NUI MAYNOOTH

Ollscoil na hÉireann Má Nuad

QUANTIFYING SOURCES OF UNCERTAINTY IN REGIONAL CLIMATE MODEL SCENARIOS FOR IRELAND

Aideen Foley

Thesis submitted for the degree of PhD.

Irish Climate Analysis and Research Units,
Department of Geography,
National University of Ireland. Maynooth.

October 2010

Head of Department:
Prof. Mark Boyle

Supervisors:
Prof. John Sweeney and Dr. Rowan Fealy

ABSTRACT

This thesis develops a novel framework for model skill assessment and the generation of probabilistic future climate scenarios. Traditional approaches to model validation assume that skill in simulating the mean climate is a valid indicator of skill in modelling the climate system. However, without information about how errors arise, conclusions cannot be drawn about whether models are genuinely skilful.

Initially, verification statistics are used to assess model skill in simulating seasonal means and variability of Irish climate for 1961-1990. Significant biases were identified, however without further analysis, these biases cannot be attributed to a cause. Therefore, a spatial analysis, including EOF analysis, was undertaken which indicated that biases may be either spatially consistent (systematic) or inconsistent (random), an important distinction. Next, representation of a key large-scale driver of Irish climate, the North Atlantic Oscillation, was examined for a representative sub-sample of models. Skill in simulating the NAO was found to vary considerably between models. Therefore, assessing statistics of mean climate may not be the optimum way to characterize model skill, as deficiencies in the representation of large-scale drivers may not be detected.

Both quantitative and qualitative information from the skill assessments was used to inform probabilistic ensemble projections of future climate using Bayesian Model Averaging. In some cases, weighting scheme variation affects the ensemble PDF shape. In other cases, PDFs are similar when different weights are used, but the relative contributions of ensemble members vary. This is a crucial finding, as this underlying variation may not be immediately apparent, but may affect the confidence attached to the PDF. Therefore, robustness of ensemble generation methods must be considered when determining the level of confidence attached to a projection.

Finally, the implications of these results for climate decision-making are discussed and recommendations for the use of climate models in decision-making are presented.

ACKNOWLEDGEMENTS

It is a pleasure to thank the people who made this thesis possible. I am deeply grateful to my supervisors, Prof. John Sweeney and Dr. Rowan Fealy. This thesis would not have been possible without their unparalleled expertise, understanding and patience. They both provided so much encouragement, sound advice and lots of good ideas and I doubt that I will ever be able to convey my appreciation fully!

Funding for this research was provided by Cycle 4 of the Higher Education Authority's Programme for Research in Third-Level Institutions (PRTLI) and I am most grateful for their support. I would also like to thank all those who shared their valuable data, making this thesis possible. The RCM data was made available through the EU PRUDENCE project data archive and I would like to thank Dr. Ole Christensen for his generous assistance with accessing the data. Gridded observational datasets generated by Met Eireann were obtained from the British Irish Council and ERA-40 data was obtained from the ECMWF Data Server.

I thank all the staff of the Department of Geography who have helped me in some way or other over the past three years. I am especially grateful to those who provided me with research advice when needed; Dr. Satish Bastola, Dr. Conor Murphy and Dr. Priscilla Mooney. A very special thank you to all my colleagues within ICARUS for filling these past three years with such comraderie, entertainment and highly educational coffee breaks!

To Ciaran, thank you for your patience and support through all the highs and lows of completing this thesis. To my parents Margaret and Donnacha and my brother Neil, thank you for being encouraging and supportive in every way and always taking an interest in what I'm working on. Last but certainly not least, I wish to thank my dog, Ben, for his constant and joyful companionship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1
INTRODUCTION

1.1 GLOBAL CLIMATE CHANGE

The existence of an anthropogenic contribution to climate change is now well-established, with the latest report from the Intergovernmental Panel on Climate Change concluding with "very high confidence" that human-induced warming of the climate is taking place (IPCC, 2007: 37). According to the Goddard Institute for Space Studies (GISS), global surface temperatures have risen by an average of 0.8^{o}C since the late 1800s (Hansen *et al.*, 2010). The initial observed increases in temperature are associated with increasing use of refined coal at the time of the Industrial Revolution. Coal, along with petroleum and natural gas, is a fossil fuel formed by the decomposition of dead organisms over millions of years. These fuels contain high levels of carbon and hydrocarbon, and when burned, the carbon combines with oxygen to produce carbon dioxide (CO_2).

Records of atmospheric observations at Mauna Loa Observatory illustrate this change (Keeling *et al.*, 1976). Figure 1.1 illustrates the upward trend in atmospheric carbon dioxide. Due to its altitude, the air surrounding this station is relatively undisturbed and additionally, the location is both remote and minimally influenced by human activity, reducing the potential for "noise" or contamination in the data. Since 1956 this station has been continuously monitoring atmospheric carbon dioxide. As such, the Mauna Loa record is an important example of the changes observed in atmospheric carbon dioxide.

Carbon dioxide is a greenhouse gas (GHG), along with methane, water vapour and nitrous oxide. When such gases are produced through human activities such as industrial and agricultural activities, they are referred to as anthropogenic GHGs and when released into the atmosphere they exceed the natural rate of uptake (le Quere *et al.*, 2009), adding to the natural levels of these gases in the atmosphere, and absorb and emit thermal radiation. Today, fossil fuels remain an integral part of

the economy of industrialized nations and as such, GHGs continue to be emitted not only through energy generation but also through industrial and agricultural processes and transportation. For example, between chemical processes and fuel consumption, cement manufacture globally produces 5% of anthropogenic CO_2 emissions and it is estimated that for every 1000kg of cement, almost 900kg of CO_2 is also produced (Mahasenan *et al.,* 2005). The Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report, published in 2007, stated that global atmospheric concentrations of carbon dioxide, methane and nitrous oxide now far exceed pre-industrial values.

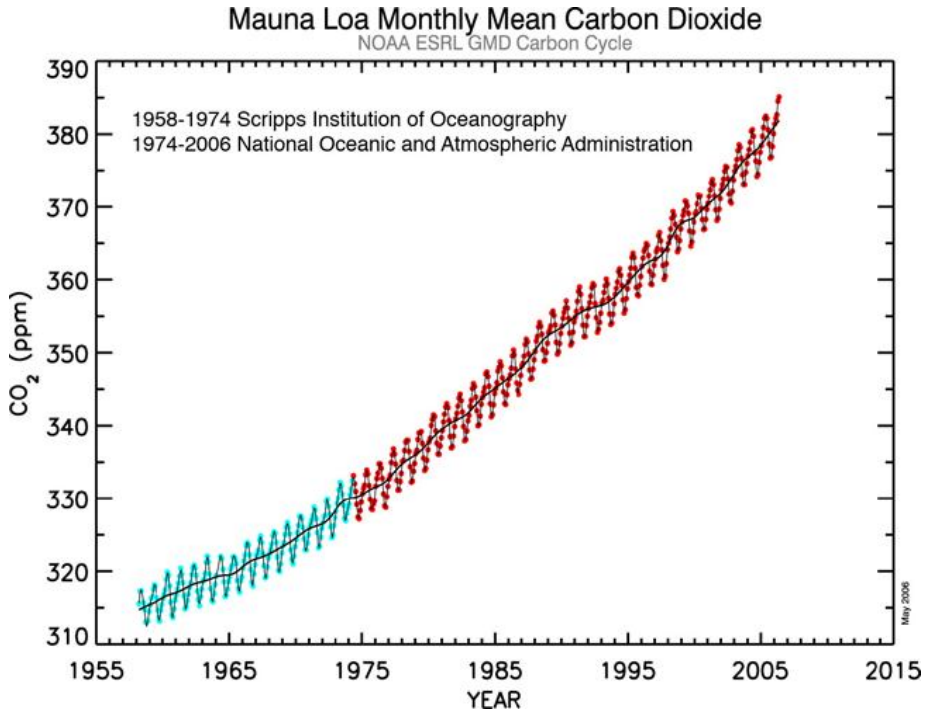


Figure 1.1: The Mauna Loa CO2 record, which indicates a two parts per million per year increase in atmospheric carbon dioxide since 1958. Smaller fluctuations indicate seasonal variations (Source: NOAA, 2006: http://celebrating200years.noaa.gov/datasets/).

The Vostok ice core provides a history of past CO_2 concentrations (Figure 1.2). The Vostok core, drilled in Antarctica, covers 420,000 years and 4 glacial-interglacial cycles. Analysis of the air contained in the ice can provide information about past atmospheric CO_2 variations and illustrates that under "natural" conditions, CO_2 concentrations varied between 180-280ppm. Yet present day measurements such as those carried out at Mauna Loa place CO_2 concentrations at approximately 380ppm and increasing. Therefore the Vostok core illustrates that present concentrations of CO_2 fall well outside what could be considered the natural range of

variability. The Vostok ice core record also demonstrates the relationship between global temperature and CO_2 concentrations, as information about temperature can be obtained through isotopic analysis of an ice core. When surface conditions are warm, there is more energy available for the evaporation of water containing ^{18}O, the heavier isotope of oxygen, which gets accumulated as snow. Ice core sections containing more ^{18}O relative to ^{16}O correspond to warmer climate phases. In this way, the ice core provides an extensive record of the co-varying relationship between atmospheric CO_2 and surface temperature. However, the initial timing of the increase in CO_2 and surface temperature is still subject to scientific study.

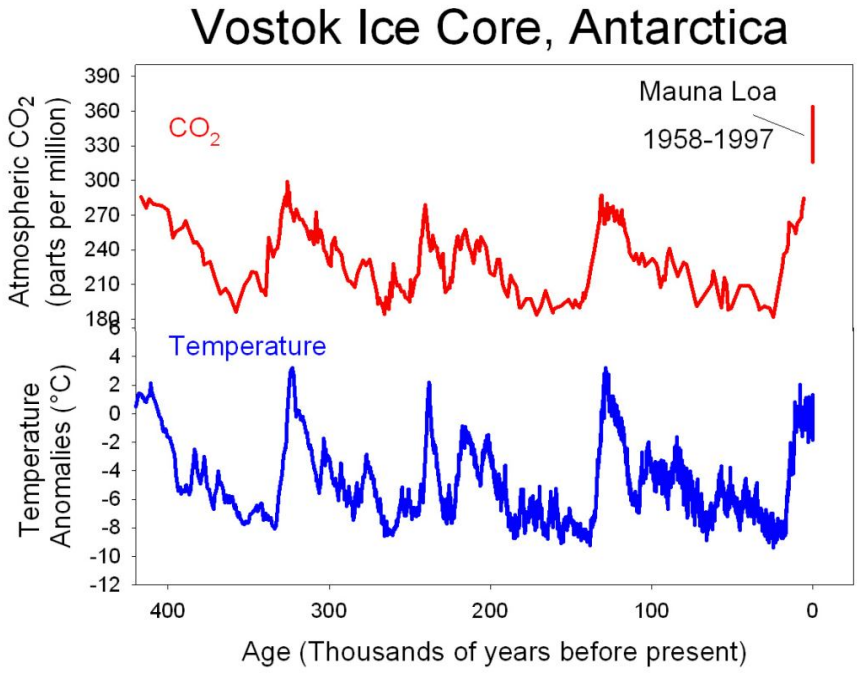


Figure 1.2: CO_2 (red) and temperature (blue) measurements from the Vostok, Antarctica ice core. Peaks of warmth occur approximately every 100,000 years. Temperature and atmospheric carbon dioxide concentrations appear to co-vary. Current CO_2 concentrations as observed at Mauna Loa are higher than at any time during the span of the ice core (Source: Petit et al., 1999).

1.2 THE NATURAL GREENHOUSE EFFECT

Without a certain naturally occurring concentration of GHGs to absorb radiation from the Sun, the Earth would be too cold to sustain life. Using the laws of blackbodies, an equation can be derived to calculate the temperature of the Earth with no greenhouse effect. A blackbody is an idealized object that is a perfect emitter and absorber of radiation at all wavelengths. The Stefan-Boltzmann law states that

the total energy radiated per unit surface area of a black body in unit time is given by:

Equation 1.1: Stefan-Boltzmann law

$$F = \sigma T^4$$

where

F = energy flux,

σ = Stefan-Boltzmann constant and

T = the blackbody's thermodynamic temperature (K).

Applying this equation to the amount of energy received by the Earth-atmosphere system and including a term to represent the Earth's albedo, the equation becomes:

Equation 1.2: Blackbody temperature of the Earth

$$T_E = \sqrt[4]{F_E / \sigma}$$

where

T_E = Earth's thermodynamic temperature (K),

σ = Stefan-Boltzmann constant and

F_E = Earth's energy flux = $\frac{(1-A)F_S}{4}$

where F_s = Sun's energy flux and

A = Earth's albedo.

Solving for an approximate planetary albedo of 0.3 and converting the result from Kelvin to degrees Celsius gives a temperature of -18°C. Clearly, there is another factor influencing the Earth's temperature and raising it to a more habitable temperature. This factor is the atmosphere, particularly its infrared (IR) emissivity which depends on atmospheric concentrations of IR active gases. The majority of the

atmosphere (O_2 and N_2 in particular) are transparent to IR radiation, but GHGs, such as CO_2 and water vapour, interact with IR radiation. Approximately half of the Sun's energy is absorbed at the Earth's surface. IR radiation is then emitted from the Earth's surface and much of this is absorbed and re-emitted by the GHGs in the atmosphere. Some of this energy is re-emitted downwards, contributing to warming and creating a natural greenhouse effect (Figure 1.3). Radiative forcing is the change in the net irradiance at the tropopause due to a change in an external driver of climate change and is a measure of the influence a given factor has in altering the energy balance (Figure 1.4). A warming influence is a positive forcing, while a cooling influence, such as certain types of aerosols (e.g. sulphur) is a negative forcing.

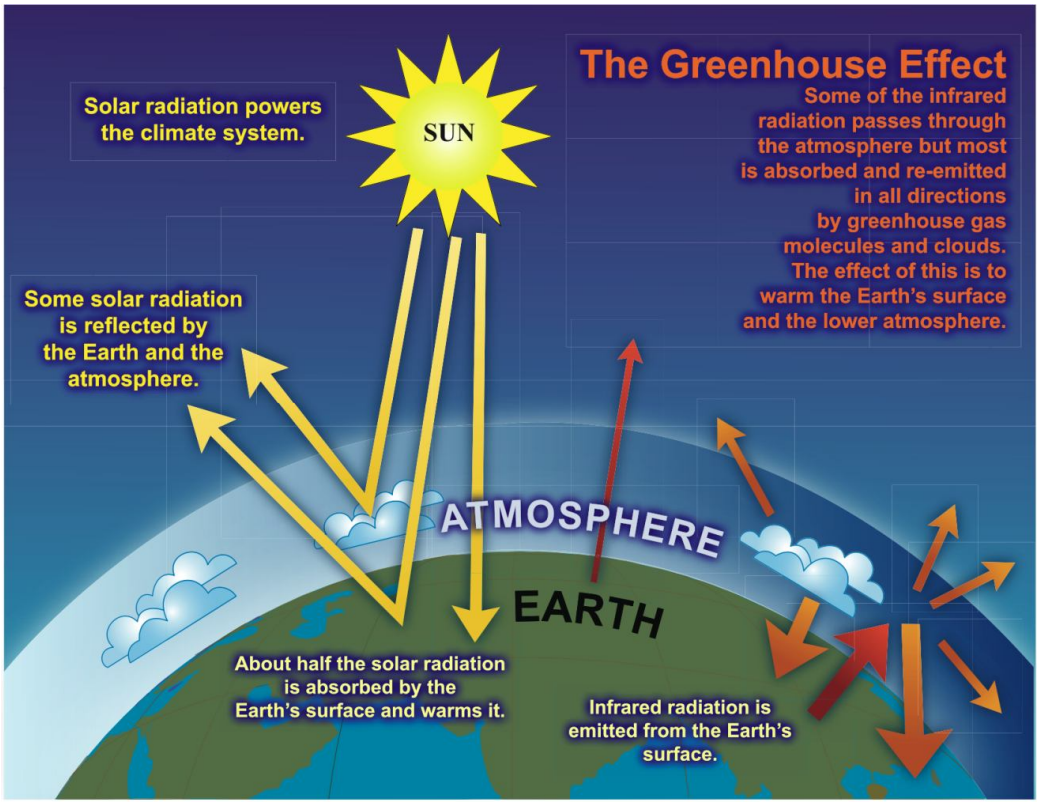


Figure 1.3: An idealised model of the natural greenhouse effect (Source: IPCC, 2007).

Unusually high concentrations of GHGs in the atmosphere can interfere with the natural balance of incoming and outgoing energy in the Earth-atmosphere system. According to the IPCC Fourth Assessment Report (2007) it is very likely (with 90% confidence or higher) that human activities since 1750 have exerted a net warming effect on the climate by increasing atmospheric GHG concentrations.

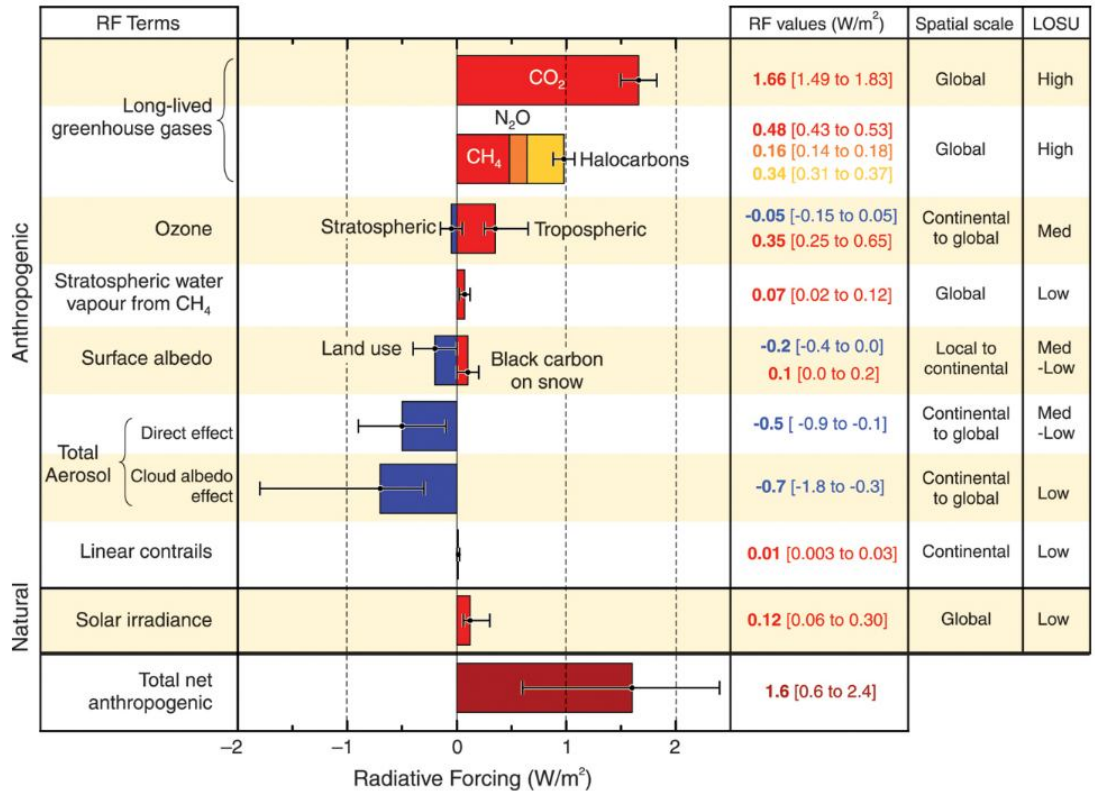


Figure 1.4: Components of radiative forcing (IPCC, 2007). Positive forcings are indicated by yellow to red bars while negative forcings are indicated by blue bars. Levels of scientific understanding (LOSU) vary for each component, however there is a high understanding of the main components, the long-lived GHGs including CO_2 (Source: IPCC, 2007).

Continued emission of greenhouse gases at or above the current rates is likely to result in further warming, and indeed even if GHG concentrations were stabilized, the timescales involved in climatic processes and feedbacks could result in continued warming (Wetherald *et al.*, 2001). Warming is evident in the melting of ice sheets in Greenland and parts of Antarctica which have very likely contributed to a rise in global average sea level. It is also evident in the widespread melting of glaciers and snow and in measurements of air and ocean temperatures (IPCC, 2007). According to the IPCC (2007: 30):

> "At continental, regional and ocean basin scales, numerous long-term changes in climate have been observed. These include changes in arctic temperatures and ice, widespread changes in precipitation amounts, ocean salinity, wind patterns and aspects of extreme weather including droughts, heavy precipitation, heat waves and the intensity of tropical cyclones."

These effects have a range of implications for human society, in relation to agriculture, water resources and flood risk, biodiversity and health. The effects of climate change can be positive or negative. For example, an increase in winter temperatures could decrease the potential for cold-related fatalities in that season. However, there are also potentially negative impacts associated with climate change, which society must be prepared for. Anthropogenic forcing has been found to have contributed to observed increases in precipitation in the Northern Hemisphere mid-latitudes and drying in the Northern Hemisphere tropic and sub-tropics (Zhang *et al.*, 2007), putting areas that are already highly vulnerable to water shortage under further threat (IPCC, 2007).

While it may be possible to prevent some of the more extreme impacts of climate change by transitioning to a low-carbon economy, the emissions already in the atmosphere make some level of climate change inevitable. How we cope currently with severe weather events such as storms, floods or dry periods highlights how we may potentially be susceptible to climate change hazards in the future. The extent to which climate change may harm a system is called its vulnerability, which is dependent on a system's sensitivity and on its ability to adapt to new climatic conditions. As the climate changes, our ability to protect sensitive systems may be tested, which is why response strategies are needed to plan for the possible impacts of climate change.

1.3 CLIMATE MODELLING

In order to develop suitable response strategies there is a need for more information about the climate changes that can be expected or considered likely, especially on the regional and local scales at which policy is formulated. Such information can be obtained using computer-driven climate models. The degree of complexity required in a model will depend on the nature of the questions asked. For example, simple energy balance models or Earth System Models of Intermediate Complexity (EMICs) may be suitable for conceptual studies, for example to simulate particular climate feedbacks (Claussen *et al.*, 2002). However, for climate change

impacts assessment, models with much higher spatial and temporal resolution are required.

1.3.1 Global climate models

Global climate models (GCMs) operate by discretising the equations for fluid motion and integrating them forward in time. They also contain parameterizations, which represent processes occurring on sub-grid scales that cannot be resolved directly. Atmospheric GCMs (AGCMs) model the atmosphere while coupled atmosphere-ocean GCMs (AOGCMs) combine both oceanic and atmospheric processes dynamically. While AOGCMs are highly complex, uncertainties still remain. The ability to model the climate system depends on having an appropriate level of scientific understanding of how the components of radiative forcing operate and influence climate. Some drivers have been the subject of more research than others (Figure 1.4) and as such are better understood. In particular, the extent to which solar activity and climate on Earth are related is strongly debated (e.g. Laut, 2003; Veizer, 2005).

Modelling future climate scenarios on a global domain ensures that the interactions between different climate regimes are handled properly. However, the computational demands of AOGCMs rule this out as a feasible option for simulating regional climate scenarios. In order to run these demanding models efficiently, their output is generally quite coarse. Typically a GCM has resolution of around 300km which is insufficient to resolve regional climate. It may seem like a contradiction that models that perform well at modelling global climate may perform poorly at the regional scale. However, the global climate is principally the response of the climate system to large-scale factors, like differences in solar forcing or in the earth's rotation, and global land–sea distribution and topography (Zorita and von Storch, 1999). The regional climates, of which there are many types, are the response of the climate system to regional details. When studying regional climate impacts, the domain of interest may be relatively small and as a result may not be represented very well on a GCM grid (Figure 1.5).

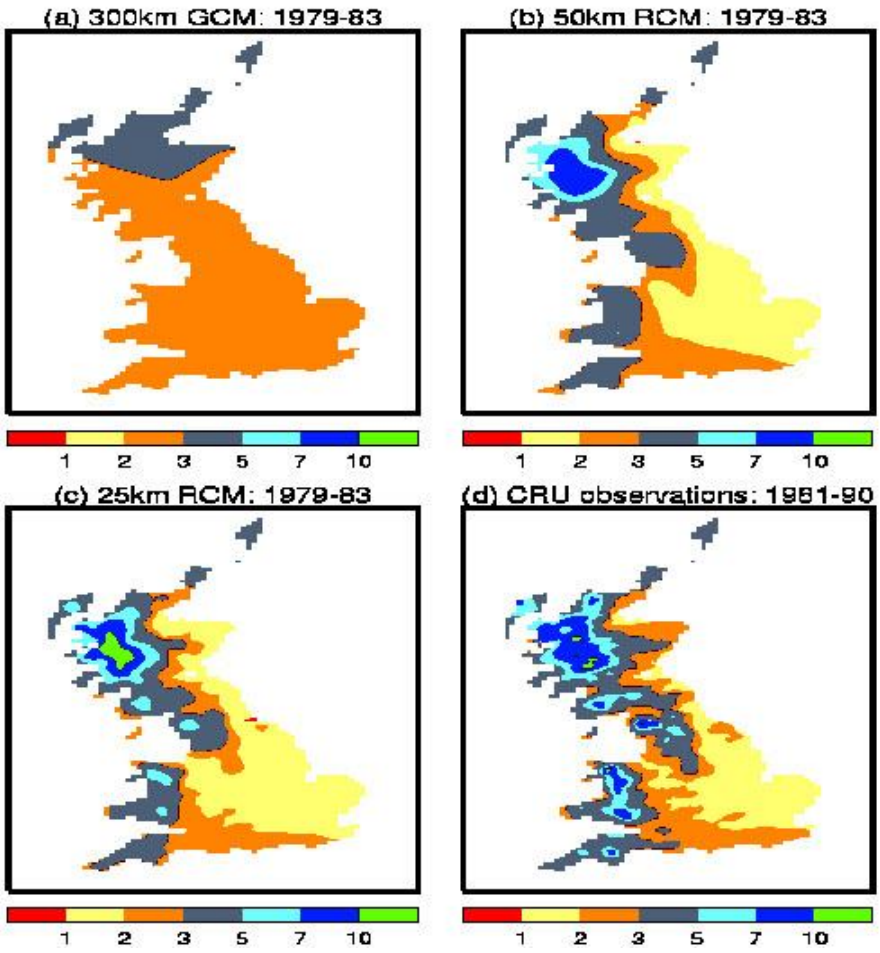


Figure 1.5: Winter precipitation over Britain as predicted by a) a GCM with resolution 300km, b) a regional model with 50km resolution and c) a regional model with 25km resolution compared to d) actual observations (Source:climateprediction.net: http://climateprediction.net/science/sci_images /RCM_improvements.jpg, accessed 22/7/2010).

Complex coastlines cannot be described at this resolution. Local topography such as mountains, and land cover information are also important forcings for local climate that cannot be represented at a coarse grid scale. In such a situation, it is useful to downscale GCM output to provide higher resolution over a smaller area.

1.3.2 Approaches to regional downscaling

There are a variety of methods which can be used to bridge the gap between global climate model output and the regional response, and these methods can be separated into two broad categories: empirical/statistical downscaling, or dynamical techniques. Essentially, empirical downscaling methodologies use statistical techniques to derive relationships between the large-scale climate and the regional

response (Benestad, 2004). A statistical model can be employed to relate the large-scale climate variables to surface environmental variables of interest, GCM output can be used as input to run the statistical model and point scale scenario information is obtained. This method has been shown by Trigo and Palutikof (2001) to reproduce the mean, variance and distribution of Iberian precipitation better than the GCM data alone. Statistical downscaling does not require a huge amount of computer resources, but large data-sets are required in the derivations. It is particularly well suited to situations where a dense network of observing stations exists, to provide the necessary datasets (Dunne *et al.*, 2008). However, the dataset used to calibrate a statistical model must span the range of natural variability, if the model is to be reliable.

The techniques applied to determine the relationships for the model can vary from linear methods such as canonical correlation analysis (CCA; e.g. Busuioc *et al.*, 2006) to non-linear techniques such as artificial neural networks (ANN), a computing approach based on human brain function (Hsieh and Tang, 1998). ANNs consist of an interconnected group of artificial neurons and process information using a connectionist approach to computation. ANNs are non-linear statistical modelling tools, which can model complex systems and find patterns in data. However, in some cases regression models and ANN models have been found to give similar results (Schoof and Pryor, 2001). Simple analogue methods have also been shown to have skill, and in a comparative study a simple analogue method of statistical downscaling was found to perform as well as more complex methods including CCA and ANN (Zorita and Von Storch, 1999). Empirical methods are inexpensive, requiring less computing power than dynamical methods. They are also less time-consuming.

A regional climate model or RCM is a dynamical downscaling technique that provides higher resolution (typically 50km) over a limited area. RCMs use a lot of computational resources as they explicitly describe the physical properties affecting climate. However, this also means that they should respond consistently to changes in forcing conditions.

RCMs take boundary inputs from AOGCMs. This means they are reliant on the results of a model that uses coarser resolution. If there are errors in the GCM, the RCM will not usually correct them. Mearns *et al.* (2003) demonstrates that in the

south-eastern United States, an RCM (RegCM2 forced by CSIRO Mk2) reproduces the observed pattern for precipitation and min/max temperature better than the GCM, for all seasons. In this situation, the RCM can clearly resolve the Florida peninsula and the Appalachian Mountains, whereas the GCM is too coarse to do so.

Many comparative studies of the two downscaling approaches have been carried out. While Murphy *et al.* (1999) found no significant differences comparing the techniques over the 1983-1994 period for Europe, Spak *et al.* (2007) found that the different techniques produce different spatial patterns of temperature across North America for the period 2000-2087, which diverge significantly from historical differences. While a climate system undergoing change is intrinsically a non-stationary system, the success of empirical downscaling requires an assumption of stationarity for the constancy of relationships to be maintained, and this is a perceived weakness of statistical methods. The relationships derived based on historical records may not hold in a future climate scenario with different forcing conditions (Wilby *et al.*, 2002). The assumption is that the statistical model that best describes the relationships between the large and small scale variables in the present will also best represent the relationships under climate change, and this assumption is inherently unverifiable.

Conversely, dynamical methods such as regional climate modelling are based on mathematical equations that describe the fundamental physical processes of the climate system. Dynamical methods are not completely free of assumption. Some processes occur on scales too small for even an RCM to resolve and may be represented instead in terms of their large scale effects. This is known as a parameterization and is may be carried out to simplify and speed up the run-time of a model, or to overcome gaps in scientific knowledge which may make the physical inclusion of a particular process infeasible. In such cases, an assumption is made that the relationship between the small scale process and the large scale effect will remain constant in time, an assumption which cannot be confirmed. That said, the core basis of climate models is immutable physics, not unverifiable assumptions, making a stronger argument for their validity under altered forcing conditions. A second advantage of RCMs as a downscaling method is that with their increased resolution (25-50km) they incorporate information about regional land-use and topography,

making them more skilful over complex terrain such as mountainous regions than statistical methods (Schmidli *et al.*, 2007).

There is no one best method of downscaling. The optimum method depends on the research questions being asked and the region of interest. However, a number of studies (Hellstrom, 2001; Murphy, 1999) have found that in the present, projections based on statistical downscaling compare well with those based on dynamical downscaling, but the future projections of the two methods vary. Murphy (1999) compared output from a GCM and a nested RCM, both configurations of the UK Met Office Unified Model, and a statistical regression-based method. The optimum model was found to vary in each season due to biases in the models themselves; e.g. in summer the RCM modelled excessively dry soil. The statistical model and the RCM perform equally well at simulating present climate, and are more accurate than the GCM. Murphy (2000) then used these methods to simulate future climate change, and found that while the methods were initially equal, evaluating present climate with equal skill, they do not produce similar projections of future climate. For temperature projections, the changes projected by the statistical and RCM methods differ by 40-50% Murphy (2000).

As these are future projections there is no empirical data that could be used to verify the projections and determine which technique is more skilful. However, one reason for this difference may be that the statistical relationships that hold at present may not hold under the different conditions of possible future climates. It has been suggested (Collins, 2007) that statistical approaches are not valid for simulating future climate change as they cannot be reliably used to make extrapolations outside the period on which they are based. On the other hand though the dynamical method, regional climate modelling, utilizes statistical approximations in its parameterization schemes, it is mainly based on physical relationships. As such, it calculates variables in an objective manner and this is its main advantage.

1.4 UNCERTAINTY IN CLIMATE MODELLING

There are many models available, each with slightly different ways of representing certain aspects of the climate, though all obey the same fundamental

equations. As such, model choice is an important consideration in any study of climate change and impacts as intermodel differences can lead to a range of different outputs, even when models are forced using the same emissions scenarios (Figure 1.6).

At every stage in the regional climate modelling process, uncertainties occur which affect the outcome of the next stage. For example, the choice of emissions scenario will have a significant impact on modelled output as the results produced by a GCM using the A1B emissions scenario (Figure 1.6 right) will be very different to the results produced by the same GCM using the B1 scenario (Figure 1.6 centre). Similarly, different GCMs will model different outcomes even when the same emissions scenario is used, due to differences in the models' construction.

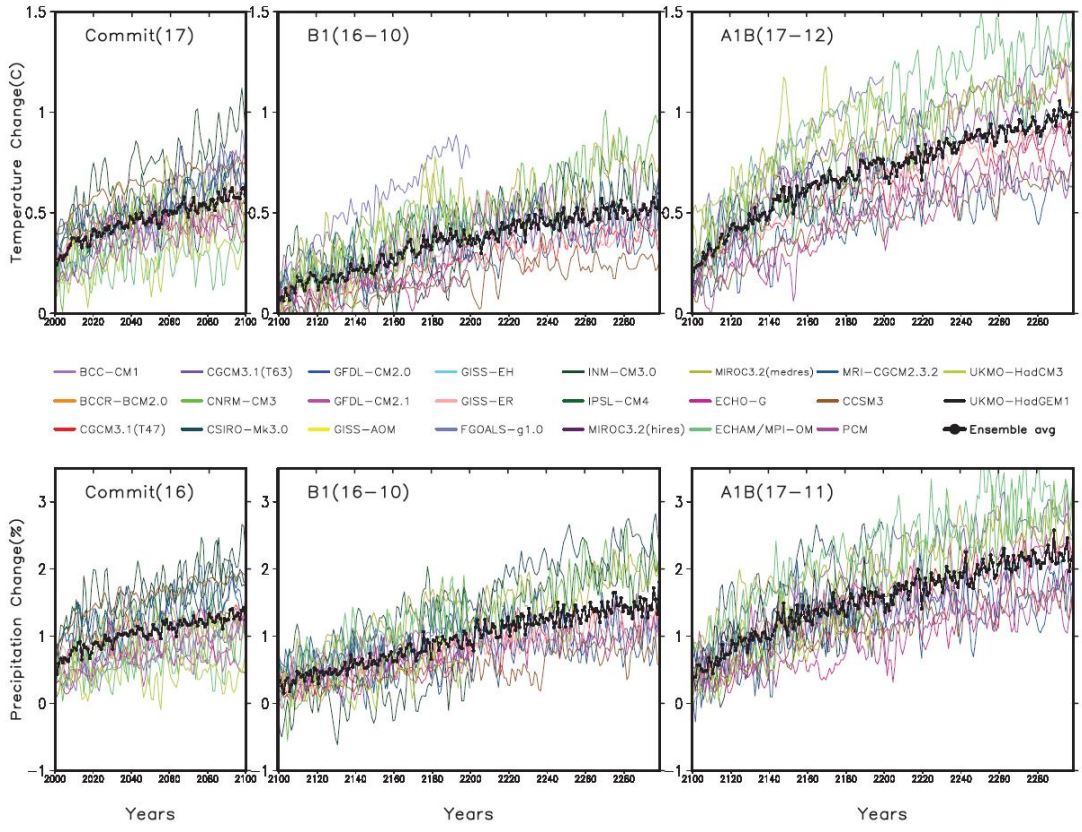


Figure 1.6: Globally averaged surface air temperature (top) and precipitation (bottom) change relative to 1980–1999 for the 20th century commitment experiment (Commit,left), for the B1 commitment experiment computed with respect to the 2080–2099 average (B1, centre) and for the A1B commitment experiment (A1B, right). The numbers in the panels denote the number of models used for each scenario and each century. Intermodel differences give rise to the spread in projections (Source: IPCC, 2007).

Various techniques exist for “nesting” the RCM in the GCM and this choice can also have an effect on the end result. Finally, the choice of which RCM to use is a source of uncertainty, Additionally, if an impacts model is used, a hydrological model, for example to determine the impacts of climate change on a river catchment, then the choice of impacts model can add an extra layer of uncertainty.

While it may be possible to reduce some of the uncertainties associated with climate modelling through further research, others remain an inherent and irreducible part of the process. For example, the behaviour of clouds is parameterized in models because cloud formation occurs on such a small scale, yet the lack of scientific understanding about clouds makes their representation in climate models a key source of uncertainty. Further study of this area may lead to better schemes to represent this variable in climate models and as such, this source of uncertainty has the potential to be reduced. However, further research may also uncover previously unknown climate processes which would then need to be accounted for in climate models, potentially increasing uncertainty. Additionally, other factors like future emissions concentrations are an irreducible source of uncertainty because they are an unknowable variable. Emissions concentrations depend on a range of social, economic and technological factors that are impossible to predict as they are the result of human action. As such, the climate models can only ever capture a range of possible potential futures as outlined by the emissions scenarios.

Uncertainty in climate modelling will be discussed in greater detail in the next chapter, which will detail the different types of uncertainty that affect climate model output, how they arise and how they are commonly managed in the climate modelling community. The specific examples of cloud uncertainties and emissions uncertainties will also be discussed in more detail.

1.5 HUMAN RESPONSES TO CLIMATE CHANGE

The continued use of carbon-based energy sources and carbon-producing industries is inherently unsustainable (Black, 1996; Poliakoff and Licence 2007; Yegulalp *et al.,* 2001). Aside from the environmental concerns, fossil fuels are a finite resource which means that there will come a point at which the rate of

production falls into irreversible decline. Additionally, the economic cost of not taking action to address climate change is now being recognized. The report of Sir Nicholas Stern in 2006 estimated that the cost of inaction would be 5%-20% of global GDP while the cost of early and effective action was estimated to be as little as 1% of GDP by 2050 (Stern, 2006) and although the report is not without its critics, it does highlight the reduced costs associated with early action.

Although much uncertainty surrounds climate model outputs, when this uncertainty is accounted for as fully as possible and communicated effectively, such models have the potential to provide valuable information to help inform these actions. When adaptive measures are put in place without taking into account potential future changes in the climate, there is a risk that the measures put in place will turn out to be insufficient to manage future climate impacts. Climate models are an important source of information about possible future changes and as such, incorporating them into the decision-making process creates the potential for more robust decision-making.

1.5.1 Climate mitigation

Mitigation refers to actions taken to lessen the effects of climate change by tackling its causes, either by reducing greenhouse gas (GHG) emissions or enhancing sinks such as forests. Research is ongoing into methods for enhancing the capacity or efficiency of carbon sinks, for example by varying crop choices or agricultural methods in croplands (Smith, 2004), or by fertilizing oceans with iron to encourage phytoplankton growth (Buesseler, 2004). With research into sink enhancement still in an early stage, policy has focused largely on GHG reduction, a task involving complex economic, political and social factors. The largest share of historical emissions originated in developed countries, yet developing countries are likely to be the most vulnerable as they lack the resources to fund adaptation strategies on the same scale as developed countries. However, if emissions in developing countries are allowed to escalate unchecked to meet their growing economic needs, this could, in effect, cancel out the efforts of those countries that implement policies to reduce GHG emission. Clearly a concerted international effort is required and this has been the approach of the United Nations Framework Convention on Climate Change

(UNFCCC) Kyoto Protocol, which is presently the only binding international agreement on climate change action. Hansen and Sato (2001) make the point that cooperative international action has already been demonstrated to be successful at tackling such environmental challenges, citing the phasing out of chlorofluorocarbons (CFCs) after the Montreal Protocol of 1989 in response to ozone layer depletion as an example. Carbon taxes, carbon offsets and emissions trading are all approaches aimed at reducing carbon emissions.

1.5.2 Climate adaptation

Adaptation refers to actions taken to prepare for the negative effects by minimizing vulnerability (Mitchell and Tanner, 2006), and also to maximize the positive impacts where they exist. For example, with increases in the intensity of extreme storms projected, communities finding themselves in the path of such storms will require superior flood defences and response plans. Low-lying communities face the threat of submersion due to rising sea levels and in the Maldives this threat is perceived to be so severe that the acquisition of new land to relocate to is being explored (Bogardi and Warner, 2008). Yet there may be some benefits to global warming for some sectors, such as the potential to grow new crops (Holden and Brereton, 2003) and maximizing such benefits requires planning also.

There are various types of adaptation, including anticipatory and reactive adaptation, private and public adaptation, and autonomous and planned adaptation (Burton *et al.*, 2001). Mitigation options have been the subject of many more studies than adaptation options (Fankhauser *et al.*, 1999; Dang *et al.*, 2003). As quantification of the anthropogenic influence on climate has improved, it has been recognized that past and present emissions have most likely already committed the Earth to at least some climate change this century. These findings have resulted in adaptation coming to the fore, but the delay has led to significant knowledge gaps. Although more mitigation measures will lead to reduced impacts and therefore reduced risks to attempt to prepare for, the full effect of measures taken now to reduce emissions will only be seen many years into the future. Therefore it is logical to pursue both strategies and also minimize the negative risks of climate change to people and property.

1.5.3 The role of RCMs in responding to climate change

Adaptation and mitigation strategies are likely to be implemented on different spatial scales, and by different people (Tol, 2005). Mitigation actions, such as taxes on less fuel-efficient cars, are for governments to put into practice. Conversely, adaptation actions will vary on a regional scale because the impacts of climate change vary from one location to the next. The needs of a community at risk of summer drought are quite different to those of a community on a low-lying coast. Thus, adaptation strategies should be the concern of local organizations such as county councils. As such, regional climate models are ideally placed to aid in the adaptation decision-making process and are particularly useful in this context.

Scheraga and Grambsch (1998) note that not only are financial resources required to implement a strategy, but also human resources and technological resources. The problem of expense is compounded by the lack of comprehensive estimates of adaptation costs and benefits (Adger *et al.*, 2007). Additionally, a society must be willing to divert the resources required away from other uses. Therefore it is very important that decisions which may involve the construction of costly infrastructure such as coastal defences or reservoirs are based on robust information.

However, if the defences put in place are ineffective they may do more harm than good. The damage to New Orleans as a result of Hurricane Katrina in 2005 highlights what is known as the 'safe development' paradox (Burby, 2006). The presence of flood protection measures led to a perception that areas of the city were safe for habitation when in fact, they were at high risk of flooding and protected by ineffective defences. When a powerful hurricane did hit, these vulnerable areas that would not have been so developed had flood protection measures not been implemented, were densely populated. These events are an example of decision-making which was not robust, as the risk posed by a very powerful hurricane was not fully accounted for in protection measures.

To make responsible choices about the level of defence required, planners at local level need information about likely changes. Previously, water resource managers have looked to past observations to inform decisions, basing design criteria on established flood return periods. However, in a changing climate, flood levels that

were previously reached every 100 years on average could occur every 50 years, or even every 20. With the past no longer the key to the future, other sources of information are required to inform the decision-making process. RCMs have the potential to provide this information, producing data on a variety of scales about how the climate could be affected by different concentrations of GHGs.

Due to the inherent uncertainty associated with modelling future climate, it cannot be assumed a priori that these high resolution scenarios are skilful and with this in mind, model outputs should not be used to identify optimal adaptation measures. However, when uncertainties are accounted for and communicated effectively, models can be quite useful for testing the sensitivity of adaptation measures and informing robust adaptation (Wilby and Dessai, 2010). As such, developing a framework for the generation of future scenarios which accounts for modelling uncertainties is a key research priority.

1.6 AIMS AND OBJECTIVES

One of the issues raised following the IPCC Fourth Assessment Report was the greater need for regional information regarding climate change, to better reflect the diversity of climate issues that concern different geographical areas (IPCC, 2007). Adaptation to climate change is by its nature a local undertaking, and as such, requires climate model output at a smaller scale than GCMs can provide. As such, RCMs are an ideal tool to help inform adaptation decisions or test adaptation strategies to ensure that they are robust to potential climate changes.

However, the climate system is highly complex, and it is impossible to mathematically model it precisely. We are fundamentally limited in our ability to represent the climate system by our level of understanding of factors and processes that influence it. We are also constrained by hardware limitations, as running climate models requires significant computational resources. As uncertainties are so often treated inconsistently in climate change projections, it can be difficult for planners to make the adaptation decisions that need to be made. Better information about the strengths and weaknesses of the various models available is needed in order for planners and decision-makers to determine what steps are necessary and implement

robust adaptation strategies. Communication of scientific uncertainty and the relevance of information obtained from the best available models is also critical.

To determine which models, if any, are best suited for a particular domain, the model must be verified by comparing a hindcast from the model to the observed climate record of the area. If the model can simulate the present climate skilfully, there is more reason to be confident in its future projections. However, verification studies often do not fully examine the complex and non-linear problem of model performance. Many studies do not explore how the RCM arrives at its projections for a key variable such as temperature or precipitation (e.g. Chen *et al.*, 2007), and as such do not distinguish between genuine model skill and skill which comes about as a result of error cancellation in the model. Model skill assessments tend to be applied to a limited selection of spatial and temporal scales. For example, by focusing on mean patterns there is potential to overlook changes in variability and extreme events (Katz and Brown, 1992). It is also possible that the average value for a season can hide information about how the models represent monthly patterns or even components of those patterns. Some verification studies compare model simulations driven by observational data with the observed climate record (e.g. Christensen and Kuhry, 2000), but this approach does not take account that in a future simulation the model is driven by boundary data from a GCM and hence does not assess how the GCM-RCM combination performs. As such, there is a real need for effective and comprehensive approaches to the assessment of uncertainties in climate modelling.

This thesis concerns itself with developing a systematic framework for the construction of robust future scenarios that accounts for intermodel uncertainties at the RCM scale, to aid decision-making and adaptation. The thesis aims to examine the problem by:

- Assessing the models' ability to simulate key aspects of the Irish climate such as means and variability and identifying the spatial and temporal scales at which different models are informative.

- Investigating how the models represent the underlying large-scale dynamics, to determine whether skill in simulating the mean climate state is a robust indicator of model performance.

- Developing procedures for assessing model skill and constructing intelligent ensemble projections and applying these procedures to create robust future climate scenarios for Ireland (Figure 1.7).

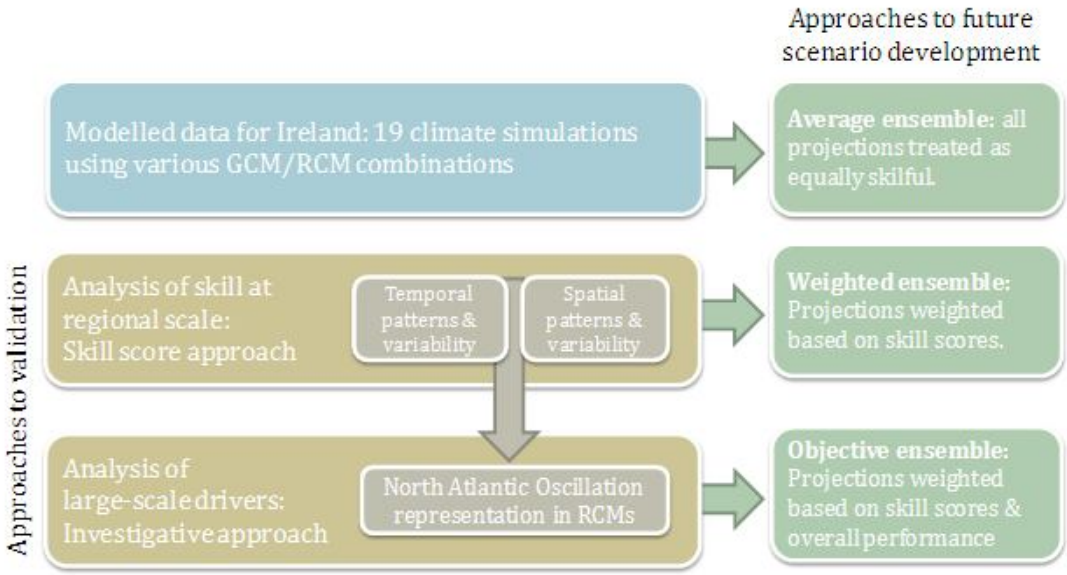


Figure 1.7: Schematic diagram of thesis.

1.7 STUDY REGION: IRELAND

Irish climate is influenced by a variety of factors, both large and small in scale. The warming current of the North Atlantic is a significant feature, as is the influence of the North Atlantic Oscillation index on prevailing westerly winds (Kiely, 1999). In the west, warm, moist winds from the Atlantic ensure that Ireland does not experience extremes of temperature like many other mid-latitude countries.

While temperature is quite homogeneous across the country, precipitation is more spatially variable. The annual precipitation pattern displays a west to east gradient, with the greatest precipitation yields occurring along the mountainous west and south-west coasts (Sweeney, 1985). As the warmest ocean areas lie to the south of Ireland, the greatest precipitation yields are associated with southerly circulations in all seasons (Sweeney, 1989). To the east, the Irish Sea is another key influence, particularly in autumn and winter with easterly winds producing high yields of precipitation along the east and south coast and cyclonic circulations resulting in

heavy precipitation on the Antrim and Down coasts (Sweeney, 1989). Local orographic also plays an important role in shaping weather patterns, with rain shadow effects discernible in certain area.

Differences in population density and geology mean that different areas of the country experience different levels of climate vulnerability. Although Ireland is an island country, it is seen as having low vulnerability to sea level rise due to its predominantly cliffed coasts (Devoy, 2008). However, winter flooding may be an issue under climate change as increased precipitation could alter both flood frequency and duration (Charlton *et al.,* 2006). Agriculture is another area which poses climate risks for Ireland as climate change could lead to losses in the yields of key crops such as barley and potatoes (Holden *et al.,* 2003). Additionally, rising temperatures could potentially enable the emergence of agricultural pests and diseases which cannot survive in the current Irish climate. For example, the recent expansion of bluetongue virus into Northern Europe has been attributed in part to changes in the climate (Gould and Higgs, 2009).

Ireland is a small country, quite poorly resolved in the coarse resolution of GCMs and as such, regional modelling is especially beneficial here. Early RCM simulations of Europe (e.g. Giorgi *et al.,* 1990; Jones *et al.,* 1995) provided valuable insights into the large-scale climate of the continent, but lacked the resolution required to provide a realistic representation of Ireland and the simulation length required to wholly capture interannual variability. However, in the last decade, nested simulations have been produced to cover the European domain (e.g. Christensen *et al,.* 2007; May, 2007; Räisänen *et al.,* 2001) that are not only finer in resolution but also longer in length. Much dynamical modelling of the Irish climate has been carried out by the Community Climate Change Consortium for Ireland (C4I) but this research has tended to utilize a single model, RCA3, leaving intermodel variability an unquantified uncertainty (Dunne *et al.*, 2008: 5).

Such simulations have the potential to provide useful data to inform Irish climate policy. However, an understanding of the uncertainty that affects data and detailed knowledge of the strengths and weaknesses of various models is essential, if model output is to be used appropriately in adaptation planning. Under climate change, the frequency of occurrence associated with the various circulation types or

the amount of precipitation associated with each type may change. As such, it is important that RCMs are able to capture these characteristics of the Irish climate in the control period, if there is to be confidence in the models' future projections.

1.8 STRUCTURE OF THESIS

Chapter 1: Introduction gives a brief overview of the science of climate change, the technique of regional climate modelling and the potential human impacts of climate change, which motivate the development of future climate scenarios.

Chapter 2: Uncertainty in regional climate modelling: A review discusses issue of uncertainty in regional climate modelling in greater detail, including how different types of uncertainty arise and how they impact modelled future scenarios. Approaches for working with climate model uncertainty are also critically assessed.

Chapter 3: Conceptual framework discusses the theoretical framework adopted throughout this research.

Chapter 4: A temporal analysis of regional climate model performance in the present day (1961-1990) presents results of an analysis of RCM skill at simulating temporal aspects of the present climate. For each model, representation of interannual variability and the mean annual climatology of temperature and precipitation are compared to the observed climate.

Chapter 5: A spatial analysis of regional climate model performance in the present day (1961-1990) presents results of an analysis of RCM skill at simulating spatial aspects of the present climate. Representation of the mean seasonal spatial patterns of temperature and precipitation are compared to the observed climate. Empirical Orthogonal Function analysis is used to assess variability of spatial patterns and to determine how skilfully models capture key components of the mean spatial patterns.

Chapter 6: An analysis of the impact of large-scale drivers on modelled climate: North Atlantic Oscillation investigates how a key large-scale driver of Northern European and Irish climate, the North Atlantic Oscillation, is captured in

RCMs, to assess whether skill at modelling means and variability of temperature and precipitation is truly an accurate indicator of model skill. For six case studies, illustrative of the broader range of GCM-RCM combinations, representation of the NAO is assessed through the study of winter mean sea level pressure and wind direction frequencies over Ireland.

Chapter 7: A comparison of approaches to future climate scenario development compares and contrasts approaches to constructing future climate scenarios. The merits and deficiencies of the average ensemble technique and the weighted ensemble based on skill scores are discussed. An approach which utilizes both skill scores and objective estimates of model reliability, based on the research presented in this thesis, is also proposed. These approaches are applied to future Irish temperature and precipitation data and the outcomes are compared and discussed.

Chapter 8: Final conclusions summarizes the key findings of the thesis and highlights areas that could benefit from future study.

CHAPTER 2
UNCERTAINTY IN REGIONAL CLIMATE PROJECTIONS: A REVIEW

2.1 INTRODUCTION

In the event that emissions of greenhouse gases continue to increase, the likely impacts of continued anthropogenic warming could include extinction risks for plant and animal species (Thomas *et al.,* 2004), and direct physical risks to people and communities, as well as economic risks. As such, climate change and climate uncertainty are relevant issues for a range of disciplines including biogeography and ecology (Diniz Filho *et al.,* 2009; Wiens *et al.,* 2009), water resource management (Buytaert *et al.,* 2009; Kay *et al.,* 2009), oceanography (Good *et al.,* 2009) and glaciology (Holland *et al.,* 2010; Vizcaino *et al.,* 2010). With the likely risks of climate change widely recognized, adaptation rather than attribution has become the chief concern. Decision-makers at all levels of governance are beginning to consider how the potential impacts of climate change can be lessened or managed.

While adaptation policy is developed at national level, differences in physical environment, land-use and population make the task of implementing adaptation strategies a task best carried out at regional and local scale. Decision-makers need to determine if adaptive capacities are robust enough to withstand the potential impacts for their region. To do this, planners require information about how human-induced warming may affect key climate parameters such as precipitation and temperature, and what effects such changes will have in their region of interest. Dynamical computer models of climate, particularly regional climate models (RCMs), can provide this information. Yet their limitations must also be understood if their outputs are to be useful in developing meaningful adaptation policy, particularly if such policies are associated with costly infrastructure such as flood defences or reservoir construction, or even relocation of populations.

The climate system is comprised of numerous complex processes and interactions and no model can ever be expected to perfectly simulate this. While many processes are represented in models by fundamental physical equations, parameterizations are also employed to approximate certain processes. The scientific knowledge on which such parameterizations are based comes from studying the current climate and proxy studies of past climate, and as such, their ability to simulate the climate under different forcing conditions may potentially be limited. Such limitations necessitate a greater understanding and awareness of the uncertainty surrounding climate model output. If such projections are to provide an effective basis for policy-making, then uncertainties must be accounted for.

Uncertainties occur at numerous spatial and temporal scales, but are classifiable according to common characteristics. The degree to which they can be successfully represented or quantified is quite variable, but methods do exist for managing certain forms of uncertainty. Nescience, that information that cannot be known, will always be a part of climate projections too, as no methodology can fully account for every uncertainty. But accounting for as much uncertainty as possible is vital if modelled climate projections are to be of benefit in decision-making and policy-making. This chapter characterizes the various forms of uncertainty, discusses how they affect each stage of the regional modelling process and considers previous attempts at dealing with uncertainty in climate modelling.

2.2 DEFINING UNCERTAINTY IN CLIMATE MODELLING

"Uncertainty" and "risk" are often taken as interchangeable concepts, but in the context of climate change assessments these are two important features with fundamental distinctions. Knight (1921) observes that they are two categorically different things. The term "risk" should only refer to measurable uncertainty, while "uncertainty" should be restricted to non-quantitative uncertainty. Hubbard (2007) expands further, illustrating that one can have uncertainty without risk but not risk without uncertainty. He defines uncertainty as "the existence of more than one possibility" (Hubbard, 2007:46), where the true outcome is not known, while risk is a

state of uncertainty where some of the possibilities involve an undesirable outcome. One definition of risk is the probability of an outcome multiplied by the loss associated with that outcome.

Various risks are also associated with climate change. The effects of anthropogenic warming could include extinction risks for plant and animal species (Thomas *et al.*, 2004), and direct physical risks to people and communities, as well as economic risks. The various losses associated with climate risk are not just financial. One cannot put a price on the loss of biodiversity associated with extinction risks, for example. The impacts associated with climate change are dependent on what degree of change emerges. This degree of change is unknown. In a system undergoing change, past observations are unlikely to be a robust estimator of future behaviour. For example, King (2004) notes that under higher emissions concentrations, flood levels that are currently expected once in every 100 years in the UK based on observational records could occur every 3 years. Therefore, long-term projections from climate models are very useful to help determine likely changes on which to base adaptation planning.

However, with less knowledge of possible outcomes, the basis for assigning probability becomes less firm (Figure 2.1). Where outcomes are poorly defined and knowledge about likelihoods is low, alternative approaches such as scenario analysis must be used, as there is no basis for probabilities. As the uncertainty surrounding the modelled output increases, confidence in the data decreases. In order to prepare strategies for managing climate risks, uncertainties must be accounted for as far as possible.

The 'types' of uncertainty commonly identified in the larger scientific community (e.g. Tannert *et al.*, 2007) are often referred to in climate science also. At its core, uncertainty in climate science is a case of "imperfect knowledge" and what Gershon (1998: 44) identifies as "causes of imperfect knowledge" are all present. However, due to the complexity of the climate system and the modelling process, the relationships between uncertainty types must also be considered. A typology of climate model uncertainties is described in Figure 2.2. The first division made is between uncertainty inherent in the climate system and uncertainty related to our

ability to model it, which can be further categorized as epistemological or ontological.

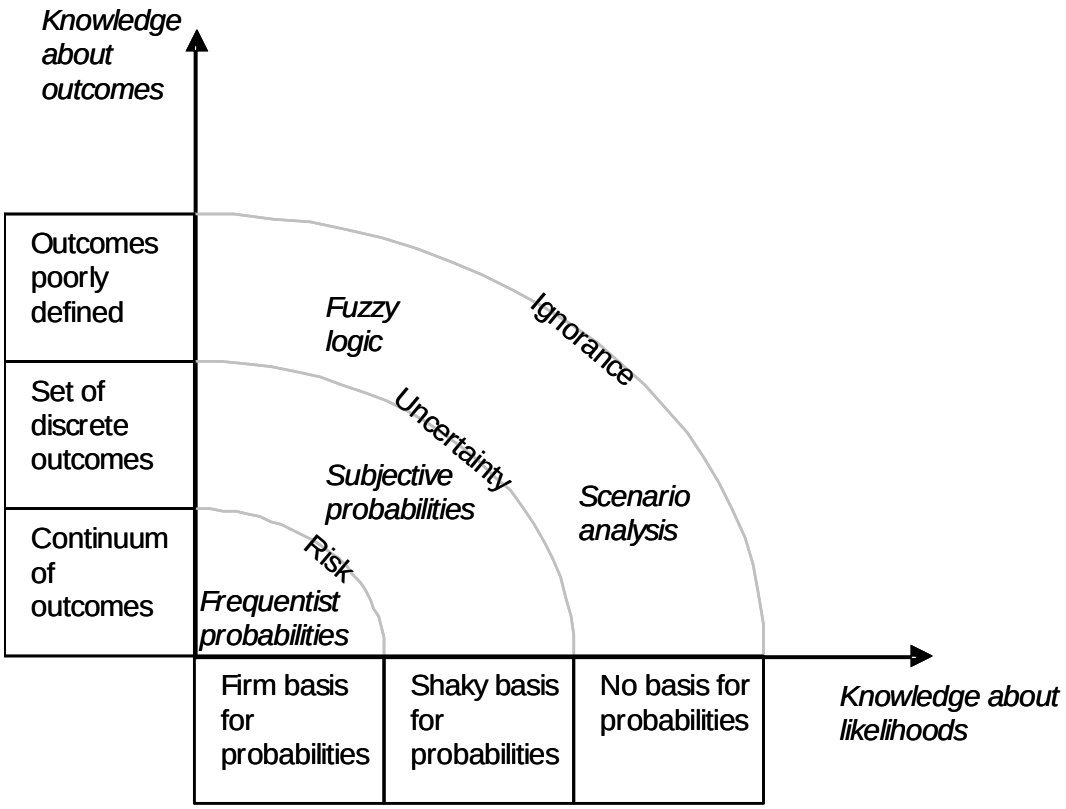


Figure 2.1: Scheme for defining risk, uncertainty and ignorance (Source: Stirling, 1998).

Uncertainty in the climate system has two main sources. Firstly, there is uncertainty over human action, including uncertainty due to unknown future emission concentrations of GHGs and aerosols. Emissions related uncertainties are what Schwierz *et al.* (2006) categorized as Type I uncertainties. This uncertainty is largely due to unknowable knowledge, and is inherently irreducible (Hulme and Carter, 1999). Secondly, there is uncertainty over how the climate system is likely to respond to our actions. Further research may reduce this uncertainty, but may also uncover previously unknown processes and lead to increased uncertainty. Also, in a complex, non-linear system the existence of unknown states or the occurrence of "surprise" events is also possible.

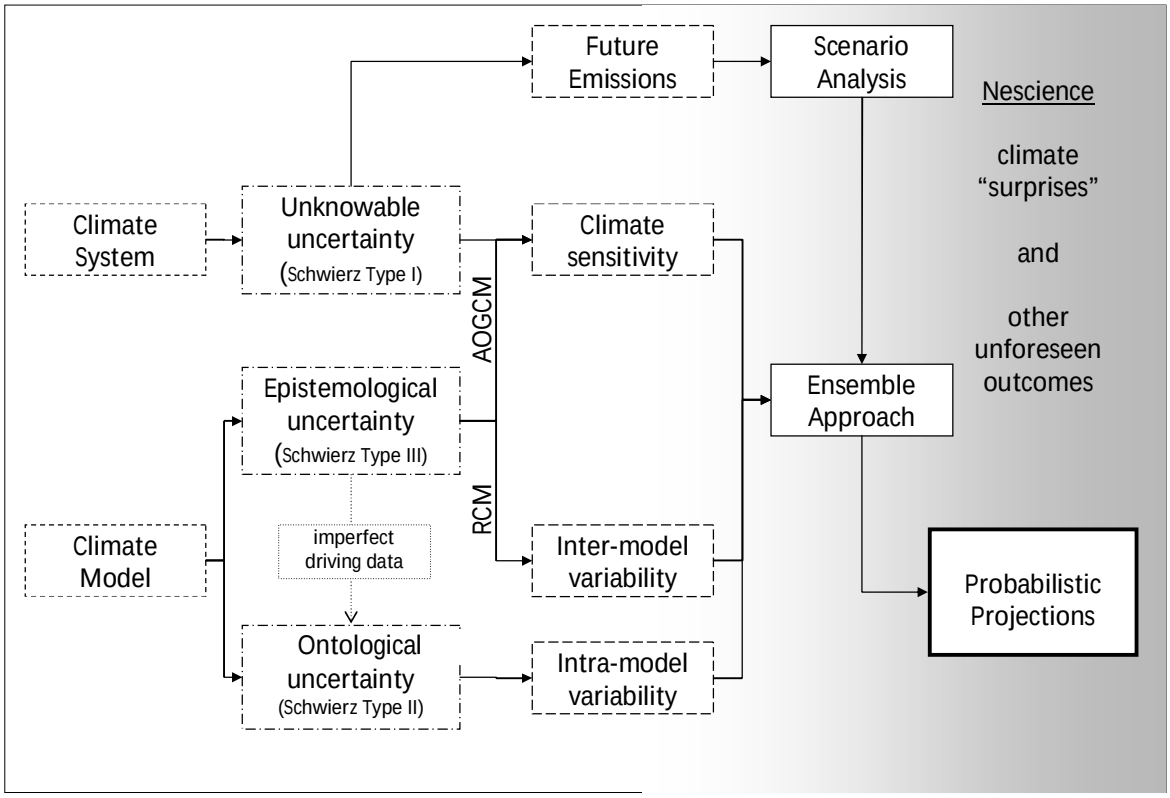


Figure 2.2: A taxonomy of climate model uncertainties

Uncertainty relating to our ability to model the climate system can be refined into two further categories. Epistemological uncertainty is that which is related to gaps in knowledge: what Hulme and Carter (1999) refer to as "incomplete" knowledge. This gives rise to what Schwierz *et al.* (2006) called Type III uncertainties, and Jenkins and Lowe (2003) called science uncertainty. These uncertainties relate to issues with modelling specific processes, and also to the issue of finite computer resources. Many climate processes have been the focus of much research for many years, and as a result they can be represented with physics in models quite well. For example, research into the role of carbon in climate change (e.g. Brown and Lugo, 1982; Maier-Reimer and Hasselmann, 1987; Houghton, 1995) enables us are to model the chemical and physical transfers of carbon from sources to sinks with a degree of confidence. Yet for a variety of reasons, other processes remain quite difficult to model accurately.

Ontological uncertainty, as it relates to climate modelling, involves the variability of the climate system and climate models (Rotmans and van Asselt, 2001; van der Keur *et al.*, 2008), what Tannert *et al.* (2007: 893) describes as "stochastic

features of the situation". The non-deterministic nature of the climate system (Mitchell and Hulme, 1999) gives rise to ontological uncertainty in climate modelling, which is characterized by a lack of predictability. Schwierz *et al.* (2006) refers to these uncertainties as Type II uncertainties. GCMs and RCMs share many of the same uncertainties and are affected to some degree by all types of uncertainty, though different sources emerge as key influencers.

2.3 UNCERTAINTY AND THE CLIMATE SYSTEM

2.3.1 Emissions scenarios

The greatest uncertainty in climate modelling, which features in all climate downscaling techniques, stems from the unpredictability of future anthropogenic greenhouse gas emissions and their resultant atmospheric concentrations. The IPCC Special Report on Emissions Scenarios (SRES) (Nakicenovic *et al.*, 2000) discusses several factors that impact on the atmospheric greenhouse gas concentrations projected over the present century: population growth, economic and social development, the development and utilization of carbon-free energy sources and technology and changes to agricultural practices and land-use. It is not possible to predict how all these influences will evolve as they depend upon future human behaviour. This information is unknowable, and as such is an inherently irreducible uncertainty. Yet the degree of climate change experienced is inextricably linked to concentrations of GHGs. No climate projections can be made without first finding a way to represent this unknowable information. As the outcomes are so poorly defined, there can be no basis for assigning probabilities to future emissions. Alternative approaches are needed to represent this uncertain factor.

A widely used approach to emissions uncertainties is scenario analysis, in which future concentrations are estimated for a range of different "storylines" representing varying combinations of populations and economic development. There are 4 socio-economic storylines for which the IPCC have defined 40 emissions scenarios and each scenario family, A1, A2, B1 and B2, has an illustrative "marker" scenario (Nakicenovic *et al.* 2000). Significant expertise goes into designing these story-lines. For example, numerical modelling may be carried out to ensure self-

consistency in assumptions (Sugiyama, 2005). Yet there has been some criticism of the manner in which they are designed. In particular, economic assumptions that the SRES scenarios make about Gross Domestic Product (GDP) have come under scrutiny (Castles and Henderson, 2003).

The limited number of scenarios also makes it more difficult to draw conclusions about the relationships between global and regional rates of climate change. Pattern-scaling in time can be used to infer climate responses for a particular degree of forcing by scaling the regional climate change signal from a future period according to global mean temperature change. But this requires the assumption that regional change occurs at the same rate as global change. Some papers have attempted to demonstrate such a relationship exists for some climate parameters by examining whether regional response and global change vary together in a model when it is forced by different scenarios (e.g. Hingray *et al*, 2007), but this is made difficult by the limited number of forced model runs available. Often there are only three emissions scenarios available for GCMs (e.g. control, A2, B2). In addition to capturing a wider range of emissions uncertainty, the validity of such relationships could be much better established were more emissions scenarios used in practice.

Emissions scenarios provide information about GHG concentrations for a range of plausible futures and cannot cover all eventualities. Outcomes are left unaccounted for even at this initial stage, introducing uncertainty to the overall projections. Since the future is not static, it is also possible that the actual outcome may be entirely unexpected, a scenario that had never been considered. It is conceivable that the very creation of particular emissions scenarios and the resulting research carried out alters the likelihood of scenarios coming to be, as humanity adopts unforeseen new strategies to avoid a negative scenario becoming reality.

2.3.2 Climate sensitivity

Climate sensitivity is a measure of how responsive the climate system is to a change in forcing. Assume that the climate system undergoes a change in forcing ΔF_{2x}, brought about by a doubling of CO_2 concentration levels. When the climate system reaches its new equilibrium, ΔT_{2x} is the resultant surface temperature

response, averaged globally. The sensitivity of the climate system to this forcing is therefore

Equation 2.1: Climate sensitivity

$$\lambda = \Delta T_{2x}/\Delta F_{2x}.$$

where

λ *=climate sensitivity,*

ΔF_{2x} *= change in forcing brought about by a doubling of* CO_2 *concentration levels and*

ΔT_{2x} *= the resultant equilibrium surface temperature response, averaged globally.*

In this way, the anthropogenic contribution to radiative forcing can be quantified as a figure of global temperature change. The magnitude and impacts of climate change are strongly dependant on climate sensitivity, so there is a real and immediate need to quantify uncertainty associated with sensitivity in climate projections. Andronova and Schlesinger (2001: 1) state:

> "If ΔT_{2x} is less than the lower bound given by the Intergovernmental Panel on Climate Change (IPCC) then AICC (anthropogenic induced climate change) may not be a serious problem for humanity. If ΔT_{2x} is greater than the upper bound given by the IPCC, then AICC may be one of the most severe problems of the 21st century."

Climate sensitivity can be estimated using a perturbed physics ensemble (e.g. Piani *et al.*, 2005) in which the same atmosphere-ocean global climate model (AOGCM) is run numerous times with slightly altered parameters, or using an ensemble of different AOGCMs (e.g. Yokohata *et al.,* 2008). In addition to inheriting the uncertainties of the emissions scenario, differences in the design of AOGCMs, such as the vertical and horizontal resolution of the atmosphere and ocean and the

parameterization of various processes, and uncertainties regarding radiative forcing (Tanaka *et al.,* 2009) introduce further uncertainty into the calculation.

AOGCM experiments provide one measure of sensitivity. Much work has been carried out on 'constraining' estimates of climate sensitivity using 20th century observations (Andronova and Schlesinger, 2001; Knutti *et al.,* 2002). Paleoclimate data has also been used to determine the sensitivity of the climate system to past changes in forcing (Watson, 2008), as past CO_2 levels and surface conditions can be estimated from sources such as ice cores or speleothems. Such research is now being used as a method of validating AOGCMs, the hypothesis being that if an AOGCM's climate sensitivity matches the climate sensitivity obtained from study of paleo data, then greater confidence can be placed in the estimate (Edwards *et al.,* 2007; Hoffert and Covey, 1992).

Of course, as the anthropogenic forcing influencing climate at present is unprecedented, non-linear feedbacks may not operate in the same manner in paleoclimates as they will under doubled CO_2 forcing. Combining constraints from different paleoclimates is likely to be more reliable that looking at single eras (Covey *et al.,* 1996), but using different constraints or combinations thereof yields different values for climate sensitivity, adding an additional layer of uncertainty. Ranges for climate sensitivity vary depending on the method employed (Figure 2.3). For the full range of emissions scenarios, the estimated range of global climate sensitivity is 1.4^{o}C-5.8^{o}C (a normal distribution, with a 5-95% probability range of 2^{o}C-4.5^{o}C, and a most likely value of around 3^{o}C) (IPCC, 2007). Wigley and Raper (2001) take account of other key uncertainties but maintain that all emission scenarios are equally likely, to show that the probabilities of warming are low on both tails of the distribution and in the absence of climate mitigation, the 90% probability of warming is more likely to be in the range 1.7^{o}C-4.9^{o}C.

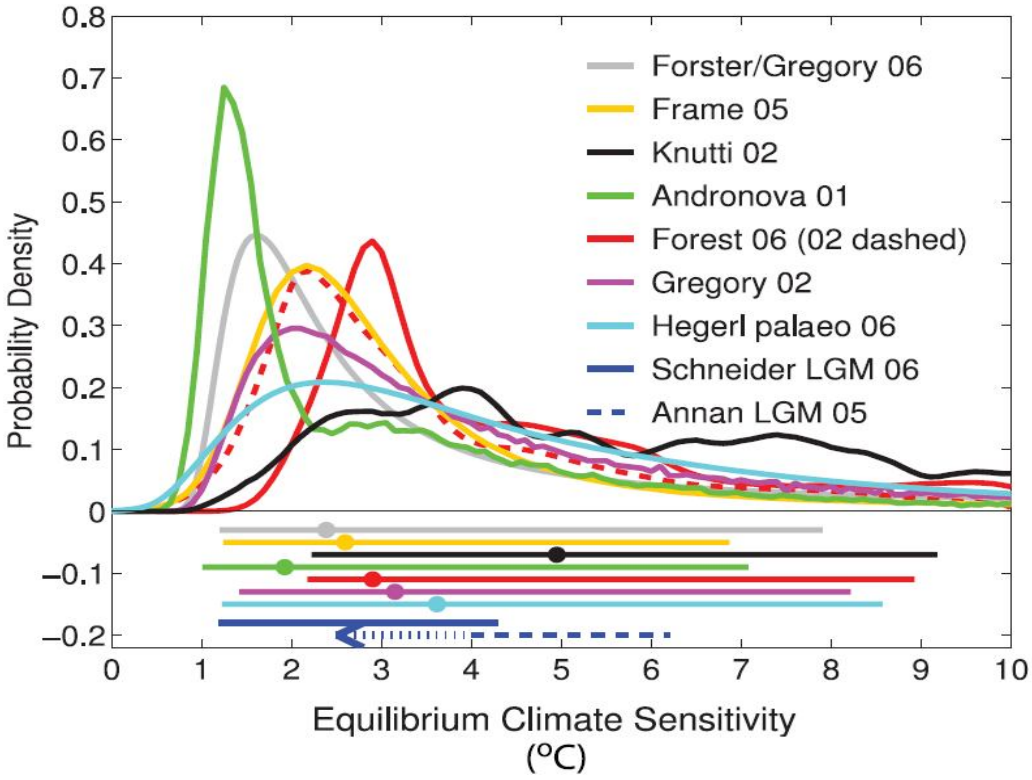


Figure 2.3: A summary of results from climate sensitivity experiments (Source: IPCC, 2007).

2.3.3 Natural variability and climate feedbacks

Even with some idea of how sensitive the climate system is to increased greenhouse gas forcing, there are barriers to understanding how the climate will ultimately respond. The climate system is a complex, non-linear, dynamical system, so understanding the behaviours of various components of the system does not imply understanding of the overall behaviour. As the system evolves it is influenced by natural variations, which are limited in their predictability. For example, the dominant influence on climate in Western Europe, the Atlantic Ocean (Sutton and Hodson, 2005), is affected by modes of variability operating on a range of time-scales from decadal (e.g. Atlantic Multidecadal Oscillation) to thousands of years (e.g. thermohaline circulation). The predictability of these modes has been a topic of study for some time (e.g. Davies *et al.*, 1997; Graham, 1994) and some modes have been shown to be quasi-predictable. For example, Griffies and Bryan (1997) found that the North Atlantic Oscillation may possess predictability in the order of a decade or longer, but not beyond that.

Such variability is naturally forced, as these oscillations of the climate system, which operate on a range of time-scales, are present even in a stable climate not undergoing any anthropogenic forcing. One of the challenges of attributing climate change is that the signals of anthropogenic climate change are superimposed on this background of natural variability, making it difficult to differentiate between the two.

It is also possible that increased GHG emissions may interfere with natural climate modes and processes. There are many ways in which such interference could manifest itself (Figure 2.4), including amplification of the effects of the change in forcing. This is referred to as a positive climate feedback. Negative climate feedback mechanisms also exist which can diminish or mask the effects of a change in forcing.

The Daisyworld scenario (Watson and Lovelock, 1983) illustrates how such feedback mechanisms work. It simulates a world composed of black daisies that absorb light, and white daisies that reflect light. Different types of daisy thrive at different temperatures and at the beginning of the simulation Daisyworld is too cold to support life, but gradually the luminosity of the sun's rays increases and the planet warms. The black daisies amplify the warming further as they absorb light, making the planet suitable for white daisies to grow also. The two types of daisy work together to maintain a surface temperature that is comfortable for both populations but as the planet continues to warm, the temperature becomes too hot for the black daisies to survive. The white daisies begin to replace them, because white daises can stay cooler due to their reflective properties. The cooling effect of the larger white daisy population keeps the surface temperature of Daisyworld habitable as the luminosity of the sun keeps increasing. Eventually, however, the temperatures become intolerable even for the white daisies and the population crashes. At this point in the simulation, solar luminosity starts to decline to a level where white daisies can grow once more. These white daisies further amplify the cooling effect. In the Daisyworld simulation, white daisies act as a negative feedback and cool the planet, while black daisies act as a positive feedback and intensify warming.

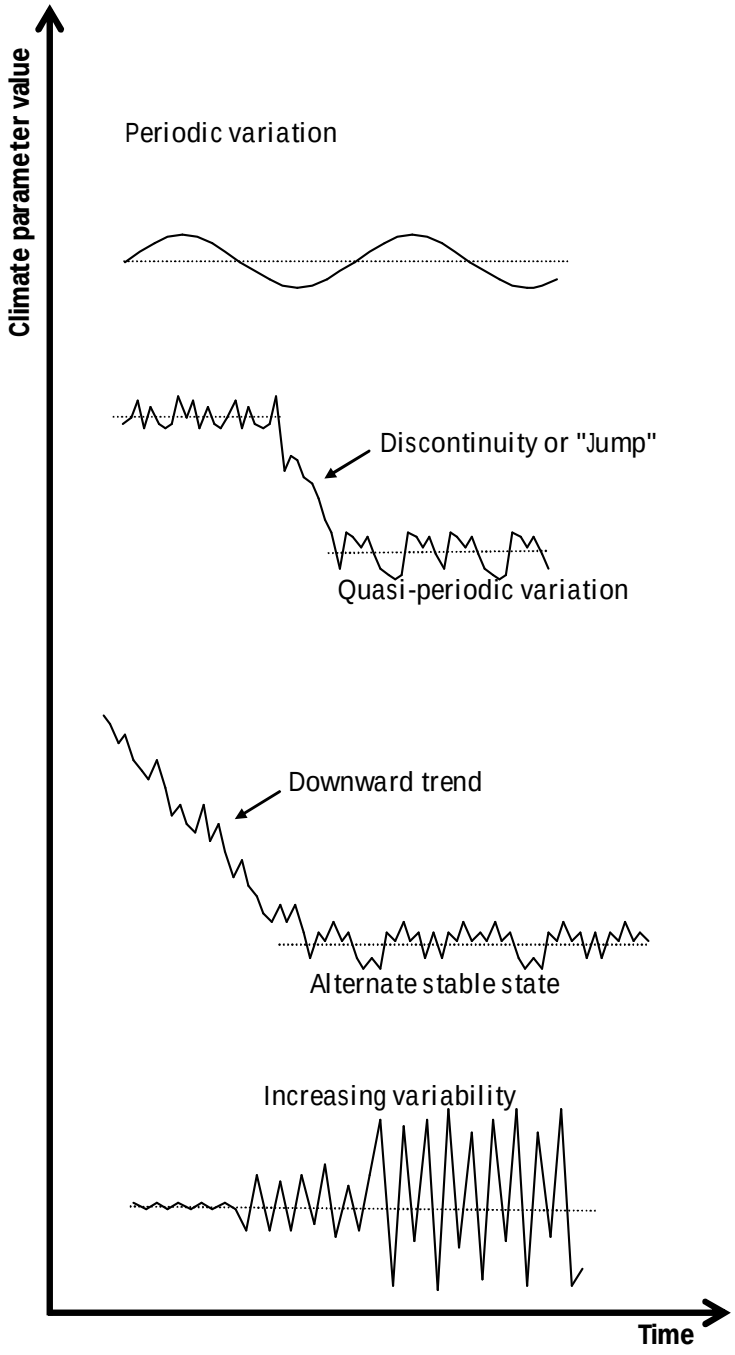


Figure 2.4: Types of climate variations (Source: Marcus and Brazel, 1984).

A similar feedback effect exists in the actual climate system; the sea ice-albedo effect, described by Curry *et al.* (1995). A decrease in snow and ice cover reduces albedo and results in a corresponding increase in surface temperature, which further decreases snow and ice coverage. Another important type of feedback are forest loss feedbacks (Laurence and Williams, 2001). A decrease in forest cover

reduces plant evapotranspirtation and, in turn, regional rainfall. This regional drying leaves the area vulnerable to forest fires and further deforestation. Melting permafrost in the Arctic and Siberia may present a similar feedback, with the release of soil carbon and methane leading to further warming (Anisimov, 2007). Indeed, Kennedy *et al.* (2008) suggest that methane released from permafrost may have been a trigger for deglaciation at the end of the Marinoan 'snowball' ice age (~635M BP).

As in the Daisyworld simulation, there is much debate about the presence of "tipping points" in the climate system at which change due to anthropogenic drivers causes sufficient new processes or change to existing processes to make any human reversal of the overall change impossible (Hansen, 2006). For example, there may be a critical threshold in the climate-carbon cycle system, where regional drying leads to the loss of large tracts of the Amazon Rainforest (Cox *et al.*, 2004). The loss of such a large carbon sink would lead to further warming, and further forest loss. Similarly, global climate model (GCM) simulations show that strong surface freshening in the North Atlantic, which may be brought about by melting glaciers, could force a reduction in the strength of thermohaline circulation (THC). Such a reduction could occur on a time-scale of decades (Hulme and Carter, 1999), or the onset could be even more rapid, taking place over just a few years (Alley *et al.*, 1993). It has been shown that THC resembles a non-linear system in many ways, becoming increasingly sensitive to small perturbations as its critical threshold is neared, and thus less predictable (Knutti and Stocker, 2002). Paleo data suggests that THC reduction, triggered by the sudden release of meltwater from Lake Agassiz (Carlson *et al.*, 2007), may have caused the Younger Dryas cold event (~11,500 BP). In addition to an abrupt climate change, there may actually be a number of stable states, of which we are unaware because they have not been observed before, that the system flips or rapidly changes between (Figure 2.4). Both in reality and in the model, a tipping point could be reached and passed without being immediately obvious. For example, at some level of temperature rise, the melt of the entire Greenland ice sheet would become inevitable as increasing amounts of melt-water further destabilize the ice, yet complete melting would not occur for thousands of years or more.

Outcomes such as the Younger Dryas cold event are often referred to as climate "surprises". Anthropogenic effects on natural climate variation could

manifest in many ways, from a slow shift from a phase of low activity to one of high activity to a sudden jump from one state to another (Figure 2.4). There may even be a number of states that the system changes between. Such jumps are also known as abrupt events or climate surprises. Paleo data and modelling can give an indication of possible outcomes, making such uncertainties ontological as they are due to the non-deterministic nature of the system but are not entirely unknowable. The more model runs considered, the larger the range of potential outcomes that can be simulated. However, models cannot be expected to reveal the full range of potential surprises as even at their most complex, they represent a simplification of the actual system.

It is also possible that future external forcings may also come from unexpected solar variability or volcanic eruptions, which can have a significant impact on the climate system. Whether this would happen, when it would happen and the magnitude of such forcings cannot be predicted with any confidence. Major volcanic eruptions such as El Chichón in 1982 and Mount Pinatubo in 1991, resulted in temperature anomalies of -0.2°C and -0.4°C respectively in the year following the eruption (McCormick *et al.*, 1995). The eruption of Mount Pinatubo in 1991 was the second largest eruption of the 20th century and was rated 6 out of 8 on the Volcanic Explosivity Index (VEI). Although there have been no VEI 8 eruptions during the entire Holocene, if such an eruption were to occur it would have a massive impact on global climate. Simulations show that the impacts of "super-eruptions" could be much greater, potentially reducing global temperatures by up to 10^{o}C (Jones *et al.*, 2005). While this initial effect may last only for a few months, it could take several decades for temperatures to return to normal. Such forcings are unlikely to ever be predictable in a deterministic sense and are thus classed as an unknowable uncertainty. Wigley and Raper (2001) do not take these uncertainties into account as to do so would lead to much wider uncertainty bounds.

The interactive effects of individual climate processes also cannot be anticipated, even with very good knowledge and understanding of the individual mechanisms. Streets and Glantz (2000) refer to these interactive effects as synergisms. In addition to the processes interactions with each other one must consider how they will interact under different forcing conditions. In the future, one may even need to consider how climate mechanisms might react to anthropogenic climate reduction measures, such as geo-engineering.

Yet some form of action is required to prevent the crossing of possible critical thresholds in the climate system that could trigger catastrophic events, and some level of information is needed to develop adaptation strategies. Although climate models can never take account of every uncertainty in the climate system, they remain the best source of information now that past observations of the climate are no longer the key to its future behaviour.

2.4 UNCERTAINTY IN CLIMATE MODELS

Emissions scenarios provide the primary input used to drive a GCM. Due to computational limitations, GCM resolution tends to be quite coarse, in the order of 1.2^{o} to 4^{o} (Genthon *et al.*, 2009). Much computer resources are needed to model the complexities of the global climate system and so running such a model at finer resolution would be quite time-consuming. Various methods can be used to bridge the gap between GCM output and regional response, but the focus of this thesis is regional climate modelling. RCMs have become an increasingly important source of information for decision-makers, providing the necessary, detailed information over a limited area. However, an RCM is but one part of the modelling process. It is part of a chain of procedures in which uncertainties and inferences at each level can impact outcomes at subsequent levels. This chain has been referred to as the "cascade of uncertainty" (Mitchell and Hulme, 1999) or the "uncertainty explosion" (Henderson-Sellers, 1993; Jones, 2000) (Figure 2.5).

However, both GCMs and RCMs are impacted by numerous sources of uncertainty such as knowledge gaps and differences in model codes. These uncertainties weaken confidence in the end projection. Decisions must be made about which GCM to use as a driver and which RCM to use. For every choice the climate modeller makes, there are options he or she did not choose and combinations left unconsidered, and so uncertainty must be recognized as an unavoidable part of climate modelling.

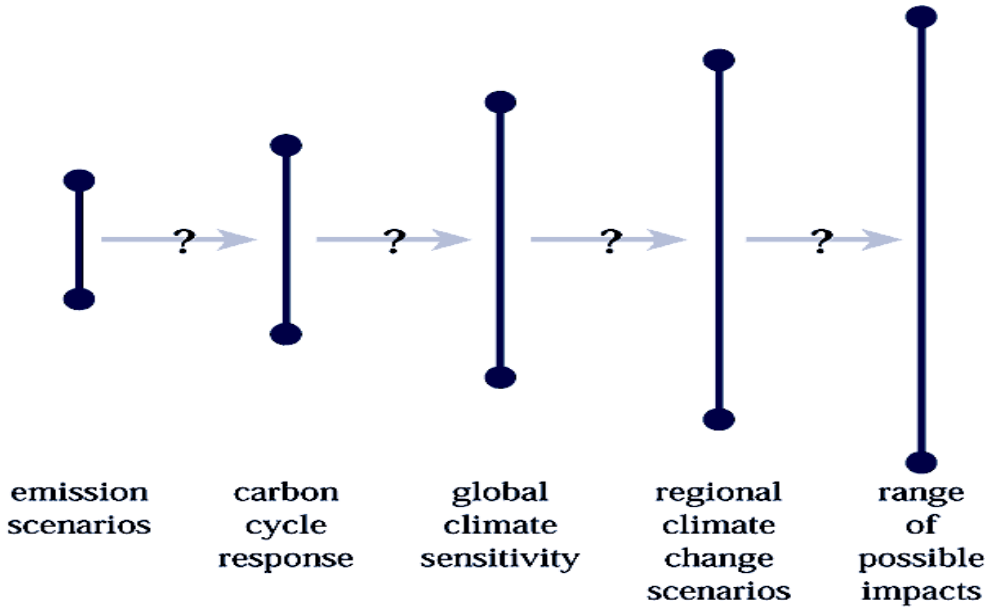


Figure 2.5: "Uncertainty explosion" of major typical uncertainties (Source: after Jones, 2000b and Schneider, 1983).

2.4.1 Epistemological uncertainty in climate modelling

Epistemological uncertainties are very influential in both GCMs and RCMs, and cloud uncertainties are a prime example of this category of uncertainty. Clouds have a variety of effects on both the radiation budget and the water balance, so it is of the utmost important that models reproduce them accurately. According to Schwarz (2008: 439) "…a 10% error in treatment of clouds in the climate model would result an error of some 4.8 W/m^2". Bony *et al.* (2006) compared results from a number of models from different cloud feedback quantification studies and found not only differences in magnitude but also differences in direction for the lapse rate feedback parameter (Figure 2.6). Clouds can have a warming effect by trapping a portion of outgoing infrared radiation and radiating it back downward or can have a cooling effect by reflecting sunlight back into space, an effect known as the cloud albedo effect. The type of cloud determines the effect it will have, with high, thin clouds having a warming effect and low, thick clouds having a cooling effect. There are also a wide range of cloud-climate feedbacks. An increase in cloud amounts is projected as a consequence of anthropogenic warming and the resultant increase in atmospheric water vapour. But the type of cloud likely to result from an increase in water vapour and the overall effect on surface temperatures is not well known.

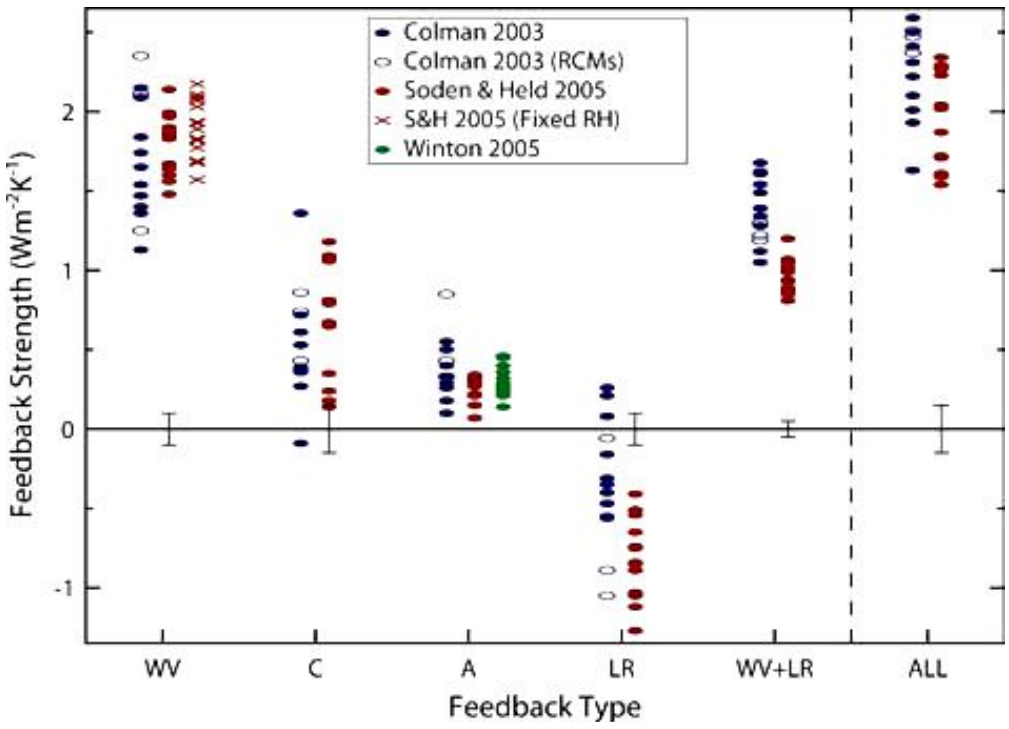


Figure 2.6: Comparison of GCM climate feedback for water vapor (WV), cloud (C), surface albedo (A), lapse rate (LR), and the combined water vapor and lapse rate (WV+ LR). ALL represents the sum of all feedbacks. Results from Colman (2003; in blue), Soden and Held (2006, in red), and Winton (2006, in green). Closed and open symbols from Colman (2003) represent calculations determined using the partial radiative perturbation method and the RCM approaches, respectively. Crosses represent water vapor feedback computed assuming no change in relative humidity. Vertical bars depict the estimated uncertainty in the calculation of the feedbacks from Soden and Held (2006) (Source: Bony et al., 2006).

There are two main reasons for the knowledge gaps surrounding clouds and cloud processes. Firstly, the more accurate satellite observation record is quite short. Conversely, the surface observation record is long but quite subjective as only clouds visible to the observer are recorded (IPCC, 2007). Higher-level clouds hidden above low-level clouds would not be noted, and this makes it very difficult to draw conclusions about cloud processes and what behaviour may be likely under different forcing conditions from the existing observations alone. But epistemological uncertainty may be reduced with further research. Baker and Peter (2008) suggest new observational and laboratory programmes are needed to fill cloud science knowledge gaps.

Secondly, the large-scale effects of clouds on the climate are actually the result of processes occurring on a much smaller scale. Increases in concentrations of anthropogenic aerosols such as sulphate and mineral dust have direct and indirect

effects on clouds by impacting on processes at this microphysical scale (Figure 2.7). Even with highly accurate observations of the large-scale cloud formations, further work would be needed to characterize these small-scale processes. Again, much new research has been carried out to determine how increases in aerosols could modify cloud behaviour (e.g. Berg *et al.*, 2008; Khain *et al.*, 2005; Lohmann, 2008). New data collection methods such as remotely-piloted aircraft (Lu *et al.*, 2008) and model experiments (e.g. Philips *et al.*, 2007) may help to close the knowledge gaps and enable better modelling of the climate system as a whole.

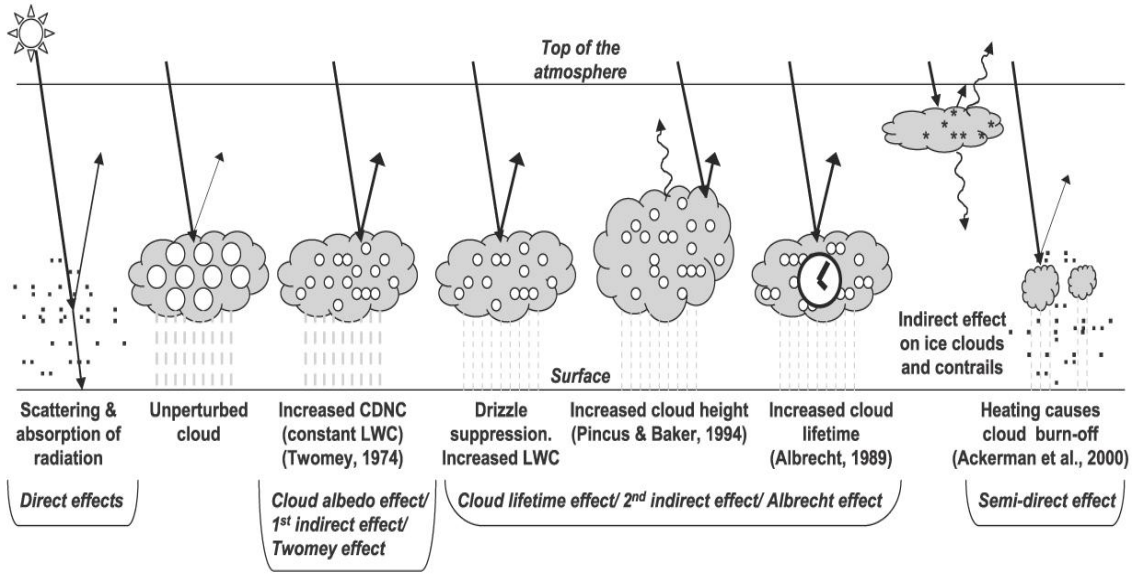


Figure 2.7: Aerosol effects on cloud. Black dots represent aerosol particles, open circles represent water droplets. CDNC = cloud droplet number concentration. LWC =liquid water content.(Source: IPCC, 2007).

However, as knowledge of the system increases, a new problem of complexity emerges. As AOGCMs require considerable computer power to run, and computer resources are not limitless, decisions must be made about how to focus the computing power and which specific processes to represent explicitly (Pope *et al.*, 2007). To maximize one attribute of the model it is necessary to compensate in other areas. Presently, to produce long and highly complex output, a model would need to be run at a coarse resolution. If high resolution output is required, it is sometimes necessary to leave out or empirically approximate processes rather than physically resolve them within the model. Such a situation arises when processes could be

physically represented in the model, but doing so would increase the complexity of the model, which would require compensation with respect to resolution or runtime.

Decisions must be made about what processes are integral, and what processes can be omitted. Of course, leaving out a process can have an effect on the model's performance, as demonstrated by Senior (1998) who found the modelled response of large-scale circulation changes significantly when interactive radiative properties are included in the model. An alternative is to parameterize a process, rather than leave it out completely. Instead of explicitly resolving the process in the model, a scheme is developed to describe the impact of the mechanism on the atmospheric system. This is achieved by formulating the effect of the subgrid-scale processes in terms of resolved grid-scale variables to make an empirical approximation. The advantage to using parameterizations is that running the model requires less computer power, as the physical equations corresponding to the processes do not need to be solved. Cumulus clouds, turbulent mixing, subgrid-scale orographic drag and moist convection are examples of such processes.

A number of issues arise from the use of such schemes. Firstly, parameterization schemes are not equally effective. Convective cloud formation is an example of a process that should not be left out of models. Deep convection generates and redistributes heat and removes and redistributes moisture, significantly affecting the stability of the large scale circulation (Emanuel *et al.*, 1994). But the scale on which convective clouds form is much smaller than that of stratiform clouds, and can be less than a kilometre. Within this very small area, the convective updrafts form narrow thermals, with slowly subsiding air in between (Bjerknes, 1934). To represent this activity physically in a model using the equations of fluid motion would require much finer grid resolution than is currently available, which would in turn require more computing power. So a scheme is created which approximates the collective effects of convective clouds in each model grid cell. In the simplest parameterization schemes, if a column of air in a model grid box is warmer at the bottom than at the top, it is overturned and the air in that column is mixed, but this is clearly not what happens in reality. The more sophisticated schemes attempt to account for more processes, for example entrainment, but they are still imprecise. Wang and Seaman (1997) assessed the performance of four such schemes, and found the skill of the four schemes variable, observed systematic error

with the schemes, and identified particular properties that more skilful schemes had and less skilful ones lacked. There is also always the possibility that more than one theory can be used to explain observations, and following on from this, that more than one parameterization scheme can be developed for a particular process. But while they may all hold true in the current climate, in a changing climate they may not all be so accurate. Empirical approximations cannot be tested under altered forcing conditions, therefore an assumption must be made that the relationships will hold in the future.

The second issue is that, as with empirical downscaling, whether or not the effect of the subgrid-scale processes will be the same under different forcing conditions is impossible to say. Parameterizations are constructed based on our knowledge of the atmospheric system as it currently is, but the processes are not physically represented in the model. Since, our knowledge of the effect of climate change on atmospheric processes is still developing, even if we were to identify an optimum set of parameterizations, the assumptions associated with them may not be valid under climate change. Under uncertain forcing conditions, different parameterizations could yield different outcomes, and in the absence of empirical data to compare them to, they must all be treated as plausible projections of future climate. This issue can be seen as a form of the problem of induction (Frame *et al.*, 2007). In inductive reasoning, a series of observations are made and a claim inferred based on them. Conversely, deductive reasoning relies on logically progressing from general laws and principles to a particular conclusion. Where our understanding of the laws and principles governing an aspect of climate system is poor, we depend on inductive reasoning to understand and represent it. But the observations made in themselves do not establish the validity of inductive reasoning. Observations that inductive reasoning has worked in the past do not imply that it will always work. To state the problem another way, the observation-based knowledge that climate models partially use relies on the uniformity of nature; the concept that the future will resemble the past. The problem is that the future will obviously not resemble the past in all respects, and *a priori* we cannot specify the respects in which the resemblance holds. Keeping in mind these considerations, the task of modelling future climate scenarios at all may at first seem quite fruitless. But barring catastrophic, abrupt events, the workings of the future climate system should resemble those of the past in

a many ways. Fundamentally, models are based on established physical laws, and have proven skill at representing important features in past and present climate, as demonstrated by the climate sensitivity experiments referred to earlier. There is good reason to be confident that models provide plausible estimates of future climate, based on various assumptions, but also much scope to improve upon epistemological uncertainties through further research.

2.4.2 Ontological uncertainty in climate modelling

As the climate system has similarities with a chaotic system, unpredictability arises in two distinct ways. The climate system could first be imagined as an initial value problem. If the system were represented by an evolution equation specifying how, given initial conditions, the system will develop over time, it would be highly sensitive to changes in initial conditions. Similarly, if a chaotic system evolves *n* number of times from slightly different starting conditions, *n* different outcomes can be expected. Although the paths taken may at first be similar, over time errors in the initial conditions amplify and make it impossible to forecast with certainty. For this reason, it is not possible to forecast individual weather events beyond the order of a week to ten days. This problem is referred to as predictability of the first kind.

There is also predictability of the second kind, which is similar to the boundary value problem. That is, a differential equation with an additional set of constraints. In a regional model, a region of interest called the domain is chosen. The domain has a certain boundary with the surrounding environment, and the model has to consider the physical processes in this boundary region in addition to the domain. So the output of the model will clearly be very sensitive to imperfect boundary conditions. Although it has been the focus of much less research than the first kind of predictability, seemingly small perturbations to boundary conditions can also lead to significantly different future behaviour (Chu, 1998; Collins and Allen, 2002).

Weather prediction was identified as an initial and boundary problem early in the 20th century. Bjerknes (1914) recognized that if one could make some simple assumptions, one could arrive at integrable systems of dynamic and thermodynamic equations to represent meteorological phenomena. He also appreciated the need for accurate, reliable information on the state of the system, to use in solving such

systems of equations. Bjerknes (1919) believed that the most important advance in weather forecasting would be the development of a close-knit, well-equipped network of weather stations to provide quality data on temperature, wind strength and direction and rainfall. Although forecasts at the time were of the order of hours, not even days, Bjerknes understood that the forecasts for the afternoon would be far more reliable if the morning's observations on which they were based were accurate. This issue persists today in climate science, but on a different scale.

For hindcasts, RCMs can take these conditions from reanalysis data, which is based on observations. But for future projections, this is not an option. RCMs must depend on a coarser GCM for these important values, which typically include wind components, temperature, water vapour and cloud variables, surface pressure, and chemical tracers (Giorgi, 2006). RCMs take initial and boundary conditions from a parent GCM, a commonly used technique known as nesting (e.g. Antic *et al.*, 2006; Ju *et al.*, 2007; Ding *et al.*, 2006). To further increase accuracy in driving conditions, a double-nesting approach uses global output to drive another model, perhaps an atmosphere-only GCM, over an intermediate domain. The output from that experiment is then used to drive the RCM (e.g. Gao *et al.*, 2006; Im *et al.*, 2006). Two-way nesting is yet another distinct variation on the technique, in which regional scale information from the RCM is allowed to feed back into the GCM (e.g. Barth *et al.*, 2005), and it has been shown to improve GCM representation of the general circulation (Lorenz and Jacob, 2005). As the GCM has its own inherent flaws, the boundary and initial conditions will always be imperfect. Although the imperfections themselves arise through the epistemological uncertainties of the parent GCM, because they detract from the predictability of the system on both counts, they are a source of ontological uncertainty at the RCM level. Closing some of the knowledge gaps at the GCM level would improve accuracy in the initial and boundary conditions, which would help to improve predictability at RCM level. But as discussed already, one cannot continuously increase GCM complexity without diverting resources from resolution or run length.

Additionally, even with "perfect" driving information, the various fluxes of heat, water and momentum need to be in dynamic and thermodynamic equilibrium for the initial conditions to be valid. In other words, it is not enough for the initial climate of the model to resemble the real climate; it also must be stable. Typically,

models are given a “spin-up” period during initialization, during which the faster adjustments (i.e. 50 year timescale) take place and stabilize. But a slower adjustment also takes place, as the deep ocean adjusts to surface heat and water flux imbalances. During initialization, models are allowed to reach a stage where this adjustment, known as “climate drift” (e.g. Bryan, 1998; Dirmeyer, 2000), is so slow as to not interfere with the interpretation of climate change signals too much. But again, the computational demands of modelling make it unrealistic to initialize the model over a timescale so long that the deep ocean adjustments fully stabilize. A flux adjustment may be required to minimize climate drift and prevent the model from sliding into unrealistic climate states. Due to improvements in the simulation of the large-scale heat balances, models have recently been developed which do not require a flux adjustment and instead maintain their own physical consistency (IPCC, 2001). While further research into the behaviour of the climate system clearly has the potential to improve the realism of climate model simulations, it is important to note that the climate system is still far too complex for a climate model to fully represent.

