382-019-05047-x

Wu, P., Wood, R., Ridley, J., Lowe, J., 2010. Temporary acceleration of the hydrological cycle in response to a CO 2 rampdown. Geophys. Res. Lett. 37, 12705.

https://doi.org/10.1029/2010GL043730

Yan, Z., Bate, S., Chandler, R.E., Isham, V., Wheater, H., 2002. An analysis of daily maximum wind speed in northwestern Europe using generalized linear models. J. Clim. 15, 2073–2088. https://doi.org/10.1175/1520-0442(2002)015<2073:AAODMW>2.0.CO;2

Yaoming, L., Qiang, Z., Deliang, C., 2004. Stochastic modeling of daily precipitation in China. J. Geogr. Sci. 14, 417–426. https://doi.org/10.1007/bf02837485

Zhu, J., Wang, S., Fischer, E.M., 2021. Increased occurrence of day–night hot extremes in a warming climate. Clim. Dyn. https://doi.org/10.1007/s00382-021-06038-7

Ziehn, T., Chamberlain, M.A., Law, R.M., Lenton, A., Bodman, R.W., Dix, M., Stevens, L., Wang, Y.P., Srbinovsky, J., 2020. The Australian Earth System Model: ACCESS-ESM1.5. J. South. Hemisph. Earth Syst. Sci. 70, 193–214. https://doi.org/10.1071/ES19035

Zscheischler, J., Seneviratne, S.I., 2017. Dependence of drivers affects risks associated with compound events. Sci. Adv. 3. https://doi.org/10.1126/sciadv.1700263

Zscheischler, J., Westra, S., Van Den Hurk, B.J.J.M., Seneviratne, S.I., Ward, P.J., Pitman, A., Aghakouchak, A., Bresch, D.N., Leonard, M., Wahl, T., Zhang, X., 2018. Future climate risk from compound events. Nat. Clim. Chang. https://doi.org/10.1038/s41558-018-0156-3