2.4.3 Intermodel variability

Uncertainty makes model design at all levels a subjective process. In addition to the variety of AOGCM drivers that could be chosen, there is a wide range of parameterizations schemes used in regional modelling also. Choices must be made about what to include in a climate model, what to exclude, what to parameterize, and how, what driver to use, what dynamical core to use, and these decisions introduce uncertainties (Tebaldi *et al.*, 2007). As a result, intermodel variability, that is variation in predictions due to the choice or model, is an important issue especially at regional modelling level, where the range of models to choose from is quite large.

The choice of which model or models to utilize is not arbitrary, as it can be based on assessing model skill. But this can never be a truly objective choice. Blyth (1972) makes a distinction between knowledge, defined as beliefs held by the entire scientific field, and subjective beliefs, defined as the personal beliefs of the individuals. A knowledge-guided decision can be made about models, using a measure of model skill acknowledged by the modelling community, but there are many such measures. Although some are used more frequently than others, there is

still no designated index for intercomparison. So the choice must be partially subjective as the decision of how to assess skill is made by the individual and not commonly agreed by the scientific field

Model intercomparison has been by far the most common technique for comprehensively determining the systematic errors in models. Many intercomparisons have been carried out on GCMs (e.g. the Coupled Model Intercomparison Project (CMIP3), the Arctic Ocean Model Intercomparison Project (AOMIP), and the Ocean-Carbon Cycle Model Intercomparison Project (OCMIP)). But it is only in more recent years that intercomparison of regional models has come to the fore, with coordinated projects such as PRUDENCE (Prediction of Regional scenarios and Uncertainties for Defining EuropeaN Climate change risks and Effects) and ENSEMBLES.

Model performance can be interpreted in different ways and quantified using a variety of metrics, using the observed climatic records for comparison. Additionally, there are a variety of skill scores that can applied, including mean square error, Brier score (Stefanova and Krishnamurti, 2002) and ignorance, amongst others. Multiple statistics of climate must be considered to provide a full picture of model skill. Often, the statistical moment relied on for such comparison is the seasonal average of particular climate variables. But this may not always provide a full picture of model skill. A change in the mean can have a disproportionate effect on the extremes of a distribution because other characteristics such as the variance are also altered by the mean change. Therefore, a model which predicts mean seasonal trends accurately may not possess the same skill at modelling extremes (e.g. Hanson *et al.*, 2007).

Aside to the subjectivity of methods that account for model differences, there are a number of philosophical arguments as to whether the results obtained from any of these methods are truly legitimate. The terms “validation”, “verification” and confirmation” are often encountered in climate modelling literature, and all are commonly used refer to the general process of comparing a climate model’s output over a control period to the observed climate record as a means of establishing reliability. But in the philosophical sense, each has a distinct meaning and it is possible for a model to be validated without essentially being verified. Validation

means that a model has met specified performance standards and is therefore suitable for a particular use (Rykiel, 1995), while verification refers to the demonstration of the "truth" of the model as a basis for reliability. However there are fundamental barriers to the validation and verification of computer models of natural systems.

First, is it impossible to demonstrate the truth of any proposition except in a closed system (Oreskes, 1994). A natural system is not closed. It is not isolated from the environment, but can instead be influenced by events outside of the conceptual boundaries imposed on it for the purposes of study. It is also dynamical, with components that change over time. Theoretically, if there are errors in the hindcast, then the future projections will have the same errors. This assumes that the errors are systematic ones, that if a model consistently underestimates a variable by a certain amount, it is possible to correct the results by that amount each time. This is sometimes referred to as tuning the model, and there are different ways this can be carried out. This route also demands the assumption the errors are constant in time and under different forcing conditions, which is a large assumption to make. Moberg and Jones (2004), having carried out such a comparison with the model HadRM3P, do not mention tuning as a next step but acknowledge that due to the presence of errors in the hindcast any future projections made using this model should be interpreted with caution.

Secondly, it has been argued that any technique which uses observations to verify models is misleading (Stainforth *et al.*, 2007), as the model is simulating a state of the system that has not been experienced before. Therefore verification of a model's performance can only ever be partial. To expand on this verification, we could also consider other criteria such as the model's ability to simulate changes in paleoclimates; a model that simulates both the recent and distant past effectively is more credible than a model that has been tested only on a 20^{th} century control run. Models can also be assessed based on how many of the characteristics desirable in a climate model they possess, such as individual treatment of GHGs, high resolution, peer reviewed publications, number of runs completed to capture natural variability (Hulme *et al.*, 2003) but this is less of a quantitative and more of a qualitative analysis.

Thirdly, in theory model intercomparison aims to identify important differences between models and the cause of such differences. However in reality, a deficiency in a model could result from a number of issues. A temperature bias, for example, could arise due to an error in how the model handles cloud cover, or in how the topography is resolved at that resolution. The error could even be the result of a summation of different errors. To definitively locate the source of a particular error, it would be necessary to run the model in question many times, varying a particular parameterization or combination of parameterizations each time while holding everything else constant. However, due to restraints on time and computer resources, such an approach is often not viable.

The delta-change method is an alternative approach, in which the differences between control and future runs for various variables are extracted and applied to an observed present-day climatology, the underlying assumption is that models simulate relative changes better than absolutes (Hay *et al.*, 2000), so the reliability of the method is not affected by the RCM's deficiencies in reproducing the current climate. But this method only accounts for the change in mean, not changes in other characteristics of the distribution. Theoretically, one could extract difference in other characteristics like variance and range of extremes to construct a fuller picture of the distribution, and adjust the observed present data based on the modelled change. However, an unstated assumption of the method is that the relative change is the climate change signal, that errors remain constant in time and are accounted for in the differencing procedure and so are not a part of the relative change. The main argument against tuning is that model biases may not be consistent over time and this argument remains valid in this context.

Even if model biases and errors cannot be comprehensively accounted for, knowing they are present is valuable information in itself. The propensity for errors could serve as a qualitative measure of how reliable a model is. However, agreement between model output and observed climate does not signify that the model is an accurate representation of the real system, and this must be acknowledged. But the model should reflect the behaviour of the real system if it is to be suitable for contributing to scenario development.

2.5. WORKING WITH UNCERTAINTY: ENSEMBLES AND PROBABILITIES

A model can have skill at modelling one climate pattern and lack skill at modelling another. The model that simulates average seasonal trends accurately may not give a true picture of future changes in extreme events, which due to their sudden nature can cause much greater damage over a short space of time compared to a gradual change. Results that vary depending on choice of model are not very reliable, and decisions need to be based on robust findings. For one particular variable or location, a single best model may perform well, but when considering all aspects of climate and uncertainty, a combination of several different models, known as an ensemble, provides better overall skill and as a result, higher reliability (Tebaldi and Knutti, 2007). Ensemble techniques are in widespread use in the climate modelling community and have been used to characterize the spread of climate responses for a range of variables, impacts and regions.

2.5.1 Multi-model ensembles

One method of producing an ensemble is to combine multiple predictions from different models. This is called a multi-model ensemble. Ideally, individual ensemble members should all possess high skill by themselves and be independent of one another. However, such ensembles are sometimes known as "ensembles of opportunity" (Stone *et al.*, 2007), as members are sometimes chosen more for their availability than their demonstrated skill, a tactic which of course has the potential to generate misleading output (Allen and Stainforth, 2002). The ensemble should have an outcome distribution similar to the natural distribution if it is to be reliable. Multi-model ensembles allow a range of different models to contribute to the overall projection so that intermodel variability is fully sampled and represented in the spread of the projections. It is also a logical approach to take in order to account for intra-model variability, as it allows a more complete range of possible future climate scenarios to be sampled.

The precise reason why an ensemble so often performs better than the individual "best" model is debatable. Doblas-Reyes *et al.* (2000) attributes the

improvement to the use of different models and the increased ensemble size, while Hagedorn *et al.* (2005) states that a large part of the ensemble's superiority is due to error cancellation, and argues that if a model existed that performed poorly in every measure, it could only add skill to an ensemble in this way. Conversely, Weigel *et al.* (2008) have argued that even a poor model can add skill, but only if the model's poor performance is due to over-confidence and not low potential predictability. If the ensemble members already have the correct spread and central value, the ensemble technique will do little to improve performance. The conclusions of Weigel *et al.* (2008) were drawn from experiments with a simple model, and the link between over-confidence and the success of the ensemble technique was verified using real model data. It seems that though these studies look at the question from different perspectives, they have come to a similar conclusion: there is nothing to be gained by including models that are fundamentally flawed in their performance. If a poor model is taken to mean an overconfident one, then this model can be compensated for using the ensemble. But if we take poor to mean a model that misrepresents the climate system, then only revisiting the mechanics of the model and looking for ways to improve its parameterizations can truly improve such a model.

2.5.2 Perturbed physics ensembles

An ensemble may also consist of different runs of the same model (Barnett *et al.*, 2006), each with perturbed versions of the original model physics. In theory, by varying the physical parameters of the model, uncertainties due to parameterization choice are represented in the spread of the output. The key advantage is that the sampling of uncertainty is more systematic than it would be in a multi-model ensemble whose members are chosen on an opportunity basis (Murphy *et al.*, 2007). One can choose a single skilful model and run many iterations rather than many models of varying skill. Of course, this approach requires a subjective decision to be made about which single model to use, and the most skilful model in the present may not remain skilful under future forcing conditions. While a perturbed physics approach is highly useful for quantifying variability within the model, it cannot characterize intermodel variability like a multi-model ensemble can.

The optimal approach to characterizing both internal model variations and intermodel variability would be to use a multi-model perturbed-physics ensemble. The traditional multi-model ensemble is formed by combining output from single iterations of many different models to construct a distribution of climate parameters. Combining perturbed physics distributions from individual models rather than single outputs would give a fuller sample of uncertainties, an approach similar to that of Christensen *et al.* (2001), which used two 8-member ensembles from different RCMs.

A larger ensemble will naturally capture a greater proportion of uncertainty. The distributed computing project climateprediction.net has been used to create multi-thousand member ensembles of GCM experiments (e.g. Piani *et al.*, 2005; Sanderson *et al.*, 2008). However to date, RCM perturbed physics ensembles have been much smaller in size, for example the 8-member ensemble of Lynn *et al.* (2009), the 10-member 10 year ensembles of Lucas-Picher *et al.* (2008) or the 25-member ensembles of Yang and Arritt (2002). Due to the time and computer resource constraints associated with regional modelling and the limitations of current computing standards, it is not feasible to produce RCM ensembles of similar size to the current suite of GCM ensembles. While the perturbed physics technique has great potential in regional climate modelling, the multi-thousand ensemble is currently more suitable for GCM use than RCM. Hawkins and Sutton (2009) note the importance of targeting investments in climate science on the areas with the greatest potential for reducing uncertainty and indeed it may be worth focusing on the problem of computer power. Better resources would enable more complex models to be run as well as larger ensembles.

2.5.3 Ensemble theory

For ensemble scenarios to be considered reliable, it is important that the performance of the individual members are assessed carefully. It is also essential that the methods used to generate such ensembles are valid (Leung *et al.*, 2003). As with RCM development, there is a level of subjectivity in ensemble construction. To formulate robust climate scenarios, assumptions need to be justified.

A key question any climate modeller must answer is whether to use information about a model's performance in the present to constrain the influence of its future output on the overall ensemble. One can consider all outcomes equally likely and blindly average the ensemble members' projections, or assign weights to models based on a measurable performance criterion. The Reliability Ensemble Averaging (REA) (Giorgi and Mearns, 2003) approach is one such quantitative approach, which assigns a weighting function to each model in an ensemble based on their performance at simulating the present climate, and their convergence. Essentially it defined a model as reliable if both its present-day bias and its distance from the simulated ensemble mean are within the range of natural variability. As bias or distance grows, the model is assumed to become less reliable. Giorgi and Mearns (2003) applied the method to a set of GCM experiments. All models in the ensemble contributed at least one maximum positive or negative regional present-day bias. Yet skilful performance in the present does not necessarily equate to a good performance in the future. It is impossible to state with certainty how a model will perform at representing climates under unprecedented forcing conditions. However, it is hard to see how a model lacking skill at representing the current climate would have better skill at modelling a future climate. Therefore, while there is an argument to be made for constraining poorly performing models based on their present-day skill, one must be careful not to mistake present-day skill for a guarantee of future skill.

Model convergence is the second criterion used in the REA method: the further a model's result is from the ensemble mean, the less reliable it is taken to be. But while present-day performance at least can be measured against empirical data, there are real issues regarding whether assessing models based on convergence can be considered a valid reliability criterion. Convergence does not immediately imply correctness, and if a model diverges from the values projected by other models, this does not mean it is wrong. In an ideal world, all ensemble members would be independent, but in reality there may be underlying similarities that lead a group of models to converge, such as sharing the same GCM driver or dynamical core, or having a key parameterization scheme in common. Alternatively, the absence or inclusion of certain parameterizations may be key. Rockel and Woth (2007) studied changes in wind speed over Europe using an ensemble of RCMs, and discovered that the absence of a gust parameterization leads to much poorer simulation of high wind

speeds or “storm peaks”. This may even contribute to the lack of agreement between models about changes in the future behaviour of mid-latitude storms, as reported by Meehl *et al.* (2000). Additionally, as the region to which a model is applied has an influence on how it performs (Haylock *et al.*, 2006; Hellstrom, 2001; Jacob *et al.*, 2007), a model can be an outlier in one region but not in another. Without empirical data to compare future projections to, it would be unwise to discount a model just because other models disagree with it. It could be very skilful, and the convergence of its peers traceable to one of the aforementioned underlying factors. The reliability of the model convergence criterion depends on the independence of the models in the ensemble, which is often difficult to establish.

As our understanding of the climate system and the climate models we design based on this understanding are incomplete, we must assume that all models provide credible future scenarios even though they differ in their design and outcomes, unless a clear and justifiable reason to omit a particular model is found. It is better to exercise caution and work with a large range that is more likely to contain the true outcome than to be overconfident and work with a smaller range that does not contain it at all. The range of outcomes supplied by climate models becomes part of a chain of inferences; regional effects are inferred from global effects which are in turn used to infer and prioritize adaptive decisions. In the words of Frame *et al.* (2007: 1986) we “*run the risk of building inferential edifices on unstable foundations,*” a situation best avoided where costly investment decisions must be made.

2.5.4 Ensembles with probability

Approaches like the REA technique are quantitative but not probabilistic. An advantage of such a technique is that one avoids making assumptions about distributions, which is required for a probabilistic approach. But probabilities are very useful in climate science. The IPCC Third Assessment Report (2001) assigned descriptive terminology to probability ranges. For example, a probability range of 10-33% was described as unlikely, while a probability range of 1-10% was described as very unlikely. Patt and Dessai (2005) investigated how people link descriptive phrases with probability ranges and found that they use intuitive heuristics rather than formal definitions. Given the same descriptive terms to describe a high

magnitude event and a low magnitude event, people interpret the language to mean the high magnitude event is less likely, leading them to actually underestimate the damage that could be expected and under-respond to the threat of the high magnitude event. A quantitative approach without probabilistic interpretation is open to subjective and possibly biased interpretation, which could result in decision-makers under-responding to climate change. In most cases, decision-makers and planners will be better served by a probability distribution of possible changes as opposed to a selection of possible scenarios. The potential for bias can be lessened by utilizing both numerical probability ranges and probability language and the recent availability of ensembles of data from multiple modelling centres makes it feasible to attach probability to scenarios.

Probabilistic methodologies have a history of use within short and medium range weather forecasting, where they are recognized as being more reliable than single deterministic forecasts. Their application to climate projections is a logical step. Räisänen & Palmer (2001) demonstrate how a GCM ensemble can be treated as a probabilistic forecast, with intermodel uncertainty characterized by the ensemble dispersion. Furthering this methodology, one can utilize probability distribution functions (PDFs) or cumulative distribution functions (CDFs) as a technique for quantifying uncertainties in RCM output as well as GCM.

The probabilities used by climate change researchers are not classical frequentist probabilities. They would be better defined as Bayesian probabilities (Dessai and Hulme, 2004). Bayesian probability is very applicable to climate change simulations as it assigns probability to propositions that are uncertain. A prior distribution is specified for the uncertain quantities of interest, which is independent of any data available for them. The prior distribution can also be expert-based. An observed distribution is then ascribed to modelled data. The likelihood of the modelled distribution as a function of parameter values is calculated, and this likelihood function is multiplied by the prior distribution. When normalized, this provides the posterior distribution, which is a distribution of unit probability over all possible values. An empirical estimate of the posterior distribution can be obtained through sampling with the Markov Chain Monte Carlo (MCMC) simulation method. The mode of the distribution is then the parameter estimate and "probability intervals" (the Bayesian analogue of confidence intervals) can be calculated.

This methodology interprets probability as a measure of a state of knowledge. But the "state of knowledge" can be subjective. For example, Bayesian statistics could be used to make a quantitative determination of climate change impacts, but it would be based on a prior assessment of the probability of anthropogenically induced climate change. This assessment would have to be subjective, and the use of different equally plausible priors would yield different priors (Barnett *et al.*, 1999). But as Berliner *et al.* (2000) asserts, Bayesian statistics acknowledges that it is imperfect by stating the assumptions and quantifying them so that the sensitivity of the results can also be assessed. Objective Bayesian probability also exists (Berger *et al.*, 2001), which utilizes a non-informative, non-subjective prior. But this can lead to paradoxes as outlined by Krieglar (2005) who notes that if you have assumed complete ignorance regarding future atmospheric CO_2 concentration, you cannot also make this assumption for the associated radiative forcing as it is logarithmically dependant. Taking a strictly objective view can also lead to the exclusion of qualitative information which has the potential to be very valuable.

Different researchers have adopted variations of the methodology, some more objective and some more subjective. An objective approach was used by Jones (2000a), which relied on properties of classic probability distributions. If the uncertainties associated with various sources are taken to be uniform and independent, then when multiplied together they will yield a peaked probability distribution. In this way, PDFs are created for key climatic variables relating to irrigation supply in Victoria, Australia. In practice, it is common to assume a uniform distribution over the appropriate range of values for the prior distribution. Jones (2000a) is particularly objective, however, in that the absence of assumptions extends to the posterior distribution. Jones' results suggest that some adaptation will be required in the area by 2030, with a theoretical critical threshold existing at around 2050. A similar approach was applied by Fealy (2010) to the Irish domain. These conclusions, with probabilities attached, are the kind of information that decision makers can begin making adaptation decisions with.

Tebaldi *et al.* (2005) proposed a Bayesian analysis approach to determining probability density functions of temperature change, which would formalize the performance and convergence criterion that the REA method first quantified. Similar to Jones (2000a), uniform, uninformative prior distributions are adopted, to avoid

making assumptions about the prior distributions that could be construed as subjective. But in the Bayesian methodology described by Tebaldi *et al.* (2005), the criteria of performance and convergence effect the posterior distribution. Performance is formalized in the likelihood function as the distance between the i^{th} simulation of a parameter and the best approximation to the truth. Convergence is formalized as the distance of the i^{th} future projection from the consensus estimate of the ensemble. As discussed, it is highly debatable whether convergence is a valid criterion to use in assessing models. Tebaldi *et al.* (2004) proposed a variant of the methodology in which convergence could be weighted differently relative to performance. There is a multiplicative factor in the likelihood function that represents a model's performance in the future, and by constraining this, it become possible to model a larger variance for future projections than present. The Bayesian method was applied to the same ensemble of GCMs as the REA method, and used to assess temperature change due to climate change in a number of case study areas. Northern Europe in both summer and winter displayed a wide range of uncertainty, due to lack of agreement between the GCMs. Other case study areas displayed tighter distributions, signalling greater certainty. This method has also been applied to simulations of regional precipitation change (Tebaldi *et al.*, 2004) using an ensemble of nine GCMs. It should be noted that in Tebaldi *et al.* (2004) "regional" has been used to describe areas of sub-continental scale. For the Northern Europe region, it was found that precipitation changes in summer were relatively small. But summer climate in Europe is quite dependant on small-scale processes that GCMs are unable to resolve (Vidale *et al.*, 2007), which may be a reason for this result.

Both the objective and subjective methodologies have their own merits. If the avoidance of assumptions is paramount, then the objective method would be the appropriate choice. For some research, this is extremely important as it is perceived that subjective choice introduces further uncertainty to the problem. Conversely, there is an argument that by treating model outcomes as equally likely, even when the evidence from control runs suggests differences in skill levels, an important opportunity for quantifying uncertainty for the benefit of the end-user has been missed. Ultimately, as the choice between objective and subjective probabilities introduces its own layer to the cascade of uncertainty. One of the challenges of

developing future climate scenarios is determining whether a probabilistic approach is merited and if so, what technique is most suitable.

2.6 CONCLUSIONS

As Collins (2007: 1958) states:

> "the very fact that a team of people can produce a simulation that bears a passing resemblance to the world we live in is, in retrospect, a significant feat".

Yet a simulation can never capture the complexities of the real system. Any numerical model is limited by the knowledge the scientist has about the real system, and the computing resources available to run it. As a result, uncertainty is unavoidable in regional climate scenarios and indeed in any geographical discipline which utilizes numerical modelling.

As adaptation strategies may require costly infrastructure it may at first seem unwise to use RCM output to inform such decisions. Strategic decisions may be flawed if decision-makers assume risks are well-characterized when they are not. However, the cost of inaction is likely to be far greater than the cost of early, adaptive measures (Stern, 2006). If climate sensitivity is at the upper end of the range specified by the IPCC, steps towards adaptation must be taken to reduce the risks to people, infrastructure and the natural environment.

The uncertainties in regional climate model output must be identified and acknowledged for the information to be put to best use using approaches appropriate to the deep uncertainty of the situation (Lempert *et al.*, 2004). By working with a range of models decision-makers can build strategies that cater for a range of plausible futures. Rather than looking for an optimum strategy which depends upon precise projections, decision-makers can build robust strategies that are open to critique and revision (Baer and Risbey, 2009) and will be beneficial under a range of different conditions (Popper *et al.*, 2005).

Uncertainty in regional climate model output cannot be eliminated. What is more, the growing and present concern of climate change means that we cannot wait until the tools are perfected before making decisions about adaptation. Fortunately, uncertainty in RCMs can be minimized, quantified and communicated effectively, and in spite of their uncertainties, regional climate models can provide valuable information for the robust decision-making process. In the next chapter, approaches to climate adaptation and the role of climate models in this process are discussed in greater detail. Key concepts from which the work in this thesis emerges will be discussed, with particular reference to the theories of knowledge that underpin these concepts and a conceptual framework for the optimal use of climate models is formed.

CHAPTER 3
CONCEPTUAL FRAMEWORK

3.1. MOTIVATION: THE ROLE OF RCMS IN CLIMATE PLANNING

There are two major schools of thought on climate adaptation planning. The top-down approach uses modelled scenarios of future climate to gauge impacts and determine what level of adaptation is required while the bottom-up approach focuses on assessing the vulnerability and adaptive capacity of the community (Figure 3.1).

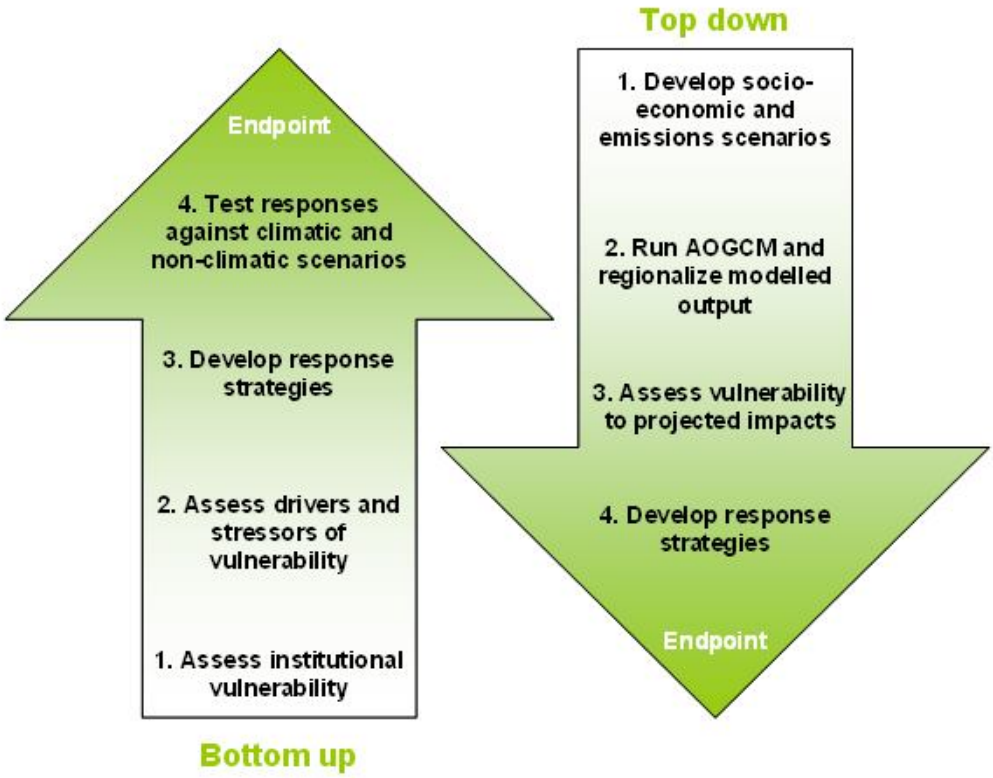


Figure 3.1: The bottom-up and top-down approaches to climate adaptation planning. While top-down planning relies heavily on robust model outcomes, bottom-up planning uses vulnerability to present climate extremes to gauge potential impacts of climate change.(Source: after Dessai et al., 2005).

This approach takes the view that as climate models and future climate scenarios are subject to much uncertainty, present or recent historic climate variability is a better proxy for near-future climate change. Such an approach focuses more on the stakeholders and at-risk population and what adaptation measures will mean for them than the top-down approach (Burton, 2002). The advantage of such an approach is that the process of adaptation is explored and factors that might constrain or render unviable certain adaptive measures are recognized and accounted for in the planning process (Smit and Wandel, 2006). Approaches like this that emphasize the societal concerns associated with the scientific question of climate change are grounded in post-positivistic research philosophy, which explores research questions through the subjective interpretations of individuals (Dyer *et al.*, 2003). Discussing the qualitative techniques often applied in post-positivist research, Kvale (1996: 239) states that:

> "Truth is constructed through a dialogue; valid knowledge claims emerge as conflicting interpretations and action possibilities are discussed and negotiated among the members of a community."

Traditionally, the top-down approach to planning, which strongly emphasizes the need for accurate climate model projections, has not highlighted the human element of adaptation. After all, adaptive recommendations are of little use if there are barriers to their implementation in a community. However, while past climate can be a important guide to understanding future changes, it is not necessarily a robust predictor of these changes and as such, may not be the best source of information on which to base adaptation decisions (Dessai and Hulme, 2004). Climate models provide valuable information about future climates and when the uncertainty surrounding their outputs is communicated correctly they can be very informative in the adaptation planning process.

3.2 DETERMINISTIC APPROACHES TO CLIMATE MODELLING

Early work in the field of climate modelling employed deterministic, single-trajectory methodologies. This approach utilizes a single RCM, providing one projection of future climate which heavily underestimates uncertainty. For example, Fried *et al.* (2004) uses GCM output to investigate changes in the behaviour of wildfires in California under climate change and notes that the results of the wildfire impacts model are sensitive to GCM choice. The model chosen is an intermediary model that lies between those producing the greatest and least change in wildfire behaviour, but this approach leaves much potential for either over- or under-adaptation. Other examples of climate impacts studies which utilize a single climate model include extreme precipitation (Jones and Reid, 2001), food impacts (Parry *et al.*, 2004) and health impacts (Tanser *et al.*, 2003).

This approach is based on the assumption that a "best" model is identifiable. However, there are a number of issues with this assumption. The deterministic approach relies on developing a single model that captures reality as accurately as possible, but there are myriad obstacles in the way of this goal. This approach is rooted heavily in positivist research philosophy, which focuses solely on the data of experience, for example, observations and experimental results. However, all such knowledge could, in principle, be mistaken. All empirical or evidence-based knowledge may have to be revised or rethought if further observations reveal previously unknown information. Knowledge is fallible and as such, certainty is impossible. For example, in a system undergoing change there are many phenomena that cannot be verified by experience. As such, the model that best approximates observed climate is not guaranteed to perform with the same skill under different forcing conditions.

Irwin (2010) describes the gaps in a positivist approach to climate science, taking the example of heat-stress to trees in the Amazon basin. A positivist approach might involve counting stressed and healthy trees to determine regional climate impacts, developing a model that can only ever approximate the real system and relying on the meteorological offices of sometimes distant countries for accurate observations and metadata to calibrate and drive the model. Irwin (2010:3) notes:

> "Positivism expects a logical formula that explains the matter of the Earth. Complexity is read as a set of complicated causal stimuli that needs to be included in the model. The unknown and the uncertain are just the yet-to-be-discovered or better still, the yet-to-be-deduced."

It is clear that the inherent uncertainties involved in understanding and modelling the climate system mean that a deterministic approach is not applicable. The attempt to develop as objective an approach as possible does not rule out all subjectivities. As noted in Chapter 2, climate modelling is based on key assumptions which are fundamentally unverifiable by experience. In modelling climate states that have not been experienced before, it cannot be stated with certainty that climate processes and feedbacks will behave as they are currently observed to behave. The climate system is a non-linear system and processes that have been replicated and investigated in laboratory conditions cannot be assumed to behave the same way in the real climate system. In utilizing climate models to generate future scenarios, it cannot be stated with certainty that a single model will perform in different forcing conditions as it has been observed to perform in present-day control simulations.

3.3 MULTI-MODEL APPROACHES TO CLIMATE MODELLING

This intersection between science and society is especially important in the field of climate modelling. The mathematical and physical representation of the climate system cannot be pursued in isolation as there are human interests and societal concerns to address. Hulme (2007: 1) suggests that the task of "making human sense of climate change" may be beyond the scope of positivist science. As such, while improving the numerical representation of the climate is clearly beneficial to climate planning, a more useful approach would also attempt to address the consequences of model uncertainty and speculate as to the best course of action in light of this uncertainty. Referring to conservation studies, Robertson and Hull

(2001) describe the need for information that is scientifically robust yet also reflective of the pragmatic nature of the adaptation decision-making process.

Where decisions are based on a single future projection, there is great potential for over- or under-estimation of the level of risk and this may lead to mal-adaptation. The multi-model method is a more pragmatic approach to regional climate modelling which acknowledges that all models are potentially lacking in skill and therefore it is unwise to rely on a single model. It also recognizes that all models represent a possible potential future and by utilizing output from many models, more of these potential futures are sampled. This technique is an improvement from deterministic studies in that it provides a more robust basis for climate adaptation and policy decisions

However, while the approach constructs a fuller picture of the range of potential futures, it does not address uncertainty inherent in the models. A significant issue with the multi-model approach is how to proceed when projections lack coherency with each other. Coherentism is a knowledge theory which holds that a complete set of beliefs form a system in which beliefs support each other. Importantly, for such a system to be justified, all the beliefs must be consistent with one another (Lightbody, 2006). Consider an ensemble of climate model projections as a belief set about future climate and it becomes apparent that such a system lacks justification as often, model outputs are distinctly inconsistent with each other. For example, if one model projects an increase in rainfall under climate change and another model projects a decrease (Figure 3.2), how should decision-makers seeking to incorporate climate change into policy proceed?

Since there is always a possibility of error in current certainties, uncertainty is unavoidable. However, it is worth considering if any of this uncertainty is reducible. One possible way to improve the multi-model approach would be to include some experiential knowledge about the individual models in the ensemble. There is a clear need to weight models based on their skill, to form a more coherent set of future projections and reduce the occurrence of contradictory scenarios which are of little use to the adaptation decision-making process. Of course, weighting model projections can be a subjective approach but it can be made more robust by using

multiple diagnostics metrics and attempting to account for as much uncertainty as possible (Tebaldi and Knutti, 2007).

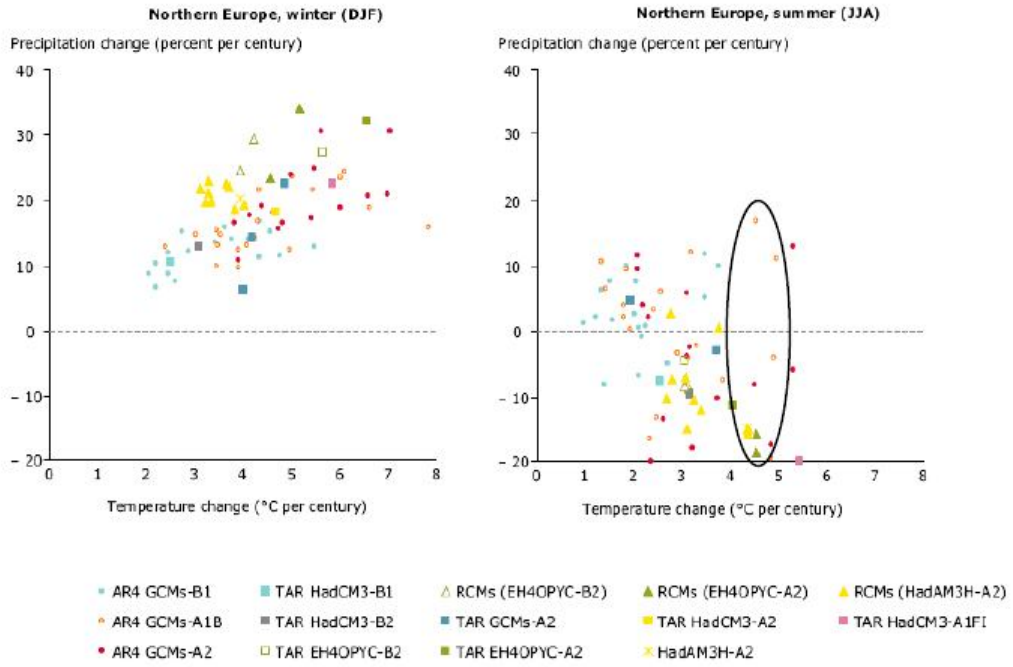


Figure 3.2: European precipitation projections for 2070-2100 relative to simulated present day climate under different SRES scenarios. While model outputs are consistent with each other for winter, the models project both increases and decreases in precipitation for summer (Source: Carter and Fronzek, 2008).

3.4 MODEL VALIDATION AND VERIFICATION

Although all information is potentially mistaken, it does not have to be viewed from the position of skepticism; that is, viewed with doubt. The work of Pierce (1868), one of the first to develop pragmatism as a philosophical theory, provides a reconciliation of the seemingly opposing concepts of fallibilism and antiskepticism of information. Peirce was a 'contrite fallibilist' (Ormerod, 2006:897), a viewpoint which argues that while current knowledge may require revision as errors emerge, this does not prevent any progress being made.

Validation, in the applied rather than the philosophical sense, provides an opportunity to progress the level of understanding of climate modelling while also improving the models themselves through further development and revision. Rather

than seeking to be absolutely certain of climate models and their projections, it is possible to conduct research in a self-correcting manner and make progress with the awareness that knowledge may need to be reviewed. Climate model validation is at its most robust when it fits this description, with validation outcomes forming the basis for on-going model development rather than being perceived as an end-point of the model development process (Figure 3.3).

However, traditional approaches to model validation have relied on a foundationalist approach to research, assuming that basic beliefs such as model skill in hindcasts support derived beliefs, such as confidence in future projections. Yet the experiential evidence of the control or hind-cast run is fallible and as such, derivations of future skill informed only be mean-based analyses may be flawed. Conversely, when all future projections are given equal weight, the result is an incomplete attempt at coherentism as outlined in the previous section, as model outcomes may significantly contradict each other, limiting their usefulness.

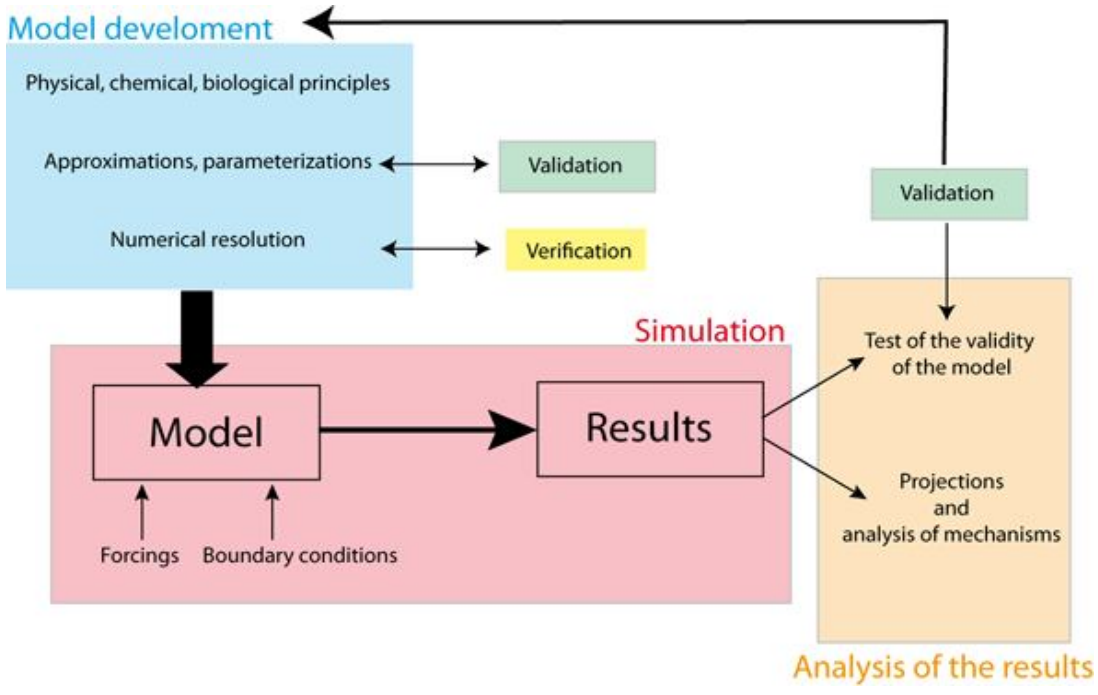


Figure 3.3: A schematic to describe the model development and validation process, with validation outcomes providing feedback for on-going model development (Source: Universite Catholique de Louvain http://stratus.astr.ucl.ac.be/textbook, accessed 23/07/2010).

An optimum approach would resemble the intermediate theory that Haack (1993) refers to as foundherentism. It would allow experiential justification as well as mutual dependence among projections. Such an approach would consider the results of present-day validations in the context of how the climate system should behave and how other climate models have been observed to perform. Considering the self-consistancy of the model in simulating the large-scale climate phenomena influencing a region and their effects on regional climate may give a better picture of how a model performs and how it arrives at its projections. The approach of using both extensive experiential evidence and an ensemble of possible projections has the potential for increased reliability, which in turn provides a more robust basis for decision-making.

However, it is also important to note that there are limits to what can be accomplished through model validation. Verifying the wealth of interdisciplinary scientific information incorporated into models and confirming the inevitable assumptions on which they are based is beyond the scope of this thesis and indeed is not its aim. Rather than concerning itself with the components of each model, this thesis aims to investigate how the model as a whole performs at simulating various aspects of and influences on Irish climate.

3.5 APPROACH USED IN THIS THESIS

Model validation studies are often grounded on the assumption that skill at modelling the key impact variables of temperature and precipitation in the present day is a robust indicator of skill at modelling these variables under future climate conditions and increased GHG forcing. While other variables may be introduced to further discuss and explain the primary results, quantitative analysis of skill tends to be limited to temperature and precipitation. Additionally, the spatiotemporal nature of climate is often neglected in validation studies. For example, while Reichler and Kim (2008) evaluate GCMs on a range of climate variables, only the time-mean climate state is considered with data limitations cited as a reason for this.

In reality, there is no reason to suppose that a model which simulates these mean characteristics well is doing so for the right reasons, without investigating the

interplay between variables as various spatial scales more closely. A model's projections may be skilful but its explanations of the climate phenomena underlying those projections may not be. Brown (2004:377) notes:

> "…limiting an assessment of uncertainty to model outputs can lead to models that appear more or less useful than their statements about the world would imply, encouraging optimism and pessimism (ignorance) in decision-making respectively"

This thesis seeks to determine an optimum approach for extracting useful information from climate models for use in regional impacts assessment, using numerical techniques to quantify error and analyze model representation of the climate phenomena underlying the mean values. Models assessed in this manner are likely to be more useful in climate adaptation and planning than models assessed on their simulation of climate averages. As such, the research presented in this thesis is both exploratory and confirmatory (Dyer *et al.*, 2003). While the comprehensive investigation of model skill in the Irish domain is exploratory, the thesis also looks to verify whether the assumptions commonly made in RCM skill assessments are truly robust and is therefore confirmatory.

CHAPTER 4
MODEL SKILL AT SIMULATING INTERANNUAL AND MEAN ANNUAL CLIMATE PATTERNS (1961-1990)

4.1. INTRODUCTION

Temperature and precipitation are two key variables which are of paramount importance in determining potential future climate impacts. Therefore an important requirement for any RCM aiming to provide useful information about future climate is that it must first be able to skilfully represent these parameters in the current climate.

However, traditional approaches to model skill assessment based on averaging data across a domain and compiling skill scores have not fully characterized the deep uncertainty that can potentially underlie mean values. For example, a change in mean is likely to directly affect areas such as agricultural and biodiversity, but may also have the indirect effect of altering other aspects of the climate distribution, such as the variability. Additionally, skilful representation of the mean climate is no guarantee that the model captures the complex dynamics of the climate system, as pseudo-skill can arise through the cancellation of different errors within the model also. In short, skill scores can be made much more useful when they are accompanied by an explanation of how skill or deficiencies arise. An examination of mean and variability together along with an investigative analysis of results obtained will therefore provide a more complete assessment of model skill. An apparently skilful simulation of mean values does not guarantee that the model is able to capture the complex climate processes and interactions that underlie these mean values.

This chapter will assess the temporal characteristics of precipitation and temperature and determining how well they are characterized by traditional

assessment methods. Model skill at representing mean trends and variability of the climate on both interannual and mean annual timescales is assessed, as these are key statistics in terms of developing climate adaptation strategy. Present-day climate simulations from the models in question, covering the baseline period of 1961-1990, are compared to the observed baseline climate to identify model errors and biases. The simulations were made available through the EC project PRUDENCE (Prediction of Regional scenarios and Uncertainties for Defining EuropeaN Climate change risks and Effects) (Christensen *et al.*, 2007). Only single simulations are available for majority of models, with the exceptions of HadRM3P, HIRHAM driven by HadAM3H and ARPEGE. Where there are more than one simulation with the same model, they are assigned a unique identifier (e.g. –H for HadCH3/HadAM3H, -E4 for ECHAM4-OPYC, -a). These are given in Table 4.1. Uppercase –H or –E signifies a variation in GCM driver while lowercase –a, -b or –c signifies different iterations of the same model.

GCM	RCM			Reference
HadCM3/HadAM3P Pope *et al.* (2000)	HadRM3P-a	HadRM3P-b	HadRM3P-c	Moberg and Jones (2004)
HadCM3/HadAM3H Pope *et al.* (2000)	PROMES RACMO CHRM CLM REGCM REMO RCAO-H HIRHAM-a	 HIRHAM-b	 HIRHAM-c	Castro *et al.* (1993) Räisänen *et al.* (2004) Vidale *et al.* (2003) Doms and Schlatter (2002) Giorgi *et al.* (1993) Jacob and Podzun (1997) Döscher *et al.* (2002) Christensen *et al.* (1996)
ECHAM4-OPYC/ECHAM5 Roeckner *et al.* (1996)	HIRHAM-E5			
Observed SSTs	ARPEGE-a	ARPEGE-b	ARPEGE-c	Déqué *et al.* (1998)
ECHAM4-OPYC Roeckner *et al.* (1996)	RCAO-E4 HIRHAM-E4			

Table 4.1: Summary of models under investigation (Source: Christensen et al., 2007).

4.2 BACKGROUND

For the purposes of this study, model skill is defined as the degree to which the modelled climate and observed climate distributions correspond. This is similar to Murphy's (1993) definition of forecast quality as the degree to which the forecast and the actual events correspond. One must then consider the attributes that

contribute to a skilful projection, and how to quantify them. The field of short-term weather-forecasting already utilizes a wealth of methodologies for validating and verifying forecasts (Barnston, 1992; Casati *et al.*, 2008; Murphy and Wilks, 1998) that can be adopted for the assessment of climate model skill. However, a challenge in the application of skill scores to climate model data is that while there are certain skill scores that are more widely used than others, there is no set of scores which is commonly agreed upon by the climate modelling community. Additionally, the suitability of a metric to assess the data and research questions must be considered. Therefore, while the score itself is a quantitative measure of skill, the process of skill assessment has an element of subjectivity also.

RCM data has both temporal and spatial components and if model outputs are to be useful in informing climate adaptation strategy and climate impacts studies, they should simulate both the temporal and spatial characteristics of the climate system. This chapter will assess temporal aspects of the modelled Irish climate. Patterns occur on a number of different time-scales in the climate system. Examples include annual climatologies and interannual variations. Even the large-scale modes of natural climate variability such as the North Atlantic Oscillation (NAO) display temporal patterns. Of course, no model can match the complexity of the real climate system and as such, perfect accuracy cannot be expected. However, for a climate model to be considered skilful, it should reasonably approximate observed statistics and characteristics such as these.

At different timescales, different skill metrics must be used due to inherent characteristics of the climate model data. For example, RCM data has no temporal consistency with observations due to the lack of temporal coherence between actual and GCM modelled climate. Different GCM data or reanalysis data can be used to derive boundary conditions for RCM and each set of boundary conditions can generate a different but mathematically valid simulation. The boundary value problem in RCMs is what is mathematically referred to as a non-homogeneous problem, meaning that there is no unique set of "correct" values. Even small differences in the boundary conditions can lead to quite different model outcomes, as demonstrated by Mooney *et al.* (2010), yet all are mathematically valid. In this way, the differences between observed climate conditions and GCM modelled conditions lead the RCM simulation to evolve differently to the observed climate, resulting in a

lack of temporal consistency between the RCM simulation and the observations. As such, one cannot look for association between observations and modelled output at the interannual timescale. However, by considering the model data as a distribution of possible outcomes other aspects can be analyzed. For example, the interannual variability can be quantified by the standard deviation of the distribution of outcomes (Giorgi *et al*,. 2004; Sato *et al.*, 2007).

Murphy (1993) specified various attributes that effect forecast quality, such as error or bias and association. Error is the difference between individual pairs of observations and model projections while bias is the difference between the average of all observed values and the average of all projected values. Association is the strength of the linear relationship between the observations and the projected data. For this assessment, metrics are required that quantify the extent to which model projections of climate means and variances display attributes such as accuracy and association.

However, a wide variety of tests exist for determining differences of means and variances and association between different datasets. The appropriate choice depends on the nature of the data being analyzed. Many of the most commonly used tests, such as the T-test for means, assume that the data in question is normally distributed. Such tests are referred to as parametric statistics. Vidale *et al.* (2007) notes that for a small dataset such as seasonal means, moment-based statistics are more robust than nonparametric statistics. This is because parametric methods have a higher statistical efficiency (Scherrer *et al.*, 2005). However, while the T-test has been found to be fairly robust to departures from normality (Boneau, 1960; Sawilowsky and Blair, 1992) certain tests, such as the F-test for variance, are less robust when applied to non-normal data (Brown and Forsythe, 1974). In such cases, non-parametric statistics may be required. If the assumptions being made are correct, parametric methods can produce more skilful results than parametric statistics, but if the assumptions are incorrect and a parametric method is applied, it can generate a misleading result. Therefore, it is important to first determine whether the data is normal in distribution or not.

It is also important to consider the focus of the analysis when choosing tests. For example, there are two commonly used coefficients for determining the

correlation between datasets. The Pearson coefficient is parametric while the Spearman coefficient is non-parametric. As such, the Pearson coefficient is more sensitive to outliers than the Spearman coefficient, as Spearman ranks the data under investigation as part of its calculation, which dampens the effect of outliers. However, in a modelled annual climatology such outliers are highly undesirable as they distort the annual pattern. Pearson's sensitivity to outliers, which could be considered a disadvantage in a different situation, is advantageous in this analysis as it penalizes models with outliers and weights models without outliers as more skilful.

4.3 INTERANNUAL VARIABILITY

4.3.1 Review of methods

The first part of this assessment focuses on interannual variability. Large-scale modes of climate variability such as the North Atlantic Oscillation (NAO) operate on somewhat regular cycles lasting from several months to several years (An *et al.*, 2005; Vallis and Gerber, 2008; Vimont, 2005). Mean temperature and precipitation can be influenced by these modes, depending on what stage of their cycles they are at. Climate change could potential disrupt these modes or heighten their effects (Latif and Keenlyside, 2009; Paeth *et al.*, 1999; Yeh and Kirtman, 2007). Therefore it is important that climate models skilfully represent interannual variability, that is, variation in temperature or precipitation from year to year, as this is the scale at which changes to the large-scale climate modes tend to be most apparent.

Climate models are not and cannot be "predictions" of the future, due to the large number of uncertainties involved (Palmer *et al.*, 2005). As the RCMs under investigation are driven by data from multiple GCMs they cannot be expected to reproduce observed climate anomalies in chronological order. The climate system is highly sensitive to changes in initial conditions, much like a chaotic system, and in any forecast errors in initial conditions grow exponentially over time (Gustafsson *et al.*, 1998; Rabier *et al.*, 1996). Even models specifically designed to study a particular climate mode experience a decrease in predictive skill as lead time increases (Latif *et al.*, 1998). As such, one cannot expect the RCMs to model positive

NAO phases, for example, exactly where they occur in the observed timeline. Therefore, directly comparing the observed and projected time series of interannual data is of no value in determining model skill.

However, the purpose of a climate model is not to give a skilful forecast of weather on a given day 20 years from now, but a skilful representation of the mean climate and variations in climate that could be expected. As such, the present-day run of the model should not be expected to model particular weather events in the correct chronological order. Instead, it is more useful to consider the possible climate outcomes projected by the model as a frequency distribution. The statistics of the distribution of modelled outcomes must be compared to the same statistics of the observed climate to examine model skill at representing interannual variability (e.g. Giorgi *et al.*, 2004).

Graphical techniques can be useful to provide a visual representation of the data and form preliminary comments about the skill of the models. A rank histogram (Hamill, 2001) can be used to assess how well the true interannual variability is represented by an unweighted average ensemble (Figure 4.1).

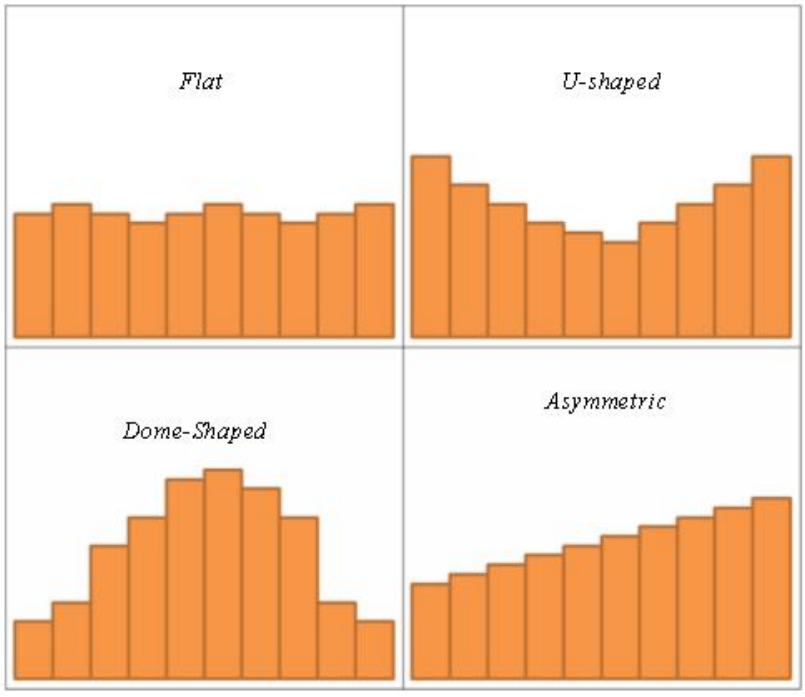


Figure 4.1: Shapes of rank histograms.

The observed annual values are ranked from lowest to highest, to give the values for the frequency bins. The values from each ensemble member are also ranked from lowest to highest and assigned to the appropriate frequency bin. If the models in the ensemble represent the observed distribution skilfully, one would expect equal frequencies across the bins. Therefore, a flat histogram indicates that the ensemble spread represents the true interannual variability well. However, exact flatness is not expected because with a finite number of model simulations it is unlikely that the true distribution will be fully sampled (Descamps and Talagrand, 2007).

Another possible outcome is overpopulation of the highest and lowest ranks, resulting in a U-shaped histogram. If the histogram is U-shaped, the majority of modelled values are falling outside the range of the observed distribution. In such a case, the ensemble spread may be too large, leading to an overestimation of interannual variability. One may also find overpopulation of the central ranks, resulting in a dome-shaped histogram. A dome shaped histogram forms when the majority of modelled values fall well inside the range of the observed distribution. This result suggests that the ensemble spread is too small and fails to capture the extremes of the true distribution. A final possibility is an asymmetric histogram, in which overpopulation occurs towards one end of the histogram but not the other. An asymmetric histogram indicates that the ensemble spread is skewed.

The observed and projected distributions for each model can also be compared using a Q-Q plot which is a plot of the quantiles of the two distributions against each other (Beirlant *et al.*, 2007). The more similar the distributions, the closer the data will fall to the 45^{O} line y=x. If the distributions are identical, the points will fall precisely on that line. A guide to the shapes the Q-Q plot may take and the inferences that can be made about each distribution from the Q-Q plot is given in Figure 4.2.

The correlation coefficient of a linear regression line through the points could be used as a numerical measure of fit to quantify how similar the distributions are, but would overlook a mean shift. Instead, a combination of metrics is used to fully capture model skill. Interannual variability can be quantified using the temporal standard deviation, an approach used by Giorgi *et al.* (2004). However, standard

deviation is not sensitive to shifts in data; for example, the standard deviation of the sequence (1, 2, 3, 4, 5, 6) is the same as the standard deviation of the sequence (11, 12, 13, 14, 15, 16). Standard deviation is a measure of the dispersion of a data set from its mean, and does not consider whether the modelled mean is skilful or correct when compared to the observed mean. Hence, using the temporal standard deviation as a metric identifies models that simulate the shape of the distribution well, but does not distinguish whether the distribution is shifted, either by consistently overestimating or underestimating the variable. To determine whether such shifts are also occurring, the means of the observed and modelled datasets must also be compared.

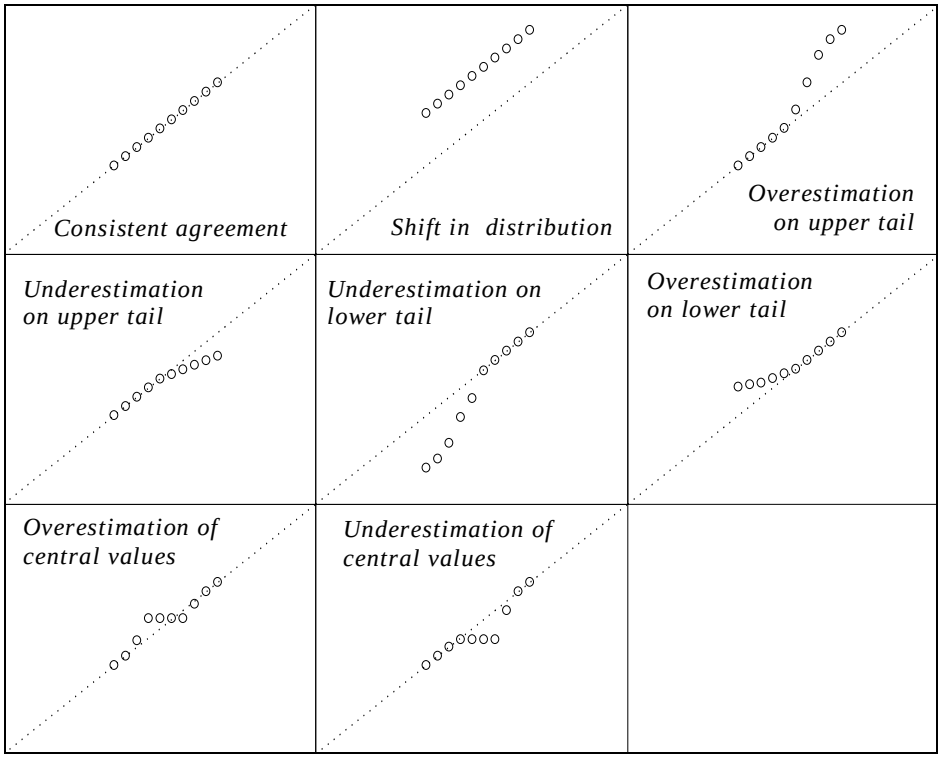


Figure 4.2: Shapes of Q-Q plots.

A significance test should also be applied to determine whether the observed and modelled data are significantly different. If the assumption of normality is fulfilled, parametric methods can be chosen for significance testing, namely the F-test for variance and the T-test for means. As is standard practice in tests of significance, a null hypothesis of "no difference" between the observed and modelled distributions is initially prescribed. A test statistic is calculated to measure

compatibility between the data and this null hypothesis. The test consists of calculating the probability of obtaining a statistic as different or more different from the observed statistic, assuming that the null hypothesis is correct. If this probability is sufficiently low, then the difference between the observed and modelled statistics is said to be "statistically significant". A confidence level alpha is selected and the test result is compared to this value to assess whether it is sufficiently low. Levels of 0.05 and 0.01 are most commonly used.

4.3.2 Methodology

As temporal consistency between modelled output and observations cannot be expected at the interannual scale, model output must be considered as a distribution of possible outcomes and assessed using statistics that quantify variability and means rather than association.

The Shapiro-Wilk test is used to determine whether data comes from a normal population (Shapiro and Wilk, 1965). The null hypothesis is that the data came from a normal population, and this is rejected at the alpha significance level if the p-value of the test statistic is less than chosen alpha level. The Shapiro-Wilk test statistic is calculated as:

Equation 4.1: Shapiro-Wilk test statistic

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where

$x_{(i)}$ *= the* i^{th} *smallest value in the sample,*

$\bar{x}$ *= the sample mean and*

a_i *= a set of constants, given by:*

$$(a_1,\ldots,a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$$

where

$m = (m_1,...,m_n)^T$, the expected values of the order statistics of independent and identically-distributed random variables sampled from the standard normal distribution and

V = the covariance matrix of those order statistics.

Interannual variability is visualized using the rank histogram and Q-Q plotting. Interannual variability is measured using the temporal standard deviation:

Equation 4.2: Temporal standard deviation

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N-1}}$$

where

N = number of years of data (30) and

$\bar{x}$ *= 30 year average*

The percentage error of $\sigma_{modelled}$ with respect to $\sigma_{observed}$ is also used to quantify agreement between observed and modelled data. For the variance metric, the two-tailed F test is used to verify whether the observed and modelled data have equal variances. The F test statistic is defined as:

Equation 4.3: F test statistic

$$F = s_o^2 / s_m^2$$

where

s_o^2 *= variance of the observed data and*

s_m^2 *= variances of the modelled data.*

Bias and percentage error of the modelled mean with respect to the observed mean is also calculated. The T-test is used to assess whether the observed and

modelled means are significantly different to each other, using the results of the F test to choose between the homoscedastic, “equal variance” T-test or the heteroscedastic, “unequal variance” T-test.

Gridded observational data provided by Met Eireann to the British Irish Council and made available via the BIC was used to evaluate the RCMs (BIC, 2003). This data is chosen for its high resolution of 5km. Other available observation-based data such as the NCAR/NCEP or ERA-40 reanalysis data, due to their global coverage, are currently available on a much coarser grid of 2.5^{o} and as such, would have to be interpolated to the finer grid used by the RCMs to create a comparable dataset. This approach has significant potential for error as the interpolated values are only approximations based on the coarser data. In the case of Ireland, conversely, transforming observation-based data from a very fine grid to the grid used by the RCMs is more desirable as the regridded values are based on a network of finely resolved data points.

4.3.3 Results

Results of the Shapiro-Wilk test for normality are given in Table 4.1. At α=0.05, the critical W value is 0.93, therefore a test statistic W of less than 0.93 indicates non-normality at the 95% confidence level. Datasets with a test statistic indicating non-normality are highlighted with grey shading. However, the majority of datasets are found to be normal. Importantly, the observed interannual datasets are all found to be normal. As the observed data that the models are to be compared with is normally distributed and the majority of the modelled data is also normally distributed, parametric tests are chosen to quantify skill at the interannual scale.

When constructing RCM ensembles, one option is an unweighted mean ensemble in which all members are taken to have equal skill. The rank histogram provides some quick insights into the performance of such an ensemble. However, as noted by Hamill (2001) in relation to the Talagrand technique, visual assessment tools can potentially be misleading as they are open to misinterpretation. Therefore, closer inspection of the data is warranted to both confirm and further characterize the issues identified from the rank histogram analysis.

	Observed	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
Yearly																				
Temperature	0.97	0.94	0.97	0.95	0.97	0.95	0.93	0.95	0.94	0.96	0.95	**0.90**	0.98	0.97	0.99	0.95	0.97	0.96	0.97	0.98
Precipitation	0.98	0.96	0.97	0.98	0.94	0.96	0.95	0.96	0.96	0.95	0.97	0.96	0.95	0.96	0.98	0.98	0.96	0.98	0.97	0.97
Winter																				
Temperature	0.95	0.97	0.96	0.94	**0.92**	0.95	0.95	0.96	0.97	0.95	0.95	0.95	0.94	0.96	0.96	**0.92**	0.97	0.97	0.97	0.96
Precipitation	0.99	**0.93**	0.99	0.98	0.98	0.96	0.98	0.97	0.99	0.97	0.96	0.94	0.99	0.93	0.98	0.97	0.96	0.97	0.97	0.98
Spring																				
Temperature	0.95	0.93	0.98	0.95	0.96	0.98	0.97	0.97	0.98	0.96	0.98	0.97	0.97	0.96	0.98	0.93	0.96	0.97	0.93	0.95
Precipitation	0.98	0.96	0.97	0.97	0.97	0.96	0.97	0.97	0.98	0.97	0.98	0.97	0.95	0.97	0.97	0.98	0.96	0.97	0.98	0.97
Summer																				
Temperature	0.94	0.98	0.96	0.97	**0.92**	0.97	0.96	0.97	0.94	0.97	0.97	0.97	0.94	**0.92**	0.96	0.94	0.98	0.97	0.96	0.96
Precipitation	0.97	0.97	0.97	0.98	0.96	0.98	0.98	0.98	0.98	0.96	0.97	**0.93**	0.98	0.98	0.97	0.96	0.98	**0.93**	0.97	0.95
Autumn																				
Temperature	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.96	0.95	0.99	**0.90**	0.96	0.96	0.96	0.97	0.95
Precipitation	0.97	0.97	0.98	0.98	**0.92**	0.94	0.96	0.97	0.95	0.98	0.96	0.95	0.97	0.95	0.98	0.96	0.99	0.95	0.94	0.94

Table 4.1: Results of Shapiro-Wilk test for normality performed on yearly and seasonal interannual data for 1961-1990 over the Irish domain. At α=0.05, the critical W value is 0.93, therefore a test statistic W of less than 0.93 indicates non-normality at the 95% confidence level. Datasets with a test statistic indicating non-normality are highlighted with grey shading.

4.3.4 Results: Temperature

The rank histogram for interannual variability of temperature is quite asymmetric, with overestimation in the highest rank and underestimation in the lower ranks (Figure 4.3a).

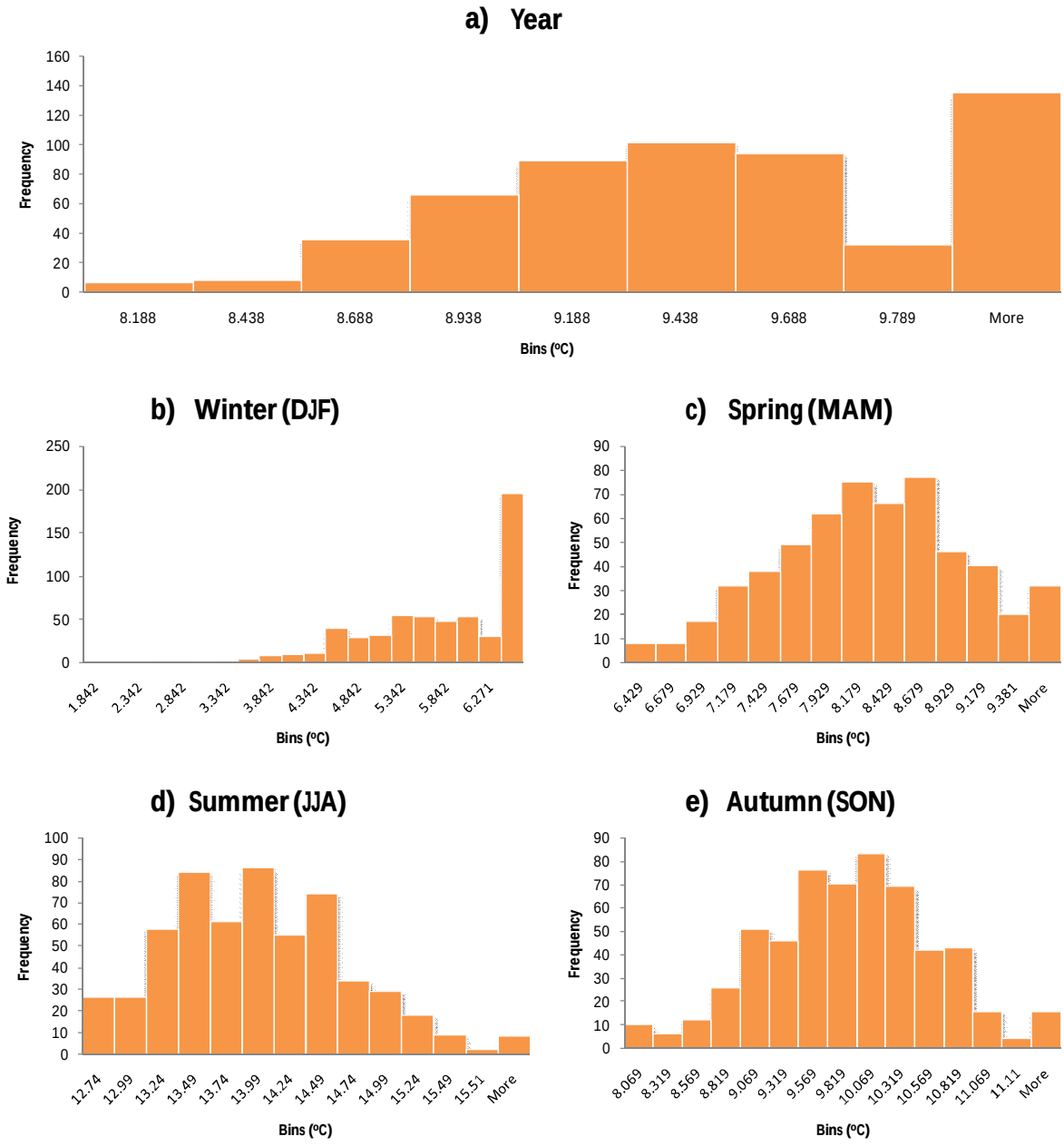


Figure 4.3: Rank histograms for interannual temperature data (1961-1990). Bins are derived from observations.

If the ensemble were skilful, one would expect approximately the same number of occurrences in each bin. This histogram suggests that the ensemble underestimates the lower extremes of the distribution while overestimating the upper

extremes. This behaviour may also be caused by a systematic bias within certain models which would shift data uniformly upwards, or may arise through the combination ensemble members with both of these types of error.

Further insights can be gained by applying the rank histogram to the seasonal interannual datasets. From these graphs it is evident that the biases identified in the year graph are most notable in winter (Figure 4.3b). In all other seasons, the histogram forms a dome-like shape which suggests that the observed range of variability of the distribution is not being captured by the models (Figure 4.3c, d and e). The Q-Q plotting technique is used to further inspect output from the individual ensemble members, to determine how closely the modelled distribution of interannual values resembles the observed distribution.

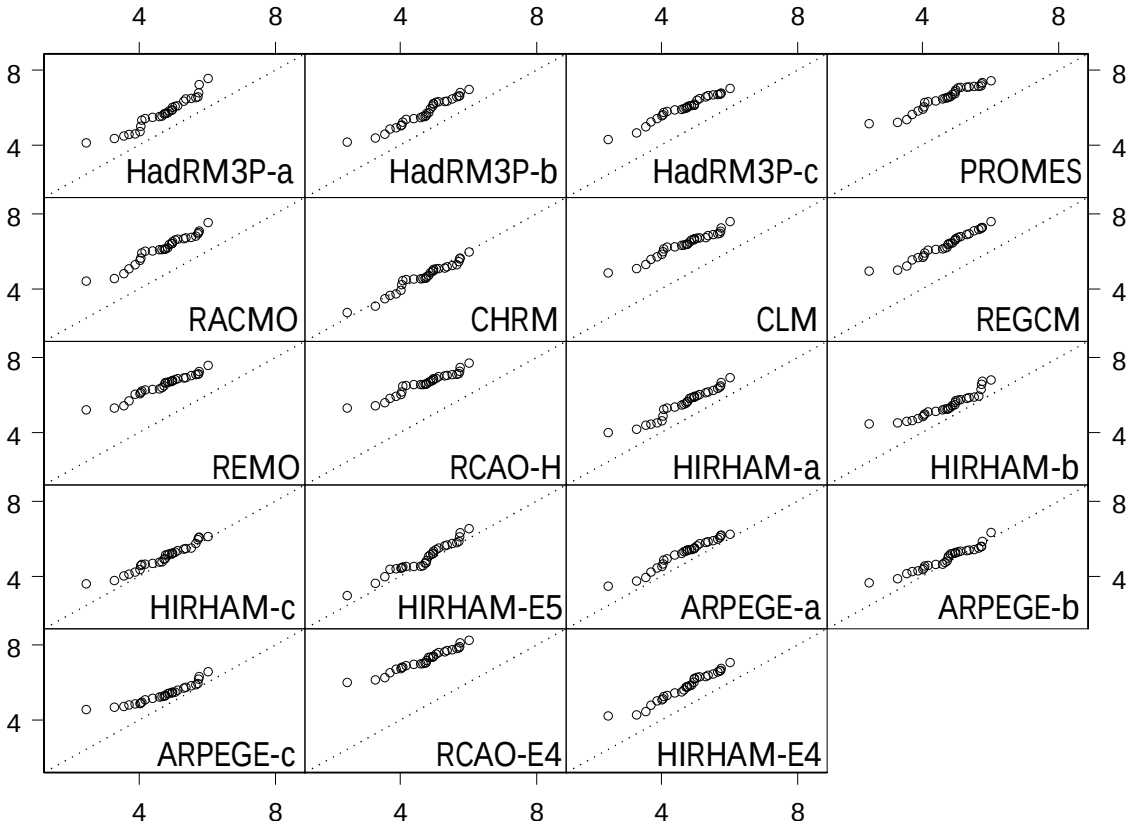


Figure 4.4: Q-Q plots of modelled versus observed interannual winter temperature (^{o}C) for the period 1961-1990.

The majority of models display a positive bias in winter, with the lower tails of the distribution particularly overestimated (Figure 4.4). There are some exceptions however. For example, CHRM and HIRHAM-E5 both capture the observed

distribution quite skilfully. Additionally, the models ARPEGE-a, ARPEGE-b and HIRHAM-c approximate the observed distribution except at the lower tail, where overestimation occurs. These results are in keeping with the rank histogram of mean-ensemble winter temperature.

In spring, a combination of model biases in the individual models gives rise to the mean-ensemble rank histogram. Many models overestimate temperature in spring, in particular the upper-central values and the extreme lower values of the distribution (Figure 4.5). Examples of this bias can be seen in the Q-Q plots of RCAO-E4, RACMO and REMO. However, certain models such as PROMES and CHRM underestimate the lower values of the distribution, balancing the overestimation errors of other models when taken as a mean-ensemble. As such, overestimation of central values is the effect that remains and is captured in the rank histogram.

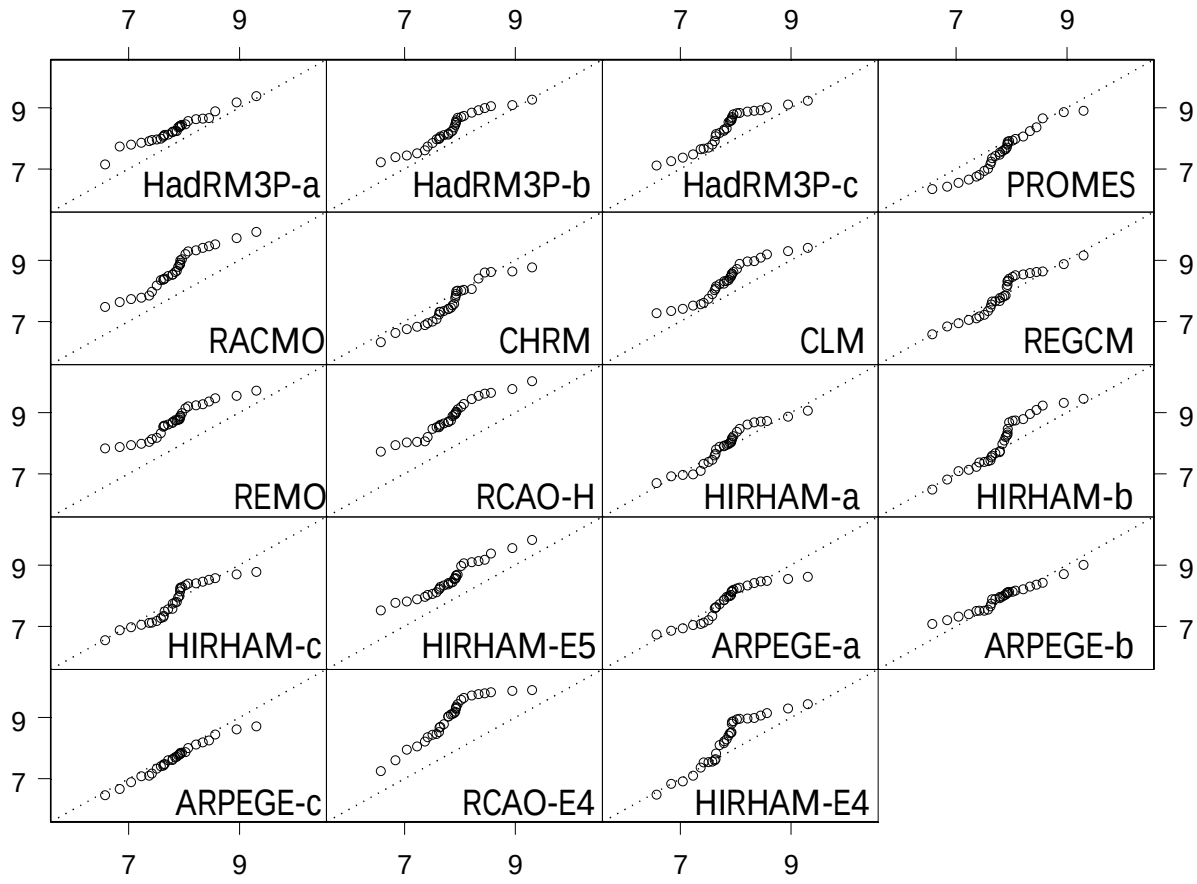


Figure 4.5: Q-Q plots of modelled versus observed interannual spring temperature (°C) for the period 1961-1990.

Most of the models capture the observed distribution in summer quite skilfully, with HadRM3P-a, HadRM3P-c and HIRHAM-E5 in particular performing well (Figure 4.6). The shape of the rank histogram is again the result of a combination of different biases amongst the individual models, which highlights the importance of examining individual model skill before attempting to construct an ensemble. Several models underestimate the upper values of the distribution, for example PROMES, ARPEGE-a and ARPEGE-b, leading to lower population in those ranks and higher population in the lower ranks. Conversely, a number of different models overestimate the lower and central values, for example RACMO and REMO. The combination of these biases results in a rank histogram for the mean-ensemble with low population in the upper ranks and higher population in the lower ranks. Rather than error cancellation, which occurs in spring data, the summer data is affected by what could be described as error accumulation.

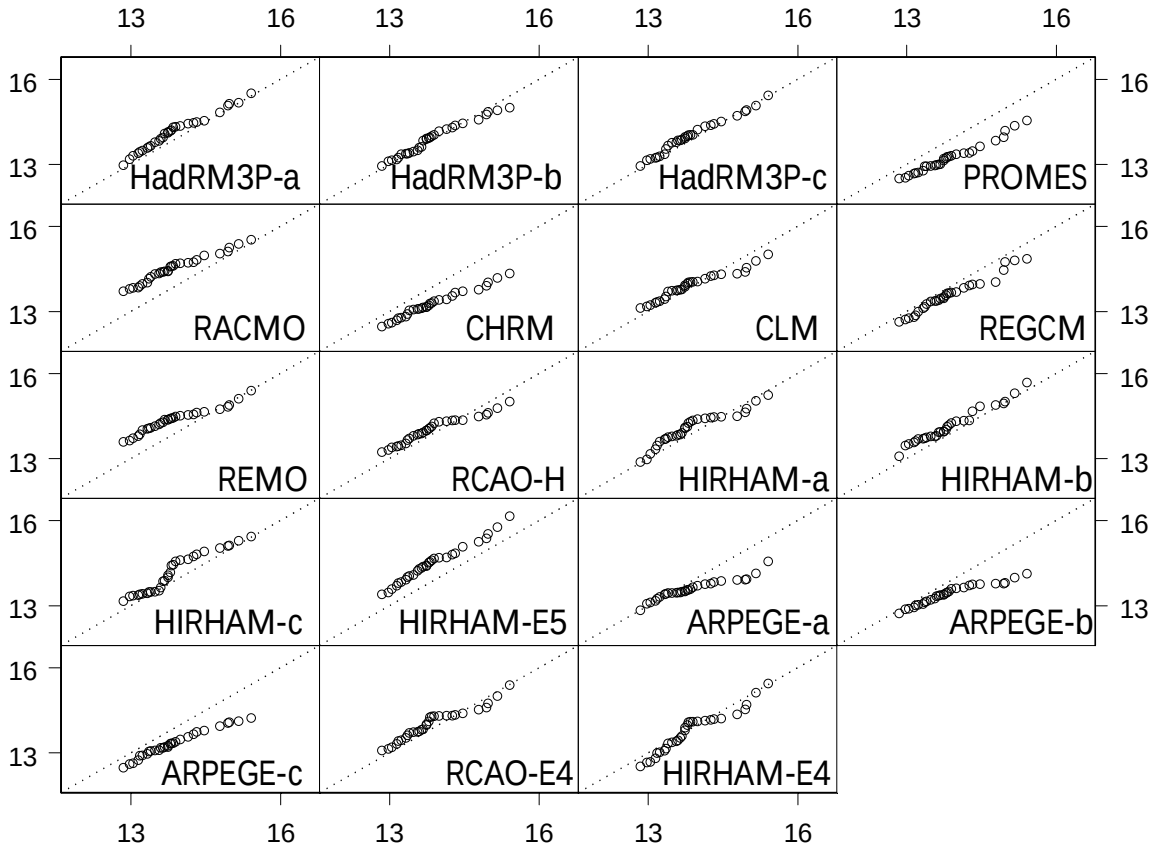


Figure 4.6: Q-Q plots of modelled versus observed interannual summer temperature (^{o}C) for the period 1961-1990.

Autumn is perhaps the season in which the greatest number of individual models best approximate the observed distribution (Figure 4.7). In particular, all five simulations using HIRHAM (HIRHAM-a, HIRHAM-b, HIRHAM-c, HIRHAM-E5 and HIRHAM-E4) represent the observed distribution well. Again, a combination of biases on the tails of the modelled distributions leads to the dome-shaped rank histogram.

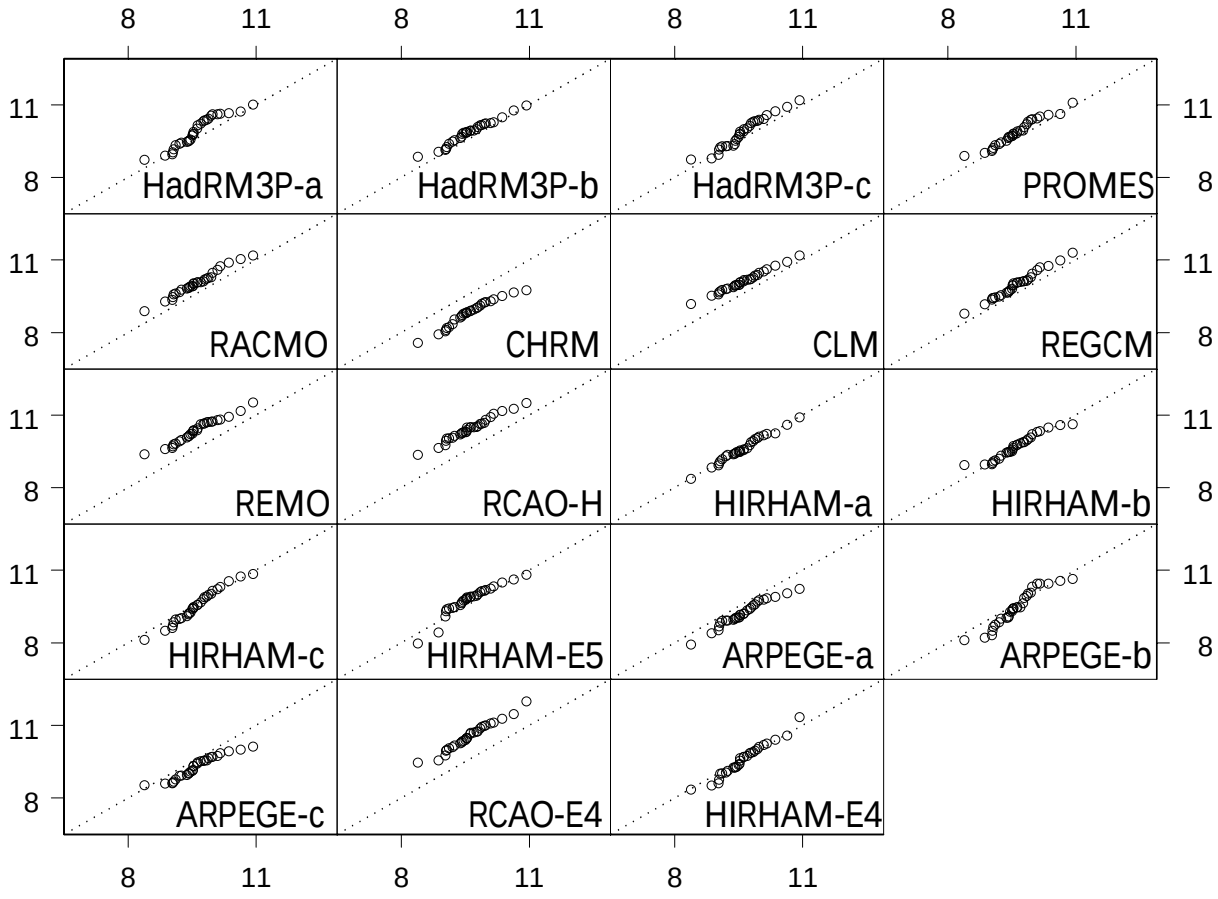


Figure 4.7: Q-Q plots of modelled versus observed interannual autumn temperature (^{o}C) for the period 1961-1990.

Additionally, skill metrics are calculated to assess the models' abilities at simulating climate means and variances. Biases are expressed as a percentage of the observed mean or variance and the significance of any differences identified is assessed using the F-test or t-test. Results are given in Table 4.2 a and b, with significant values for the F-test and t-test marked as bold. These metrics quantify the errors observed in the rank histograms and Q-Q plots.

Winter

	HadCM3/HadAM3P			HadCM3/HadAM3H										ECHAM5	Observed SSTs			ECHAM4-OPYC	
	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
S.Dev.	0.89	0.78	0.69	0.66	0.80	0.81	0.67	0.72	0.61	0.61	0.79	0.65	0.69	0.88	0.77	0.65	0.53	0.59	0.78
Bias	-0.03	-0.14	-0.22	-0.26	-0.11	-0.11	-0.24	-0.19	-0.30	-0.30	-0.12	-0.27	-0.22	-0.04	-0.15	-0.26	-0.38	-0.32	-0.13
% Bias	-3.1%	-15.1%	-24.4%	-28.1%	-12.6%	-11.8%	-26.4%	-20.7%	-33.1%	-33.3%	-13.0%	-29.0%	-24.4%	-4.1%	-15.9%	-28.4%	-42.0%	-35.3%	-14.3%
F-test	0.87	0.38	0.14	0.08	0.47	0.50	0.10	0.22	**0.03**	**0.03**	0.46	0.07	0.14	0.82	0.36	0.08	**0.00**	**0.02**	0.41
Mean	5.70	5.71	5.98	6.51	6.12	4.61	6.35	6.34	6.49	6.58	5.56	5.42	4.98	4.95	5.21	4.94	5.36	7.17	5.73
Bias	1.01	1.02	1.29	1.82	1.43	-0.08	1.66	1.65	1.80	1.89	0.87	0.73	0.29	0.26	0.52	0.25	0.67	2.48	1.04
% Bias	21.5%	21.8%	27.5%	38.8%	30.5%	-1.6%	35.4%	35.2%	38.4%	40.4%	18.5%	15.7%	6.1%	5.6%	11.1%	5.4%	14.4%	52.9%	22.3%
T-test	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.74	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.18	0.27	**0.02**	0.23	**0.00**	**0.00**	**0.00**

Spring

	HadCM3/HadAM3P			HadCM3/HadAM3H										ECHAM5	Observed SSTs			ECHAM4-OPYC	
	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
S.Dev.	0.51	0.58	0.61	0.71	0.67	0.68	0.61	0.70	0.54	0.61	0.65	0.82	0.64	0.59	0.57	0.46	0.56	0.73	0.84
Bias	-0.10	-0.03	0.00	0.10	0.06	0.07	0.00	0.09	-0.07	0.00	0.04	0.21	0.03	-0.02	-0.04	-0.15	-0.05	0.12	0.23
% Bias	-16.5%	-5.7%	0.1%	15.8%	9.6%	11.1%	0.6%	14.3%	-12.3%	0.4%	6.7%	34.2%	5.5%	-2.5%	-6.9%	-24.5%	-8.9%	19.9%	37.8%
F-test	0.34	0.76	1.00	0.43	0.62	0.57	0.98	0.48	0.49	0.98	0.73	0.12	0.77	0.89	0.70	0.14	0.62	0.33	0.09
Mean	8.26	8.26	8.29	7.54	8.66	7.57	8.31	7.86	8.71	8.82	7.90	8.02	7.76	8.51	7.79	7.92	7.64	8.94	8.18
Bias	0.42	0.42	0.45	-0.30	0.82	-0.27	0.46	0.01	0.87	0.97	0.06	0.18	-0.08	0.67	-0.05	0.08	-0.20	1.10	0.34
% Bias	5.3%	5.4%	5.7%	-3.8%	10.4%	-3.5%	5.9%	0.2%	11.0%	12.4%	0.8%	2.3%	-1.0%	8.6%	-0.6%	1.0%	-2.6%	14.0%	4.4%
T-test	**0.01**	**0.01**	**0.01**	0.09	**0.00**	0.11	**0.01**	0.94	**0.00**	**0.00**	0.71	0.34	0.64	**0.00**	0.76	0.59	0.19	**0.00**	**0.08**

Table 4.2a: Skill metrics for interannual temperature (Winter and Spring), for 1961-1990. Significant test values are given in bold. Grey titles represent driving GCMs/boundary data.

Summer

	HadCM3/HadAM3P			HadCM3/HadAM3H										ECHAM5	Observed SSTs			ECHAM4-OPYC	
	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
S.Dev.	0.67	0.60	0.66	0.56	0.50	0.50	0.50	0.62	0.45	0.48	0.59	0.65	0.72	0.72	0.39	0.37	0.49	0.59	0.76
Bias	-0.03	-0.10	-0.04	-0.15	-0.20	-0.20	-0.21	-0.08	-0.26	-0.23	-0.11	-0.05	0.01	0.01	-0.31	-0.33	-0.21	-0.11	0.05
% Bias	-4.8%	-14.2%	-6.2%	-20.9%	-28.6%	-28.9%	-29.4%	-11.9%	-36.5%	-32.1%	-15.9%	-7.4%	2.0%	1.6%	-44.3%	-47.1%	-30.5%	-15.5%	7.7%
F-test	0.79	0.41	0.73	0.21	0.08	0.07	0.07	0.50	**0.02**	**0.04**	0.36	0.68	0.92	0.93	**0.00**	**0.00**	**0.05**	0.37	0.69
Mean	14.11	13.87	13.98	13.22	14.49	13.24	13.89	13.56	14.34	13.99	14.03	14.13	14.13	14.48	13.56	13.39	13.30	14.00	13.74
Bias	0.23	-0.01	0.10	-0.65	0.61	-0.63	0.01	-0.32	0.46	0.12	0.15	0.25	0.25	0.60	-0.32	-0.49	-0.57	0.12	-0.14
% Bias	1.7%	-0.1%	0.7%	-4.7%	4.4%	-4.6%	0.1%	-2.3%	3.3%	0.8%	1.1%	1.8%	1.8%	4.3%	-2.3%	-3.5%	-4.1%	0.9%	-1.0%
T-test	0.21	0.96	0.57	**0.00**	**0.00**	**0.00**	0.93	0.07	**0.00**	0.47	0.38	0.16	0.18	**0.00**	**0.04**	**0.00**	**0.00**	0.48	0.47

Autumn

	HadCM3/HadAM3P			HadCM3/HadAM3H										ECHAM5	Observed SSTs			ECHAM4-OPYC	
	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
S.Dev.	0.66	0.52	0.66	0.59	0.56	0.55	0.48	0.60	0.54	0.52	0.59	0.51	0.71	0.64	0.56	0.74	0.47	0.62	0.74
Bias	0.07	-0.07	0.07	0.00	-0.02	-0.03	-0.11	0.01	-0.05	-0.07	0.01	-0.08	0.12	0.05	-0.02	0.16	-0.12	0.03	0.15
% Bias	12.0%	-11.6%	12.4%	0.2%	-4.2%	-5.4%	-18.2%	2.3%	-8.6%	-11.7%	1.0%	-13.7%	20.8%	8.4%	-3.7%	26.7%	-19.7%	5.4%	25.8%
F-test	0.54	0.51	0.53	0.99	0.82	0.77	0.29	0.90	0.63	0.51	0.96	0.43	0.32	0.67	0.84	0.21	0.24	0.78	0.22
Mean	9.87	9.83	9.82	9.83	10.01	8.77	10.08	9.94	10.33	10.40	9.58	9.67	9.49	9.74	9.18	9.41	9.26	10.53	9.56
Bias	0.26	0.83	0.82	0.83	1.01	-0.23	1.08	0.94	1.32	1.39	0.58	0.66	0.48	0.73	0.18	0.41	0.25	1.53	0.56
% Bias	2.7%	9.2%	9.1%	9.2%	11.2%	-2.6%	12.0%	10.4%	14.7%	15.5%	6.4%	7.4%	5.3%	8.1%	2.0%	4.5%	2.8%	17.0%	6.2%
T-test	0.12	0.13	0.21	0.16	**0.01**	**0.00**	**0.00**	**0.04**	**0.00**	**0.00**	0.85	0.71	0.46	0.44	**0.01**	0.25	**0.01**	**0.00**	0.77

Table 4.2b: Skill metrics for interannual temperature (Summer and Autumn), for 1961-1990. Significant test values are given in bold. Grey titles represent driving GCMs/boundary data.

In winter, the t-test confirms statistically significant differences in means in the majority of models, with CHRM, HIRHAM-c, HIRHAM-E5 and ARPEGE-b the only exceptions (Table 4.2). However, the models capture the observed variability of winter temperature, quantified be the temporal standard deviation, more skilfully. Only four models are found to have significantly different standard deviations to the observed data, namely REMO, RCAO-H, ARPEGE-c and RCAO-E4. The small, insignificant differences in variance coupled with a more significant change in mean all point towards a systematic model bias in which the entire distribution is shifted, preserving the variance but altering the mean.

In spring and autumn, variability is also well-described by the models, with no individual model displaying a significantly different variance to the observed. However, in summer, five models are found to have significant variability biases, ranging from -47.1% to -30.5% of the observed standard deviation. Conversely, significant differences in means occur in all seasons, in ten models in spring, eight in summer and nine in winter. While the t-test identifies these differences as significant, they are much smaller than the errors observed in winter.

4.3.5 Results: Precipitation

The rank histogram for interannual values of precipitation is quite different in shape to the temperature histogram (Figure 4.8a). It is approximately flat across the first six bins. However, the seventh bin has exceptionally low frequency and the final bin has much higher frequency of occurrences. Examination of the seasonal data reveals different patterns in each season which average each other out to give the annual average pattern. In winter and spring, the graph is dome-shaped, suggesting that the models fail to capture the range of observed variability (Figures 4.8b and c). In summer, the dome shape is still apparent but is much less pronounced (Figure 4.8d). In autumn, two key patterns emerge. The histogram is quite asymmetric, suggesting a downwards shift in the modelled distributions (Figure 4.8e). However, there is also overestimation in the upper bin which suggestions that some models may have a wet bias in autumn.

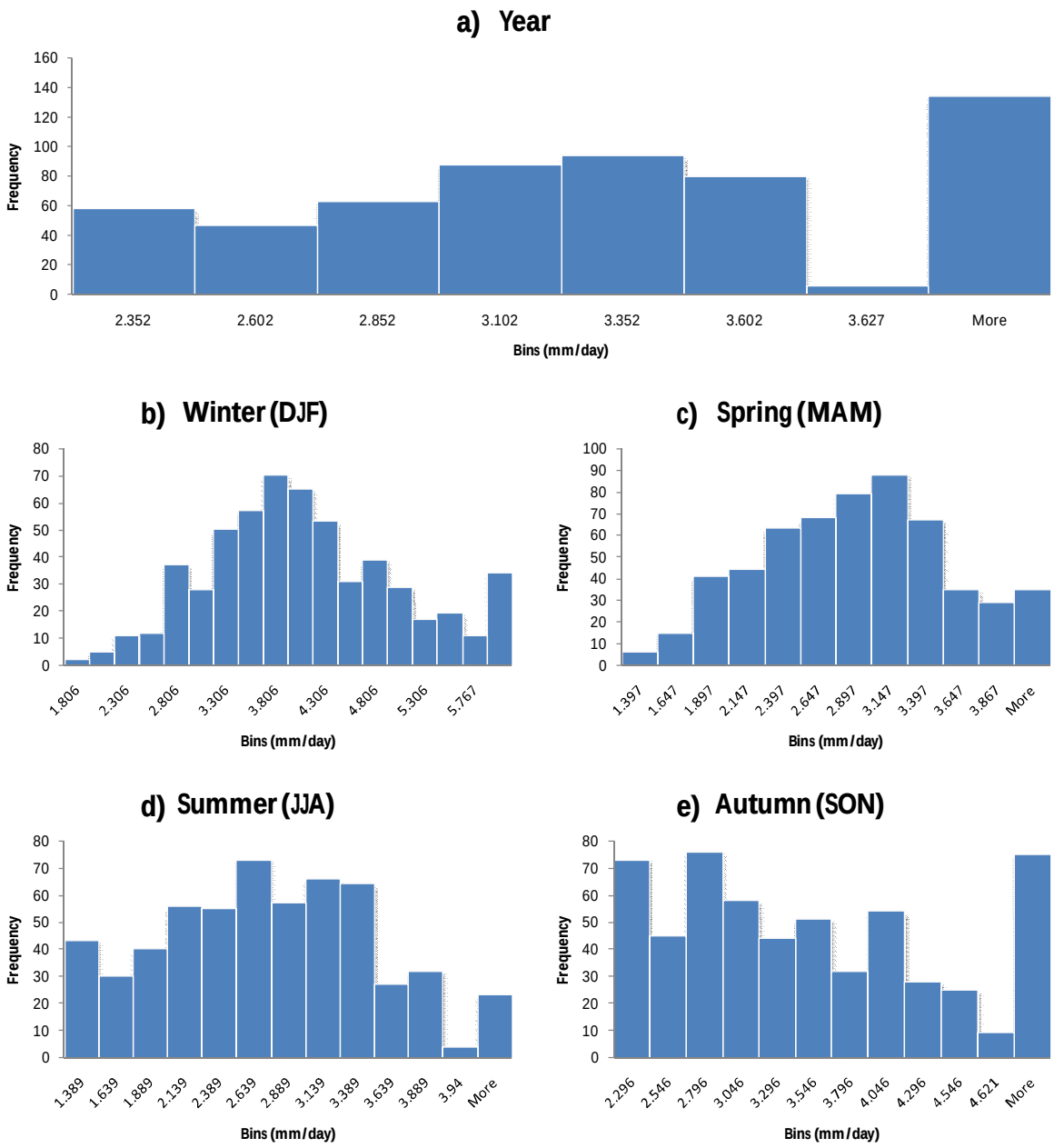


Figure 4.8: Rank histograms for interannual precipitation data. Bins are derived from observations.

A dominant feature across the majority of winter Q-Q plots is the dense clustering of central values, suggesting that the central values of the model distributions are less dispersed than the observed (Figure 4.9). Additionally, certain models overestimate values on the lower tail of their distributions. For example, RCAO-H, ARPEGE-b and ARPEGE-c all overestimate lower tail values leading to low population in the lowest bins of the histogram and extra population in the middle bins. Much greater overestimation that occurs on the upper tails of models

such as CLM, RCAO-H and RCAO-E4, resulting in low populations in the upper bins and overpopulation in the uppermost bin, with values falling beyond the observed range. Similar patterns can be observed in spring. However, in spring, the central values are more evenly dispersed along the expected line, which may account for the less pronounced shape of the rank histogram (Figure 4.10).

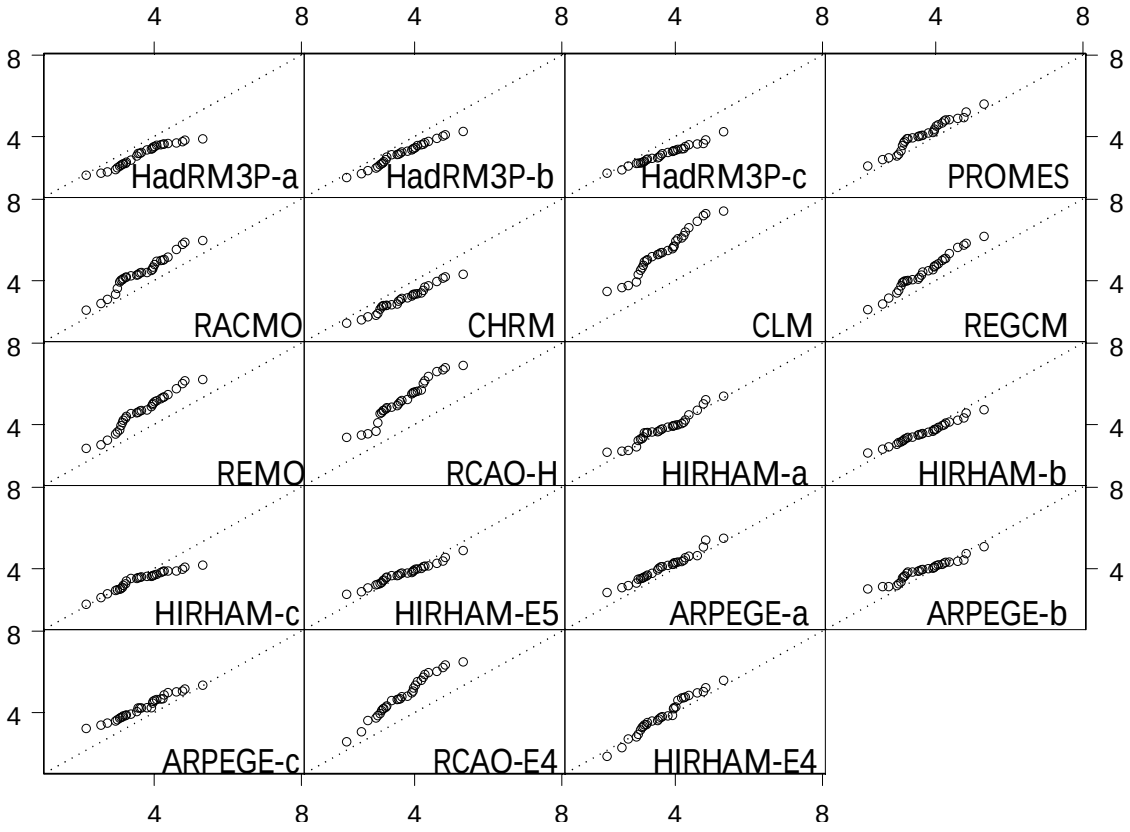


Figure 4.9: Q-Q plots of modelled versus observed interannual winter precipitation (mm/day) for the period 1961-1990.

In summer, many of the models capture the observed distribution quite well (Figure 4.11). For example, the Q-Q plots of HIRHAM-b, HIRHAM-c, HIRHAM-E5 and ARPEGE-b follow the expected line, indicating a high level of agreement between observed and modelled distributions. As such, this rank histogram is somewhat flatter than winter and spring. Underestimation on the upper tails of HadRM3P-a, HadRM3P-b and HadRM3P-c may account for the low population in the penultimate bin of the rank histogram. Conversely, the Q-Q plots of REMO,

REGCM and CLM suggest that their upper tails are more dispersed than observed, which may account for the overpopulation in the final bin of the rank histogram.

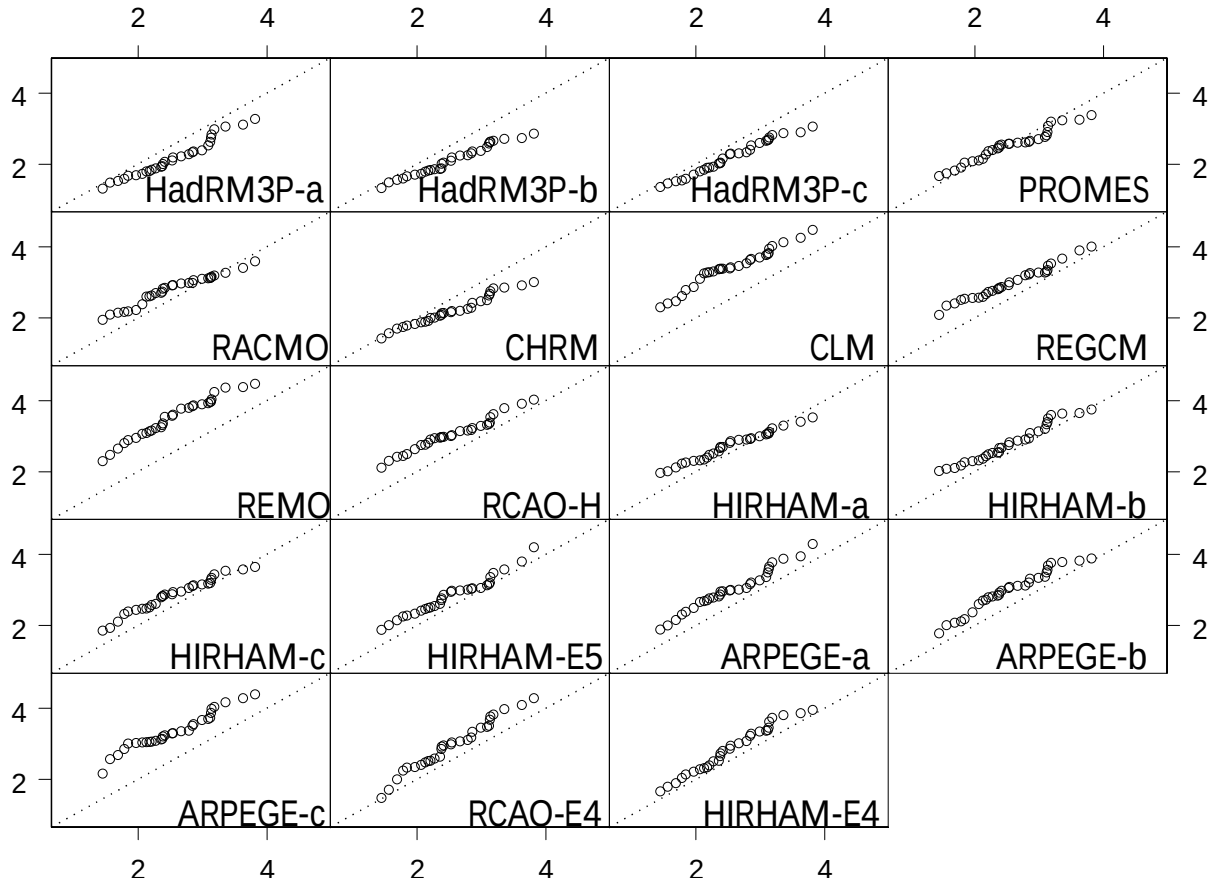


Figure 4.10: Q-Q plots of modelled versus observed interannual spring precipitation (mm/day) for the period 1961-1990.

In autumn, many of the models such as CHRM, HadRM3P-a, HadRM3P-b and ARPEGE-a display a downwards shift in their distributions, relative to the observed distribution, which may explain the somewhat asymmetric shape of the rank histogram (Figure 4.12). However, several other models display overestimation on the upper tail of their distributions, such as CLM and RCAO-H. Additionally, one model, RCAO-E4 appears to overestimate the distribution of precipitation. The combination of these error patterns may explain the overpopulation spike in the upper bin of the rank histogram.

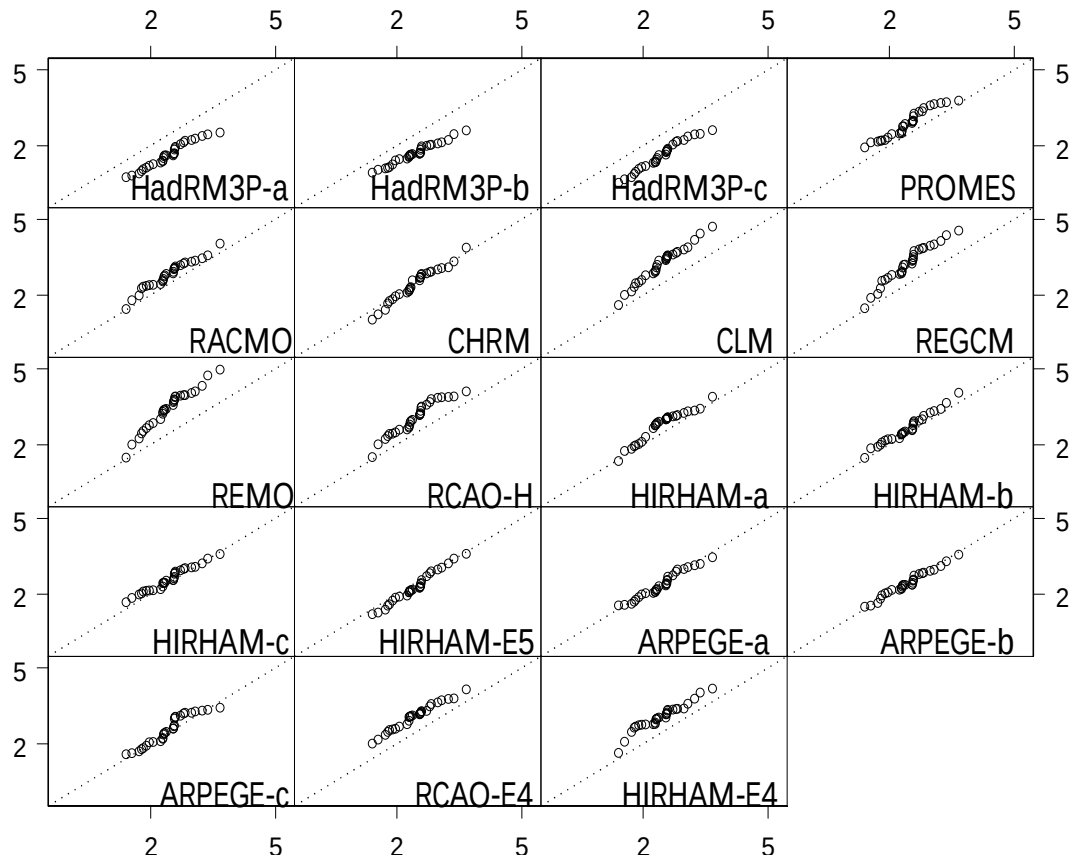


Figure 4.11: Q-Q plots of modelled versus observed interannual summer precipitation (mm/day) for the period 1961-1990.

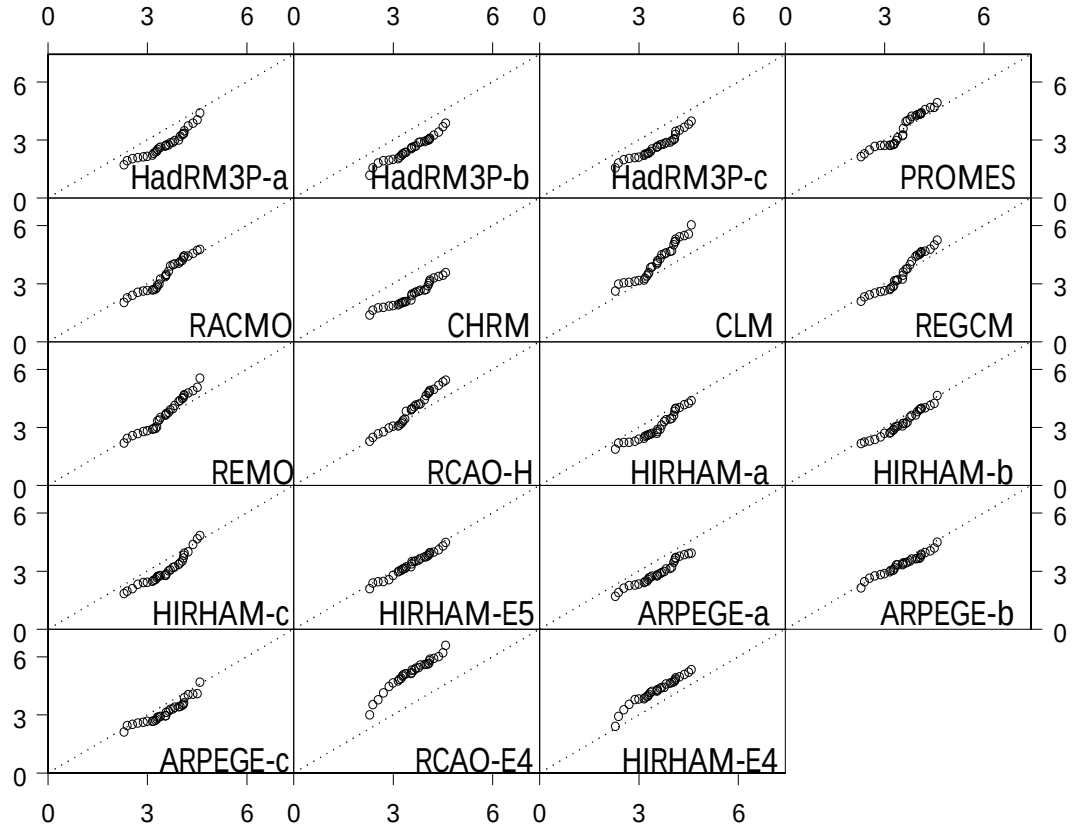


Figure 4.12: Q-Q plots of modelled versus observed interannual autumn precipitation (mm/day) for the period 1961-1990.

Skill metrics are used to quantify the biases identified through visual inspection. Results are given in Table 4.3a and b. In winter, the t-test confirms statistically significant differences in means in 11 out of 19 models. However, only five are found to have a significantly different standard deviation to the observed. Although most of the differences in variability are not significant, it is clear that on the whole the models underestimate variability, with seven models overestimating variability compared to the twelve that underestimate it. These results agree with the interpretation of the dome-shaped rank histogram.

In spring, though no statistically significant differences in variability are identified, the biases are again mostly negative, with only two models overestimating variability. Again, several models are found to have significantly different means to the observed, with both positive and negative biases occurring. Results for summer are quite similar, though one model, REMO, is found to have a statistically significant difference in variability to the observed for this season. In autumn, four models, CLM, REGCM, REMO and RCAO-H, display a statistically significant difference in variability. Again, several models are found to have significantly different means to the observed.

4.3.6 The effect of GCM driver choice on interannual variability

Table 4.4 summarizes the seasonal percentage bias of the interannual data standard deviation and mean with respect to the observed dataset. Vertical patterns within GCM-driver groupings, indicating a bias common to all RCMs driven by the same GCM, suggest errors due to issues with the GCM output used to drive the RCMs. For example, as temperature variability in winter and summer appears to be underestimated in all the HadAM3H-driven models, this suggests an error with the HadAM3H data used to drive the models. However it is also possible that RCMs driven by the same GCM exhibit similar biases for different underlying reasons. Many processes affect modelled temperature so it is conceivable that a cloud parameterization deficiency in one model and an albedo error in another could lead to the same average error in temperature.

Winter

	HadCM3/HadAM3P			HadCM3/HadAM3H										ECHAM5	Observed SSTs			ECHAM4-OPYC	
	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
S.Dev.	0.56	0.61	0.49	0.78	0.88	0.64	1.10	0.92	0.92	1.01	0.72	0.54	0.50	0.52	0.68	0.52	0.60	1.03	0.95
Bias	-0.24	-0.19	-0.30	-0.01	0.08	-0.16	0.30	0.12	0.12	0.21	-0.08	-0.26	-0.30	-0.28	-0.12	-0.27	-0.20	0.23	0.15
% Bias	-30.2%	-23.8%	-38.1%	-1.8%	10.0%	-19.5%	38.0%	15.6%	15.4%	26.6%	-10.0%	-32.8%	-37.3%	-34.6%	-15.1%	-34.3%	-25.2%	28.5%	19.3%
F-test	0.06	0.15	**0.01**	0.92	0.61	0.25	0.09	0.44	0.45	0.21	0.57	**0.04**	**0.01**	**0.03**	0.38	**0.03**	0.12	0.18	0.35
Mean	3.07	3.15	3.09	4.06	4.39	3.07	5.41	4.37	4.62	5.17	3.80	3.57	3.42	3.71	4.02	3.88	4.22	4.78	3.84
Bias	-0.62	-0.55	-0.60	0.36	0.69	-0.63	1.71	0.67	0.93	1.47	0.10	-0.12	-0.28	0.01	0.32	0.19	0.53	1.08	0.14
% Bias	-16.8%	-14.8%	-16.3%	9.9%	18.8%	-17.0%	46.3%	18.3%	25.1%	39.8%	2.8%	-3.3%	-7.5%	0.4%	8.7%	5.0%	14.3%	29.3%	3.8%
T-test	**0.00**	**0.00**	**0.00**	0.08	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.61	0.50	0.12	0.94	0.10	0.30	**0.01**	**0.00**	0.54

Spring

	HadCM3/HadAM3P			HadCM3/HadAM3H										ECHAM5	Observed SSTs			ECHAM4-OPYC	
	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
S.Dev.	0.55	0.44	0.50	0.47	0.44	0.44	0.57	0.49	0.60	0.49	0.43	0.54	0.49	0.58	0.62	0.60	0.56	0.74	0.68
Bias	-0.08	-0.19	-0.13	-0.16	-0.19	-0.20	-0.07	-0.14	-0.03	-0.14	-0.20	-0.10	-0.14	-0.06	-0.01	-0.03	-0.07	0.10	0.04
% Bias	-13.2%	-29.9%	-20.9%	-25.3%	-29.8%	-31.1%	-10.5%	-21.9%	-5.0%	-22.8%	-31.5%	-15.5%	-22.3%	-9.1%	-1.9%	-4.5%	-11.4%	16.3%	6.7%
F-test	0.45	0.06	0.21	0.12	0.06	0.05	0.55	0.19	0.78	0.17	0.05	0.37	0.18	0.61	0.92	0.81	0.52	0.42	0.73
Mean	2.17	2.09	2.19	2.50	2.78	2.20	3.39	2.97	3.48	3.03	2.71	2.79	2.83	2.84	3.00	2.96	3.32	2.94	2.82
Bias	-0.39	-0.46	-0.37	-0.05	0.22	-0.36	0.84	0.42	0.92	0.48	0.15	0.24	0.28	0.28	0.44	0.40	0.77	0.38	0.27
% Bias	-15.2%	-18.1%	-14.4%	-2.0%	8.7%	-14.0%	32.9%	16.3%	36.2%	18.7%	6.1%	9.3%	10.8%	11.0%	17.3%	15.8%	30.2%	15.1%	10.4%
T-test	**0.02**	**0.00**	**0.02**	0.73	0.13	**0.01**	**0.00**	**0.01**	**0.00**	**0.00**	0.28	0.13	0.07	0.08	**0.01**	**0.02**	**0.00**	**0.04**	0.13

Table 4.3a: Skill metrics for interannual precipitation (Winter and Spring), for 1961-1990. Significant test values are given in bold. Grey titles represent driving GCMs/boundary data.

Summer

	HadCM3/HadAM3P			HadCM3/HadAM3H										ECHAM5	Observed SSTs			ECHAM4-OPYC	
	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
S.Dev.	0.51	0.43	0.58	0.56	0.62	0.70	0.77	0.78	0.85	0.69	0.65	0.63	0.49	0.64	0.54	0.52	0.59	0.55	0.57
Bias	-0.06	-0.14	0.02	0.00	0.05	0.13	0.20	0.21	0.28	0.13	0.08	0.06	-0.07	0.07	-0.03	-0.04	0.02	-0.02	0.00
% Bias	-9.9%	-24.6%	2.9%	-0.7%	9.4%	22.9%	35.1%	37.4%	49.4%	22.3%	14.7%	10.0%	-13.1%	12.0%	-4.6%	-7.8%	4.4%	-3.8%	-0.2%
F-test	0.58	0.13	0.88	0.97	0.63	0.27	0.11	0.09	**0.03**	0.28	0.47	0.61	0.45	0.54	0.80	0.66	0.82	0.84	0.99
Mean	1.62	1.68	1.59	2.84	2.79	2.42	3.19	3.18	3.39	3.02	2.74	2.62	2.56	2.27	2.36	2.41	2.53	3.07	3.03
Bias	-0.83	-0.78	-0.86	0.39	0.34	-0.03	0.74	0.73	0.93	0.57	0.29	0.16	0.11	-0.18	-0.09	-0.04	0.08	0.61	0.57
% Bias	-33.8%	-31.6%	-35.2%	16.0%	13.7%	-1.3%	30.0%	29.7%	38.0%	23.2%	11.7%	6.7%	4.3%	-7.4%	-3.7%	-1.6%	3.3%	25.0%	23.4%
T-test	**0.00**	**0.00**	**0.00**	**0.01**	**0.04**	0.85	**0.00**	**0.00**	**0.00**	**0.00**	0.08	0.30	0.45	0.26	0.54	0.79	0.60	**0.00**	**0.00**

Autumn

	HadCM3/HadAM3P			HadCM3/HadAM3H										ECHAM5	Observed SSTs			ECHAM4-OPYC	
	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
S.Dev.	0.70	0.66	0.63	0.87	0.84	0.62	0.99	0.95	0.93	0.92	0.74	0.70	0.80	0.63	0.64	0.57	0.66	0.87	0.72
Bias	0.07	0.03	0.00	0.24	0.22	-0.01	0.36	0.32	0.30	0.29	0.12	0.07	0.18	0.00	0.01	-0.06	0.03	0.24	0.09
% Bias	11.0%	4.8%	0.6%	38.6%	34.4%	-1.3%	56.9%	51.0%	47.3%	46.5%	18.5%	10.9%	27.9%	0.8%	1.1%	-9.1%	5.4%	37.9%	13.9%
F-test	0.58	0.80	0.98	0.08	0.12	0.95	**0.02**	**0.03**	**0.04**	**0.04**	0.37	0.58	0.19	0.97	0.95	0.61	0.78	0.09	0.49
Mean	2.79	2.54	2.67	3.51	3.45	2.41	4.16	3.58	3.71	3.89	3.03	3.25	3.05	3.32	2.87	3.33	3.15	5.11	4.21
Bias	-0.76	-1.01	-0.88	-0.04	-0.10	-1.14	0.61	0.03	0.15	0.33	-0.52	-0.30	-0.50	-0.23	-0.68	-0.22	-0.40	1.56	0.66
% Bias	-21.3%	-28.5%	-24.9%	-1.1%	-2.8%	-32.2%	17.1%	0.9%	4.4%	9.4%	-14.7%	-8.5%	-14.0%	-6.5%	-19.2%	-6.2%	-11.2%	43.9%	18.5%
T-test	**0.00**	**0.00**	**0.00**	0.85	0.61	**0.00**	**0.01**	0.88	0.46	0.11	**0.01**	0.09	**0.01**	0.17	**0.00**	0.17	**0.02**	**0.00**	**0.00**

Table 4.3b: Skill metrics for interannual precipitation (Summer and Autumn), for 1961-1990. Significant test values are given in bold. Grey titles represent driving GCMs/boundary data.

Table 4.4: Seasonal percentage bias of standard deviation (variability) and mean for interannual temperature and precipitation data.

However, sharing a GCM driver is an important commonality and where the same bias occurs across RCMs driven by the same GCM, the driving model must be considered as a source of bias. Conversely, horizontal patterns indicate errors arising due to differences in model construction. For example, HadRM3P and CHRM show a marked underestimation of precipitation quantities across the year while the other HadAM3H driven models show positive biases or smaller mixed biases. This suggests that certain processes governing precipitation amounts in the Irish domain may be represented less skilfully in HadRM3P and CHRM.

It is clear that error in the interannual variability between modelled and observed temperature is greatest in winter and summer. Biases are entirely negative in winter, mostly negative in summer, mostly positive in spring and mixed in autumn. Error in the mean of the interannual values is greater in winter than all other seasons and biases are almost all positive. The exceptions are CHRM (all seasons) and PROMES (spring and summer only).

For precipitation, biases are more mixed. Precipitation variability is underestimated in spring in all the HadAM3H-driven models. While precipitation variability is overestimated in all other seasons, it is most greatly overestimated in autumn in the majority of models. Conversely, in the two models driven by ECHAM4-OPYC, summer is the only season when underestimation occurs. In addition to the errors relation to HadRM3P and CHRM, there are two further errors to note regarding precipitation amounts. The ARPEGE and HIRHAM sub-ensembles both underestimate precipitation amounts in autumn in all three ensemble members, suggesting internal RCM error in ARPEGE and HIRHAM.

Variability of temperature is underestimated in winter and summer, overestimated in spring and biases are mixed in autumn. Conversely, variability of precipitation is underestimated in spring, overestimated in summer and autumn and biases are mixed in winter. Positive temperature biases tend to correspond to positive precipitation biases also, possibly due to the higher temperatures promoting excessive evaporation. This behaviour is seen in the majority of models, but not all. In an analysis of limited area model output over Europe, Christensen *et al.* (1996) notes that overestimation of precipitation is generally found in areas of orographic precipitation. This may be another factor to consider and will be discussed further in

the next chapter, which will examine spatial performance. Overall, these findings suggest the possibility of a link between precipitation amounts and temperature, though experimentation with the individual models would be required to establish whether this is truly a causal relationship. However there does not seem to be any link between variability of temperature and precipitation.

Lateral boundary conditions are known to significantly impact RCM output (Denis *et al.*, 2003; Wu *et al.*, 2005). Giorgi *et al.* (2004) notes that large-scale mid-tropospheric circulations in a regional model are primarily controlled by the forcing boundary conditions, and this in turn effects the surface variables. However, a skill scoring approach to model assessment does not indicate that GCM choice is the source of this error. To determine the extent to which boundary conditions influence the interannual variability of these models, interannual values from models with the same drivers are plotted together for temperature (Figure 4.13) and precipitation (Figure 4.14). Although the GCM-driven models will not simulate climate events in the same chronological order as is observed, if GCM input is a key factor then RCMs with the same driver should have similar patterns of interannual variability.

Only models driven by HadAM3H and ECHAM4-OPYC are included in this analysis as those models which are part of a perturbed sub-ensemble cannot be compared in this manner, due to the small differences in driving conditions which are introduced to create the different sub-ensemble members. Even the members of the ARPEGE sub-ensemble are very different to both each other and the observations. One would expect these models to come closest to matching the observed interannual variability as they are driven by observed SSTs, but they do not. However, even between observed datasets there can be differences (Betts *et al.*, 2006; Grotjahn, 2008; Ma *et al.*, 2008). While the observations on which they are based may be the same, different interpolation techniques could be used to fit observations to the grid used by the model. While the resulting difference in the final dataset could be very small, it has been shown by Chu (1999) and Collins and Allen (2002) that even seemingly small perturbations to boundary conditions can lead to significantly different projections.

In winter, almost all models driven by HadAM3H display very similar patterns for interannual temperature. This suggests that driving information is the

predominant influence in determining both the variability and magnitude of modelled winter temperatures. There is one exception, CHRM, which displays a similar pattern of interannual evolution to the other models but a different magnitude of temperatures. This error is still small relative to the observed temperature, and hence was not noticeable in the seasonal analysis. However, it does suggest that while variability in CHRM is governed by the boundary conditions supplied by HadAM3H, there is a systematic bias in winter that is unique to this model. Similarly, autumn temperatures appear to be predominantly influenced by boundary conditions in the HadAM3H-driven models, with the exception of CHRM. In spring and summer, the general pattern of interannual variability is similar in these models, but results do not fall in as tight a range as for winter and autumn. This suggests that while boundary conditions are still important in spring and summer, differences in internal regional model physics becomes more influential than in autumn and winter.

All of the single experiments share these characteristics, as well as one experiment from the HIRHAM sub-ensembles driven by HadAM3H. This suggests that the interannual variability of temperature in these models is quite sensitive to boundary conditions, as even the small perturbations applied to create the sub-ensembles result in very different results. As there is only one model driven by ECHAM5 and only a sub-ensemble of one model driven by HadRM3H, no conclusions can be drawn about the behaviour of these simulations.

Conversely, the two models driven by ECHAM4-OPYC agree with each other the most in summer. While the evolution of the interannual pattern is similar in all seasons, the magnitudes of temperature agree most in summer and least in winter. This suggests that while the lateral boundary conditions are highly influential for the ECHAM4-OPYC-driven models, other factors must also be considered. As there are only two models driven by this GCM, it is more difficult to draw conclusions about them. RCAO-H and HIRHAM-a driven by HadAM3H did not display any great differences to the other HadAM3H models, which would suggest that internal RCM differences are not the source of this discrepancy.

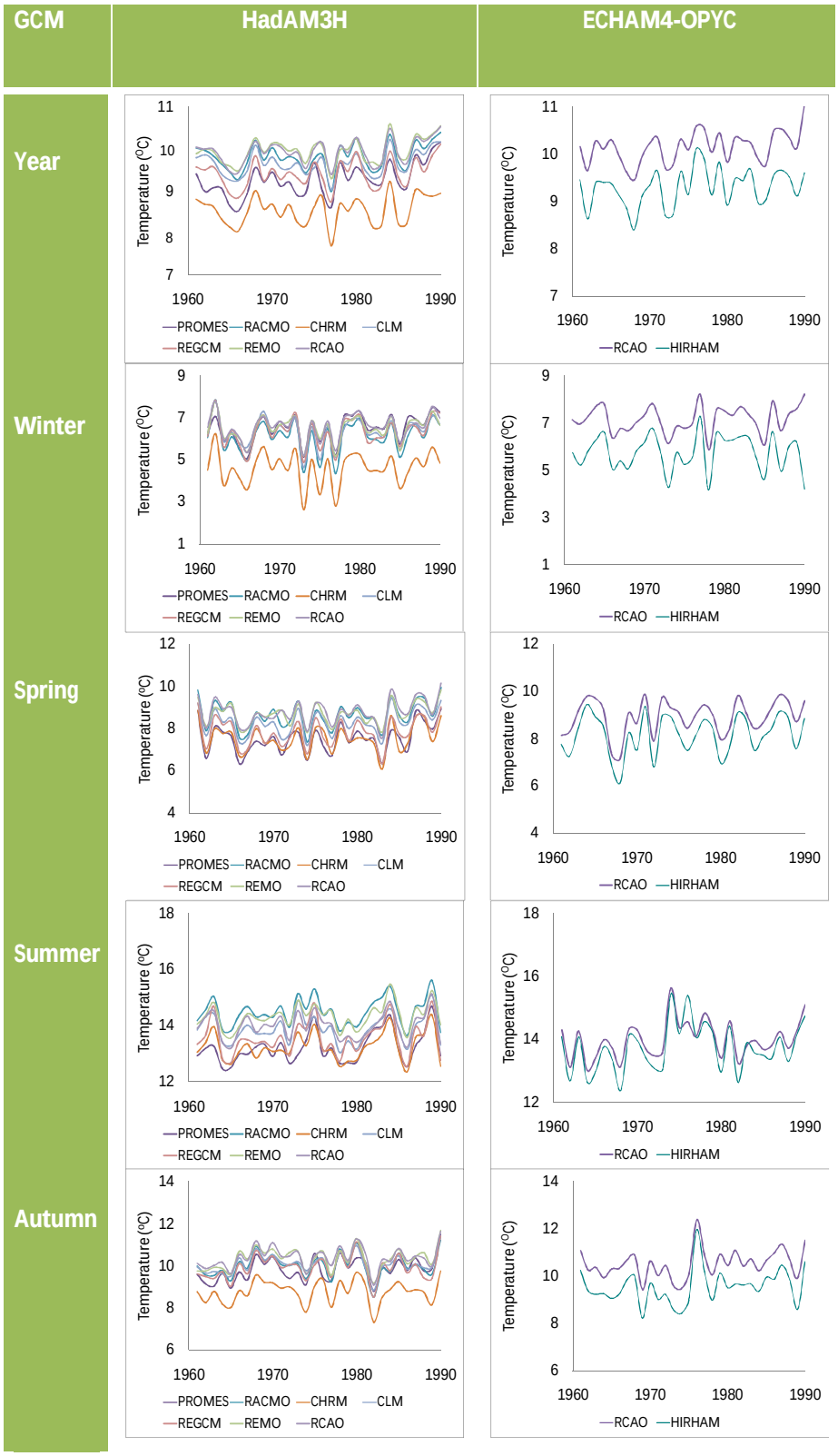


Figure 4.13: Interannual evolution of temperature over 1961-1990, categorized by driving GCM.

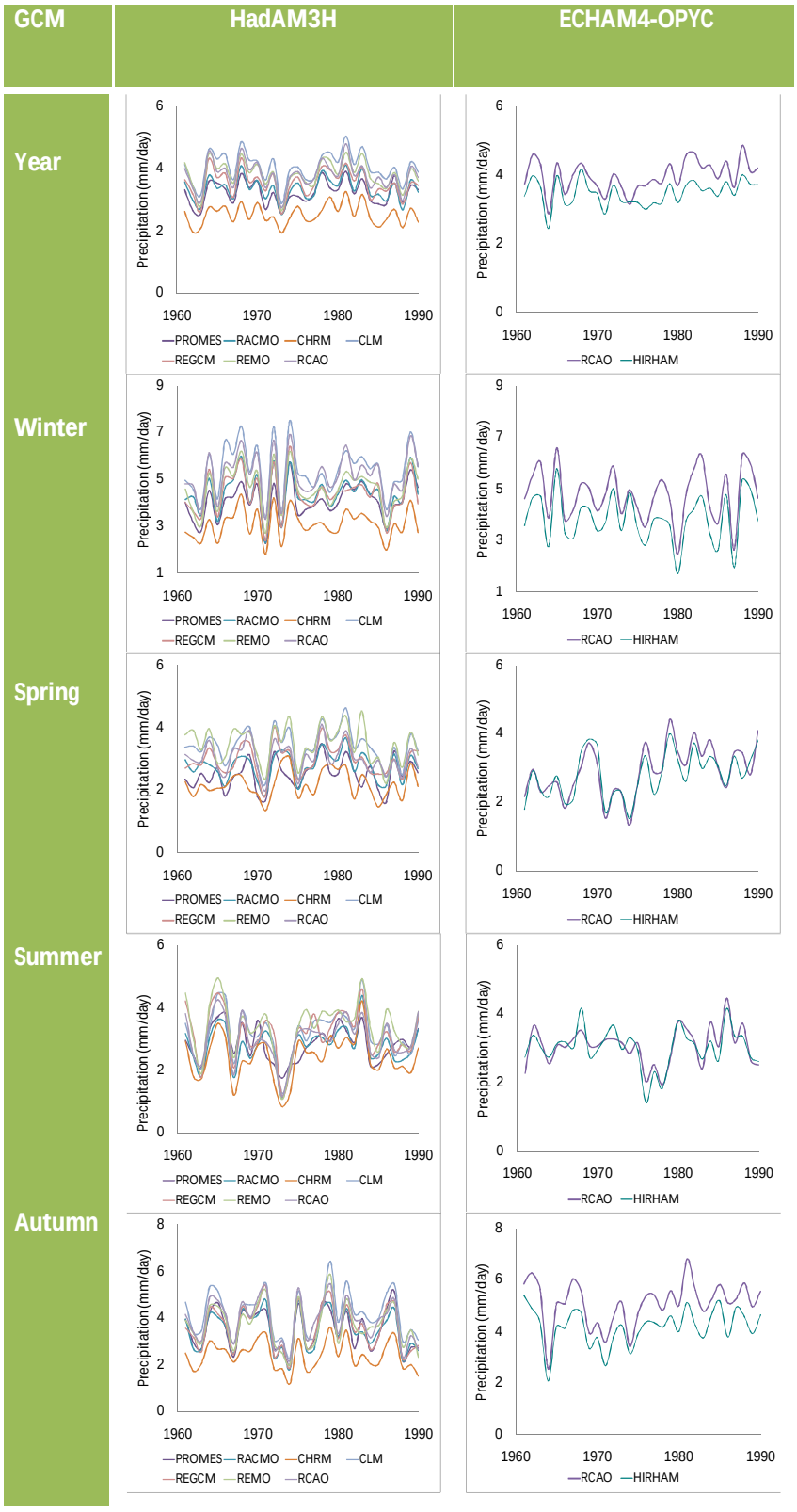


Figure 4.14: Interannual evolution of precipitation over 1961-1990, categorized by driving GCM.

But the GCM driver is the same, therefore there are no differences in boundary conditions to account for a systematic difference between the RCMs.

However, the error of a particular GCM-RCM combination is not just composed of GCM error and RCM error. There is also a GCM-RCM interaction component, which could account for the results obtained here. Ferro (2004) decomposed the spatial variation of several PRUDENCE experiments into its component parts and found that GCM-RCM interaction was a significant component at a number of gridpoints within the Irish domain.

The results obtained when the same models are plotted together for precipitation are slightly different to those obtained for temperature. With respect to the yearly, winter and autumn values, CHRM shares the same interannual pattern of evolution as the other HadAM3H-driven models, but is systematically biased. This suggests that while variability in CHRM is controlled by the boundary conditions supplied by the GCM, precipitation amounts are influenced by some characteristic of the RCM itself. The range of the models is greatest in winter and smallest in summer for both the HadAM3H and ECHAM4-OPYC driven models. Overall, results suggest that like temperature, precipitation variability in the various RCMs is significantly influenced by the lateral boundary conditions supplied by the GCM.

4.4 MEAN ANNUAL CLIMATOLOGY

4.4.1 Review of methods

The next consideration in this analysis is how well the models represent means and variations across the year. For the purpose of this analysis, seasons are defined as: December, January, February (DJF); March, April, May (MAM); June, July, August (JJA) and September, October, November (SON). The climatological year is calculated by averaging the 30 January datasets from the 30 years of data, the 30 February datasets and so on to produce a 30-year average for each month, which forms a series of twelve average values.

Murphy (1993) notes that traditionally, forecast verification methods have been "measure-oriented", focusing on quantifying model bias. Mean square error, which penalizes larger errors more heavily than smaller ones, and mean absolute error are cited as examples of this approach. However, while both of these metrics are quite widely used, they do not give any indication of the direction of model

biases, only the magnitude. Therefore, if both magnitude and direction of model errors need to be quantified, mean percentage error would be a more appropriate choice. Though lesser used, this metric gives a fuller picture of model error.

In addition to bias, association is an important skill at the mean annual scale and the correlation coefficient can be used to quantify the level of similarity between the observed and modelled climatologies. Zheng and Frederiksen (1998) note that the correlation coefficient can be a misleading indicator of skill at the interannual scale, as due to the chaotic nature of the climate system, even small differences in initial conditions can lead to very different climate projections. However, unlike interannual datasets, one can compare observed and modelled mean climatologies directly. As the comparison is between mean values over 30 years rather than absolute values for a single year, the models should simulate the annual cycle skilfully. For example, Barnston *et al.* (1999) uses temporal correlation along with root-mean-square error to quantify model skill in ENSO prediction, by measuring the level of association between model forecasts and observed El Niño episodes

A number of different coefficients are used for different situations. In a modelled annual climatology outliers are highly undesirable as they distort the annual pattern. As discussed, in this instance, the Pearson correlation coefficient is chosen as due to its sensitivity to outliers, it weights models with outliers as less skilful than those without outliers. Additionally, the Pearson coefficient specifically tests a linear relationship while the Spearman rho is a test of monotonic association. Two identical datasets will clearly have the highest degree of linear dependency. If a model simulates the climatological year with a high degree of skill, there should be minimal departure between the modelled output from the climatological year as calculated using observed data. Therefore the optimum metric is the coefficient which tests linear association, another reason why the Pearson coefficient is better suited to this analysis. Examples of the Pearson coefficient being applied to climate data include Cohen and Fletcher (2007), who used the Pearson coefficient to measure association between observed winter climate and hindcasts from a statistical model and Lal *et al.* (2007), who applied the Pearson coefficient to observed and simulated 10-year time series for temperature and precipitation in Fiji.

However, this metric is insensitive to differences in mean and variance. A model might consistently overestimate temperature or precipitation throughout the year, but as long as the annual cycle of the modelled variable mirrors the pattern of the observational climatological year, the model would score highly using the correlation coefficient as a metric. Therefore, another validation statistic is required. The Nash-Sutcliffe model efficiency coefficient, originally developed to assess the predictive power of hydrological models, can also be used as a validation statistic for climate models (Nash and Sutcliffe, 1970). An example of its application to non-hydrological data is Chen *et al.* (2004), who used the efficiency coefficient as a measure of skill for global radiation models. It has also been used as a regional climate model skill metric. For example, Evans *et al.* (2004) used the Nash-Sutcliffe efficiency coefficient to quantify the skill of the regional model RegCM at modelling temperature and precipitation in the Middle East, using observational data for comparison.

Significance of the efficiency coefficient for efficiencies greater than zero is also calculated. As computed values of the efficiency coefficient are sample values, the underlying population may in fact have a different true distribution value. Hypothesis testing is therefore carried out to determine the likelihood of a calculated efficiency coefficient having come from a population whose true efficiency is 0.5 or higher (McCuen *et al.*, 2006).

Possible outcomes for the annual climatology metrics are summarized in Table 4.5. The optimum outcome would be high correlation and efficiency scores, accompanied by low model bias. This would indicate a model with little systematic error that also simulates the annual trend skilfully. Failing that, high correlation accompanied by lower efficiency and high model bias scores would indicate a model that simulates the annual trend well but has a systematic bias. The final possibility is a low correlation and efficiency scores accompanied by a low model bias. A low bias score is not useful in the absence of a high correlation score. This indicates that the model overestimates or underestimates the magnitude of the variable by a different amount each month, resulting in a low average bias but distorting the annual trend so that it does not represent the observed pattern.

	High bias ***(low efficiency)***	***Low bias*** ***(high efficiency)***
High correlation	Good representation of annual pattern, with systematic bias	Optimum result: Good representation of annual pattern, with no systematic bias
Low correlation	Least desirable result: Poor representation of annual pattern, and systematic bias	Poor representation of annual pattern, but no systematic bias

Table 4.5: Outcomes for correlation/efficiency tests

4.4.2 Methodology

The second part of the assessment examines the mean monthly time series of climatological means for each model. Modelled mean annual climatologies were compared to the observed mean annual climatology using four validation statistics: bias, correlation, efficiency and trend. In addition to identifying any model errors or systematic biases using the mean squared error (MSE) metric, two "goodness of fit" metrics are applied, namely the Pearson correlation coefficient and the Nash-Sutcliffe efficiency. The Pearson correlation coefficient is sensitive to extreme values or outliers in the data while the Nash-Sutcliffe efficiency is also sensitive to differences in mean and variance. The significance of differences between the modelled and observed climatologies is also calculated.

The Pearson correlation coefficient is chosen as a metric of association as it is known to be highly sensitive to outliers in the data. The Pearson correlation coefficient is less suited to identifying unknown relationships between smaller datasets, due to the possibility of a high r value occurring by chance. However, in this case the relationship between the datasets is known, a covarying relationship is expected between the observed and modelled data and the Pearson correlation coefficient is used to quantify the known relationship rather than identify an unknown relationship. Additionally, the significance of all r values is tested to ensure

that high r values are truly indicative of covariation. For the annual climatology of a particular climatic variable *x*, the Pearson coefficient is calculated as:

Equation 4.4: Pearson correlation coefficient

$$r = \frac{\frac{1}{T}\sum_{t=1}^{T}\left(x_o^t - \overline{x_o}\right)\left(x_m^t - \overline{x_m}\right)}{\sigma_o \cdot \sigma_m}$$

where,

σ_o *= standard deviation of the observed data,*

σ_m *= standard deviation of the modelled data,*

$\overline{x_o}$ *= mean of the observed data,*

$\overline{x_m}$ *= mean of the modelled data,*

T = number of entries in the datasets,

x_o^t *= observed data at time t and*

x_m^t *= model output at time t.*

Pearson correlation coefficients range from −1 to 1, with a perfect score of 1. A score of 0 indicates that there is no linear relationship between the observations and the modelled data, while a negative score indicates a decreasing linear relationship.

Significance of the correlation is determined by the result of an F-test and the associated P value for the F-test is given. For a confidence level of 95%, if $P<0.05$, then the null hypothesis (that there is no statistically significant association between observed and modelled data) is rejected. Conversely, if $P >0.05$, then the null hypothesis is accepted.

Nash–Sutcliffe efficiencies can range from $-\infty$ to 1, with a perfect score of 1. An efficiency score less than zero occurs when the residual variance is larger than the data variance and is indicative of a very deficient model. The Nash-Sutcliffe coefficient is calculated as:

Equation 4.5: Nash-Sutcliffe coefficient

$$E = 1 - \frac{\sum_{t=1}^{T}\left(x_o^t - x_m^t\right)^2}{\sum_{t=1}^{T}\left(x_o^t - \overline{x_o}\right)^2}$$

where

$\overline{x_o}$ *= mean of the observed data,*

T = number of entries in the datasets,

x_o^t *= observed data at time t and*

x_m^t *= model output at time t.*

Significance of the efficiency coefficient is calculated for efficiencies greater than zero. A series of equations is used to transform the theoretical distribution of the efficiency index to a normal distribution. From McCuen *et al.* (2006)

Equation 4.6: Significance of the Nash-Sutcliffe coefficient

$$z = \frac{(\varepsilon - m_\varepsilon)}{S_\varepsilon}$$

where

$\varepsilon = 0.5 \ln_e\left(\frac{1 + E_f^{0.5}}{1 - E_f^{0.5}}\right)$ *and* $E_f^{0.5} =$ *computed efficiency index,*

$m_\varepsilon = 0.5 \ln_e\left(\frac{1 + \varepsilon_0^{0.5}}{1 - \varepsilon_0^{0.5}}\right)$ *and* $\varepsilon_0^{0.5} =$ *underlying population value,*

$S_\varepsilon = (n-3)^{-0.5}$ *and*

n = sample size

The standard normal distribution table can then be used to compare the z-score to a critical value. For a confidence level of 95%, the critical value for the normal distribution is 2.576. Therefore, when z is greater than 2.576, P is less than 0.05. The null hypothesis (that the calculated efficiency coefficient is based on a population whose true efficiency is 0.5 or less) is rejected and the efficiency

coefficient is considered significant. Conversely, if z<2.576 then P>0.05, and the null hypothesis is accepted.

One final analysis is carried out to further characterize fluctuations in model skill across the year. Bias is represented using the mean percentage error (MPE). Mean percentage error is chosen over mean absolute error or mean squared error as it maintains the sign of the bias. Annual mean percentage error E_A of each model is calculated as

Equation 4.7: Mean percentage error

$$E_A = \frac{1}{12}\sum_{t=1}^{12}\left(\frac{x_m^t - x_o^t}{x_o^t}\right)$$

where

x_o^t *= observed data at time t and*

x_m^t *= model output at time t.*

MPE of zero means that the model is very skilful. While values of mean error other than zero are in themselves meaningless, they are useful for comparative purposes. A seasonal breakdown of contribution to MPE is also assessed. To calculate contribution to annual MPE for a particular season, all other seasons are assumed to have perfect skill. Therefore the seasonal contribution E_S is calculated as:

Equation 4.8: To determine the seasonal contribution to MPE

$$E_S = \frac{1}{12}\sum_{t=1}^{3}\left(\frac{x_m^t - x_o^t}{x_o^t}\right)$$

where

x_o^t *= observed data at time t and*

x_m^t *= model output at time t.*

From the seasonal analysis, models with seasonal biases can be sorted from those with random biases. A model that is biased in a particular season alone has the potential to provide useful information about the rest of the annual cycle. However, a model with biases that do not follow a set pattern is less useful as the source of those types of biases are harder to diagnose.

4.4.3 Results: Temperature

An initial analysis of the annual climatology data, the ensemble climatology plot, indicates the skill of the overall ensemble of models at capturing the annual cycles of temperature and precipitation. However, ensembles can give skilful results for the wrong reasons if their skill is a by-product of bias cancellation rather than real predictive skill. Therefore it is also necessary to look at the climatology plots of the individual ensemble members to gauge how well they simulate the observed mean annual pattern. Additionally, skill metrics are calculated to assess the models' abilities at simulating these patterns.

The mean annual climatology plot for temperature indicates that the ensemble members simulate the observed annual pattern well, except in winter months (Figure 4.15).

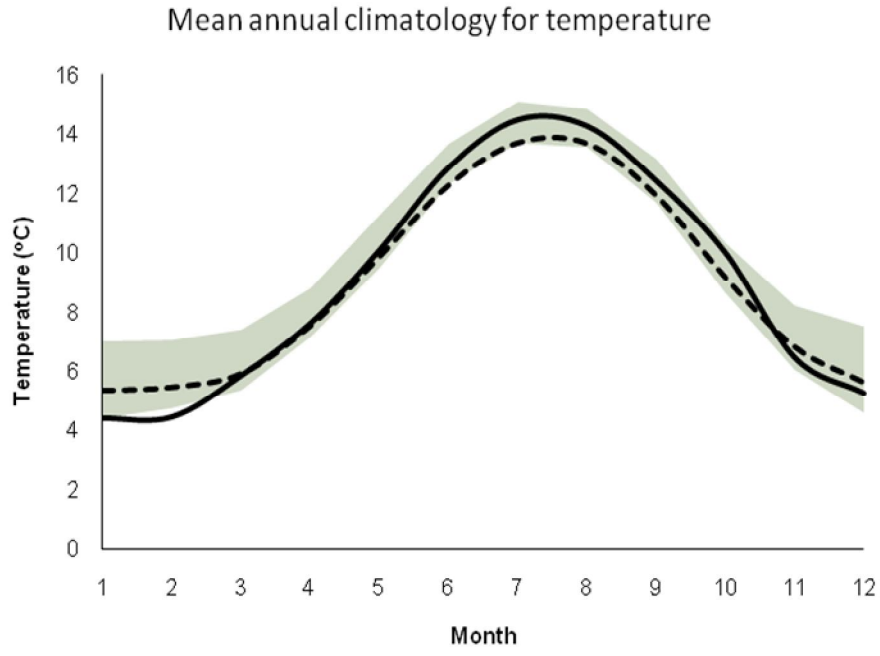


Figure 4.15: Ensemble mean annual climatology for temperature (dashed line), including ensemble range (green) and observed mean annual climatology (solid line) for 1961-1990.

The mean annual temperature follows a typical cycle, cooler in winter and warmer in summer, and the models all reflect this. From November until March, the models tend to overestimate monthly temperature values. The bias is as high as 2^{o}C in some ensemble members and is most noticeable in February, when all models overestimate the observed temperature. It is also notable that from March to May, the ensemble average almost perfectly corresponds to the observed climatology, yet the ensemble member range indicates both overestimation and underestimation amongst the individual models. This behaviour has two possible causes. One possibility is that the majority of models simulate the observed pattern for springtime well, and the range is caused by a small number of outliers. The other possibility is that the majority of models are biased and these biases cancel each other out to give a favourable average. The individual ensemble members must be examined in more detail to determine which is the case.

Next, mean annual climatology plots are generated for individual models, with the observed climatology included for reference (Figure 4.16). As indicated by the ensemble climatology, most of the models overestimate temperature in winter. This lack of skill in winter greatly contrasts with the high level of skill the models possess throughout the rest of the year. It is most noticeable for RCAO-E4, driven by ECHAM4-OPYC. This model simulates summer temperatures with high accuracy, but starting in late autumn it becomes increasingly biased, peaking with a maximum bias of almost 3^{o}C in January. The bias becomes smaller again over spring.

The three ARPEGE experiments, which are all driven by observed SSTs, perform markedly better in winter than many of the other models. Only very small biases occur in all three simulations. The greater accuracy and realism of the driving information for these models may be a reason for their higher levels of skill. However, a small minority of GCM-driven models in this ensemble also demonstrate skill in winter. HIRHAM driven by ECHAM5 simulates the observed climatology almost perfectly, with very minimal bias. The skill of this simulation may indicate that ECHAM5 provides more skilful driving information that the other GCMs, but as only one model in the ensemble is driven by ECHAM5, this cannot be confirmed. However, the third member of the HIRHAM ensemble, driven by HadAM3H, also provides very skilful output. Experiments in this sub-ensemble are driven by

perturbed data from the same GCM, and the other members of this ensemble exhibit the same winter bias as the majority of the models. It is possible that the perturbations applied for the third simulation resulted in more realistic driving data, or that the different outcomes represent the range of internal model variability for this model.

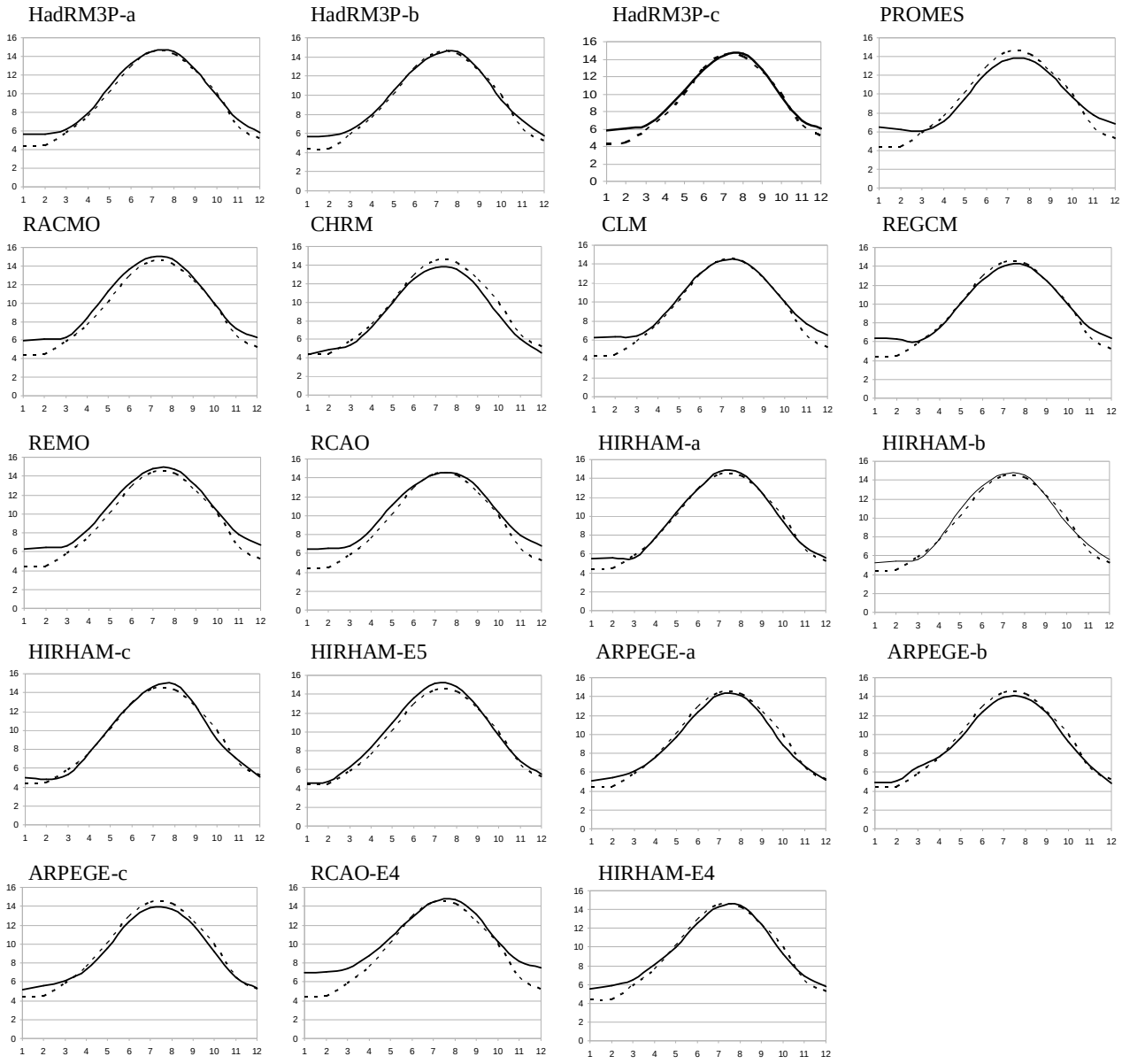


Figure 4.16: RCM mean annual climatology for temperature for 1961-1990. Observations are represented by the dotted line. Months are displayed on the x-axis and temperature in ^{o}C is displayed on the y-axis.

The agreement between the experiments in the ARPEGE ensemble could correspond to a lower level of internal variability in the ARPEGE model. CHRM, which is also driven by HadAM3H, simulates winter temperatures more skilfully than its peers, but unlike the other models it underestimates temperature during the rest of the year by approximately 1^{o}C. It is difficult to attribute the winter skill of this model. Multiple RCMs can interact with the same GCM in different ways, and the overall response is a combination of GCM response, RCM response and interactive effects (Kaufman *et al.*, 2008). The skill of these particular simulations may be a result of their unique GCM-RCM interaction.

Skill is quantified using the skill metrics outlined previously and are summarized in Table 4.6. Significant correlations are marked in bold. The optimum outcome is a high correlation and efficiency score, accompanied by low model error. Failing that, high correlation accompanied by lower efficiency and high model error scores is preferable to a low correlation and efficiency scores accompanied by a low model error.

All the models have near-perfect Pearson correlation values of 0.98 or higher. This indicates that the models simulate the annual pattern of temperature well, and that there are no outliers in the data. The Nash-Sutcliffe scores are also high, though not as high as the Pearson scores. Nash-Sutcliffe scores range from 0.84 to 0.99, which suggests that while the pattern may be well-represented, there is a bias in magnitude in some models. The bias, quantified using the mean absolute error, is greatest in the models with the lowest Nash-Sutcliffe scores.

The most notable biases are RCAO driven by ECHAM4-OPYC, which has a Nash-Sutcliffe score of 0.84, RCAO driven by HadAM3H which has a Nash-Sutcliffe score of 0.90 and REMO driven by HadAM3H which has a Nash-Sutcliffe score of 0.91. These Nash-Sutcliffe scores correspond to annual mean percentage errors of 20.55%, 16.27%, and 15.77% respectively. However, the largest portion of this error stems from the winter months. Mean percentage error contributions from spring, summer and autumn are much smaller.

Temperature

	HadCM3/HadAM3P			HadCM3/HadAM3H										ECHAM5	Observed SSTs			ECHAM4-OPYC	
	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
r	**1.00**	**0.99**	**0.99**	**0.98**	**0.99**	**0.99**	**0.99**	**0.99**	**1.00**	**1.00**	**0.99**	**0.99**	**0.99**	**1.00**	**0.99**	**1.00**	**0.99**	**0.99**	**0.99**
E	**0.97**	**0.97**	**0.96**	**0.91**	**0.93**	**0.97**	**0.93**	**0.94**	**0.91**	**0.90**	**0.98**	**0.98**	**0.99**	**0.98**	**0.98**	**0.98**	**0.97**	**0.84**	**0.97**
% Error																			
Year	8.25%	7.99%	9.49%	9.09%	12.76%	-4.53%	12.39%	9.82%	15.77%	16.27%	5.22%	5.30%	1.46%	5.07%	1.45%	0.91%	1.47%	20.55%	6.94%
DJF	5.54%	5.63%	7.05%	9.83%	7.80%	-0.24%	9.03%	9.00%	9.77%	10.26%	4.82%	4.05%	1.67%	1.39%	2.96%	1.55%	3.78%	13.38%	5.75%
MAM	1.31%	1.45%	1.56%	-0.77%	2.52%	-1.00%	1.58%	0.10%	2.84%	3.25%	0.05%	0.32%	-0.48%	2.11%	0.02%	0.60%	-0.43%	3.96%	1.36%
JJA	0.08%	-0.02%	0.18%	-1.17%	1.11%	-1.13%	0.03%	-0.57%	0.84%	0.22%	0.26%	0.46%	0.44%	1.09%	-0.59%	-0.89%	-1.04%	0.21%	-0.27%
SON	-2.68%	0.94%	0.69%	1.20%	1.32%	-2.17%	1.75%	1.30%	2.32%	2.54%	0.08%	0.47%	-0.17%	0.48%	-0.95%	-0.34%	-0.84%	2.99%	0.09%

Table 4.6: Annual climatology skill metrics for temperature for 1961-1990e, where r denotes Pearson correlation coefficient and E denotes Nash-Sutcliffe efficiency index. Significant values of r and E are marked in bold print.

4.4.4 Results: Precipitation

The mean annual climatology plot for precipitation indicates that the ensemble members are less skilful at simulating this parameter (Figure 4.17). The range of the individual ensemble members is quite large throughout the year. The model range reaches 3 mm/day in September, corresponding to a range of 55.9 mm to 145.8 mm total precipitation for the month. Conversely, observed precipitation lies between 2 mm/day and 3.8 mm/day, a range of just 1.8 mm/day, for the entire year. This suggests that intermodel variability is much greater than the annual variability of precipitation. Additionally, while the ensemble average compares favourably with the observations in terms of magnitude, it follows a much smoother pattern than is observed. While the general trend of decreasing precipitation from January to July and increasing precipitation from July to December is identifiable, the ensemble average does not capture in detail the differences in precipitation from month to month. Again, closer examination of the individual ensemble members is required to determine if the skill of the ensemble can be improved.

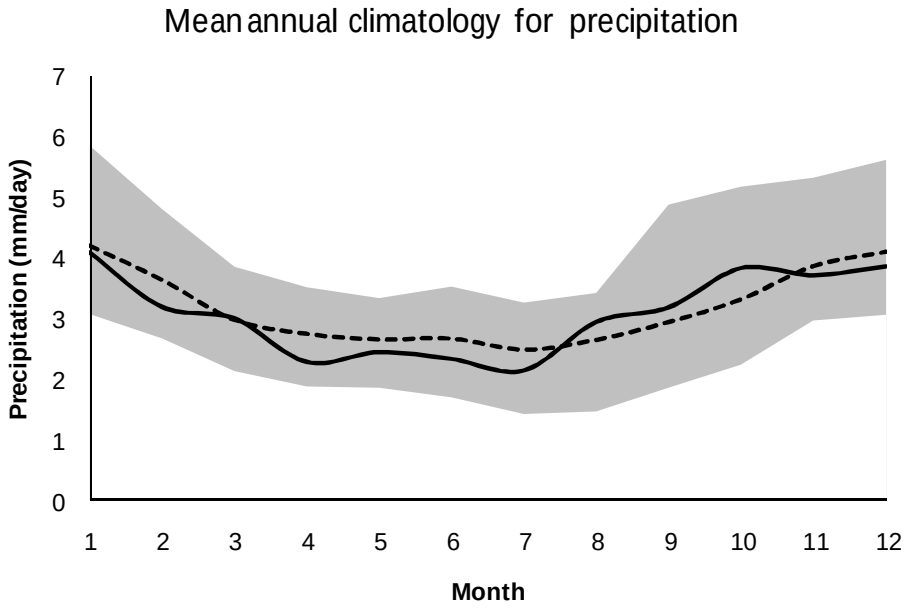


Figure 4.17: Ensemble mean annual climatology for precipitation (dashed line), including ensemble range (grey) and observed mean annual climatology (solid line) for 1961-1990.

The observed annual pattern for precipitation is characterized by a general decrescendo-crescendo pattern. However, there is a lot of month-to-month variation in the precipitation pattern which the models have difficulty capturing (Figure 4.18).

RACMO and the ARPEGE sub-ensemble simulate a much smoother pattern than observed, which captures the overall trend but not the month-to-month variation. Several models capture the general trend but model specific features at the wrong time. For example, HadRM3P-a, HadRM3P-b and HadRM3P-c model August, not July, as the driest month. Conversely, PROMES models May as the driest month, with precipitation increasing throughout the summer.

CHRM models precipitation minimums in March/April and September/October, with an increase in precipitation over summer. This pattern is quite unlike the observed pattern, with winter being the only part of the cycle which is represented somewhat skilfully. CLM, RCAO-H and RCAO-E4 overestimate precipitation throughout the year, but the greatest biases occur in autumn and winter months, possibly as a result of their warm temperature biases.

Skill varies more between models for precipitation than for temperature (Table 4.7). Nash Sutcliffe scores are very low. Out of 19 experiments, 12 have Nash-Sutcliffe scores of less than zero. This indicates that the variance of the errors in the modelled data is larger than the observed variance of the parameter. These models lack skill at representing the magnitude of precipitation.

However, the models with the lowest Nash-Sutcliffe scores have some of the highest Pearson scores. The model with the highest Pearson score is RCAO-E4, which has a very low Nash-Sutcliffe score of -3.89. This indicates that while this model simulates the annual pattern of precipitation well, it has a large systematic bias. In fact, significant association was found between all of the RCMs and the observed data, using the Pearson coefficient and associated significance test.

However, none of the Nash-Sutcliffe efficiencies were found to be significant. Considering magnitude and pattern together, the most skilful models are HIRHAM-b and ARPEGE-b. Both models have Pearson scores greater than 0.70 and positive Nash-Sutcliffe scores, though the low to moderate Nash-Sutcliffe scores indicate that there are some biases in the models.

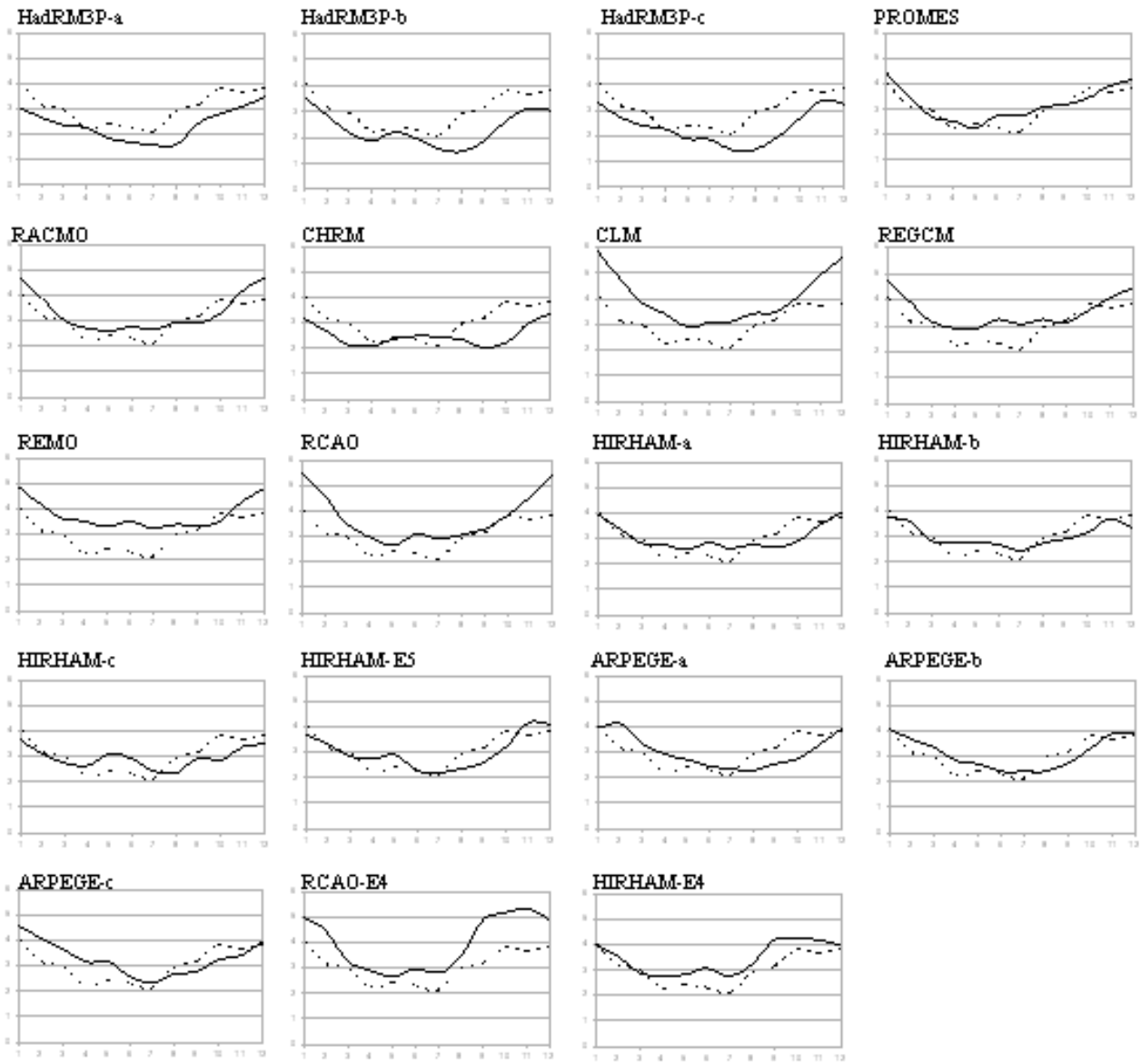


Figure 4.18: RCM mean annual climatology for precipitation for 1961-1990. Observations are represented by the dotted line. Months are displayed on the x-axis and precipitation in mm/day is displayed on the y-axis.

Precipitation

	HadCM3/HadAM3P			HadCM3/HadAM3H										ECHAM5	Observed SSTs			ECHAM4-OPYC	
	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
r	**0.87**	**0.82**	**0.83**	**0.90**	**0.86**	**0.61**	**0.87**	**0.85**	**0.76**	**0.86**	**0.78**	**0.85**	**0.70**	**0.81**	**0.64**	**0.85**	**0.72**	**0.92**	**0.87**
E	0.07	-0.18	-0.27	-0.07	0.01	-0.13	-2.98	-0.59	-1.96	-1.81	0.54	0.61	0.47	0.57	0.17	0.49	-0.06	-3.89	-0.24
% Error																			
Year	-16.84%	-18.44%	-17.97%	12.22%	16.29%	-10.65%	39.56%	23.61%	33.98%	30.34%	7.92%	7.58%	5.02%	5.59%	7.38%	9.67%	15.88%	36.36%	21.40%
DJF	-2.92%	-2.38%	-2.77%	4.10%	6.50%	-2.97%	13.82%	6.41%	8.23%	12.11%	2.30%	0.83%	-0.45%	1.69%	4.14%	2.98%	5.48%	9.41%	2.63%
MAM	-2.69%	-3.46%	-2.51%	0.78%	3.55%	-2.26%	9.85%	5.62%	10.94%	6.12%	2.99%	3.86%	4.30%	4.29%	5.76%	5.36%	9.19%	5.19%	4.13%
JJA	-7.29%	-6.61%	-7.66%	5.80%	5.24%	1.43%	9.56%	9.55%	11.85%	7.84%	4.79%	3.36%	3.00%	-0.33%	0.82%	1.29%	2.44%	8.10%	7.71%
SON	-3.94%	-5.99%	-5.03%	1.53%	1.00%	-6.85%	6.33%	2.02%	2.96%	4.26%	-2.15%	-0.47%	-1.84%	-0.06%	-3.34%	0.04%	-1.23%	13.66%	6.92%

Table 4.7: Annual climatology skill metrics for precipitation for 1961-1990, where r denotes Pearson correlation coefficient and E denotes Nash-Sutcliffe efficiency index. Significant values of r and E are marked in bold print.

4.5 DISCUSSION AND CONCLUSIONS

This chapter presented a preliminary analysis of model skill using a skill-scores approach. Interannual variability and the mean annual climatology were the focus of this analysis.

The RCMs have a tendency to overestimate interannual variability for temperature in spring and precipitation in summer and autumn. There is also a tendency to underestimate interannual variability for temperature in winter and summer, and precipitation in spring. Biases for temperature in autumn and precipitation in winter are more mixed. Interannual variability of both temperature and precipitation is found to be largely governed by the choice of GCM driver. This is expected as lateral boundary conditions are known to be a dominant factor in determining RCM interannual variability (Giorgi *et al.,* 2004). This analysis highlights the importance of considering both GCM and RCM choice in nested climate modelling experiments with care, as GCM choice is a highly significant factor in determining interannual variability. If the GCM does not capture the true range of interannual variability, then corresponding RCM simulations cannot be expected to simulate interannual variability correctly either.

An issue that was raised in previous chapters is the concept that uncertainty, when improperly accounted for in climate model scenarios, can lead to an over or underestimation of risk, which in turn leads to mal-adaptation. This analysis of interannual variability illustrates this point. There are 19 simulations included and one would expect that an ensemble based on this number of simulations would adequately capture the range of possible future outcomes. Yet with regards to interannual variability, the RCMs that share a GCM driver give such similar results that they cannot possibly be considered independent experiments. With regards to this particular aspect of the climate, the ensemble approach gives a sense that more outcomes are captured and as such, that uncertainty is reduced. Only with closer inspection does it become apparent that the RCMs with shared GCM drivers are effectively the same in terms of interannual variability. These findings suggest that the multi-model ensemble approach has the potential to be much more robust if all ensemble members are driven by a different GCM, or by perturbed versions of the same GCM where only a single GCM is available.

To identify any patterns in the model errors, seasonal errors for temperature and precipitation are calculated and graphed. Seasonal errors are assessed by comparing the seasonal contributions to total annual mean error for each model. A colour-block graph was compiled which enables comparison among GCM driver groups and between climate parameters and seasons (Figure 4.19). In general, a vertical pattern indicates a possible GCM-related error while horizontal patterns indicate errors unique to a specific RCM. The models with the greatest bias are immediately apparent in Figure 4.19 and in the majority of cases, winter error contributes most to overall error.

CHRM, which was one of the few models to simulate winter temperature skilfully, has a negative bias of almost -0.5^{o}C due to errors in its representation of summer and autumn temperature. Seasonal contribution to annual average error is again very different for precipitation than for temperature. CHRM, which displayed a notable overall cold temperature bias, models drier conditions also. CHRM's greatest underestimation of precipitation occurs in autumn, which was also the season when it underestimated temperature the most. This behaviour is to be expected, due to the link between temperature and atmospheric moisture amounts. Conversely, the HadRM3P sub-ensemble models consistently drier conditions than observed, despite exhibiting a warm temperature bias. This suggests that HadRM3P's precipitation bias is not linked to its temperature bias, but is a result of another model error. Winter bias accounted for the largest share of annual temperature bias. However, precipitation biases in most models are instead quite equal in magnitude across the year. RACMO, CLM and RCAO, both driven by HadAM3H, are the only models with a winter bias that is somewhat large relative to spring, summer and autumn bias.

Overall, the RCM simulations analyzed here are quite skilful in their representation of temperature, but less adept at representing precipitation and particularly the annual climatology of precipitation. Using skill scores it is possible to identify a number of errors in this selection of RCMs (Table 4.8). However, while analysis of the data can give some clues as to the source of the error, without closer investigation there can be no conclusive statement made about the nature of model errors or the robustness of model skill. As discussed in previous chapters, skill at representing the mean climate may not indicate skill at simulating the climate phenomena underlying the mean climate. Models which appear to simulate the

climate skilfully based on a skill-scores assessment may derive their apparent “skill” from error cancellation. Conversely, models which appear to be significantly biased may capture the dynamics of the climate system quite well. Without examining the RCM data in more detail, there is no way of knowing whether the RCMs are genuinely skilful or highly uncertain. Additional analysis is required, as formulating future scenarios without deeper investigation of the models’ skill levels would result in a high degree of uncharacterized uncertainty associated with those future scenarios.

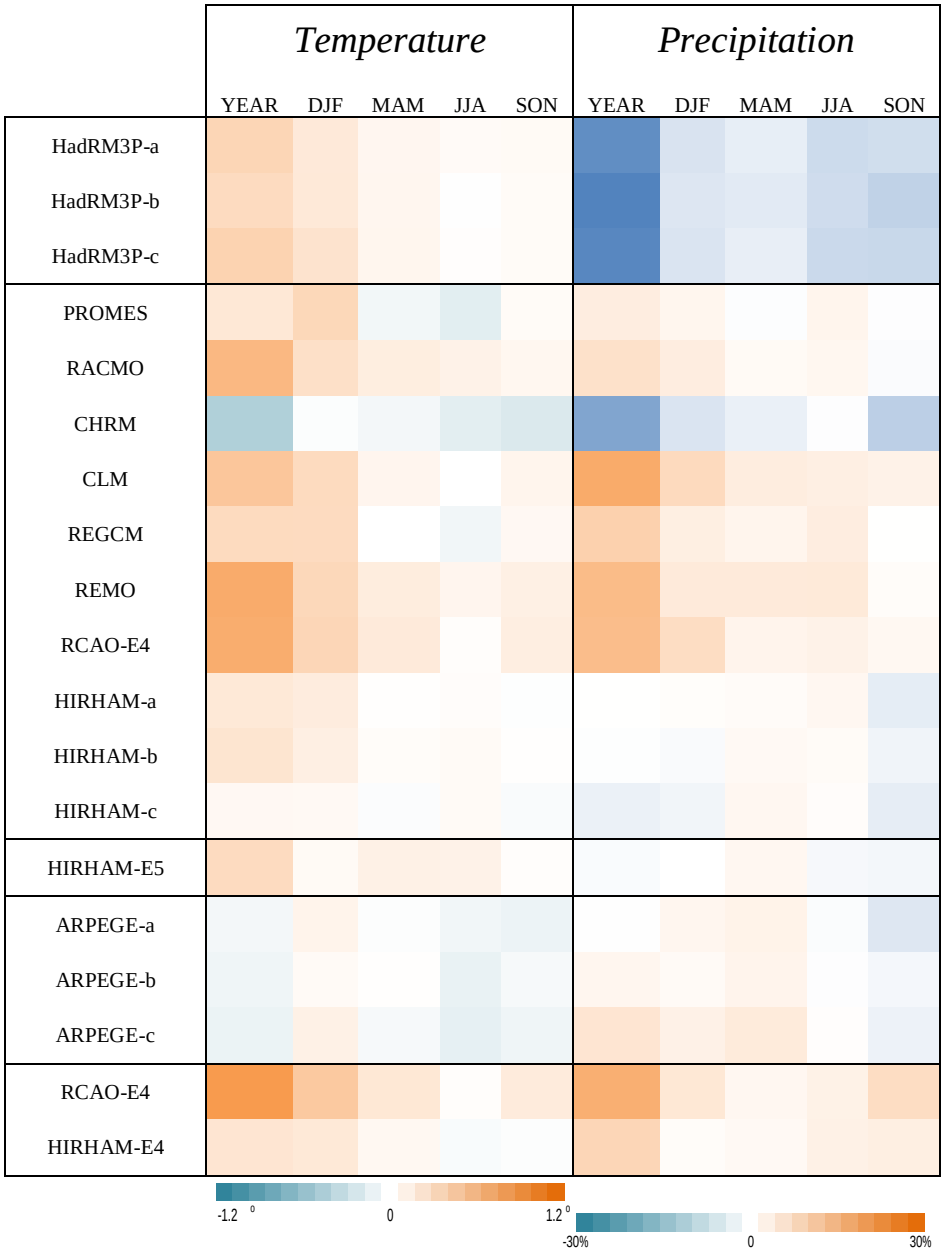


Table 4.8: Seasonal errors on mean annual climatology of temperature and precipitation for 1961-1990

While this skill score analysis has identified a number of model errors, further research is required to help explain how these errors arise and to identify models which are genuinely skilful. The next step is an analysis of the simulated spatial patterns, to determine how skilful the models are at capturing the dominant spatial patterns of temperature and precipitation. As different patterns are influenced by different factors, this will help to identify sources of model errors. A spatial analysis will also indicate how well the models represent the regional details that arise due to differences in topography, such as orographic precipitation. With greater knowledge of how the RCMs perform, there is greater potential to account for the uncertainty associated with their outputs, which in turn creates the potential for more robust decisions about climate planning and adaptation.

CHAPTER 5
MODEL SKILL AT SIMULATING SPATIAL CLIMATE PATTERNS (1961-1990)

5.1 INTRODUCTION

In this chapter, model skill at representing the mean seasonal spatial patterns is assessed, along with the spatial variability of key climate patterns on monthly timescales. Having analyzed certain temporal characteristics of the modelled data, the next aspect to examine is the monthly spatial data. National figures for means and variances are useful at a national level to inform policy decisions. However, as a result of locational effects, the climatological means of temperature and rainfall as calculated for the entire domain may not fully describe climate at the regional/local scale. Similarly, the impacts associated with an average, domain-wide change in climate may not accurately reflect the location-specific change and related impacts at the regional scale. A spatial analysis is therefore important to assess spatial skill and to help identify areas of low skill in the individual models.

The first part of this assessment focuses on mean seasonal spatial patterns for both temperature and precipitation. Recalling Murphy's (1993) definitions of error and bias, the average of the differences between individual observed and modelled gridcells is used to quantify spatial bias rather than the difference between the spatial averages, which was used in the previous chapter. The spatial correlation is used to quantify agreement between the observed and modelled mean patterns. The Pearson correlation coefficient is used for its sensitivity to extreme values, which makes it useful for identifying models with a geographically specific bias. The significance of differences between the modelled and observed climate patterns is also calculated.

Mean spatial patterns are the result of many processes and factors, and a model may be skilful at simulating a particular subset of those processes and the associated component pattern although it may lack skill at representing the average

climate pattern. Therefore, empirical orthogonal function (EOF) analysis is used to investigate interannual variability of spatial patterns. Model EOF patterns are compared to the EOF patterns of the observed data and the explained variance of each mode is compared also.

An analysis of spatial patterns may provide further insights into whether a model is displaying genuine skill or skill due to error cancellation. In Chapter 4, certain models emerged as having smaller levels of error than others when their nationally averaged values for temperature and precipitation are compared to the observed over annual timescales. However, an agreeable average is not necessarily a sign of a skilful model, as the right combination of both positive and negative biases in the individual gridcells would also lead to an apparently skilful average. Therefore, it is important to check that spatial patterns are represented skilfully, to be certain whether a model is exhibiting genuine skill. If the models can be shown to possess high skill when compared with present-day observations, they are much more likely to be useful tools for planning and decision-making

5.2 MEAN SEASONAL SPATIAL PATTERNS

5.2.1 Methodology

Spatial correlations are widely used for assessing model skill at simulating spatial patterns. Essentially, a spatial correlation treats a two-dimensional data field as a list of individual points, from which the Pearson correlation coefficient is calculated. The spatial correlation has been demonstrated to be a useful method for quantifying model performance at representing a variety of spatial patterns. Spatial correlation has been used as a model skill metric by Leung and Ghan (1999), who used spatial correlations to demonstrate the increased skill of an RCM over a GCM in the Pacific Northwest. Spatial correlations were also employed by Pan *et al.* (2001) to determine how skilfully RCMs simulate the observed precipitation patterns of the United States and also to determine the similarity between present-day and future modelled precipitation patterns. Additionally, Miller *et al.* (2006) used spatial correlations to determine AOGCM skill at representing the Northern and Southern hemisphere annular modes of variability.

Seasonal spatial patterns are extracted by calculating the mean temperature and precipitation at each gridcell across all relevant months. For example, the spatial pattern of winter temperature is derived from the averages of all December, January and February values at each grid cell. Both the modelled patterns and the bias of the modelled patterns with respect to the observed are mapped. To obtain comparable observed spatial data, the BIC data is regridded to the grid resolution of the RCMs and the RCM land-sea mask is applied to the regridded observational data. To quantify agreement between the observed and modelled patterns, the spatial correlation coefficient is calculated as follows:

Equation 5.1: Spatial correlation coefficient

$$\rho = \frac{\frac{1}{G}\sum_{g=1}^{G}\left(x_o^g - \overline{x_o}\right)\left(x_m^g - \overline{x_m}\right)}{\sigma_o \cdot \sigma_m}$$

where

G = number of gridcells,

σ_o *= standard deviation of the observed data,*

σ_m *= standard deviation of the modelled data,*

$\overline{x_o}$ *= mean of the observed data,*

$\overline{x_m}$ *= mean of the modelled data,*

x_o^g *= observed data at gridcell g and*

x_m^g *= model output at gridcell g.*

The mean spatial bias of each model is also calculated as the average of the differences between observed and modelled values at each gridcell:

Equation 5.2: Spatial bias

$$Bias = \frac{1}{G}\sum_{g=1}^{G}(x_m^g - x_o^g)$$

where

G = number of gridcells,

x_o^g *= observed data at gridcell g and*

x_m^g *= model output at gridcell g.*

5.2.2 Results: Temperature

Spatial patterns for temperature in spring and the associated bias with respect to observations are given in Figures 5.1 and 5.2. Several models are notably warmer than observed, namely RACMO, REMO, RCAO-H, HIRHAM-E5 and RCAO-E4. Warm biases are particularly apparent in the midlands in certain models, such as HadRM3P-a, b and c. However, several other models are slightly cooler than observed, such as PROMES, CHRM and REGCM. The unweighted ensemble average is also given for illustrative purposes. When all the models are treated as an ensemble, with equal skill assumed and no weighting to account for model differences, the net result is a spatial pattern that is close to the observed, with minimal bias. This is a clear example of an ensemble getting the right answer for the wrong reasons, as adopting such an approach would leave uncertainties associated with the individual model's outputs unexplored.

Spatial patterns for summer temperature and the associated bias are given in Figures 5.3 and 5.4. The models which displayed a cool bias in spring, such as PROMES, CHRM and REGCM, display an enhanced negative bias with CHRM in particular displaying biases of up to -1.49^{o}C in some gridcells. ARPEGE-a, ARPEGE-b and ARPEGE-c also exhibit an increased cool bias in summer compared with spring. Some of the warm biased models, namely RACMO, REMO and HIRHAM-E5 remain positively biased in most gridcells but RCAO-H and RCAO-E4 model much smaller biases in summer than in spring, suggesting that RCAO as an RCM has greater skill at modelling summer climate than spring climate for Ireland.

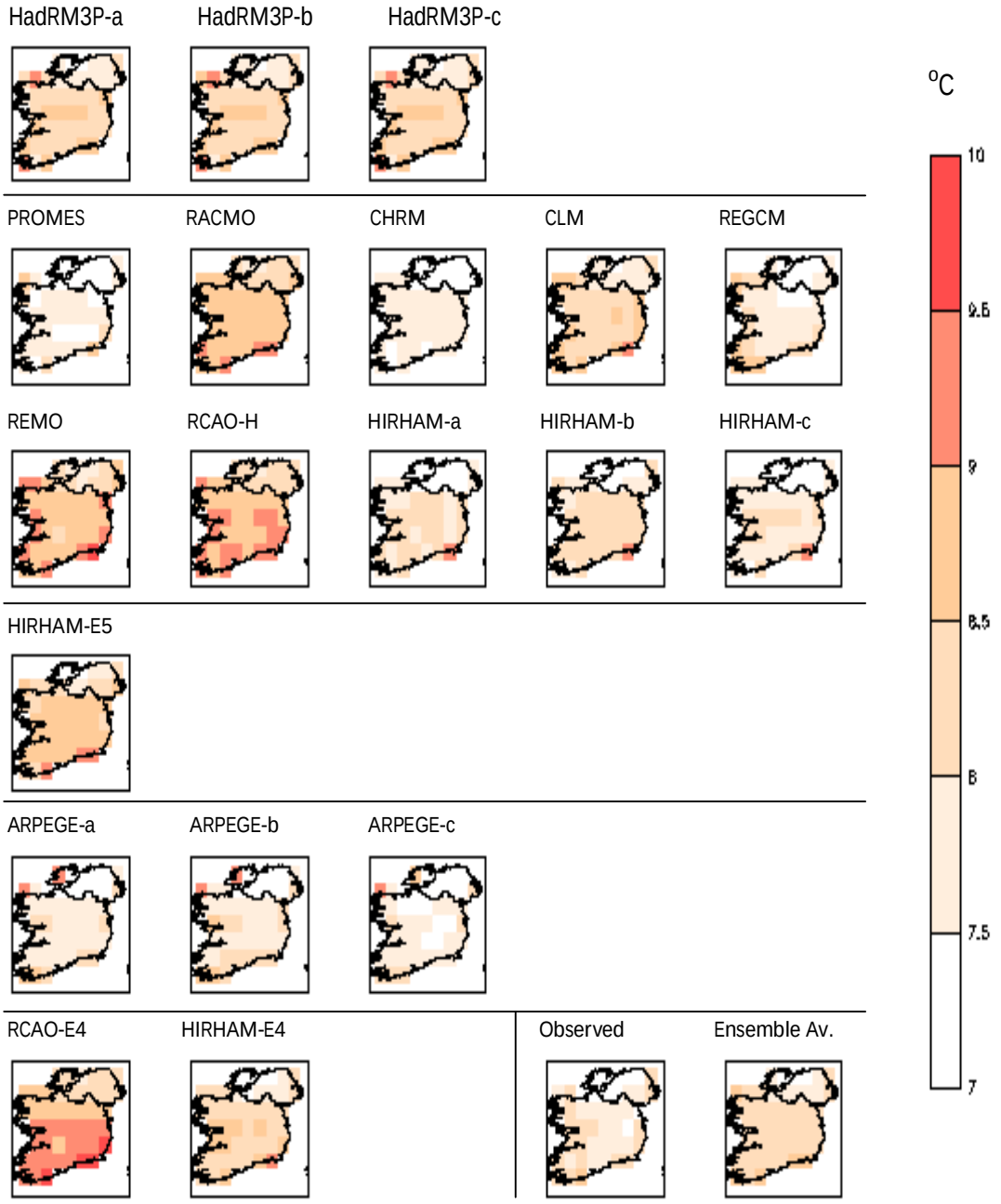


Figure 5.1: Modelled temperature over Ireland in spring (MAM) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row-HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

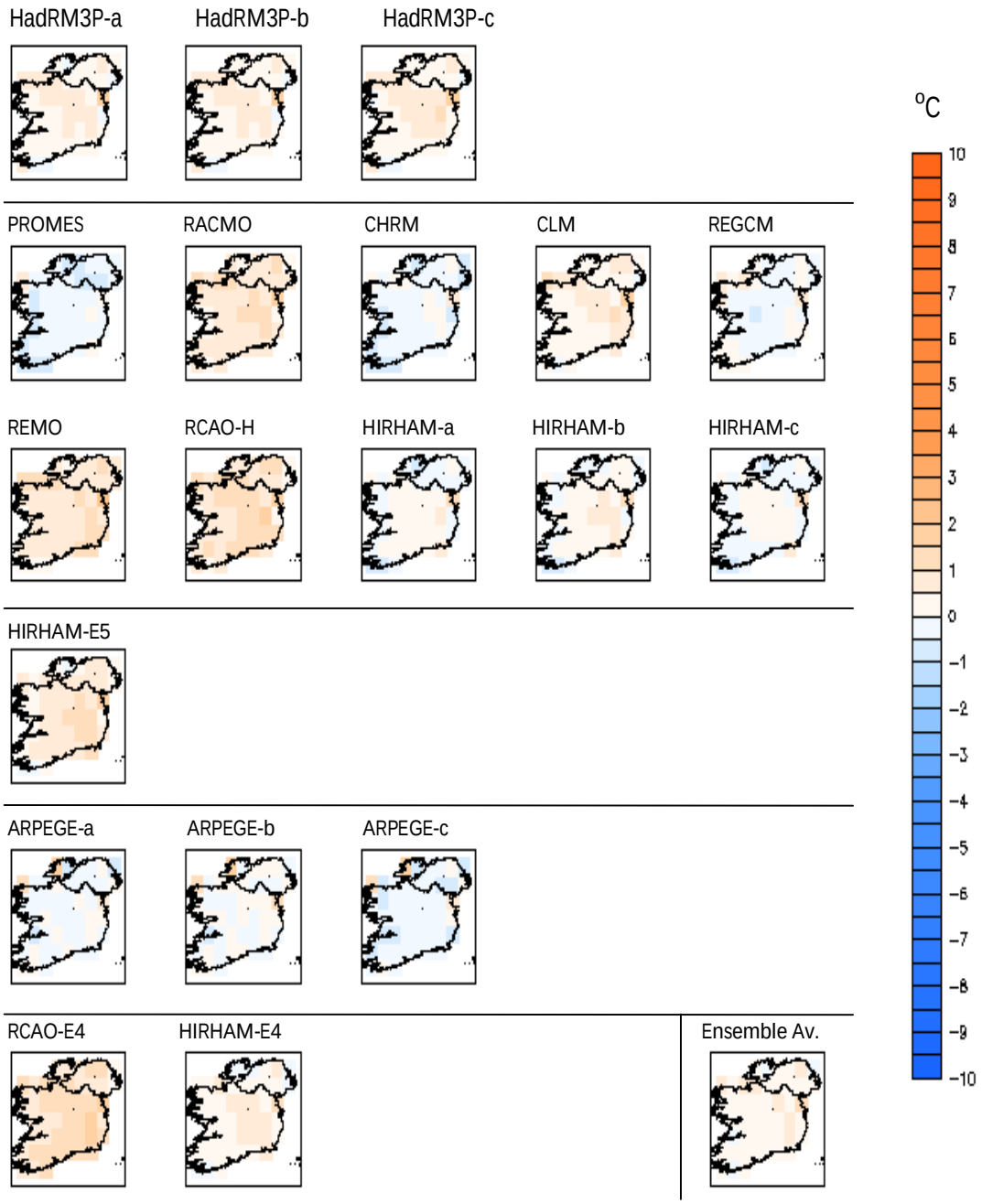


Figure 5.2: Bias of modelled temperature compared with observed over Ireland in spring (MAM) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row-HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

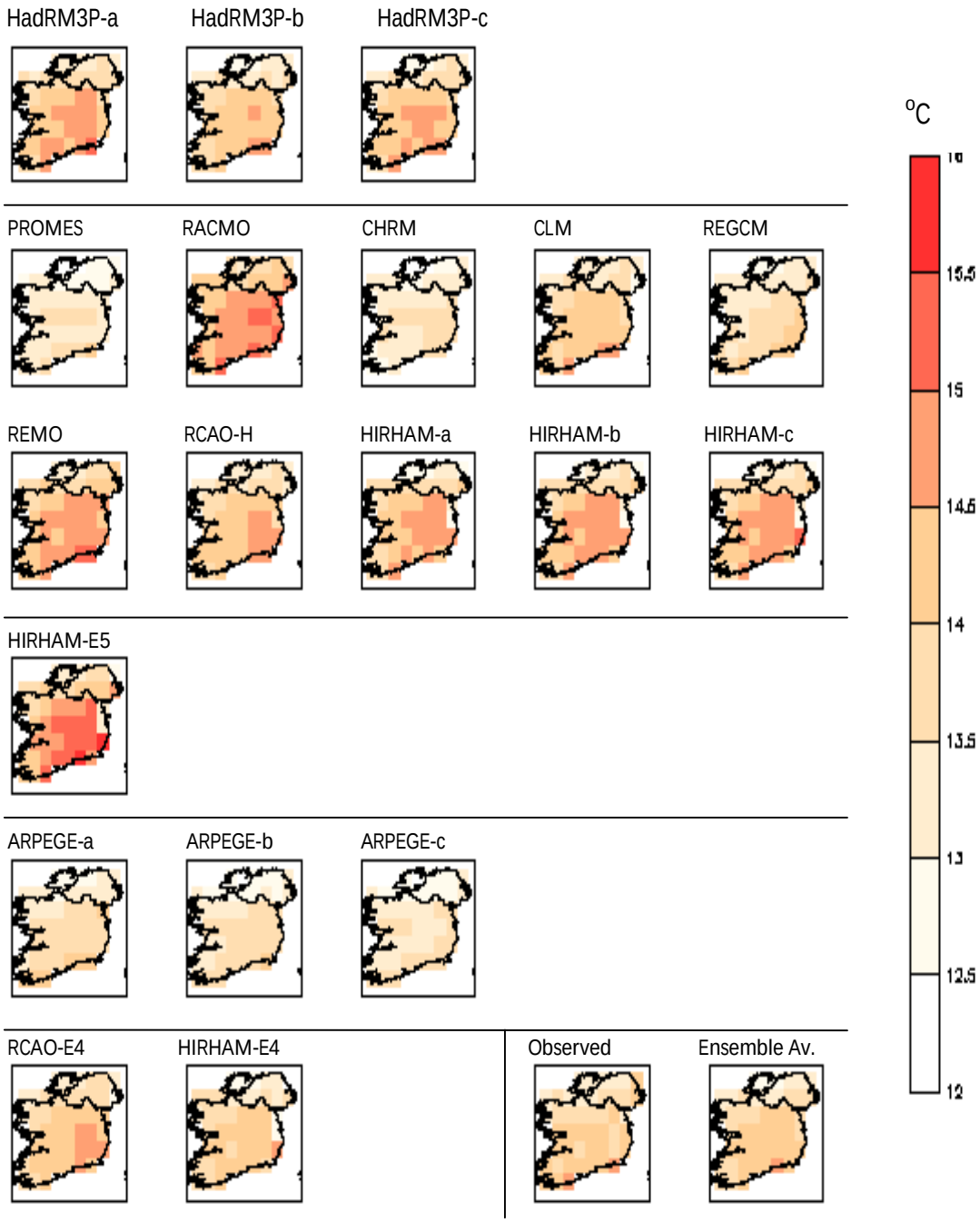


Figure 5.3: Modelled temperature over Ireland in summer (JJA) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row-HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

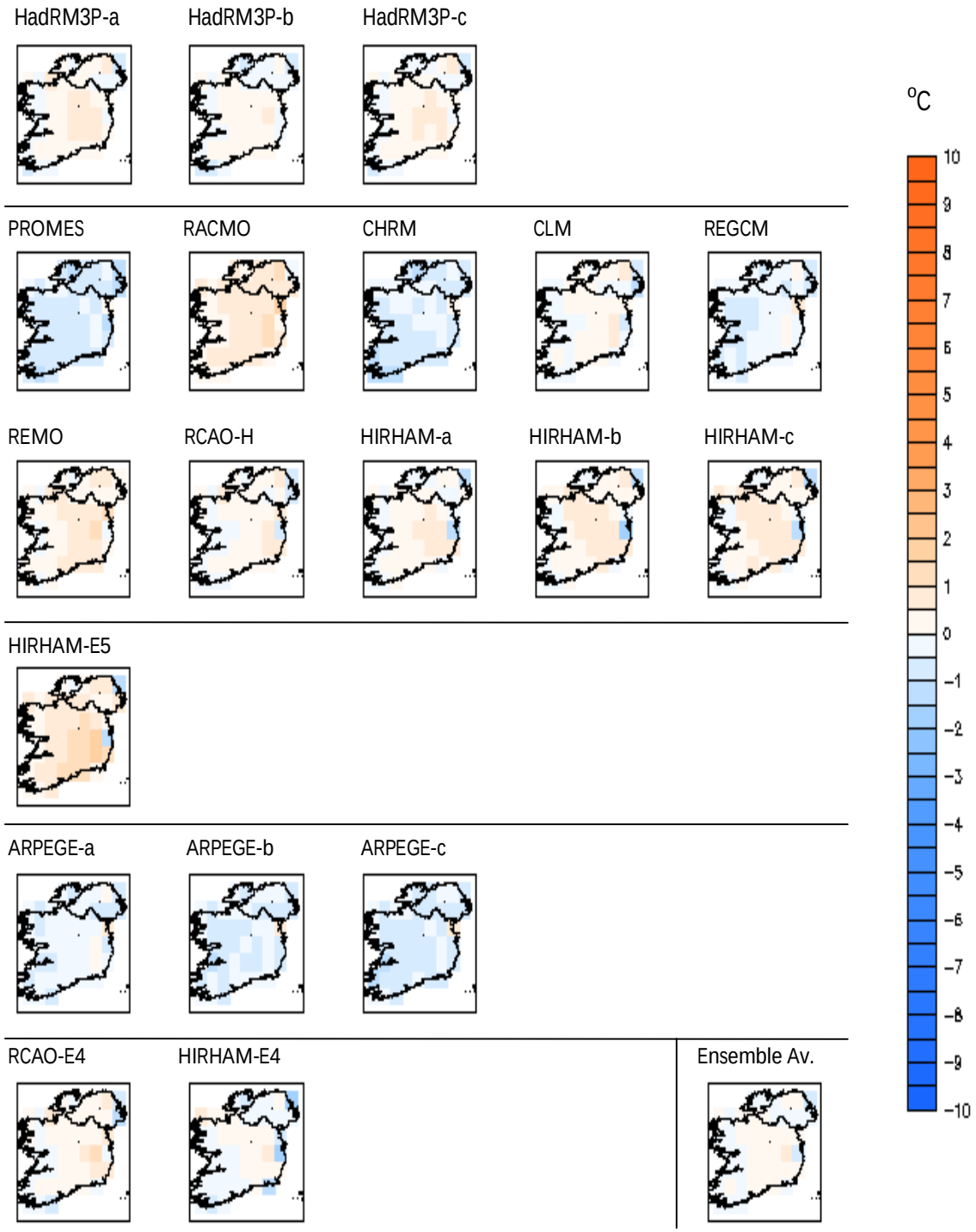


Figure 5.4: Bias of modelled temperature compared with observed over Ireland in summer (JJA) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row- HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

Several models which exhibit a mostly warm bias display a cool bias over one to two gridcells in the east to north-east. Examples include CLM, RCAO-H, HIRHAM-a, HIRHAM-b, HIRHAM-c and HIRHAM-E5. Again, when the models are treated as an unweighted ensemble, the ensemble average yields a spatial pattern that is close to the observed, with minimal bias as the warm and cool models cancel each other out.

Spatial patterns for autumn temperature and the associated bias are given in Figures 5.5 and 5.6. PROMES and REGCM, which are predominantly cool in summer, model autumn climate with much less bias. However, CHRM, ARPEGE-a, ARPEGE-b and ARPEGE-c are model cooler temperatures than observed, similar to their performance in summer. The cool bias in CHRM is systematic across the Irish domain while the ARPEGE simulations exhibit a warm bias across some north-western grid-cells also. Overall, the majority of models tend towards slightly warmer temperatures than observed in autumn, with the greatest warm biases occurring in REMO, RCAO-H and RCAO-E4. Again, the unweighted ensemble average is close to the observed pattern, with minimal bias.

Lastly, spatial patterns for temperature in winter and bias with respect to observations are given in Figures 5.7 and 5.8. The most important feature to note is that when treated as an unweighted ensemble, winter temperature is significantly overestimated, suggesting deficiencies that do not cancel out in the averaging process (Figure 5.7). The majority of the models tend towards significantly warmer temperatures in winter and while some models, such as CHRM and HIRHAM-E5 display cool biases in specific gridcells, there are no models which give systematically cooler output. The ARPEGE simulations, which were cooler than observed in autumn, model winter temperatures more skilfully, though they too display warm biases over certain gridcells in the north-west. The unweighted ensemble average displays low levels of bias overall, however the errors in the individual ensemble members are reflected in the unweighted ensemble. For example, there is a slight negative bias in certain gridcells on the east coast, a result of the large negative bias in these gridcells in the various simulations using HIRHAM. The bias is diminished when these models are averaged with the other ensemble members, but the error is still apparent.

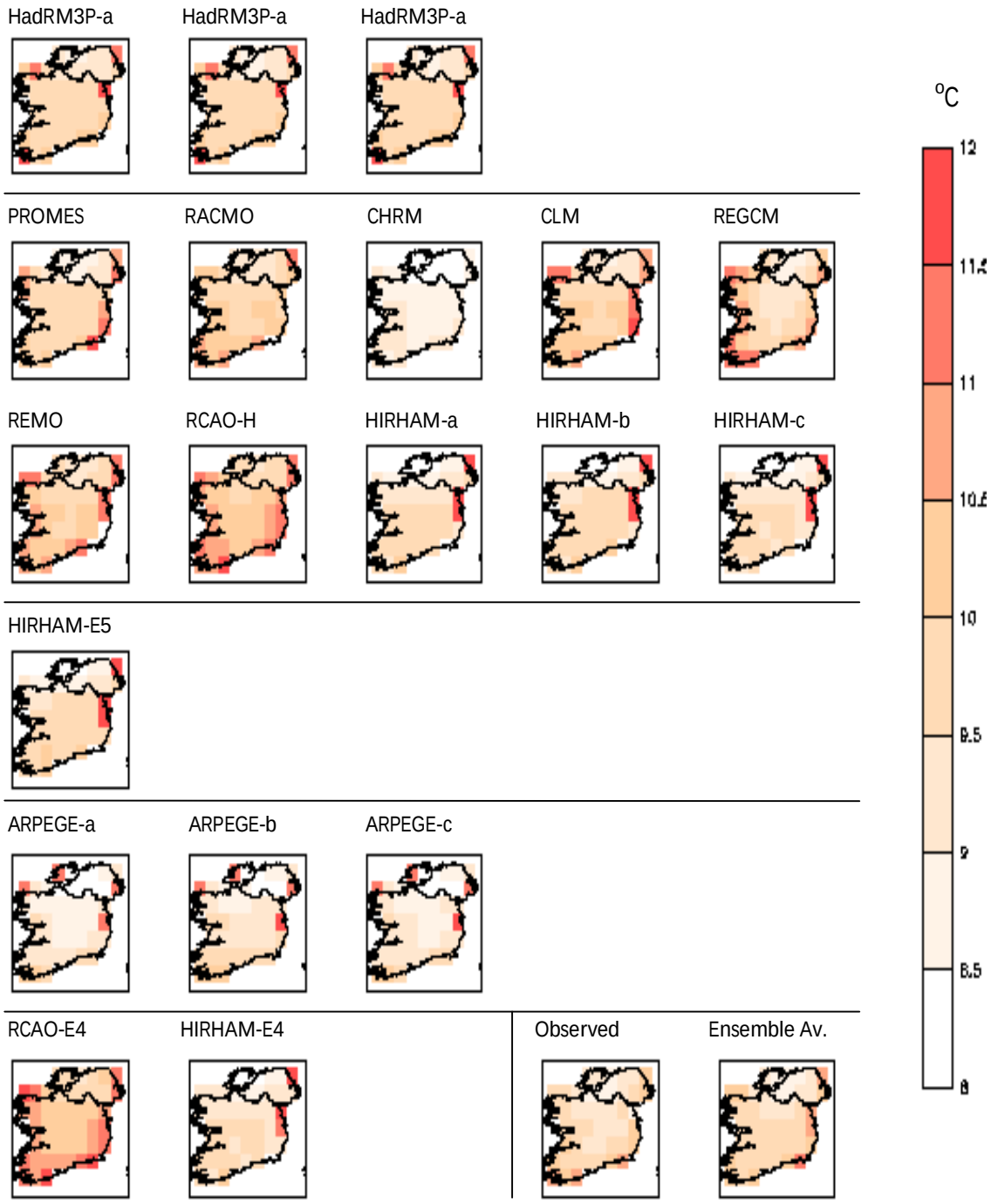


Figure 5.5: Modelled temperature over Ireland in autumn (SON) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row-HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

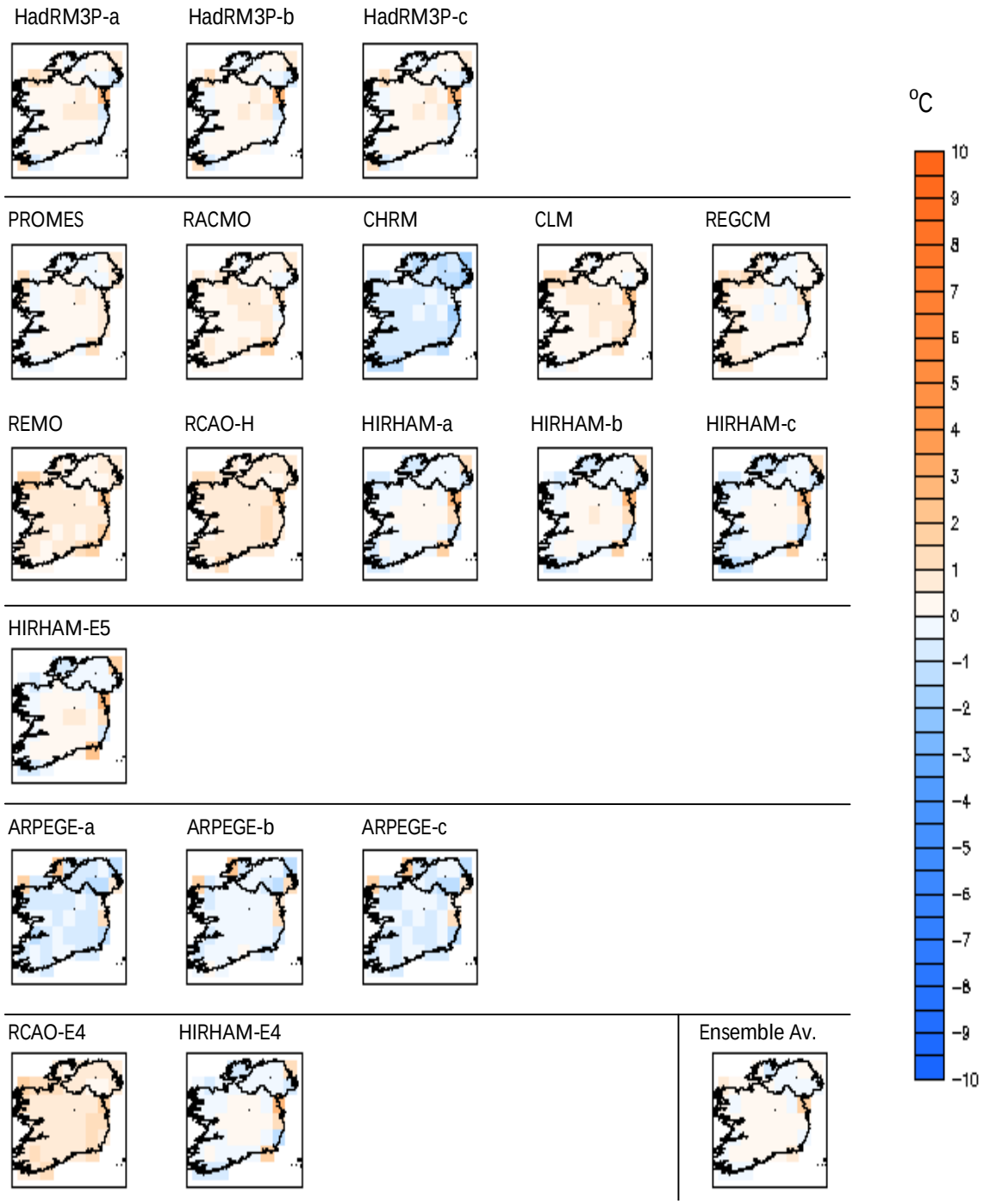


Figure 5.6: Bias of modelled temperature compared with observed over Ireland in autumn (SON) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row-HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

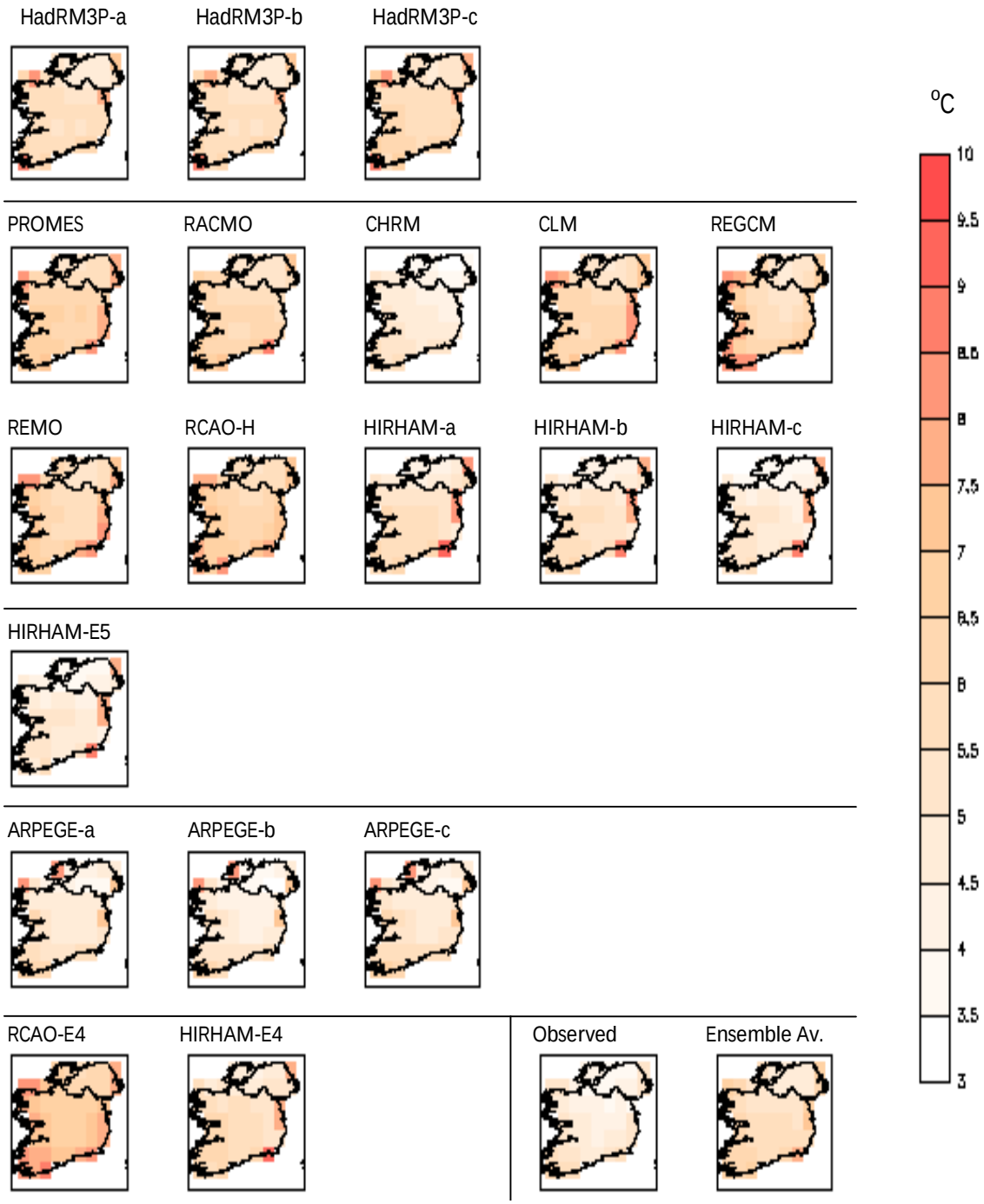


Figure 5.7: Modelled temperature over Ireland in winter (DJF) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row-HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

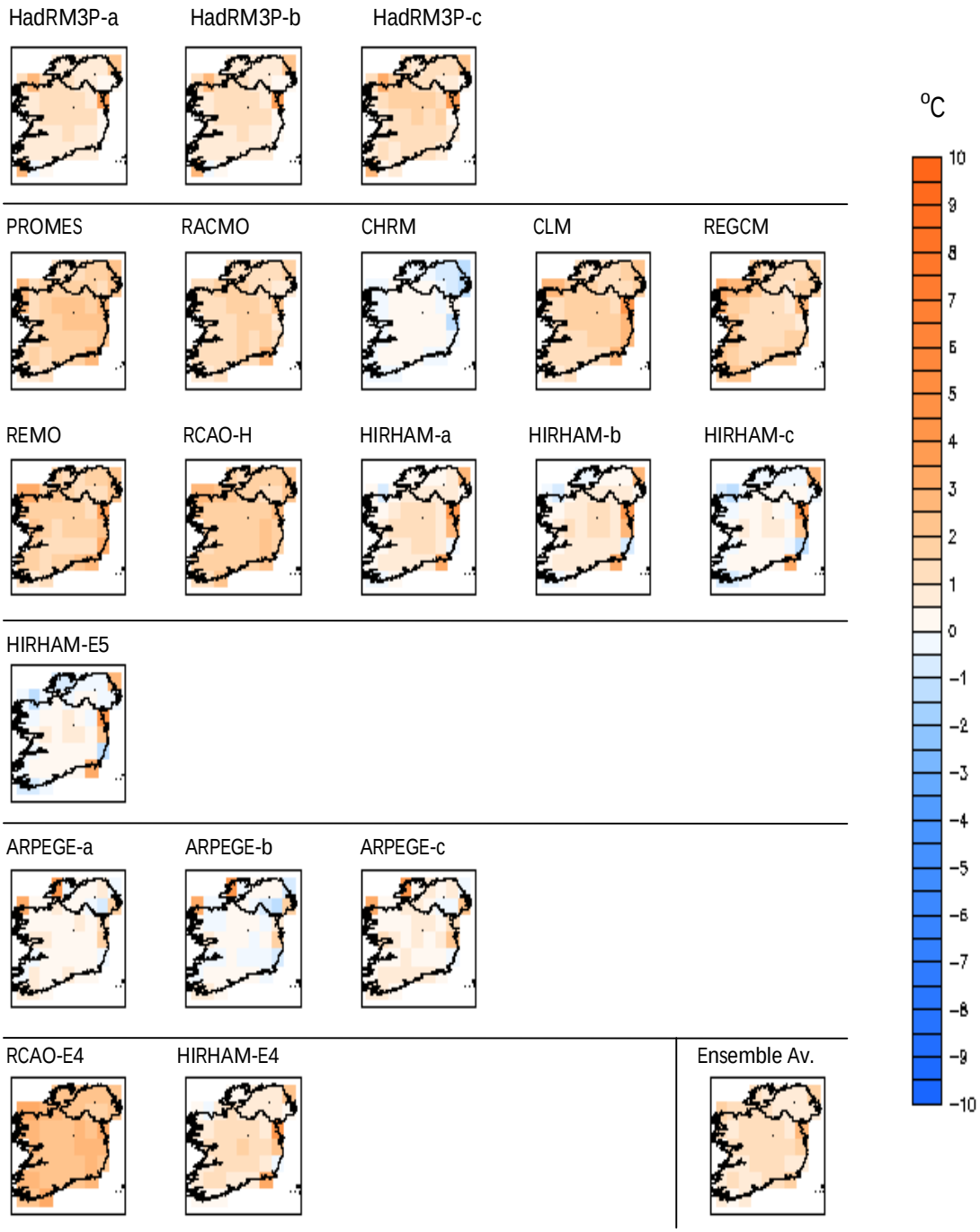


Figure 5.8: Bias of modelled temperature compared with observed over Ireland in winter (DJF) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row- HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

5.2.3 Results: Precipitation

Spatial patterns for precipitation in spring and the associated bias with respect to observations are given in Figures 5.9 and 5.10. Several models are wetter than observed. Some of these models also display a positive temperature bias, which may offer an explanation for the positive precipitation biases. In a warmer climate, there is more heat available for evaporation, which is turn leads to increased availability of atmospheric water vapour for precipitation in the model. In the case of RACMO, REMO, RCAO-H, HIRHAM-E5 and RCAO-E4, this may contribute to the wet biases exhibited in spring. Similarly, PROMES and CHRM, which were slightly cooler than observed, are also slightly drier. However, REGCM, which displayed a cool temperature bias, is now found to simulate wetter conditions than observed. Conversely, the three simulations using HadRM3P were warmer than observed yet tend towards drier conditions. The processes underlying these biases are not so apparent. The unweighted ensemble average is again close to the observed pattern, as drier models nullify the effects of wetter models.

Spatial patterns for summer precipitation and the associated bias with respect to observations are given in Figures 5.11 and 5.12. There are some differences between the spring and summer spatial maps. The HadRM3P simulations remain drier than observed. However the ARPEGE simulations, which tended towards wetter conditions in most gridcells in the previous season, now display a much more limited wet bias, affecting only the north-eastern gridcells. PROMES, which was a cool model in summer is now found to tend towards wetter conditions, particularly along the west coast. Similarly, REGCM, another cool model in summer, displays a significant wet bias across the Irish domain.

CHRM, also a cool model, simulates a small dry bias across some gridcells but wetter conditions to the north-east. Indeed, the wet bias over the north-eastern gridcells is a feature of most models and as a result, while the unweighted ensemble average is close to observed in most areas, the wet bias is still present in this area. Sweeney (1989) states that based on observed data, increased rainfall in this area is associated with cyclonic circulation activity, which may indicate that the models lack skill in simulating the large-scale circulations or their effects on regional climate patterns. Another concern is the range of individual model errors and deficiencies that an unweighted ensemble approach would disguise.

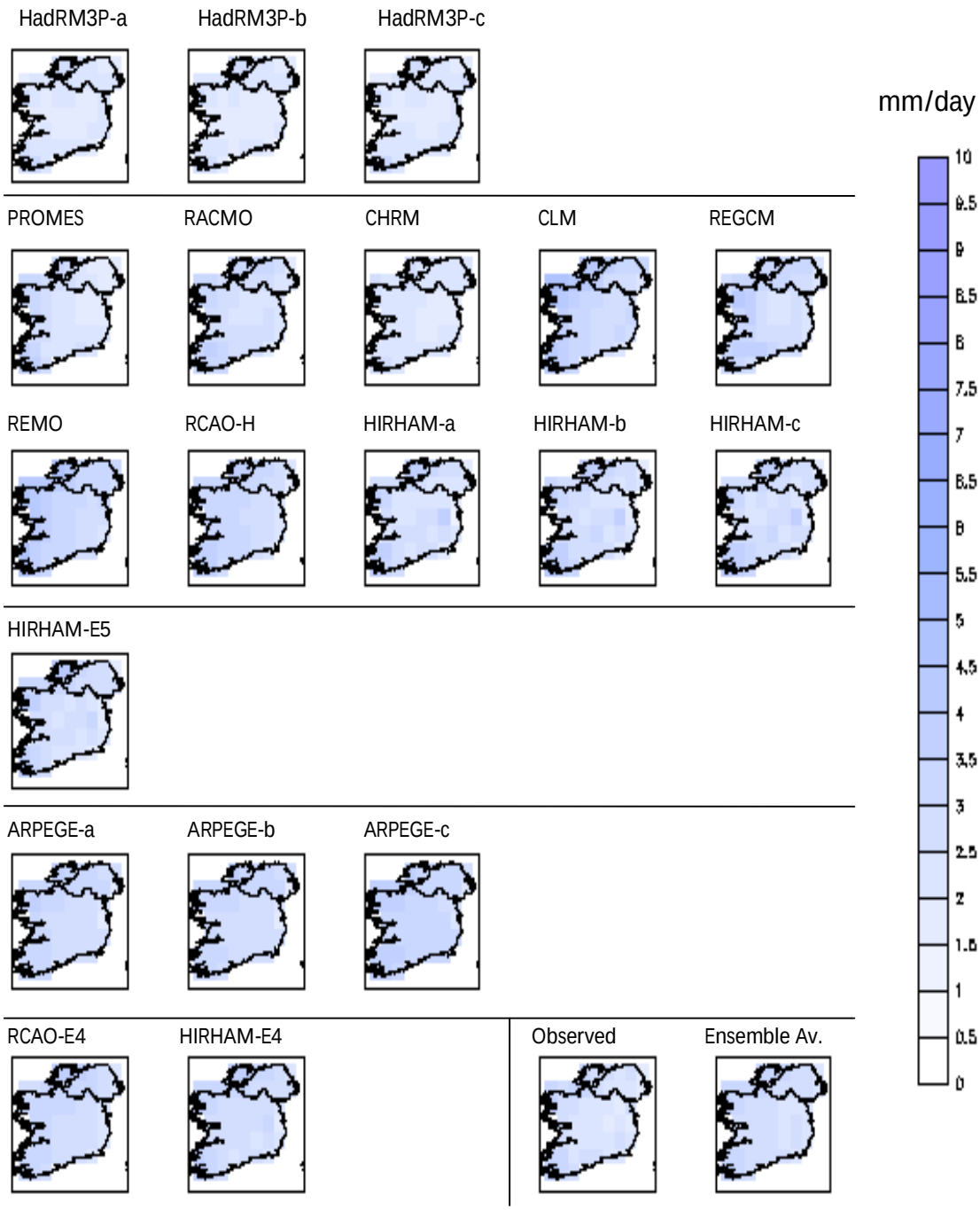


Figure 5.9: Modelled precipitation over Ireland in spring (MAM) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row-HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row-Observed SSTs, sixth row (left) – ECHAM4-OPYC.

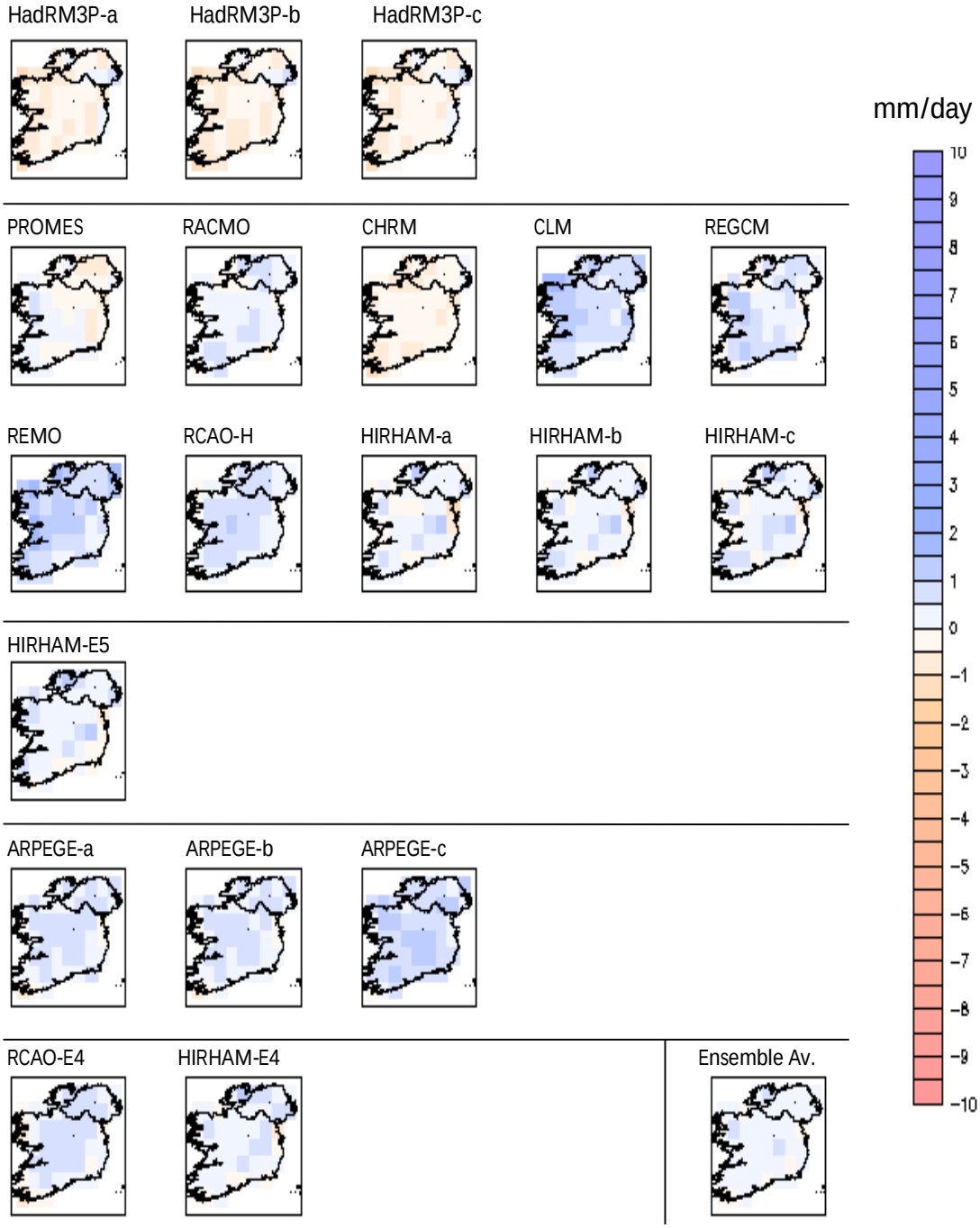


Figure 5.10: Bias of modelled precipitation compared with observed over Ireland spring (MAM) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row- HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

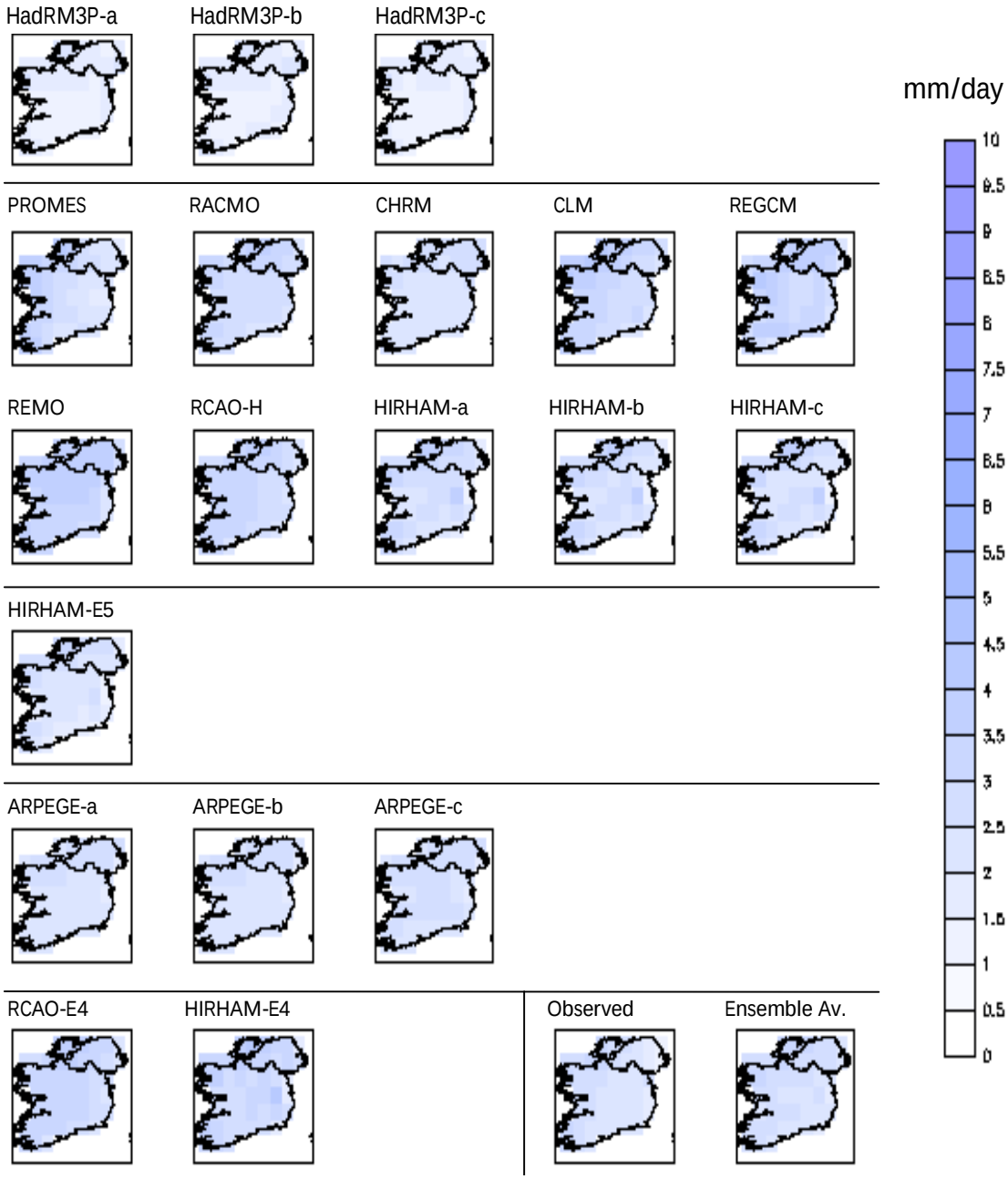


Figure 5.11: Modelled precipitation over Ireland in summer (JJA) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row-HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row-Observed SSTs, sixth row (left) – ECHAM4-OPYC.

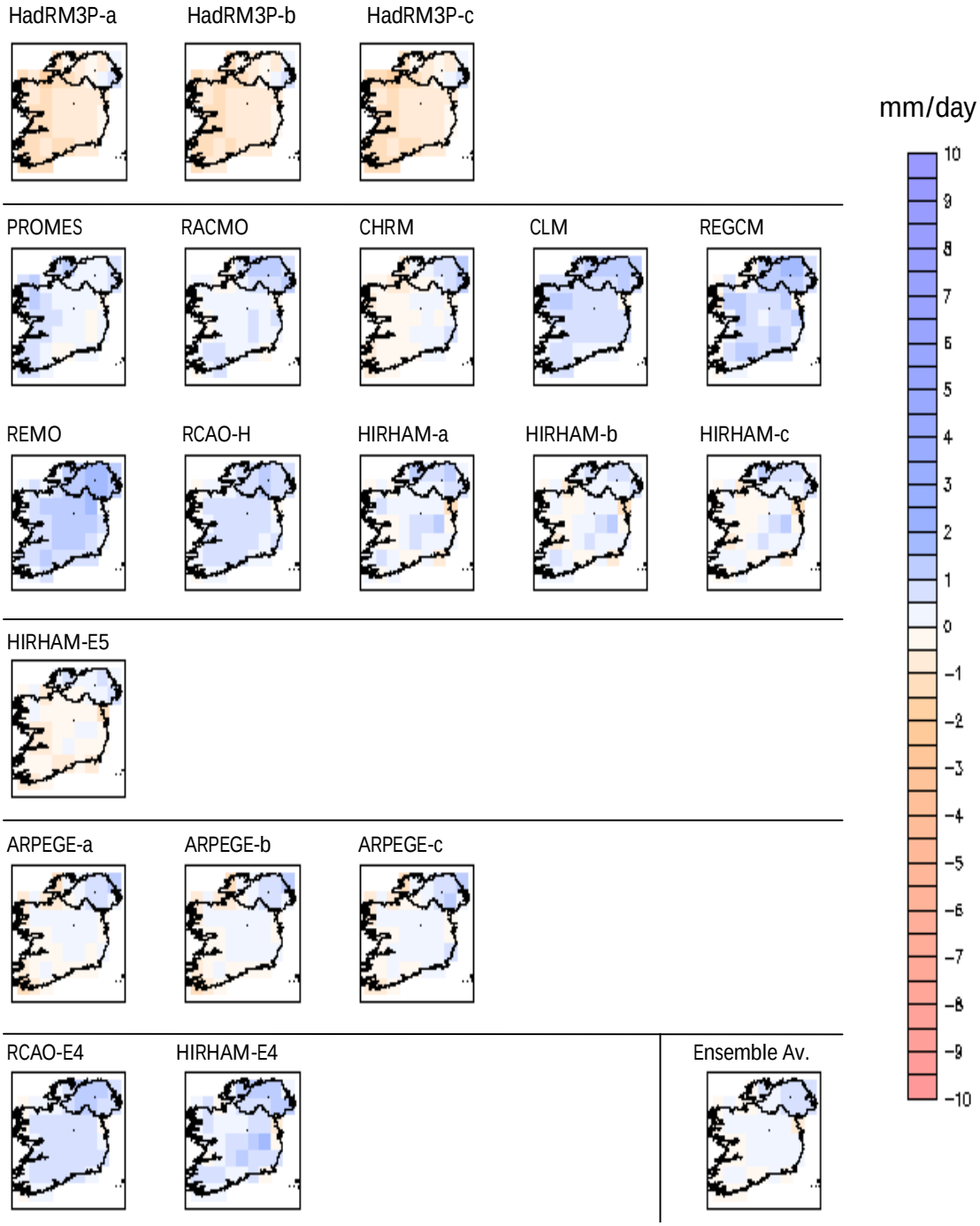


Figure 5.12: Bias of modelled precipitation compared with observed over Ireland summer (JJA) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row- HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

Spatial patterns for autumn precipitation and the associated bias are given in Figures 5.13 and 5.14. PROMES and REGCM model autumn temperature with much less bias than in summer and similarly, their outputs for precipitation are closer to observed in autumn. The driest models are the HadRM3P models and CHRM, although the HadRM3P models simulate wetter conditions than observed in some north-eastern gridcells. Such behaviour is to be expected in CHRM, as it is quite a cool model in autumn. However, the three simulations of HadRM3P are all warmer than observed in this season and as such, a dry bias across so much of the country is an unexpected outcome. One model, RCAO-E4 is particularly wet in autumn. This model was also found to have a warm bias Simulations using HIRHAM, such as HIRHAM-a, HIRHAM-b, HIRHAM-c and HIRHAM-E5, tend towards wetter conditions in a specific gridcell which lies one cell inland from the east coast. Due to proximity to the Wicklow mountains, a rain shadow effect is expected in this area. Therefore, the positive precipitation bias may indicate that this model fails to capture certain orographic effects.. When taken as an unweighted ensemble, the model biases largely cancel each other out with the exception, again, of the wet bias over the north-eastern part of the domain.

Spatial patterns for precipitation in winter and bias with respect to observations are given in Figures 5.15 and 5.16. Although winter temperature in the unweighted ensemble was significantly overestimated, precipitation in the unweighted ensemble appears much closer to the observations. While several of the warmer models such as RACMO, CLM, REGCM, RCAH-H and RCAO-E4 also tend towards wetter conditions, other models that were warmer than observed in the temperature analysis are either drier than observed or exhibit a combination of wet and dry biases.

For example, the HadRM3P simulations are all predominantly dry in winter, but PROMES and REMO, which were systematically warmer across the domain, exhibit an unsystematic pattern of precipitation bias. Several models, such as ARPEGE-a, ARPEGE-b, HIRHAM-E5, HIRHAM-E4 and HIRHAM-a are wetter in the midlands than in coastal regions, contributing to a slight wet bias in this area when the models are taken as an ensemble.

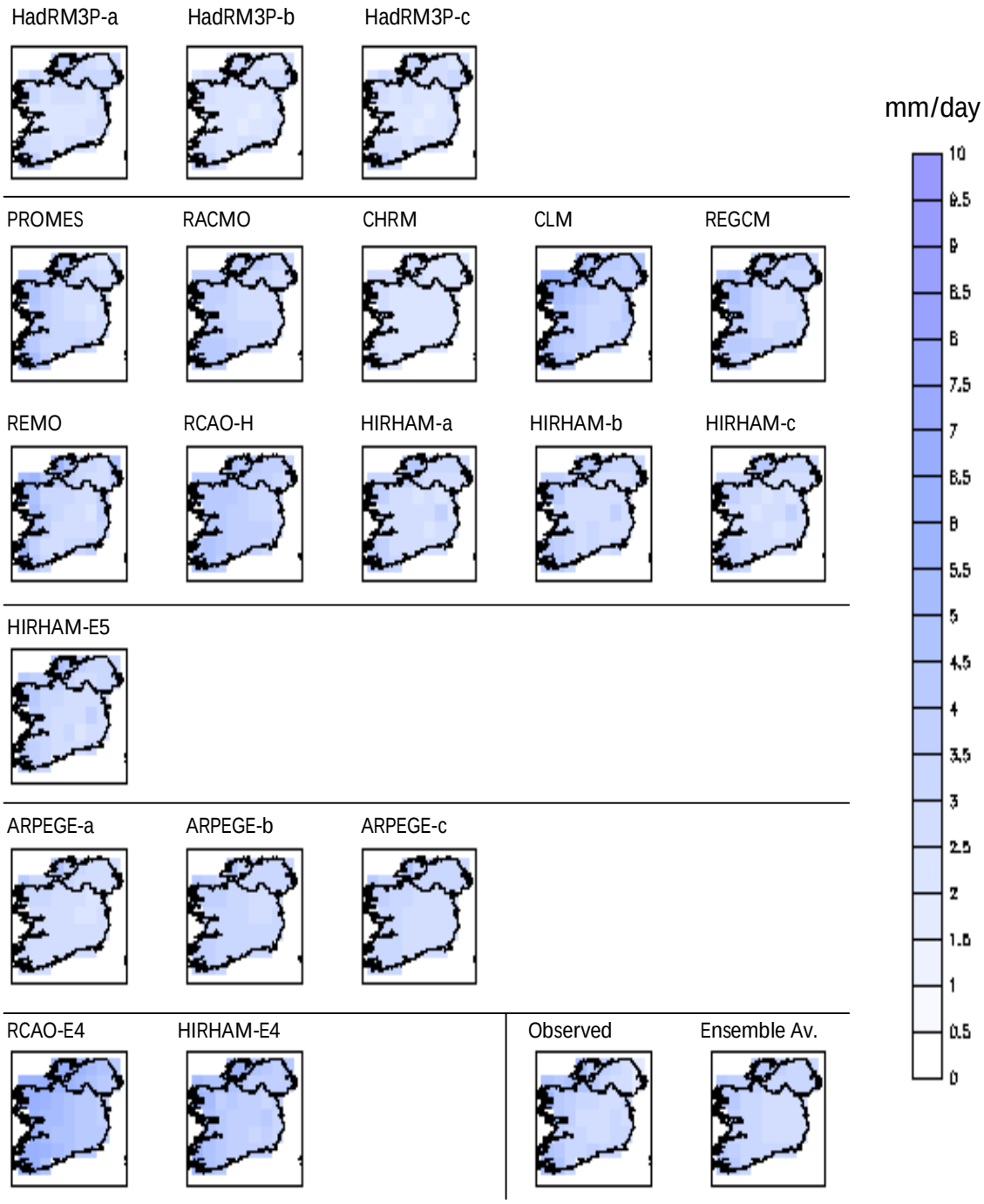


Figure 5.13: Modelled precipitation over Ireland in autumn (SON) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row-HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row-Observed SSTs, sixth row (left) – ECHAM4-OPYC.

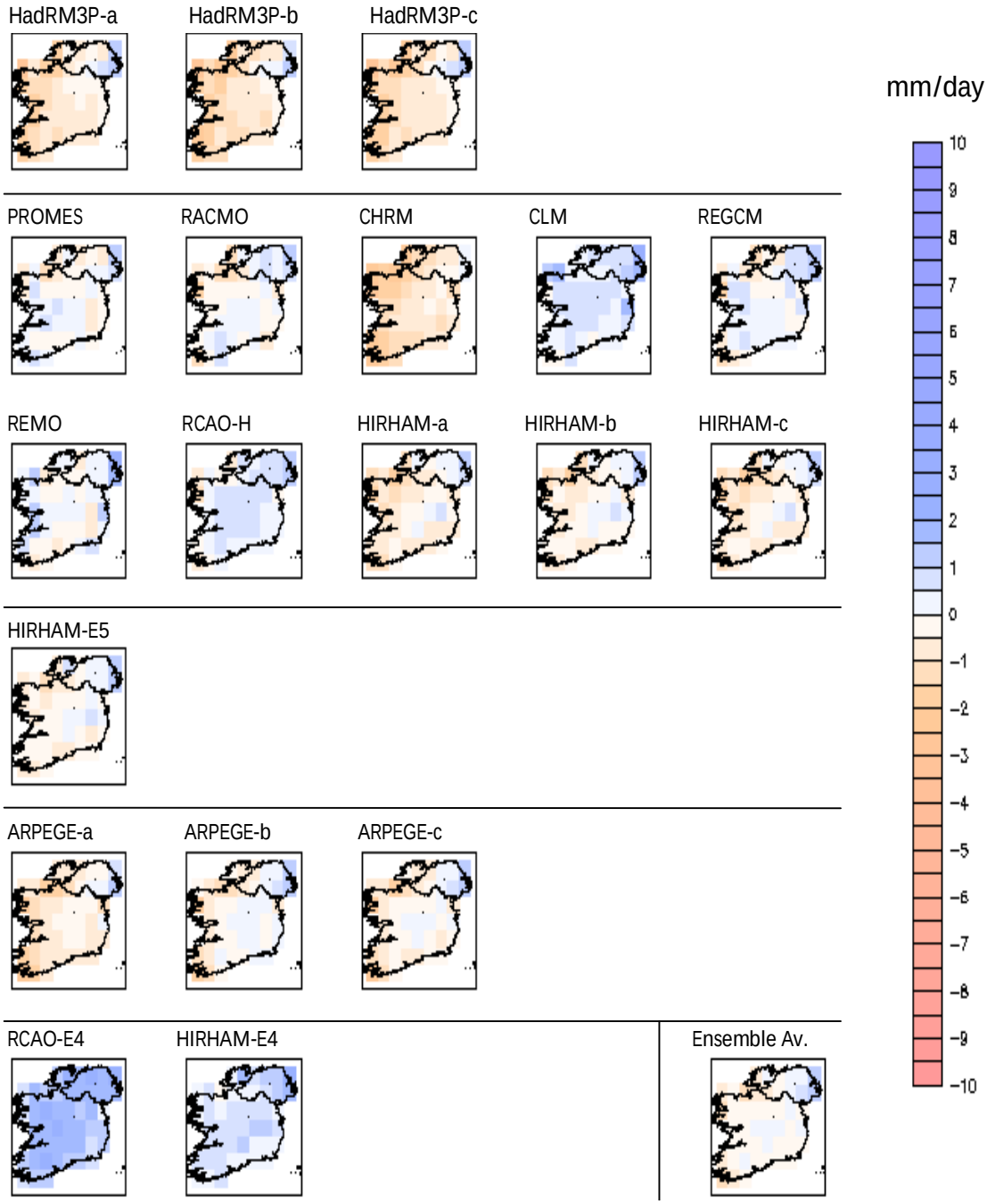


Figure 5.14: Bias of modelled precipitation compared with observed over Ireland autumn (SON) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row- HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

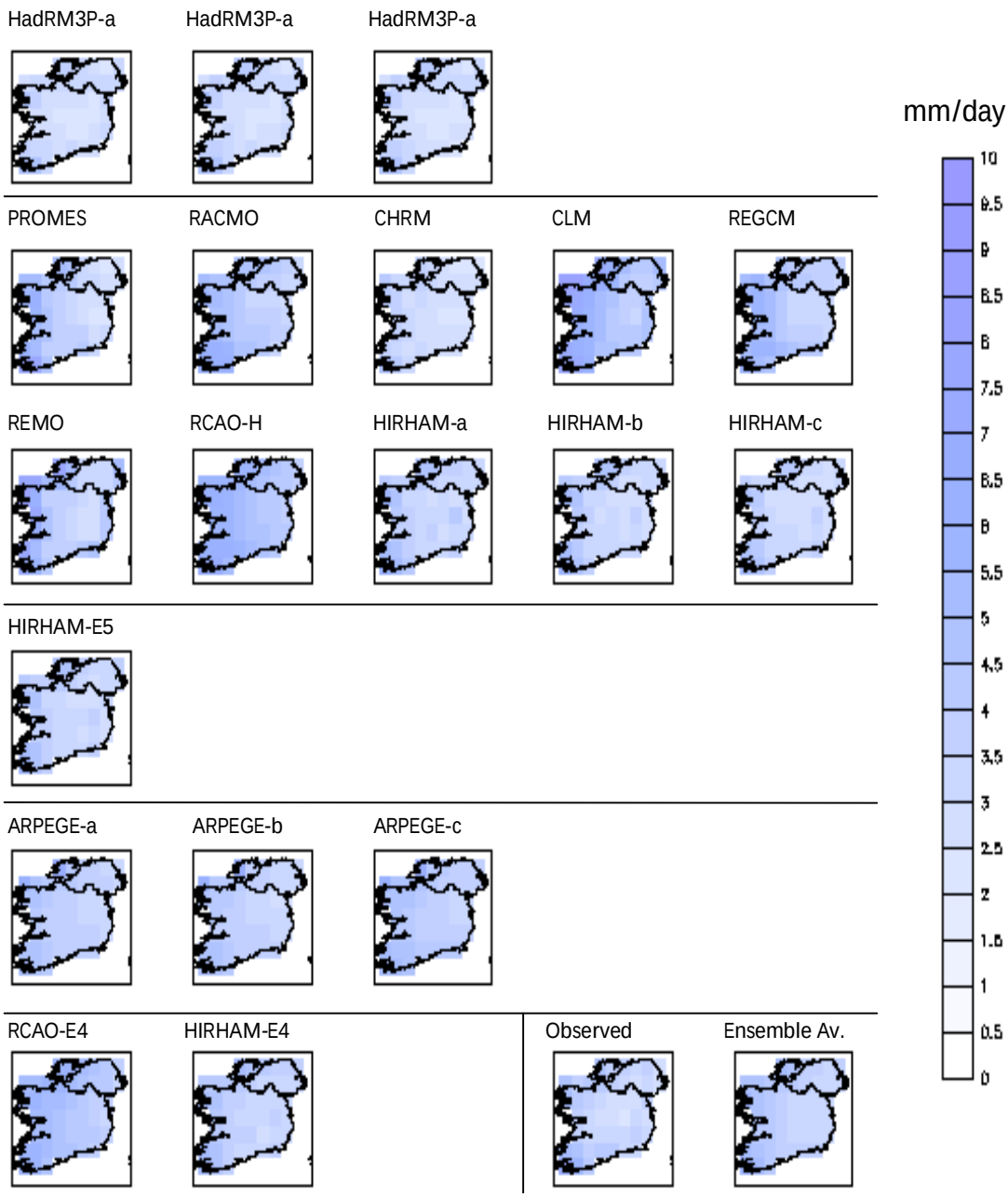


Figure 5.15: Modelled precipitation over Ireland in winter (DJF) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row-HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row-Observed SSTs, sixth row (left) – ECHAM4-OPYC.

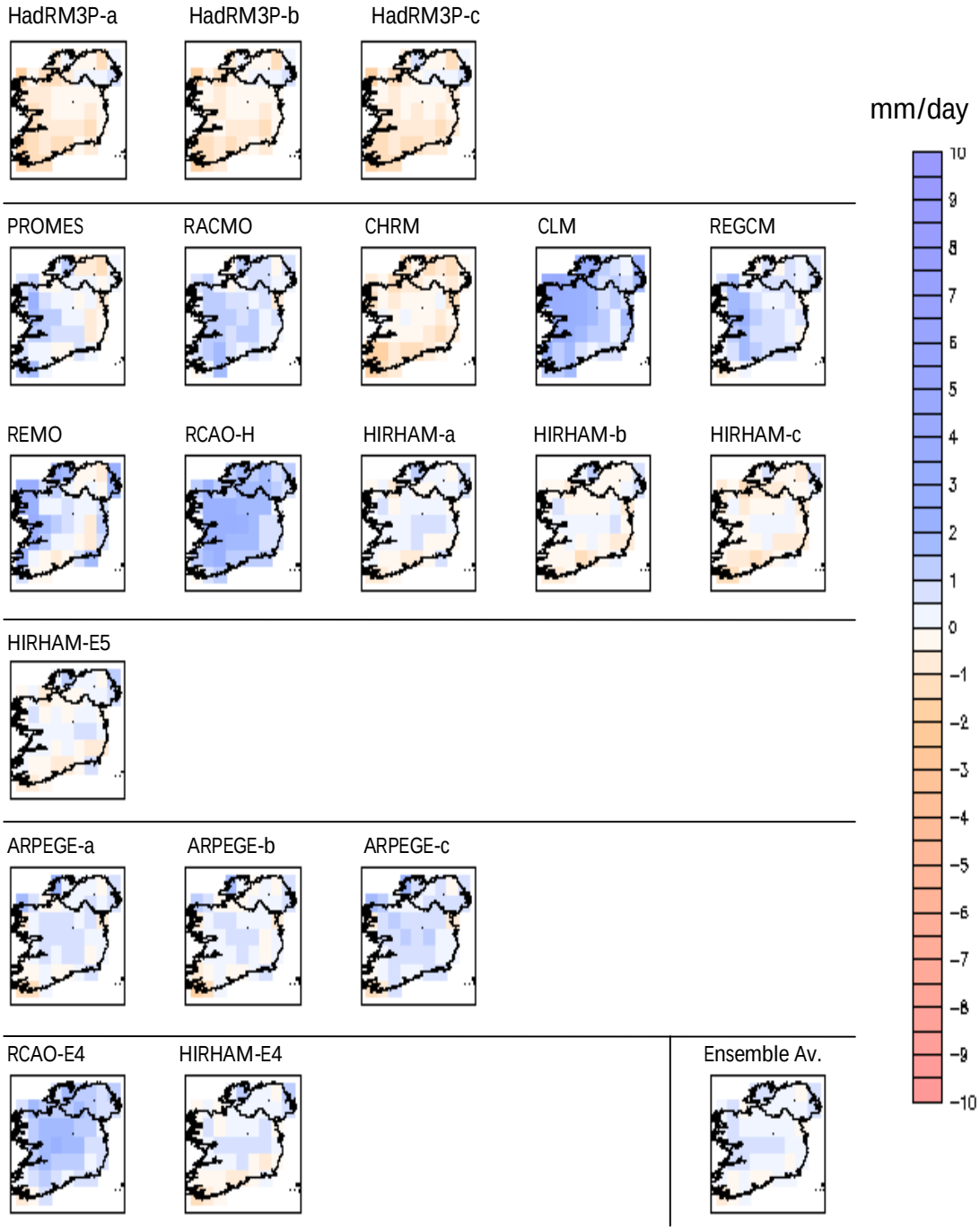


Figure 5.16: Bias of modelled precipitation compared with observed over Ireland winter (DJF) for the 1961-1990 baseline period. Models are classified according to GCM driver group: First row- HadCM3/HadAM3P, second and third row – HadCM3/HadAM3H, fourth row- ECHAM4-OPYC/ECHAM5, fifth row- Observed SSTs, sixth row (left) – ECHAM4-OPYC.

5.2.4 Further analysis and discussion

Table 5.1 gives the average bias, highest bias for a single gridcell, lowest bias for a single gridcell and correlation between modelled and observed patterns for each model in each season. These values pose a number of questions about what constitutes model skill and what characteristics to look for in a skilful model.

While bias scores indicate how much a model over or underestimates climate parameters, correlation scores indicate how skilfully they model spatial patterns, omitting systematic bias. A model may have low average bias, but if the corresponding correlation score is also low, this indicates that the bias is not spatially consistent. Conversely, a model with high average bias could have a high spatial correlation score, indicating an ability to model the spatial pattern and the presence of a systematic model bias.

Ideally, a model would capture both the spatial pattern and the magnitude of the climate variable with skill. However, it is evident from this table that no models behave in this way. The models with the highest correlation scores, such as RCAO-H and RCAO-E4 also have high average bias scores. For example, in winter, both models have correlation scores greater than 0.88 for temperature and greater than 0.73 for precipitation, signifying a strong association between the observed and modelled spatial pattern.

Yet RCAO-H overestimates temperature by an average of 1.88^{o}C and precipitation by an average of 1.47mm/day, while RCAO-E4 overestimates temperature by 2.46^{o}C on average and precipitation by 1.07mm/day.Conversely, several models with low average bias such as the ARPEGE simulations for temperature also have lower correlation scores, signifying a lesser level of association between the observed and modelled pattern. Although the average bias is low, from gridcell to gridcell, biases vary from approximately -1 to +4^{o}C in these three simulations. As the biases vary in space and are not systematic across the domain, the spatial pattern is distorted, resulting in output that does not correlate well with the observed spatial pattern.

Figures 5.17 and 5.18 illustrate this problem. These graphs show both absolute values for bias and values of r from the climate model skill analysis for temperature and precipitation.

Temperature

	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
MAM Av.	0.46	0.46	0.46	-0.32	0.80	-0.29	0.45	0.00	0.85	0.96	0.05	0.17	-0.09	0.66	-0.06	0.07	-0.22	1.08	0.33
MAM Max	1.56	1.56	1.56	0.65	1.73	0.62	1.70	0.90	2.01	1.80	1.27	1.26	1.22	1.55	1.77	1.84	1.71	1.94	1.41
MAM Min	-0.34	-0.44	-0.34	-0.89	0.33	-0.98	-0.35	-0.53	0.24	0.29	-0.68	-0.65	-0.89	-0.06	-0.63	-0.52	-0.83	0.15	-0.45
MAM r	0.66	0.63	0.64	0.74	0.77	0.67	0.65	0.71	0.74	0.68	0.68	0.66	0.52	0.67	0.55	0.54	0.58	0.68	0.71
JJA Av.	0.26	-0.04	0.06	-0.67	0.60	-0.65	0.00	-0.33	0.45	0.10	0.14	0.23	0.24	0.59	-0.33	-0.50	-0.59	0.10	-0.16
JJA Max	0.96	0.66	0.86	0.15	1.76	0.29	0.68	0.51	1.30	0.99	0.89	0.95	0.96	1.53	0.80	0.63	0.53	1.06	0.54
JJA Min	-0.74	-0.84	-0.84	-1.27	-0.49	-1.49	-1.37	-0.90	-0.10	-0.86	-1.44	-1.53	-1.46	-1.19	-0.83	-1.02	-1.15	-1.30	-1.89
JJA r	0.73	0.76	0.75	0.80	0.68	0.68	0.66	0.74	0.73	0.64	0.66	0.62	0.76	0.65	0.77	0.77	0.58	0.54	0.56
SON Av.	0.26	0.26	0.16	0.20	0.38	-0.85	0.46	0.31	0.70	0.77	-0.05	0.04	-0.14	0.11	-0.44	-0.22	-0.37	0.90	-0.07
SON Max	2.66	2.56	2.56	1.24	1.54	-0.24	2.12	1.39	2.17	1.59	2.83	2.89	2.77	3.00	2.21	2.27	2.21	1.76	2.70
SON Min	-0.64	-0.74	-0.74	-0.50	-0.47	-1.81	-0.23	-0.30	0.06	0.26	-0.99	-0.95	-1.14	-0.94	-1.39	-1.13	-1.27	0.36	-1.01
SON r	0.53	0.55	0.54	0.80	0.79	0.71	0.72	0.81	0.80	0.88	0.54	0.53	0.52	0.55	0.57	0.60	0.58	0.88	0.55
DJF Av.	1.06	1.06	1.26	1.80	1.41	-0.09	1.65	1.64	1.79	1.88	0.86	0.73	0.24	0.25	0.50	0.24	0.67	2.46	1.02
DJF Max	3.86	3.86	3.96	2.91	2.88	0.38	4.05	3.05	3.53	2.80	4.36	4.28	4.02	4.09	4.21	4.05	4.24	3.29	4.05
DJF Min	-0.04	-0.04	0.26	0.64	0.51	-1.36	0.65	0.75	0.90	1.24	-0.54	-0.66	-1.37	-1.20	-0.68	-1.02	-0.48	1.83	-0.31
DJF r	0.57	0.59	0.59	0.73	0.82	0.77	0.66	0.85	0.76	0.88	0.51	0.49	0.62	0.48	0.60	0.62	0.58	0.91	0.55

Precipitation

	HadRM3P-a	HadRM3P-b	HadRM3P-c	PROMES	RACMO	CHRM	CLM	REGCM	REMO	RCAO-H	HIRHAM-a	HIRHAM-b	HIRHAM-c	HIRHAM-E5	ARPEGE-a	ARPEGE-b	ARPEGE-c	RCAO-E4	HIRHAM-E4
MAM Av.	-0.39	-0.46	-0.37	-0.05	0.22	-0.36	0.84	0.41	0.93	0.48	0.15	0.23	0.27	0.28	0.44	0.40	0.77	0.38	0.26
MAM Max	0.59	0.57	0.70	0.62	0.87	0.28	1.89	1.30	1.76	1.03	1.17	1.31	1.45	1.43	0.93	0.90	1.22	0.93	1.19
MAM Min	-1.07	-1.28	-1.09	-0.86	-0.47	-1.01	0.09	-0.68	0.12	-0.30	-1.06	-0.97	-0.93	-0.86	-0.62	-0.74	-0.18	-0.75	-0.61
MAM R^2	0.71	0.66	0.65	0.79	0.73	0.87	0.77	0.51	0.75	0.71	0.75	0.77	0.75	0.78	0.68	0.66	0.70	0.59	0.71
JJA Av.	-0.83	-0.78	-0.87	0.39	0.34	-0.03	0.74	0.73	0.93	0.57	0.29	0.17	0.11	-0.18	-0.09	-0.04	0.08	0.62	0.58
JJA Max	0.57	0.59	0.50	1.28	1.19	1.31	1.36	1.80	1.74	1.19	1.45	1.34	1.23	0.86	1.07	1.02	1.16	1.29	1.76
JJA Min	-1.51	-1.47	-1.54	-0.33	-0.49	-0.80	-0.27	-0.62	0.09	-0.25	-1.07	-1.15	-1.21	-1.10	-1.27	-1.20	-1.21	-0.26	-0.88
JJA R^2	0.42	0.41	0.42	0.84	0.62	0.24	0.79	0.30	0.47	0.64	0.56	0.58	0.59	0.62	0.22	0.35	0.24	0.60	0.51
SON Av.	-0.77	-1.02	-0.89	-0.04	-0.10	-1.15	0.61	0.03	0.16	0.33	-0.53	-0.31	-0.50	-0.23	-0.69	-0.22	-0.41	1.56	0.65
SON Max	1.38	1.08	1.21	1.12	1.09	0.34	2.63	1.38	2.63	1.32	1.08	1.18	1.09	1.44	1.00	1.33	1.23	2.32	2.11
SON Min	-2.55	-2.75	-2.59	-1.29	-1.51	-2.64	-0.27	-1.72	-0.97	-1.28	-1.69	-1.53	-1.74	-1.31	-2.44	-2.01	-2.13	0.47	-0.36
SON R^2	0.49	0.48	0.53	0.82	0.76	0.89	0.79	0.67	0.73	0.83	0.78	0.79	0.77	0.81	0.65	0.74	0.70	0.85	0.85
DJF Av.	-0.63	-0.56	-0.63	0.36	0.69	-0.63	1.72	0.68	0.92	1.47	0.10	-0.14	-0.31	0.00	0.31	0.18	0.54	1.07	0.13
DJF Max	0.70	0.62	0.57	2.12	1.96	0.17	4.65	2.14	3.98	2.53	1.19	1.07	0.68	1.17	2.19	1.99	2.16	2.10	1.24
DJF Min	-1.81	-1.65	-1.62	-1.44	-0.64	-1.69	0.19	-0.80	-0.96	0.00	-0.72	-1.01	-1.17	-0.92	-1.44	-1.55	-1.11	-0.82	-1.06
DJF R^2	0.77	0.81	0.78	0.82	0.81	0.86	0.76	0.71	0.67	0.78	0.87	0.85	0.86	0.84	0.72	0.73	0.75	0.73	0.83

Table 5.1: Average bias (Av.), maximum individual gridcell bias (Max), minimum individual gridcell bias (Min) and Pearson coefficient of correlation between observed and modelled spatial pattern (r) for temperature (^oC, top) and precipitation (mm/day, bottom) for 1961-1990.

The green line denotes the range above which r values can be considered as high and indicative of skill at simulating the spatial pattern. Pearson r values can be between -1 and 1 but values closer to 1 signify a strong covariation. Therefore, 0.7 is chosen as the threshold above which models are considered as skilful. Immediately it is clear that with the exceptions of summer temperature, autumn and winter precipitation, most seasons do not have a large selection of models with Pearson r values of 0.7 or greater. The problem is compounded by the fact that in many cases, the highest r values are also accompanied by the largest biases, for example in the case of PROMES for summer temperature, RCAO for autumn and winter temperature and CLM for summer and winter precipitation The question from a climate planning and decision-making point of view is, in light of the various model uncertainties that have presented themselves, what constitutes the most reliable set of models? Arguably, a systematic bias is more desirable than a random one, as it has more potential to be accounted for in subsequent scenario development.

This analysis of mean spatial patterns for temperature and precipitation has uncovered a variety of spatial errors, but further analysis is required to determine the cause of those errors. Differences in how the various models resolve complex orography, how they respond to different types of land-use or how accurately the relevant large-scale circulations are simulated in both RCM and driving GCM can all influence the skill of the model and lead to better skill in one gridcell than another. There are a great number of processes that could potentially lead to the various model biases found in this analysis and verifying each parameterization and process that could be the cause falls beyond the scope of this thesis.

In light of this limitation, it is worth considering model spatial skill on a more macroscopic scale. Rather than attempting to validate models on a gridcell by gridcell basis, an alternative approach would be to consider how the models perform at simulating the key components of spatial variability. It has been shown that all models exhibit a range of biases in their mean seasonal patterns and as such, based on these results, none of the models can be considered “correct”. However, if they are able to capture the key components of spatial variability in spite of those biases, they may be potentially useful. Therefore, spatial variability will be assessed in more detail using Empirical Orthogonal Function (EOF) analysis.

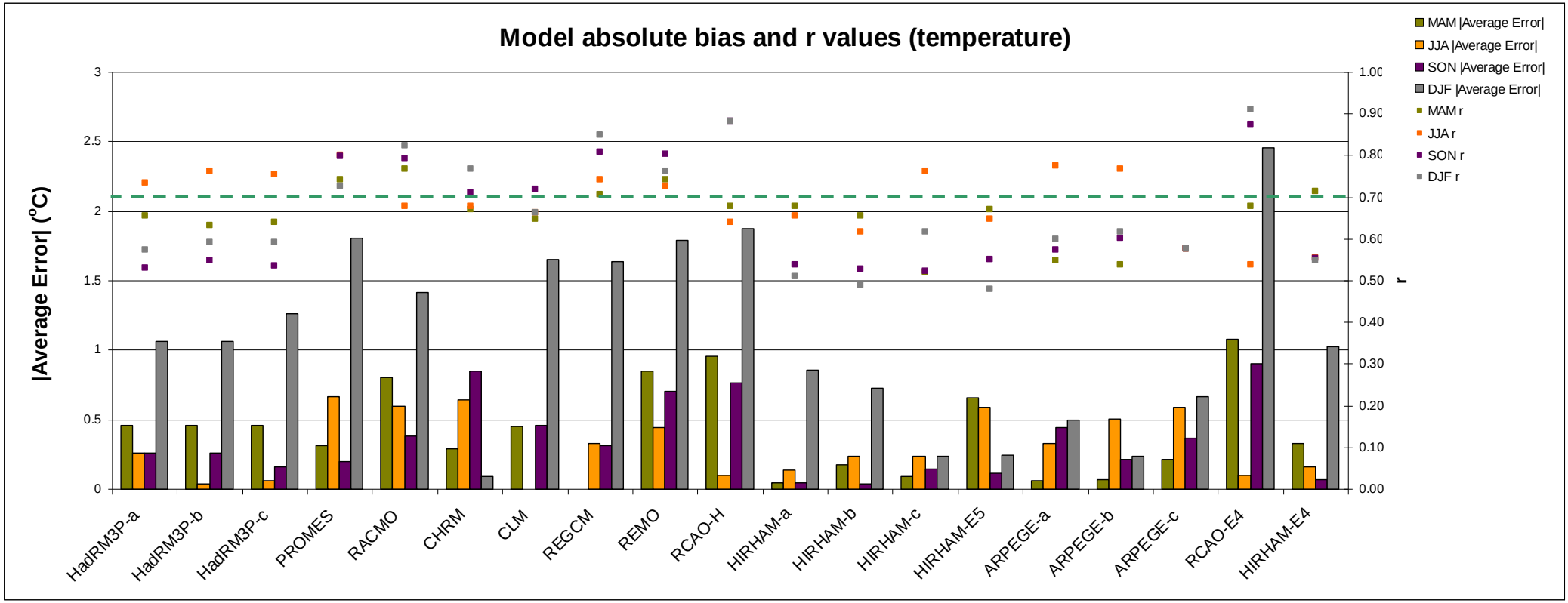


Figure 5.17: Graphs of model absolute bias and r values for temperature (1961-1990). The green line denotes the range above which r values can be considered to denote a strong association (greater than 0.7) between modelled and observed spatial patterns. Bars illustrate absolute value of average error (x-axis). Dots illustrate r values (y-axis).

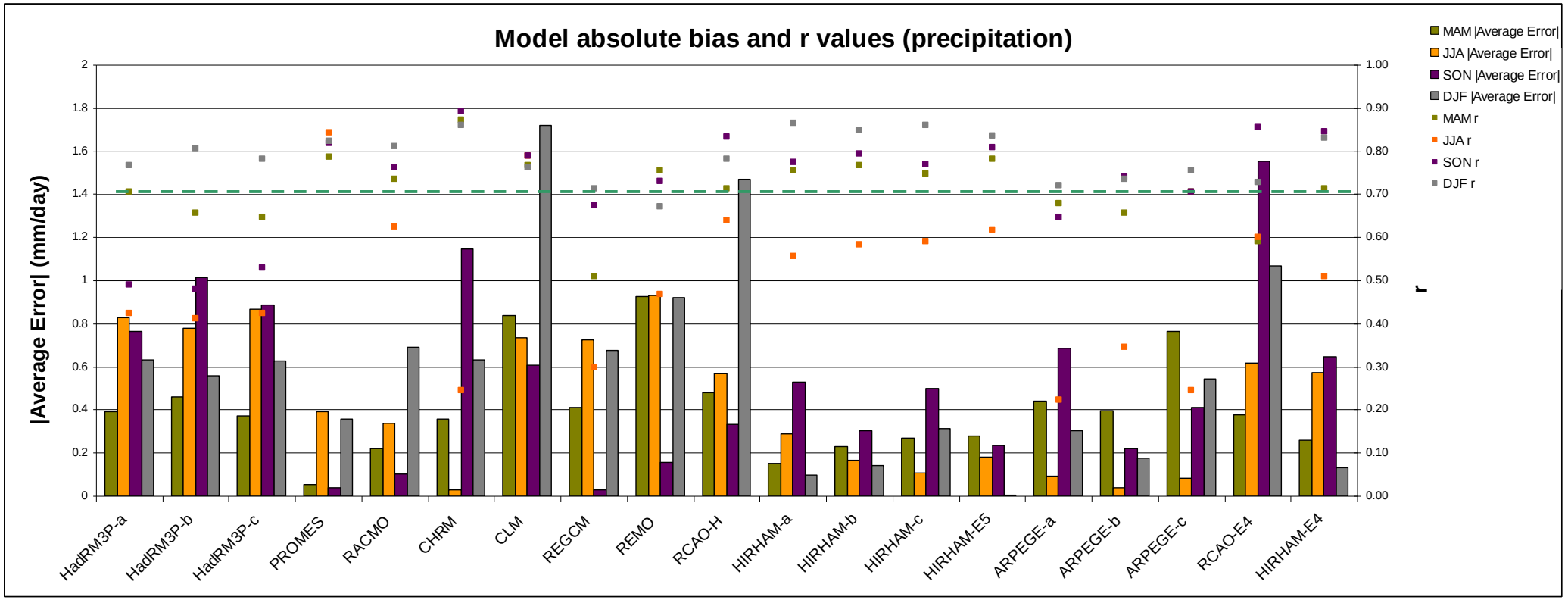


Figure 5.18: Graphs of model absolute bias and r values for precipitation (1961-1990). The green line denotes the range above which r values can be considered to denote a strong association (greater than 0.7) between modelled and observed spatial patterns. Bars illustrate absolute value of average error (x-axis). Dots illustrate r values (y-axis).

5.3 EOF ANALYSIS OF MONTHLY SPATIAL DATA

5.3.1 Background

A model may possess skill at representing key components of the pattern that is not apparent when looking at the mean distribution pattern alone. For example, it may capture the gradients of precipitation of the west and north-west while failing to simulate observed conditions in other parts of the country. Empirical orthogonal function (EOF) analysis provides a method for determining model skill in representing component patterns. Essentially, EOF analysis deconstructs interannual climate data into component patterns, each of which explain a certain amount of variance in the original data. These patterns can be extracted for both the observed and modelled data and compared. The method is widely used in climate science. For example Kim and North (1993) use the technique to study temperature variability in a stochastic climate model. Fyfe *et al.* (1999) applied EOF analysis to modelled data to determine how well models capture the Arctic and Antarctic Oscillations. In particular, Wang *et al.* (2006) applies this method to Irish climate data. However, only a single simulation from one climate model was used in that instance and only precipitation output was considered. As such, the work presented in this thesis provides a useful opportunity to identify whether the EOF technique can also be useful for Irish temperature data.

Peixoto and Oort (1992: 492-495) provide a clear introduction to the mathematical background of this method. There are some important nuances in the preprocessing of data for EOF analysis, which lead to different variants of the method and which will be discussed later. Consider a spatial climate data-set composed of N maps, each defined at a particular step in time, and each with M elements. Each map can be represented by a Mx1 column vector f_n. If we consider the full time series as an array of these column vectors, we arrive at an MxN matrix:

Equation 5.3: Spatial climate data matrix

$$F=\begin{bmatrix} f_{11} & f_{11} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{bmatrix}$$

where

rows represent the M points on each map and

columns represent the N maps.

Therefore, element f_{mn} represents the model output for gridpoint *m* at time *n*. This is the data matrix employed in EOF analysis.

The N vectors exist in an M-dimensional linear vector space, directed from the origin to some point in that space. If there is correlation between the vectors, the expected result would be the formation of clusters, as the vectors tend towards a preferred direction. To determine whether such correlations exist, the orthogonal basis in the vector space $\{e_1, e_2, \ldots, e_M\}$ is defined so that each vector e_m best represents the clustering of the f_n map vectors. This set of vectors, {e} are the empirical orthogonal functions.

Equation 5.4: {e}, the set of empirical orthogonal functions

To find {e} the expression

$$\frac{1}{N}\sum_{n=1}^{N}\left[f_n \cdot e_m\right]^2$$

is maximized for m=1,2,…,M subject to the conditions

$e^T_m e_j = 0$ *and*

$e^T_m e_m = 1$ *and for all* $j \neq m$.

As the vectors of {e} are assumed to be mutually orthonormal (i.e. both orthogonal and of unit length), these conditions are derived from the definition of the

dot product for spatial vectors. Noting that the transpose of the product of two matrices is the same as the product of the transposes in reverse order and considering the data matrix F, the expression becomes:

Equation 5.5: To calculate the diagonal matrix of eigenvalues

$$\frac{1}{N}[F \cdot e_m]^2 = \frac{1}{N}[e_m^T F\ F^T e_m] = e_m^T R\, e_m$$

where

R =the covariance matrix of the data.

Maximizing the expression results in the equation:

$$(R-\lambda I)e_m=0$$

As *I* is the unit matrix of order M, λI is a diagonal matrix with the eigenvalues as diagonal elements. The sum of the eigenvalues is the total variance of the data. To calculate the associated eigenvectors, the system of equations associated with the equation $(R-\lambda I)e_m=0$ must be solved for each eigenvalue. To solve, we utilize the point that as $e_m \neq 0$, it follows that $|R-\lambda I| = 0$ if solutions are to exist. By arranging the eigenvectors in descending order based on their associated eigenvalues, one can determine which eigenvectors explain the greatest amount of variance. The eigenvalue can be expressed as a percentage, which is the percentage variance explained by each eigenvector.

The eigenvectors are independent and mutually orthonormal. They represent key components of the spatial patterns, such that any of the original data vectors f_n can be expressed as a linear combination of eigenvectors. As a different combination is required to represent each original data vector, each eigenvector e_m has a different coefficient in each time-step, which denotes the weight of the component represented by e_m, at times-step N. Plotting these coefficients over time for a particular eigenvector gives a visual indication of how the importance of that mode changes over time. This time amplitude series accompanies each spatial pattern and also illustrates how the actual pattern at each time step compares with the spatial loading pattern in question. Positive peaks signify time steps at which the actual pattern was

very similar to the spatial loading pattern, and negative troughs signify timesteps at which the patterns were very different.

With the mathematics of the analysis established, the first stage in EOF analysis is preparing the data, and this is the point at which the mode of decomposition is selected. Various methods are available, which use either the covariance matrix or the correlation matrix. Covariance is a measure of the degree to which variables vary together, and when generated, the covariance matrix describes this linear coupling for the elements of the input data matrix. The correlation matrix is generated by standardizing the data prior to the generation of the covariance matrix.

S-mode analysis uses the covariance matrix. It requires that that the climatological mean at each point be subtracted from the data, to generate time-centred data, before the analysis. The aim of this operation is to remove the component of the data which is relatively easy to discover, i.e. the mean, so that the attention of the subsequent analysis is entirely focused on the variability around the mean, i.e. the anomaly (van den Dool, 2007: 12). R-mode analysis (Davis, 1973: 503), so named because it uses the correlation matrix, is identical to S-mode apart from the standardization procedure.

Both methods are used when the aim of the analysis is to examine the spatial patterns in the data. Other methods exist, such as T-mode analysis, for analyses where the focus is on the temporal evolution of the spatial patterns. T-mode uses the transpose of the S mode input data, making the focus of the analysis patterns in time periods rather than space (Mohapatra *et al.*, 2003). As the focus of this work is spatial patterns, the choice is between S-mode and R-mode, and the covariance or correlation matrix.

The correlation matrix is a necessity when data with different units is being analyzed. Wilks (2006: 471) explains that if one performs an EOF analysis using precipitation data in inches and temperature data in ^{O}F with no prior standardization, the variance of the temperature data dominates. In these units, the range of variation of temperature appears larger than the range of variation in precipitation. Standardizing the data and using the correlation matrix instead makes the two datasets comparable. A unit change in one is equivalent to a unit change in the other,

and the analysis can be carried out without interference from artefacts of different measurement systems. As the input data in the analysis does not fall into this category, the mode of decomposition is not preordained by the data and the merits of the two methods are assessed to determine the more suitable one.

In a discussion of eigenvector-based map-pattern classifications, Yarnal (1993: 82) advocates using the correlation matrix so that variance gradients do not impact the analysis. If the variances from point to point in the input data differ greatly, the regions with greatest variance will dominate the analysis, and potentially distort the spatial patterns. Overland and Preisendorfer (1982) explain that this sensitivity of the covariance matrix to variance gradients is not necessarily a negative point, as the method can be employed specifically to uncover regions of unusually high variance.

Overland and Preisendorfer (1982) also states that the correlation matrix is particularly suited to detecting spatial patterns. When using a covariance matrix, the diagonal elements are the variance of each point. The result is that in addition to covariance between points, variance at each point contributes to the formation of the eigenvector also. Conversely, the diagonal elements of the correlation matrix will always equal one, as these are the correlations of each point with itself, so only off-diagonal elements, i.e. the correlations between points, will contribute when determining eigenvectors.

An issue encountered when performing an EOF analysis is whether or not to rotate the eigenvectors. By removing less significant axis, the principal axes are allowed to rotate further, such that their loadings are either significantly high (plus or minus 1) or insignificant (0). This makes the resulting spatial modes somewhat easier to interpret (Wilks, 2006). If one chooses to rotate there are multiple methods that can be applied. Rotation can be either orthogonal (varimax rotation scheme) or oblique. Oblique rotation is more extreme than orthogonal, making it even easier to interpret the resulting EOFs, but because orthogonality is not preserved, it is possible for axes to be correlated with each other. Using oblique methods, it would be possible to recapture the original variables as factors (Davis, 1973: 517), a futile result in exchange for a time-consuming analysis. Using varimax rotation keeps the axes at right angles with each other, preserving orthogonality. However, the first

mode is no longer the pattern that explains the most variance, as the axes do not coincide with the principal axes of the variance-covariance ellipsoid. The variance is spread more uniformly amongst the rotated eigenvectors (Wilks, 2006).

Sometimes it is necessary to rotate axes. For example, rotation can be applied to remove Buell patterns (Yarnal, 1993) which are statistical artefacts unrelated to spatial variation in the data. Buell patterns can occur when using a rectangular domain in which variation at each grid point is strongly correlated with that of the neighbouring grid point. They manifest as a monopole first mode, as the central grid points have the highest correlations with their neighbours, and a dipole second mode orientated in the longest spatial dimension, as the points furthest apart appear have the strongest negative correlations (Figure 5.19). As the Irish domain is significantly longer along its y axis that on its x axis, it may be susceptible to Buell patterns and this issue will be examined in further detail.

Varimax rotation is also used to obtain localized modes (e.g. Hannachi *et al.*, 2007; Wang *et al.*, 2005). Yet Dommenget *et al.* (2001) demonstrate that both regular EOF methods and varimax methods do not necessarily reproduce the centres of action of the real modes. So while there are situations in which varimax rotation has its uses, it does not guarantee "improved" results. Climate model skill assessment is already a subjective science, with many assumptions being made. Without a reasoned argument for applying a rotation, the procedure is likely to add another layer of subjectivity in an already biased process (Davis, 1973: 517). One strategy would be to apply EOF analysis first and then assess whether rotation could be beneficial.

Once generated, the principle EOFs, those that explain the highest proportion of variance in the original data, can be extracted and plotted as contour maps. One can then identify centres of activity and determine whether activity in particular areas is related, or inversely related, or not at all related. Two key patterns are the monopole and the dipole (Figure 5.20). The monopole describes a pattern with positive loadings on all the grid points, meaning that activity at all points is related. The dipole describes a pattern with both positive and negative loadings on various areas of the grid, meaning that activity is inversely related in those areas. However, it is important to keep in mind that the patterns revealed in the modes do not always

represent physical phenomena. The investigator must use his or her own discretion to determine whether the pattern is significant or simply a statistical artefact.

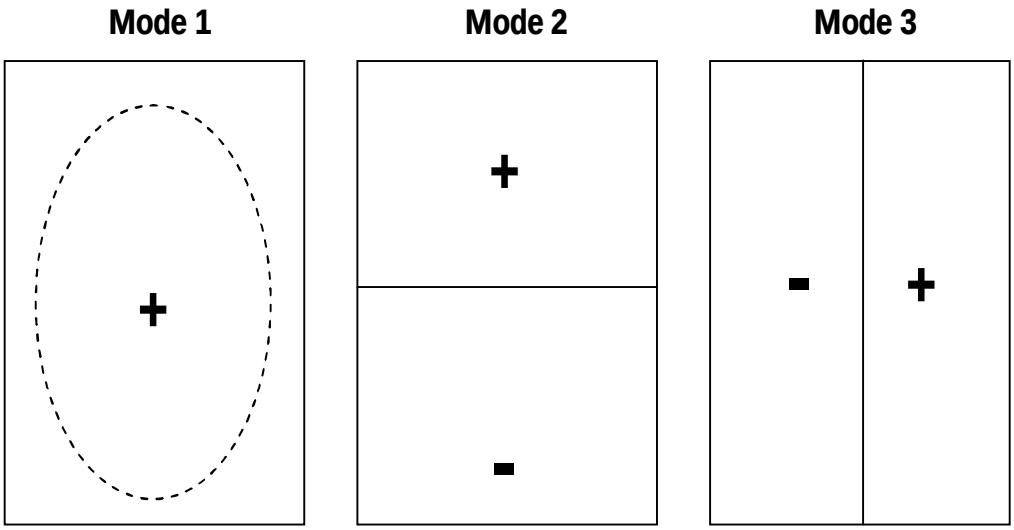


Figure 5.19: Example of typical Buell patterns.

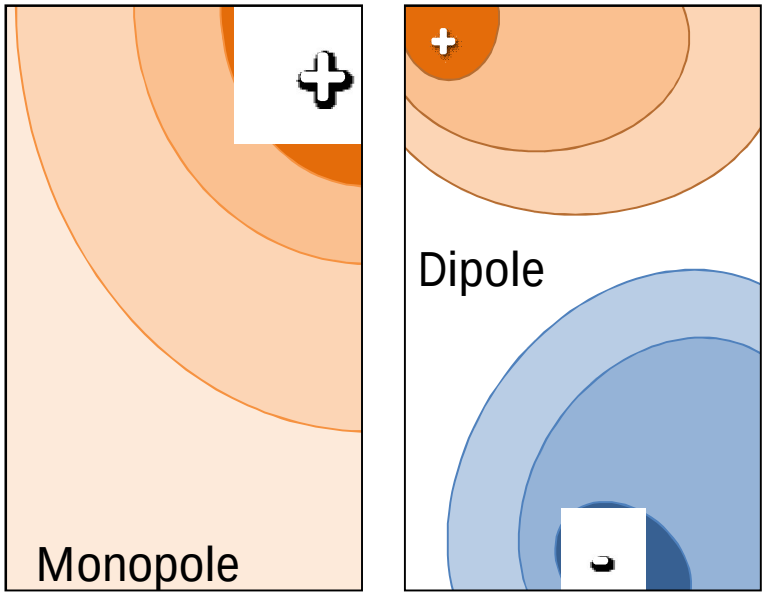


Figure 5.20: Example of monopole (left) and dipole (right) patterns.

5.3.2 Methodology

The EOF approach applied in this analysis is a correlation matrix, rotated approach. While the covariance matrix is useful for identifying areas of greatest variance or “action centres” within data, the correlation matrix is used here as the

goal is to extract the key spatial patterns within the data. The correlation matrix is generated by standardizing the data prior to the generation of the covariance matrix.

To determine whether or not the modelled output is susceptible to Buell patterns, EOF analysis was carried out on the January and June datasets for both temperature and precipitation. The output was both rotated and left unrotated and compared to the patterns described by Buell (1975) and further discussed by Richman (1986).

Figure 5.21 shows the first three EOF patterns from the HadRM3P simulation using January precipitation data, first without rotation and second with rotation applied. Without rotation of the eigenvectors, the EOF patterns are typical Buell patterns consisting of a "duck egg" positive pattern for mode 1 and positive/negative patterns centred along the longest horizontal and vertical axes of the domain for modes 2 and 3.

Several models were checked at random to determine whether a rotated approach was required. The non-rotated patterns in Figure 5.21 are illustrative of the behaviour of all models tested. That is, without rotation, all models tested produced Buell patterns. Therefore, a rotated approach was adopted to remove Buell patterns from the EOF modes. The EOF analysis was applied to the January, April, July and October monthly data for each model. One month from each season was chosen. The first five EOF patterns are mapped and similarities or differences between these and the first five observed EOF patterns are discussed.

For January and July, the associated time amplitude functions are graphed also. The temporal component of the EOF analysis is examined in less detail here than the spatial component as the focus of this chapter is spatial patterns and the temporal analysis carried out in the previous chapter identified that interannual variability is highly influenced by the choice of driving GCM.

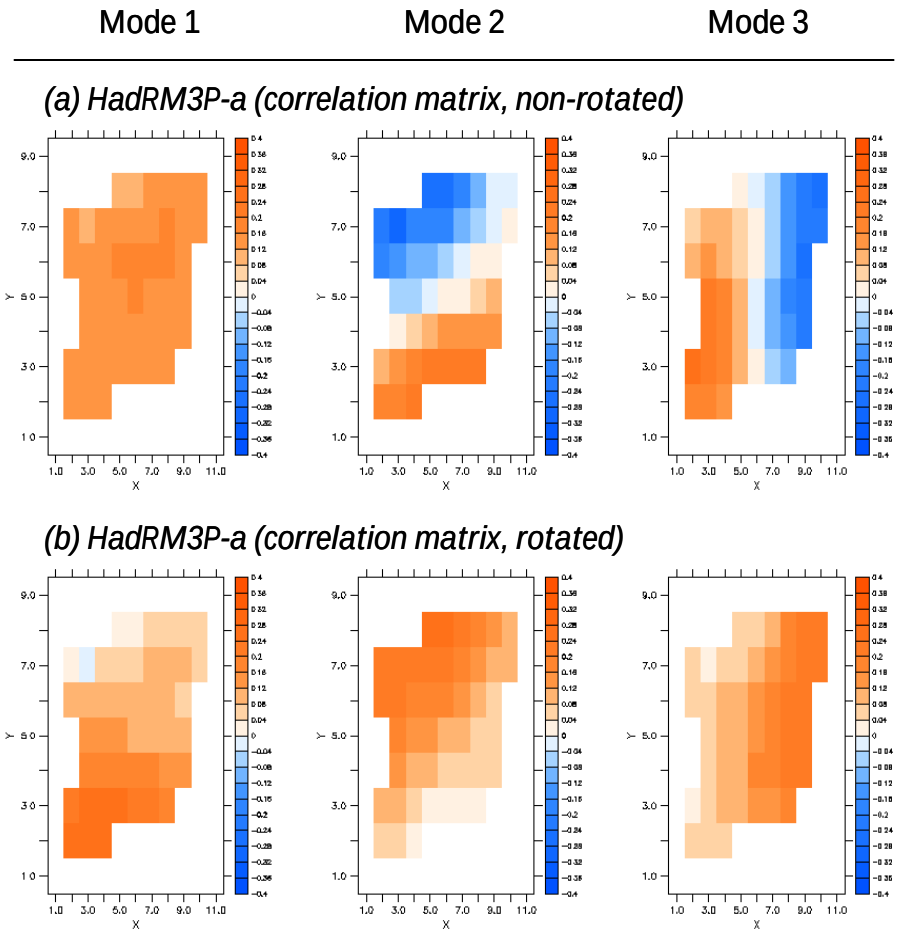


Figure 5.21: The first three EOF patterns from the HadRM3P simulation using January precipitation data, first with rotation omitted (a) and second with rotation applied (b). Patterns obtained without rotation of the axes are typical Buell patterns.

However, examination of the time amplitude data will help to confirm the findings of the previous chapter. Models with the same GCM driver are plotted together, to determine whether their time amplitude series have any shared characteristics. As such, only the models driven by HadAM3H and ECHAM4-OPYC are included as the limited number of simulations available for the other drivers limits the ability to draw such comparisons. Similar spatial patterns from different RCMs are identified visually and plotted together, regardless of the order they appear in the EOF analysis of the individual model, allowing the time evolution of particular patterns to be examined.

5.3.3 Results: Observed patterns versus modelled patterns

Percentage variance explained was plotted against EOF mode number to determine the "nick point" within the datasets, that is, the EOF mode number at which percentage variance explained falls off to insignificant levels. From the graph (Figure 5.22) it is clear than in most cases, percentage variance explained falls off to less than 5% after the third EOF mode. The only exception is precipitation in July. Modes 3 and 4 in this dataset have similar levels of percentage variance explained and after mode 4, the percentage variance explained falls to less than 6% after that. This may reflect the more localized nature of summer precipitation patterns. While large-scale drivers play a major role in determining winter precipitation in Ireland, in summer, local factors such as sub-gridscale effects are likely to dominate. As such, modes after the third mode are omitted from the analysis, with the exception of July precipitation.

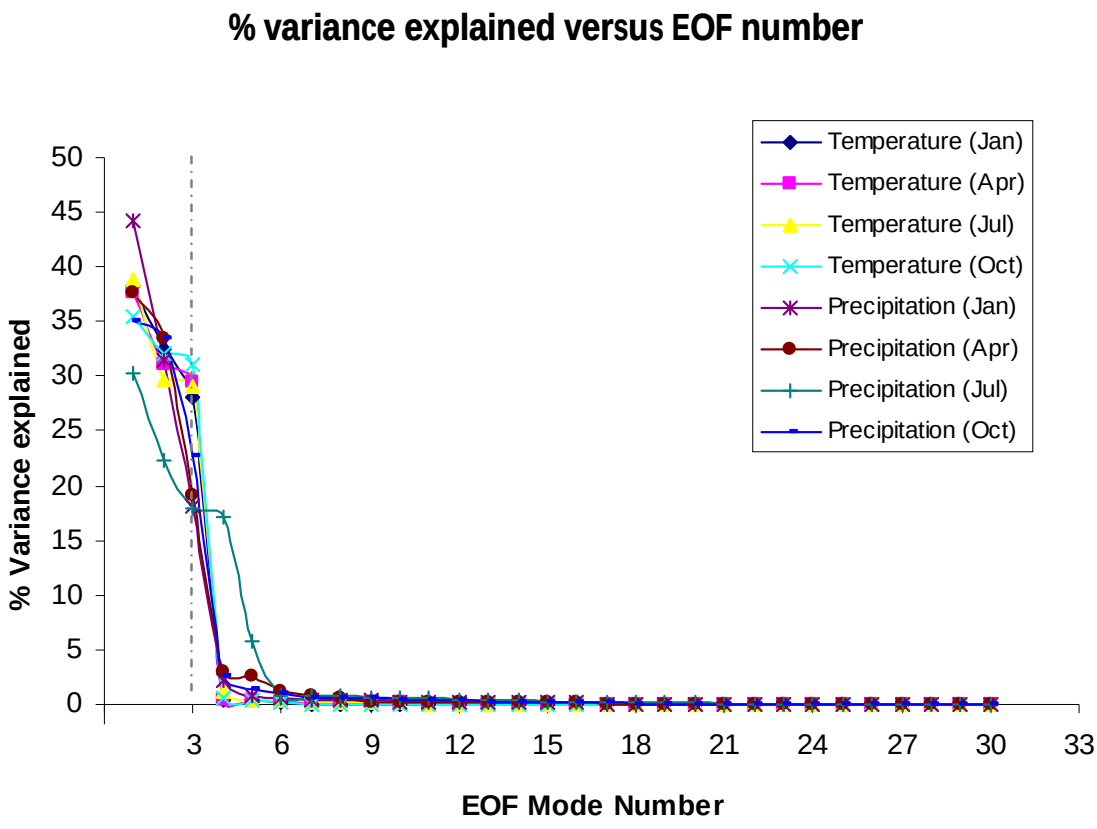


Figure 5.22: Percentage variance explained versus EOF mode number for observed temperature and precipitation. With the exception of July precipitation, percentage variance explained falls off to less than 5% after the third EOF mode.

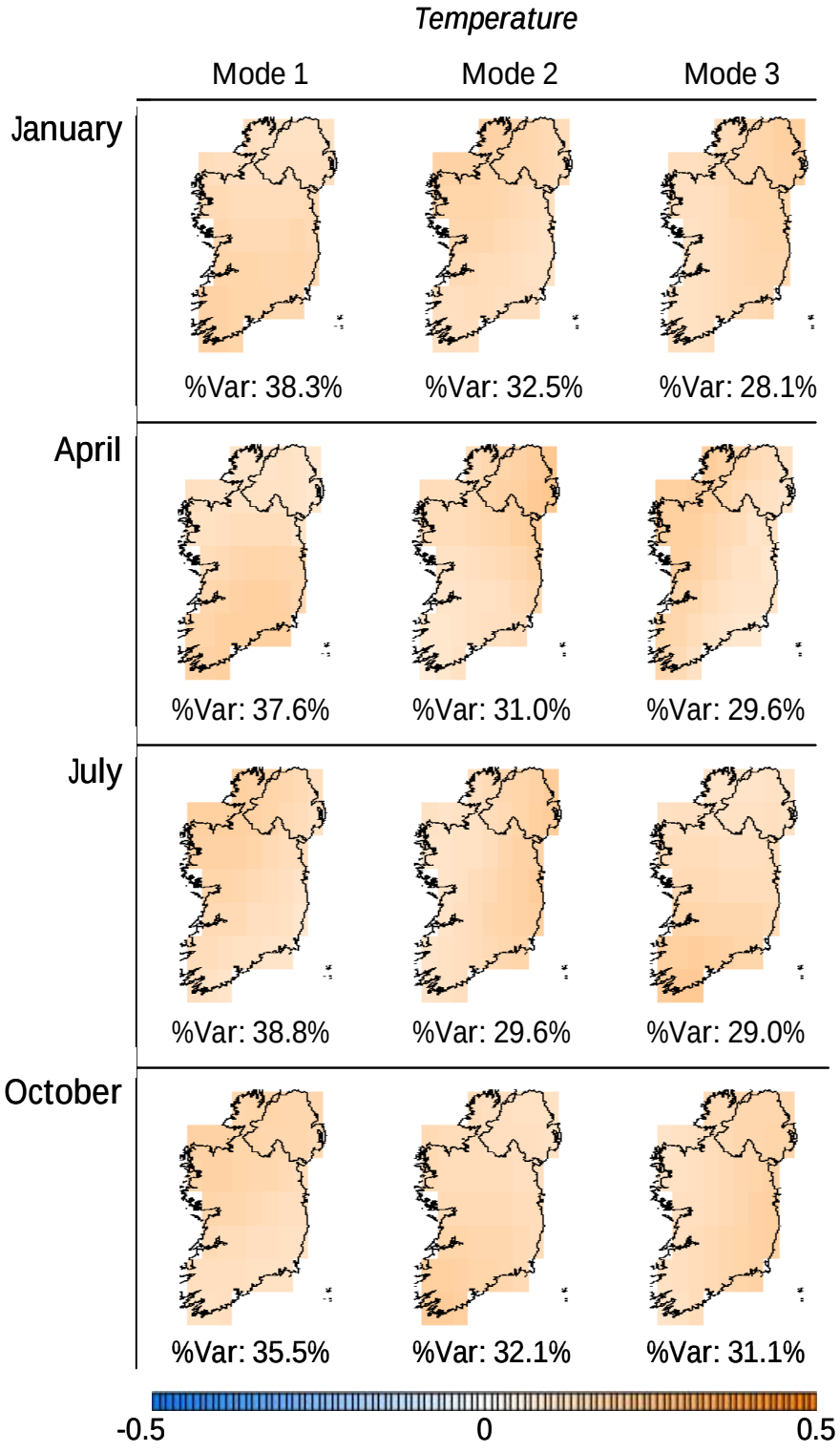


Figure 5.23: First three EOF patterns for observed temperature for 1961-1990. Patterns after the third are omitted as the nick point in percentage variance explained is reached at 3 modes.

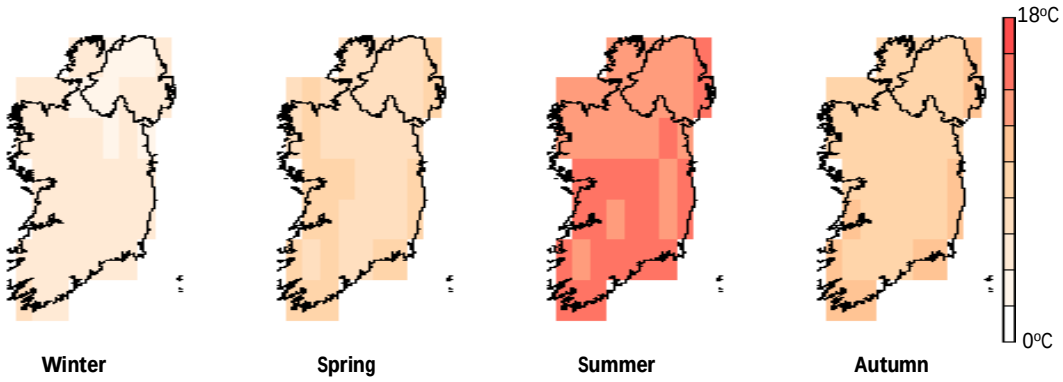


Figure 5.24: Observed seasonal temperature patterns for 1961-1990.

EOF patterns for observed temperature are shown in Figure 5.23. Temperatures do not vary much across Ireland (Figure 5.24) and as such, the EOF patterns for temperature do not have clear action centres and are quite uniform across the country. There are no particular areas which contribute significantly to a single EOF pattern, but rather all gridcells contribute suggesting that the temperature EOF patterns are based on large-scale rather than regional processes and effects.

The percentage variance explained by each pattern also does not vary significantly from month to month, suggesting underlying processes which do not have a temporal component. For example, in January and April, the first EOF mode pattern is a north-south low to high gradient, potentially indicative of the effects of latitude on temperature. This pattern is also evident in Mode 3 for July and Mode 2 for October.

Another pattern which appears in each month is a east or north-east to west high to low gradient. This is the third mode in January and October and the second pattern in April and July. Such a pattern may be related to orographic details. Temperature decreases with altitude and terrain in the west and south-west is more mountainous than in the east.

In July and October, there is an action centre situated in the north-west. This pattern also occurs in January, where it is the second pattern and in April where it is the third pattern. While the south coast is warmer than the north coast, the coolest areas of Ireland are those inland (Figure 5.24). Therefore, this pattern may represent a portion of the coastal-inland temperature difference that was not captured in the south-north gradient pattern.

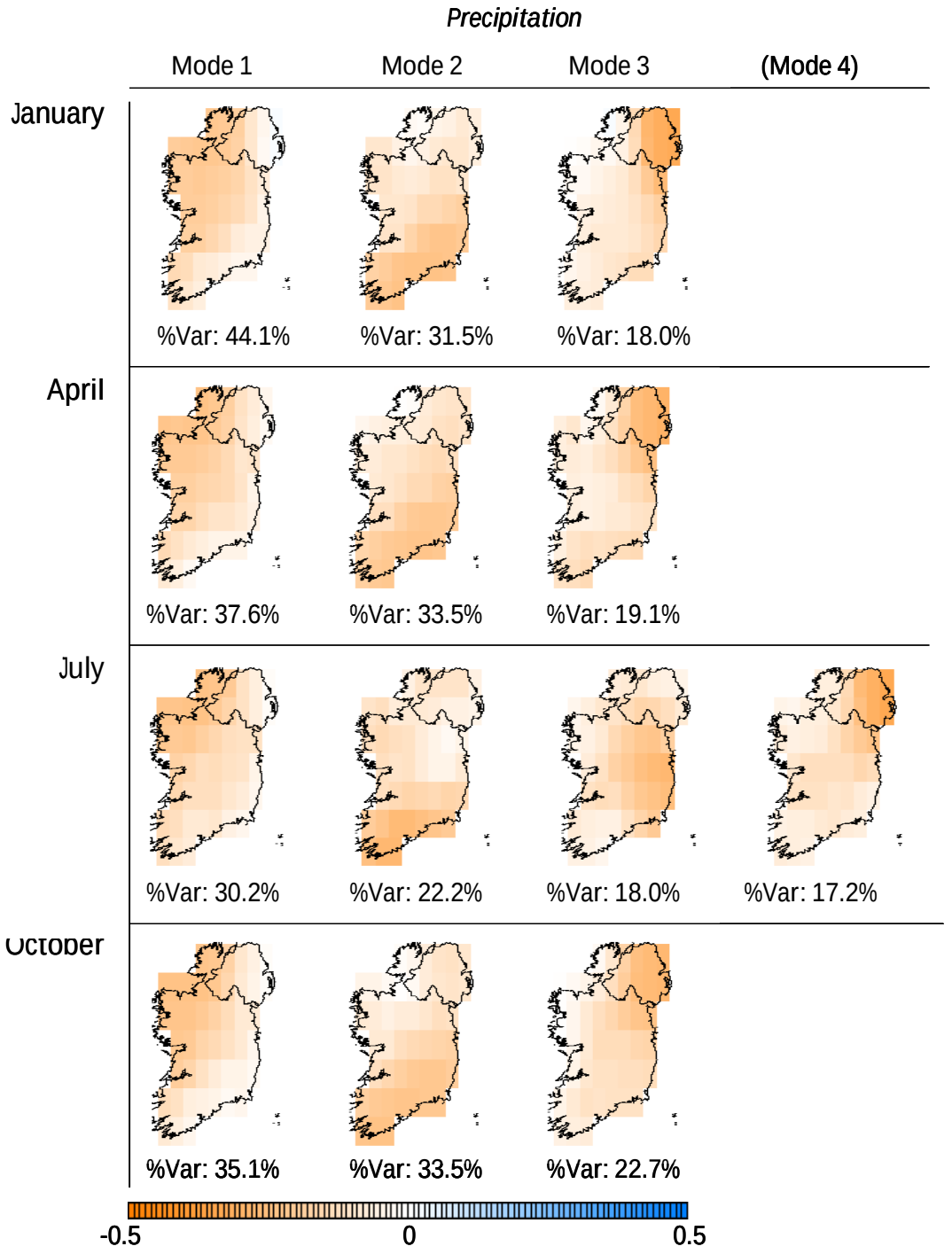


Figure 5.25: First three EOF patterns for observed precipitation for 1961-1990. Patterns after the third are omitted as the nick point in percentage variance explained is reached at 3 modes, except in the case of July, where a fourth mode is included.

Both EOF patterns based on monthly observed precipitation data and the actual observed seasonal spatial patterns for precipitation are given in Figures 5.25 and 5.26 respectively. Observed patterns are given to help elucidate the physical processes, if any, underlying the EOF patterns. The first EOF pattern consists of an action centre in the north-west of the country. This pattern is consistent in all four months analyzed The combination of prevailing westerly winds and mountainous terrain in this region results in high levels of precipitation in this area, particularly in winter months, so this pattern does have a physical basis. The second EOF pattern is also consistent across all four months and consists of an action centre situated in the south to south-west. Again, the combination of westerly airflow from the Atlantic and mountainous terrain in this area leads to high levels of precipitation variability, particularly in winter.

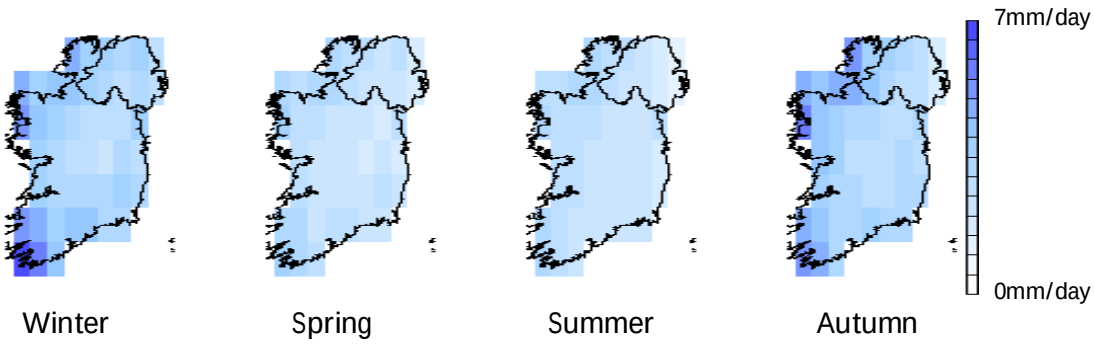


Figure 5.26: Observed seasonal precipitation patterns for 1961-1990.

The percentage variance explained by these patterns varies in each month, reflecting the relative influence of the underlying processes at different times of the year. For example, the Mode 1 pattern explains 44.1% of variance in January, but this decreases to 37.6% in April, 30.2% in July and increases again to 35.1% in October. This may indicate the strengthening in winter of westerlies due to the NAO and the weakening of the westerlies again in spring and summer as the NAO breaks down. Conversely, percentage variance explained by the Mode 2 pattern varies slightly less, remaining at 31.5% to 33.5% in January, April and October, but falling to 22.2% in July. Wang *et al.* (2007) note that this increased spatial variability in Irish summer precipitation may be due an increase in convective rainfall in summer, which tends to be intense in nature and distributed over a small area.

Convective rainfall may be the process underlying the third mode of July variability, which explains 18% of variance in July precipitation. This pattern has an action centre situated on the east coast, approximating the greater Dublin area. This is likely to be due to the proximity of this area to the Wicklow Mountains. As the Wicklow Mountains are a plateau, their orography and the rainfall associated with them is likely to be better resolved than mountain ranges to the west which contain tall peaks. Urban effects may also play a role. Bornstein and Lin (1999) suggest that the urban heat island effect of a city can induce a convective zone in summertime, leading to increased precipitation.

The third precipitation EOF pattern in January, April and October and the fourth pattern in July is one with an action centre situated to the north-east. This again may be due to the orographic rainfall effects produced by certain mountain ranges in this area. As this area is sheltered from the prevailing winds to a certain extent by the more mountainous west coast, this pattern explains a smaller percentage of the overall variability.

An interesting test of model skill is to examine whether the models capture these spatial patterns and the percentage variance associated with each pattern. The spatial patterns of precipitation relate strongly to terrain type and large-scale circulation patterns. Failure to capture these patterns may indicate a lack of skill in resolving orographic or in modelling how the prevailing winds change in strength throughout the year. In particular, several studies have demonstrated the link between the NAO and precipitation variability in the north-west of Ireland (e.g. Wibig, 1999; Murphy and Washington, 2001), so failure to simulate spatial patterns and percentage variance explained with skill may indicate deficiencies in the models' representations of large-scale drivers such as this.

5.3.4 Results: RCM temperature

Spatial EOF patterns and the variance associated with them for modelled January temperature data are given in Figure 5.27. Although the lack of spatial variability makes specific patterns difficult to discern, it appears as though the models capture the observed EOF patterns well. However, the percentage variance associated with the spatial patterns differs from the observed.

While ARPEGE-a, ARPEGE-b and ARPEGE-c have a first EOF pattern similar to the observed, with a low to high gradient running from north to south, in most models this pattern is the second of third EOF pattern (for example, HIRHAM-a, HIRHAM-b, HIRHAM-c, HIRHAM-E5, REGCM, REMO, RCAO-H, CLM). As such, the percentage explained variance associated with this pattern is lower in many of the models than in the observed EOFs. For example, while 38.3% of variance is associated with the north to south pattern in the observed data, in REGCM this pattern is the third EOF pattern and is only associated with 24.4% of variance.

Results suggest that while the patterns themselves are represented in the models, the extent to which they reflect temperature variability is not represented as well. In the observed data, each of the first three patterns for January is associated with a significant portion of explained variance, with 28.1% of variance attributed to the third EOF pattern. In many of the models, these proportions are approximated well. However, in some cases the variance values do not match the observed values. For example, ARPEGE-b and ARPEGE-c overestimate the variability associated with the first pattern, with 46.6% and 42.3% of variance associated with their first EOF pattern but only 10.8% and 11.3% respectively associated with their third. The ARPEGE simulations scored quite highly in the temporal skill analysis of Chapter 4, yet there are indications that the processes underlying the skilful mean values may not be well-represented in the model.

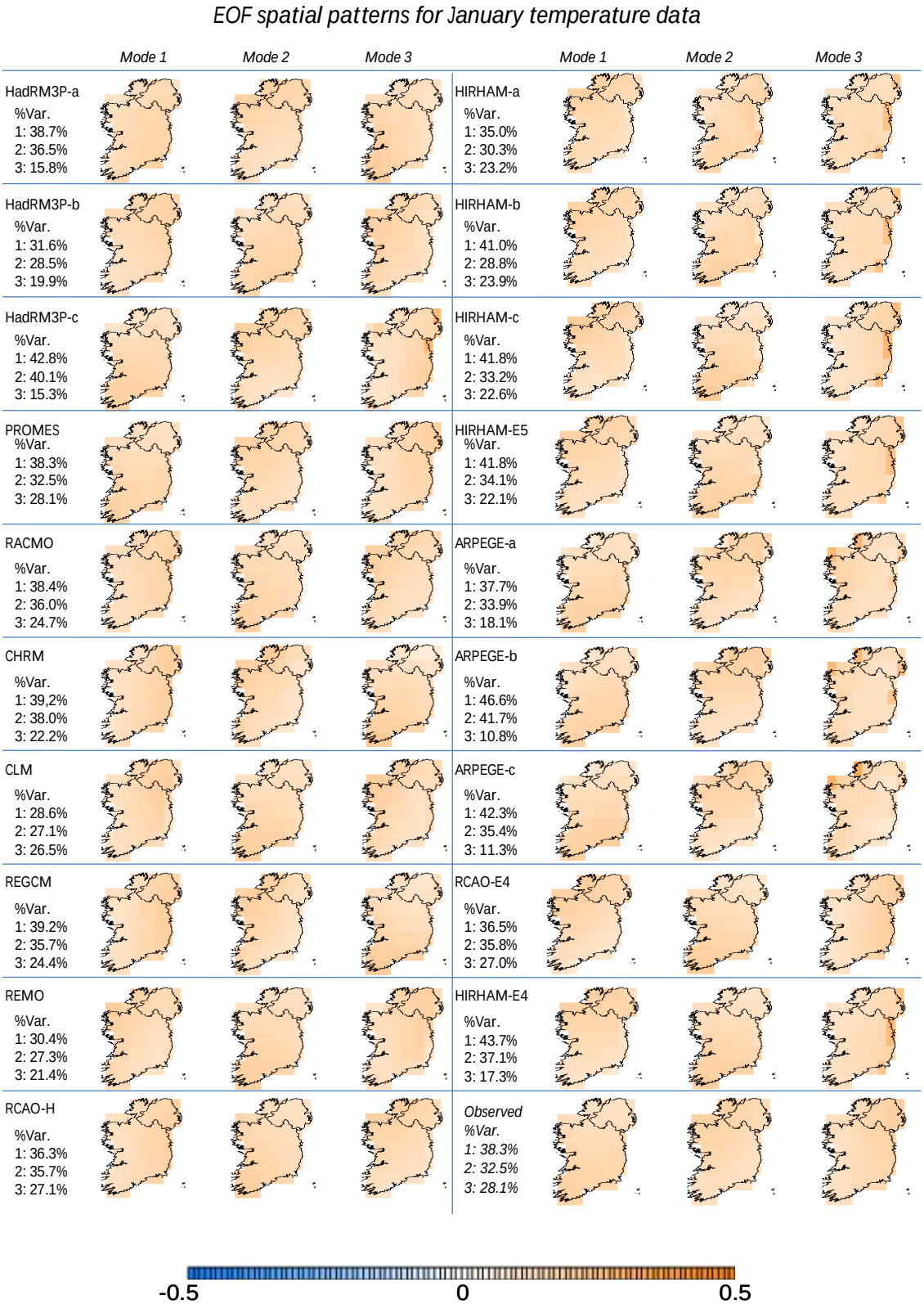


Figure 5.27: Modelled EOF patterns and associated percentage variance explained for January temperature data (1961-1990).

Observed EOF patterns for April are quite similar to those observed in January and the observed percentage variance explained values are also very close in both months (Figure 5.28). Interestingly, the model values for the percentage variance explained by April EOF patterns are very different to the observed, with many models displaying much higher values associated with their first pattern and much smaller values associated with their second pattern.

The most extreme example is HIRHAM-c, which has an associated percentage variance value of 80.7% attached to its first EOF pattern, compared with the observed EOF value of 37.6%. The models which associate very high levels of variance with their first mode also display a different and very specific pattern that is not present in the observed EOF patterns. HIRHAM-b, HIRHAM-c, HIRHAM-E5 and HIRHAM-E4 all have strong action centres in specific gridcells along the east coast in their mode 2 pattern and these same gridcells contribute less than the others in the mode 1 pattern. Results suggest that the processes or phenomena dominating temperature variability in April in these simulations are very different to the processes that influence observed temperature variability. Additionally, it is interesting to note that all these simulations are variations using the HIRHAM model. This strongly suggests an issue specific to this RCM, as even with different driving GCMs the same pattern appears.

However, other models that do capture the observed patterns also tend to misrepresent the associated variance explained. For example, ARPEGE-a, ARPEGE-b, ARPEGE-c and CLM all approximate the correct pattern for EOF mode 1 but associate excessive percentages of variance with the pattern when compared to the observed EOFs. Again, this suggests that the importance of this pattern and the processes that give rise to it are overestimated by these models. Alternatively, it may also suggest that other processes that influence temperature in April are not as dominant as they should be within the models.

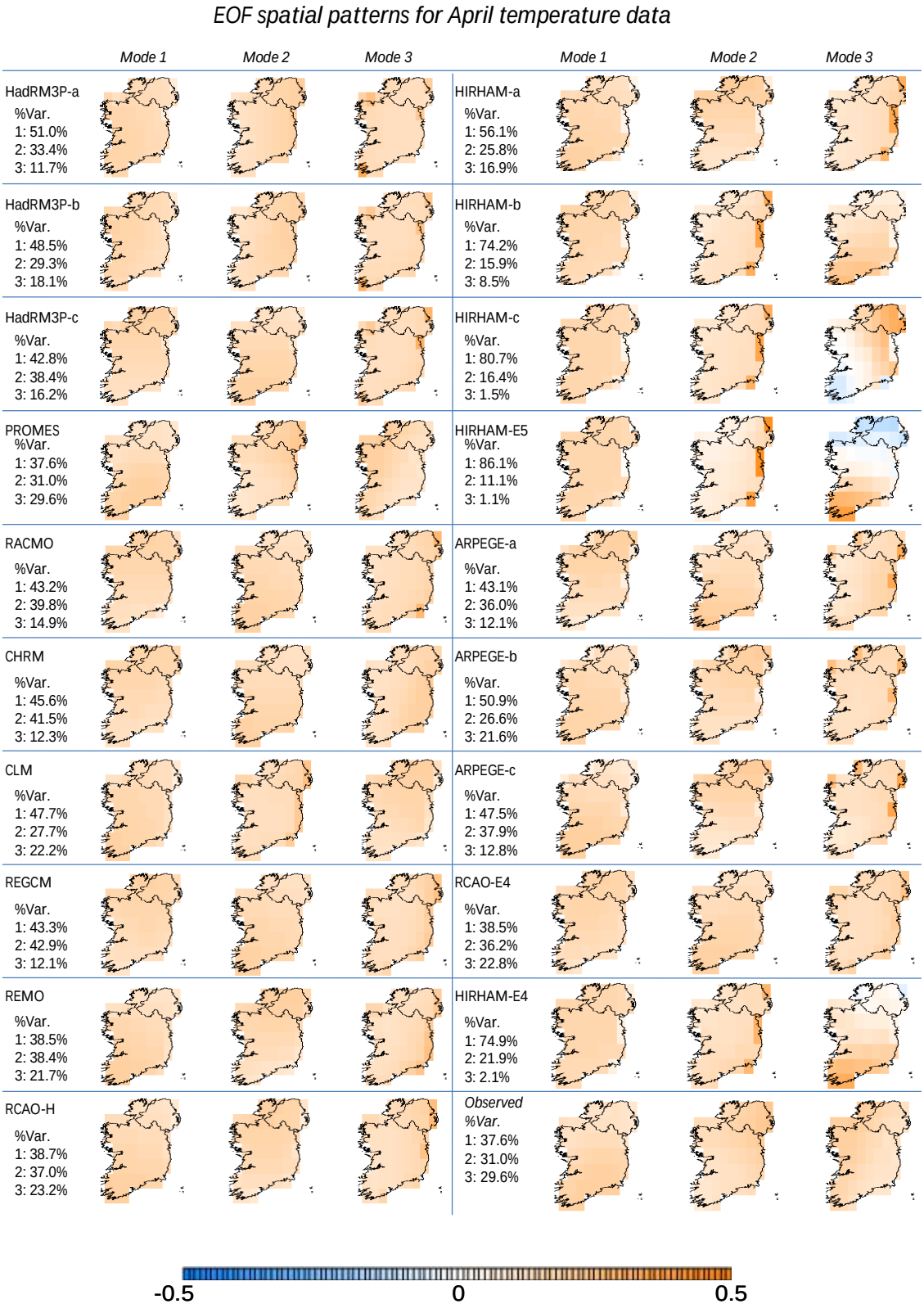


Figure 5.28: Modelled EOF patterns and associated percentage variance explained for April temperature data (1961-1990).

Summer temperature is known to be more homogeneous than winter temperature so it is not surprising that the modelled EOF patterns display more spatial variability in July than in previous months (Figure 5.29). However, the observed EOF patterns for July did not display this level of spatial variability and were in fact very similar to the patterns obtained in previous months. This suggests that the models in some cases are overestimating the importance of small-scale regional temperature effects, leading to excessive levels of variance being associated with patterns and processes that are not that dominant in reality.

In the HIRHAM simulations, specific gridcells on the east coast form a spatial pattern. It is worth noting that when the average spatial patterns for summer were examined, the HIRHAM simulations were found to overestimate temperature except in those gridcells, where temperature is underestimated. As such, the variability of these gridcells is so different to the others that the EOF technique identifies a separate mode for them.

However in July, there is a similar issue with the ARPEGE simulations. ARPEGE-a, ARPEGE-b and ARPEGE-c all have a third EOF mode with strong actions centres in specific gridcells in the north-west and north-east. This pattern is associated with between 9.7% and 13.6% of variance in these simulations. In the seasonal spatial patterns analysis, these gridcells were found to behave differently to their surrounding gridcells, modelling slightly warmer temperatures, but this difference in behaviour is actually much more pronounced in the winter and spring mean seasonal patterns. As such, it is surprising that the EOF patterns did not capture it in January and April.

EOF patterns for October temperature data are given in Figure 5.30 and display similar characteristics to other months. In general, the spatial patterns are well represented, though the associated percentage variance varies. Additionally, HIRHAM and ARPEGE simulations still display very specific EOF patterns which suggest that these models have difficulties resolving temperature correctly in these particular gridcells.

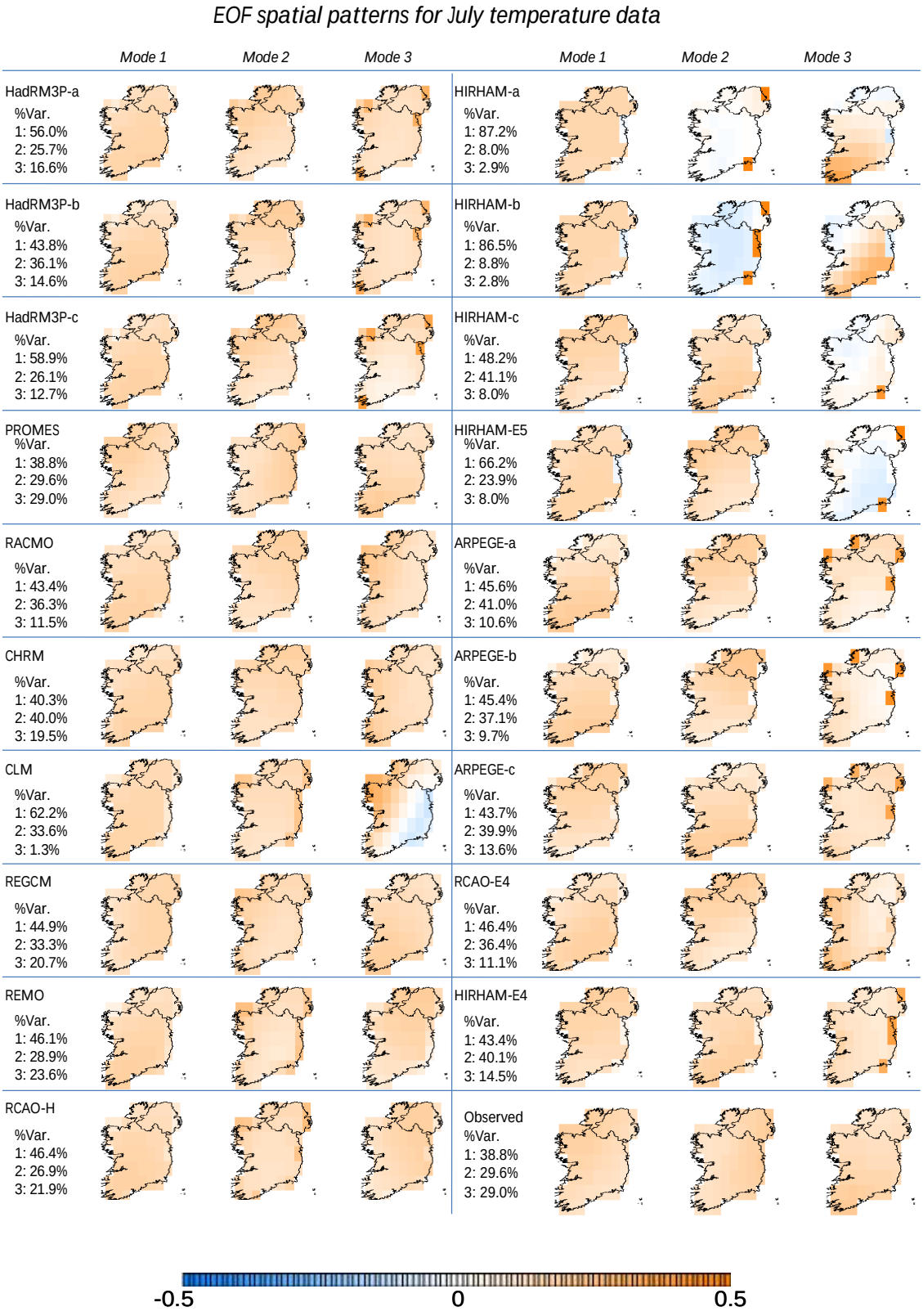


Figure 5.29: Modelled EOF patterns and associated percentage variance explained for July temperature data (1961-1990).

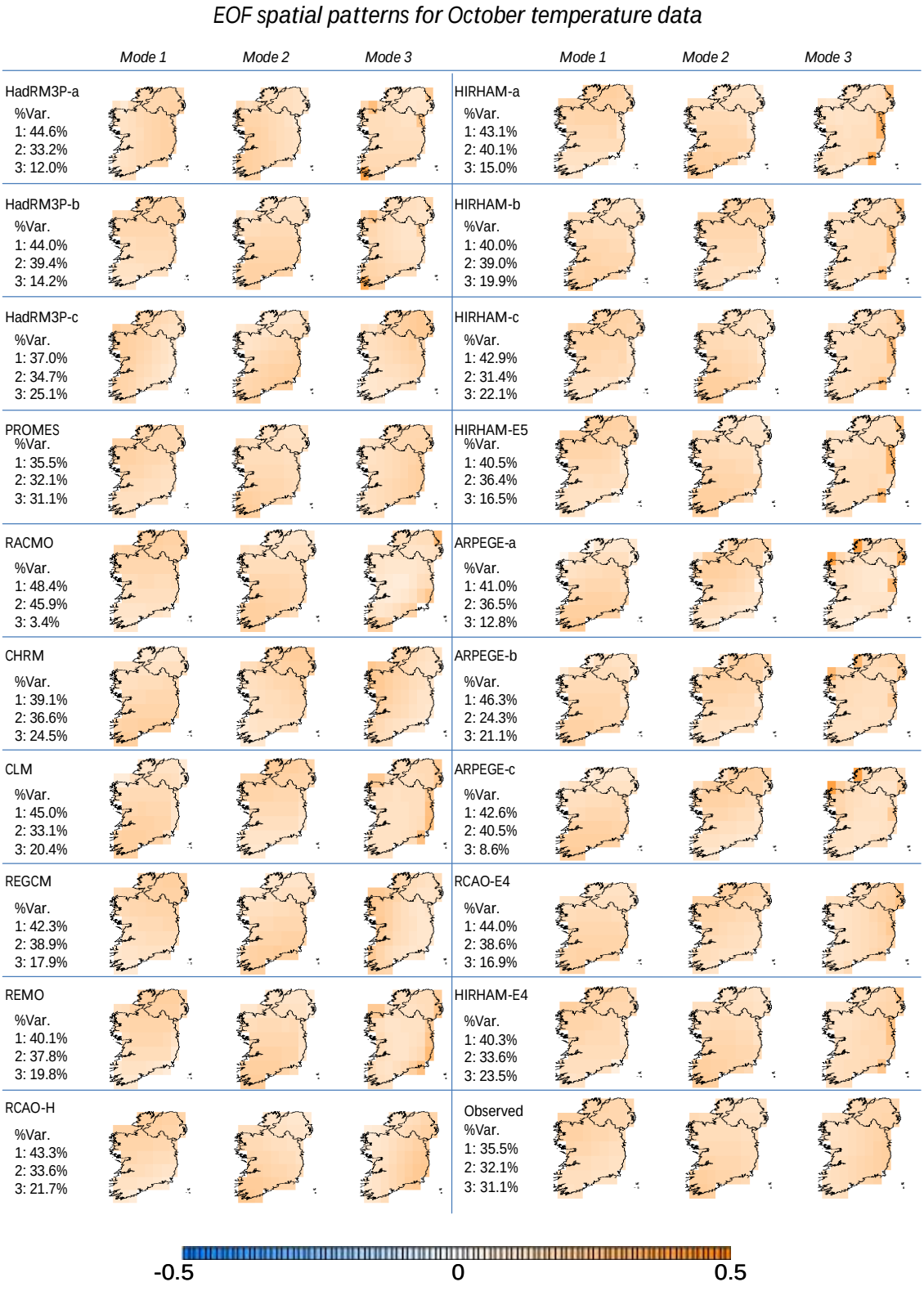


Figure 5.30: Modelled EOF patterns and associated percentage variance explained for October temperature data (1961-1990).

5.3.5 Results: Precipitation

In January (Figure 5.31), HadRM3P-a, HadRM3P-b, HadRM3P-c and PROMES all capture the first and second mode patterns and their associated percentage variability quite well. However, the third EOF pattern is not the same as the observed third EOF pattern, as the action centre in the models is situated to the east rather than the north-east.

The majority of models capture the first EOF pattern with skill, however there are exceptions. RACMO, CHRM, REGCM, REMO, HIRHAM-E5, ARPEGE-b and ARPEGE-c all have a first EOF pattern that is unlike the observed in January. Instead, the second EOF pattern best resembles the observed first EOF pattern.

The percentage variance associated with the north-west action centre pattern, whether it occurs as the first or second EOF pattern, varies from approximately 35% in ARPEGE-c, CHRM and REGCM to 48% in HadRM3P-b. In fact, the observed value is 44%. Most of the models underestimate the variance associated with this pattern, suggesting that westerly airflow, which would be most responsible for bringing precipitation to this area, is underestimated in the models. Some models do overestimate the associated variance of this pattern, namely the HadRM3P simulations and RCAO-E4, suggesting an overestimation of the underlying airflow type.

In April (Figure 5.32), observed patterns are similar to January, but notably, the percentage variance associated with the first pattern falls from 44.1% to 37.6%. Some models simulate associated percentage variance of less than 40%, such as HadRM3P-a, HIRHAM-E5 and RCAO. However, many models significantly overestimate the variance associated with the north-west centred pattern in April, with CHRM in particular displaying an associated variance value of over 50%. This suggests that this precipitation pattern occurs much more frequently than observed in certain models.

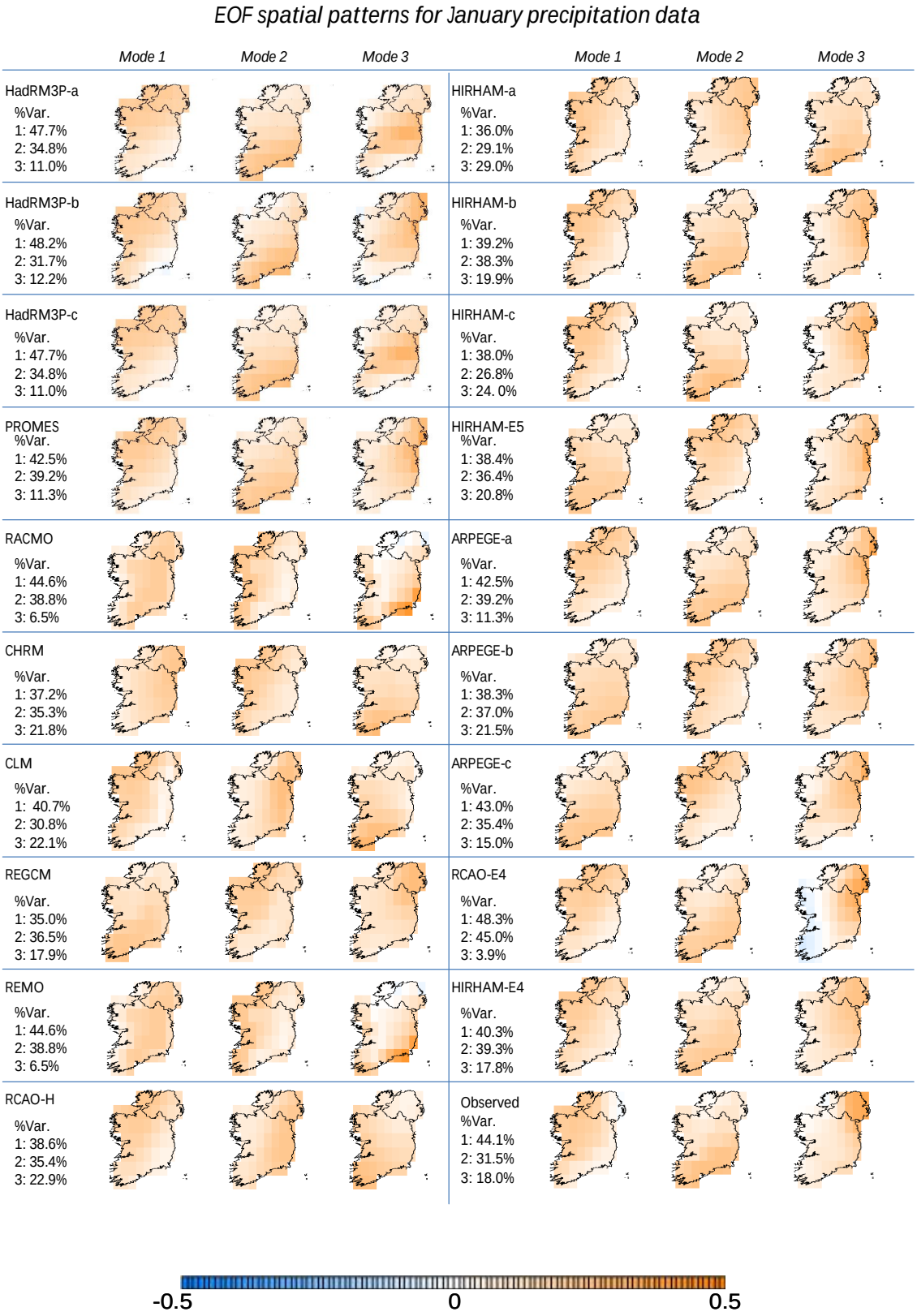


Figure 5.31: Modelled EOF patterns and associated percentage variance explained for January precipitation data (1961-1990).

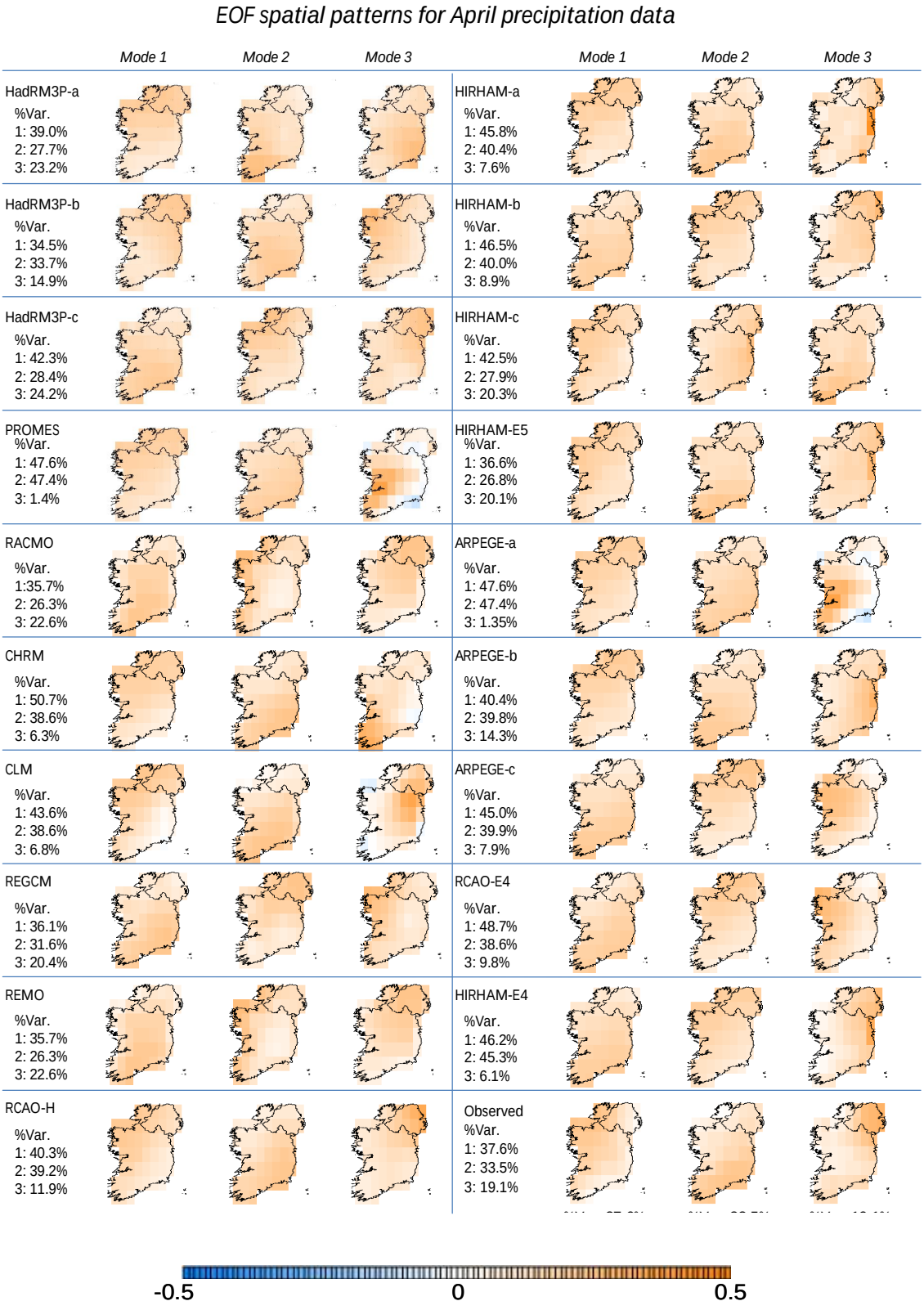


Figure 5.32: Modelled EOF patterns and associated percentage variance explained for April precipitation data (1961-1990).

For July (Figure 5.33), observed percentage variance explained did not fall to insignificant levels until the fourth EOF mode. This most likely reflects the greater levels of spatial variability associated with summer precipitation. If the models are to represent summer precipitation well, they should capture the same four EOF patterns and similar levels of associated variance. However, many of the models reach insignificant levels of variance after the third mode. HadRM3P-b, CHRM, HIRHAM-E5, ARPEGE-b and RCAO-E4 all have less that 5% variance associated with their fourth mode. Yet the observed fourth EOF pattern was a significant pattern with 17.2% of variance associated with it. These results suggest that these particular models simulate July precipitation as being more homogenous, dominated by fewer patterns than it actually is.

Several models simulate the spatial patterns skilfully, for example, CLM, REMO and REGCM. However, no models simulate the EOF modes in the correct order. Only HIRHAM-E5 has a first mode pattern similar to the observed first mode pattern, though it overestimates the variance associated with it. This suggests that the pattern that dominates the observed data is not as influential as it should be in the models, except in HIRHAM-E5 where its influence is overestimated.

Finally, precipitation EOF patterns for October modelled data are given in Figure 5.34. As autumn precipitation is more heterogeneous than summer precipitation, once again only three modes are required to capture the significant spatial patterns. Many models capture the spatial patterns quite skilfully and in the correct order. HadRM3P-c, PROMES, RACMO, CLM, REGCM, RCAO-E4 and the three ARPEGE simulations all capture the appropriate spatial patterns in the correct order, although the actual percentage variance associated with each pattern varies between models. Other models simulate the correct patterns but in the wrong order, but this is still a positive outcome as the key components of spatial variability are represented.

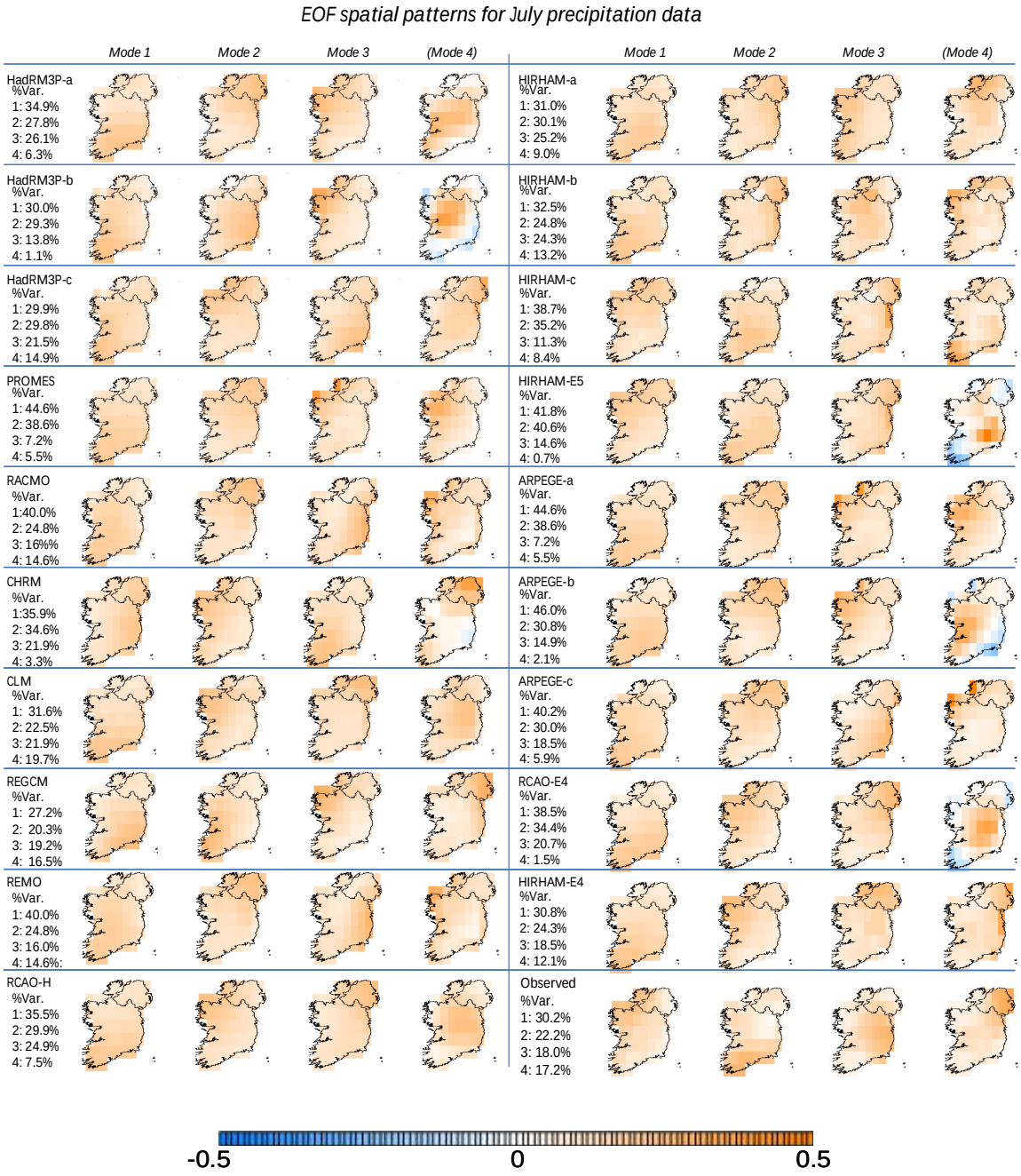


Figure 5.33: Modelled EOF patterns and associated percentage variance explained for July precipitation data (1961-1990).

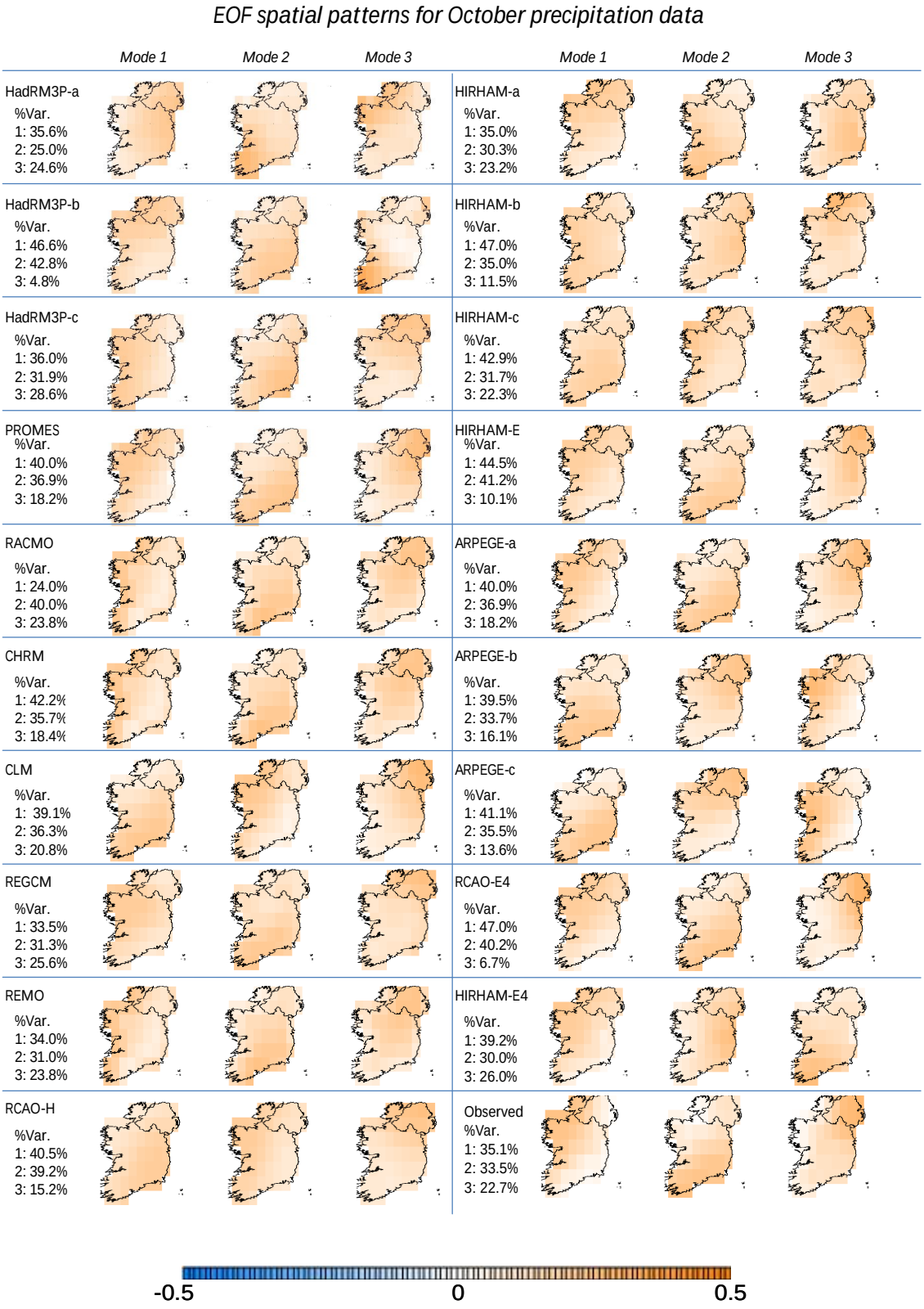


Figure 5.34: Modelled EOF patterns and associated percentage variance explained for October precipitation data (1961-1990).

5.3.6 Results: Modelled time amplitude series

The findings of Chapter 4 suggested that interannual variability in the RCMs is strongly governed by the driving GCM. The time amplitude series from the EOF analysis are used to further elucidate this point. Where a single GCM was used to drive more than one RCM, the time amplitude functions from those GCM groups are plotted together to determine whether there are similarities in the temporal evolution of the modelled time amplitude series.

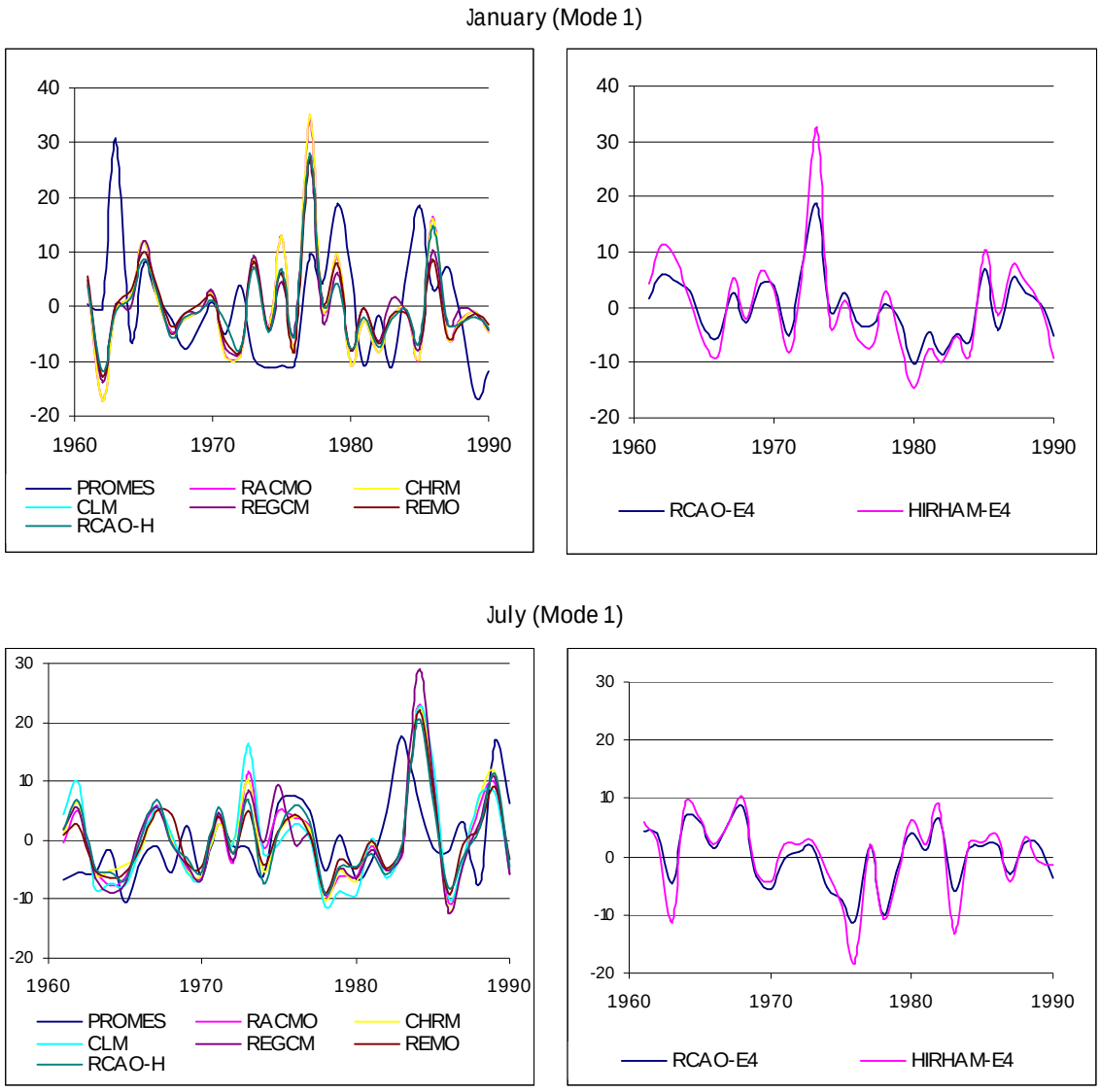


Figure 5.35: Time amplitude series of temperature EOF mode 1 for HadAM3H-driven models (left) and ECHAM4-OPYC-driven models (right) in January (top) and July (bottom).

It is clear from the plots in Figures 5.35, 5.36, 5.37 and 5.38 that there is a GCM influence on the time amplitude series of both the first and second EOF modes.

RCMs driven by the same GCM follow very similar temporal evolutions in their time amplitude series, suggesting that these components of the spatial pattern are heavily influenced by GCM boundary conditions. The slight variations between RCMs are most likely due to differences in how the individual RCMs parameterize different processes.

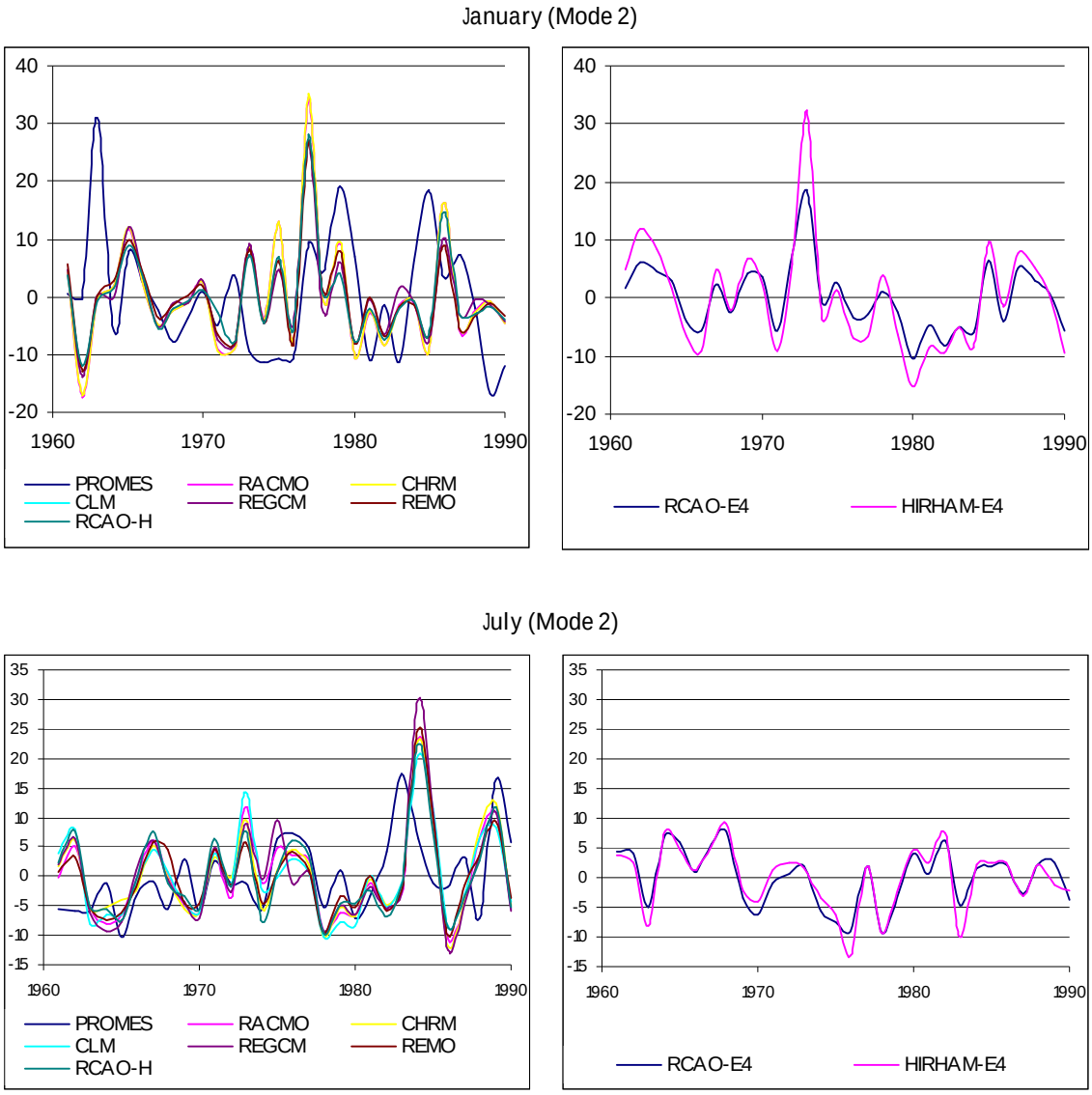


Figure 5.36: Time amplitude series of temperature EOF mode 2 for HadAM3H-driven models (left) and ECHAM4-OPYC-driven models (right) in January (top) and July (bottom).

While the boundary conditions from the GCM appear to be the most dominant factor influencing interannual variability, the internal model physics of the individual model also plays a part. One particular model, PROMES, behaves slightly different to the other HadAM3H driven models. While it does follow approximately

the same temporal evolution as the other models driven by this GCM, the other models cluster far more tightly together, with very little individual variation.

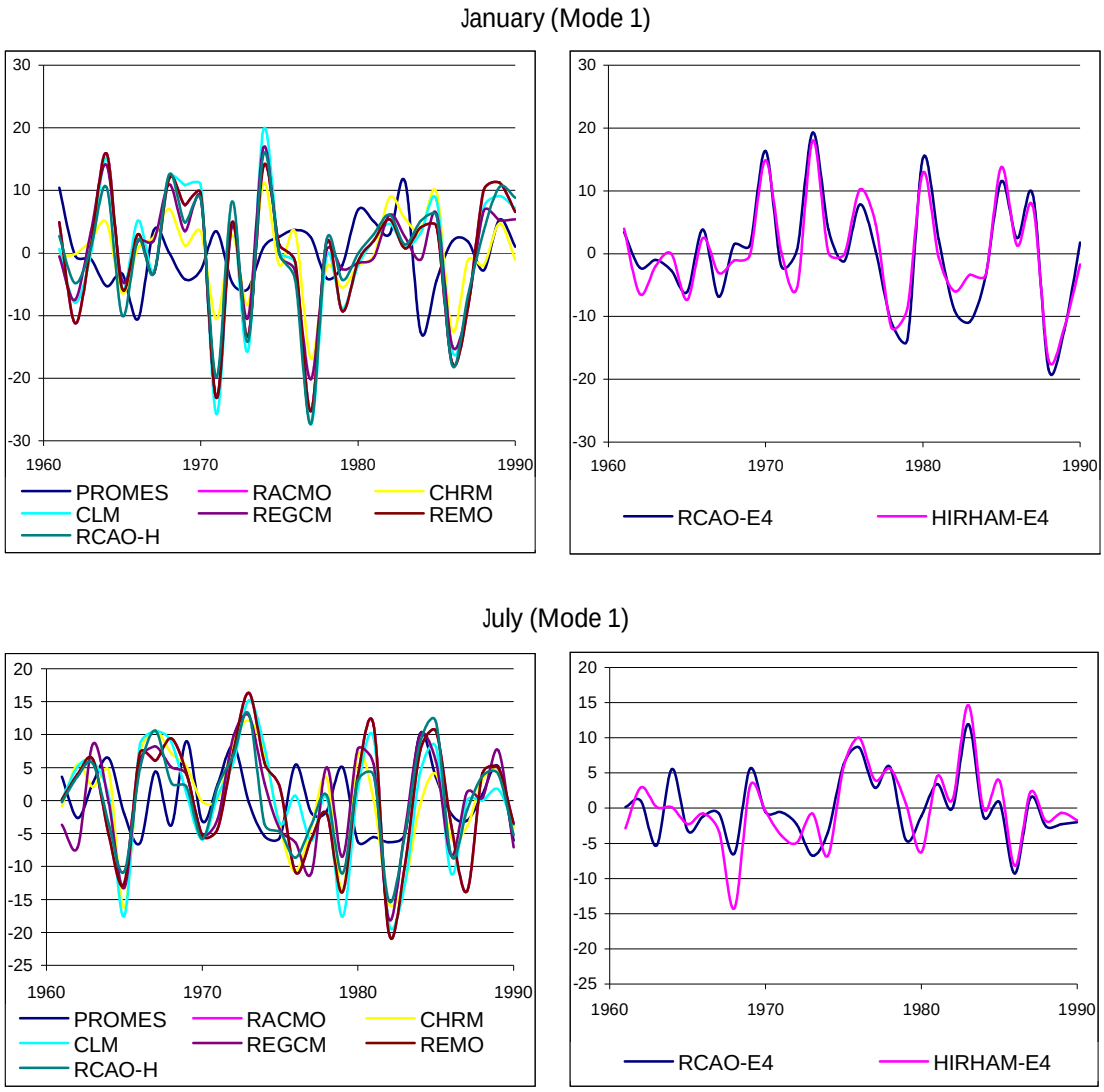


Figure 5.37: Time amplitude series of precipitation EOF mode 1 for HadAM3H-driven models (left) and ECHAM4-OPYC-driven models (right) in January (top) and July (bottom).

PROMES exhibits much more variation, suggesting that its internal model physics influence its interannual variability more so that in the other models. This outcome is interesting as it reaffirms the importance of both GCM and RCM choice in scenario development. If interannual variability were determined by GCM choice alone, the RCM choice, concerning this climate parameter at least, would be arbitrary. However, as both GCM and RCM have an influence, both choices must be carefully considered.

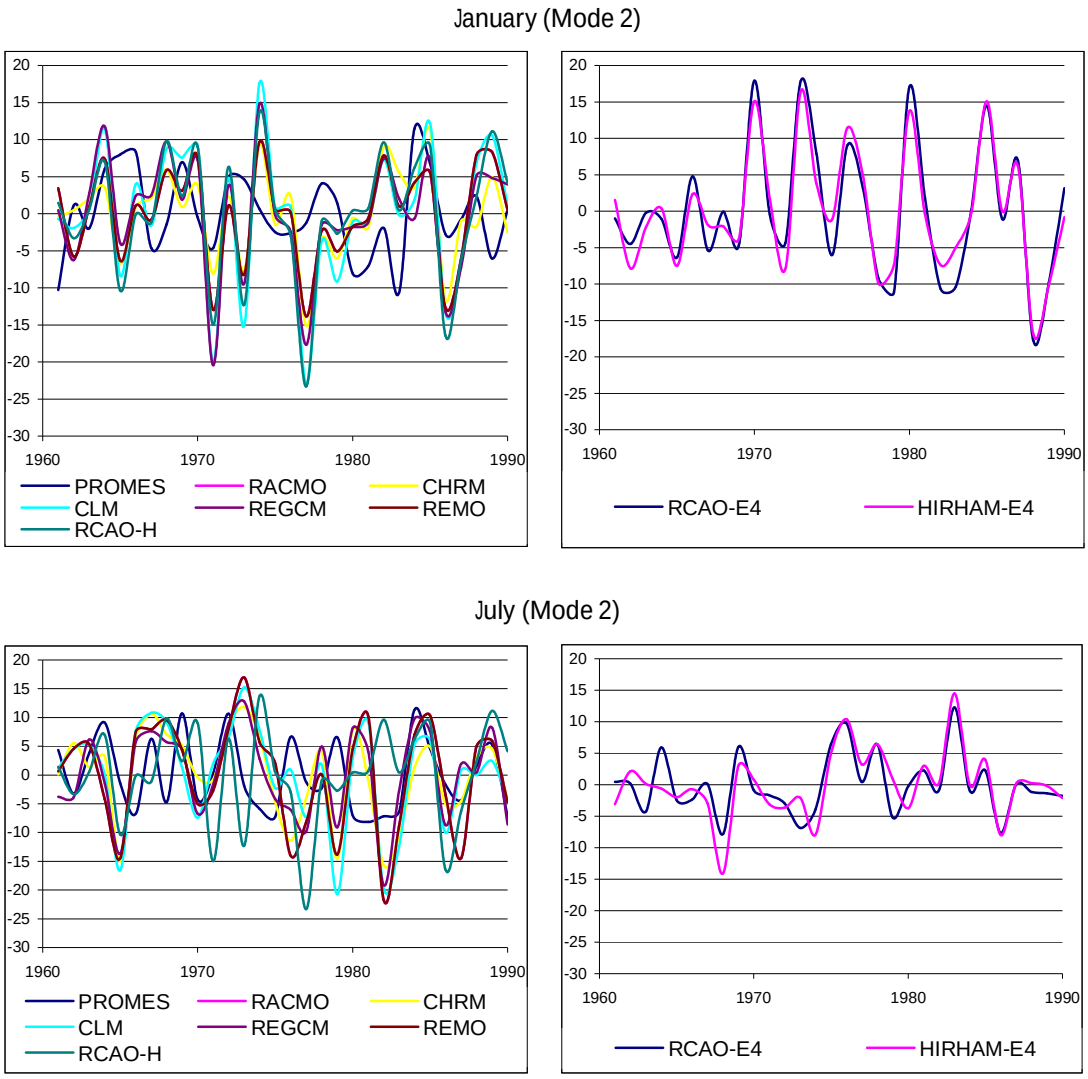


Figure 5.38: Time amplitude series of precipitation EOF mode 2 for HadAM3H-driven models (left) and ECHAM4-OPYC-driven models (right) in January (top) and July (bottom).

5.4 DISCUSSION AND CONCLUSIONS

Some key questions arise out of the results of this chapter, surrounding the nature of model skill. Firstly, the results of the seasonal pattern assessment indicate that on the surface, an ensemble can appear as skilful, simulating climatic patterns close to those observed. Yet in reality, the average ensemble can hide a variety of errors among the individual models. The analysis illustrated how models which over or underestimate temperature or precipitation can be combined to create an apparently "skilful" ensemble. However, such skill does not result from genuine model skill. It arises from error cancellation and if such errors were to change in any

way under different forcing conditions, the skill of the ensemble would not hold. Ensembles are proposed as a means of increasing reliability and confidence in model projections, yet if they are constructed in an opportunistic manner, they have the potential to negatively affect decisions made about climate planning and sustainability. Such ensembles may have the outward appearance of decreasing uncertainty through the inclusion of multiple projections, although in fact the uncertainty associated with the individual ensemble members can be quite large and unaccounted for.

Another issue which arises is the suitability of domain-wide means and variances as a metric for model skill. As evidenced in this analysis, a model may overestimate a climate parameter in one part of the country and underestimate it in another part of the country and none of this information is represented in a spatial average. For example, ARPEGE-a, ARPEGE-b and ARPEGE-c appear to be skilful models when considered based on their domain-wide skill scores alone (Table 4.4). Yet in the seasonal spatial pattern analysis and the EOF analysis, questions arose about how these models resolve temperature in particular gridcells to the north-west of the country. The prevailing westerlies which are controlled to a large extent by the North Atlantic Oscillation are a dominant influence on climate. Therefore, not only do these spatial errors reduce confidence in these simulations' projections of climate in these specific gridcells, they also raise questions about the representation of the underlying large-scale processes which govern climate in this area.

The results of this chapter also raise questions about what constitutes a skilful climate model. When no model emerges as skilful in every regard examined, the question is no longer which models are right, but which models are most useful. In the analysis of seasonal spatial patterns, two key categories of model emerged: those which simulate a low level of bias but fail to capture the spatial pattern with skill, and those which simulate a high level of bias but do capture the observed pattern well. Errors that have a potential physical explanation are more desirable than those that have no discernable explanation, as these can be communicated and accounted for in subsequent impacts analysis. For example, the absence of any representation of urban effects may cause the models to simulate cooler temperatures in urbanized regions than is observed (Giorgi *et al.*, 2004) and this may account for the cool biases around the Dublin area in the summer temperature simulations of some models.

Arguably, a systematic bias, which does not interfere with the spatial pattern, is a more desirable bias to have in a model as there is potential to account for such a bias in subsequent scenario development. If a model overestimates temperature, for example, by the same amount in every gridcell in the present day simulation, the error can potentially be corrected in future simulations by subtracting that amount from each grid cell. Of course, there is the possibility that the error will not remain constant over time and under different forcing conditions, that it may change in magnitude, making the correction applied insufficient, or become more random in space. However, if errors are random in space to start with, if they occur randomly in the present-day control run and as such interfere with spatial pattern resulting in output that does not resemble the observed pattern, there is even less potential to work with such errors as they are likely to be the result of several different deficiencies in the internal model physics rather than a single deficiency. As one attempts to correct for more and more deficiencies, there is inevitably growing uncertainty surrounding how these errors will behave under different forcing conditions and through time. As such, although a model with systematic bias is less than ideal, it has more potential to be useful and less potential for uncertainty than a model that exhibits random errors that differ in magnitude and sign in different parts of the country.

Finally, the EOF analysis suggests that models can generate the correct spatial pattern and even the correct EOF component patterns, but that the influence of one pattern versus another may not be captured. This may indicate that while the physical processes that control temperature and precipitation are represented well in the models, their strength or dominance in the region is not represented well. As such, the next step is to examine these physical processes in greater detail, investigating the large-scale physical processes and phenomena that control temperature and precipitation on the regional Irish scale.

CHAPTER 6
AN ANALYSIS OF THE IMPACT OF LARGE-SCALE DRIVERS ON MODELLED CLIMATE: NORTH ATLANTIC OSCILLATION

6.1. INTRODUCTION

The previous two chapters highlighted the possible role of large-scale forcing mechanisms on spatial and temporal patterns of temperature and precipitation for Ireland. The simulation of winter temperatures appeared to be a deficiency across the majority of models and as the North Atlantic Oscillation (NAO) is a key driver of winter climate in Ireland, further analysis is required to investigate how skilfully the NAO is represented in the models and determine whether representation of this large-scale driver is a contributing factor in the regional errors observed in the earlier analysis.

While the assessment techniques applied so far, skill scoring and EOF analysis, have identified a number of errors in the RCM simulations of Irish climate, they do not indicate with any certainty the sources of these errors. Examination of the data may suggest whether a bias is GCM or RCM related, whether it is random or systematic, but a fuller analysis is required to determine how both the deficiencies and the abilities identified in the models so far arise. In addition to explaining errors identified in the models, this analysis may also help to determine if the skill demonstrated by the models in the earlier assessments arises out of skilful simulation or the cancellation of errors in different processes within the models.

This question is of paramount importance as models whose skill is derived from error cancellation cannot be relied upon under different climate scenarios. Varying emissions concentrations may affect different processes to different degrees, so error cancellation cannot be depended upon to remain constant through time and

under different forcings conditions. Additionally, for testing adaptation strategies and aiding in the climate planning process, models with genuine skill are far more preferable to models whose skill is due to error cancellation. A model that captures the large-scale drivers of climate in a region with skill is likely to be a much more robust and reliable tool. Even if there are biases in such a model, biases can be corrected but deficiencies in the model's representation of key processes may only be resolved through model improvements.

This chapter focuses on the representation of the NAO at various temporal scales. At the European domain scale, mean sea level pressure patterns in each season are examined to identify potential errors in the representation of the Icelandic Low – Azores High pressure gradient. Then, for a sub-set of case study models that are illustrative of the key differences between the models, the 1961-1990 data is categorized into NAO positive (NAO+) and NAO negative (NAO-) years. Spatial patterns of mean sea level pressure, temperature and precipitation across the UK and Ireland are examined to determine whether the models capture the effects of NAO phase on the relevant climate parameters. Finally, the effect of large-scale errors on regional climate is assessed. A modified weather classification approach is used to examine the frequency of wind directions across Ireland in the six case study models and to determine associated precipitation amounts. Errors in the representation of the regional climate and their relationship, if any, to errors in the representation of the NAO are discussed.

6.2 THE NORTH ATLANTIC OSCILLATION

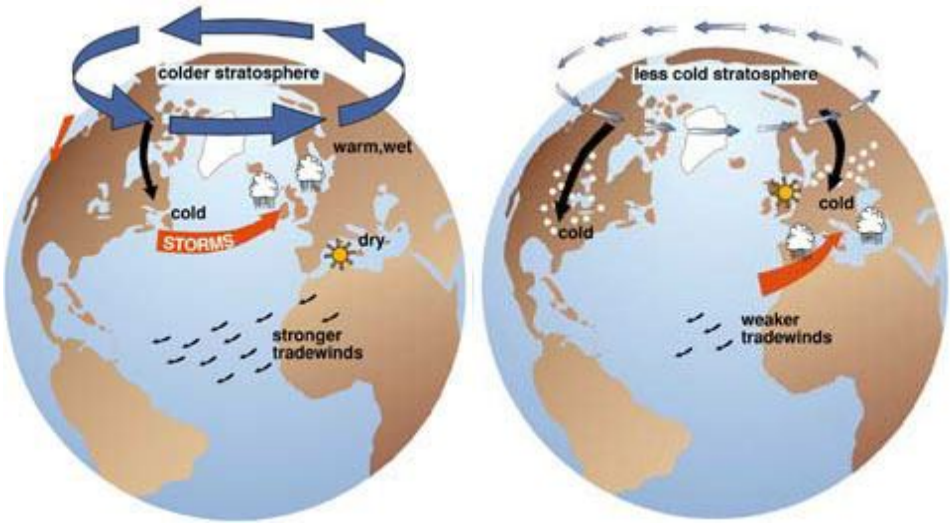
The North Atlantic Oscillation is a large-scale decadal mode of natural climate variability that influences climate in the whole North Atlantic region and in Europe particularly. It is most dominant in the northern hemisphere winter months. The NAO is largely an atmospheric mode of variability, unlike the El Nino Southern Oscillation (ENSO) in the Pacific which demonstrates a coupling between the ocean and atmosphere. The NAO is closely related to the Arctic Oscillation, however the NAO is viewed as the more relevant and robust pattern of Northern Hemisphere

variability, particularly as the NAO is represented in a more physically consistent way through principal components analysis than the AO (Ambaum *et al.*, 2001).

The NAO arises due to the east-west oscillatory motions of the Icelandic Low and the Azores high permanent pressure systems. The NAO can be defined using differences in meteorological station data and for this approach, a measure of the Icelandic Low is taken as sea level pressure at Reykjavik, Iceland, as this is the only station in this area with a sufficiently long record on which to base calculation. The Azores High may vary, as Lisbon, Ponta Delgada and Gibraltar are all weather stations near the centre of the high pressure system, but the choice of location makes little difference to the overall calculation (Osborn, 2001). The relative strengths and positions of the two systems vary from year to year, affecting climate in the North Atlantic region in a number of ways.

The phases of the NAO and their effects on North Atlantic climate are illustrated in Figure 6.1. In positive NAO years, the pressure is below average towards Iceland and above average towards the Azores. The resulting difference in pressure between the two centres is large, resulting in stronger westerly and southwesterly winds. These winds bring warm maritime air from the North Atlantic to Central and Western Europe, leading to mild and wet winters in this region and cool summers. Conversely, the Mediterranean experiences drier conditions in a positive NAO year.

In negative NAO years, the Azores high pressure is weaker than average, with above average pressure in the Iceland area. In this case, westerlies are suppressed and without the moderating influence of the mild maritime winds, temperatures in Northwest Europe become more extreme. The region instead experiences a greater frequency of north or northeasterly winds. Winters become colder and drier for Northwest Europe, including Ireland, with the possibility of snow and severe frosts. Additionally, the Atlantic storm tracks which so greatly influence European rainfall are diverted towards Spain and Portugal. As a result, the Mediterranean experiences wetter than average conditions and increased storm activity in winter.



Positive NAO phase **Negative NAO phase**

Figure 6.1: Effects of positive NAO phase (left) and negative NAO phase (right) on climate in the North Atlantic region. In positive NAO phases, Central and Northwestern Europe experience warmer, wetter conditions due to stronger westerly winds from the Atlantic. A positive NAO phase also results in drier conditions in the Mediterranean and can contribute to warmer conditions in North America. In negative NAO phases, conditions are reversed, with Northern Europe experiencing very cold winters and extremely warm summers, while the Mediterranean experiences increased rainfall and storminess (Source: USGCRP, 2000: http://www.usgcrp.gov/usgcrp/seminars/000320FO.html, accessed 03/08/2010).

In addition to its impact on European climate, the NAO is also believed to have an impact on weather in eastern North America. In positive NAO years, this region experience stronger southerly winds (Hurrell, 1995). This can suppress the flow of cold air from the Arctic to the north, which can contribute to warmer winter conditions in much of North America, especially if other modes of variability with a greater affect on the area, such as ENSO, are also in a warming phase. Additionally, there is a strong correlation between the positioning of the Azores High pressure system and the direction of storm paths for North Atlantic hurricanes. When the system is positioned further south, storms tracks are diverted towards the Gulf of Mexico, while a northern position allows storms to travel upward towards the North American Atlantic Coast (Scott *et al.*, 2003).

Predictability of the NAO has been a matter of much investigation, as knowledge of future NAO behaviour can provide an indication of future European winter climate. Saunders *et al.* (2003) find that summer extent of snow cover in North America and Northern Europe can be used as a predictor of upcoming winter

NAO state while Rodwell *et al.* (1999) find that North Atlantic sea surface temperatures can be used to reconstruct much of the multiannual to multidecadal variability of the winter NAO over the past half century.

Quasi-decadal variability of the NAO has been especially pronounced over the period of 1960-1990 (Figure 6.2). There has been an observed trend in winter over this period towards a positive NAO phase, contributing to warmer winter temperatures in Europe. Precipitation has also been greater than average over northern Europe, and this has also been linked to NAO behaviour (Hurrell, 1995). The climatic conditions that the NAO brings also have a range of ecological effects. For example, NAO-induced weather conditions impact the population dynamics of bird species in both North America (Nott *et al.*, 2002) and Europe (Saethar *et al.*, 2000) by affecting their food resources. The effects of the recent positive NAO phase on sea temperatures also have impacts on marine species. The NAO positive phase has brought colder temperatures to the Labrador Sea but warmer conditions to the North Sea. Different species have different optimum temperatures and so are affected to varying degrees by changes in sea temperature. For example, the snow crab population in the Labrador Sea has thrived as it has a lower optimum temperature, while survival rates for the cod larvae population in the same area are reduced as they are at their lower temperature threshold in the colder water (Pearson, 2009).

The effect of climate change on the NAO is a matter of debate. Paeth *et al.* (1999) finds that radiative forcing due to increased CO_2 concentration influences the variability of the NAO in climate models, regardless of model version, on time scales of 60 years and longer. However, Gillett *et al.* (2003) notes that while the majority of climate models simulate an increase in winter NAO index strength in response to increased anthropogenic forcing, there are some exceptions. Hartmann *et al.* (2000) suggests that increased greenhouse gas concentrations in their climate model simulation induce an enhancement of the meridional temperature gradient in the lower stratosphere. Such a mechanism could be responsible for a shift towards a positive NAO trend such as that which has been observed in recent decades.

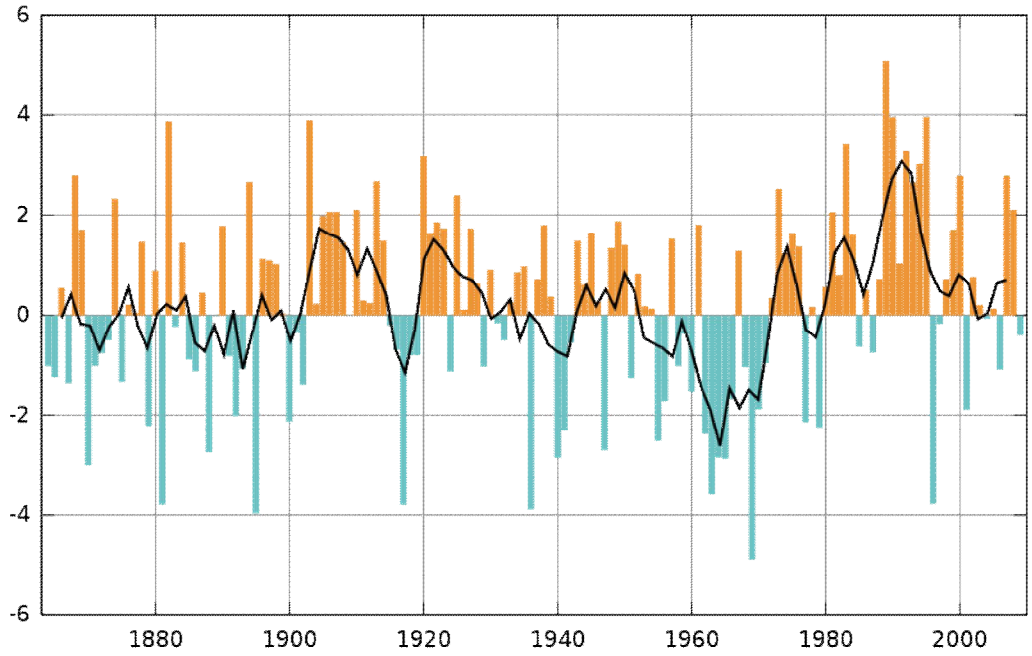


Figure 6.2:Winter (December through March) index of the North Atlantic Oscillation (NAO) based on the difference of normalized sea level pressure (SLP) between Lisbon, Portugal and Stykkisholmur/Reykjavik, Iceland since 1864, with a five year moving average (black) (Source: CGD's Climate Analysis Section.)

Due to the importance of the NAO as an influence on climate in the North Atlantic region, the range of impacts associated with its phases and the potential for climate change to alter its behaviour, it is highly desirable that the models used to inform and test climate adaptation strategies in Ireland are able to capture this mode of variability in a skilful manner.

The representation of the NAO in GCMs has been the focus of some study. Osborn *et al.* (1999) investigated the realism of the NAO in the Hadley Centre GCM HadCM2, a predecessor of the current HadCM3. The HadCM2 GCM was compared to observations and was found to be largely skilful in a 1400 year control integration, with the exception of the period from the 1960s to the 1990s. An analysis of 30 year trends showed that the five observed trends starting between 1962 and 1966 exceeded the highest modelled trend. However, Collins *et al.* (2001) examined the representation of NAO in HadCM3 and found that recent absolute values of the NAO did lie within the range of natural variability for the updated model, although the recent rate of change was inconsistent with model variability. Stephenson *et al.* (2006) assessed the NAO in control and transient GCM simulations for 18 GCMs,

including HadCM3, and found that HadCM3 was a notable exception in terms of its representation of NAO. While the majority of models overestimated the observed mean wintertime NAO index, HadCM3 underestimated it. These findings suggest that while there may be some issues with representation of the NAO in the Hadley AOGCMs, overestimation of the absolute values of NAO is not one of those issues.

The RCMs driven by HadCM3 are driven using a double nested technique in which the AOGCM is used to drive an AGCM, which in turn drives the RCM. Jacob *et al.* (2007) examined performance at the European scale of the PRUDENCE RCMs driven by the HadAM3H atmosphere-only GCM. This driving model was also found to display a stronger pressure gradient than observed across much of Europe, along with excessively high winter temperatures and precipitation rates. Therefore there is a strong possibility that mean sea level pressure bias in the driving AGCM propagates through to these RCMs, potentially leading to similar effects on temperature and precipitation.

6.3. ANALYSIS OF WINTER MEAN SEA LEVEL PRESSURE ACROSS EUROPE

The RCMs in this study have been found to display consistent errors with the representation of winter temperatures, modelling milder winters than observed. To further investigate the cause of this bias, seasonal mean sea level pressure (MSLP) maps for each RCM are compared to the ERA-40 reanalysis dataset, as the dataset used previously does not contain MSLP data. Bias between the RCMs and the observations is plotted to determine whether pressure systems are accurately represented.

The ERA-40 dataset is produced by the ECMWF and describes global atmosphere and surface conditions from 1957 to 2001. To produce the reanalysis dataset, relevant meteorological observations from a range of different sources, such as meteorological stations, ship and buoy measurements and satellite observations, are assimilated using a variation of the ECMWF/Météo-France Integrated Forecasting System (IFS) which outputs data at 2.5^{o} latitude by 2.5^{o} longitude

resolution. As such, reanalysis data is observation-based, through a model is used to reanalyze the observed data and outputs data on a gridded format. The ERA-40 data is interpolated from its native grid to the finer grid used by the RCMs, but as this analysis is focused on a larger domain and on large-scale patterns, error at the regional scale is less of a concern than in previous chapters.

6.3.1 Analysis

Overall, biases in spring are generally much smaller than in winter, with the exception of the ARPEGE sub-ensemble (Figure 6.3a and b). In spring, all three iterations of this model markedly underestimate MSLP across the northern half of the continent, including Ireland. This lower-than-observed pressure may account for the overestimation of rainfall in this model in spring. In summer, biases are again small in most models and mostly appear related to orography (Figure 6.4a and b). Consider the positive MSLP bias present over the Alps in all models except the three HadRM3P simulations. MSLP is pressure at the given elevation reduced to sea level. Therefore errors in MSLP may be linked to errors in the representation of pressure and elevation. As pressure decreases with altitude, pressure over the Alps is quite low. However, as the Alps are fold mountains they are more difficult for a climate model to resolve compared to other types of mountain, such as plateau mountains. The Alps contain numerous high peaks whose resolution poses a challenge as they occur on such a small scale relative to the grids used by climate models. Inadequate resolution of these peaks may lead to an underestimation of elevation in certain gridcells, which would in turn result in higher pressure values than expected and higher MSLP values than expected, when compared with observations.

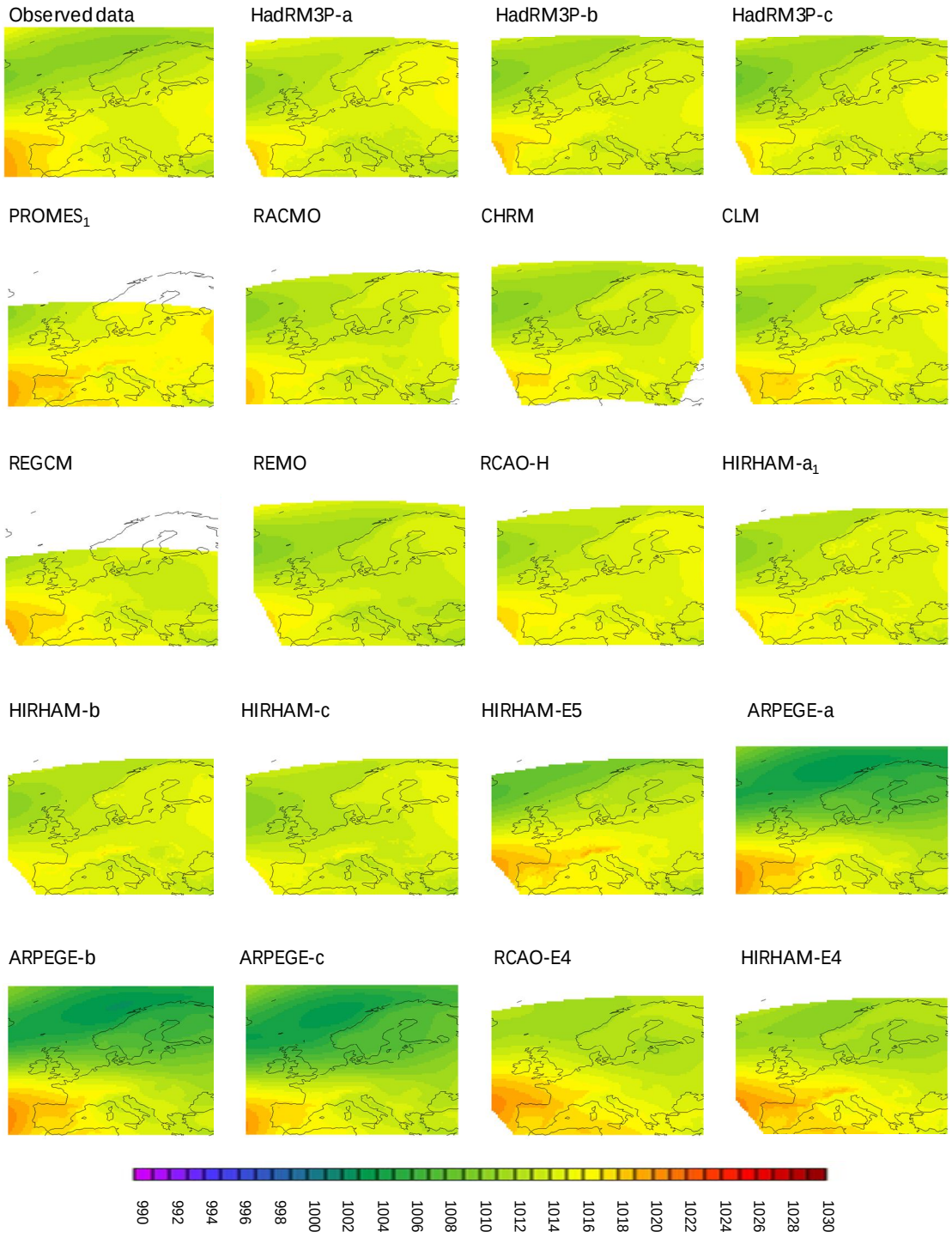


Figure 6.3a: Modelled mean sea level pressure for spring (MAM), measured in hPa for 1961-1990.

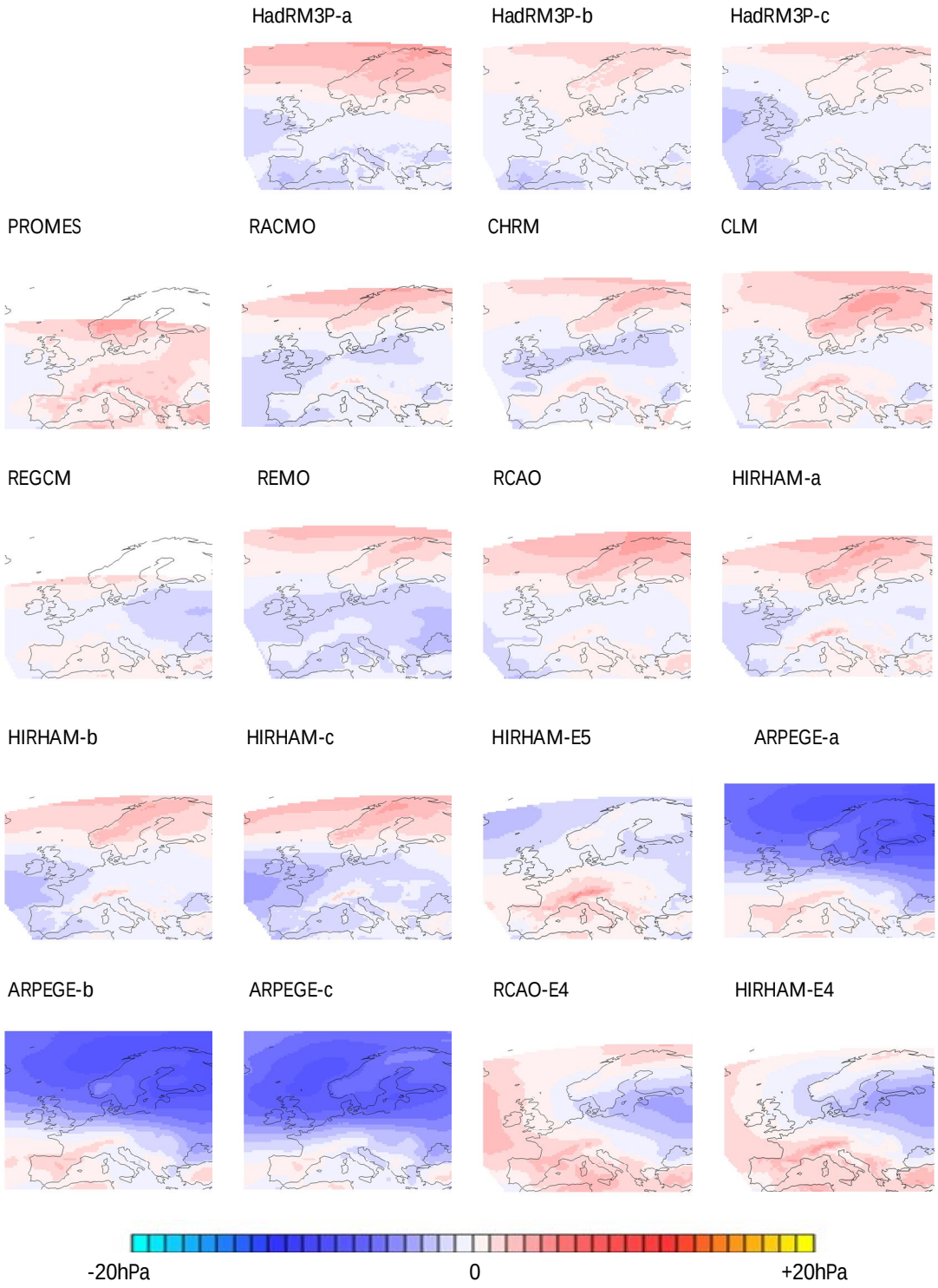


Figure 6.3b: Bias of modelled mean sea level pressure relative to observed for spring (MAM), measured in hPa for 1961-1990.

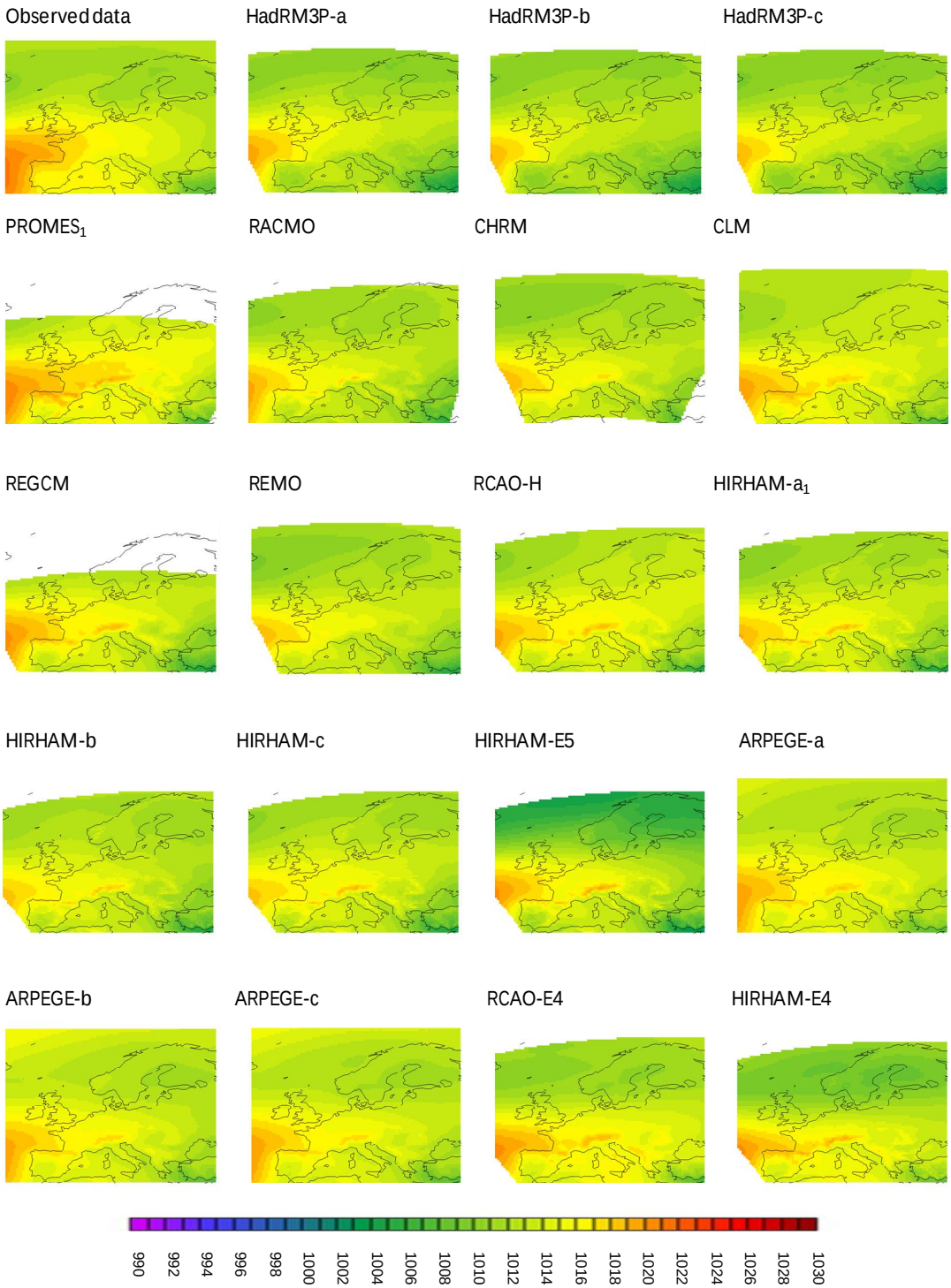


Figure 6.4a: Modelled mean sea level pressure for summer (JJA), measured in hPa for 1961-1990.

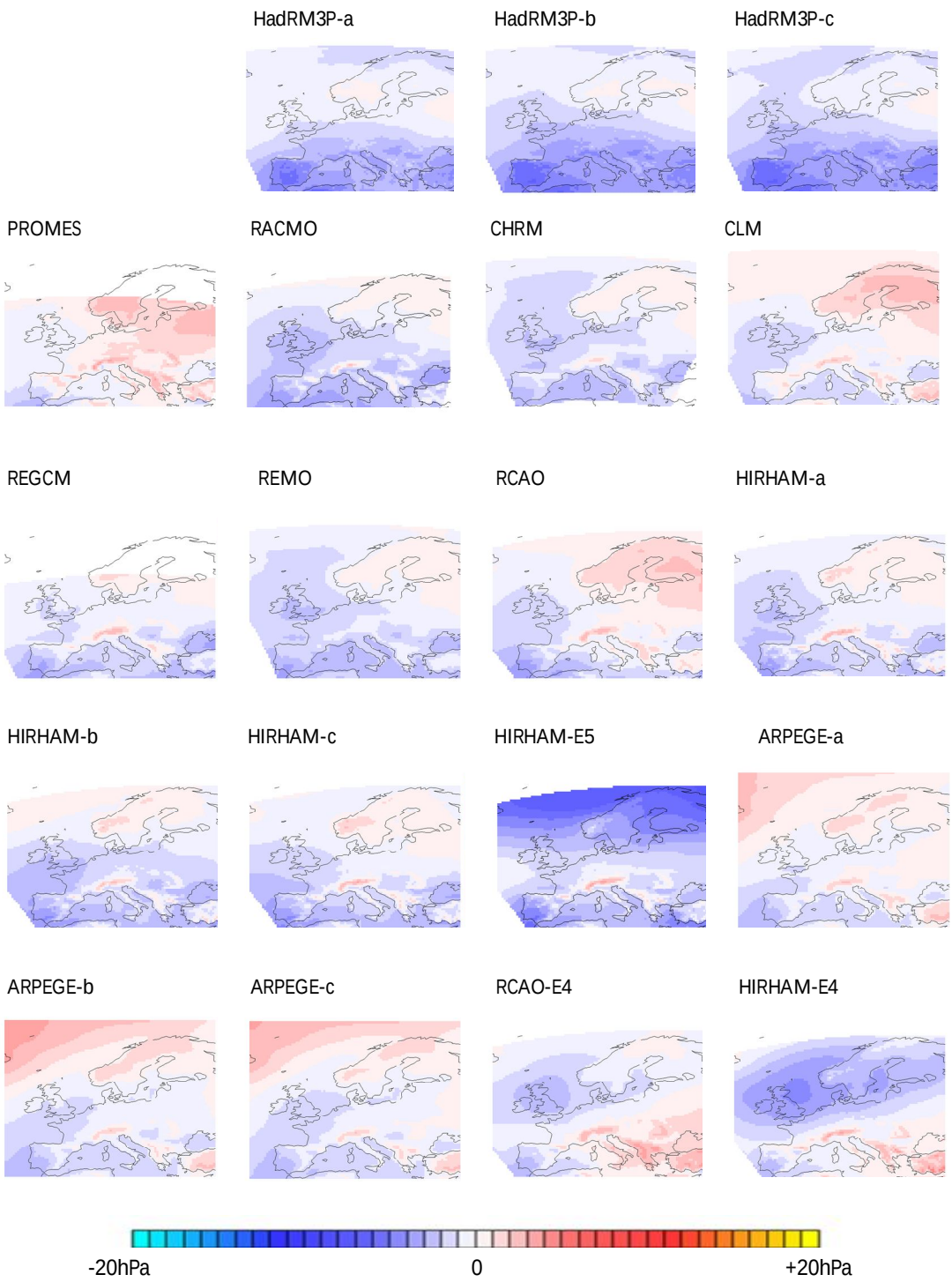


Figure 6.4b: Bias of modelled mean sea level presuure relative to observed for summer (JJA), measured in hPa.), measured in hPa for 1961-1990.

In autumn, observed MSLP shows the north-south pressure gradient developing (Figure 6.5a). However, modelled MSLP in many of the RCMs is overestimated in low pressure areas and underestimated in high pressure areas,

leading to a smaller difference between pressure centres than is observed. Some orographic bias is still present in mountainous regions, but becomes less noticeable in the context of the larger-scale bias across the domain as a whole (Figure 6.5b).

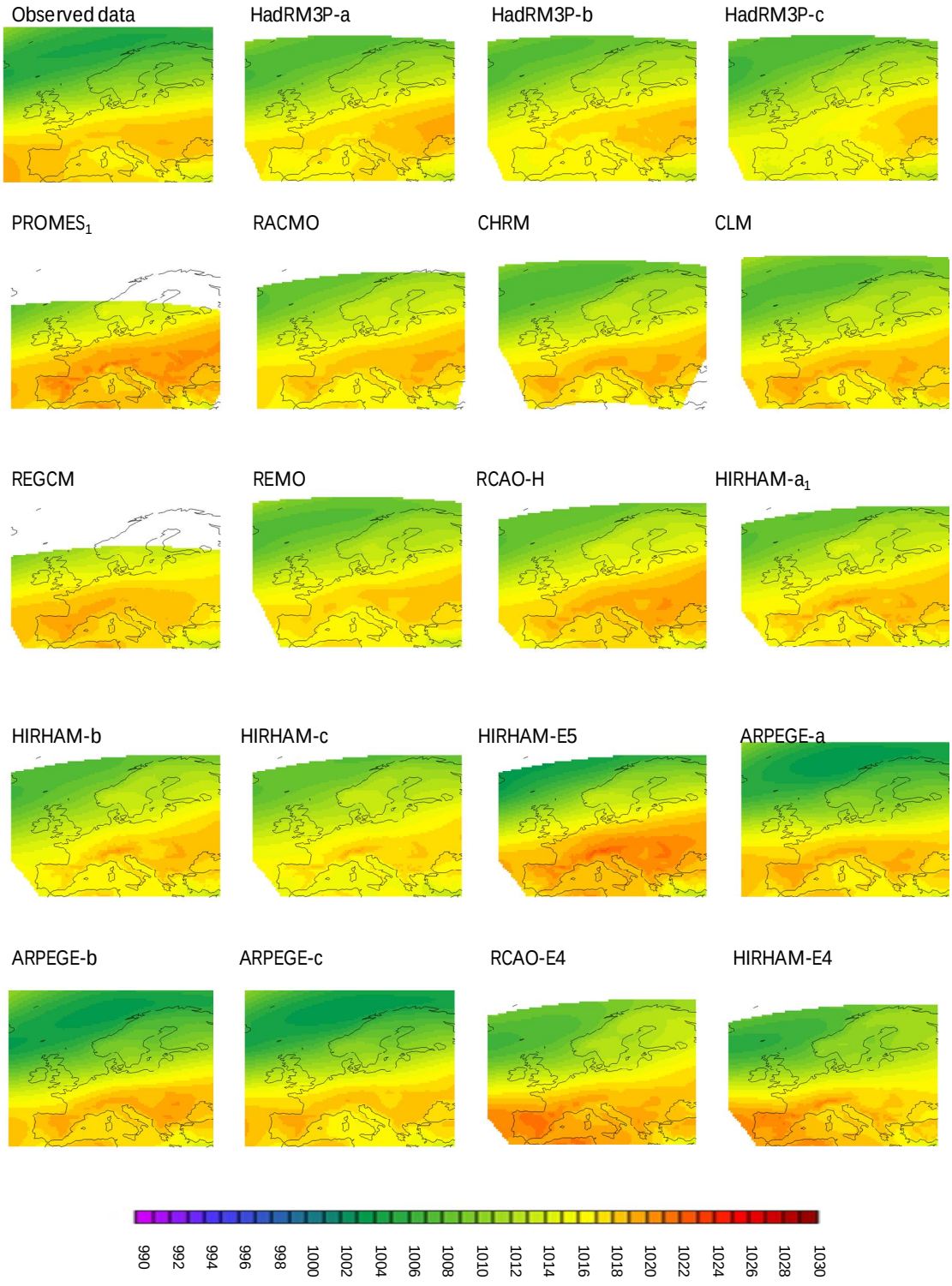


Figure 6.5a: Modelled mean sea level pressure for autumn (SON), measured in hPa for 1961-1990.

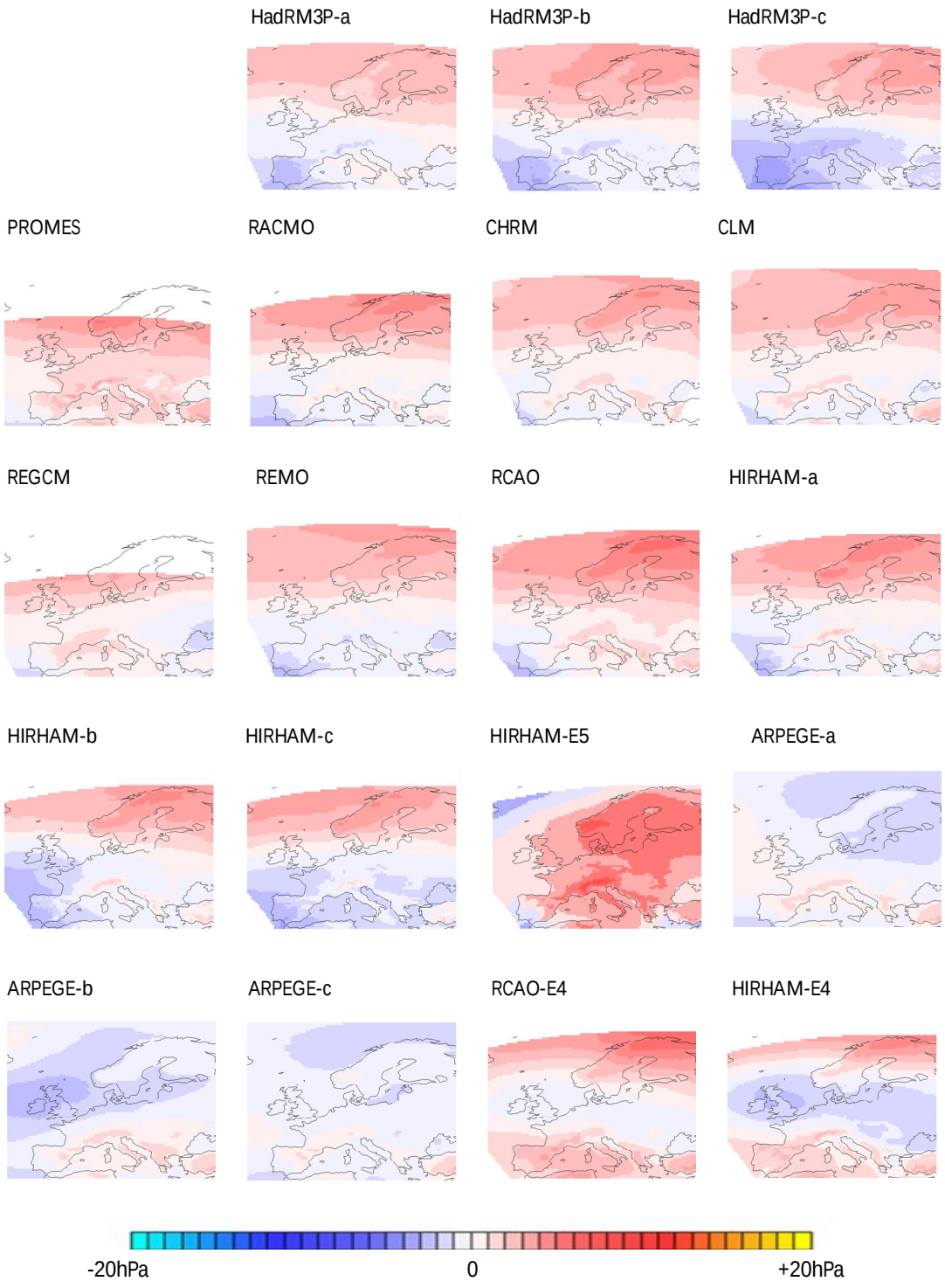


Figure 6.5b: Bias of modelled mean sea level pressure relative to observed for autumn (SON), measured in hPa for 1961-1990.

In winter, observed mean sea level pressure forms a low to high gradient from north to south (Figure 6.6a). However, many of the models simulate lower mean sea level pressure than observed in the northern half of the domain, including Ireland, the United Kingdom and Scandinavia, and higher mean sea level pressure

than observed across central Europe and the Mediterranean (Figure 6.6b). All models driven by HadAM3H and ECHAM5 follow this pattern. The result is a greater difference been the high and low pressure centres.

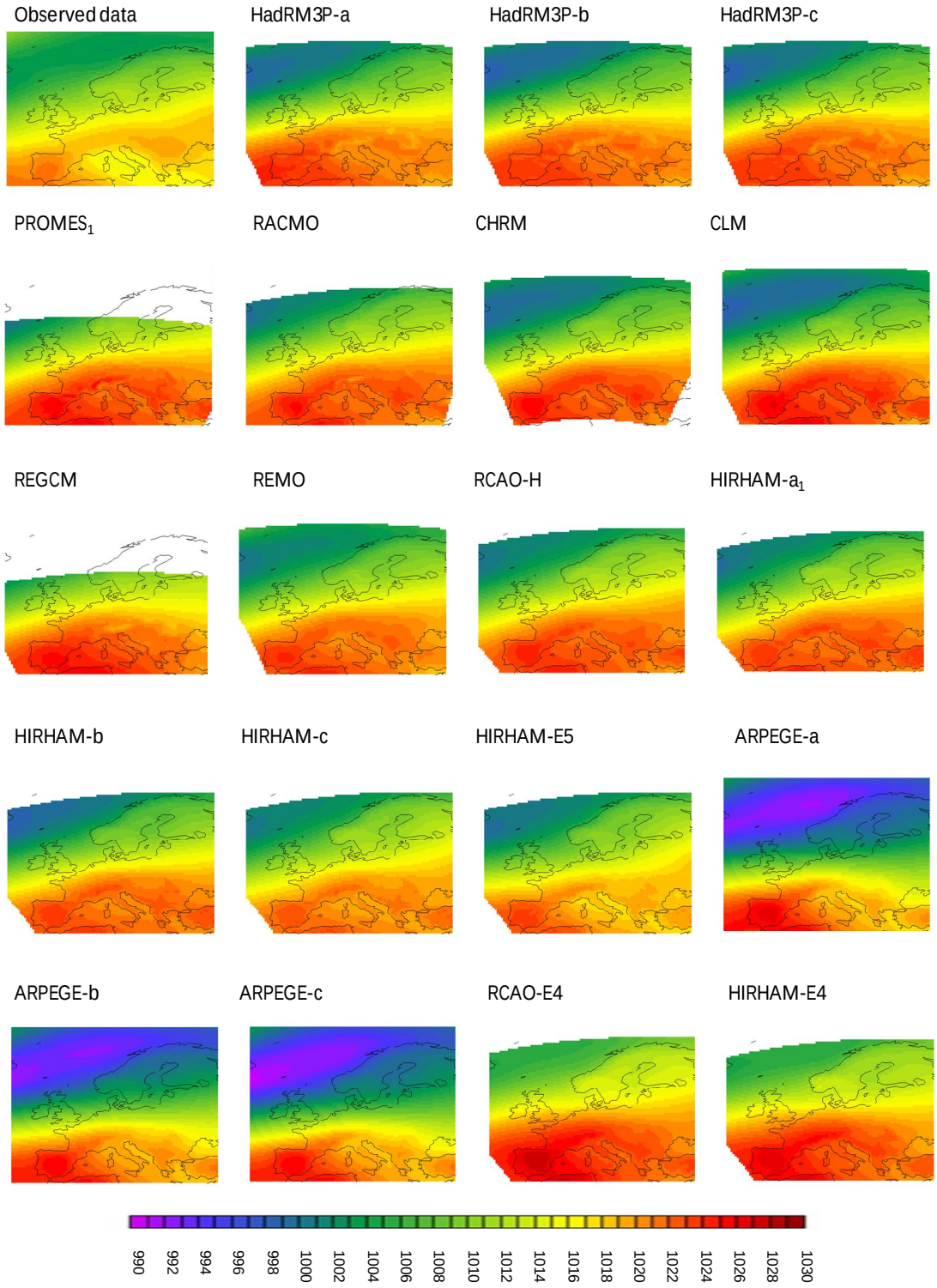


Figure 6.6a: Modelled mean sea level pressure for winter (DJF), measured in hPa for 1961-1990.

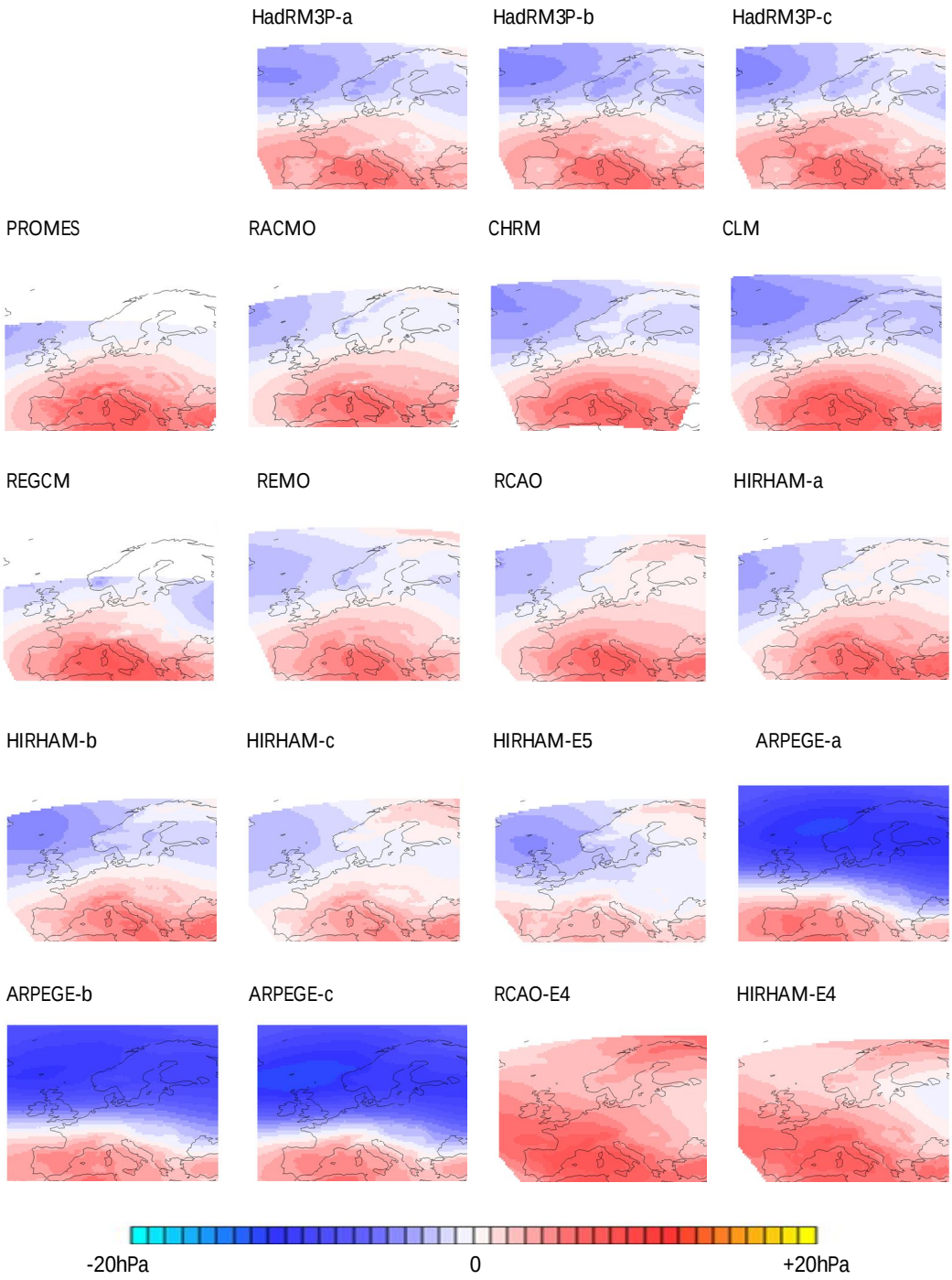


Figure 6.6b: Bias of modelled mean sea level pressure relative to observed for winter (DJF), measured in hPa for 1961-1990.

The pattern of mean sea level pressure bias and the effects on simulated Irish winter temperature and precipitation appear to correspond with positive North Atlantic Oscillation (NAO) behaviour. In positive NAO years, the pressure gradient across Europe is enhanced, resulting in increased westerlies and leading to warmer, wetter winters in Western Europe. In most of the models, the winter pressure

gradient is steeper than observed and there are positive precipitation and temperature biases, consistent with positive NAO behaviour. The HadRM3P sub-ensemble and the CHRM model are the only exceptions as both simulate drier conditions than observed in winter.

Overall, these findings suggest that in winter, the RCMs are greatly influenced by errors in the large-scale pressure systems that constitute the NAO. In spring and summer, MSLP errors relating to orography occur in certain areas, as the resolving limitations of RCMs can lead to elevations and correspondingly atmospheric pressure in mountainous areas being misrepresented. As the NAO is a large scale mode of climate variability, it is logical to examine the driving GCMs as a source of error in mean sea level pressure. The bias patterns identified in the RCMs tend to be quite similar for RCMs driven by the same GCM and as RCM output in winter is governed by large-scale processes is it quite possible that the mean sea level pressure biases identified in the RCMs are due to errors in the driving GCMs.

6.4 ANALYSIS OF UK AND IRISH CLIMATE PATTERNS IN NAO POSITIVE AND NEGATIVE YEARS

6.4.1 Data and methods

As the error patterns that emerge from the analysis of seasonal mean sea level pressure appear to be consistent across GCM driver groups, a sub-set of models representative of the overall ensemble is chosen for further investigation. The simulations chosen are HadRM3P-a, RCAO-H, HIRHAM-E5, ARPEGE-a and RCAO-E4, driven by HadAM3P, HadAM3H, ECHAM5, observed SSTs and ECHAM4-OPYC respectively. The case study models include one RCM driven by each of the GCMs. The focus of this section is assessing models' abilities to capture climate patterns over the UK and Ireland in NAO+ and NAO- years. As the NAO is most dominant in winter, the NAO indices and all spatial patterns are based on winter data.

The observed NAO index cannot be used to identify NAO+/- years in the models, as differences in the initial and boundary conditions of the models lead the individual simulations to evolve differently to what is observed. However, if the models are to be reliable, one would expect the modelled frequency distribution of NAO+/- years to be similar to observations. A model NAO index is calculated for each model based on the simulated pressure difference between the Icelandic Low and Azores High, to identify model NAO+/- years. As the domain of the RCMs does not include Reykjavik, the closest available point (14W, 64N) was used to represent this station. Lisbon (9W, 38N) is chosen as the southern point. Therefore the pressure centres are defined as:

Equation 6.1: Definition of pressure centres

$$P_R = P_{Reykjavik} = P_{[14W,65N]}$$

$$P_L = P_{Lisbon} = P_{[9W,38N]}$$

where P = mean sea level pressure.

The model NAO index was than calculated as follows:

Equation 6.2: To calculate NAO index

At each pressure centre:

$$P_{DJF} = \frac{\left(P_{Dec} + P_{Jan} + P_{Feb}\right)}{3}$$

$$P_{Anomoly} = P_{DJF} - \frac{\sum_{t=0}^{T} P_{DJF_t}}{T}$$

$$P_{St.Dev.} = \sqrt{\operatorname{var} P_{DJF_t}}$$

$$P_{Normalized} = \frac{P_{Anomoly}}{P_{St.Dev.}}$$

$$NAO = P_{R\ Normalized} - P_{L\ Normalized}$$

To eliminate noise in the data, years with an NAO index value of between -1 and +1 were omitted from the analysis. NAO+ years were defined as years with an NAO index greater than +1, while NAO- years were defined as years with an NAO index of less than -1.

As this method uses different pressure centre points to those commonly used to calculate the NAO index, validity of the method was tested by applying it to ERA40 mean sea level pressure data and comparing results with the NAO index calculated by the Climate Analysis Section at the National Centre for Atmospheric Research (NCAR). As illustrated in Figure 6.7, the difference in location makes little difference to the calculation and the method used to calculate NAO index in this thesis yields a very similar result when compared to the true NAO index.

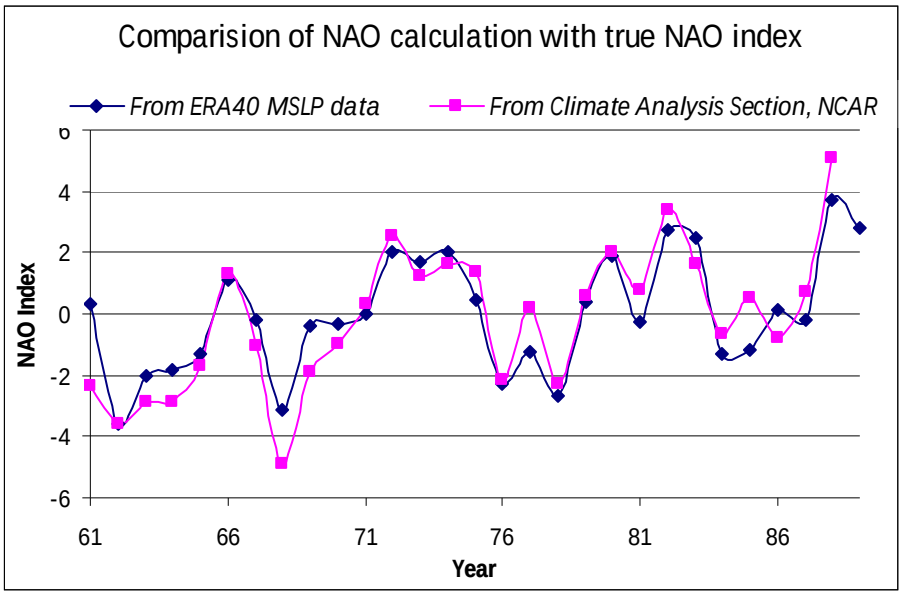


Figure. 6.7: Comparison of calculated NAO index based on ERA40 data (blue) with NCAR NAO index (pink). Indices are comparable in terms of both magnitude and temporal pattern.

For mean sea level pressure, temperature and precipitation, the mean spatial pattern associated with NAO+ years was obtained by averaging the NAO+ datasets at each point. The same calculation was carried out for NAO- years and maps of these mean spatial patterns were produced.

6.4.2 Results: Modelled frequency of NAO+/- years

Figure 6.8 illustrates the differences in frequency of NAO+/- years in a sub-sample of different RCMs. The models capture the number of positive NAO years quite skilfully, though all models underestimate the number of negative NAO years. However, overall these differences are quite small and results suggest that these RCMs are able to capture the distribution of positive and negative modes of the NAO over the period of analysis. This is a welcome outcome as the years 1961-1990 were marked by a shift towards predominantly positive NAO activity and the models' ability to capture the observed frequencies suggests that they have captured this large-scale mode of variability. However, this does not guarantee that the effects of NAO activity on climate will also be simulated well. Further analysis is required to determine this.

	ERA40	HadRM3P-a	RCAO-H	ARPEGE-a	HIRHAM-E5	RCAO-E4
NAO+	9	9	10	9	11	9
NAO-	10	7	9	9	8	9

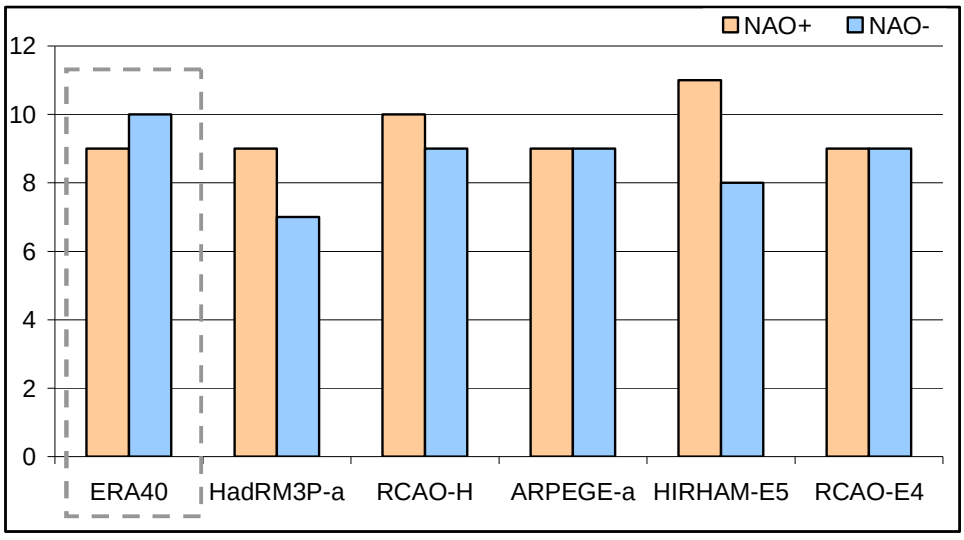


Figure 6.8: Frequency of NAO+/- years in RCMs, compared with observed frequencies over 1961-1990. The observed dataset, ERA40, is highlighted in grey. Orange bars denote NAO+ years and blue bars denote NAO- years.

6.4.3 Results: Observed NAO+/- patterns

For the observed patterns, ERA40 temperature and precipitation data was used, as this dataset, though coarser than the dataset used previously, includes sea gridcells. As such, it was necessary to interpolate the ERA40 data from 2.5^{o} to 0.5^{o}

resolution, making this data a less precise representation of Irish climate on a regional scale.

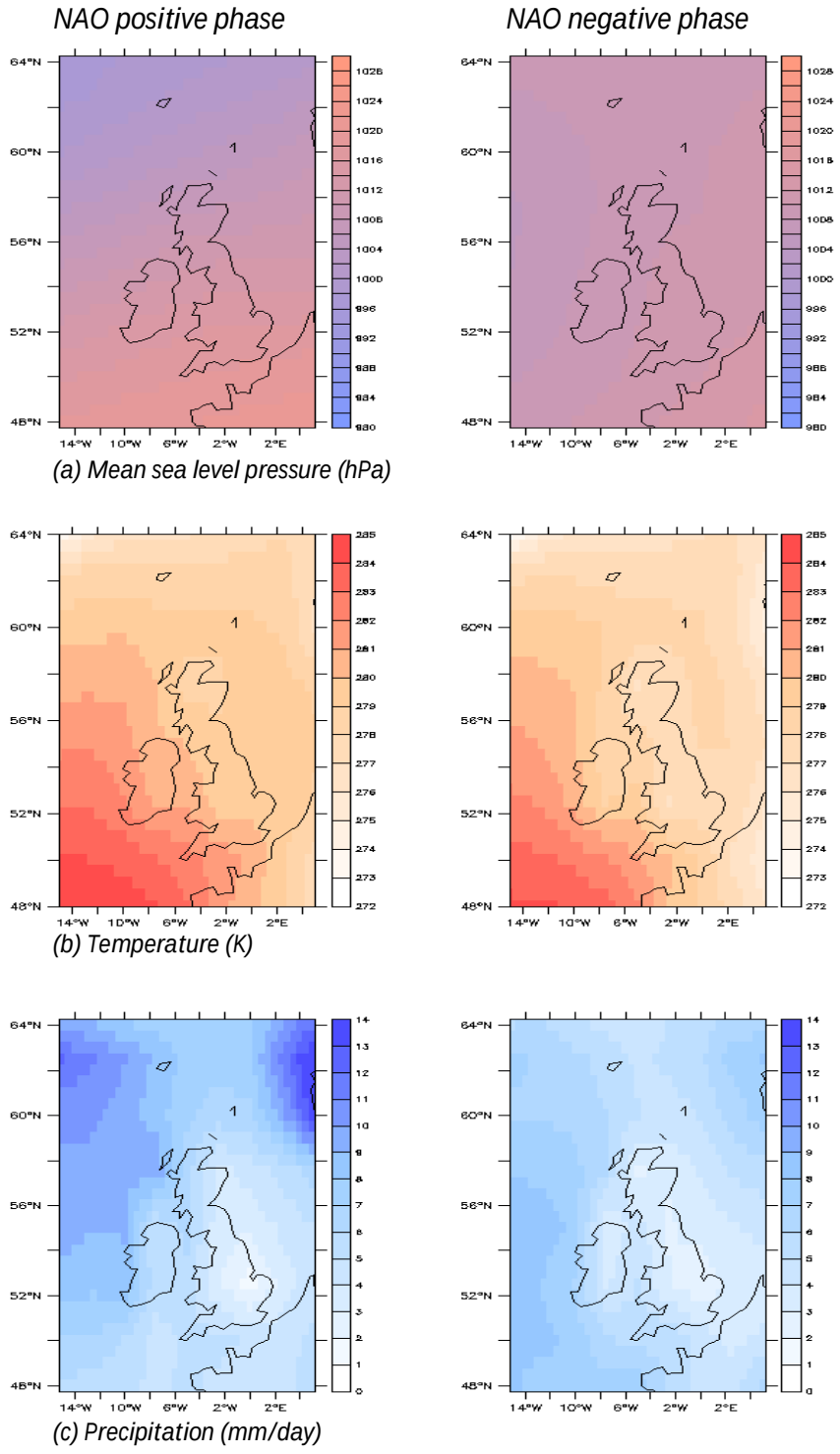


Figure 6.9: Observed average spatial patterns of mean sea level pressure (a), temperature (b) and precipitation (c) over the UK and Ireland in positive NAO years (left) and negative years (right) over 1961-1990.

Spatial patterns based on this data are provided as an indicator of expected NAO+/- behaviour on a large scale, rather than for comparison with modelled output on a fine scale. Observed winter spatial patterns associated with NAO phases are given in Figure 6.9. In NAO+ years, there is a distinct pressure gradient across the area, while in NAO- years MSLP across the area is more uniform (Figure 6.9a). The effects of NAO activity on regional climate are evident. In positive NAO years, temperatures are warmer (Figure 6.9b) and there is more precipitation (Figure 6.9c). The increased precipitation is especially noticeable in areas which are more exposed to the Atlantic, such as the west coast of Ireland.

The models have demonstrated an ability to capture the frequency of NAO+/- years, but an ability to simulate the effects of NAO activity on regional climate would be more valuable. Models which capture these regional effects provide a much fuller picture of this large-scale driver. As the NAO influences much of winter climate in this region, skilful representation of not only the frequency of occurrence of its phases but also its regional climatic effects would be a very desirable ability in a climate model.

6.4.4 Case Study 1: HadRM3P-a driven by HadAM3P

Spatial pattern maps for HadRM3P-a are given in Figure 6.10. It is clear from Figure 6.10a, the map of mean sea level pressure patterns, that HadRM3P does capture the enhanced pressure gradient associated with a positive NAO phase. However, MSLP in negative NAO years is not as uniform across the domain as in observations. Instead, a slight gradient is associated with NAO negative years.

The mean sea level pressure maps for the European domain, with no NAO division of data, showed a marked negative bias to the north of the domain and a marked positive bias to the south in HadRM3P-a. The analysis of the UK and Irish domain shows that this error arises through the combination of a slight pressure gradient in NAO negative phases and an enhanced gradient in NAO positive phases (Figure 6.10a). MSLP to the north of the domain is lower than observed, while MSLP to the south is higher.

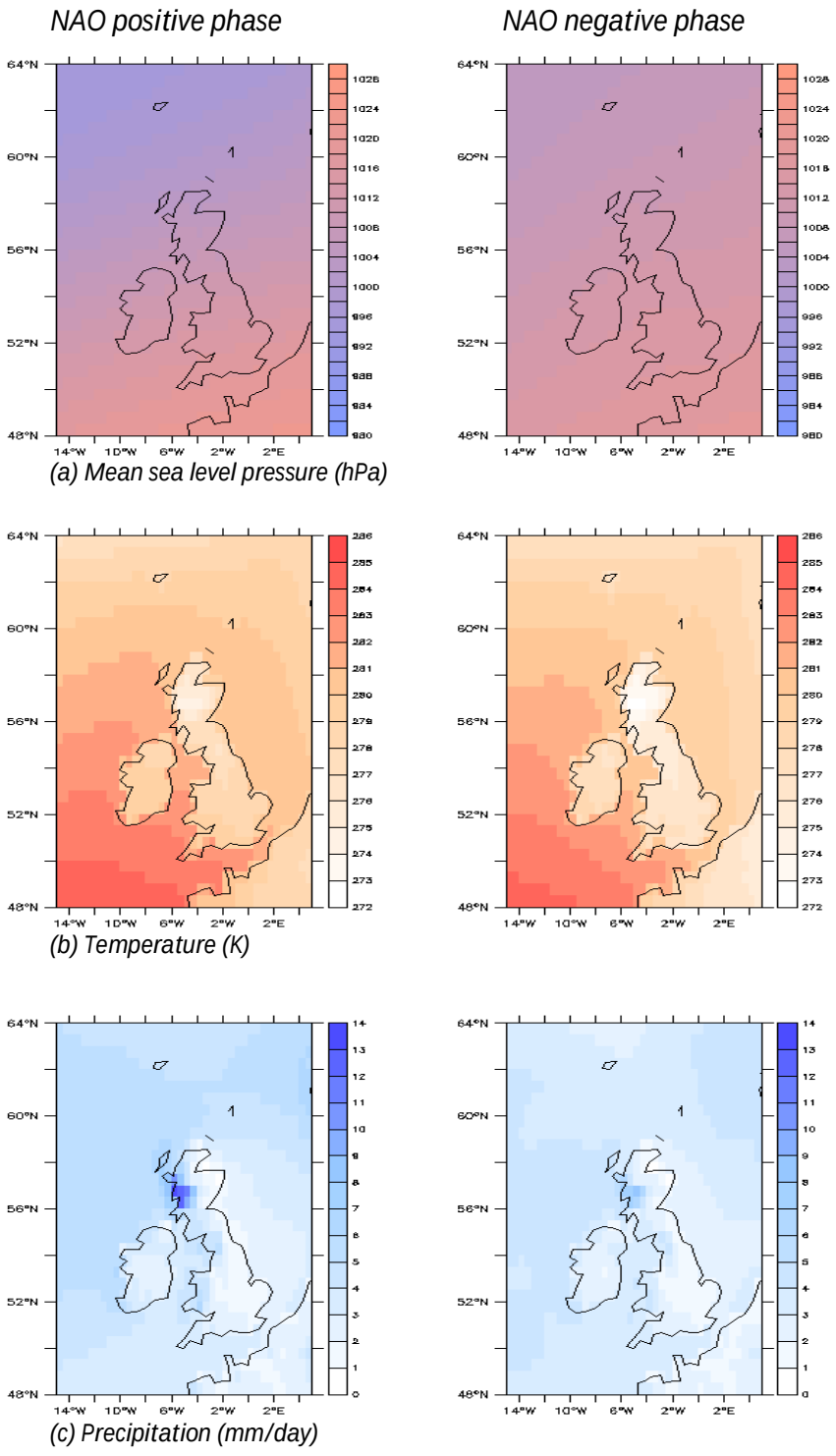


Figure 6.10: Average spatial patterns of mean sea level pressure (a), temperature (b) and precipitation (c) over the UK and Ireland in positive NAO years (left) and negative years (right), as modelled by the HadRM3P-a RCM simulation over 1961-1990.

With an enhanced pressure gradient, one might expect the HadRM3P-a simulation to also model enhanced NAO effects on regional climate. While temperature (Figure 6.10b) over the ocean in NAO positive years is warmer than in NAO negative years, with the warm temperatures clearly reaching further north, the difference on land is somewhat less pronounced. The skill score assessment in Chapter 4 identified that for the Ireland, HadRM3P-a models warmer winter temperatures than observed and so representation of the NAO may contribute to this error.

The model captures increased precipitation of the NAO positive phase over Western Scotland (Figure 6.10c), however in the observed NAO patterns the increased precipitation is a domain-wide characteristic and not restricted to this specific area. For Ireland, there is minimal difference between positive and negative NAO years. This may explain the drier than observed winter conditions simulated by this model for Ireland.

6.4.5 Case Study 2: RCAO-H driven by HadAM3H

Spatial patterns for RCAO-H are given in Figure 6.11. In the skill scores assessment, this model simulated warmer and wetter average winter conditions than observed and it also models a steeper average pressure gradient for winter, conditions which could correspond to enhanced NAO effects.

Like HadRM3P-a, this model simulates a slight pressure gradient across the UK and Ireland in NAO negative years, rather than the more uniform conditions expected based on the observational patterns. In NAO positive phases, the pressure gradient is quite pronounced and in combination, this could explain the gradient seen in the average winter MSLP maps (Figure 6.11a).

Again, an important consideration is whether the enhanced NAO positive MSLP conditions are associated with the biases in temperature and precipitation. However, temperature over both ocean and land exhibits a very similar pattern in both NAO positive and negative years with only slight differences in parts of the domain (Figure 6.11b). If the positive bias were the result of an overactive NAO in

the model, one would expect NAO positive years to be significantly warmer than NAO negative years, but this is not the case.

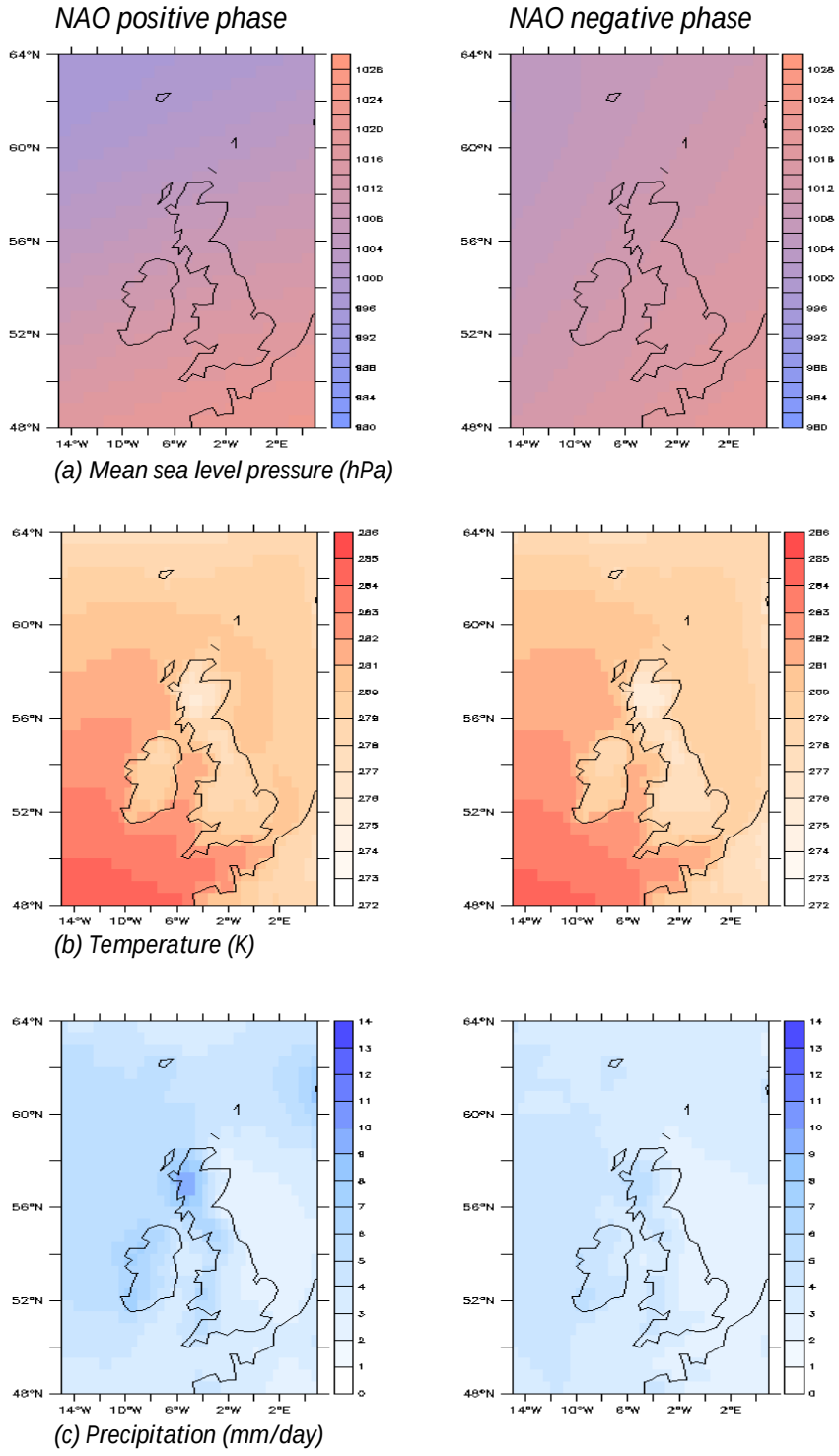


Figure 6.11: Average spatial patterns of mean sea level pressure (a), temperature (b) and precipitation (c) over the UK and Ireland in positive NAO years (left) and negative years (right), as modelled by the RCAO-H RCM simulation over 1961-1990.

However, this model captures the increased precipitation of the NAO positive phase well particularly over Western Ireland and Scotland (Figure 6.11c). Since the pressure gradient is enhanced in RCAO-H and the model captures the effects of NAO activity on precipitation with skill, the wetter than observed winter conditions simulated by this model may be attributable to its representation of the NAO. To determine this more conclusively, the next section of the analysis will investigate whether the excess modelled precipitation is associated with westerly winds, which are largely controlled by the NAO, or other sources.

6.4.6 Case Study 3: HIRHAM-E5 driven by ECHAM5

Spatial patterns for HIRHAM-E5 are given in Figure 6.12. In the skill scores assessment, this model simulated average winter conditions for temperature and precipitation that were quite close to the observed, but it also models a much lower MSLP across the north of the European domain in winter, and a higher MSLP across the Mediterranean. The result is a steeper average pressure gradient for winter. One would expect such an error to influence the simulation of average temperature and precipitation, yet this model simulates the averages of these climate parameters with apparent skill. Further analysis is required to determine the level of confidence with which output from this model should be considered.

Although there is a notable bias in the model's representation of seasonal average MSLP, it captures the difference in pressure patterns between NAO positive and negative years quite well (Figure 6.12a). NAO positive years are characterized by a noticeable pressure gradient across the UK and Ireland, while MSLP in NAO negative years is more uniform across this sub-section of the model domain. The model appears to capture the pressure differences that underlie the NAO, but it is important to also consider whether these pressure differences have the effect on temperature and precipitation that is expected. Temperature does appear to be warmer in NAO positive years than in NAO negative years both over the ocean and on land and this pattern is especially apparent over ocean gridcells (Figure 6.12b).

This model does capture increased precipitation in the NAO positive phase, particularly over Western Ireland and Scotland (Figure 6.12c). These results suggest that HIRHAM-E5 captures the NAO and its associated effects on regional climate

reasonably well. However, if the NAO is captured accurately, such large biases in MSLP should lead to an amplification of NAO behaviour, yet this has not occurred here.

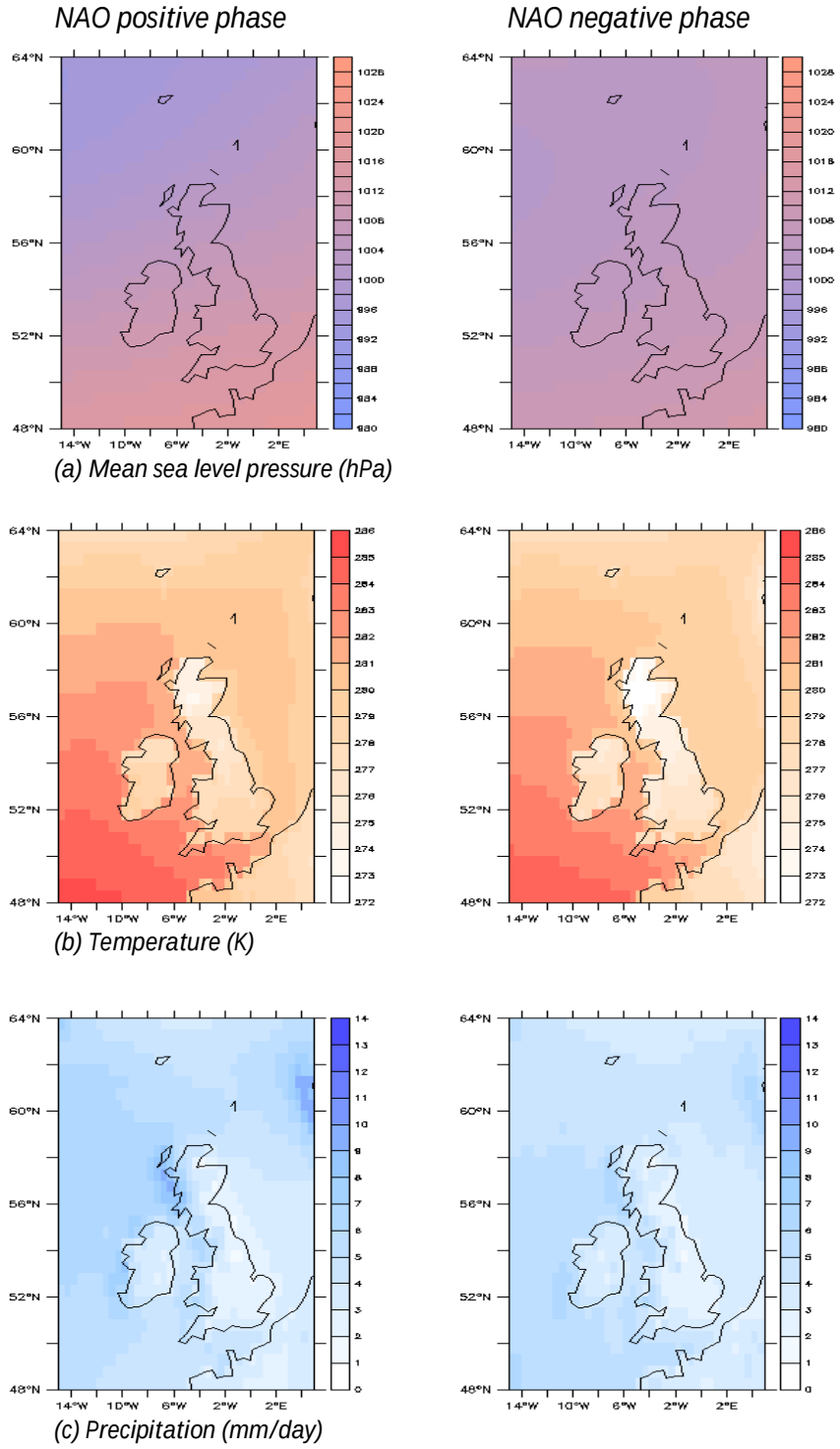


Figure 6.12: Average spatial patterns of mean sea level pressure (a), temperature (b) and precipitation (c) over the UK and Ireland in positive NAO years (left) and negative years (right), as modelled by the HIRHAM-E5 RCM simulation over 1961-1990.

6.4.7 Case Study 4: ARPEGE-a driven by Observed SSTs

Spatial patterns for ARPEGE-a are given in Figure 6.13. In the skill scores assessment, this model simulated average winter conditions for temperature and precipitation that were quite close to the observed, but it also models a much steeper average pressure gradient for winter than any other model, which is a cause for decreased confidence in its skill.

ARPEGE-a appears to model a pronounced pressure gradient regardless of NAO phase (Figure 6.13a). Although pressure across the north of the UK and Irish domain is lower in NAO positive years, it is only slightly higher in NAO negative years. Certainly, the modelled NAO negative pattern could not be described as uniform as the observed pattern was.

Correspondingly, there is minimal difference between the temperature patterns associated with NAO positive and NAO negative years in this model (Figure 6.13b). However there is an effect on precipitation, with NAO positive years tending towards wetter conditions, particularly over Western Scotland (Figure 6.13c).

This model produces a range of different results with little consistency between them. There is a significant mean winter pressure gradient bias, but little preservation of the temperature patterns associated with NAO phases at the regional scale. This suggests that the NAO is not properly represented in the model. One would expect an error in such an important large-scale driver to have impacts on regional climate. Without the moderating effect of the NAO, simulated winter climate should be colder and drier than observed, yet ARPEGE-a simulated winter climate averages close to the observed. With such a range of errors occurring and problems emerging in the model's ability to simulate a driver as important as NAO, it is possible that ARPEGE-a derives its skill from error cancellation rather than genuine ability.

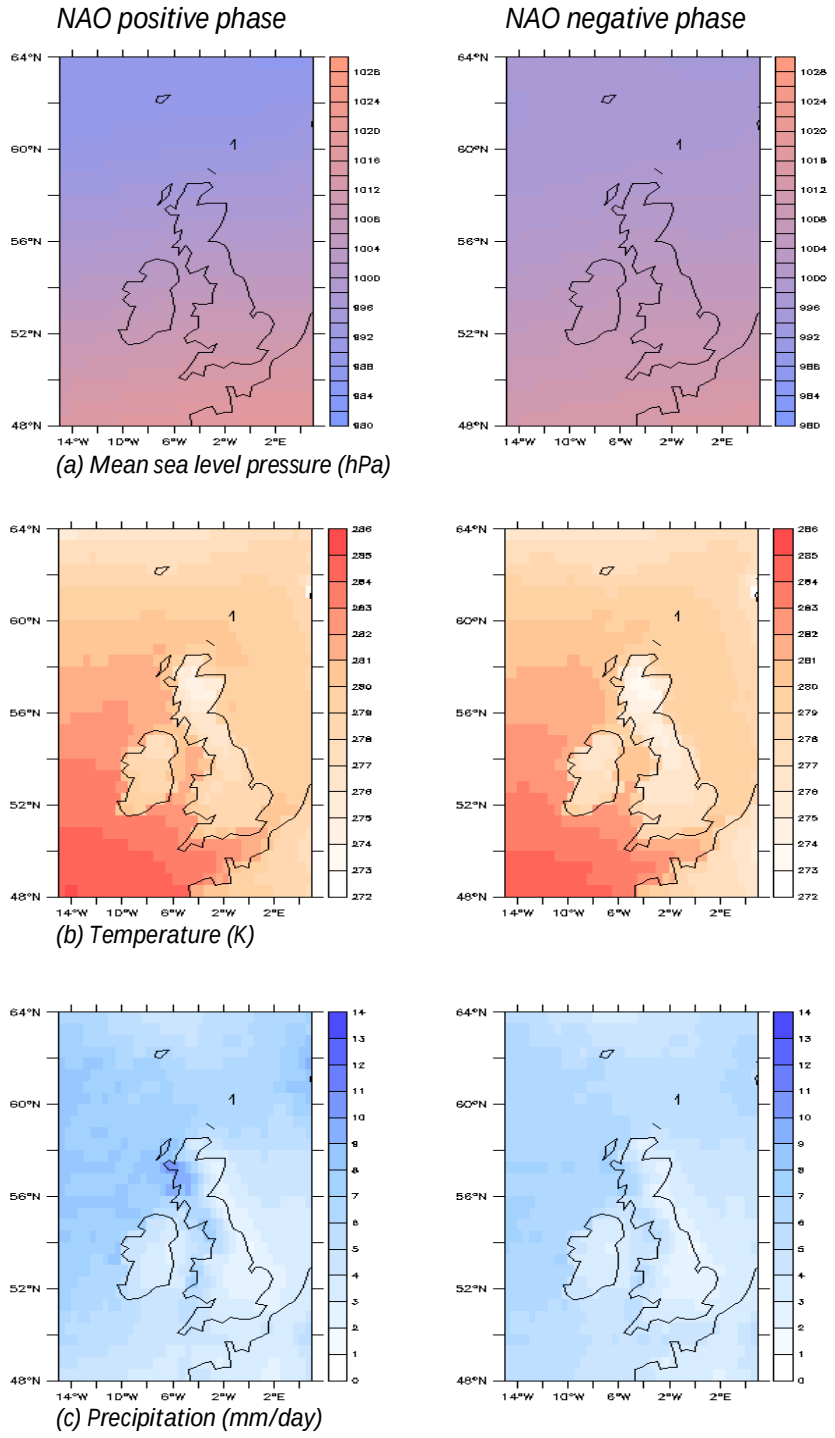


Figure 6.13: Average spatial patterns of mean sea level pressure (a), temperature (b) and precipitation (c) over the UK and Ireland in positive NAO years (left) and negative years (right), as modelled by the ARPEGE-a RCM simulation over 1961-1990.

6.4.8 Case Study 5: RCAO-E4 driven by ECHAM4-OPYC

Spatial patterns for RCAO-E4 are given in Figure 6.14. In the skill scores assessment, this model simulated much warmer and wetter winter conditions for temperature and precipitation than the observed. Based on those biases, this model appears to be a less skilful model than the other ensemble members. However, this model was one of only two that did not simulate a heightened pressure gradient across Europe in winter. Instead, RCAO-E4 displayed a systematic positive MSLP bias across the domain.

It is evident from the NAO-related MSLP patterns that pressure in this model is indeed much higher than in the others. However, the model does capture the difference in pressure patterns for positive and negative NAO years (Figure 6.14a), with a north-south gradient occurring in positive years and more uniform conditions occurring in negative years.

There is a notable difference between the temperature patterns associated with NAO positive and NAO negative years in this model (Figure 6.14b), with NAO positive years simulated as warmer. Additionally, this warming is more apparent over land gridcells than in some of the other models. There is also an effect on precipitation, with NAO positive years tending towards wetter conditions (Figure 6.14c).

Although this model simulates erroneous values for mean temperature and precipitation, it captures the dynamics of the NAO quite well. While the systematic pressure bias should not interfere with the gradient of pressure across the domain, further analysis may indicate whether this error has caused amplified NAO effects, which could in turn explain the systematic errors in temperature and precipitation.

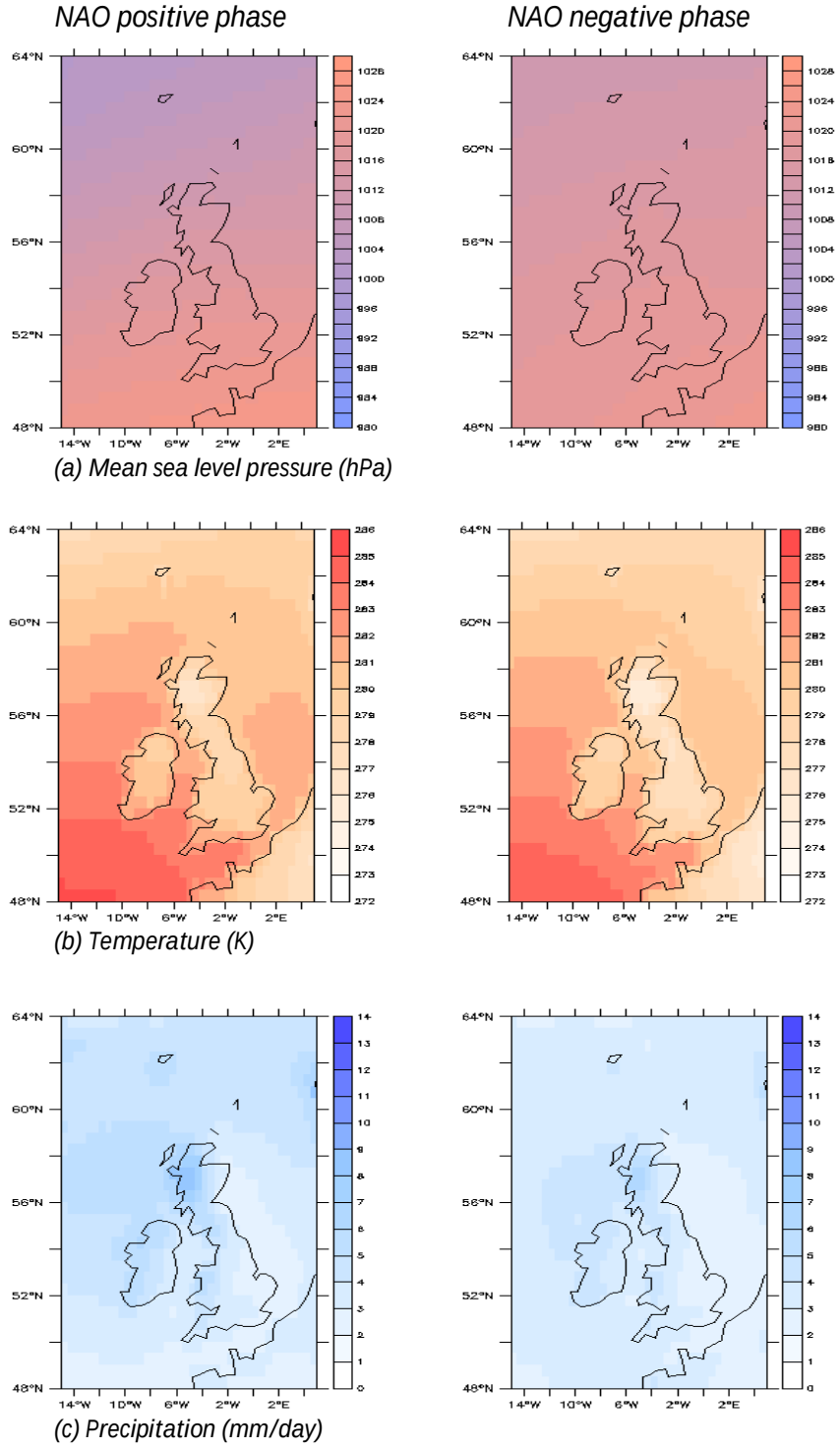


Figure 6.14: Average spatial patterns of mean sea level pressure (a), temperature (b) and precipitation (c) over the UK and Ireland in positive NAO years (left) and negative years (right), as modelled by the RCAO-E4 RCM simulation over 1961-1990.

6.4.9 Summary

The RCMs examined in this section display varying levels of skill in simulating the effect of an important large-scale climate driver, the NAO, on regional climate patterns.

HadRM3P displays the least skill. In both NAO+ and NAO- years, a north-south MSLP gradient can be observed. However, such a pattern should not be present in NAO- years. There is little difference in precipitation patterns over the UK and Ireland in NAO+ and NAO- years, suggesting that the influence of this driver on precipitation may not be represented well in the model. However, there is a slight increase in temperature in NAO+ years, suggesting that the model may possess some skill in representing this aspect of the NAO's influence on regional climate.

RCAO-H and ARPEGE-a also exhibit a MSLP gradient in both NAO+ and NAO- years. Both of these models also fail to capture the effects of NAO variability on temperature patterns. However, both RCAO-H and ARPEGE-a tend towards significantly wetter conditions in NAO+ years. This feature is particularly noticeable over mountainous areas. While these two models do not possess skill in simulating all aspects of the NAO's influence on regional climate, they are quite successful in capturing the precipitation patterns associated with NAO phases.

Finally, both HIRHAM-E5 and RCAO-E4 successfully simulate the MSLP, temperature and precipitation effects of NAO+ and NAO- years. NAO+ years are characterized by a noticeable pressure gradient which is absent in NAO- years. NAO+ years are also warmer and wetter on average in these models than NAO- years. As such, HIRHAM-E5 and RCAO-E4 display the most skill at simulating the effects of the large-scale driver on regional climate patterns.

6.5 THE IMPACT OF LARGE-SCALE VARIABILITY ON SIMULATED IRISH CLIMATE

6.5.1 Data and methods

To further explore the underlying cause of the RCM errors and assess whether errors in the representation of the large-scale driver effect the simulation of regional climate, an objective weather classification was applied to 5 RCMs. The RCMs chosen are HadRM3P (a) and RCAO driven by HadCM3/HadAM3H, HIRHAM driven by ECHAM4-OPYC/ECHAM5, RCAO driven by ECHAM4-OPYC and ARPEGE (a). An automatic Lamb classification (Jenkinson and Collison, 1977) uses a set of simple rules applied to gridded pressure maps to determine westerly flow, southerly flow and so forth. Applications of the technique include Goodess and Palutikof (1998) who applied automatic Lamb classification tosouth-east Spain and Linderson (2001) who used the technique to analyse data for southern Scandinavia. Here it is used to classify the monthly data underlying the seasonal MSLP maps. Equations are calculated using the points indicated in Figure 6.15, to determine the predominant wind direction in each winter month. Monthly data rather than seasonal data is used in this analysis as wind direction varies on a much finer temporal scale than seasonal. If W is positive and S is negative, 360^o is added to D. In all other cases, 180^o is added. Equations are adapted from Jones *et al.* (1993):

Equation 6.3: To determine wind direction frequency

Note: $\underline{2}$ is a multiplier, not a point reference.

$$W = \tfrac{1}{4}(7+8+9+10) - \tfrac{1}{4}(1+2+3+4)$$

$$S = \frac{1}{\cos(53.5^\circ)}\left[\left(\tfrac{1}{4}(4+10+(\underline{2}x6)\right) - \left(\tfrac{1}{4}(1+7+(\underline{2}x5)\right)\right]$$

$$D = \tan^{-1}(W/S)$$

where

W = westerly flow,

S = southerly flow,

D = wind direction and

53.5^o = the bisecting latitude of the analysis grid (Figure 6.15).

Although wind direction can vary on a daily basis, a comparison of monthly data from RCAO driven by ECHAM4 and daily data from the ECHAM4 GCM shows that applying the method to monthly data adequately captures the overall shape of the wind direction frequency distribution (Figure 6.16). The frequencies with which the various wind directions occur and precipitation amounts associated with each classification are extracted from the data. As precipitation output is in units of mm/day, this figure is multiplied by 30, the number of days in the 'modelled month', to determine total monthly rainfall and summed over a number of months to determine the total precipitation associated with a wind direction.

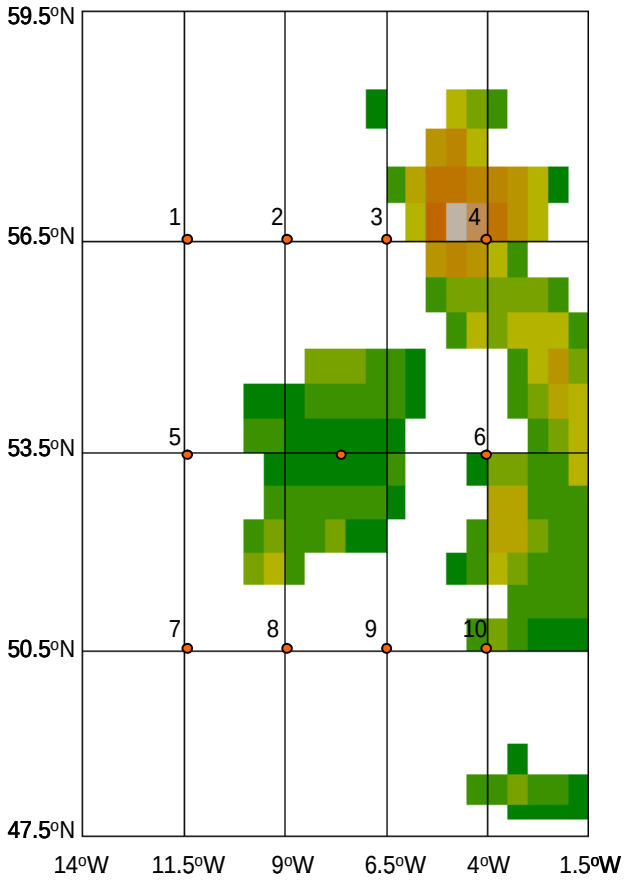


Figure 6.15: Grid used for wind direction calculations.

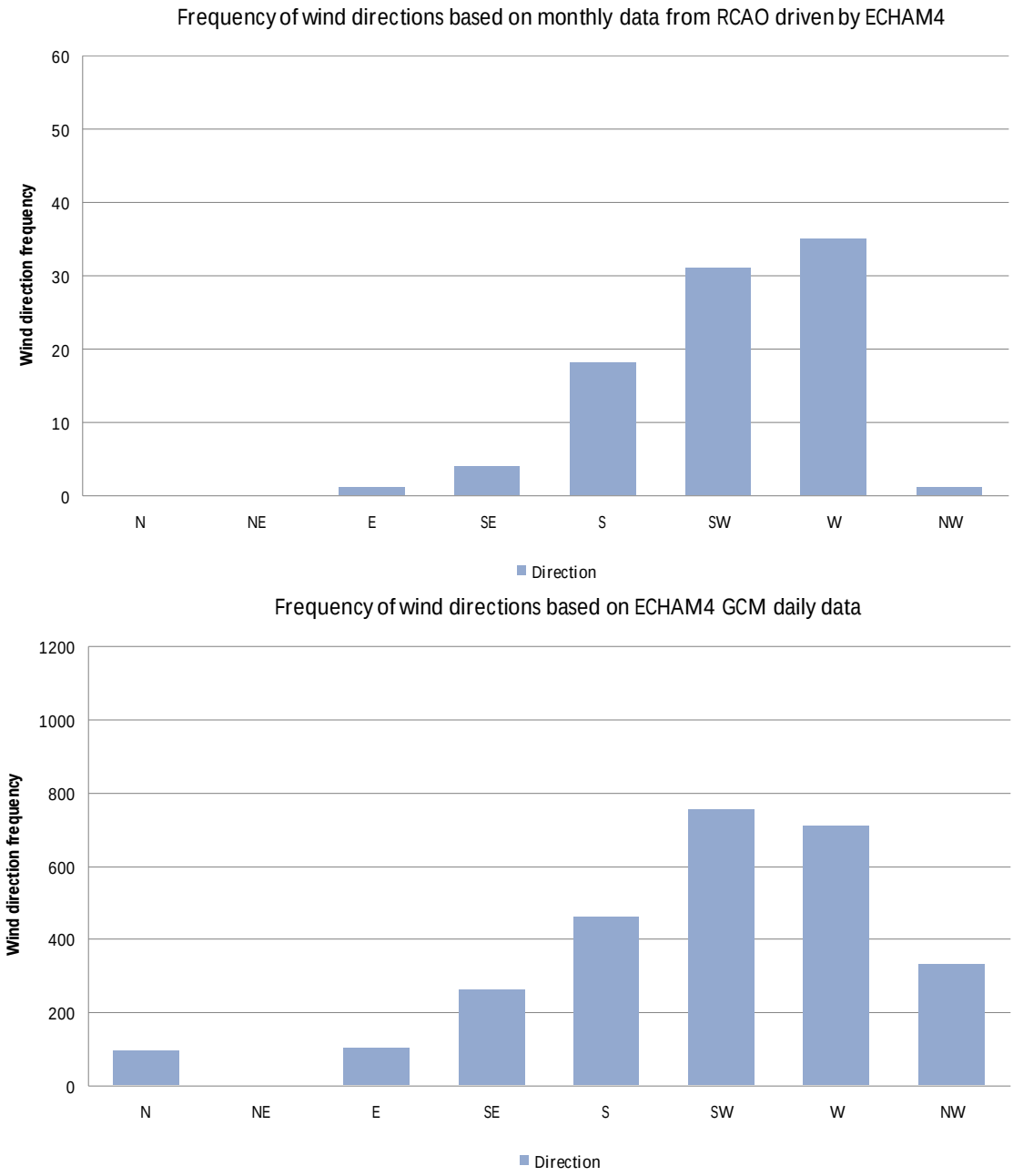


Figure 6.16: Wind frequency distributions from monthly data from RCAO driven by ECHAM4 (top) and daily ECHAM4 GCM data (bottom) for 1961-1990.

6.5.2 Results

Figures 6.17a and b displays observed and modelled wind direction frequencies and associated precipitation amounts for the case study models. The most noticeable characteristic of these graphs is that none of the models skilfully simulate the observed wind direction distribution. While HadRM3P-a and RCAO-H display a similar wind direction distribution, perhaps due to sharing GCM drivers

from the same model centre, the associated precipitation amounts are quite different. Most of the rainfall in these models is associated with south-westerly winds and this wind direction occurs with similar frequency in both models.

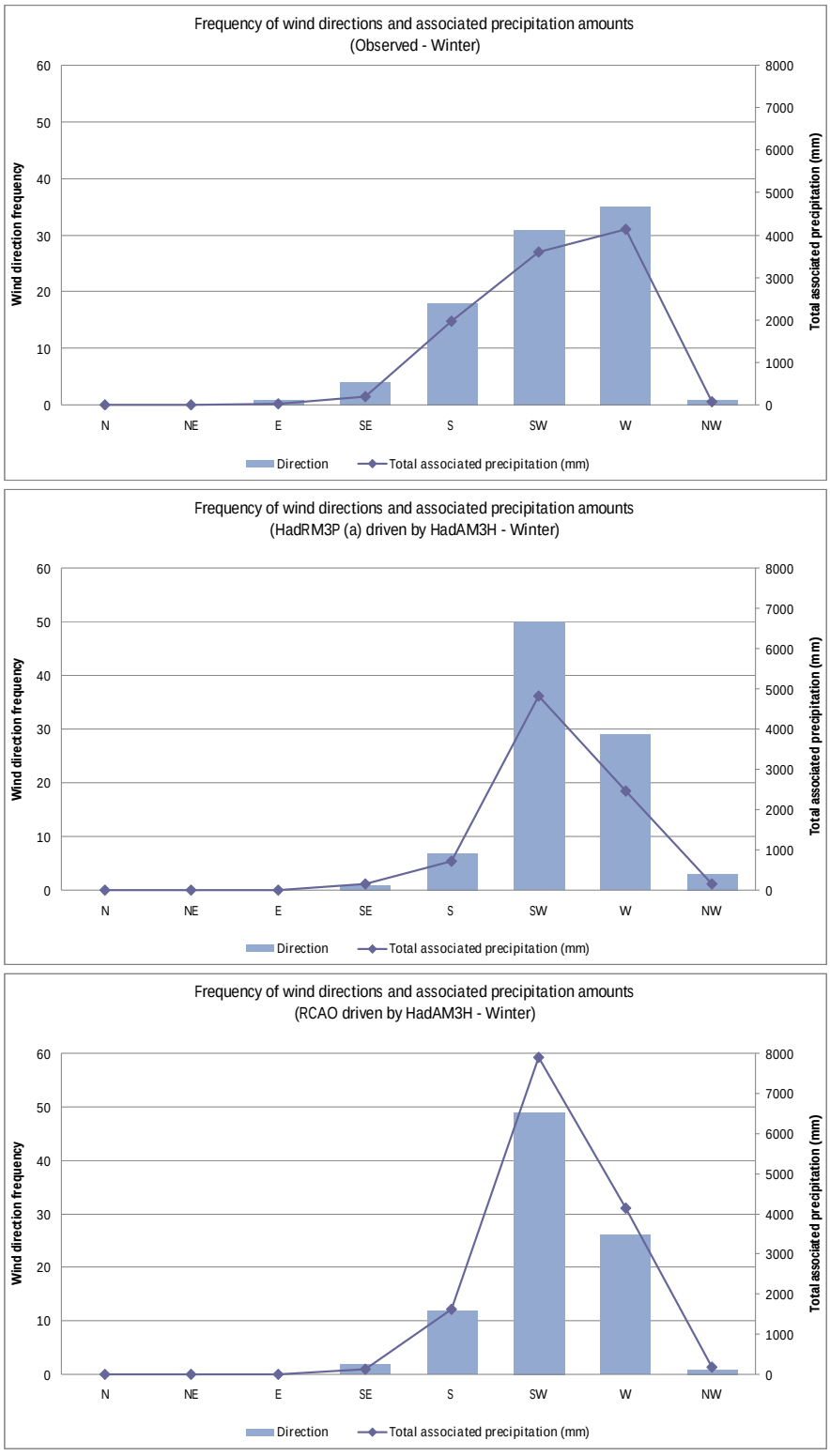


Figure 6.17a: Observed and modelled wind frequency distributions from monthly data

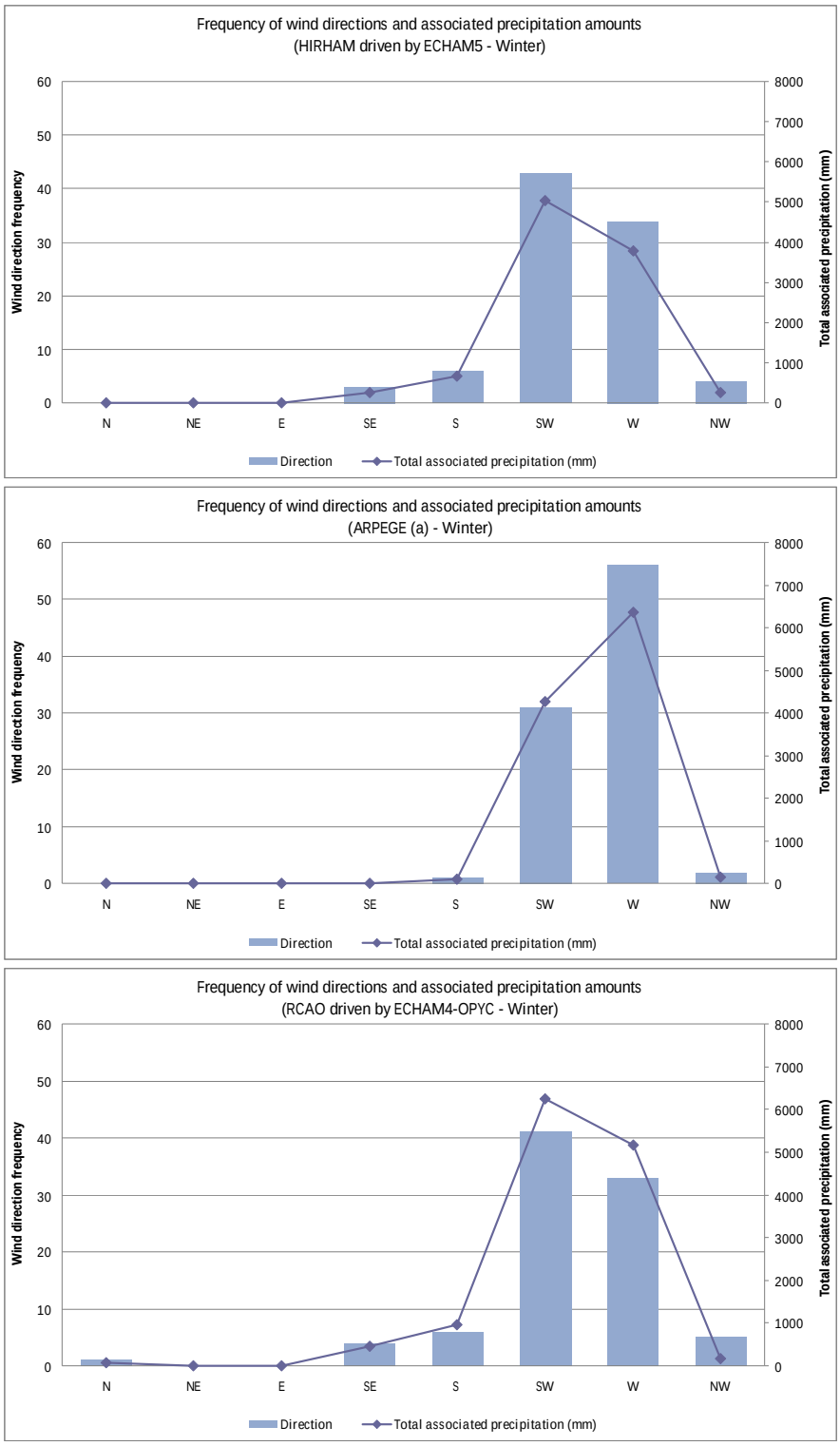


Figure 6.17b: Observed and modelled wind frequency distributions from monthly data

However, in RCAO-H more rain is associated with these winds, making it a wetter model overall. Although HadRM3P-a slightly overestimates south-westerly rain compared to the observed, it underestimates rain associated with all other win directions, resulting in a drier model overall than observed.

In both HIRHAM-E5 driven by ECHAM5 and RCAO-E4 driven by ECHAM4-OPYC, the difference in frequency between south-westerlies and westerlies is less pronounced. However, more precipitation is associated with these wind directions in RCAO-E4, making it a wetter model. In HIRHAM-E5, although the fractions of precipitation that can be attributed to the different wind directions are not the same as the observed, when totaled they amount to a similar level of rain, making HIRHAM-E5 apparently more skilful model overall.

Unlike the other models, ARPEGE considerably overestimates the frequency of westerlies, and has a correspondingly large amount of associated precipitation. However, while the other models also have some contributions from the south-east, south and northwest directions, these winds are greatly underestimated in the ARPEGE simulation. As a result, these wind directions contribute only a small amount of rain to the ARPEGE total. Therefore, on balance, it appears to model the observed winter precipitation skilfully. In fact, it produces the right overall precipitation amount for the wrong reasons.

ECHAM4-OPYC, which was used to drive RCAO-E4, is an AOGCM and exhibited systematic MSLP bias. This suggests that the differences in RCAO-E4's modelled wind direction frequencies and precipitation amounts are a result of RCAO-E4's internal model construction, as a systematic MSLP bias should not impact the pressure difference across Europe. As both simulations of RCAO, with different driving GCMs, tend towards wetter conditions, and all simulations of HIRHAM, again with different driving GCMs, tend to represent overall precipitation skilfully, it would seem that RCAO is simply a wetter RCM. However, analysis of wind frequency in one HIRHAM simulation does suggest that the model may simulate overall precipitation well for the wrong reasons.

It is interesting to note that the models driven by HadAM3P, HadAM3H and ECHAM5 have similar spatial patterns of MSLP bias. All three of these models are AGCMs, used as part of a double-nested technique to drive the RCMs. This may

indicate that for a maritime country like Ireland, a fully-coupled AOGCM is a better choice of driver.

6.6 DISCUSSION AND CONCLUSIONS

The results of this chapter illustrate that model averages are not a good indicator of a model's ability to simulating the climate phenomena that underpin mean temperature and precipitation. Five case study models were chosen, to represent the RCM/GCM combinations of the full ensemble. Results are summarized in Table 6.1.

The NAO is associated with changes in temperature and precipitation when variation in the pressure difference between the Azores High and Iceland Low alter the monthly mean flow over the Atlantic, shifting storm tracks northwards (Hurrell and van Loon 1997). Only two of the case studies, HIRHAM-E5 and RCAO-E4, capture the MSLP, temperature and precipitation patterns associated with NAO activity. As such, errors in representation of the pressure systems could have significant impacts on the winter climate of the other case study models. However it is important to note that mean temperature and precipitation values in a RCM are affected by many factors within the model and it is unlikely that any one source can account for all the errors identified.

As noted earlier, the similarity of MSLP error patterns within GCM driver groups may indicate that errors in the representation of the NAO may arise in the GCM and cascade through to the RCM outputs. In an analysis of mean circulation indices in GCMs, van Ulden *et al.* (2007) found positive westerly biases in of the HadAM3H AGCM, ARPEGE (included as a variable resolution AGCM) and the ECHAM4-OPYC AOGCM, which also suggests that RCM errors in NAO representation arise from the boundary conditions supplied by the GCM drivers.

However, it is clear from the results of this chapter that RCMs can respond quite differently to GCM deficiencies. Only in ARPEGE-a is the westerly bias reported in the literature carried through to the regional simulation. In all the other case study models, the south-westerlies are the notable bias. The HadAM3H GCM,

which is used to model many of the simulations in the ensemble, has a tendency to model a steeper pressure gradient than the observed (Jacob *et al.*, 2007), which in turn would impact how the RCMs driven by this model, such as case study 2, RCAO-H, driven by HadAM3H, represent temperature and precipitation.

Model	Mean winter temperature	Mean winter precipitation	Winter MSLP bias	Winter wind frequencies	NAO+/- patterns
Case study 1: HadRM3P-a	Warmer than observed	Drier than observed	Enhanced gradient	More SW winds but associated precip is lower than expected.	MSLP gradient still present in NAO- years. Slight NAO+/- temp effect but little precip effect over Ireland.
Case study 2: RCAO-H	Warmer than observed	Wetter than observed	Enhanced gradient	More SW and associated precip.	MSLP gradient still present in NAO- years. Captures precip effects but little temp effect.
Case study 3: HIRHAM-E5	Close to observed	Close to observed	Enhanced gradient	Overestimated SW winds and precip, nullified by underestimated S winds and precip.	Captures MSLP, temp and precip effects.
Case study 4: ARPEGE-a	Close to observed	Close to observed	Much enhanced gradient	Much more W precip but other directions underestimated.	Captures precip effects but little temp effect.MSLP gradient still present in NAO- years.
Case study 5: RCAO-E4	Much warmer than observed	Wetter than observed	Systematic bias	More SW and W precip.	Captures MSLP, temp and precip effects.

Table 6.1: Summary of case study results.

Although there is no information in climate modelling literature regarding HadAM3P specifically, it is closely related to HadAM3H and the models driven by it display the same issues regarding MSLP and temperature. This suggests that this driver is also less skilful at representing the observed pressure gradient. However, the

RCMs driven by HadAM3P, such as case study 1, HadRM3P-a, tend towards drier conditions, which is a marked departure from the behaviour of the other Hadley-driven RCMs such as RCAO-H. Whether this is attributable to the RCM or the AGCM is unknown.

The implications of these results on climate scenario development are very important. Many previous studies have used skill in representing the averages of key climate parameters such as temperature and precipitation as an indicator of overall model skill, but these results highlight the potential difficulties in this approach.

For example, ARPEGE-a, which has the greatest bias for MSLP also models the mean winter temperature and precipitation most skilfully. On closer investigation, this model fails to capture the spatial patterns of MSLP and temperature that are associated with NAO activity. ARPEGE-a does capture the effects of NAO activity on precipitation and in fact overestimates westerly winds and their associated precipitation. As such, one would expect this model to tend towards wetter conditions than observed in winter. Yet due to the under representation of other wind directions in this model, an apparently 'skilful' average winter precipitation amount is acquired. Based on average temperature and precipitation values, this model would appear to be one of the most skilful in the overall ensemble. However, on closer analysis it becomes apparent that this model derived much of its skill in regards to winter precipitation from error cancellation. Models that behave in this manner do not provide a robust basis for informing and testing climate strategy, as these type of errors cannot be relied upon to remain constant through time. If a change in forcing conditions effects the frequency of other winds and the precipitation associated with them but not the westerlies, the errors would no longer nullify each other and the resulting value for precipitation in that future forcing scenario would be flawed. Additionally, without investigating how the NAO is represented in the model, this uncertainty and the potential error associated with it would be unknown and unaccounted for in scenarios based on this model.

Conversely, the ECHAM4-OPYC-driven models, such as case study 5, RCAO-E4, display a number of systematic biases in winter. MSLP is higher across the European domain and average values for Irish temperature and precipitation in winter are also higher. Yet this model captures all the spatial patterns associated with

NAO activity. The MSLP patterns associated with positive and negative NAO years are represented well and the temperature and precipitation patterns suggest that the effects of NAO activity on regional climate are also captured by the model. Based on climate averages, RCAO-E4 appears to be a less skilful model, yet it captures the climate dynamics that underpin mean winter temperature and precipitation quite well.

In light of these results, an important question is which model, if any, provides the most useful information. In this instance, as RCAO-E4 captures the dynamics of the climate system more realistically, it is arguably a more robust model than ARPEGE-a. Systematic errors in mean climate parameters can be overcome to varying degrees. For example, one technique, the delta change method, is to subtract the model's present-day values from its future values to calculate the temperature or precipitation response to climate change and add this signal to the present-day observed climate to determine future scenarios. Such an approach may eliminate model systematic bias, although there is a possibility of the model bias changing through time and under different forcing conditions. Conversely, if there are errors in a model's representation of important large-scale climate drivers, there is no way to overcome this apart from revising the model itself.

In short, though there is uncertainty associated with RCAO-E4's output, it has more potential usefulness than ARPEGE-a. As ARPEGE-a's skill comes from error cancellation rather than genuine modelling ability, it is a much less reliable tool. The next consideration is whether this information can be used to inform future climate scenarios developed using the ensemble technique and what impact the inclusion or exclusion of information about model performance has on the resulting projection.

CHAPTER 7
APPROACHES TO DEVELOPING FUTURE CLIMATE SCENARIOS

7.1 INTRODUCTION

The work presented thus far in this thesis has generated much information about the 19 RCMs under investigation, in particular the five case study models chosen for the NAO analysis. However, if this knowledge is to enhance the reliability of future scenarios and help account for the uncertainty surrounding modelled scenarios, a method must be used to create those scenarios which take the known information about RCM skill into account.

To determine the difference in projections when varying levels of knowledge are used to inform model choice, a Bayesian model averaging (BMA) approach is used with the uncertainty surrounding the model reflected in the weights associated with each model. The BMA approach takes account of uncertainty in model selection, reducing the potential for over-confident projections (Hoeting *et al.*, 1999). The technique can be used to construct a skill score-weighted ensemble probability distribution function (PDF) from the outcomes projected by different RCMs, which accounts for variations in model skill and reliability, to determine the most probable outcome. Various weighting schemes are applied, each of which is informed to a varying degree about model performance. Spatiotemporal skill scores, objective estimates of NAO representation skill and a combination of both spatiotemporal skill scores and NAO information are used to weight models, resulting in three different weighting schemes. Projections are also calculated using an unweighted approach, representative of a case in which all model projections are assumed to be equally likely, an assumption that has been proven to be unlikely based on the results of previous chapters. The outcomes obtained using each method are then compared and discussed.

7.2 APPROACHES TO GENERATING ENSEMBLE CLIMATE MODEL PROJECTIONS

Simple ensemble methods have a history of use within short-term weather forecasting and a widely used approach is to treat the ensemble mean as a single projection or best estimate of future conditions (Whitaker and Loughe, 1998). This approach is often found to provide a more skilful projection compared with any single projection from an individual ensemble member.

Examples of the mean ensemble method applied to climate model data include Gates *et al.* (1998), who assessed the skill of an ensemble of AOGCMs and Rinke *et al.* (2006), who used an ensemble of RCMs to investigate Arctic spatiotemporal patterns for a range of climate parameters. Both studies found that the ensemble means outperformed the individual models for certain climate parameters. However, Tracton and Kalnay (1993) note that the increase in skill is in part due to the cancellation of errors in the individual forecasts when the ensemble members are averaged.

Error in the mean ensemble projection depends on the level of independence of model errors as well as the error associated with the individual models that compose the ensemble (Goerss, 2000). However, Abramowitz (2010) notes that the independence of models employed in ensembles is rarely quantified. Additionally, as demonstrated in Chapter 4 and Chapter 6, different RCMs driven by a common GCM can potentially give very similar projections for certain climate parameters, in this case the interannual variability. As such, there can be a high degree of uncertainty attached to mean ensemble projections if the error of the individual models and the independence of the projections is not assessed.

If one model is particularly lacking in simulative skill, the ensemble mean forecast will be affected by this and so differences in skill should be considered when formulating ensemble projections (Grimit and Mass, 2002). Much information can be generated about model skill through validation and verification studies using present-day observational data and incorporating this information into the ensemble projection provides an important opportunity to account for uncertainty and increase the confidence of the ensemble.

Weighting systems often rely on skill scores calculated by comparing the modelled climate parameter and the observed. For example, Sanchez *et al.* (2009) weighted models based on the similarity of the modelled precipitation cumulative distribution function (CDF) to the observed, resulting in a more skilful ensemble simulation of precipitation. Similarly, Yun *et al.* (2004) used a weighted multi-model ensemble using EU DEMETER (Development of a European Multi-Model Ensemble System for Seasonal to Interannual Prediction) output to generate projections of seasonal climate, in which the weights were calculated based on statistical relationships between individual AOGCM output and past observations. However, as Brown (2004) notes, simply comparing climate model outputs can result in model skill being under or overestimated. Skill in representing the mean field or a single key climate parameter does not guarantee that the processes and drivers that give rise to mean temperature or precipitation, for example, are adequately represented in the model. Lucarini *et al.* (2007) notes that the focus on mean fields has greatly influenced the development of GCMs and suggests that as the climate system is essentially a non-linear system it would be appropriate for model validation to include analysis of the representation of dynamical processes. This aspect of the model output was examined in Chapter 6 and results indicated that assessing the mean climate may not be the optimum way to characterize model skill, as deficiencies in the representation of large-scale drivers may not be detected. As such, weighting models based on seasonal mean skill-scores alone would leave much uncertainty arising from intermodel variability unaccounted for.

Another approach is to weight models according to their relative agreement. This is a component of the REA approach (Giorgi and Mearns, 2001) and was subsequently applied in Tebaldi *et al.* (2004). Yet weighting according to relative agreement is also a potentially flawed approach. As noted by Abramowitz (2010), model independence is rarely quantified in climate modelling studies and shared parameterizations or GCM drivers may lead to a high degree of similarity between models. In such an instance, weighting by relative agreement may actually promote non-independent models over independent ones. The resulting convergence of outcomes would be highly overconfident. The application of ensemble methods is often understood to generate an increase in reliability (Tebaldi and Knutti, 2007), but

this reliability is optimised when ensemble members are independent and more limited when they are not.

There are some examples in the literature of weighting approaches based on the models' abilities to simulate the dynamics of the climate system rather than the mean fields. For example, Schmitter *et al.* (2005) used model skill in representing key hydrographic properties and circulation estimates to weight members in an AOGCM ensemble to form a best estimate of the future meridional overturning circulation (MOC) in the Atlantic. Yet the mean-based skill scores approach is the more widely-used technique. The research presented so far in this thesis has illustrated that assessing models based on statistics of temperature and precipitation alone can potentially result in a misleading conclusion about model skill, while assessing models based on their representation of key climate drivers can give a more comprehensive picture of model performance. As such, it follows that weighting model output for future time periods based on spatiotemporal skill scores alone may result in less reliable projections. Additionally, there may be an opportunity to reduce the uncertainty associated with future projections and improve their reliability by incorporating information about how key climate drivers are represented in the models.

7.3 DATA AND METHODOLOGY

RCM output for 2071-2100 was again obtained from the PRUDENCE data archive. As part of the PRUDENCE project, the same RCMs which have been analysed in previous chapters were also run for the period 2071-2100 using forcing conditions associated with a particular SRES scenario. For future projections, the ARPEGE RCM, which was forced using observed SSTs in the control period, was run using driving data from the HadCM3 AOGCM. 19 projections were generated for PRUDENCE using the A2 emissions scenario and eight were generated using the B2 emissions scenario. HadRM2P, HIRHAM and ARPEGE, which are each run three times to create sub-ensembles for the control and A2 scenarios, are only run once for the B2 scenario. Therefore, there are more simulations available for the A2 scenario than the B2 scenario (Table 7.1).

GCM	RCM	A2	B2
HadCM3/HadAM3P	HadRM3P-a	✓	✓
	HadRM3P-b	✓	
	HadRM3P-c	✓	
HadCM3/HadAM3H	PROMES	✓	✓
	RACMO	✓	
	CHRM	✓	
	CLM	✓	
	REGCM	✓	✓
	REMO	✓	
	RCAO-H	✓	✓
	HIRHAM-a	✓	✓
	HIRHAM-b	✓	
	HIRHAM-c	✓	
ECHAM4-OPYC/ ECHAM5	HIRHAM-E5	✓	
Observed SSTs	ARPEGE-a	✓	✓
	ARPEGE-b	✓	
	ARPEGE-c	✓	
ECHAM4-OPYC	RCAO-E4	✓	✓
	HIRHAM-E4	✓	✓

Table 7.1: Availability of modelled data for the future emissions scenarios A2 and B2.

The first selection of projections (Section 7.5) illustrates the difference in projections that occurs when different weighting schemes are used. For these projections, output from the five case study models only is used so that information from the NAO analysis can be incorporated into the weighting process. Additionally, these models are all driven by different GCM drivers and as such, provide a more independent sample than the full suite of models.

Figure 7.1 outlines the various ensemble generation approaches that have been applied. Projections are calculated using the deterministic AEM approach, the BMA approach using equal weights (BMA-EQ), the BMA approach using spatiotemporal skill scores (BMA-SS), the BMA approach using skill in representing the NAO (BMA-NAO) and the BMA approach using both spatiotemporal skill scores and skill in representing the NAO to form a combined objective skill estimate (BMA-COM).

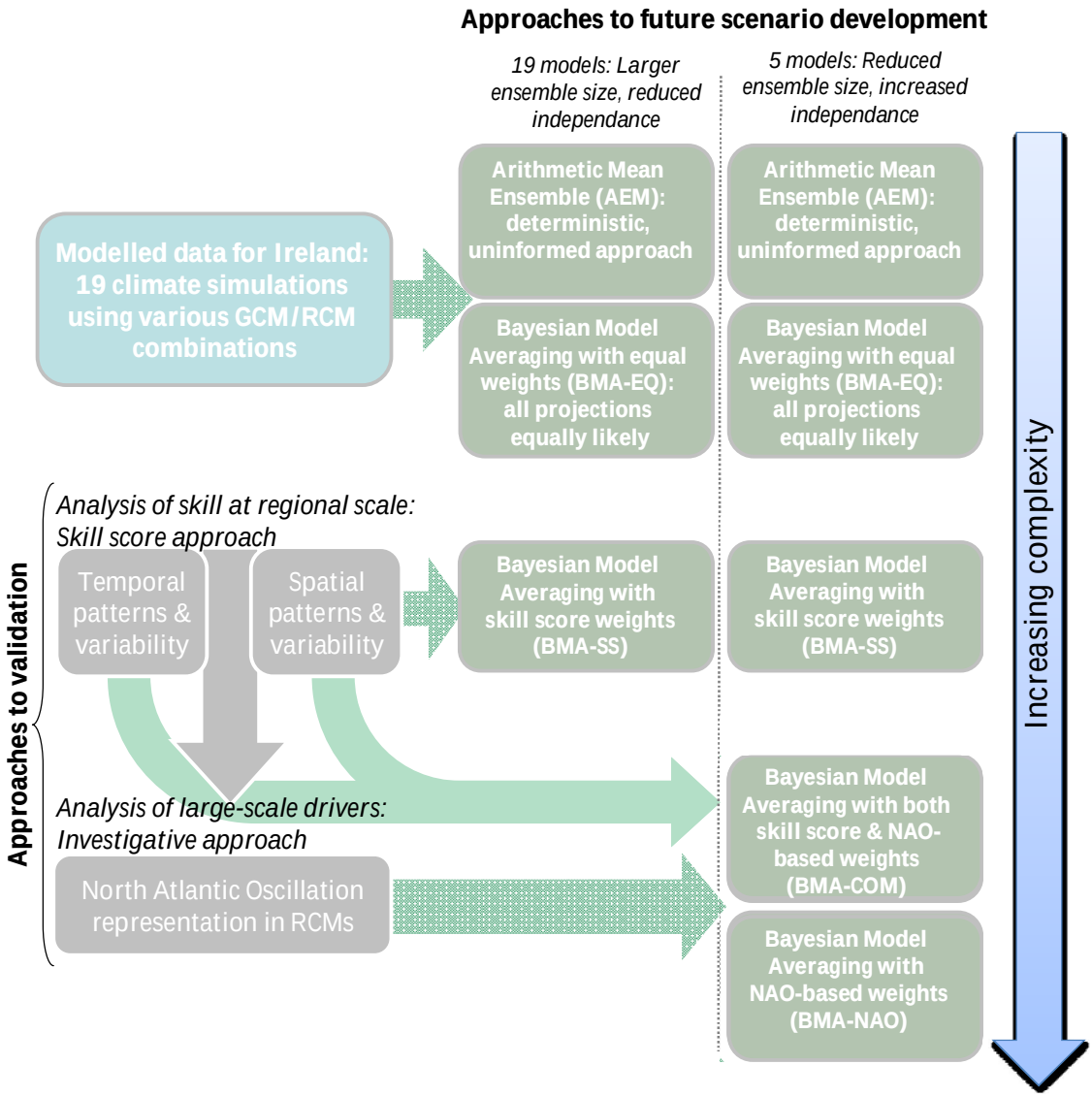


Figure 7.1: Diagram of ensemble generation approaches

The second selection of projections (Section 7.6), utilizes the full suite of climate models and illustrates how the projection varies when a larger selection of models is used. The AEM, BMA-EQ and BMA-SS approaches are applied to the full selection of RCMs. However, the reduction in both model independence and information about skill must also be taken into account when assessing these projections.

Projections are calculated for winter (DJF) and summer (JJA) temperature and precipitation data, under the A2 and B2 emissions scenarios. Projections are also

formulated for the A2 and B2 data combined, to attempt to capture some of the variability that is not accounted for by the individual scenarios. The A2 and B2 combined projections using BMA-SS, BMA-NAO and BMA-COM approaches are given in Appendix A.

7.4 OVERVIEW OF ENSEMBLE METHODS

7.4.1 Arithmetic ensemble mean (AEM) approach

The AEM approach assumes that all models are equally skilful and their projections are equally probably. Such an approach contains no information about model performance. While such an approach lacks the subjectivity that weighting inevitably introduces, there is also significant uncertainty regarding intermodel variability that is left unaccounted for. Projections from each model are treated as equally probable and given projections from N different models, the AEM is:

Equation 7.1: Arithmetic ensemble mean

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

where

$\bar{x}$ *= the arithmetic ensemble mean,*

x_n *= the individual model projection for the climate parameter and*

N = the number of individual models.

The resulting projection has no probability attached to it and although information about the range of potential future outcomes is communicated through graphical representation, no information about the likelihoods attached to the various projections is included in the AEM method.

7.4.2 Bayesian model averaging (BMA) approach using skill scores

Another option is to use skill scores to weight output from different models based on how they perform. For this method, the skill-scores generated in Chapters 4 and 5 are used to weight model projections using the BMA approach. Projections in which the weighting factors associated with the individual models are all equal are also calculated to consider the effects of omitting model skill.

The bias associated with simulations of mean Irish temperature and precipitation (Table 5.1), the temporal r values (Tables 4.6 and 4.7) and the spatial r value of the underlying seasonal spatial data (Table 5.1) are used to inform future projections of Irish temperature and precipitation. First, the issue of systematic bias will be addressed by applying a correction to those models that require it.

Developing scenarios based on the relative difference between future simulations and control simulations requires acceptance of the assumption that the difference between simulations is the climate change signal. This assumption holds only if model biases and errors stay constant over time and do not change under different forcing conditions. For example, if the bias of a model decreases under different forcing scenarios, the climate change signal of that model could be perceived to be smaller than it actually is. The results presented in this thesis have demonstrated that errors can have either systematic or random characteristics. Where errors are inconsistent over time or space in the control period (i.e. random errors), there is little reason to be confident that these patterns will remain constant under different forcing scenarios. Conversely, where errors are systematic in the control period, it is more conceivable that such errors may remain constant over time and under different forcing scenarios. As such, an approach which corrects errors while distinguishing between systematic and random bias was applied.

An r value of 0.7 or higher is regarded as evidence of a strong association between the observed and modelled patterns, while values of less than 0.7 represent weak to moderate association. Models with a bias of greater than 10% of observed precipitation and temperature (0.37mm/day or 0.47^{o}C in winter, 0.25mm/day or 1.39^{o}C in summer) and which display a Pearson r of greater than 0.7 were assumed to be systematically biased and were therefore corrected. Spatial r scores of less than 0.7 are considered indicative of potentially random bias. This bias is not corrected as

one cannot assume that the bias will remain constant in time. For example, the results of the spatial analysis in Chapter 5 show that for winter temperature, RCAO has a spatial r value of 0.91 and a bias of +2.46^{o}C, therefore this bias is corrected. Conversely, for summer, this model has a temperature bias of +0.1^{o}C but a lower spatial r value of 0.54, therefore this bias is not corrected. The figures from which bias corrections are calculated can be found in Table 5.1.

Where required, biases are corrected by subtracting the measured bias based on the present day simulation from the future value. Models with an r less than 0.7 are not altered. This step minimizes bias in the models with a significant systematic error only. Models with random errors in the underling spatial data are left unchanged as those errors are less likely to retain the same spatial distribution pattern under different forcing conditions. For example, in the control period in winter, RCAO-H and RCAO-E4 have spatial r scores of greater than 0.7 and temperature biases of 1.88^{o}C and 2.46^{o}C respectively. Therefore, these biases are corrected in the future A2 scenario data, but the other models remain unchanged (Table 7.2).

Winter temperature 2071-2100

Model	**HadRM3P-a**	**RCAO-H**	**HIRHAM-E5**	**ARPEGE-a**	**RCAO-E4**
Mean	7.42	8.12	6.63	6.96	10.12
Spatial r	0.53	0.88	0.55	0.57	0.88
Bias Correction	-	1.88	-	-	2.46
Corrected Mean	7.42	6.24	6.63	6.96	7.66

Table 7.2: Example of bias correction using winter temperature data for the A2 scenario.

This approach was assumed to be less subjective than assuming all model biases are systematic, but it is important to recognize that the decision of how to treat data before developing scenarios is another potential source of uncertainty.

As outlined in Chapter 2, Bayesian statistics differs from frequentist statistics in that subjective information regarding the "level of knowledge" about projections can be incorporated into the ensemble projection through the use of an informative prior. In this case, the priors are the weights, determined based on the present-day skill scores. Models are weighted using their seasonal spatial r value and annual temporal r value for the parameter being examined. For each model the two scores

are averaged and the resulting score is squared so that very low skill is heavily weighted against. These scores are normalized across the models contributing in each scenario so that they sum to one, giving the weights used in the spatiotemporal skill scores-based approach (BMA-SS). An example of the weightings calculation for the BMA-SS approach is given in Table 7.3. RCAO-E4 has both a high spatial r and temporal r score which combines to give a high weighting, while HadRM3P-a has a high temporal r score but a lower spatial r score and when combined, gives a lower weighting.

Winter temperature 2071-2100

Model	HadRM3P-a	RCAO-H	HIRHAM-E5	ARPEGE-a	RCAO-E4
Weighting information: BMA-SS					
Spatial r	0.53	0.88	0.55	0.57	0.88
Temporal r	1.00	1.00	1.00	0.99	0.99
Overall squared skill score	0.58	0.88	0.60	0.61	0.87
BMA-SS weights	0.16	0.25	0.17	0.17	0.25

Table 7.3: Example of BMA-SS weightings using winter temperature data for the A2 scenario.

For values within the combined range of all the models, the likelihood for each model is calculated. The likelihood associated with each model is the probability density associated with the climate parameter value for a normal distribution specified using the mean and standard deviation of that model:

Equation 7.2: BMA model likelihood

$$p(x \mid x_n) = g_n(x \mid \bar{x}_n, \sigma^2)$$

where

$p(x \mid x_n)$ *= the likelihood associated with a value of climate parameter x, given projections from model xn and*

$g_n(x \mid \bar{x}_n, \sigma^2)$ *= a theoretical normal distribution defined by mean* $\bar{x}_n$ *and variance* σ^2 *from the future projections of each model.*

As part of the assessment of temporal variability in Chapter 4, normality of both the temperature and precipitation 30 year seasonal datasets in the control period was tested using the Shapiro-Wilks test. For both parameters, the data was found to be largely normally distributed with only a small number of exceptions. Additionally, the Central Limit Theorem shows that as variables that are not normally distributed are summed, as it the case when daily precipitation, which is usually best approximated using the gamma or lognormal distribution, is accumulated and averaged into a seasonal figure, the PDF of the sum approaches the normal distribution. As such, it is valid to use the normal distribution to generate likelihood functions for future seasonal temperature and precipitation. These likelihoods are then multiplied by the respective priors to form the posterior distribution. In this case, the posterior is a weighted ensemble PDF, which takes account of intermodel uncertainty and information about model skill:

Equation 7.3: BMA weighted ensemble PDF

$$p(x \mid x_1,\ldots,x_N,x^T) = \sum_{n=1}^{N} w_n g_n(x \mid \bar{x}_n,\sigma^2)$$

where

$p(x \mid x_1,\ldots,x_N,x^T)$ = *the ensemble PDF for the climate parameter x, given projections from N models* $x_1,\ldots,x_N$ *and present-day data* x^T,

w_n = *the weight for each model and*

$g_n(x \mid \bar{x}_n,\sigma^2)$ = *a theoretical normal PDF for each model defined by mean* $\bar{x}_n$ *and variance* σ^2 *from the future projections of each model.*

The inclusion of weights effectively constrains the extent to which less skilful models contribute to the ensemble projection and allows the models with the greatest level of reliability, based on performance in the present day, to contribute most.

7.4.3 BMA approach using skill scores and objective skill estimates

The NAO analysis generated much information about model skill that is not as well quantified as the skill scores. Yet this information could further account for some of the uncertainty associated with the future climate scenario. Therefore, the models' ability to capture the NAO effect on the temperature and precipitation will be used to add further weighting to the models.

The five case study models are ranked based on their skill at simulating the NAO and its effects, as analysed in Chapter 6. HIRHAM-E5 and RCAO-E4 were found to be the most skilful, capturing the MSLP, temperature and precipitation patterns associated with NAO behaviour. The next best models are RCAO-H and ARPEGE-a, which capture the precipitation effects of the NAO but only slight temperature differences and incorrect MSLP patterns. Finally, the least skilful model is HadRM3P-a, which does not capture the precipitation or MSLP patterns associated with NAO behaviour for Ireland and only simulates a slight temperature difference between NAO+ and NAO- years.

Scores of 0.99, 0.66 and 0.33 are assigned based on performance at simulating the NAO, ranging from 0.99 for the most skilful models to 0.33 for the least skilful, as outlined in Table 7.4.

Legend
✓= NAO+/- effects are modelled with skill
? = Slight NAO+/- effect is modelled
✖= No NAO+/- effect is modelled

Overall skill levels	Skill score
HIGH	0.99
MODERATE	0.66
LOW	0.33
NONE	0.00

	Model	HadRM3P-a	RCAO-H	HIRHAM-E5	ARPEGE-a	RCAO-E4
Presence of NAO+/- patterns	MSLP	✖	✖	✓	✖	✓
	Temperature	?	?	✓	?	✓
	Precipitation	✖	✓	✓	✓	✓
	Overall skill	LOW	MODERATE	HIGH	MODERATE	HIGH
	Skill score	0.33	0.66	0.99	0.66	0.99

Table 7.4: NAO skill estimates based on analysis of model-simulated NAO in the control period.

This range was chosen to match the 0 to 1 range of the correlation coefficients and allows the different skill scores to be easily combined. These scores are again squared and normalized so that they sum to one to form the weights for the NAO-based approach (BMA-NAO).

To generate scenarios that take both skill scores and NAO performance into account, normalized NAO skillscores are added to the normalized spatiotemporal skillscores to form a combined objective skill estimate. The objective skill estimates are then normalized to form the BMA weights for the combined approach (BMA-COM). An example of the weightings calculations for the BMA-NAO and BMA-COM approaches is given in Table 7.5.

Winter temperature 2071-2100

Model	HadRM3P-a	RCAO-H	HIRHAM-E5	ARPEGE-a	RCAO-E4
Weighting information: BMA-SS					
Spatial r	0.53	0.88	0.55	0.57	0.88
Temporal r	1.00	1.00	1.00	0.99	0.99
Overall squared skill score	0.58	0.88	0.60	0.61	0.87
BMA-SS weights	0.16	0.25	0.17	0.17	0.25
Weighting information: BMA-NAO					
NAO skill score	0.33	0.66	0.99	0.66	0.99
NAO squared skill score	0.11	0.44	0.98	0.44	0.98
BMA-NAO weights	0.04	0.15	0.33	0.15	0.33
Weighting information: BMA-COM					
SS+NAO	0.20	0.40	0.50	0.32	0.58
BMA-COM weights	0.10	0.20	0.25	0.16	0.29

Table 7.5: Example of BMA-NAO and BMA-COM weightings using winter temperature data for the A2 scenario.

7.5 RESULTS: FUTURE CLIMATE PROJECTIONS USING DIFFERING WEIGHTING SCHEMES

7.5.1 Winter (DJF) temperature projections: A2 scenario

Figure 7.2 shows a deterministic AEM projection for winter temperature under the A2 emissions scenario for 2071-2100. In this approach, no information generated from the control period assessments are used to inform the projection and all subjective decisions about whether bias is correctable or how to weight models are avoided. However, there is no likelihood attached to the projection. Should a decision-maker choose the mean ensemble or a particular model for developing adaptation policies, any decision based on either of these selections is likely to lead to under- or over-adaptation. Additionally, none of the uncertainty attached to the projection is accounted for. In this case, it is clear that one model, RCAO-E4, projects a very different outcome from the other models, which in turn affects the AEM.

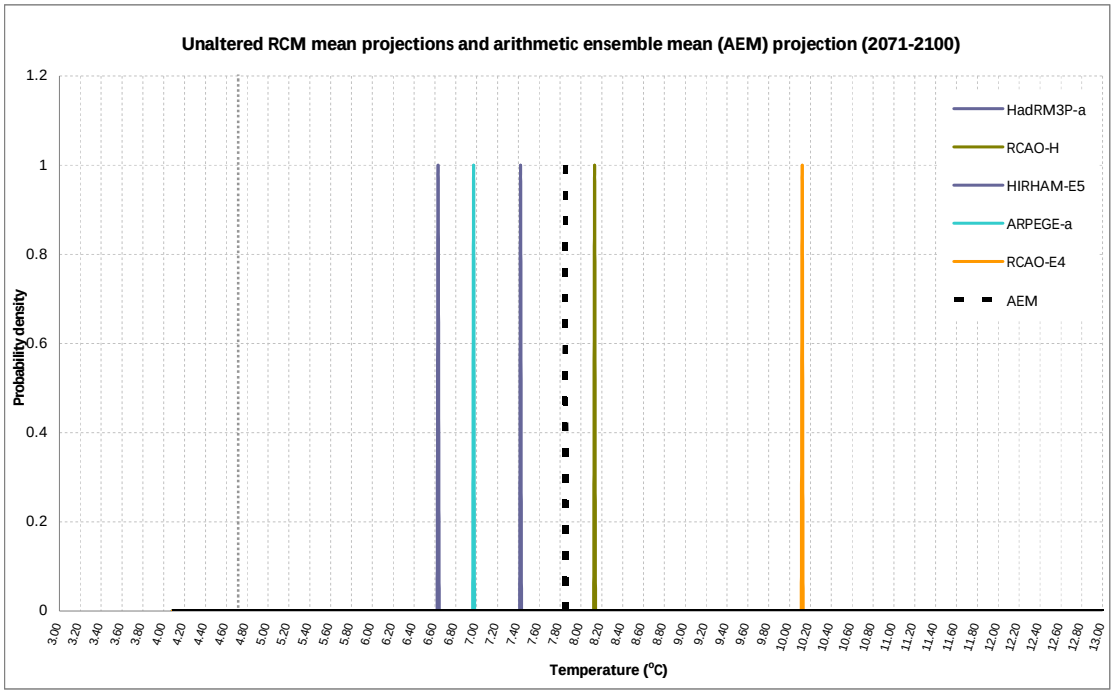
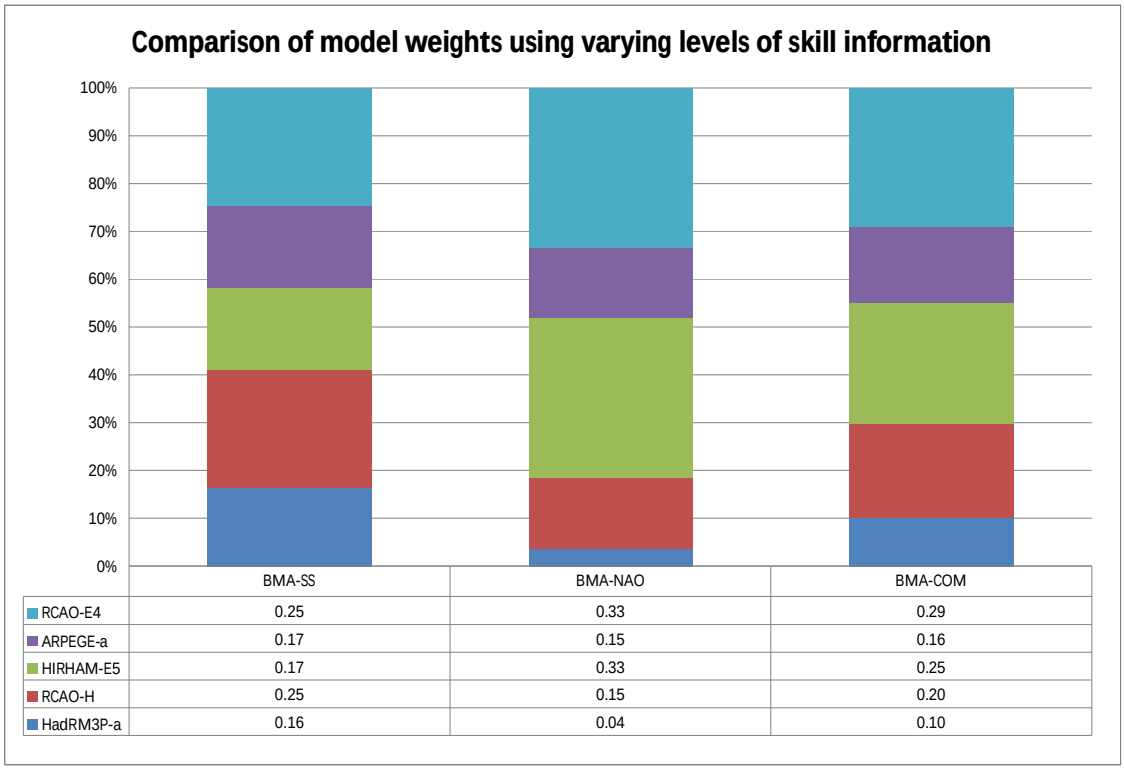


Figure 7.2: Unaltered RCM mean projections and arithmetic ensemble mean (AEM) projection based on these models for winter mean temperature under the A2 emissions scenario. Systematic bias has not been corrected and no skill information is used to construct the likelihood of individual projections. The grey line denotes observed winter mean temperature in the control period 1961-1990.

In the control period, RCAO-E4 exhibited a systematic bias and correcting for this in the future projections may result in a more robust outcome. Therefore, although the decision to correct or weight output is a subjective one, it may be justifiable to develop more reliable projections.

Figure 7.3 illustrates the contributions of the models under different weighting schemes. HIRHAM-E5 has a low spatiotemporal skill score, but a high NAO skill estimate. As such, in the BMA-SS projection, this model contributes less than in the BMA-NAO or BMA-COM projections. Conversely, RCAO-H scores highly in terms of its spatiotemporal skill score but is not as skilful at representing the NAO.



	BMA-SS	BMA-NAO	BMA-COM
RCAO-E4	0.25	0.33	0.29
ARPEGE-a	0.17	0.15	0.16
HIRHAM-E5	0.17	0.33	0.25
RCAO-H	0.25	0.15	0.20
HadRM3P-a	0.16	0.04	0.10

Figure 7.3: Comparison of model weights using varying levels of skill information for winter mean temperature under the A2 emissions scenario. Systematic bias is corrected and scores indicate model skill based on performance in the present day.

Figures 7.4 shows the BMA-EQ(a), BMA-SS(b), BMA-NAO(c) and BMA-COM(d) ensemble projections. For this parameter, season and emissions scenario, the shape of the ensemble PDF varies little when different weighting schemes are applied.

A2 winter temperature projections (2071-2100)

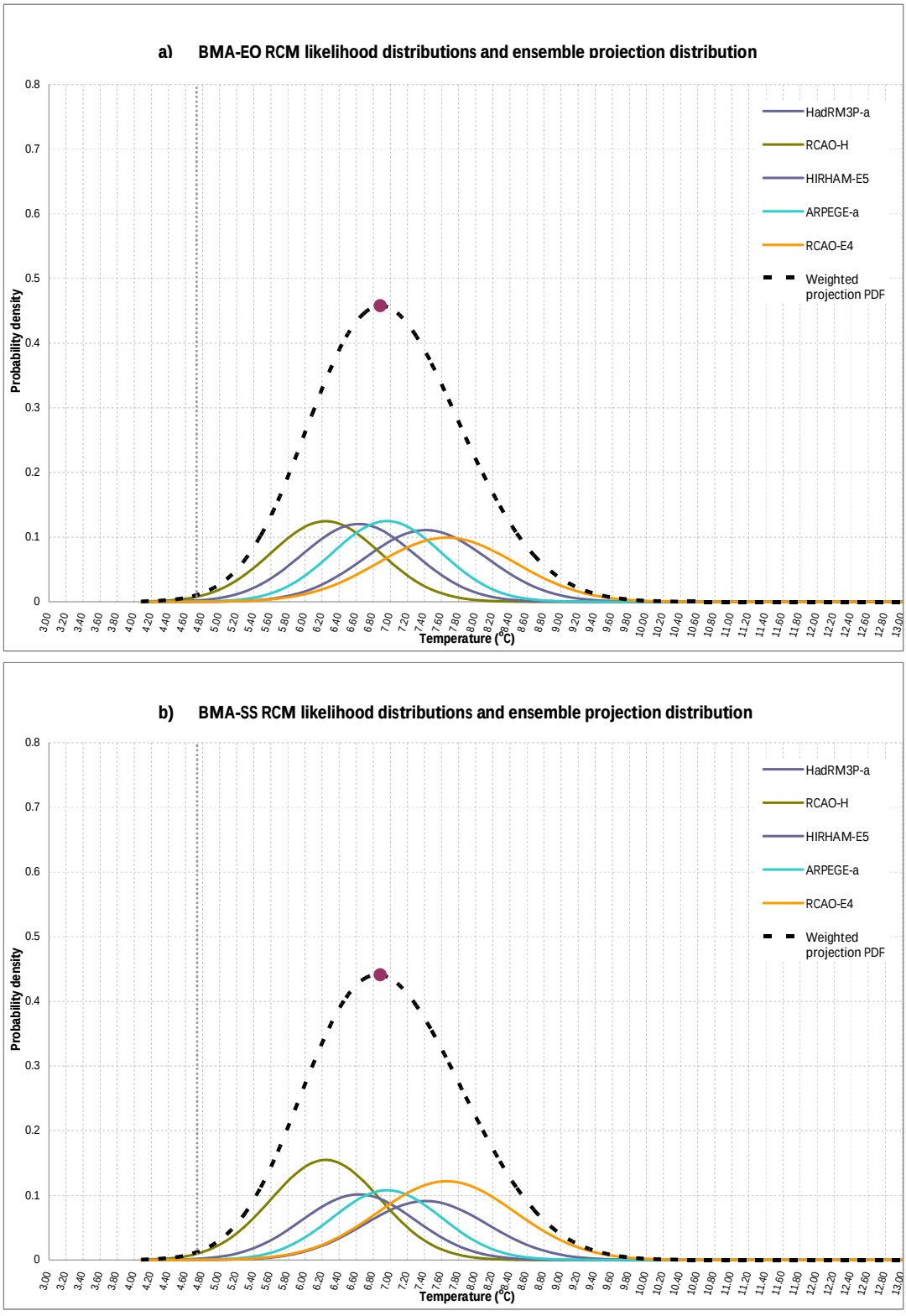
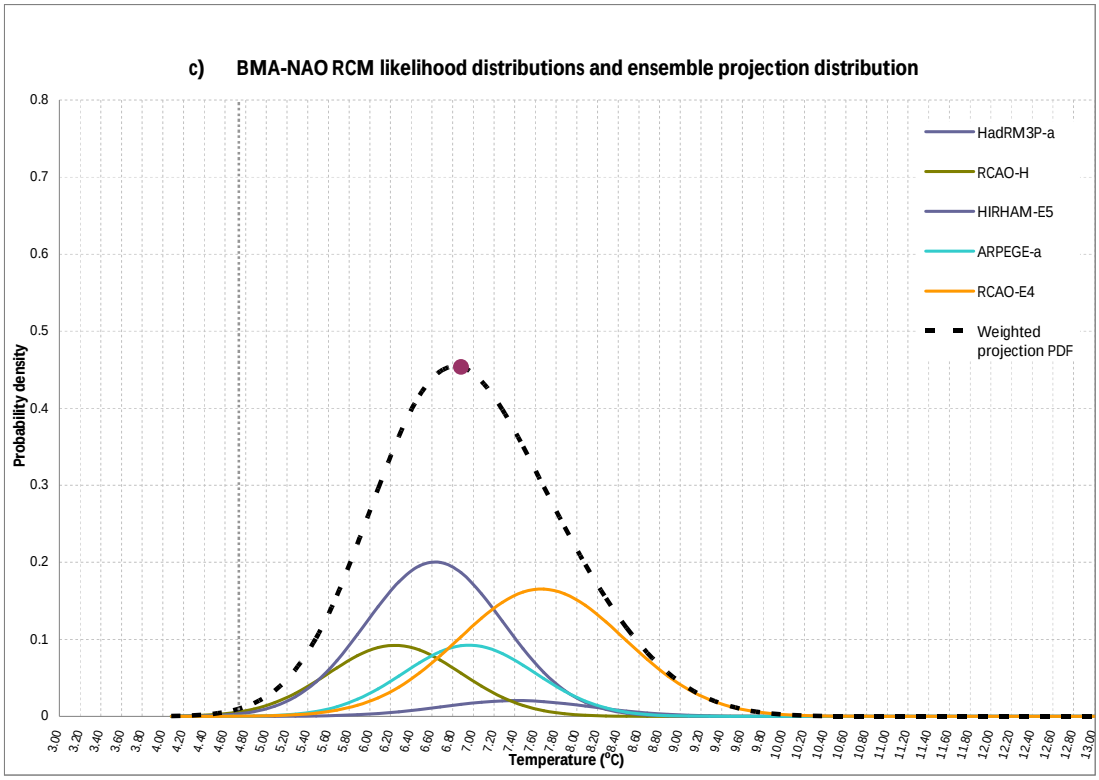


Figure 7.4: Weighted RCM likelihood distributions and weighted ensemble projection distribution for winter mean temperature under the A2 emissions scenario, using a) BMA-EQ weighting and b) BMA-SS weighting. Systematic bias is corrected. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.



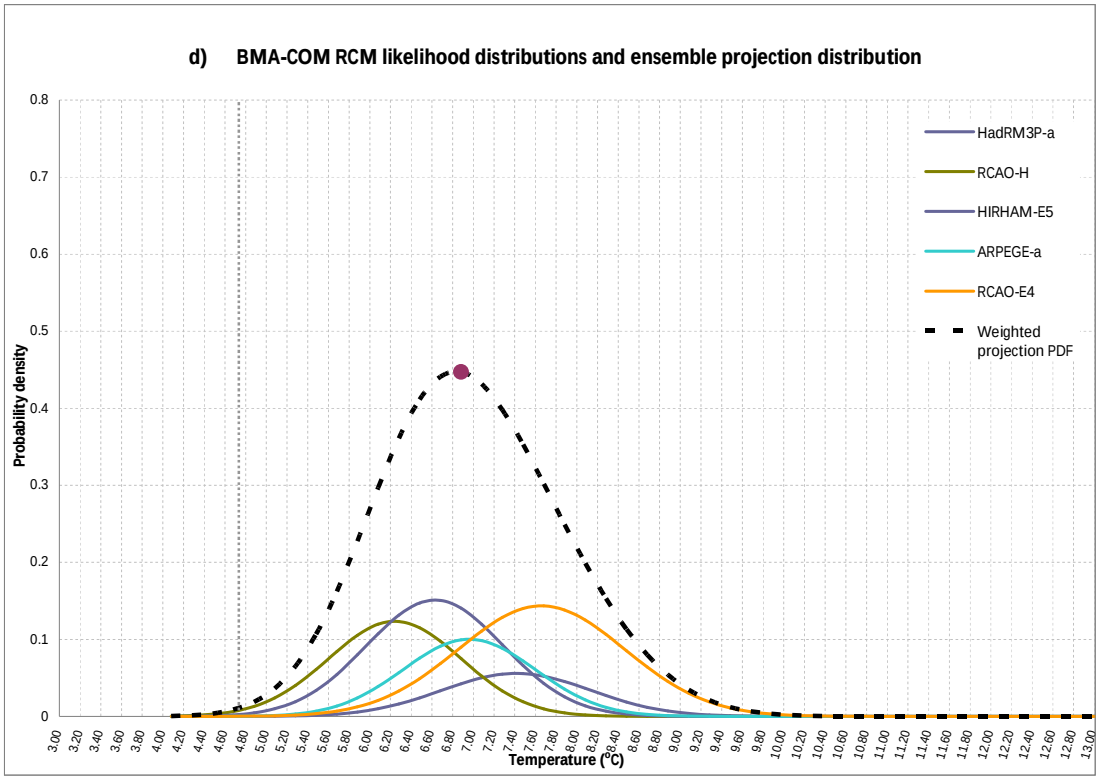


Figure 7.4 (continued): Weighted RCM likelihood distributions and weighted ensemble projection distribution for winter mean temperature under the A2 emissions scenario, using c) BMA-NAO weighting and d) BMA-COM weighting. Systematic bias is corrected. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

The most likely temperature value, represented by the red dot, falls between 6.8 and 7^{o}C in all cases. However, the data underlying the averages changes significantly when different weightings are used, with different models emerging as the most influential in each weighting scheme. For example, HIRHAM-E5 is more influential when BMA-NAO weighting is used compared with BMA-SS weighting. Conversely, HadRM3P-a has a higher weight under BMA-SS weighting than it does under BMA-NAO weighting (Figure 7.3). Due to these differences, the contribution of each model to the ensemble PDF varies under each weighting scheme. Though similar results can be obtained for a mean projection even when the underlying data varies, but if there is to be confidence in the mean projection, the underlying data must be assessed and combined according to the relative merits and deficiencies of the models.

7.5.2 Winter (DJF) temperature projections: B2 scenario

Figure 7.5 shows the AEM projection for winter temperature under the B2 emissions scenario. As no HIRHAM-E5 B2 data is available, only four simulations are used, changing the weights and contributions of each model. Again, RCAO-E4 has a significant effect on the AEM. Although a lack of convergence with other models is not a reason to disregard a model, when one projection has such a significant effect on the AEM, it is important to be as certain as is possible of the skill of that model. There is a definite argument for correcting the systematic bias RCAO-E4 appeared to exhibit in the control period. Figure 7.6 illustrates the contributions of the various models under different weighting schemes. RCAO-H has high spatiotemporal skill scores but is not as skilful as other models at representing the NAO. RCAO-E4 is very skilful at capturing the effects of the NAO and has high spatiotemporal skill scores and when the scores are combined, it is the most reliable and therefore the most influential model in the ensemble.

Figures 7.7 show the BMA-EQ(a), BMA-SS(b), BMA-NAO(c) and BMA-COM(d) ensemble projections. Systematic bias is corrected in all these projections. The choice of weighting has a large effect on the shape of the ensemble PDF. The four models appear to form two separate peaks, with RCAO-E4 and HadRM3P-a

both peaking at approximately 7.2 to 7.6°C while RCAO-H and ARPEGE-a have most likely values of approximately 5.7°C and 6.3°C respectively.

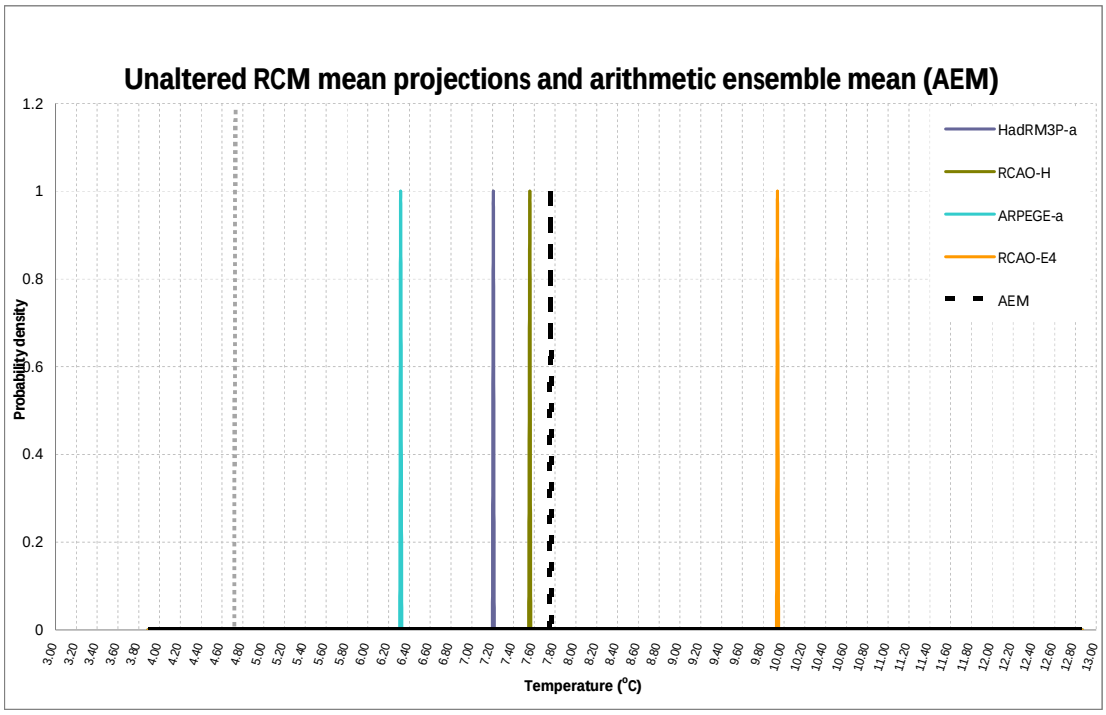
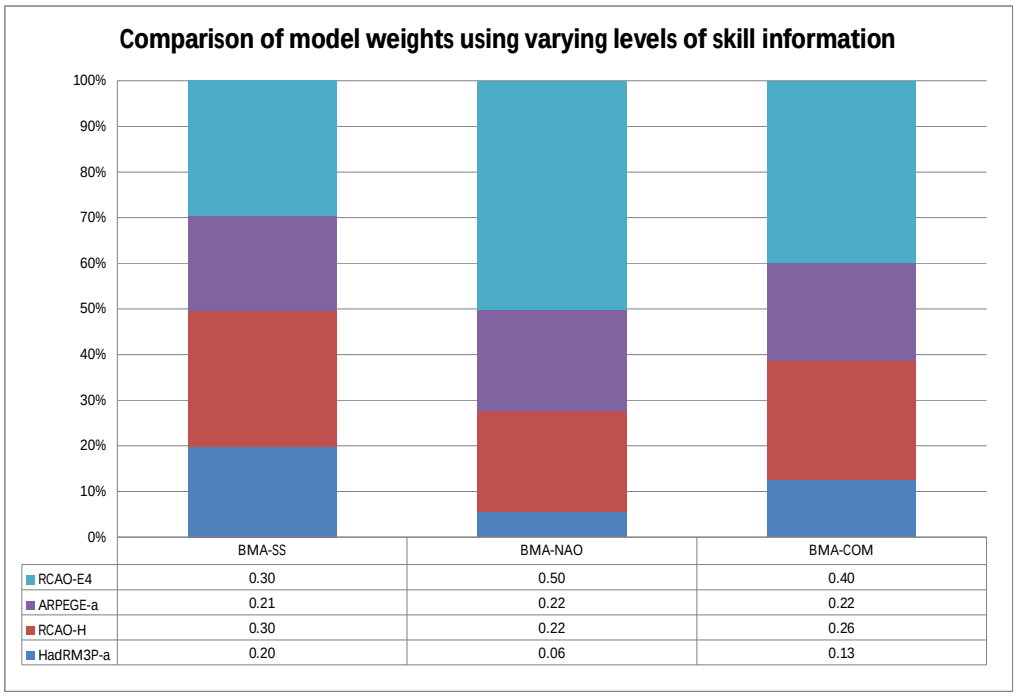


Figure 7.5: Unaltered RCM mean projections and arithmetic ensemble mean (AEM) projection based on these models for winter mean temperature under the B2 emissions. Systematic bias has not been corrected and no skill information is used to construct the likelihood of individual projections. The grey line denotes observed winter mean temperature in the control period 1961-1990.



	BMA-SS	BMA-NAO	BMA-COM
RCAO-E4	0.30	0.50	0.40
ARPEGE-a	0.21	0.22	0.22
RCAO-H	0.30	0.22	0.26
HadRM3P-a	0.20	0.06	0.13

Figure 7.6: Comparison of model weights using varying levels of skill information for winter mean temperature under the B2 emissions scenario. Systematic bias is corrected and scores indicate model skill based on performance in the present day.

B2 winter temperature projections (2071-2100)

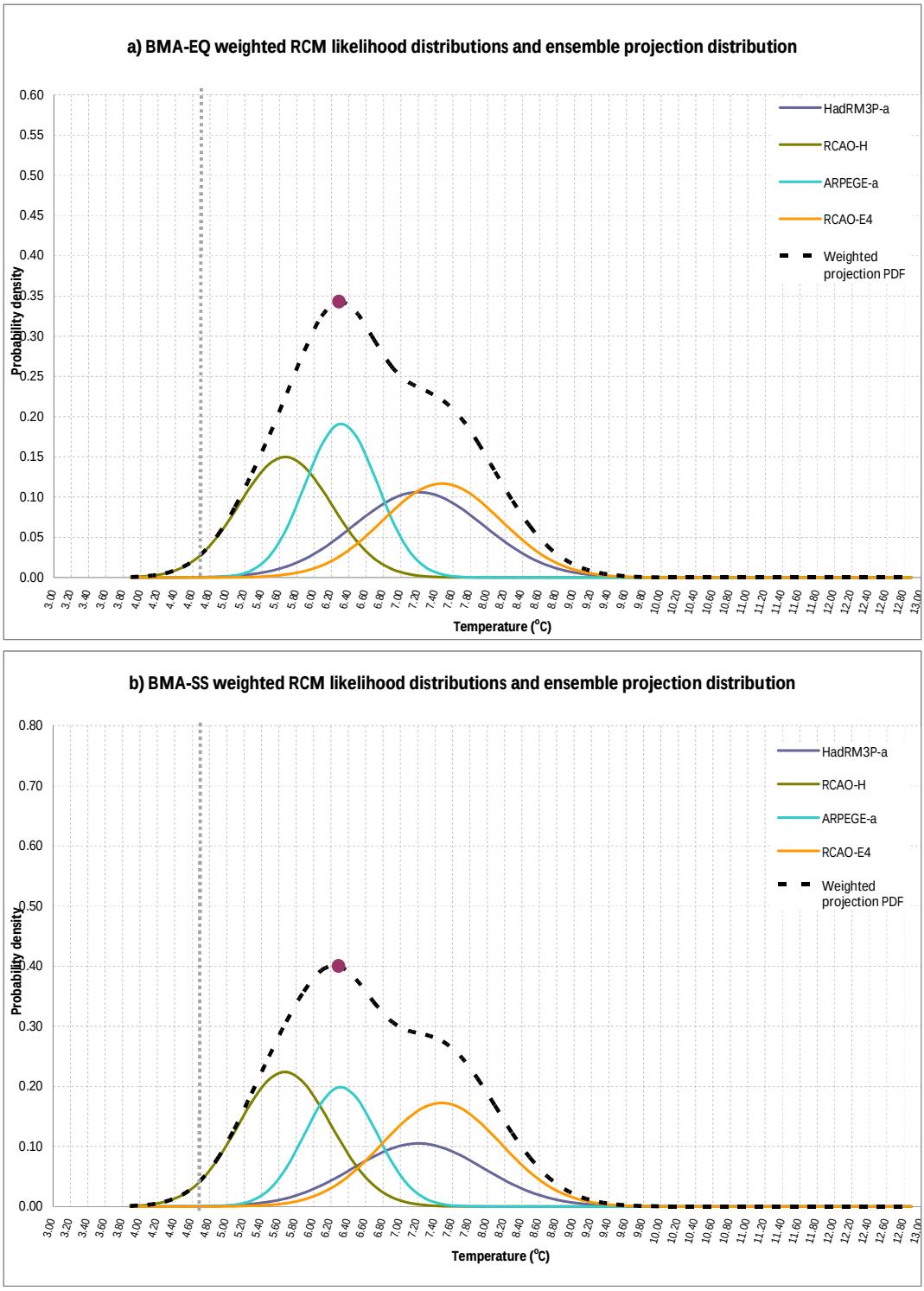
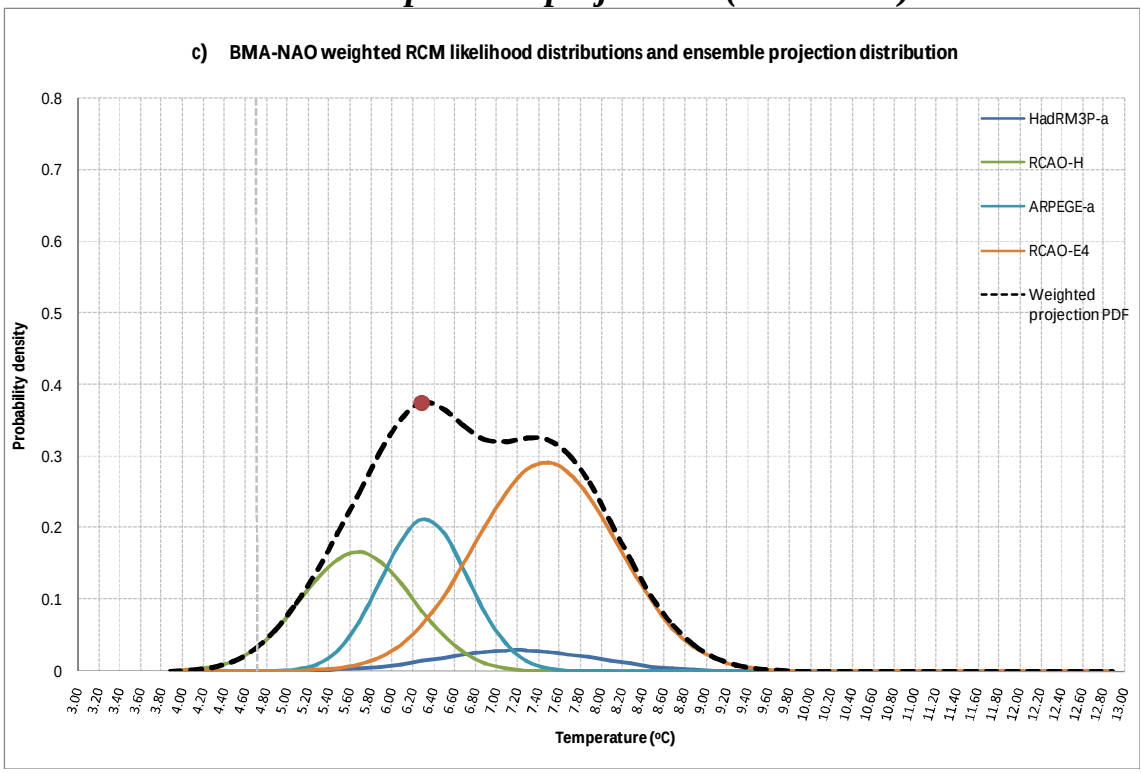


Figure 7.7: Weighted RCM likelihood distributions and weighted ensemble projection distribution for winter mean temperature under the B2 emissions scenario, using a) BMA-EQ weighting and b) BMA-SS weighting. Systematic bias is corrected. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

B2 winter temperature projections (2071-2100)



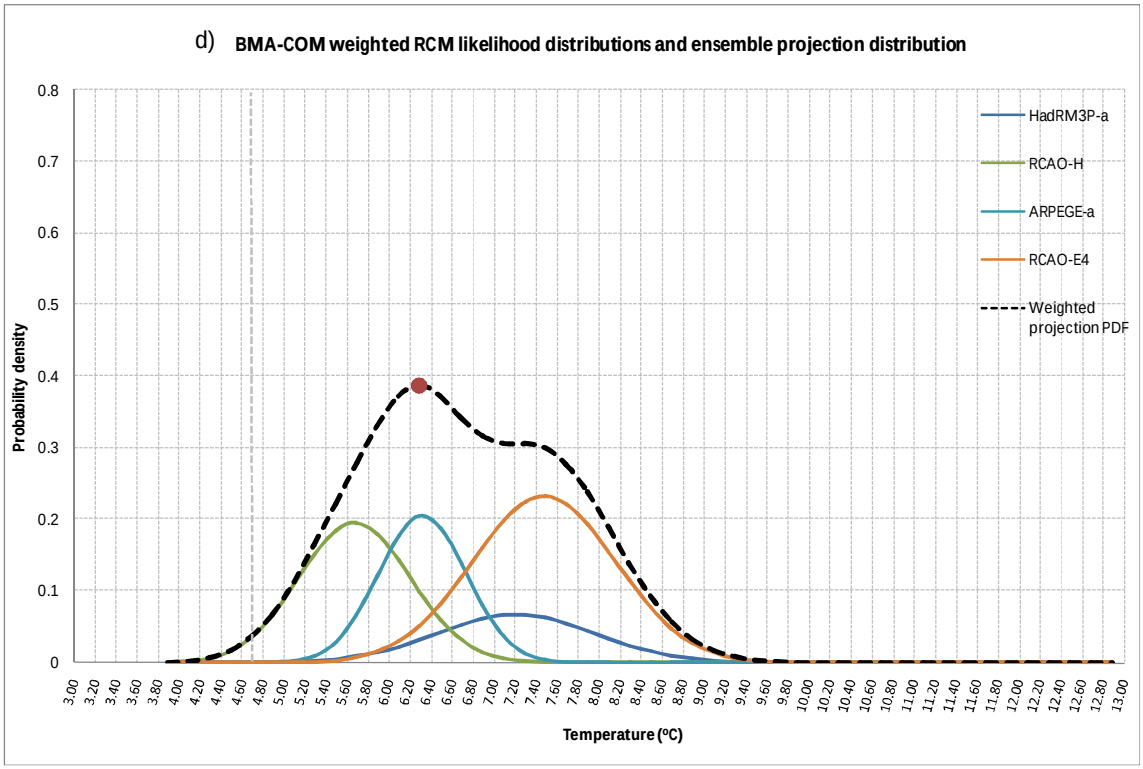


Figure 7.7 (continued): Weighted RCM likelihood distributions and weighted ensemble projection distribution for winter mean temperature under the A2 emissions scenario, using c) BMA-NAO weighting and d) BMA-COM weighting. Systematic bias is corrected. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

When these models are combined, the resulting distribution is bimodal, but the degree of bimodality depends on the weighting scheme applied. For example, when all models are weighted equally (Figure 7.7a), the influence of RCAO-E4 is constrained by the equal inclusion of the other models and as such, the peak at 7.5^{o}C is not very well developed. However, when weights are introduced, the high scores of RCAO-E4 in both spatiotemporal metrics and NAO representation (Figure 7.6) make it a more influential model and the ensemble PDF takes on a much more pronounced bimodal shape, which is most apparent using the BMA-NAO approach (Figure 7.7c). For this parameter, season and emissions scenario, the shape of the ensemble PDF and the conclusions that might be drawn from it vary significantly when different weighting schemes are applied, making it especially important that the weightings chosen are genuinely representative of the predictive skill of the model and do not occur because of error cancellation.

7.5.3 Summer (JJA) mean temperature: A2 emissions scenario

Figure 7.8 shows a deterministic AEM projection for summer temperature under the A2 emissions scenario. For this parameter, season and emissions scenario, the AEM, represented by the black dashed line, is not overly influenced by any one model. There is a cluster of models around 16.4 to 17.2^{o}C and one model each side of that cluster that could be considered an outlier. However, as there is a diverging model on both sides, the projections converge towards a central value. Of course, the deterministic approach reflects little information about the potential range of projections, which is why information about the likelihoods associated with the individual models needs to be included in the ensemble projection.

Figure 7.9 illustrates the contributions of the various models under different weighting schemes. Skill at modelling the summer spatial pattern for temperature is now incorporated into the weighting, which alters the BMA-SS and BMA-COM weights. RCAO-E4 and RCAO-H are less skilful at simulating the summer spatial pattern than at simulating the winter spatial pattern and so have less influence in the calculation of summer projections. Figure 7.10 shows the BMA-EQ(a), BMA-SS(b), BMA-NAO(c) and BMA-COM(d) ensemble projections.

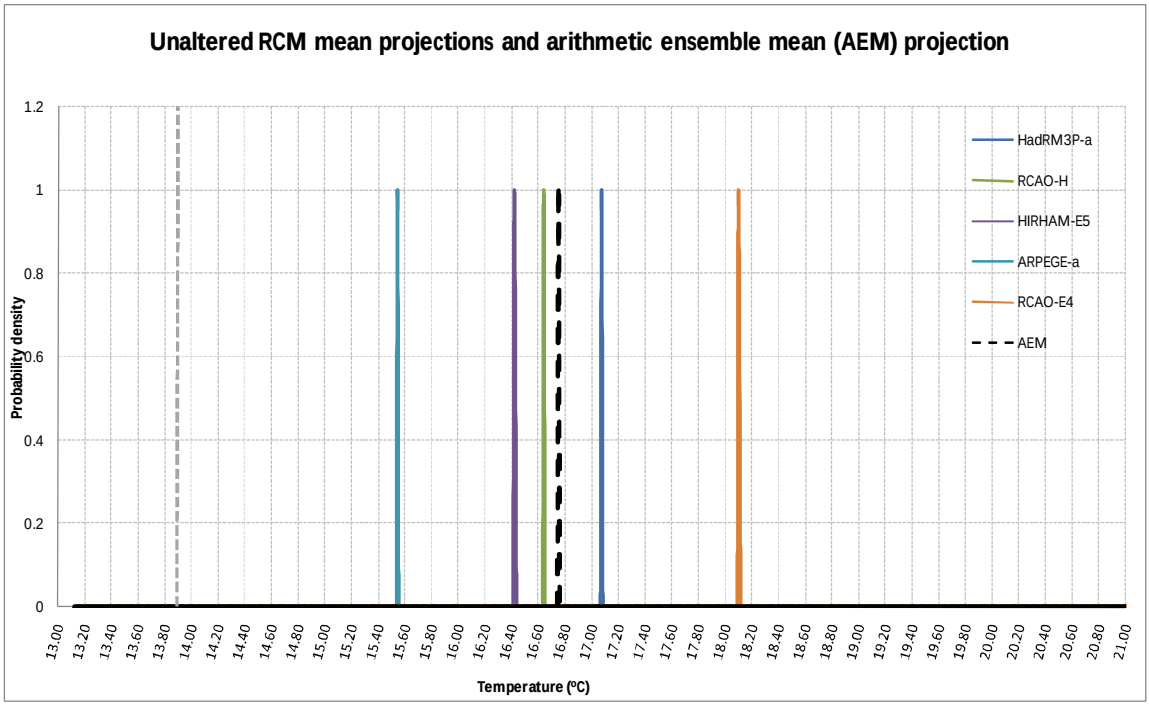
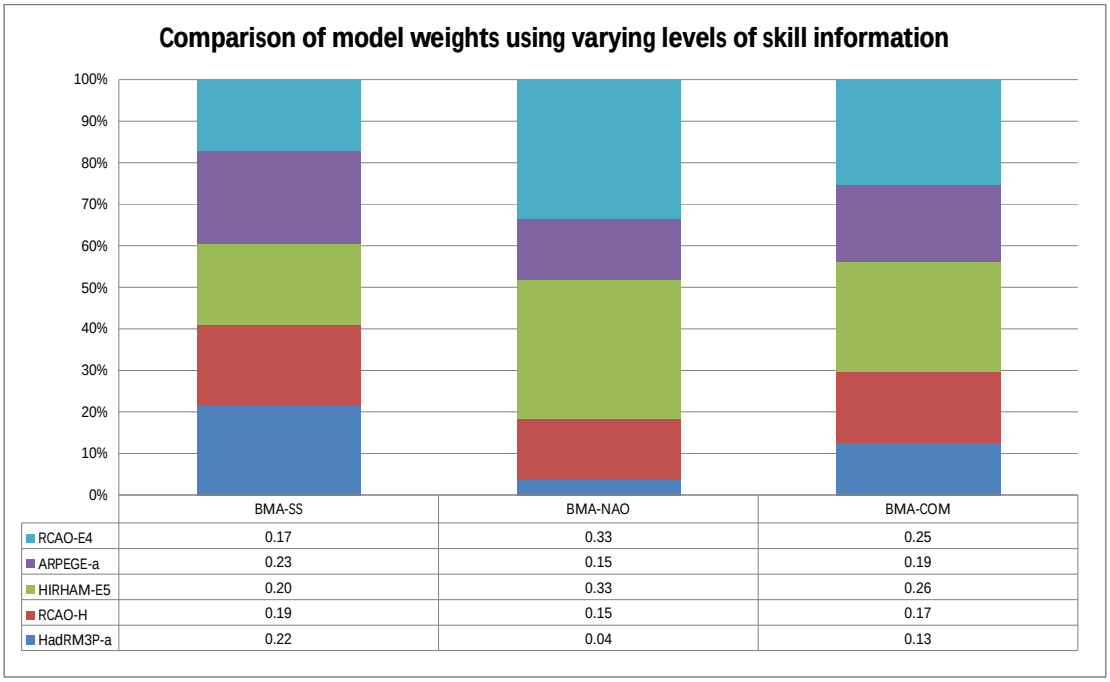


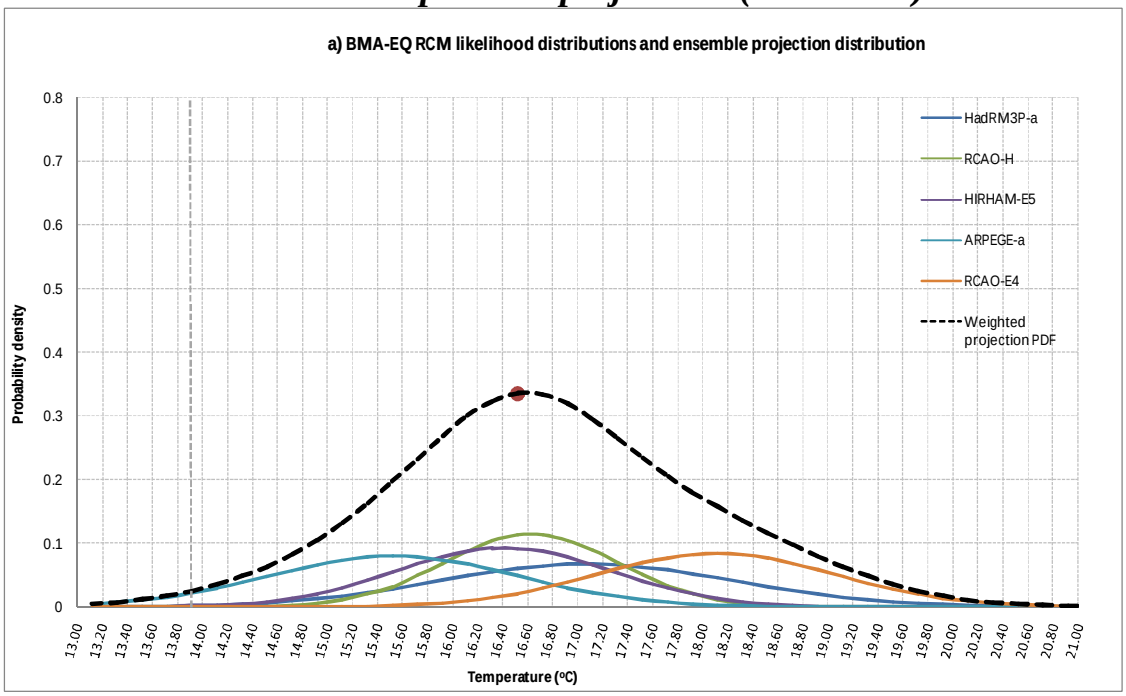
Figure 7.8: Unaltered RCM mean projections and arithmetic ensemble mean (AEM) projection based on these models for summer mean temperature under the A2 emissions scenario. No skill information is used to construct the likelihood of individual projections. The grey line denotes observed winter mean temperature in the control period 1961-1990.



	BMA-SS	BMA-NAO	BMA-COM
RCAO-E4	0.17	0.33	0.25
ARPEGE-a	0.23	0.15	0.19
HIRHAM-E5	0.20	0.33	0.26
RCAO-H	0.19	0.15	0.17
HadRM3P-a	0.22	0.04	0.13

Figure 7.9: Comparison of model weights using varying levels of skill information for summer mean temperature under the A2 emissions scenario. Systematic bias is corrected and scores indicate model skill based on performance in the present day.

A2 summer temperature projections (2071-2100)



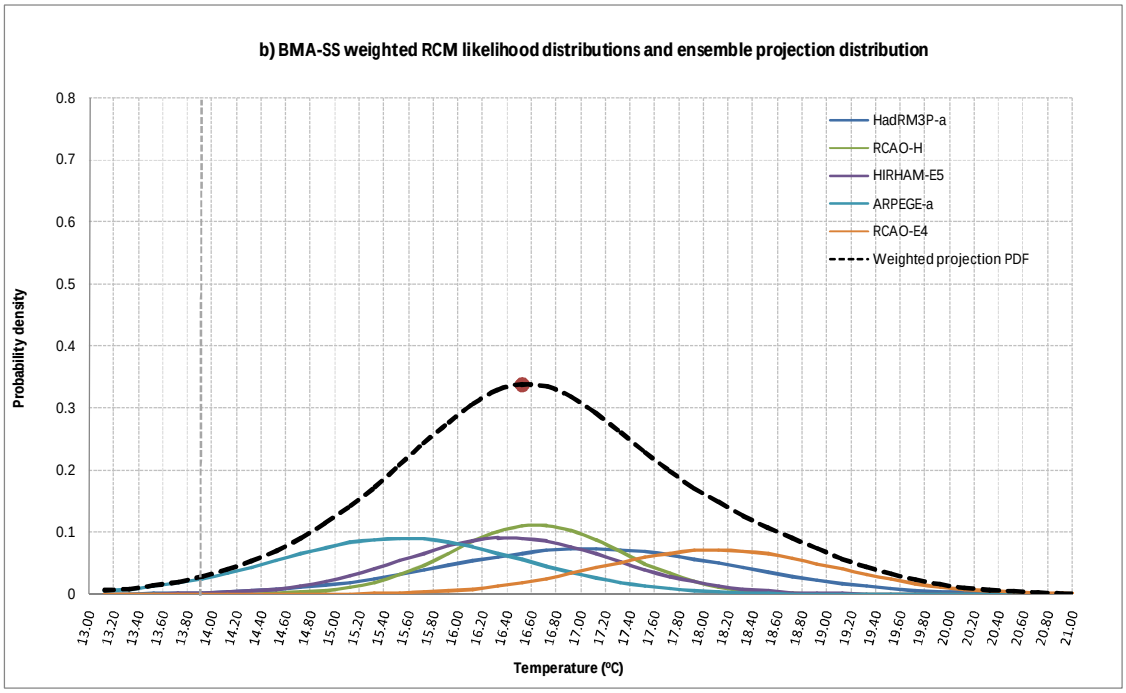


Figure 7.10: Weighted RCM likelihood distributions and weighted ensemble projection distribution for summer mean temperature under the A2 emissions scenario, using a) BMA-EQ weighting and b) BMA-SS weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

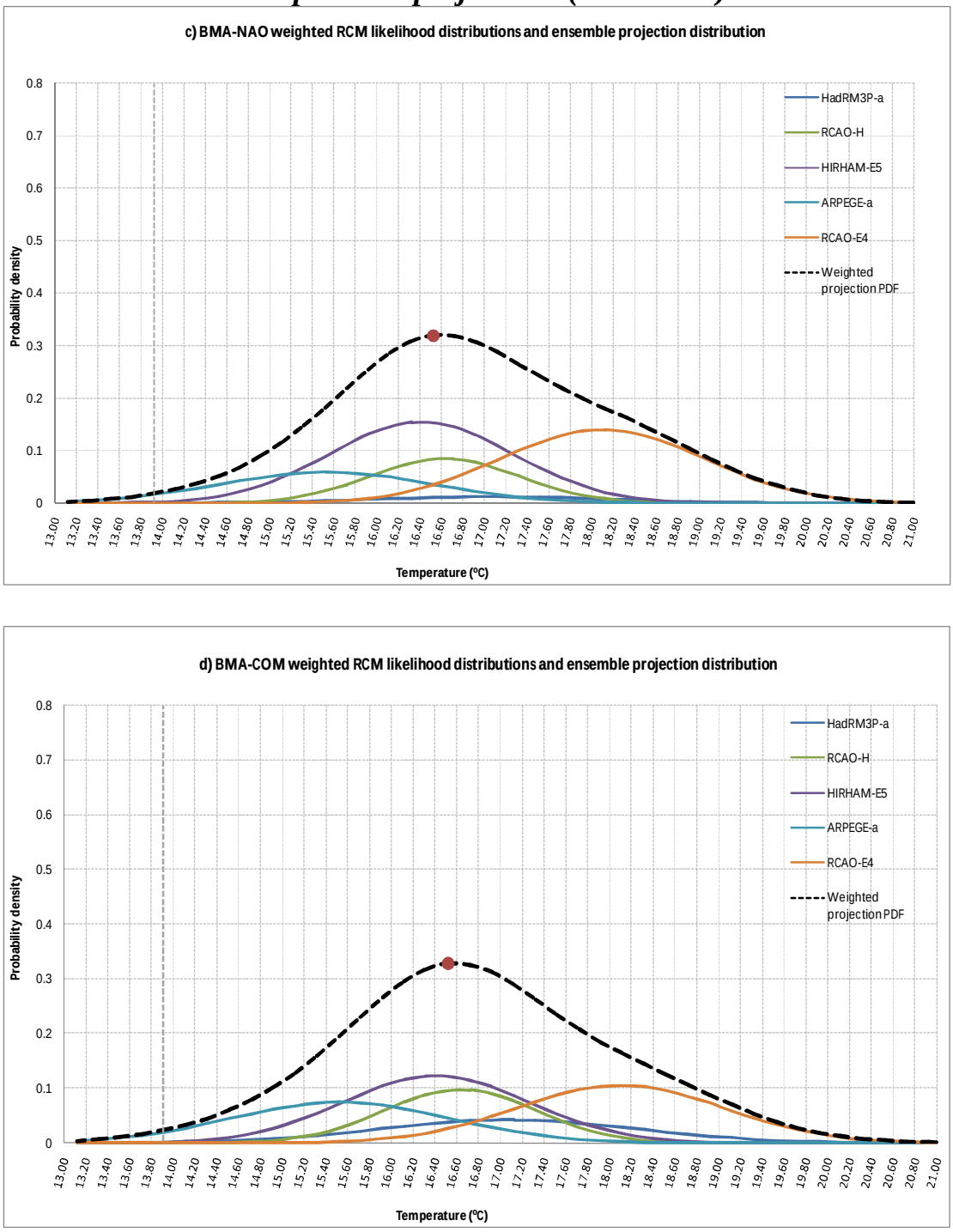


Figure 7.10 (continued): Weighted RCM likelihood distributions and weighted ensemble projection distribution for summer mean temperature under the A2 emissions scenario, using c) BMA-NAO weighting and d) BMA-COM weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

Systematic bias is corrected in all these projections. As for the winter A2 projections, the distribution PDF has a single peak. The tails of the distribution PDF are quite long, ranging from 13 to 21^{o}C with a most likely projection of approximately 16.5^{o}C regardless of the weighting system used. However, an interesting feature is that the heaviness of the upper tail varies depending on the weighting system used. Under BMA-SS weighting, the upper tail is thinner (Figure 7.10b), suggesting that these higher projections for temperature have a low likelihood associated with them, yet weighting based on skill scores alone has the potential to be quite unreliable. Under BMA-NAO weighting, the influence of RCAO-E4 is greater (Figure 7.9), contributing to a significantly heavier tail (Figure 7.10c). This means that higher levels of probability are attached to the upper extremes of the ensemble PDF. An increase in summer temperatures under climate change is likely to have a range of impacts for areas such as water resource management and health, with the extent of the impacts depending largely on the degree of temperature change. Therefore it is important that the likelihoods associated with summer temperature projections are robust, particularly on the upper tail of the distribution.

7.5.4 Summer (JJA) mean temperature: B2 emissions scenario

Figure 7.11 shows a deterministic AEM projection for summer temperature under the B2 emissions scenario. Similar to the A2 summer projection, the AEM is not particularly influenced by any one model and the individual model projections converge towards a central value, however there is no information contained in such a projection about the likelihood associated with the outcome. Figure 7.12 illustrates the contributions of the various models under different weighting schemes. There is a significant difference between the weights associated with RCAO-E4 under the BMA-SS and BMA-NAO weighting schemes. As HIRHAM-E5 is not available with B2 forcing, RCAO-E4 becomes the most skilful model for representing the NAO. However, this model is not as skilful when it is scored based on spatiotemporal metrics. The varying weights have significant effects on the probabilistic climate projections. Figure 7.13 shows the BMA-EQ, BMA-SS, BMA-NAO and BMA-COM ensemble projections. Under this emissions scenario, the choice of weighting has a large effect on the shade of the ensemble PDF.

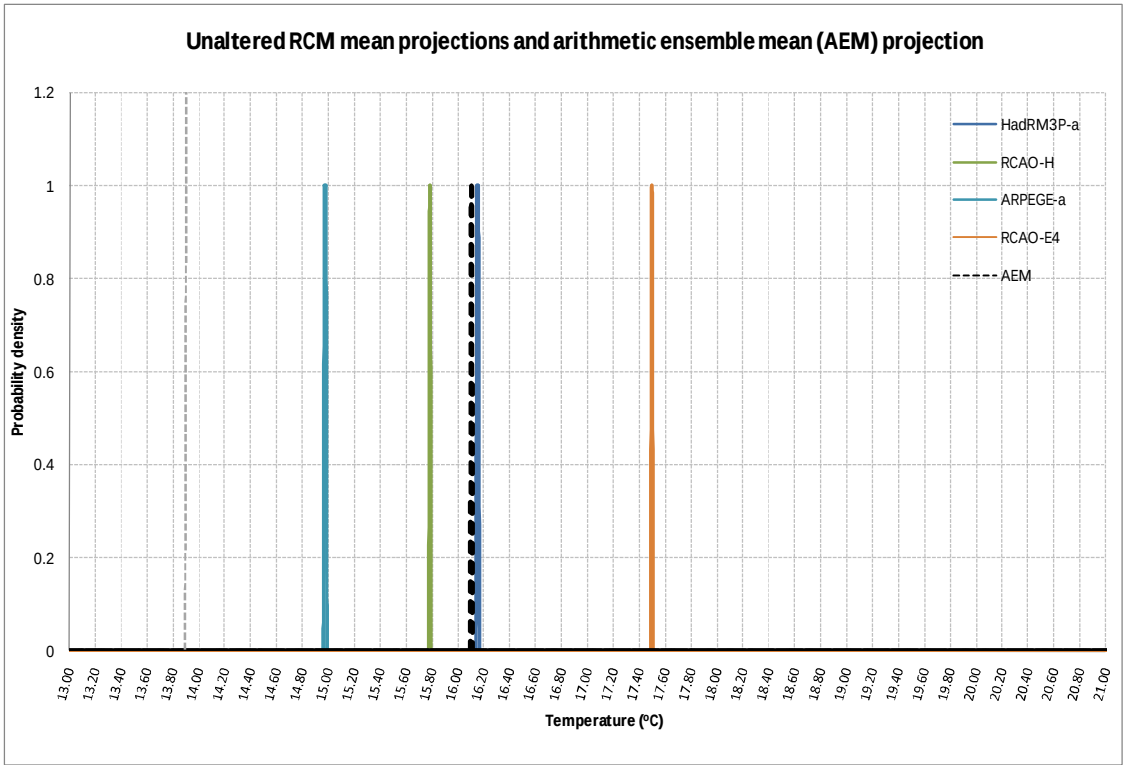
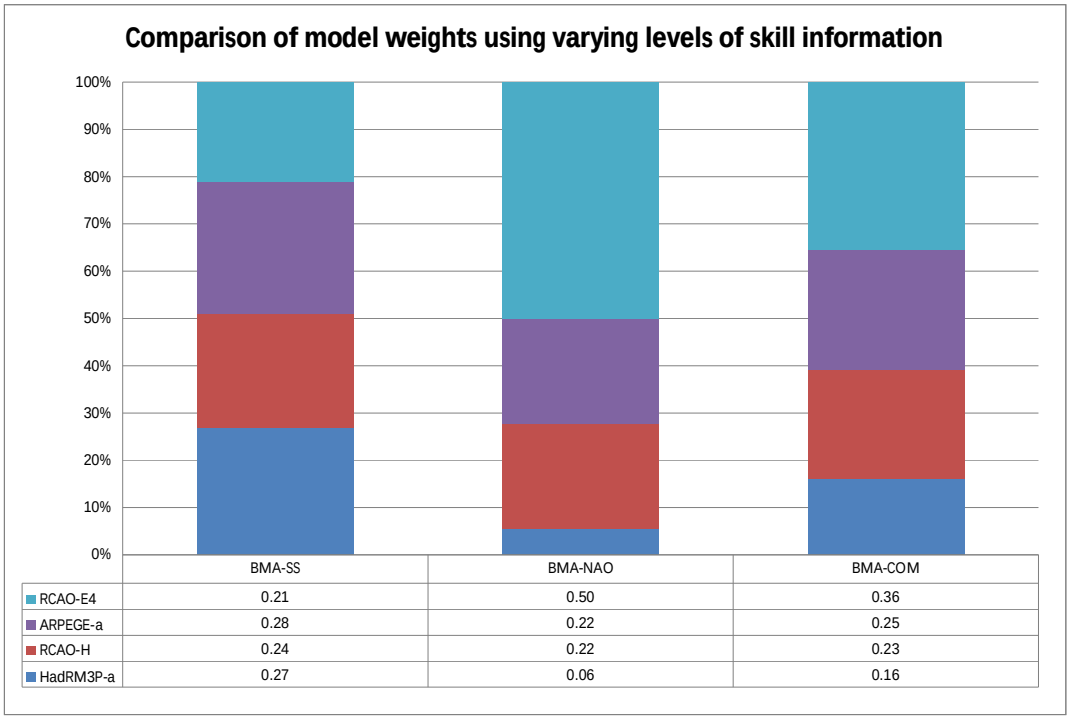


Figure 7.11: Unaltered RCM mean projections and arithmetic ensemble mean (AEM) projection based on these models for summer mean temperature under the B2 emissions scenario. No skill information is used to construct the likelihood of individual projections. The grey line denotes observed winter mean temperature in the control period 1961-1990.



	BMA-SS	BMA-NAO	BMA-COM
RCAO-E4	0.21	0.50	0.36
ARPEGE-a	0.28	0.22	0.25
RCAO-H	0.24	0.22	0.23
HadRM3P-a	0.27	0.06	0.16

Figure 7.12: Comparison of model weights using varying levels of skill information for summer mean temperature under the B2 emissions scenario. Systematic bias is corrected and scores indicate model skill based on performance in the present day.

B2 summer temperature projections (2071-2100)

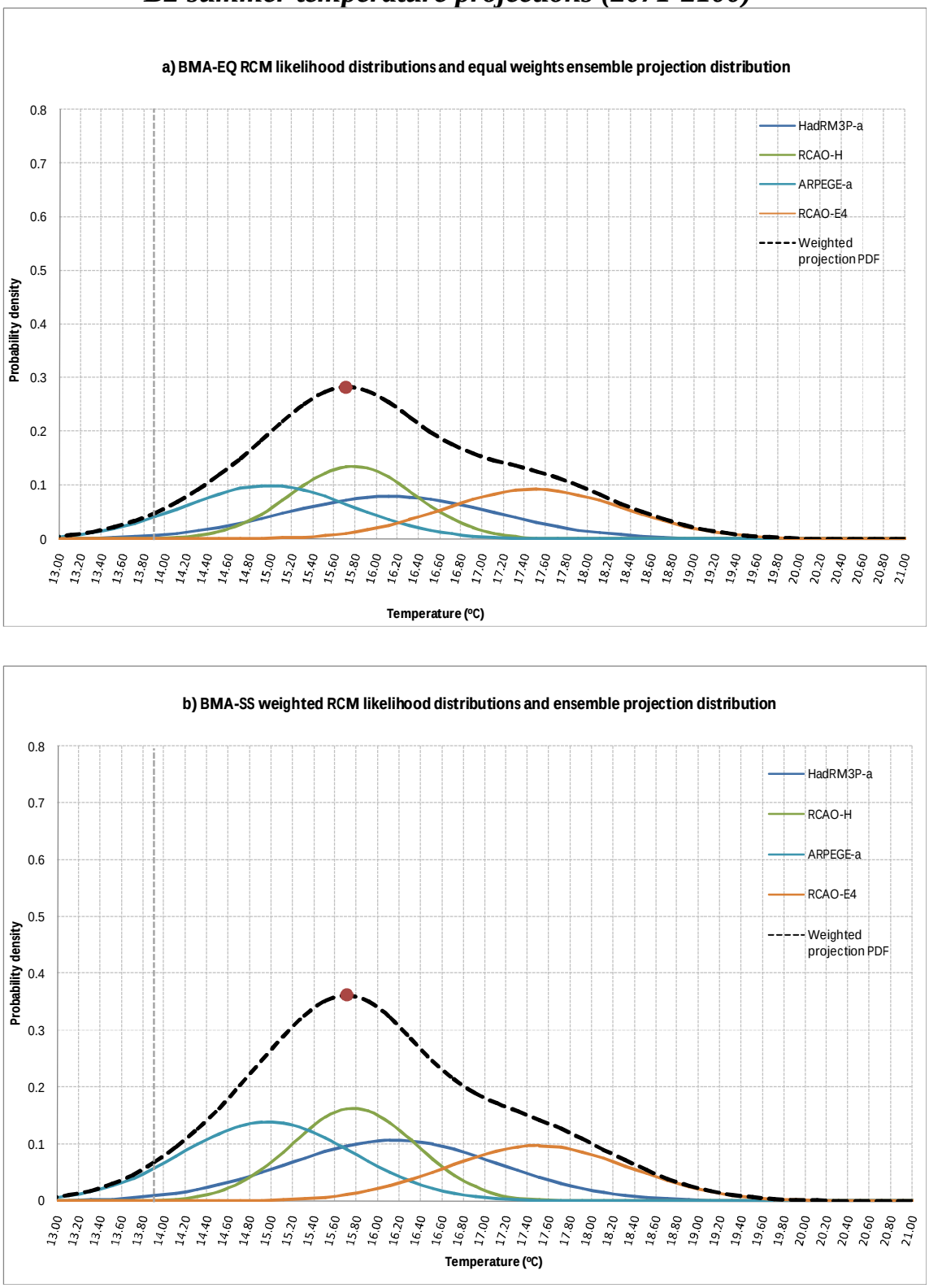
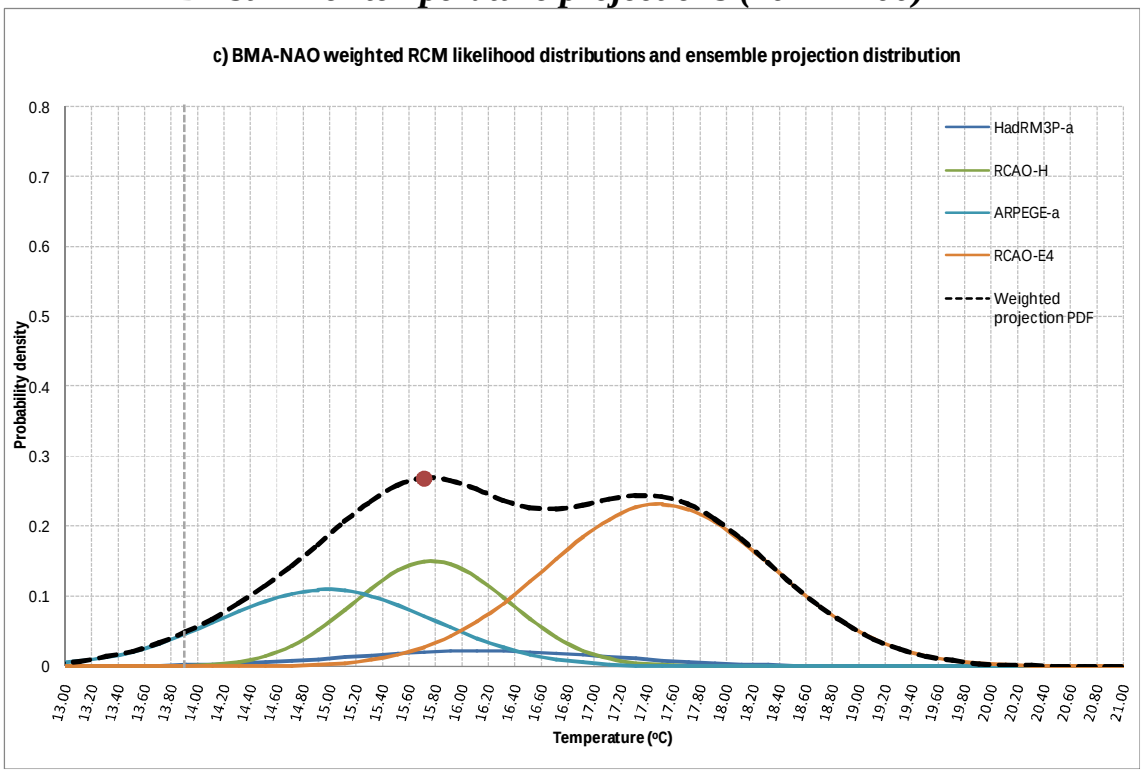


Figure 7.13: Weighted RCM likelihood distributions and weighted ensemble projection distribution for summer mean temperature under the B2 emissions scenario, using a) BMA-EQ weighting and b) BMA-SS weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

B2 summer temperature projections (2071-2100)



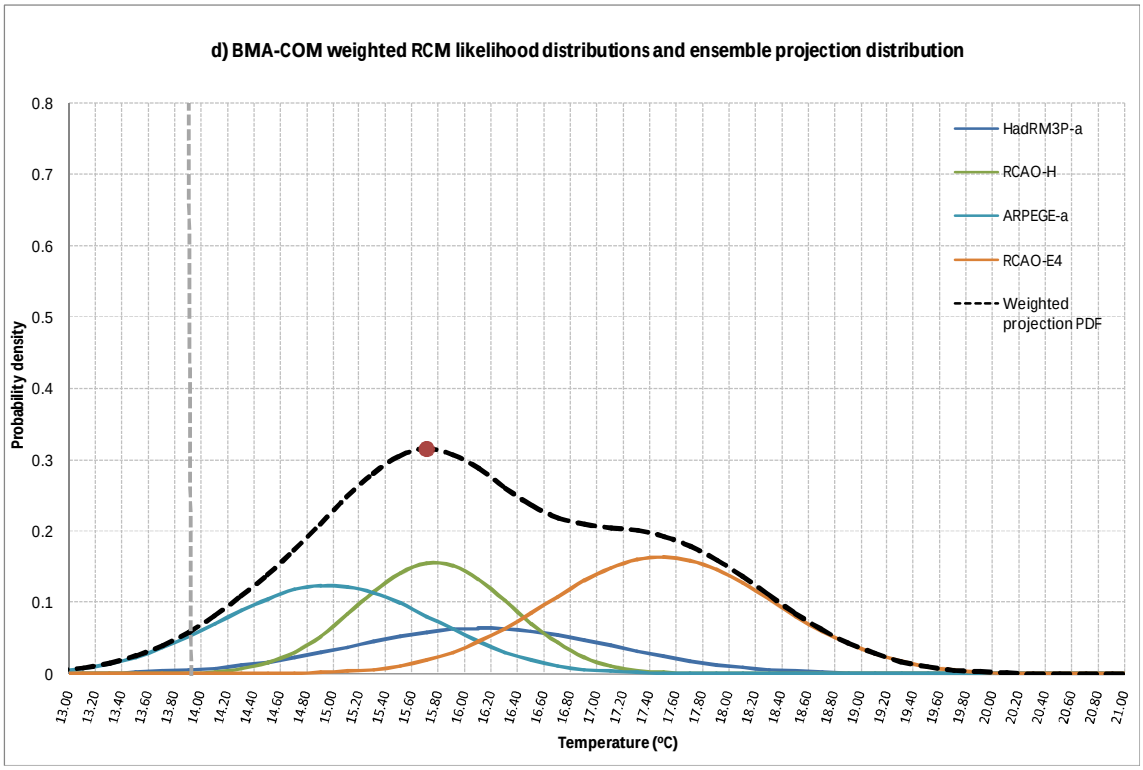


Figure 7.13 (continued): Weighted RCM likelihood distributions and weighted ensemble projection distribution for summer mean temperature under the B2 emissions scenario, using c) BMA-NAO weighting and d) BMA-COM weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

When equal weights or skill-score-based weights are applied, the ensemble PDF has a heavy upper tail and a most likely value for future temperature of 15.7^{o}C. This value is slightly lower than the AEM of 16.1^{o}C. However, when BMA-NAO or BMA-COM weighting is applied, the contribution of RCAO-E4 becomes much greater and the influence of other models which are less skilful at representing the NAO is constrained (Figure 7.12). As a result, the weighted ensemble PDF becomes bimodal, with peaks at 15.7^{o}C and 17.6^{o}C. The bimodal PDF is most pronounced for BMA-NAO weighting (Figure 7.13c) and less pronounced using BMA-COM weighting (Figure 7.13d).

It appears that under the B2 forcing scenario, the shape of the temperature ensemble PDF in both winter and summer is significantly influenced by the choice of weighting scheme. Therefore it is vital that the weightings chosen are genuinely representative of the predictive skill of the model. There is a clear argument for choosing the BMA-COM approach as more information about model skill in the present day is incorporated into this weighting scheme that the others.

7.5.5 Winter (DJF) precipitation projections: A2 scenario

Figure 7.14 gives the deterministic AEM projection for winter precipitation under the A2 emissions scenario. As with previous deterministic projections, the range of possible outcomes is not reflected. Figure 7.15 illustrates the contributions of the models under different weighting schemes. When only skill scores are considered, the models appear to have similar levels of skill, with the exception of ARPEGE-a. When skill at representing the NAO is also taken into account, the weightings vary more.

Figure 7.16 shows the BMA-EQ(a), BMA-SS(b), BMA-NAO(c) and BMA-COM(d) ensemble projections. HadRM3P-a, RCAO-H and RCAO-E4 all receive a systematic bias correction as their spatial r scores in the present day were 0.7 or above and their bias was greater than 10% of observed winter precipitation.

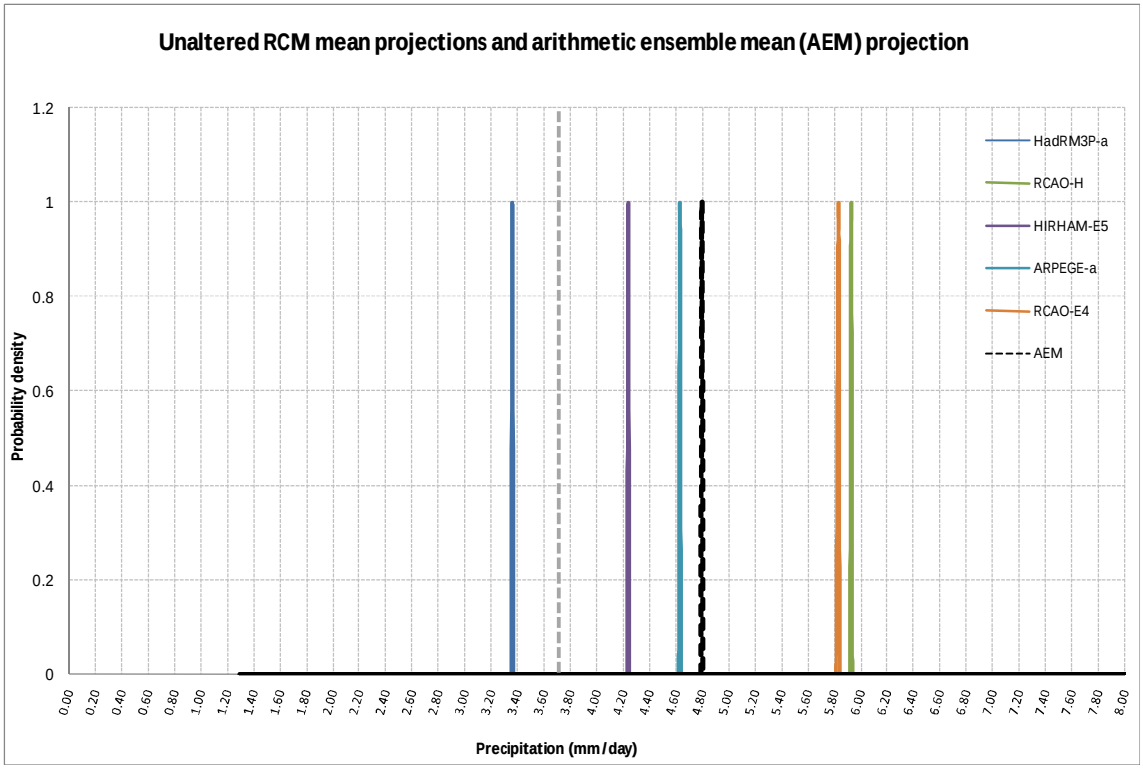
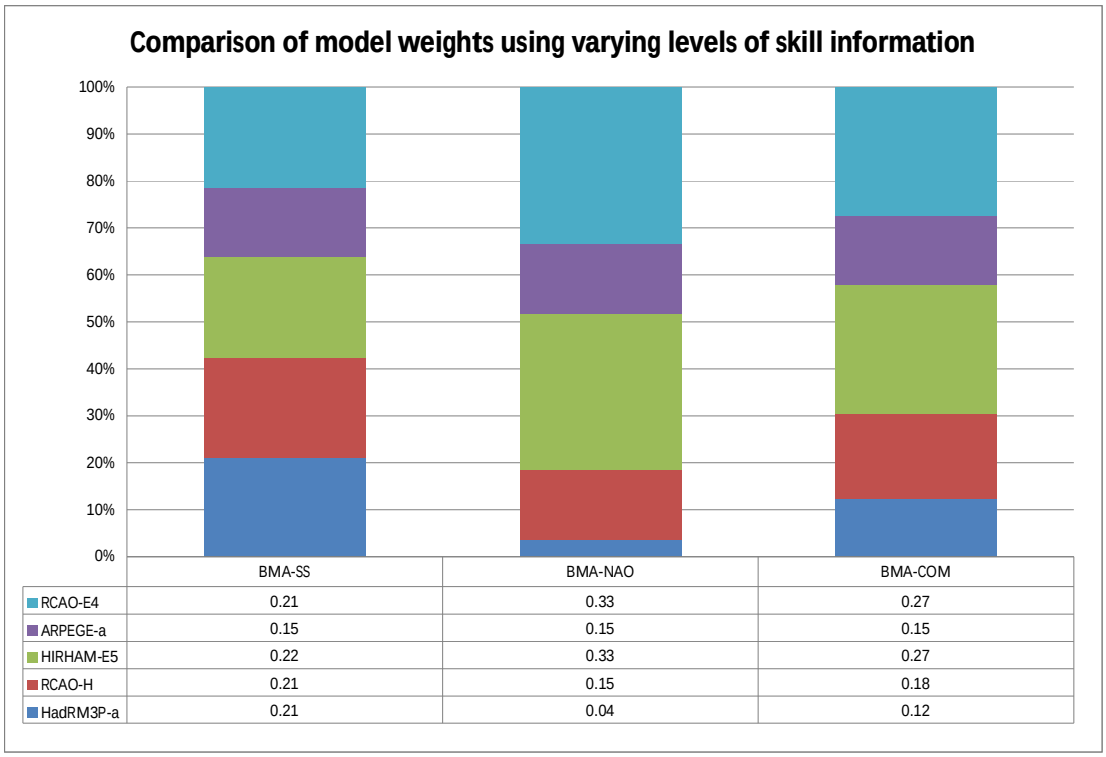


Figure 7.14: Unaltered RCM mean projections and arithmetic ensemble mean (AEM) projection based on these models for winter mean precipitation under the A2 emissions scenario. Systematic bias has not been correction and no information is used to construct the likelihood of individual projections.



	BMA-SS	BMA-NAO	BMA-COM
RCAO-E4	0.21	0.33	0.27
ARPEGE-a	0.15	0.15	0.15
HIRHAM-E5	0.22	0.33	0.27
RCAO-H	0.21	0.15	0.18
HadRM3P-a	0.21	0.04	0.12

Figure 7.15: Comparison of model weights using varying levels of skill information for winter mean precipitation under the A2 emissions scenario. Systematic bias is corrected and scores indicate model skill based on performance in the present day.

A2 winter precipitation projections (2071-2100)

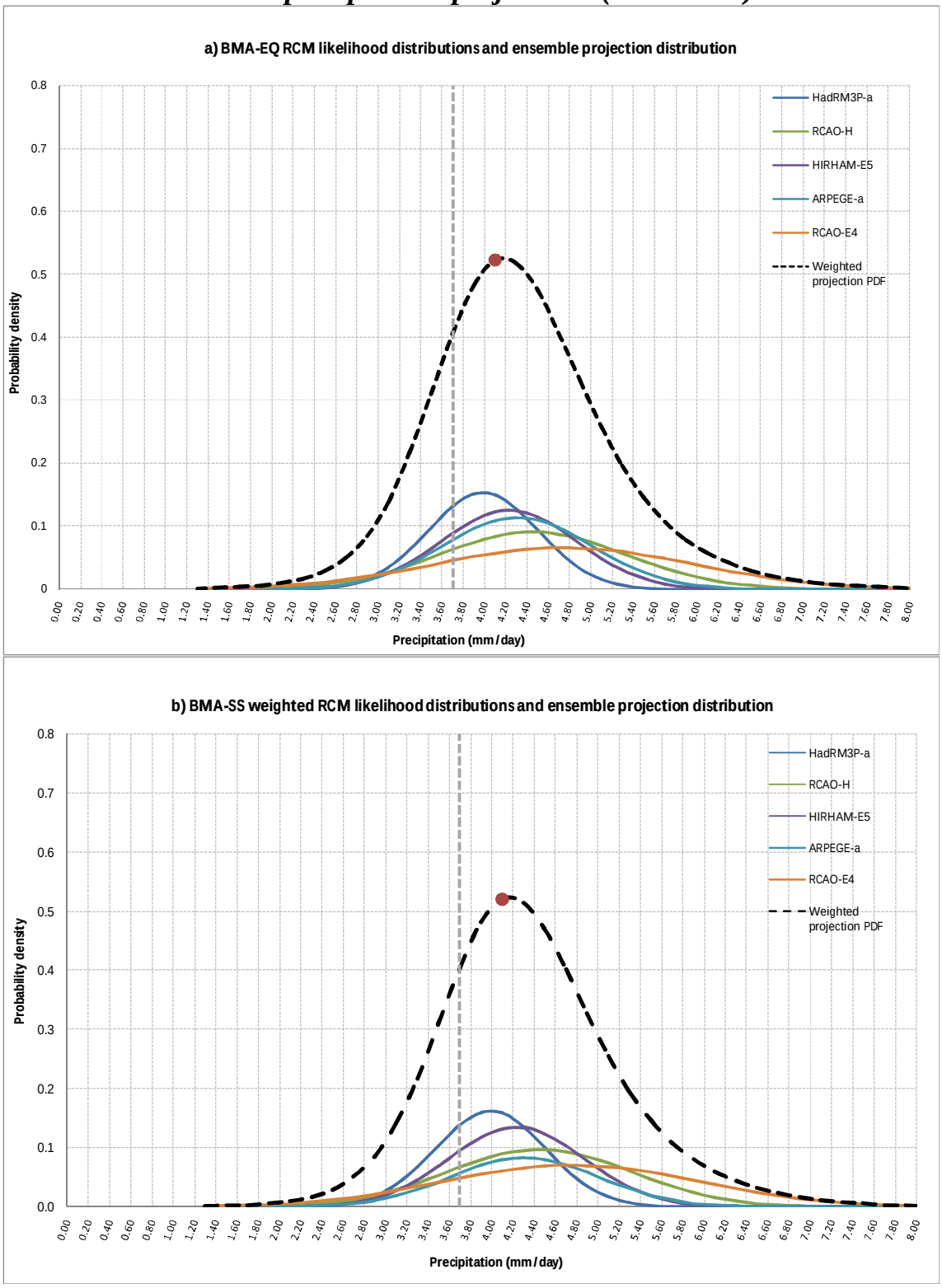
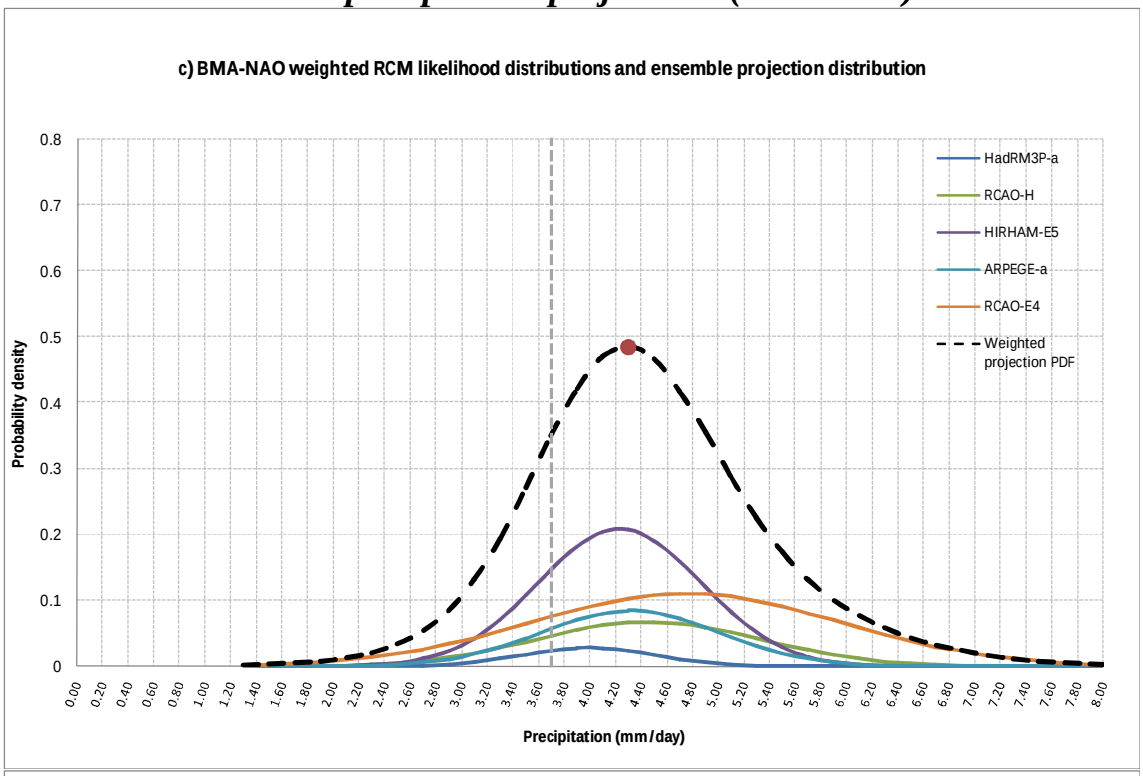


Figure 7.16: Weighted RCM likelihood distributions and weighted ensemble projection distribution for winter mean precipitation under the A2 emissions scenario, using a) BMA-EQ weighting and b) BMA-SS weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

A2 winter precipitation projections (2071-2100)



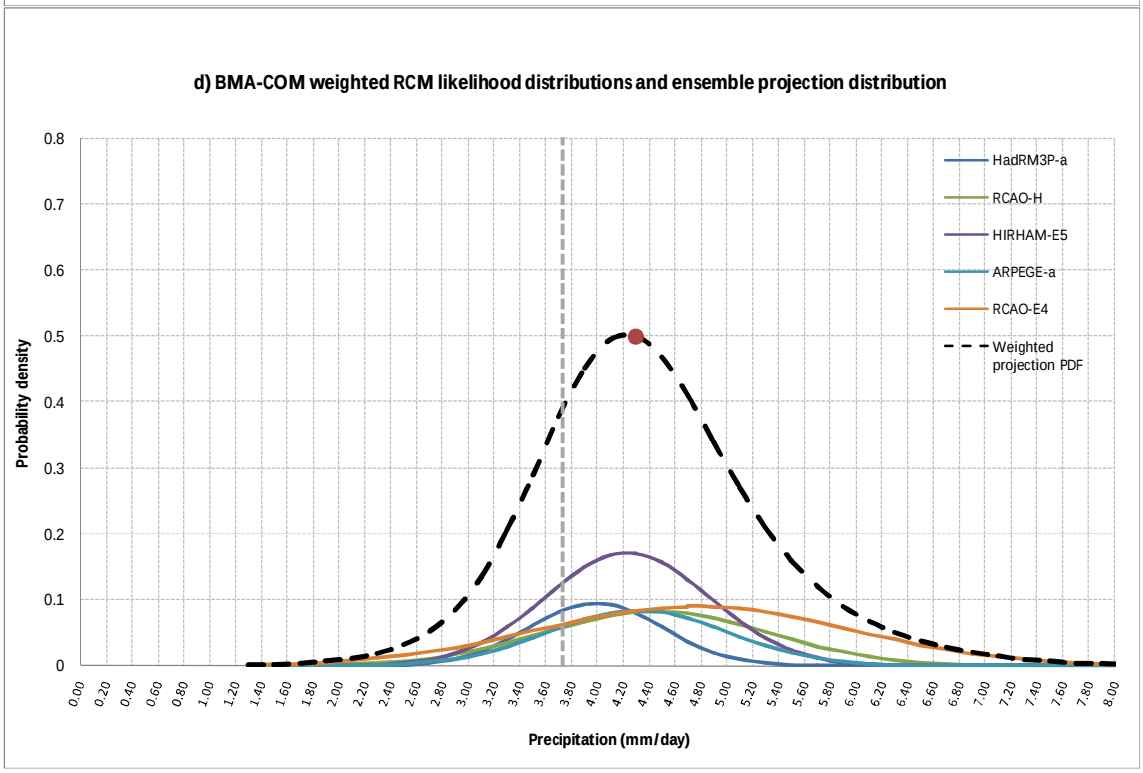


Figure 7.16 (continued): Weighted RCM likelihood distributions and weighted ensemble projection distribution for winter mean precipitation under the A2 emissions scenario, using c) BMA-NAO weighting and d) BMA-COM weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection

When equal weights or skill-score-based weights are applied, the ensemble PDF is normal-shaped. When BMA-NAO or BMA-COM weightings are used, this distribution becomes slightly less peaked. In the first two projections (Figures 7.16 a and b), the HadRM3P-a projection dominates. The 2071-2100 output from this model has a smaller standard deviation compared to other models, meaning that precipitation projections from this model are more likely to fall nearer the mean value. This results in a likelihood distribution with a greater peak relative to other models.This feature has an impact on the shape of the ensemble PDF under the BMA-EQ and BMA-SS approaches. However, HadRM3P-a demonstrated little skill in the NAO assessment in Chapter 6. As such, its weighting under the BMA-NAO and BMA-COM approaches is lower than the other models (Figure 7.15) and the lower weighting dampens its influence on the final weighted projection.

In this instance, likelihood associated with the ensemble PDF was influenced by the projections of HadRM3P-a, yet the NAO assessment indicates that this model does not perform well in simulating the large-scale drivers of Irish climate. As such, its likelihood function is potentially over-confident and the reliability of the weighted ensemble PDF is in turn compromised. However, the advantage of the Bayesian approach is that further information about model skill, such as performance at simulating the NAO, can be incorporated into the projection to reflect more fully the state of knowledge about model skill.

7.5.6 Winter (DJF) precipitation projections: B2 scenario

Figure 7.17 illustrates the deterministic AEM projection for winter precipitation under the B2 emissions scenario. While the AEM reflects the median of the individual ensemble projections, the method offers little information about the range and likelihood of projections. Figure 7.18 illustrates the contributions of the various models under different weighting schemes. Figure 7.19 shows the BMA-EQ, BMA-SS, BMA-NAO and BMA-COM ensemble projections. The overall shape of the weighted ensemble PDF in all four cases is a normal shape with an elongated upper tail. However, the precise characteristics of the curve vary depending on the weighting used.

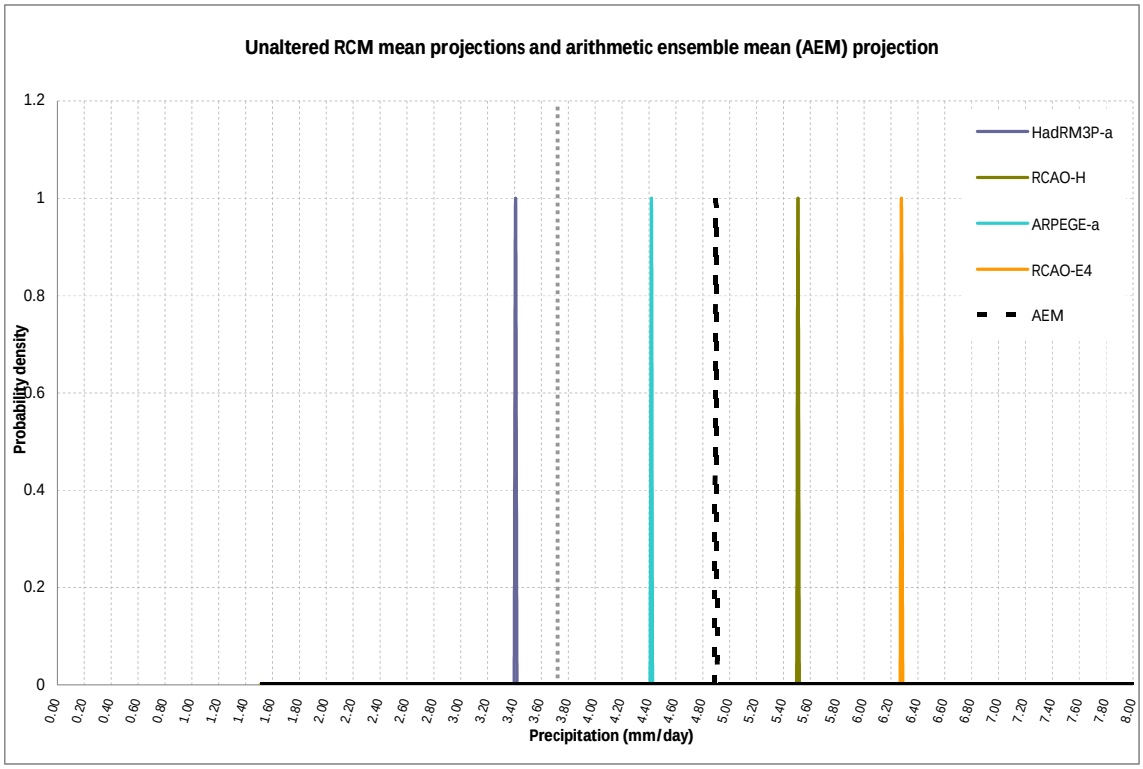
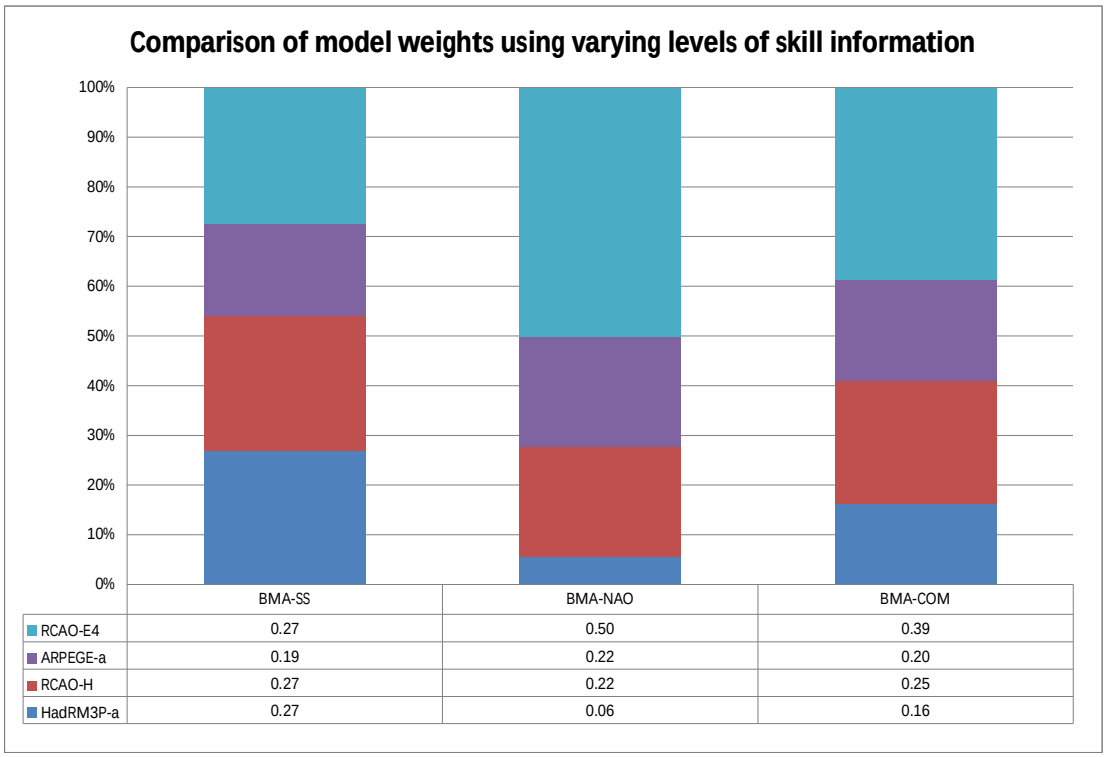


Figure 7.17: Unaltered RCM mean projections and arithmetic ensemble mean (AEM) projection based on these models for winter mean precipitation under the B2 emissions. Systematic bias has not been correction and no information is used to construct the likelihood of individual projections.



	BMA-SS	BMA-NAO	BMA-COM
RCAO-E4	0.27	0.50	0.39
ARPEGE-a	0.19	0.22	0.20
RCAO-H	0.27	0.22	0.25
HadRM3P-a	0.27	0.06	0.16

Figure 7.18: Comparison of model weights using varying levels of skill information for winter mean precipitation under the B2 emissions scenario. Systematic bias is corrected and scores indicate model skill based on performance in the present day.

B2 winter precipitation projections (2071-2100)

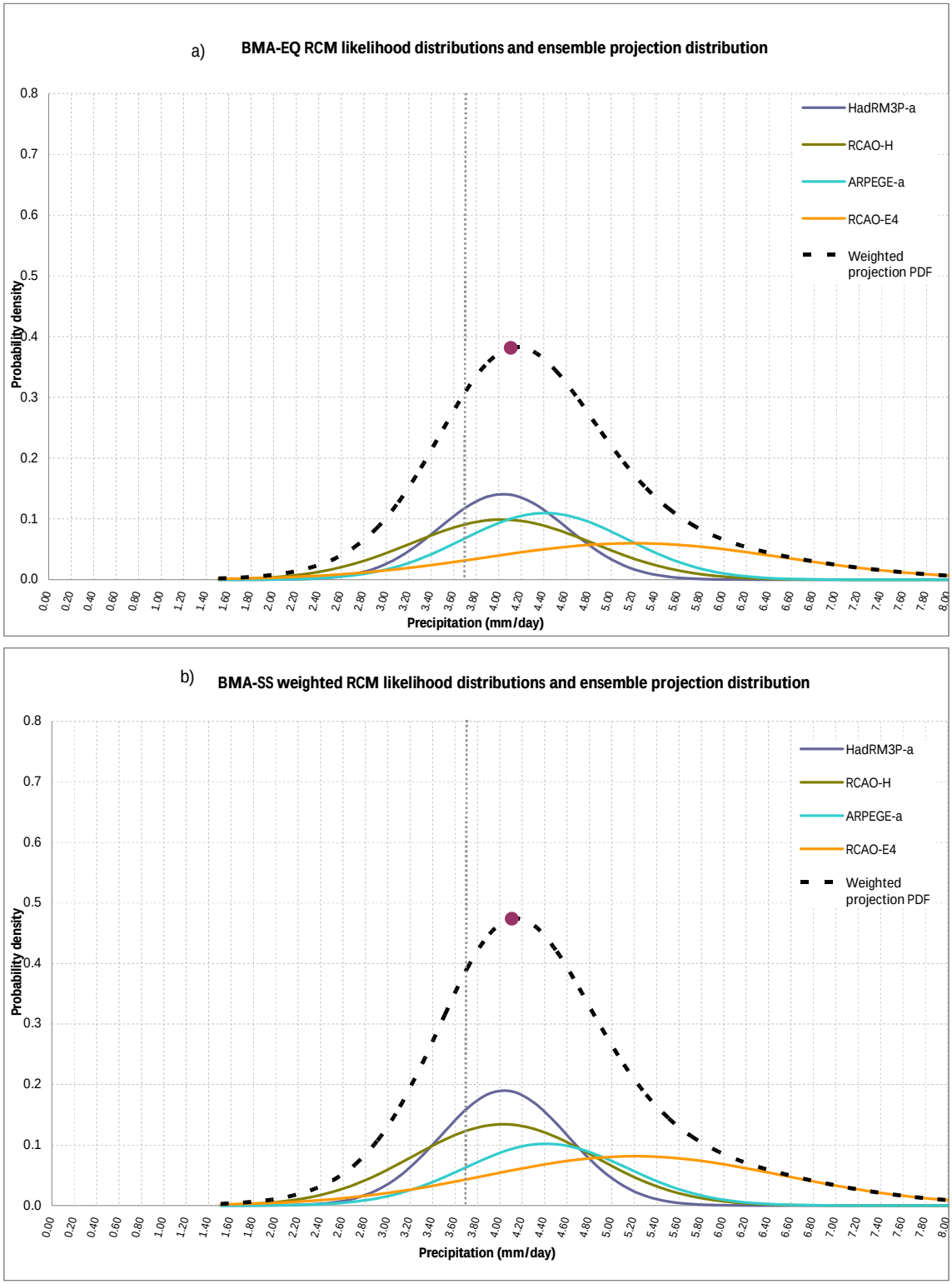


Figure 7.19 : Weighted RCM likelihood distributions and weighted ensemble projection distribution for winter mean precipitation under the B2 emissions scenario, using a) BMA-EQ weighting and b) BMA-SS weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

B2 winter precipitation projections (2071-2100)

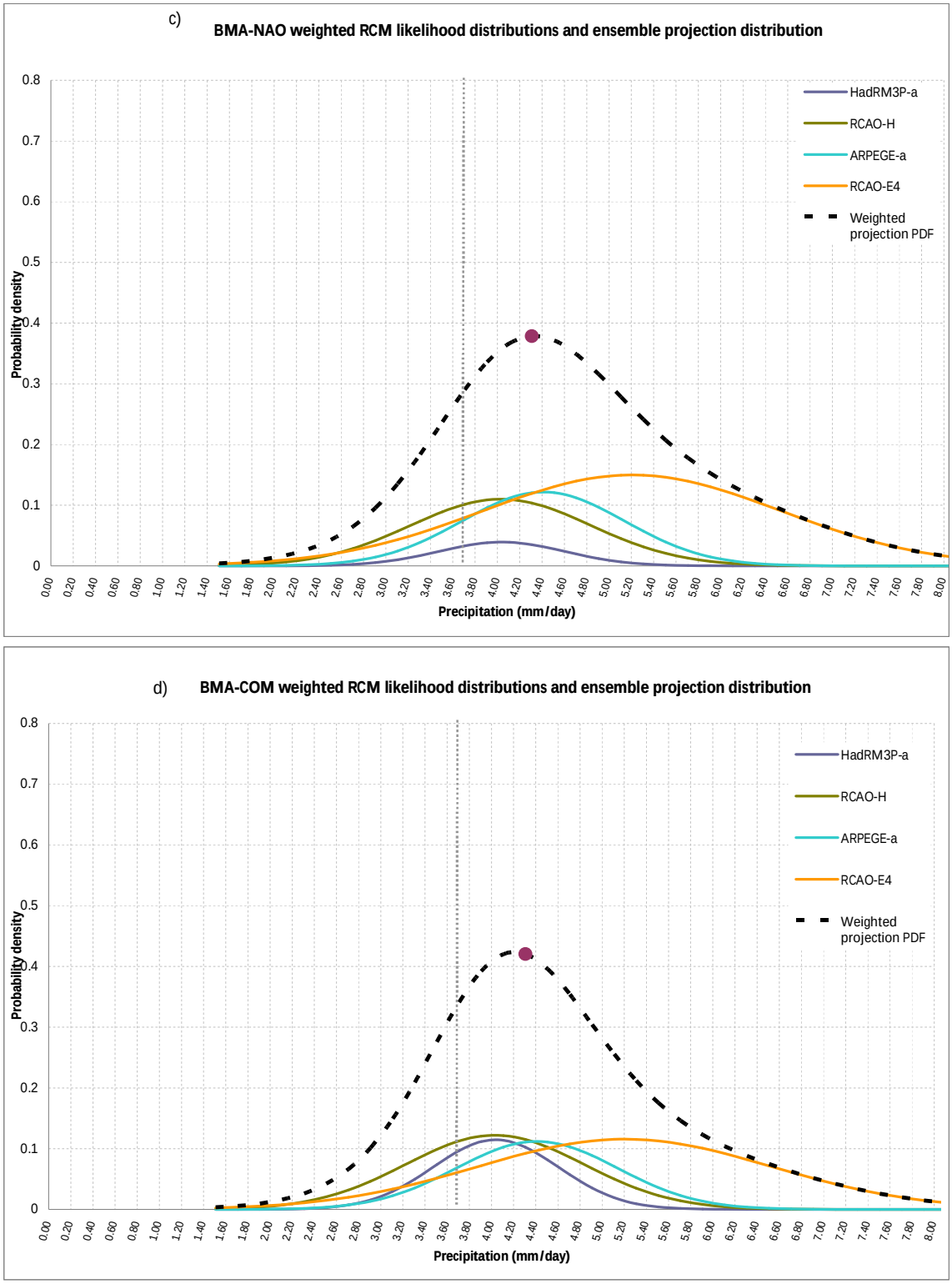


Figure 7.19 (continued): Weighted RCM likelihood distributions and weighted ensemble projection distribution for winter mean precipitation under the B2 emissions scenario, using c) BMA-NAO weighting and d) BMA-COM weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

When equal weights or skill-score-based weights are applied, the ensemble PDF is quite similar with a most likely value of between 4.0 and 4.2mm/day, although the BMA-SS PDF is more peaked. (Figure 7.19 a and b) Again, HadRM3P-a has a very peaked likelihood distribution and when combined with a skilful spatiotemporal skill score, the resultant ensemble PDF appears more confident, with increased likelihood associated with the most likely projection.

However, the projection may actually be over-confident, as the addition of NAO skill information into the weighting changes the projection considerable. HadRM3P-a demonstrated a low level of skill in simulating the NAO, therefore its lower weighting when this information is included dampens the initial high confidence associated with its projection, limiting its influence on the weighted ensemble PDF. Instead, RCAO-E4 becomes the key contributor to the ensemble PDF, as this model captured the NAO quite well (Figure 7.18). As such, the projection becomes heavier in the upper tail, signifying an enhanced level of likelihood associated with values at the upper end of the projection range (Figure 7.19c). The most likely projected precipitation value under BMA-NAO weighting is 4.3mm/day. The PDF is less peaked, signifying a decrease in the level of likelihood associated with the most likely projection. When both skill scores and NAO information are combined in the BMA-COM weightings, the most likely projected precipitation value is approximately 4.1mm/day (Figure 7.19d).

7.5.7 Summer (JJA) mean precipitation: A2 emissions scenario

Figure 7.20 illustrates the deterministic AEM projection for summer precipitation under the A2 emissions scenario. The AEM projects a decrease in rainfall, with the projection of HadRM3P-a lowering the average projection. However, in the control period, HadRM3P-a was found to be a dry model. Therefore, the application of the BMA technique has the potential to greatly increase the reliability of this projection. Figure 7.21 illustrates the contributions of the various models under different weighting schemes. Figure 7.22 shows the BMA-EQ(a), BMA-SS(b), BMA-NAO(c) and BMA-COM(d) ensemble projections. The overall

shape of the weighted ensemble PDF in all four cases is a normal shape, though the precise characteristics of the curve vary depending on the weighting used.

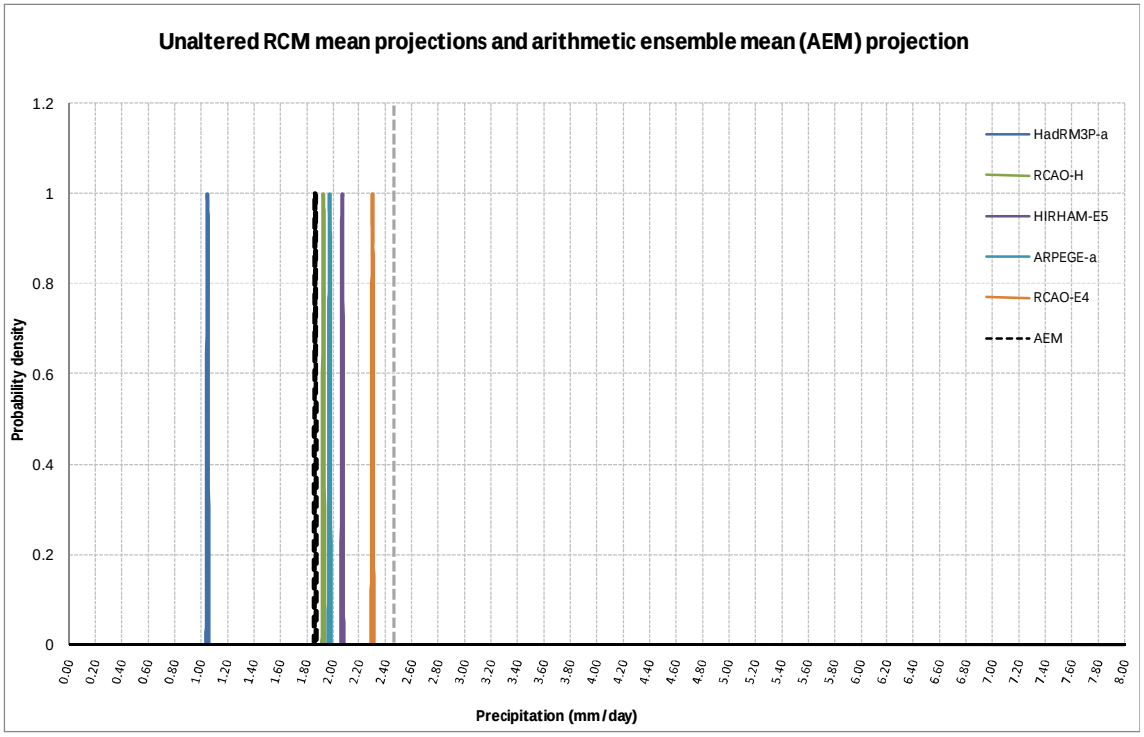
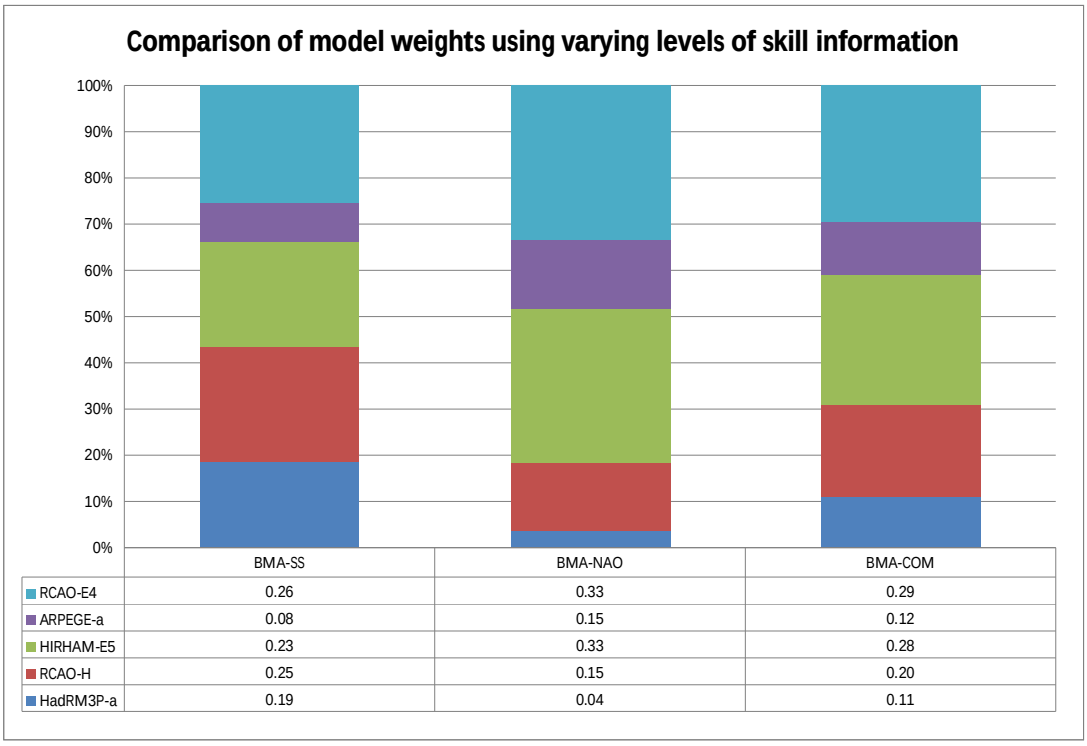


Figure 7.20: Unaltered RCM mean projections and arithmetic ensemble mean (AEM) projection for summer mean precipitation under the A2 emissions scenario. No skill information is used to construct the likelihood of individual projections.



	BMA-SS	BMA-NAO	BMA-COM
RCAO-E4	0.26	0.33	0.29
ARPEGE-a	0.08	0.15	0.12
HIRHAM-E5	0.23	0.33	0.28
RCAO-H	0.25	0.15	0.20
HadRM3P-a	0.19	0.04	0.11

Figure 7.21: Comparison of model weights using varying levels of skill information for summer mean precipitation under the A2 emissions scenario. Systematic bias is corrected and scores indicate model skill based on performance in the present day.

A2 summer precipitation projections (2071-2100)

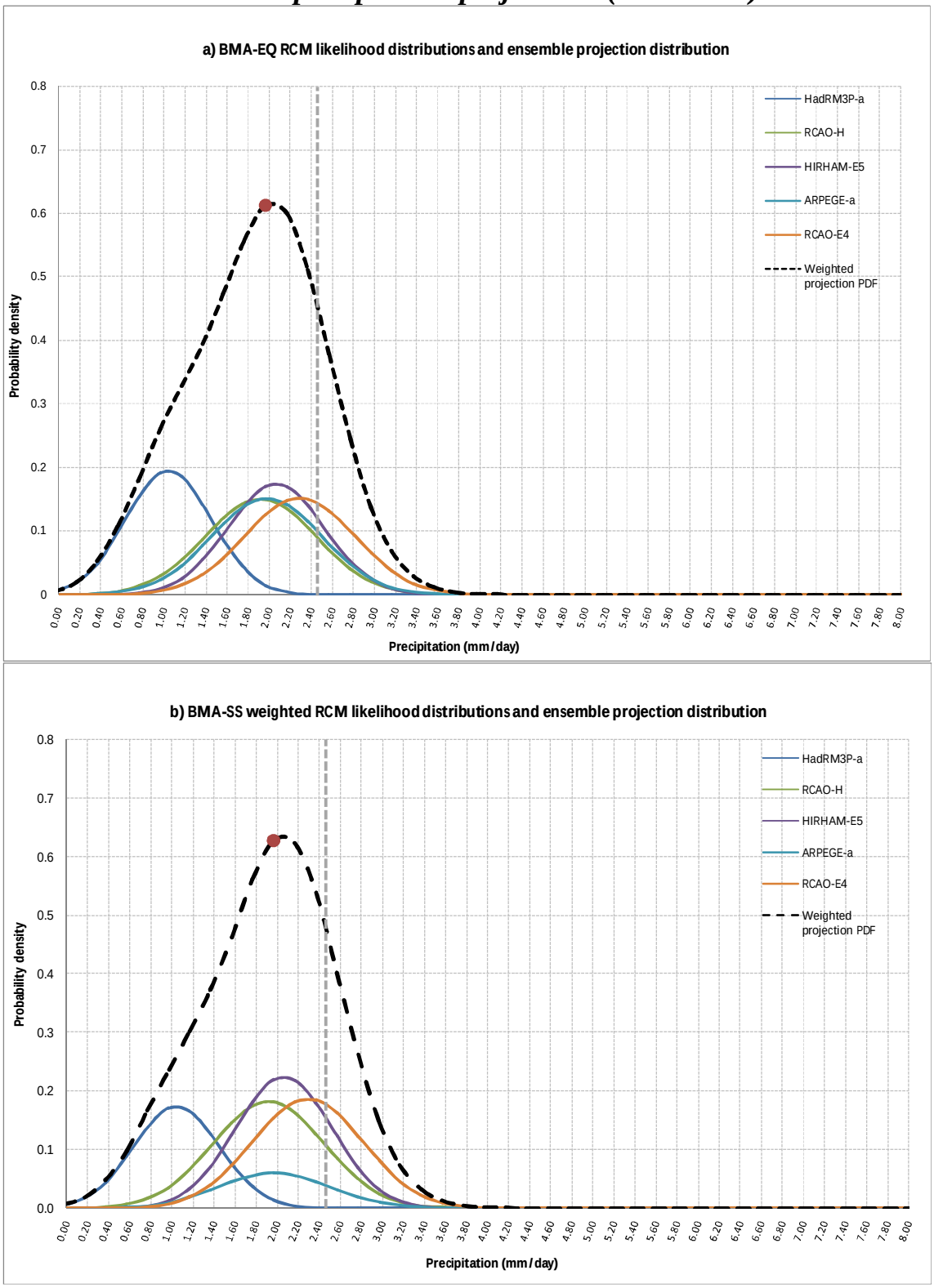
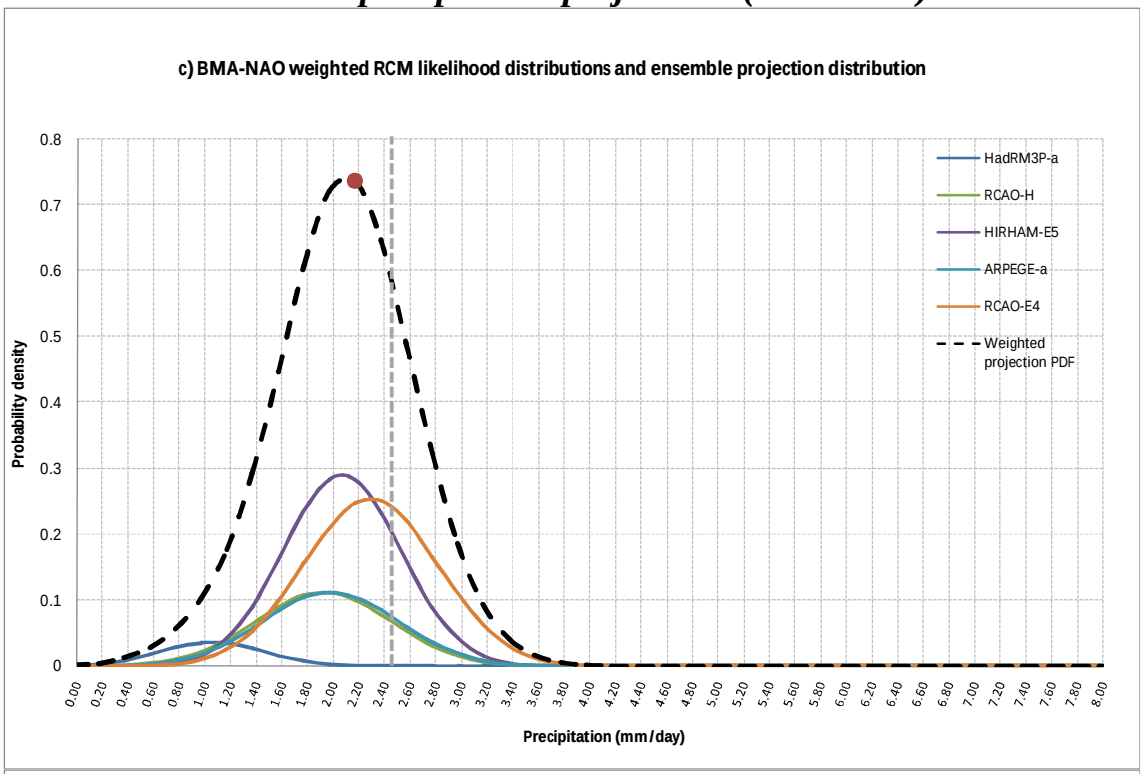


Figure 7.22: Weighted RCM likelihood distributions and weighted ensemble projection distribution for summer mean precipitation under the A2 emissions scenario, using a) BMA-EQ weighting and b) BMA-SS weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection
n.



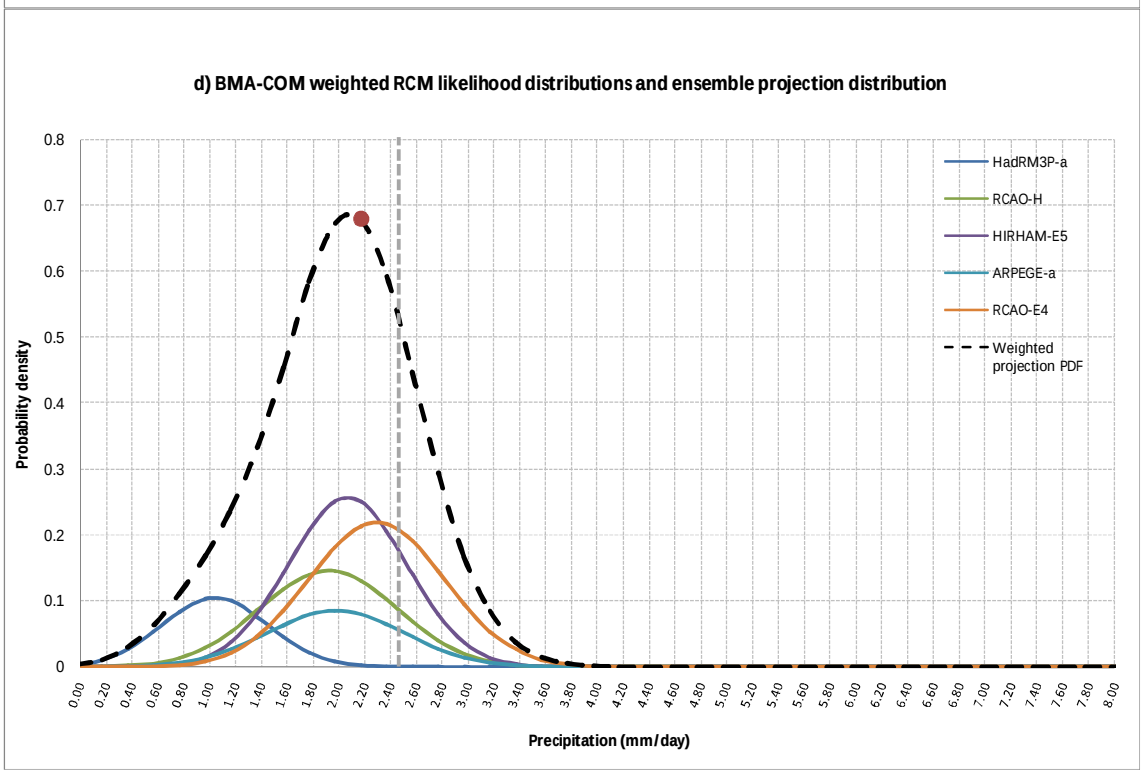


Figure 7.22 (continued): Weighted RCM likelihood distributions and weighted ensemble projection distribution for summer mean precipitation under the A2 emissions scenario, using c) BMA-NAO weighting and d) BMA-COM weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection

.

When equal weights or skill-score-based weights are applied, the ensemble PDF is heavier on the lower tail, suggesting that extreme low values of precipitation are more likely than extreme high values of precipitation under the A2 scenario (Figure 7.22a and b). The influence of HadRM3P-a is evident here. This model tended towards drier conditions in the control period, though there was no significant systematic bias found or corrected for. However, it is important to note that bias in the control period may not be representative of bias under future forcing conditions. Model errors and biases may not stay constant under different forcing conditions and as such, the skill scores calculated for the control period may not reflect the skill of the model in a future time period. Incorporating other forms of skill assessment into the projection may add to the reliability of the projection

The addition of NAO skill information into the weighting results in a more confident projection, illustrated by enhancement of the distribution peak. As HadRM3P-a demonstrated a low level of skill in simulating the NAO (Figure 7.21), its influence on the weighted ensemble PDF is dampened. The most likely projected precipitation value under BMA-NAO weighting is approximately 2.15mm/day, a decrease of 0.3mm/day compared with the control period (Figure 7.22c). This is a 12% decrease, amounting to 9mm less precipitation over the course of a month. However, it must be noted that while this is the value with the highest likelihood associated with it, there is a range of both higher and lower values modelled by the RCMs and the full range of these outcomes must be considered for the purposes of robust climate planning.

7.5.5 Summer (JJA) mean precipitation: B2 emissions scenario

` Figure 7.23 illustrates the deterministic AEM projection for summer precipitation under the B2 emissions scenario. The model RCAO tends towards wetter conditions than observed in the control period, while HadRM3P-a tends towards much drier conditions. As such, the projection converges to a central value. The model RCAO-E4 tended towards slightly wetter conditions in the control period while HadRM3P-a exhibited a dry bias.

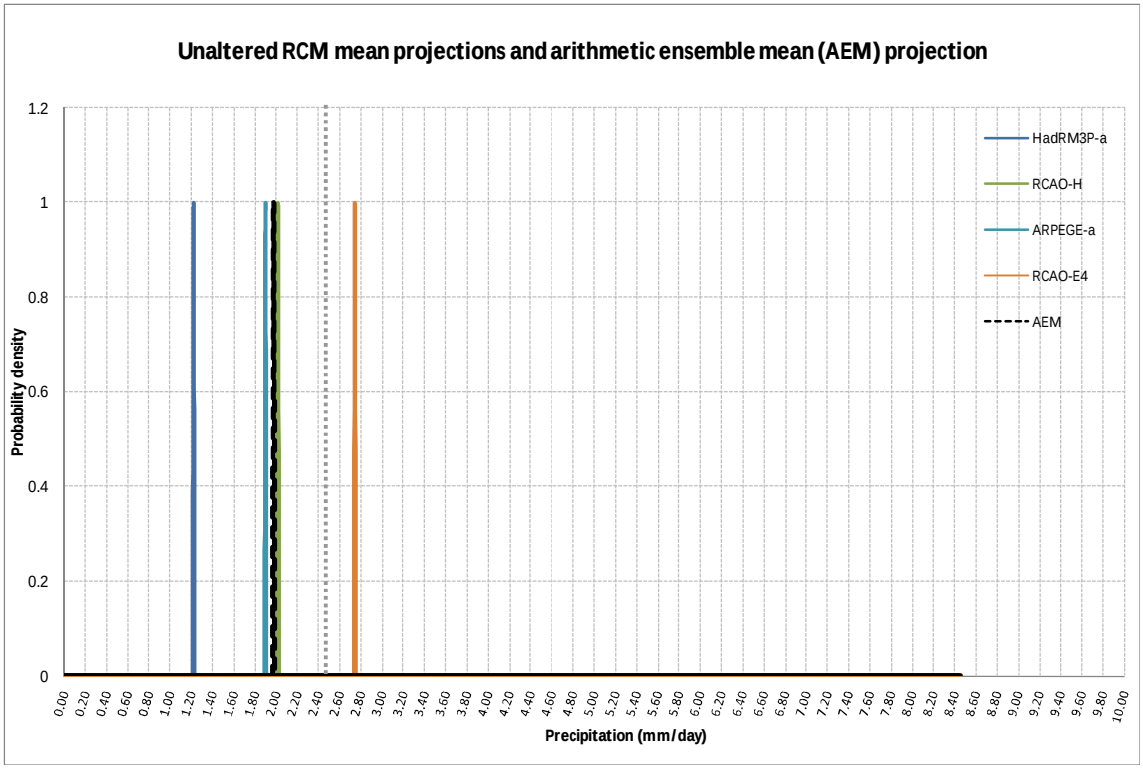
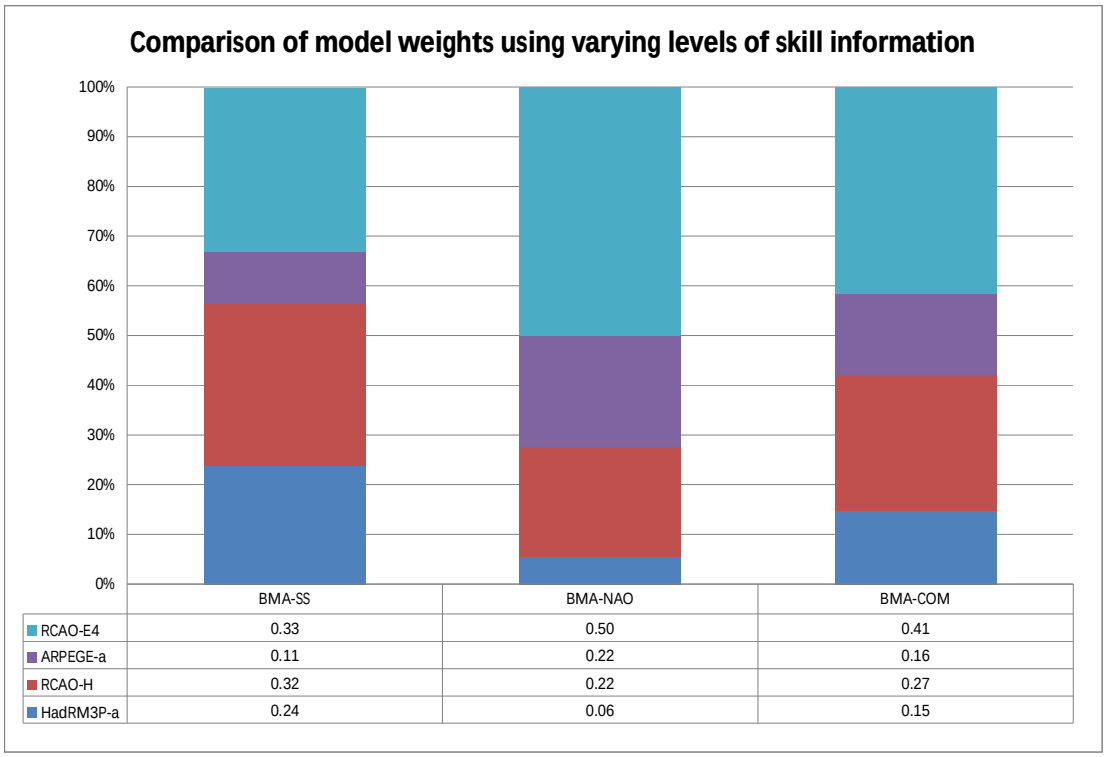


Figure 7.23: Unaltered RCM mean projections and arithmetic ensemble mean (AEM) projection based on these models for summer mean precipitation under the B2 emissions scenario. Systematic bias has not been correction and no information is used to construct the likelihood of individual projections.



	BMA-SS	BMA-NAO	BMA-COM
RCAO-E4	0.33	0.50	0.41
ARPEGE-a	0.11	0.22	0.16
RCAO-H	0.32	0.22	0.27
HadRM3P-a	0.24	0.06	0.15

Figure 7.24: Comparison of model weights using varying levels of skill information for summer mean precipitation under the B2 emissions scenario. Systematic bias is corrected and scores indicate model skill based on performance in the present day.

B2 summer precipitation projections (2071-2100)

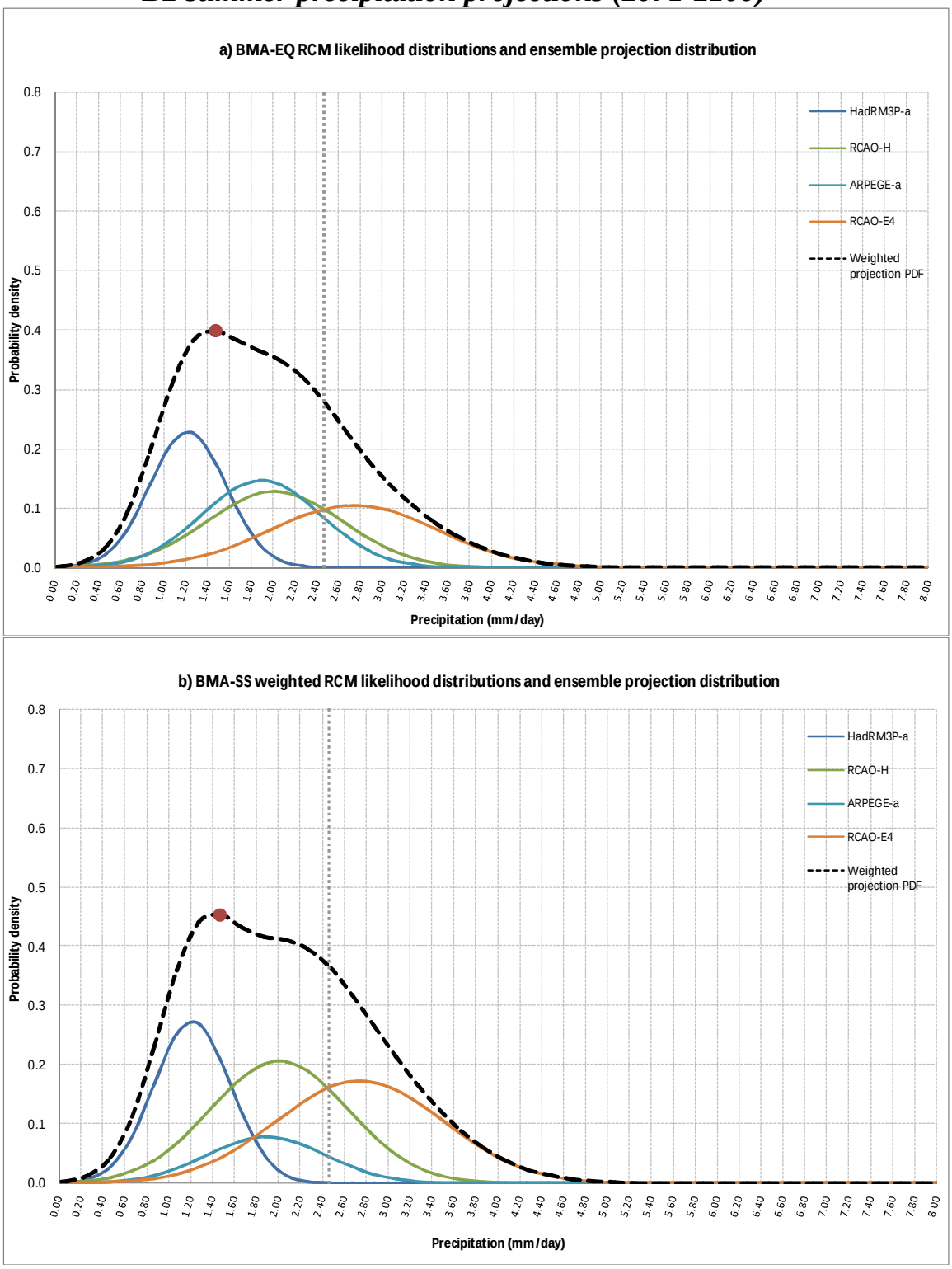
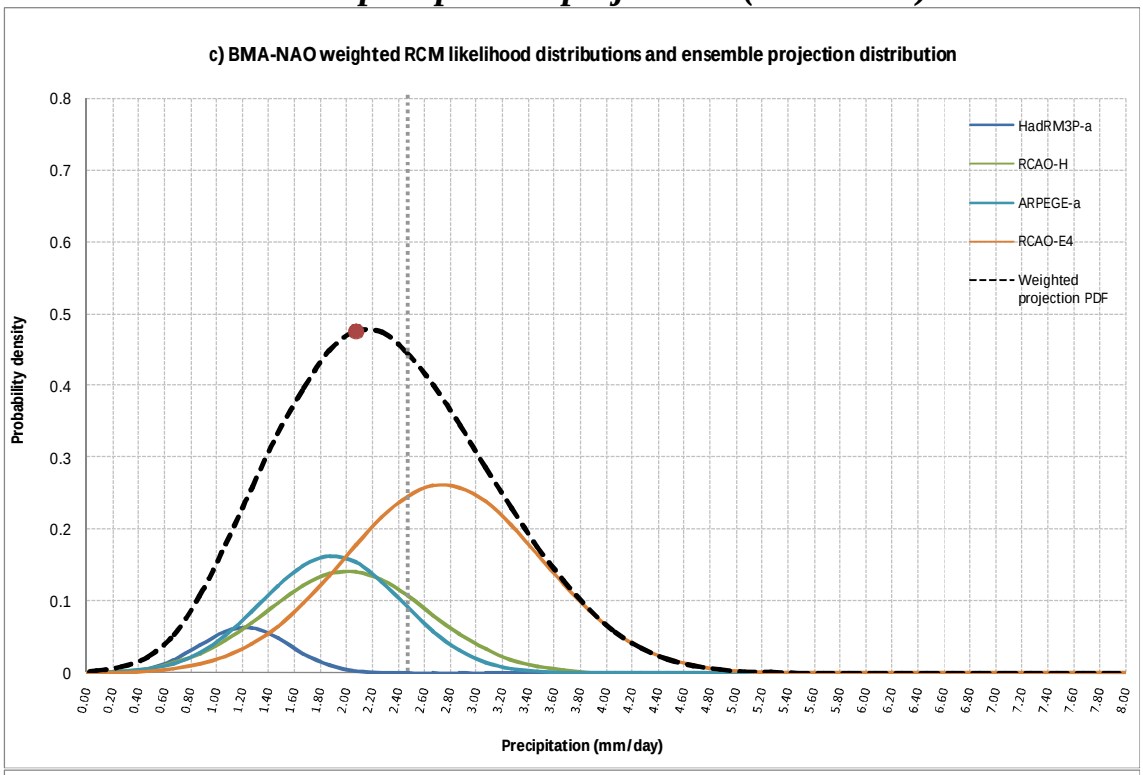


Figure 7.25: Weighted RCM likelihood distributions and weighted ensemble projection distribution for summer mean precipitation under the B2 emissions scenario, using a) BMA-EQ weighting and b) BMA-SS weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

B2 summer precipitation projections (2071-2100)



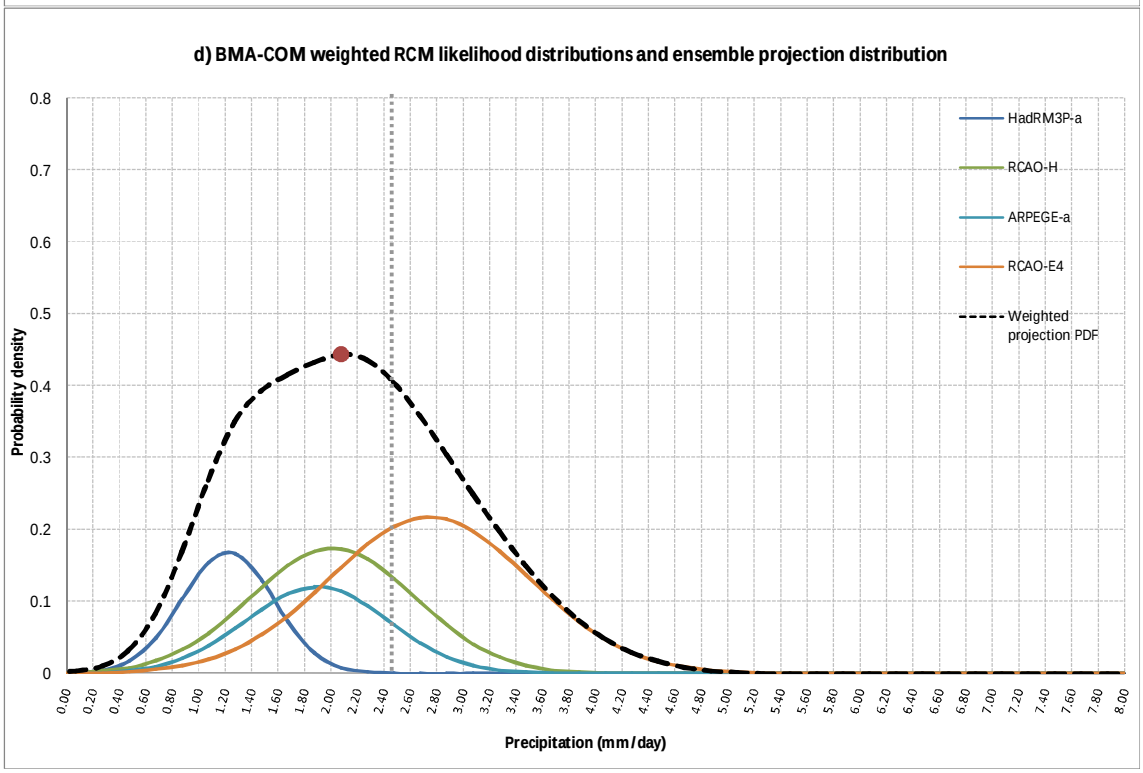


Figure 7.25 (continued): Weighted RCM likelihood distributions and weighted ensemble projection distribution for summer mean precipitation under the B2 emissions scenario, using c) BMA-NAO weighting and d) BMA-COM weighting. The grey line denotes observed winter mean temperature in the control period 1961-1990 and the red dot denotes the most likely future projection.

However, when the spatial r score signifies that there is not a strong co-variation between the observed and modelled pattern in the present day, the bias patterns cannot be considered systematic and are not corrected for.While forming projections using the relative change within the model would possibly overcome such errors, such an approach makes the unverifiable assumption that the relative change within the model is the climate change signal and that errors will not fluctuate over time. Inevitably, any method used to develop projections of future climate is susceptible to uncertainties of different kinds and communication of these uncertainties becomes the key issue.

Figure 7.24 illustrates the contributions of the various models under different weighting schemes. Figure 7.25 shows the BMA-EQ(a), BMA-SS(b), BMA-NAO(c) and BMA-COM(d) ensemble projections. The overall shape of the weighted ensemble PDF varies considerably depending on the weighting used. When equal weights are applied (Figure 7.25a), the ensemble PDF is heavy on the upper tail, but when skill scores are used (Figure 7.25b), the distribution starts to take on a bi-modal shape. HadRM3P-a, which has a smaller standard deviation in the future than the other models and simulated drier summer conditions in the control period that the other models, is a key contributor to this projection. The addition of NAO skill information into the weighting results in a distribution curve that is approximately normal. As RCAO-E4 modelled the NAO with considerable skill in the control period, it becomes the key contributor to the projection when NAO weights are used (Figure 7.24). As HadRM3P-a demonstrated a low level of skill in simulating the NAO, its influence on the weighted ensemble PDF is dampened.

In this instance, neither skill scores or NAO information alone offer an optimal approach to generating a projection, as the most influential models in both cases are outlying models with a degree of uncertainty attached to their projections. The combined weighting approach may offer enhanced reliability and the Bayesian methodology offers the ability to quantify the uncertainty associated with these divergent precipitation signals and combine them into a single PDF. The most likely projected precipitation value under BMA-COM weighting is approximately 2.07mm/day, a decrease of 0.38mm/day or 15% compared with the control period (Figure 7.25d). Again, while this is the value with the highest likelihood associated

with it under the BMA-COM weighting approach, there is a range of both higher and lower projections to be taken into account.

7.6 RESULTS: FUTURE CLIMATE PROJECTIONS WITH INCREASED ENSEMBLE SIZE

7.6.1 Winter (DJF) mean temperature: A2 emissions scenario

Projections of future winter temperature for the A2 emissions scenario using the full suite of available PRUDENCE simulations are given in Figure 7.26. These projections are calculated using the AEM(a), BMA-EQ(b) and BMA-SS(c) approaches. Weights which include NAO skill estimates cannot be applied as the NAO analysis carried out in Chapter 6 was conducted for a selected sample of models. The AEM approach (Figure 7.26a) shows how the deterministic means of each model lie relative to each other. Most models converge towards a central value but there are some clear outliers, namely the models driven by ECHAM-4/OPYC (RCAO-E4 and HIRHAM-E4) and CHRM. However, the convergence of the majority should not be taken as an indication of skill by itself, as the prevalence of models driven by HadAM3H makes the ensemble somewhat lacking in independence. Both the BMA-EQ and the BMA-SS (Figure 7.26 b and c) approaches produce a normal-shaped ensemble PDF with a most likely value of 6.8^{o}C. These results are quite similar to the PDFs obtained using the selected sub-sample of models. The inclusion of a larger number of simulations in the calculation of the ensemble projection is often understood to provide an increase in reliability. However, this is not always the case and a smaller sample of more independent simulations should be considered a more robust ensemble than a larger sample of simulations that share common characteristics.

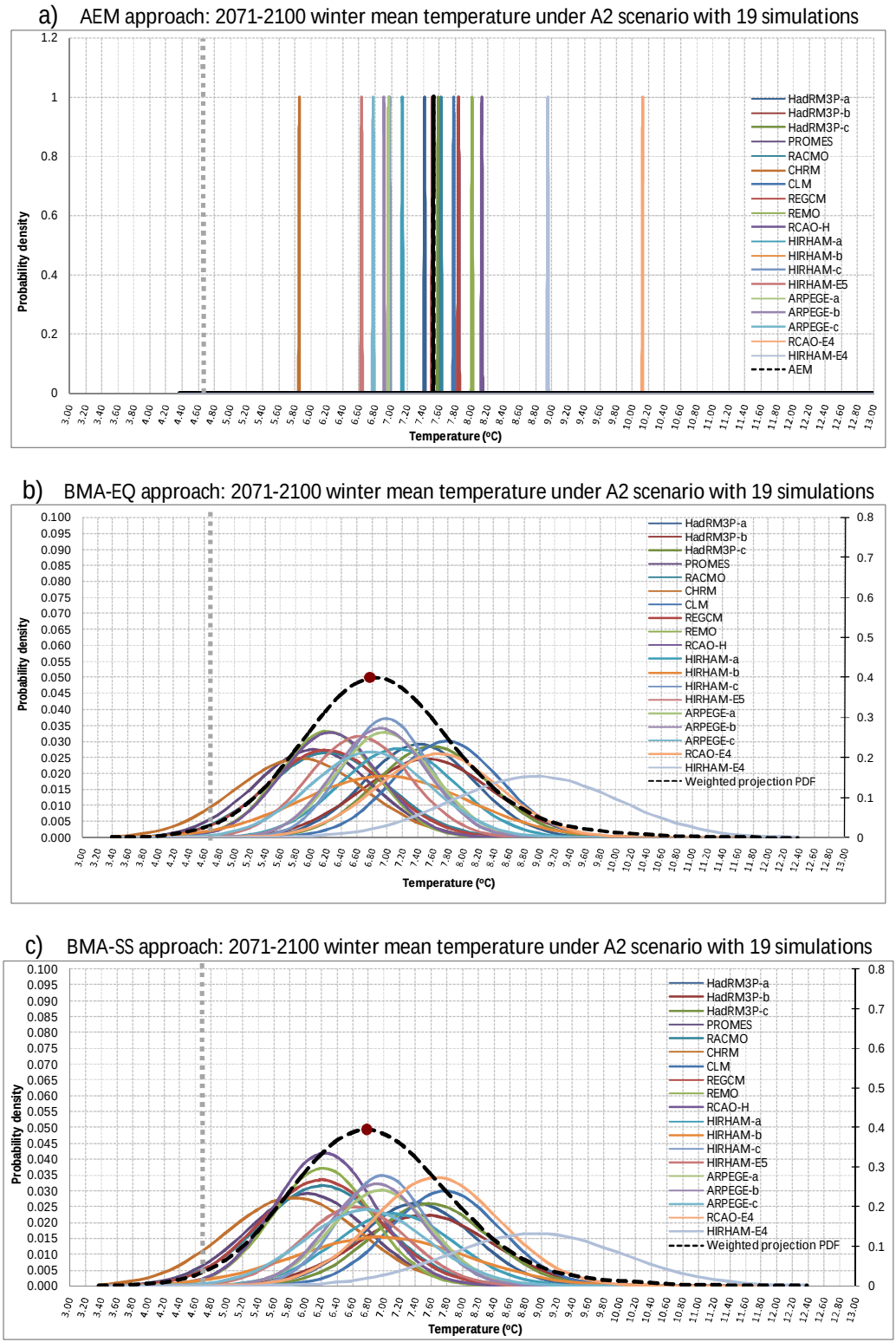


Figure 7.26: Projections for 2071-2100 winter mean temperature using a) AEM, b) BMA-EQ and c) BMA-SS approaches and 19 simulations. For a) and b), probability density for individual projections are plotted on the left, to allow model PDFs to be distinguished. Red dot denotes most likely projection. Grey dashed line denotes observed mean in the control period.

7.6.2 Winter (DJF) mean temperature: B2 emissions scenario

Projections of future winter temperature for the B2 emissions scenario, calculated using the AEM(a), BMA-EQ(b) and BMA-SS(c) approaches respectively and using the full suite of available PRUDENCE simulations, are given in Figure 7.27. With fewer simulations available for the B2 scenario, and particularly less models driven by HadAM3H, this ensemble is more independent that the A2 ensemble. Under the AEM approach (Figure 7.27a) the deterministic means converge less than they do for the A2 scenario.

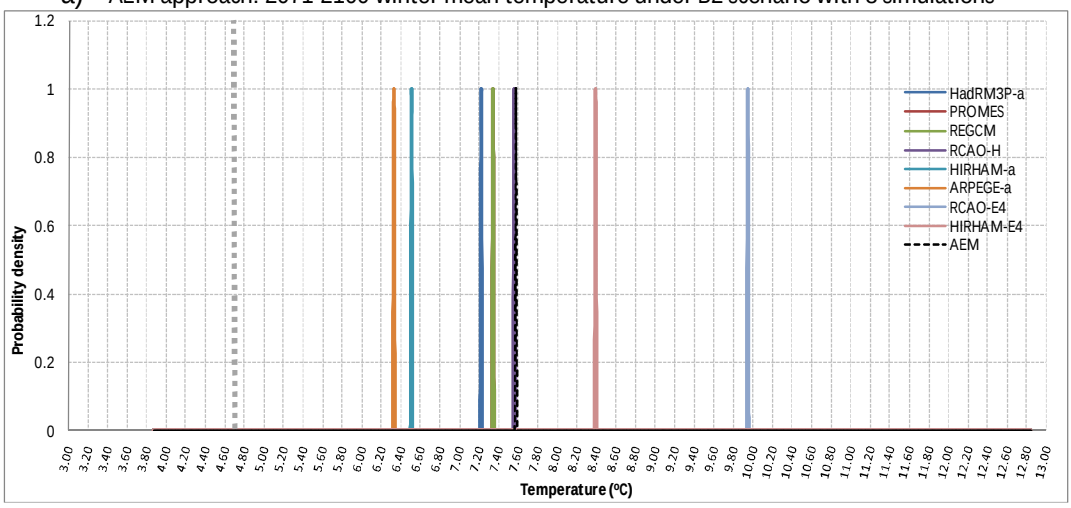
Both the BMA-EQ (Figure 7.27b) and the BMA-SS (Figure 7.27c) approaches produce a ensemble PDF which is approximately bell-shaped, though the upper tail is quite heavy. This is a very different shape to the bimodal PDF which emerges when the smaller sample is used. However, despite the increase in ensemble size, the lack of independence in the additional models means that these PDFs should be treated with caution. As there are more simulations available which using the HadCM3 AOGCM as a driver and less that used ECHAM4-OPYC, the influence of the models driven by ECHAM4-OPYC is constrained. In this case, the smaller sample of more independent simulations should be considered a more reliable projection.

7.6.3 Summer (JJA) mean temperature: A2 emissions scenario

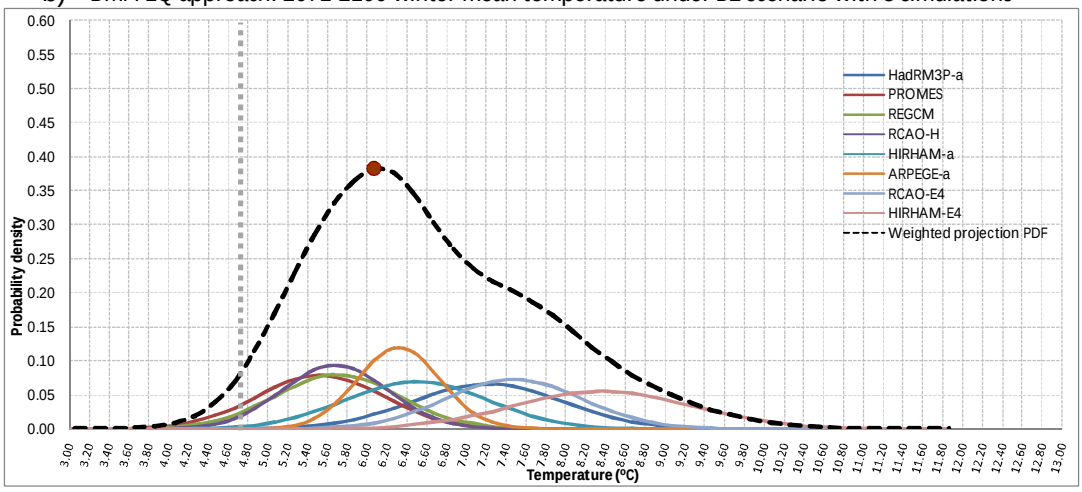
Projections of future summer temperature for the A2 emissions scenario using the full suite of available PRUDENCE simulations are given in Figure 7.28. Weighted projection results are quite similar to the PDFs obtained using the select sample of models, with the BMA-EQ (Figure 7.28b) and BMA-SS (Figure 7.28c) approaches resulting in a bell-shaped PDF with long tails at both ends.

As evidenced by the AEM graph (Figure 7.28a), the additional models tend to converge towards the same central value; therefore the inclusion of more models does not extend the range of the climate projections. Using the AEM approach (Figure 7.28a), the individual deterministic projections fall mostly between 15.4^{o}C and 17.3^{o}C, with the RCAO-E4 and HIRHAM-E4 appearing to diverge from the other models. However, again, this should not be considered a measure of skill as intermodel similarities have a great impact on how models converge or diverge.

a) AEM approach: 2071-2100 winter mean temperature under B2 scenario with 8 simulations



b) BMA-EQ approach: 2071-2100 winter mean temperature under B2 scenario with 8 simulations



c) BMA-SS approach: 2071-2100 winter mean temperature under B2 scenario with 8 simulations

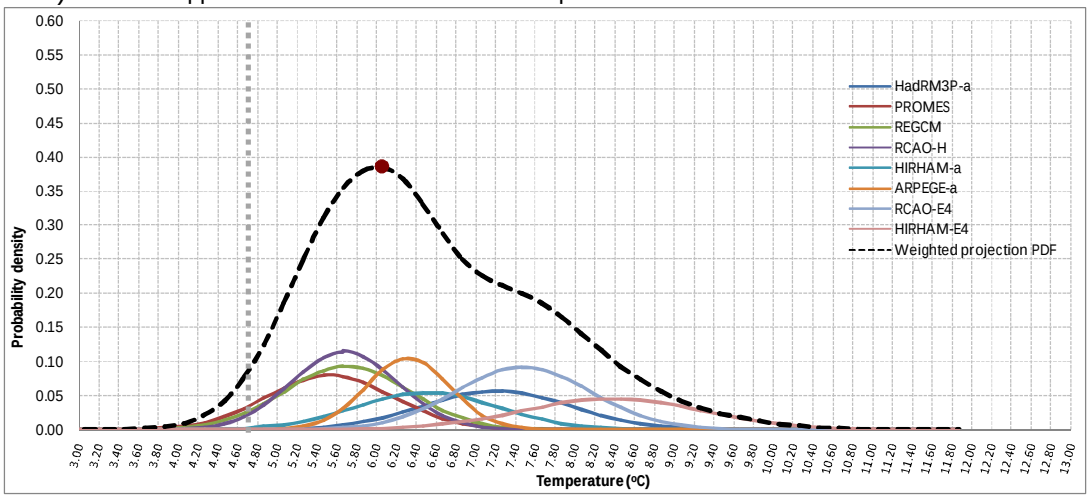
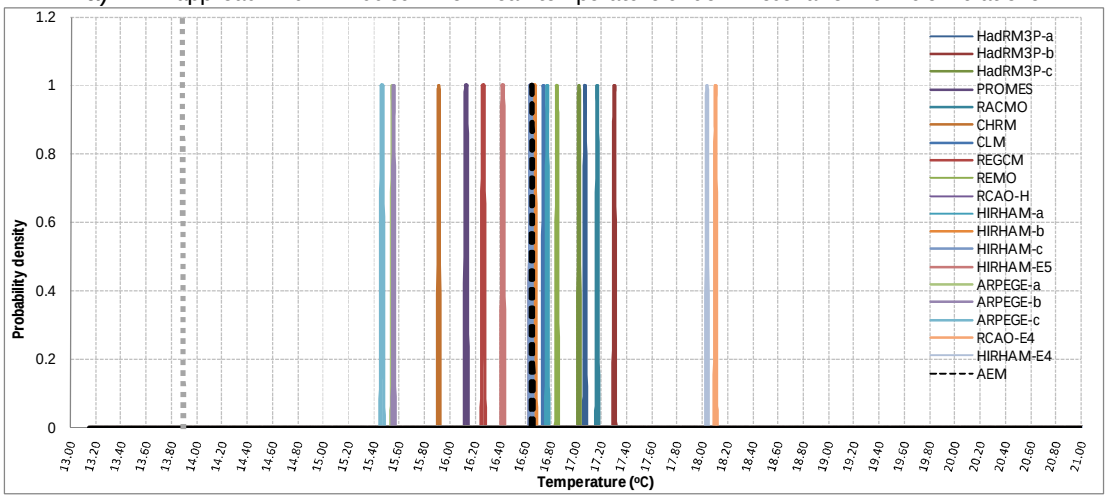
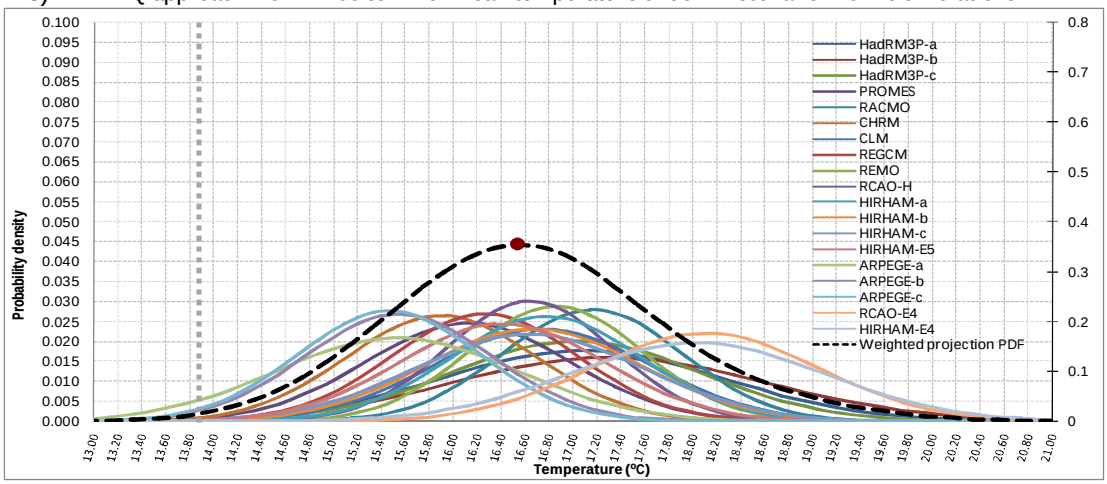


Figure 7.27: Projections for 2071-2100 winter mean temperature using a) AEM, b) BMA-EQ and c) BMA-SS approaches and 8 simulations. Red dot denotes most likely projection. Grey dashed line denotes observed mean in the control period.

a) AEM approach: 2071-2100 summer mean temperature under A2 scenario with 19 simulations



b) BMA-EQ approach: 2071-2100 summer mean temperature under A2 scenario with 19 simulations



c) BMA-SS approach: 2071-2100 summer mean temperature under A2 scenario with 19 simulations

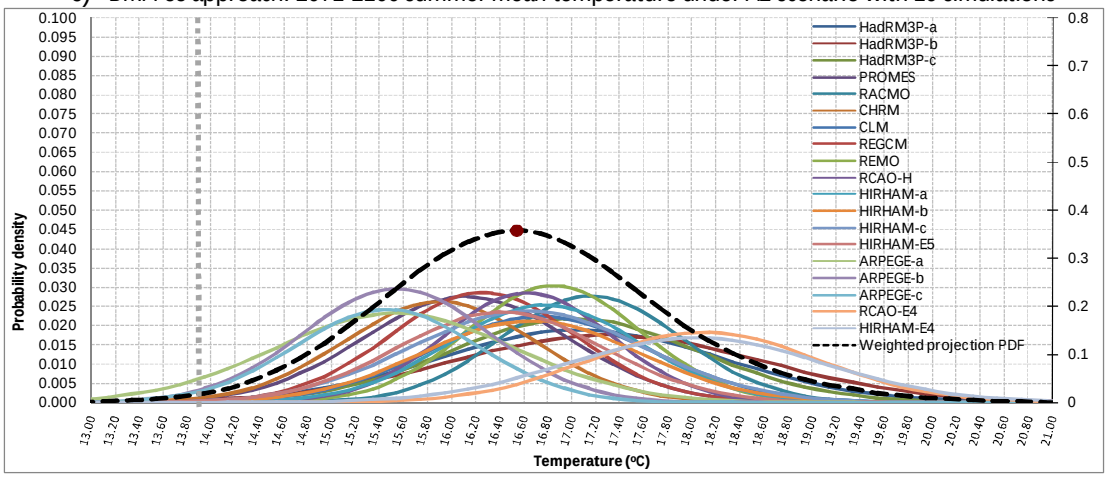


Figure 7.28: Projections for 2071-2100 summer mean temperature using a) AEM, b) BMA-EQ and c) BMA-SS approaches and 19 simulations. For a) and b), model likelihood PDFs are plotted on the left, to allow them to be distinguished. Red dot denotes most likely projection. Grey dashed line denotes observed mean in the control period.

Additionally, the convergence of many models towards the same value has the potential to induce overconfidence in the data. If the models were independent, then their convergence could be interpreted as strong evidence that the projection they converge towards is the most likely outcome. However as the models have significant similarities, their convergence is more likely to be a result of this lack of independence. This result illustrates the importance of interpreting the ensemble projection properly and understanding and communicating the uncertainties associated with the projection.

7.6.4 Summer (JJA) mean temperature: B2 emissions scenario

Projections of future summer temperature for the B2 emissions scenario, using the full suite of available PRUDENCE simulations, are given in Figure 7.29. Using the AEM approach, the ECHAM-4/OPYC models RCAO-E4 and HIRHAM-E4 again appear to be outliers compared to the rest of the ensemble (Figure 7.29a). The BMA-EQ and the BMA-SS approaches produce a PDF which is approximately bell-shaped, with a heavy upper tail (Figure 7.29b and c). The bimodal shape obtained with the smaller sample is indiscernible. Yet when one looks closely at the individual model likelihood functions, two distributions are evident. The ECHAM4-driven models simulate a very different distribution yet this characteristic is not evident in the ensemble PDF. The abundance of models driven by HadRM3P effectively cancels this feature out. Such an outcome provides strong motivation to select models for independence rather than sheer numbers. If models with shared characteristics must be used to generate ensemble projections, model independence should be quantified in some way and incorporated into the projection.

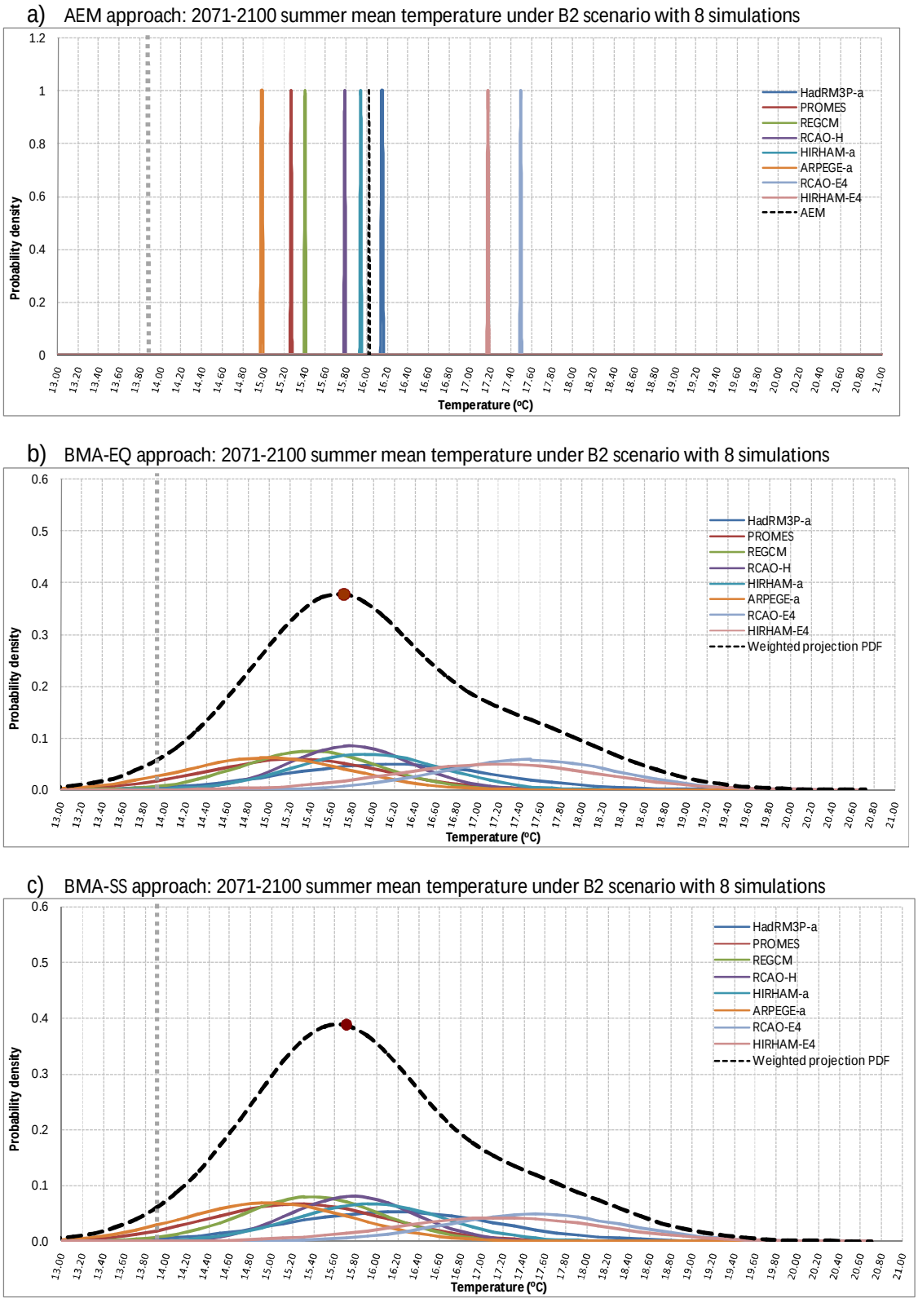


Figure 7.29: Projections for 2071-2100 summer mean temperature using a) AEM, b) BMA-EQ and c) BMA-SS approaches and 8 simulations. Red dot denotes most likely projection. Grey dashed line denotes observed mean in the control period.

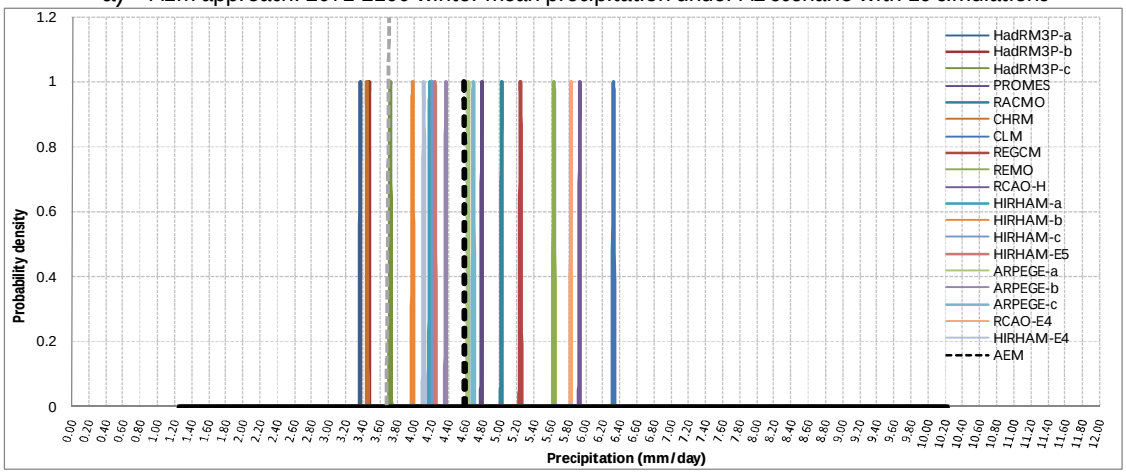
7.6.5 Winter (DJF) mean precipitation: A2 emissions scenario

Projections of future winter precipitation for the A2 emissions scenario using 19 simulations are given in Figure 7.30. Before systematic bias is corrected for (Figure 7.30a), two of the HadRM3P simulations model drier conditions than observed in the control period. After systematic bias is corrected, all models agree on an increase in winter precipitation. However, the degree of change varies between models. When likelihood distributions are constructed for each model, first with BMA-EQ weighting (Figure 7.30b) and then with BMA-SS weighting (Figure 7.30c), those driven by ECHAM-4/OPYC have a markedly different distribution to the other models. These distributions are flatter, signifying that precipitation values in these models are more dispersed about the mean than in the other models. This is to be expected as the standard deviation from which the distributions are constructed is dependent on the variability of the 2071-2100 data. As illustrated in Chapters 4 and 6, the interannual variability of the models in the control period is influenced greatly by the choice of GCM, therefore while the RCMs vary in terms of projection mean, projected standard deviation is similar for models with the same GCM driver, resulting in similar likelihood distributions.

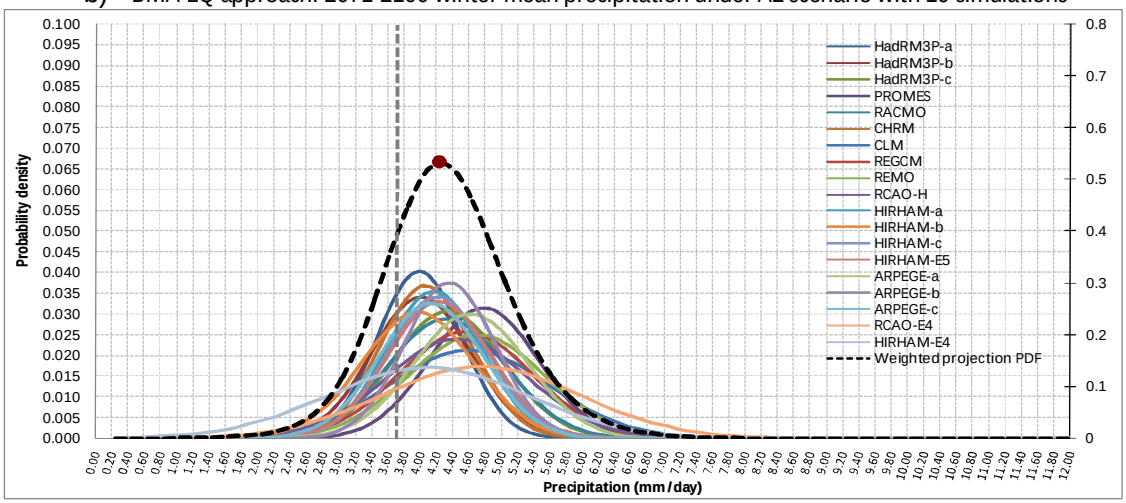
7.6.6 Winter (DJF) mean precipitation: B2 emissions scenario

Projections of future winter precipitation for the B2 emissions scenario using 8 simulations are given in Figure 7.31. These projections are calculated using the AEM(a), BMA-EQ(b) and BMA-SS(c) approaches. With less simulations available, there is greater independence between the simulations used in this ensemble. Comparing Figure 7.31a and Figure 7.31b, correcting the systematic bias reduces the range of the individual projections, but the range of potential increase is still quite large. When systematic bias is corrected, the models simulate a precipitation increase of between 0.1mm/day and 1.5mm/day. This is a considerable range, amounting to between 3mm and 45mm of excess accumulated precipitation over a month. Consider the model HIRHAM-a, which modelled spatial precipitation patterns in the control period with skill, exhibiting only a 0.1mm/day bias.

a) AEM approach: 2071-2100 winter mean precipitation under A2 scenario with 19 simulations



b) BMA-EQ approach: 2071-2100 winter mean precipitation under A2 scenario with 19 simulations



c) BMA-SS approach: 2071-2100 winter mean precipitation under A2 scenario with 19 simulations

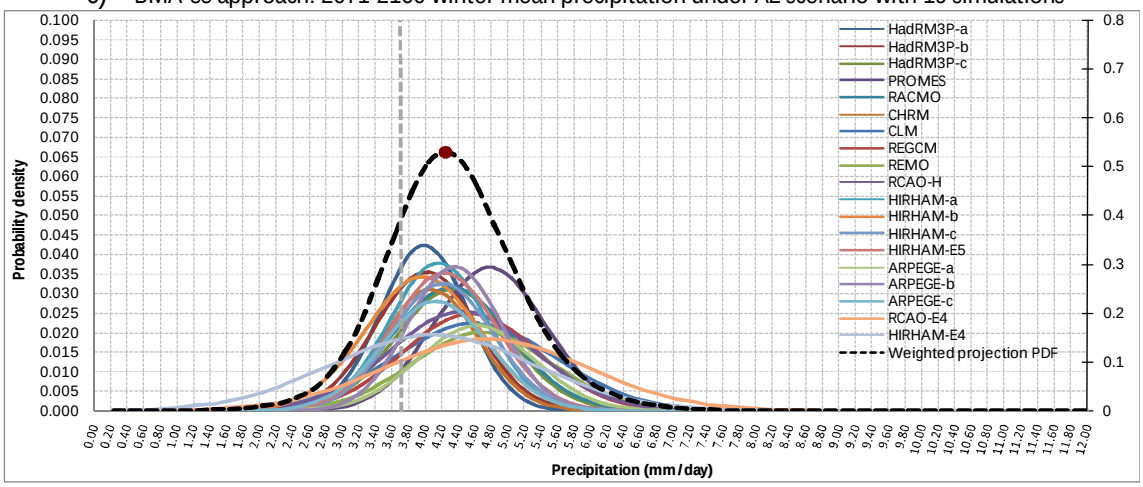
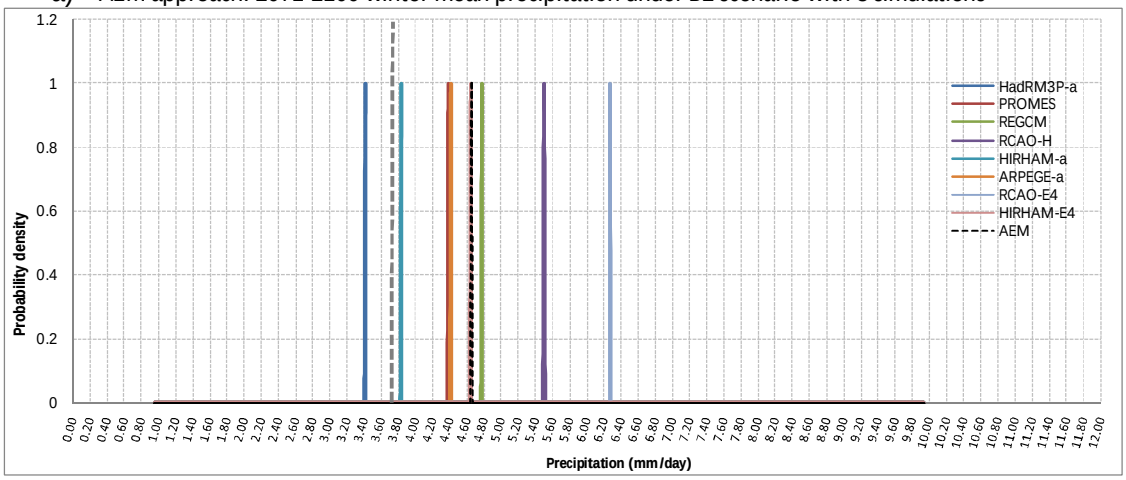
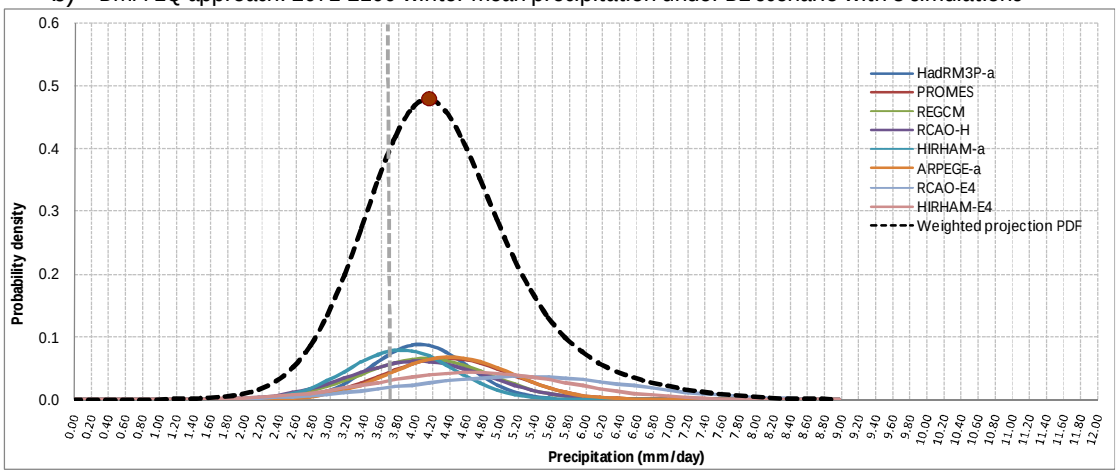


Figure 7.30: Projections for 2071-2100 winter mean precipitation using a) AEM, b) BMA-EQ and c) BMA-SS approaches and 19 simulations. For a) and b), individual likelihood PDFs are plotted on the left, to allow them to be distinguished. Red dot denotes most likely projection. Grey dashed line denotes observed mean in the control period.

a) AEM approach: 2071-2100 winter mean precipitation under B2 scenario with 8 simulations



b) BMA-EQ approach: 2071-2100 winter mean precipitation under B2 scenario with 8 simulations



c) BMA-SS approach: 2071-2100 winter mean precipitation under B2 scenario with 8 simulations

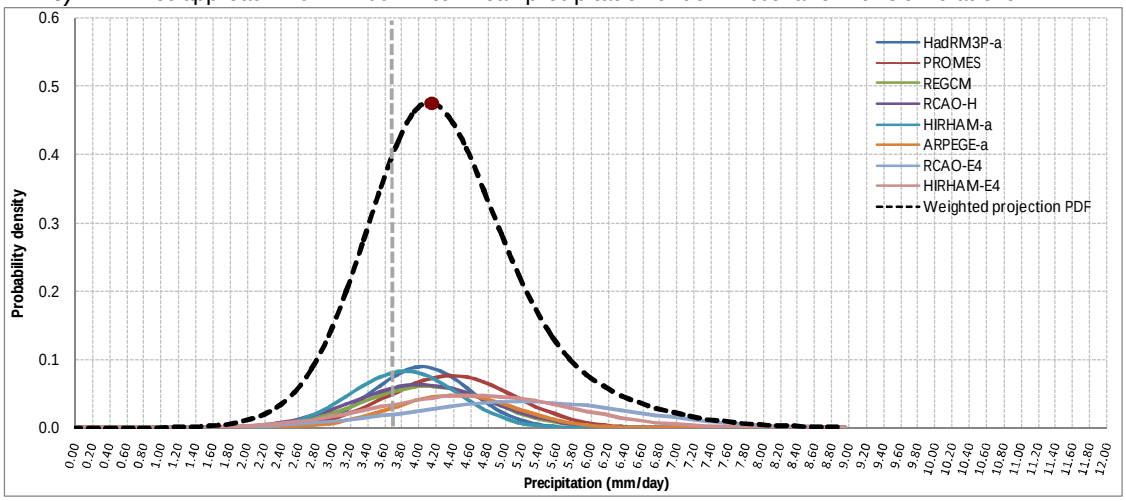


Figure 7.31: Projections for 2071-2100 winter mean precipitation using a) AEM, b) BMA-EQ and c) BMA-SS approaches and 8 simulations. Red dot denotes most likely projection. Grey dashed line denotes observed mean in the control period.

As this bias amounted to less than 10% of observed winter mean precipitation, it was not corrected for. Based on its skill in the control period, there is reason to have confidence in HIRHAM-a's projections. However, HIRHAM-a models no discernible difference between winter mean precipitation for 2071-2100 under the B2 emissions scenario and winter mean precipitation for 1961-1990.

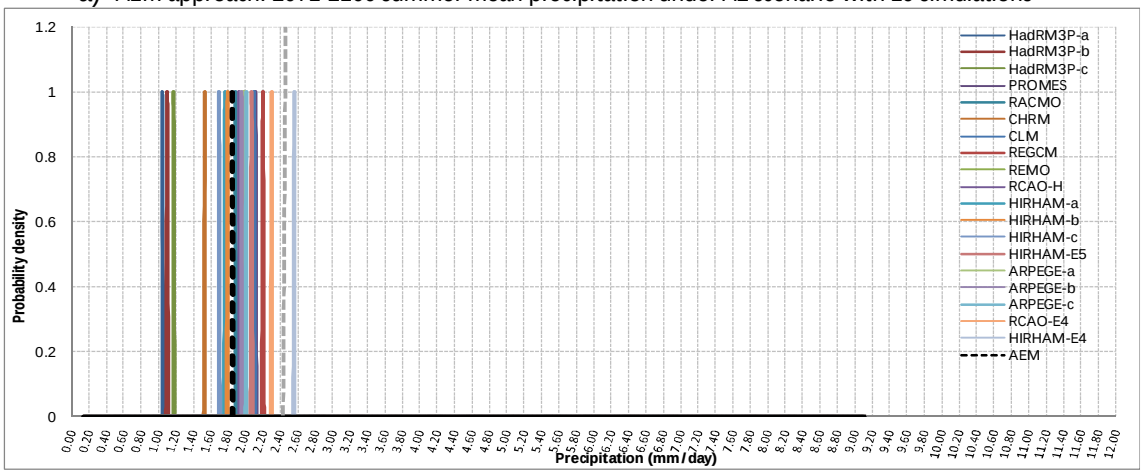
As there are no observations to compare future projections to, there is no way to determine which projections are closest to simulating the actual response of the climate system to the B2 forcing scenario. Interestingly, all the models modelled winter spatial precipitation patterns in the control period with skill and those that exhibited significant bias in the control period had this bias corrected. Therefore, the spread in projections reflects the varying response of each model to increased forcing. Results suggest that the response of the parameterizations governing winter precipitation in these models may not remain the same through time and under different forcing conditions, leading to a divergence of projections. As such, the representation of winter precipitation is clearly a key uncertainty for the RCMs. Projections using the BMA-EQ and BMA-SS approaches yield similar results. In light of the uncertainty surrounding this climate parameter and the limited information supplied by skill score-based assessments, the BMA-COM approach of combining skill scores with skill estimates based on the ability of models to capture large-scale atmospheric drivers may be more suitable and provide increased reliability in this instance.

7.6.7 Summer (JJA) mean precipitation: A2 emissions scenario

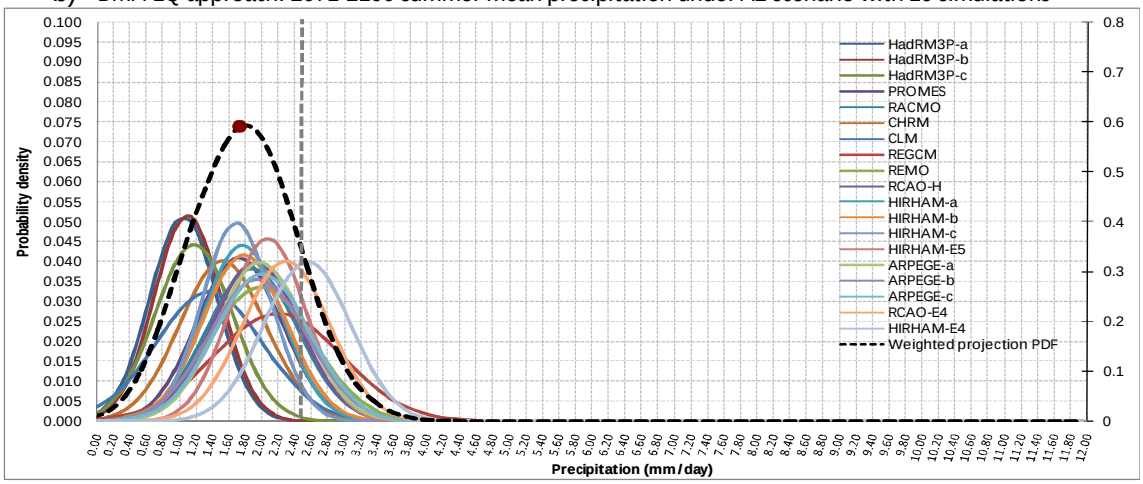
Projections of future summer precipitation for the A2 emissions scenario using 8 simulations are given in Figure 7.32. As spatial r values in the control period for summer mean precipitation were low to moderate in the majority of models, only two models, PROMES and CLM have their bias corrected in the future projections. As such, there is a degree of uncertainty in the A2 summer mean precipitation projections that cannot be accounted for. In the absence of evidence of systematic behaviour, applying a systematic correction to a random bias could lead to overconfidence, as the outcomes are constrained yet as the response of the model

parameterizations to the change in forcing is potentially non-stationary, uncertainty is still not accounted for.

a) AEM approach: 2071-2100 summer mean precipitation under A2 scenario with 19 simulations



b) BMA-EQ approach: 2071-2100 summer mean precipitation under A2 scenario with 19 simulations



c) BMA-SS approach: 2071-2100 summer mean precipitation under A2 scenario with 19 simulations

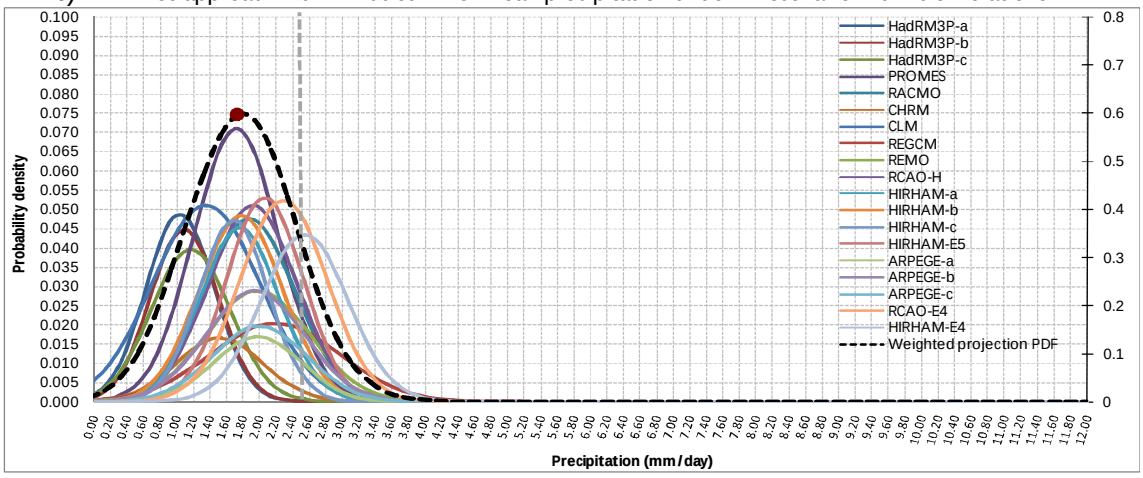


Figure 7.32: Projections for 2071-2100 summer mean precipitation using a) AEM, b) BMA-EQ and c) BMA-SS approaches and 19 simulations. For a) and b), individual likelihood PDFs are plotted on the left, to allow them to be distinguished. Red dot denotes most likely projection. Grey dashed line denotes observed mean in the control period.

The majority of models simulate a decrease in precipitation relative to observed mean precipitation in the control period. Individual model projections for this climate parameter tend to converge to a central value (Figure 7.32a) and when likelihood functions are calculated, they tend to be quite confident. This is due to the small simulated standard deviations for this parameter in the 2071-2100 period. All models have a future projection standard deviation of 0.77mm/day or less, indicating that modelled precipitation values fall close to the mean modelled value in these RCMs. Both BMA-EQ and BMA-SS weighting produce a slight decrease in summer mean precipitation under the A2 scenario (Figures 7.32b and c), although the relative influence of the various ensemble members varies depending on the weighting used.

7.6.8 Summer (JJA) mean precipitation: B2 emissions scenario

Projections of future summer precipitation for the B2 emissions scenario using 8 simulations are given in Figure 7.33. Again, results highlight the complexity of determining which models are most skilful in simulating an altered climate state. Consider RCAO-E4 and HIRHAM-E4: RCAO-E4 simulates 0.45mm/day or 13.5mm/month less precipitation under A2 forcing than it does under B2 forcing, while HIRHAM-E4 simulates 0.27mm/day or 8.1mm/month less. These responses are quite different, yet their skill scores for this parameter in the control period do not indicate that either model is significantly more skilful than the other.

Both BMA-EQ and BMA-SS weighting produce a slight decrease in summer mean precipitation under the B2 scenario (Figures 7.33b and c), which is smaller than the decrease calculated for the A2 scenario. Again, the relative influence of the various ensemble members varies depending on the weighting used. Most noticeably, PROMES becomes much more influential under BMA-SS weighting, due to its high spatial r score of 0.84 in the control period.

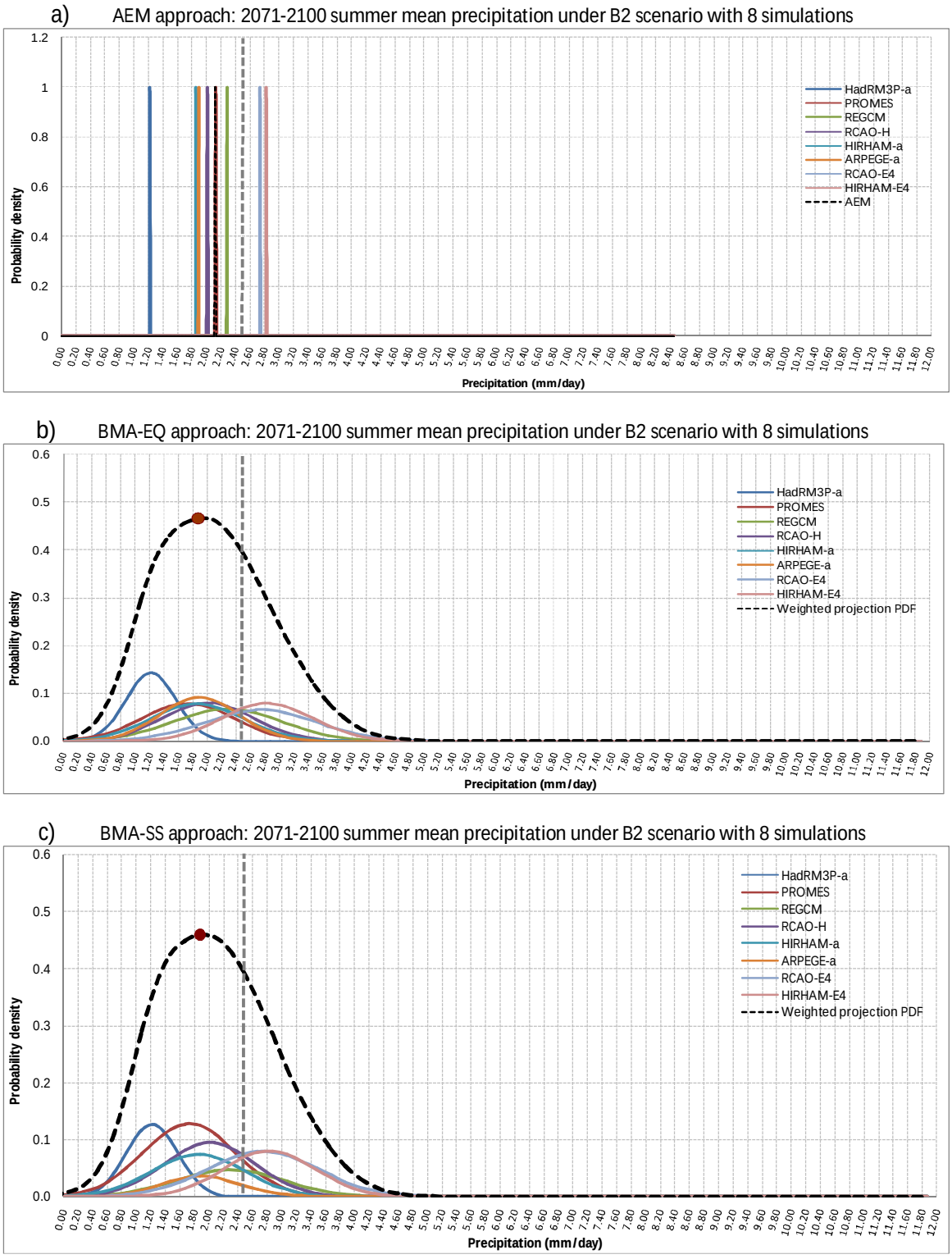


Figure 7.33: Projections for 2071-2100 summer mean precipitation using a) AEM, b) BMA-EQ and c) BMA-SS approaches and 8 simulations. Red dot denotes most likely projection. Grey dashed line denotes observed mean in the control period.

7.7 DISCUSSION AND CONCLUSIONS

In this chapter, a systematic framework for the development of future climate scenarios was proposed which take into account information about different aspects of model skill. The results of the skill assessments presented in earlier chapters were used to inform the ensemble generation process. The BMA-EQ approach assumes that all models are equally skilful or could potentially be equally skilful under a future forcing scenario, given that a model's performance under a difference forcing scenario may be more as well as less skilful compared to the control period. The BMA-SS approach assumes that skill in the control period is indicative of skill at modelling future climate states and weights projections according to the skill scores obtained from an assessment of the model control simulations. The BMA-NAO approach utilizes objective estimates of model skill at simulating the NAO, having identified that the skill score-based assessments did not detect deficiencies in the simulation of large-scale climate drivers. Finally, the BMA-COM approach recognizes the potential value of both skill scores and further analysis of the large scale drivers and weights models using a combination of the two assessment techniques.

The first section of results in this chapter focused on the impact of using different weighting schemes based on varying levels of information about model performance. To assess the impact of different weighting schemes, ensemble projections were generated using the five case study models assessed in Chapter 6, so that the information gained about their representation of the NAO could be incorporated into the Bayesian Model Averaging technique. Under the A2 emissions scenario, the mean ensemble PDF for both winter and summer temperature changes little when different weighting schemes are used. However, the relative contributions of the ensemble members underlying that mean projection do vary (Figures 7.4 and 7.10).

An important finding is that under the B2 scenario, the mean ensemble PDF changes considerably for both winter and summer temperature, when different weighting schemes are applied (Figures 7.7 and 7.13). In summer especially, the BMA-SS approach of weighting based on aspects of performance in the control period results in an approximately normal curve with a single peak. However, when

weights are employed that reflect the ability to capture the effects of the NAO in the control period (BMA-NAO), the resulting ensemble PDF is bi-modal. In this instance, the models which perform best at simulating the NAO model a very different distribution to the models that perform skillfully at simulating the mean seasonal spatial pattern.

Interestingly, in the control period the models were largely skilful at simulating the domain-wide mean summer temperature value, with modelled values falling within a 1.2^{o}C range, but less skilful at simulating the actual spatial patterns that underlie that mean. However, under the A2 emissions scenario, individual model average projections are spread over a 2.6^{o}C range. The increased divergence of mean projections for the 2071-2100 period suggests that the spatial bias patterns that gave rise to a skilful mean in the control period do not remain constant through time.

The projections for 2071-2100 precipitation were also calculated. Again, the mean projection for both winter and summer precipitation under the B2 scenario changes considerably when different weighting schemes are used. While there is little difference in the shape of the ensemble PDF under the A2 emissions scenario, the relative contribution of the individual ensemble members underlying that mean projection varies when different weighting schemes are used. As such, there is uncertainty attached to these projections that is not immediately apparent.

In summer, precipitation projections are affected to a greater extent by varying the weighting scheme. As such, when different weighting schemes are used, different models emerge as the most influential in the ensemble. Due to the differences in their projections, this variability greatly alters the shape of the ensemble PDF. Under the A2 emissions scenario, varying the weighting scheme impacts the overall likelihood and the shape of the lower tail of the precipitation distribution (Figure 7.19). Using the BMA-EQ approach, the lower tail of the precipitation ensemble PDF under A2 forcing is heavy, meaning that a greater likelihood of occurrence is attached to these extremely low values. Using the BMA-NAO approach, the lower tail of the precipitation ensemble PDF is considerably lighter. As such, there is uncertainty over the likelihood attached to extremely low levels of precipitation in summer. Although these values are not the most probable, having much lower likelihood attached to them than the most likely projection, such

low levels of summer precipitation would have considerable impacts associated with them if they were to occur. Therefore, it is in the interests of robust climate planning to take these values into consideration.

Under the B2 emissions scenario the ensemble precipitation PDF varies with respect to both shape and maximum likelihood when different weightings are used (Figure 7.25). Under BMA-NAO and BMA-COM weighting, the distribution is approximately normal while under BMA-EQ and BMA-SS weighting, the distribution is skewed to the left. Projections for the B2 forcing scenario have a greater degree of uncertainty attached to them, particularly with regards to the upper extremes of the precipitation distribution, as the choice of weighting scheme introduces another layer of variability into the climate modelling process. Again, this creates the potential for significant mal-adaptation to occur if inappropriate skill metrics are employed or uncertainties are not suitably quantified.

These findings highlight the need to move towards more comprehensive weighting approaches that incorporate information about model skill in areas beyond the mean climate state. The research presented in previous chapters has already demonstrated that skill score-based assessments of the mean climate state may not identify deficiencies in the simulation of large-scale climate drivers. As such, formulating future climate scenarios based on skill scoring assessments was found to be an unreliable approach and techniques which incorporate a fuller picture of model skill into the formulation of projections are required.

The effect of variation in the weighting scheme is not always visible in the mean ensemble PDF, which highlights another important point. Even when the choice of weighting scheme does not change the shape of the ensemble PDF considerably, the data underlying the projection may be altered as different models become the key contributors to the projections. When such uncertainties are associated with the underlying data, it is highly important that these uncertainties are characterized and communicated, even if the mean projection is not largely affected. For example, the confidence that might be attached to a ensemble PDF which is dominated by a single RCM projection would be very different to the confidence that might be attached to a ensemble PDF in which all RCM converge and contribute equally, yet the overall shape of the ensemble PDF could be very similar.

As such it is important, when providing information for decision-makers, to consider not only the ensemble PDF but also how that PDF is formed. Rather than placing emphasis on the climate scenario as an end product, scenario development should be viewed as a process in which each step merits understanding and clarity. Not only is this in the interests of robust decision-making, it is in the interests of scientific credibility to ensure that any ensemble averaging technique offers a genuine increase in reliability rather than masking uncertainties in the data.

The impact of ensemble size was also examined and while ensemble size has little impact on the A2 climate projections for temperature, incorporating the full suite of model projections into the B2 scenario further dampens the potential bi-modal shape. This is due to the abundance of models driven by the HadAM3H AGCM. Similarly, the addition of extra models smoothes the B2 summer precipitation ensemble PDF, making it approximately normal in shape.

As illustrated in earlier chapters, while RCMs driven by the same GCM vary in terms of spatial patterns, their temporal evolution over the simulation period is very similar, leading to a lack of independence between model projections. The application of ensemble techniques relies on the assumption that each simulation represents a different potential future outcome and that be including more projections, more of these potential future climates are sampled. When there is a lack of independence in the projections, the same outcomes are being sampled repeatedly, leading to over-confidence rather than increased reliability and this is illustrated quite effectively by these results.

There are still benefits to be had from expanding ensemble size, as including more potential futures in an ensemble naturally increases the level of confidence attached with the ensemble projection. However, ensemble size should not be increased at the expense of independence. The addition of models that are not independent of each other does not result in greater sampling of potential future outcomes but only serves to falsely increase confidence in the shared characteristics of the projection.

The independence of ensemble members is not often assessed or discussed in climate modelling studies, yet it has a large effect on how robust the ensemble projection actually is. If an ensemble projection is overly-influenced by one family of

models or one GCM driver and problems are later found within those models, the reliability of the projection as a whole is subsequently compromised. If ensemble members are independent, the effects of individual model errors are isolated and can be accounted for. There is a clear argument for placing a greater focus on model independence and for including an assessment or discussion of model independence in climate modelling studies that utilize ensemble approaches.

CHAPTER 8
FINAL CONCLUSIONS

RCMs have the potential to be a very useful tool to inform adaptation decision-making and test subsequent adaptation strategies. Yet RCMs cannot model climate precisely due to the inherent complexity of the climate system. If uncertainties in climate change projections are left unaccounted for, it is likely that any subsequent decisions made may lead to mal-adaptation. Therefore, improved information about the relative strengths and weaknesses of various models is essential and such information needs to be quantified and incorporated into ensemble projections of future climate

The focus of this thesis was to develop a systematic framework for the construction of future scenarios utilizing regional climate models, which takes account of model uncertainty. To evaluate model skill and identify sources of uncertainty, the RCMs' ability to simulate key aspects of the climate such as means and variability were initially assessed. The models were then assessed in their ability to represent the underlying large-scale climate dynamics, to determine whether skill in simulating the mean climate state is a reliable indicator of model performance.

By identifying the spatial and temporal scales at which different models are informative and determining whether model skill is genuine or a result of error cancellation, ensemble projections were then constructed which take account of uncertainties stemming from model variability (Figure 8.1). Deterministic approaches such as the Arithmetic Ensemble Mean (AEM) incorporate no information about model skill or the likelihood associated with individual members in the ensemble and as such, do not provide a reliable basis for testing adaptation strategy. For example, if one model in an ensemble projects an increase in precipitation and another model in the ensemble projects a decrease in precipitation, this presents an issue for adaptation decision-making as the optimum strategies for such projections are very different. However, when information about the likelihood associated with each projection is included, projections can provide a more robust basis for decision making.

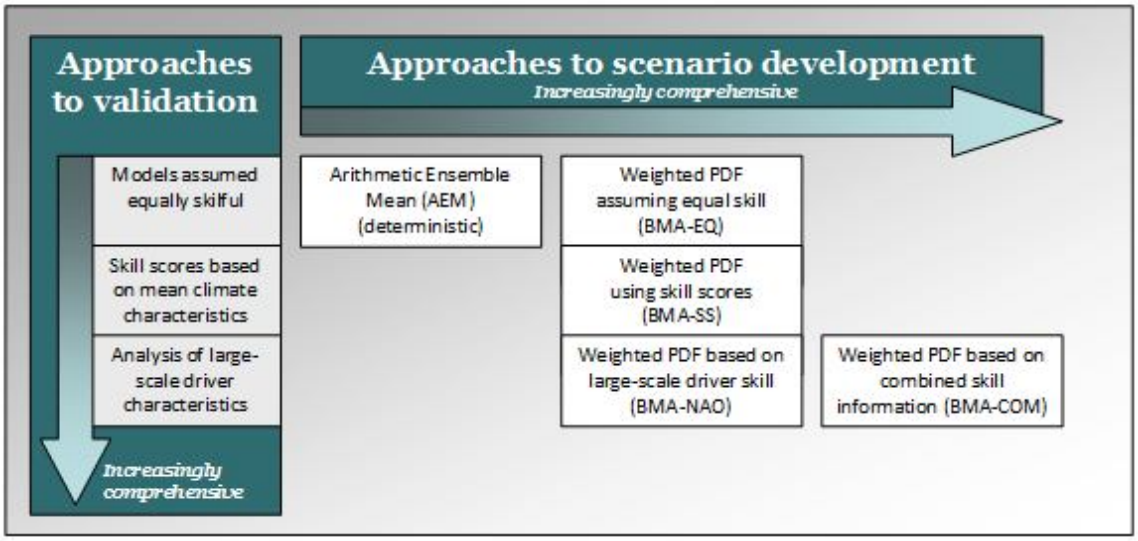


Figure 8.1: Schematic diagram of approaches to scenario development

When skill-scores such as the spatial and temporal r scores are used to weight individual ensemble members using a Bayesian Model Averaging approach (BMA-SS), the ensemble PDF is also potentially unreliable as the findings from Chapter 6 have shown that such skill scores may not reflect the skill of the model in simulating large-scale drivers of climate. Weighting models based on their ability to capture large-scale drivers (BMA-NAO) offers a more reliable approach. The optimum approach suggested by this research is to combine traditional skill scores with analysis of the large-scale drivers and weight individual ensemble members using both forms of skill information (BMA-COM). Looking forward, there may be opportunities to further refine this methodology by including other forms of skill assessment in the Bayesian approach to create future climate scenarios that have a greater level of confidence associated with them, which can aid decision-making and adaptation.

Following on from the work presented in this thesis, a number of recommendations can be made that have the potential to make RCMs a more reliable tool for decision-making and enable robust adaptation decisions to be developed and tested. This chapter summarizes the key findings and recommendations of this thesis and highlights areas that could benefit from further research.

8.1 SUMMARY OF RESULTS

Throughout this thesis, a range of regional climate models have been assessed based on their ability to simulate aspects of Irish climate in the control period 1961-1990. Spatial and temporal characteristics of two key impact variables, temperature and precipitation, have been assessed, along with the model-simulated North Atlantic Oscillation. Table 8.1 summarizes the findings of the skill assessments carried out.

Across the full range of metrics applied to both temperature and precipitation, and considering also the representation of the simulated NAO, the most skilful model for simulating the Irish domain is HIRHAM-E5. This model possesses moderate to high skill in almost all metrics, with the exception of wintertime interannual variability. As interannual variability has been shown to be highly dependant on the choice of driving GCM, it is important that GCM skill and GCM-RCM interaction be considered when choosing a GCM-RCM configuration.

RCAO-E4 is also a skilful model in many regards, particularly in the simulation of NAO behaviour and impacts on regional climate. It models seasonal spatial patterns for both temperature and precipitation with skill in all seasons. Although it exhibits systematic bias on the seasonal means, this kind of bias is much easier to account for than randomly occurring biases. The model is less skilful at simulating patterns of interannual variability. As this aspect of the climate is governed by GCM choice, coupling the regional model RCAO with the driving combination used by HIRHAM-E5 may potentially result in a more skilful simulation.

As the analysis delved further into the data, errors emerge that are not apparent from an initial overview. For example, all models simulate the annual climatology for 1961-1990 with very high skill for temperature and less skill for precipitation.

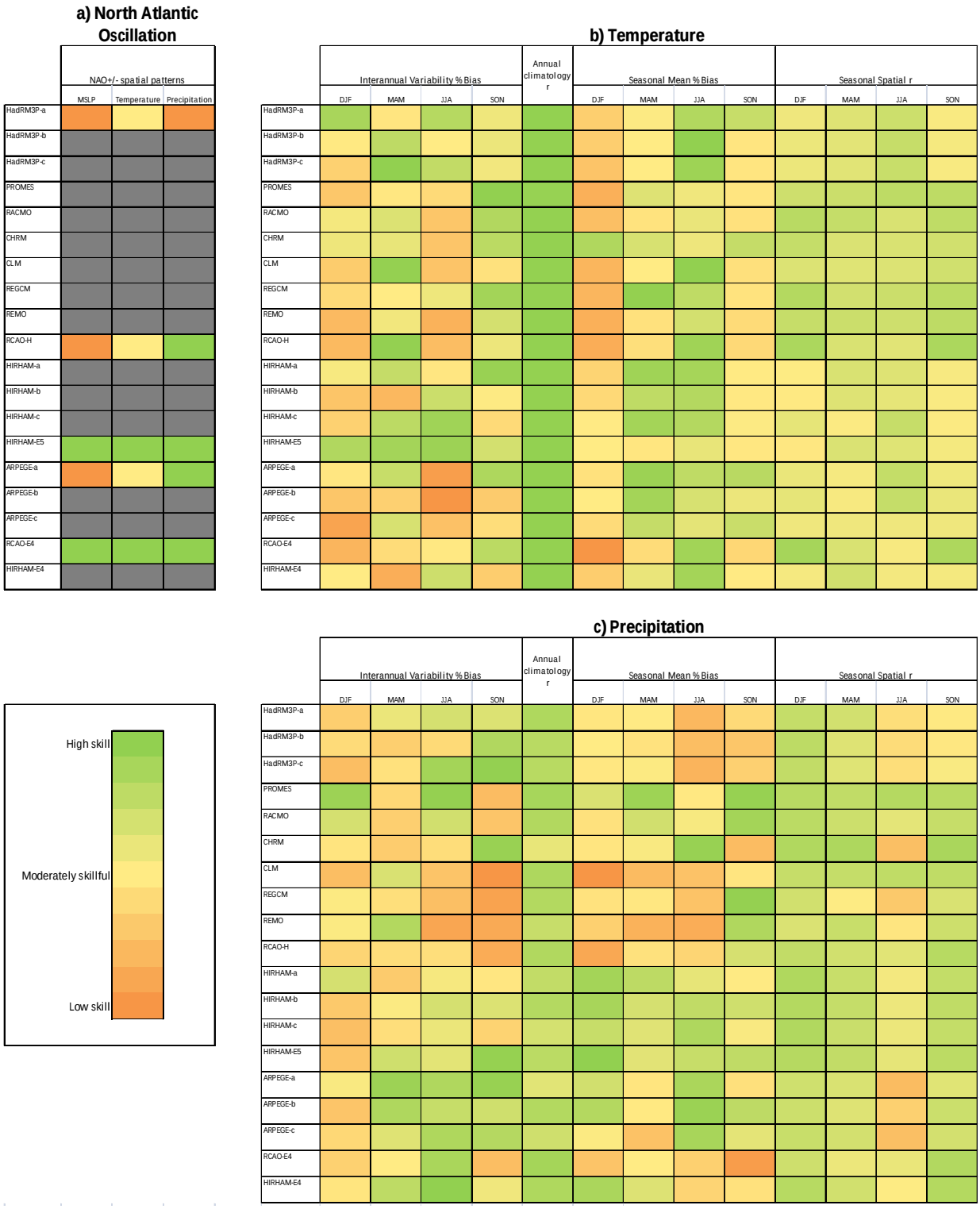


Table 8.1: Summary of model skill assessments showing (a) skill in simulating the North Atlantic Oscillation, (b) temperature metrics and (c) precipitation metrics.

When the analysis is undertaken on a seasonal basis, biases emerge in particular seasons and when the seasonal spatial data is examined, diverse spatial errors patterns emerge. These findings highlight the critical importance of scale in model verification. Even if the model output is to be used at a national scale, it is important to assess model skill on a range of spatial and temporal scales as a model may appear skilful at the national scale due to error cancellation at a grid-scale level. For example, all simulations in the ARPEGE sub-ensemble simulates summer mean precipitation with a low overall bias, yet simulates the spatial pattern of summer precipitation with very low skill, indicating that the skill these models display in simulating the summer mean precipitation is actually derived from error cancellation in the summer spatial data.

Additionally, skill in simulating mean climate patterns may not be a robust indicator of model performance. HadRM3P-a displayed particularly low skill in simulating the effects of the NAO on regional winter climate. Yet it captured the mean spatial pattern of winter temperature and precipitation with skill. Evidently, a model may simulate the correct mean climate state for the wrong reasons, failing to capture the climate dynamics which underlie the mean state.

As such, it is important when assessing climate model projections to look not only at the level of agreement between modelled and observed parameters, but also at how that skill arises. If a model is to provide a credible, robust basis for informing policy, it is not sufficient for it to simulate the current climate with skill; it must also be skilful for the right reasons. Such an assessment needs to quantify not only how modelled climate parameters such as temperature and precipitation compare to the observed, but also how skilfully the dynamics of the climate system are represented in a model.

Given the range of skill displayed by the various models, it is essential to incorporate a measure of model performance into ensemble generation techniques. However, as skill scores do not provide a comprehensive picture of model skill, it is necessary to move past ensemble generation techniques which rely on this type of skill assessment alone. Chapter 7 presented a systematic framework for the construction of future scenarios which takes account of both spatio-temporal skill scores and skill in simulating the large-scale dynamics of the climate. By adopting a

combined weighting approach in which both skill scores and the ability to simulate large-scale dynamics are taken into account, uncertainty associated with model skill may be accounted for more fully, resulting in a more robust climate projection.

Ensemble PDFs generated using NAO skill information (BMA-NAO) and the combined weighting approach (BMA-COM) were compared to ensemble PDFs generated using the assumption of equal likelihood (BMA-EQ) and spatio-temporal skill scores alone (BMA-SS) (Figures 7.4, 7.7, 7.10, 7.13, 7.16, 7.19, 7.22 and 7.25). In several cases the choice of weighting scheme significantly affects characteristics of the mean ensemble PDF. The likelihood associated with the most probable projection, the shape of the distribution tails and even the shape of the entire PDF can be affected by this choice.

Planning for future climate scenarios may involve the building of costly infrastructure such as reservoirs and water pipelines and when there is uncertainty surrounding the level of change, there is the potential for under- or over-adaptation to occur. As such, the choice of ensemble generation method is an important consideration with implications for the decision-making process.

8.2 ISSUES FOR IMPACT ASSESSMENT

8.2.1 Model development and validation issues

The results of the skill assessments carried out in earlier chapters and summarized in section 8.1 illustrate that assessments of the mean climate state may not reveal deficiencies in the simulation of the large-scale climate. Similarly, models that display significant systematic biases in the control period may capture the patterns of large-scale variability quite well. For example, while ARPEGE-a simulated the mean climate state with skill, it failed to capture certain regional climate patterns associated with NAO positive and negative phases. Conversely, RCAO-E4 modelled considerable systematic temperature, precipitation and MSLP biases in the mean winter climate, but captured the effects of NAO variability with skill. Evidently, a large bias is not necessarily indicative of a less reliable model and more importantly, little to no bias does not always guarantee a skilful simulation. In light of these findings, a key question is how to identify a skilful model.

Assessments of the mean climate state are a less reliable way of determining skill. As data is averaged, errors in the data underlying the average value can be masked, as is the case when overestimation of a climate parameter in one part of the domain combines with underestimation in another part of the domain to create an apparently skilful average. When skill is derived from error cancellation rather than genuine ability to model the dynamics of the climate system, it cannot be depended upon to remain constant through time and under different forcing scenarios. As such, the assumption that the relative difference between the control and future period is the signal of climate change, an approach which is often applied in climate change studies, is inherently flawed. Knowledge of the bias patterns and the nature of errors in the control period can enable the reduction of systematic bias in the future period, though some uncertainty still remains as the bias patterns diagnosed in the control period may change over time. However, when a model exhibits random bias rather than systematic bias in the control period, it becomes more difficult to account for the uncertainty associated with such a model.

When a model is assessed based on its ability to simulate a mean state, no information is obtained about how the model arrives at that mean state. The model may generate the right answer for the wrong reasons and as long as skill assessment focuses on the mean climate state rather than the process through which that simulation is generated, the true predictive skill of the model might never come to light. Though there is no way to have full confidence in a model's ability to simulate a future climate state, assessing models on their ability to capture the dynamics of the climate system is likely to be a more comprehensive approach to skill assessment. Such an approach focuses on how the model captures the dynamics rather than the mean state of the climate. Of course, a model that simulates the dynamics of the climate system and the effect of large-scale drivers with skill may be less skilful at simulating the mean climate state, but it is much more preferable for a model to simulate the dynamics of the climate system with skill and model a biased mean climate state than model the dynamics of the climate system incorrectly and model an unbiased mean climate state.

RCMs are dynamical models of the climate system and their main advantage is that they are supposed to simulate the response of the climate system to changes in forcing in a physically consistent manner. Therefore it is only logical that verification

focuses not on the mean climate state but on whether the model is approximating the dynamics and interactions of the real climate system.

8.2.2 Ensemble methods

Unlike many other climate modelling studies, this thesis has focused less on the actual scenarios generated and more on the uncertainty associated with them. Arithmetic Mean Ensembles (e.g. Gates *et al.*, 1998; Rinke *et al.*, 2006) are limited in their usefulness by the lack of probability associated with them. Therefore, probabilistic approaches using various weighting schemes were compared.

The future scenarios generated in Chapter 7 demonstrate that using an "equal likelihoods" weighting scheme, which allows randomly biased models and skilful or systematically biased models to contribute equally to an ensemble projection, can produce the same mean future projection as a skill-based weighting scheme depending on the parameter and season in question. However, in some cases, the weighting scheme chosen has a noticeable effect on the shape of the ensemble PDF. The choice of which weighting scheme to use is somewhat subjective, but the least reliable must be the equal weighting scheme as it contains no information on the skill of ensemble members.

When a model that is randomly biased is used in ensemble generation, the only way such a model can contribute to the overall skill of the ensemble is through error cancellation. These findings support the view of Hagedorn *et al.* (2005) that a large part of the ensemble's superiority is due to error cancellation and suggests that the use of different models and the increased ensemble size is not as significant a factor, as proposed by Doblas-Reyes *et al.* (2000). This kind of "skill" is not due to genuine predictive ability and ensembles generated in this manner are potentially unreliable, as the error patterns in different models may change over time and may not cancel each other so effectively in a future climate simulation. By using a weighting scheme that penalizes random bias, the influence of these models can be constrained.

Weighting schemes based on skill scores, such as the REA method (Giorgi and Mearns, 2003; Tebaldi *et al.*, 2004) offer an improvement, but as already

discussed, skill score-based assessments of the mean climate state do not offer information on the ability of the model to simulate the climate system as a whole. Incorporating information about the skill of the model in simulating large-scale drivers and patterns of variability, which can be accomplished using the Bayesian Model Averaging approach, offers a methodology for further constraining models that lack predictive skill, allowing models that perform well at simulating the dynamics of the climate system in the control period to contribute most to the future ensemble PDF.

It is also worth noting that even when the mean ensemble PDF is similar using different weighting methods, often the data underlying the mean ensemble PDF varies considerably, posing a question not just of ensemble reliability but of scientific credibility. It is not sufficient to communicate the final outcome, the ensemble PDF, to climate decision-makers without also including information on the data underlying that outcome. When projections generated using different techniques converge, the natural response is increased confidence in the projections. Therefore it is important that the climate scientist communicates the underlying uncertainties and internal variability of the ensemble projection, to minimize the potential for overconfidence.

Climate scientists and decision-makers are largely aware of the concept of cascading uncertainty in climate impacts assessment (Henderson-Sellers, 1993; Jones, 2000b; Mitchell and Hulme, 1999), in which uncertain emissions concentrations, GCM variability and RCM variability lead to uncertainty in climate impacts and the methods by which modelled data is transformed into climate scenarios and impacts assessments are another procedure in that chain of uncertainties and inferences. As such, variability in ensemble generation methods is a source of uncertainty that must be accounted for. Climate projections cannot be presented as definitive outputs and must be understood to be one plausible outcome from a chain of subjective decisions.

The projections also highlight the importance of model independence when generating climate scenarios. As illustrated in Chapters 4 and 5, models driven using the same GCM produce very similar simulations of interannual variability, which could result in overconfidence in the variability projected by an ensemble, if the

ensemble is dominated by models with the same GCM driver. To comprehensively assess the independence of ensemble members, other factors should also be taken into account, such as key parameterizations. Additionally, the sensitivity of model output to key parameterization schemes should be assessed, to determine how significant a factor parameterization choice is for various climatological processes and identify those parameterizations that are of critical importance to independence. It may in fact be inadvisable for different modelling institutions to share code and parameterizations, as this limits the independence of model simulations and in turn, limits the range of possible futures that can be sampled.

8.2.3 Robust decision making

A key question that emerges from this research is how to optimize the usefulness of regional climate models and simultaneously quantify the uncertainty associated with their outputs in order to reduce the potential for mal-adaptation.

Certainly, treating uncertain projections as "predictions" of future climate is a flawed approach and this kind of deterministic thinking is best avoided. As such, the "top-down" approach to developing climate adaptation strategies, in which scenarios from climate models provide input on the climate impacts to be accounted for in adaptation strategy is an unsatisfactory approach to adaptation. Such an approach ultimately leads to poor decision-making where model outputs are conflicting or contradictory, with no associated likelihoods. Given the uncertainty associated with climate model outputs, founding climate policy and adaptation strategies on model outputs alone is inadvisable and it is especially important that a single model is not used as a basis for decision-making. Given the range of different projections generated by different models, utilizing a single model would lead to a large degree of uncertainty and a high potential for mal-adaptation. Additionally, given that communities are not homogeneous in terms of exposure to vulnerability, such as age and socio-economic status, individuals and groups within a community are likely to be affected to varying degrees by climate impacts (Yamin *et al.*, 2005). Therefore, the diverse needs the community must take a central role in determining adaptation strategy (Smit and Wandel, 2006). However, adopting the "bottom-up" approach to adaptation is not without limitations either, as formulating adaptation measures based

on the vulnerability and adaptive capacity of a community does not address the issue of whether the strategies implemented will be sufficient to withstand the impacts of a projected change in climate.

Perhaps the optimum approach combines ideas from both top-down and bottom-up thinking. Decisions for future adaptation cannot be made without information about likely future climate, but if model outputs are considered as "predictions" of climate, without any regard for the associated uncertainties, this may lead to over or under-adaptation. An alternative approach would be to use climate model outputs to test the robustness of a range of adaptation measures to different degrees of climate change, to identify measures that are likely to be beneficial under a range of potential future scenarios. In the words of Rummukainen (2010):

> "RCMs are not a panacea, but a tool in the arsenal of climate science."

Uncertainty is likely to always remain a part of modelling climate impacts because while new research increases our understanding of the climate system, aiding in the development of climate models, there is also the potential for new research to reveal previously unknown processes and interactions, which would then need to be accounted for in models also. As such, rather than attempting to reduce uncertainty through research before adaptation decisions can be made, strategies are needed that enable robust decision-making even in the presence of uncertainty.

As climate change is an on-going concern, in order for decision making to be robust, it must also be an on-going process, as suggested by Baer and Risbey (2009). Correspondingly, projections from climate models should not be treated as static information but considered as subject to change and further refinement, as our understanding of the climate system and our ability to model it increases., Vulnerability should be reassessed at regular intervals and adaptive measures re-evaluated using the most up-to-date data, to ensure that strategy evolves and adapts to emerging environmental hazards, such as climate change impacts.

8.3 FUTURE PERSPECTIVES

Following on from the work presented in this thesis, various recommendations can be made for further refining climate models and for utilizing them responsibly in decision-making.

8.3.1 Model development

Where a model lacks predictive skill, only further development of the model dynamics and refinement of the parameterizations can result in a genuine improvement in model skill. In light of the errors identified in this study, the representation of large-scale climate drivers particularly stands out as an area where model development could potentially be focused to improve overall skill.

As several models failed to capture the pressure, temperature and precipitation patterns associated with the North Atlantic Oscillation, further study of the large-scale drivers of regional climate and their representation in models is clearly a key research area. The variability and effects of the NAO has been the focus of much research (e.g. Hurrel, 1995; Lamb, 1987; Rodwell *et al.*, 1999) yet the analysis presented in Chapter 6 suggests that representing this information in a realistic simulation of Northern European climate remains a challenge.

In some regions, the nature of the large-scale circulation patterns and drivers themselves are still being investigated, making their inclusion in climate models even more challenging. For example, due to the complex orography and the location of the Alps in a transition region between the Atlantic Ocean, the Mediterranean Sea and the European continent, links between Alpine climate variability and large-scale circulations are still being explored (e.g. Efthymiadis *et al.*, 2007; Quadrelli *et al.*, 2001). Additionally, the response of these large-scale drivers to anthropogenic forcing needs to be considered.

As such, further research on the large-scale dynamics that influence the climate of a region may bring about important improvements in regional climate modelling. Models which capture the processes and dynamical properties of the climate system with genuine skill would provide a much more robust platform for subsequent decision-making and strategy testing.

8.3.2 Model verification

A framework for robust model assessment is required, which distinguishes genuine skill from skill derived from error cancellation. A variety of metrics have been applied in this thesis for the purpose of assessing model skill, but the most insightful approach has been the investigative approach adopted in Chapter 6 to examine the model-simulated NAO.

This is not to say that skill scores or assessments of the mean climate state do not have a place in model validation studies. These approaches provide a quick and comparable indication of skill across multiple models and form a quantitative basis for weighting model projections. However, that basis is only robust if the skill levels indicated by the scoring approach actually reflect the level of predictive skill the models possess. This is where a more in-depth analysis may prove more helpful.

As large-scale modes of variability play such an important role in shaping regional climate patterns, it would be remiss not to include them in an analysis of model skill. One of the ways in which climate change may manifest is through a change in the variability or patterns associated with such climate modes, therefore if there is to be confidence in a model's ability to simulate the future climate, it is desirable that the model capture such effects in the control period. This thesis took the example of the North Atlantic Oscillation as an important large-scale driver of climate over Western Europe. Similarly, in a regional climate model being used to simulate North American, South American or Asian climate, an ability to capture the effects of the El Niño-Southern Oscillation would be an important measure of skill.

Yet even if a model captures the mean state and dynamic properties of the current climate quite well, there is no guarantee of its skill under different forcing conditions. Skill in simulating paleoclimates may offer a means of assessing ability under different forcing conditions and that information could also be incorporated into the Bayesian Model Averaging approach. Looking again beyond the mean climate state, the ability of models to capture the frequency, magnitude and duration of various extreme events may offer another more rigorous skill metric. Given the potential for climate change to exacerbate the already adverse effects of extreme

events, it is especially important that climate models capture their characteristics skilfully.

8.4 LIMITATIONS OF RESEARCH

Although this thesis has aimed to account for a greater portion of the uncertainty attached to model projections, utilizing a BMA approach to weight model outputs using different skill metrics and further refine projections, there is still potential for mal-adaptation if the underlying assumptions on which these projections are based are not understood.

The focus of this thesis has been regional climate model variability and skill, but as discussed in previous chapters, the RCMs are part of a chain of processes and inferences. At the outset, the choice of emissions scenarios used would lead to different projections. The PDFs produced are a set of possible scenarios based on the A2 and B2 emissions scenarios, but there are other potential scenarios that give rise to different levels of anthropogenic emissions, and these potential futures are not accounted for.

This thesis has highlighted that different metrics of skill may give different results for a selection of models. Models may be skilful in one statistic of climate, but lack skill in another aspect. This thesis has aimed to provide a comprehensive skill assessment; however it also acknowledges that the choice of which metrics to use to measure model skill is a subjective choice. Variations in skill scores may occur when a different set of metrics is applied and this could impact the priors used in the BMA approach, which could in turn yield different projections to those presented.

Measures of model skill are also conditional on the specific boundary data, GCM and RCM approach. If the same RCMs are driven by different GCMs or forced using different observational data sets in the control period, the resulting simulations may display more or less skill that those analysed in this study. Similarly, even slight variations in the technical details or parameterizations of a model may yield different results and assessing variations such as these fell outside the scope of this thesis.

Another assumption underlying the BMA approach is the choice of distribution for the modelled likelihood functions. In this thesis, that normal distribution is chosen as the parameters under investigation were assessed using the Shapiro Wilks test for normality and were found to be normally distributed at the interannual timescale in the control period. As such, there is a logical basis for choosing the normal distribution for the modelled likelihood functions. However, users of these scenarios much be aware of this key assumption, as varying the distribution used is likely to lead to different outcomes also.

8.5 APPLICATIONS OF PROBABILISTIC SCENARIOS

In the past, when the observed record provided some indication of future events, return periods could be determined from the observation and this information could be used to inform decision-making. However, under increased anthropogenic emissions, this is no longer a valid option. Large-scale projects such as ENSEMBLES (Weisheimer *et al.*, 2009) and climateprediction.net (Stainforth *et al.*, 2005) have produced a wealth of modelled output at both GCM and RCM scale. However, as there is no single "best" model, a key question is how to utilize output from different models in the most intelligent way.

Various studies have utilized probabilistic methods to combine model output and develop climate impacts assessments, for example Fronzek *et al.* (2010) utilize probabilities of climate change based on AOGCM output to assess the impacts of climate change on palsa (peaty permafrost mounds containing permanently frozen ice lenses) disappearance in Fennoscandia. Similarly, New *et al.* (2007), utilized climateprediction.net data to form impacts assessments for the water resources sector using a probabilistic framework and illustrated that with probabilistic rather than deterministic information, a PDF can be estimated and utilized to calculate risk, for example the risk of high flow events in a particular catchment area, and inform risk-based judgments. Another example is the work of Bouwer *et al.* (2010), which uses flood scenarios and projections of flooding probabilities under climate change along with projections of socioeconomic change and a simple damage model to assess changes in future flood risk under climate change for a case study area in The

Netherlands. Bouwer *et al.* (2010) finds that using loss-probability curves may be a more appropriate approach than using single loss estimates, as such estimates may lead to underestimation of the impact of very high losses.

Where probabilities form a part of the risk assessment and decision-making processes, it is important to understand the uncertainties associated with probabilistic methods. The research presented in this thesis uses the results from a comprehensive skill assessment to weight probabilistic projections of climate change and demonstrates that the different approaches to measuring model performance may yield different weights and subsequent projections. Of course, any weighting scheme adds a further layer of uncertainty to the decision-making process, due to the subjective choices inherent in model skill analysis (Christensen *et al.*, 2010; Kjellstrom and Giorgi, 2010). As such, there are inevitable subjective components to these and indeed any other probabilistic scenarios. Clearly, probabilistic scenarios of climate change derived from the modelled data are a key factor in the impacts assessments and variation in the methods used to formulate probabilistic scenarios may yield different results, making uncertainty in probabilistic scenarios a key issue for the impacts assessment and adaptation communities.

Where climate adaptation requires investment in costly infrastructure, decision-makers are likely to look to models for the relevant probabilities for key variables in future climate scenarios, to provide a similar quantitative basis for decision-making. Yet if that information is based on a flawed model or ensemble of models, then it does not actually make a firm basis for establishing policy or making adaptation decisions. Clearly, an alternative approach to decision-making, which acknowledges the uncertainty of the situation, is required. Given the subjective elements and various assumptions underlying the PDFs of future climate, it is vital that any climate adaptation decisions informed by these scenarios are guided by the framework of robust decision-making. The framework outlined by Wilby and Dessai (2010) for developing climate adaptation options is an example of such a framework, in which adaptation options are formulated based on assessing institutional vulnerability, emphasizing both observed climatic and non-climatic drivers and then tested using narratives of both climatic and non-climatic change (Figure 8.3).

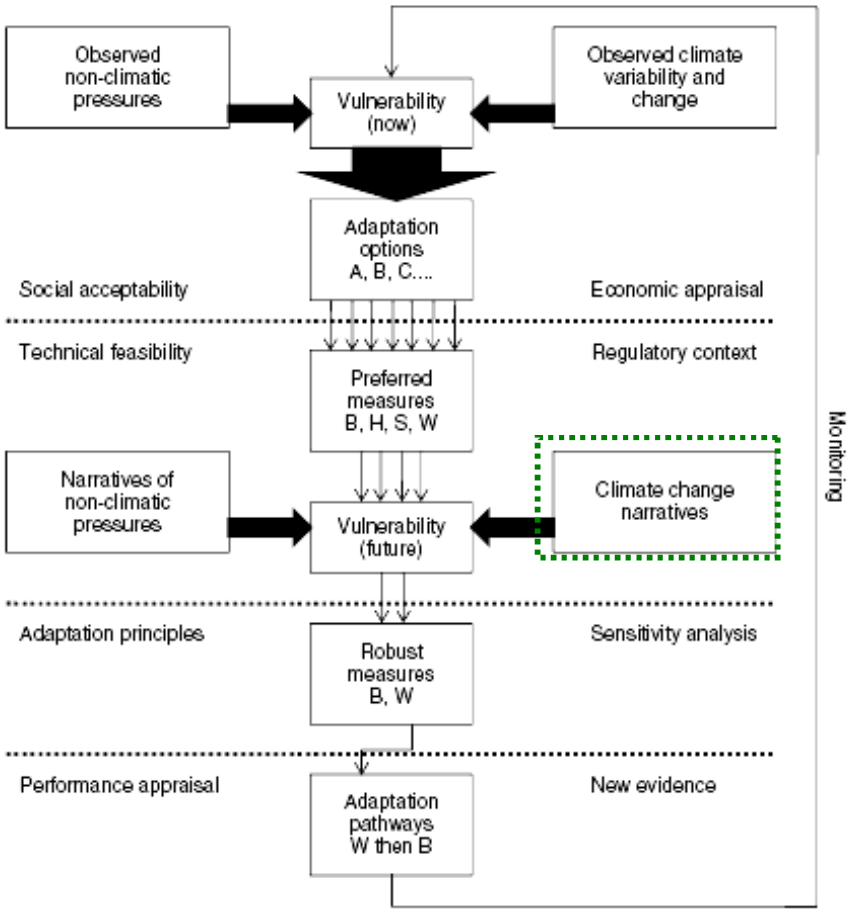


Figure 8.2: Conceptual framework for a robust decision-making approach to climate adaptation planning, illustrating the role of climate change scenarios.(Source: afterWilby and Dessai, 2010).

As such, the probabilistic scenarios of future climate produced in this thesis have very real applications, providing the climate change narratives required to test adaptation options developed through the robust decision-making framework. Rather than applying models to determining what the mean climate state will be towards the end of this century, it is more appropriate to use models to assess which adaptation measures, based on the vulnerability and adaptive capacity of the community, are likely to remain robust under different scenarios of anthropogenic climate change. In the words of Tukey (1962: 13):

> "Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise."

Decisions can be further refined and strategy can be allowed to evolve based on testing using models and periodic reassessments of potential impacts, as certain scenarios become more or less likely.

Given that the ensemble projections presented in Chapter 7 indicate changes in temperature and precipitation under anthropogenic forcing, for all the approaches applied, adaptation measures are certainly required. The question is not whether to act, but how to act, and how climate models can best be utilized in informing those actions. Following the robust decision-making framework, a number of 'low-regret' or 'no regret' adaptation measures, which are beneficial under a wide range of climate scenarios, could be tested against these climate change scenarios and subsequent decisions could be made with regards to implementation. Given that all probabilistic climate scenarios have inherent subjective or conditional components, the potential for mal-adaption can be reduced by adopting 'low-regret' or 'no regret' adaptation measures.

Consider the water resources sector as an example. Lopez *et al.* (2009) found that ensemble approaches provide a better understanding of potential future conditions as they relate to water resources management, compared with deterministic approaches. Therefore, there is much potential for the PDFs of future climate presented in this thesis to be beneficial in relation to impacts assessments for this sector. Ireland is known to be vulnerable to regional water shortages at certain times of the year (Charlton *et al.*, 2006) and the PDFs of future summer precipitation in Ireland show that the most likely outcome is a decrease in precipitation under both A2 and B2 emissions scenarios. As such, there are several potential adaptation measures that could be considered and tested against the climate change scenarios. For example, water efficiency across various sectors could be improved by managing and maintaining pipe and drainage systems more effectively. Additionally, water sources could be better protected from pollution, to maximize the available resources. If measures to improve the utilization of resources are found to be insufficient, further options may be explored. For example, the construction of new reservoirs or desalinization plants might be considered to create new water sources (Arnell and Delaney, 2006).

Climate model projections provide a viable way of assessing potential futures and informing impacts assessment and adaptation strategy. However, as the results of impacts assessments or adaptation strategy testing are inevitably influenced by the choice of modelled data, the choice of probabilistic methods and the inherent subjectivities of these techniques, it is vital that decision-makers are aware of the uncertainties associated with the data. Similarly, the climate modelling community must seek to combine and present modelled data such that much of the associated uncertainty is quantified and accounted for, that the underlying assumptions and conditions of the data are clear and the projections themselves are accessible and interpretable. To optimize the usefulness of regional climate models, an independent, rigorously verified ensemble should be used, weighted to reflect skill at simulating a variety of statistics and characteristics of the climate system.

8.6 PRINCIPAL FINDINGS

- The mean annual climatology cycle of temperature is simulated with high skill by all models. However, models exhibit less skill in simulating the annual climatology of precipitation. The majority of models exhibit a positive temperature bias in winter, while modelling the rest of the seasons with skill.
- Empirical Orthogonal function analysis shows that the majority of models capture the key component patterns of temperature and precipitation, though the percentage variance attached to them may differ from the observed.
- Interannual variability across all seasons is largely influenced by the choice of GCM driver and less so by the choice of RCM.
- Bias patterns for mean sea level pressure in winter are largely influenced by the choice of GCM driver.
- Representation of the effects of NAO positive and negative behaviour on regional climate is a key area of uncertainty, with only two out of the five case study models simulating the expected mean sea level pressure, temperature and precipitation patterns.

- Simulation of wind direction frequency and also the precipitation associated with different wind directions is also a key area of uncertainty.
- Error cancellation is identified as a major source of uncertainty. In addition to spatial errors patterns resulting in a skilful mean, errors in the precipitation amounts associated with different wind directions are also observed to cancel out in certain models.
- Validating models using skill scores and assessments of the mean fields of temperature and precipitation does not detect errors in the representation of large-scale climate drivers.
- Incorporating information about skill in modelling the large-scale climate drivers into future climate projections, rather than weighting projections based on skill-scores alone, can provide a more robust basis for generating future climate scenarios.
- For certain combinations of season, climate parameter and forcing scenario, the choice of weighting scheme has a noticeable effect on the shape of the future ensemble PDF, making this another source of uncertainty.
- For other combinations of season, climate parameter and forcing scenario, the shape of the future ensemble PDF is not changed by applying different weighting schemes, but the relative contributions of the underlying models are altered.
- Model independence is more important than the number of models when generating robust ensemble projections.

8.7 FUTURE WORK

- Further study of large-scale drivers of regional climate, such as the NAO, and their representation in models is a key research area.
- Increased awareness of model independence is critical and the sensitivity of model output to key parameterization schemes should be assessed.

- To increase the confidence associated with ensemble projections, a wider range of model skill information should be systematically incorporated into ensemble weighting schemes. In addition to skill in simulating large-scale drivers, skill in modelling extreme events could also be used to assess skill and weight model output. Skill in modelling paleoclimates could also be assessed, though due to the uncertainties surrounding paleoclimate data and the limited availability of observations, a detailed assessment such as the analysis carried out in this thesis may not be feasible.

- This thesis has presented a framework which accounts for RCM uncertainties in probabilistic future scenarios, however uncertainties remain at the emissions scenario and GCM scale which are beyond the scope of this thesis. While emissions-related uncertainties will always remain an unknowable factor, by applying the same framework to the driving GCMs, uncertainty stemming from GCM errors and variability may also be accounted for in the future scenarios.

APPENDIX A: COMBINED A2 AND B2 FUTURE CLIMATE PROJECTIONS

A.1 WINTER TEMPERATURE (2071-2100)

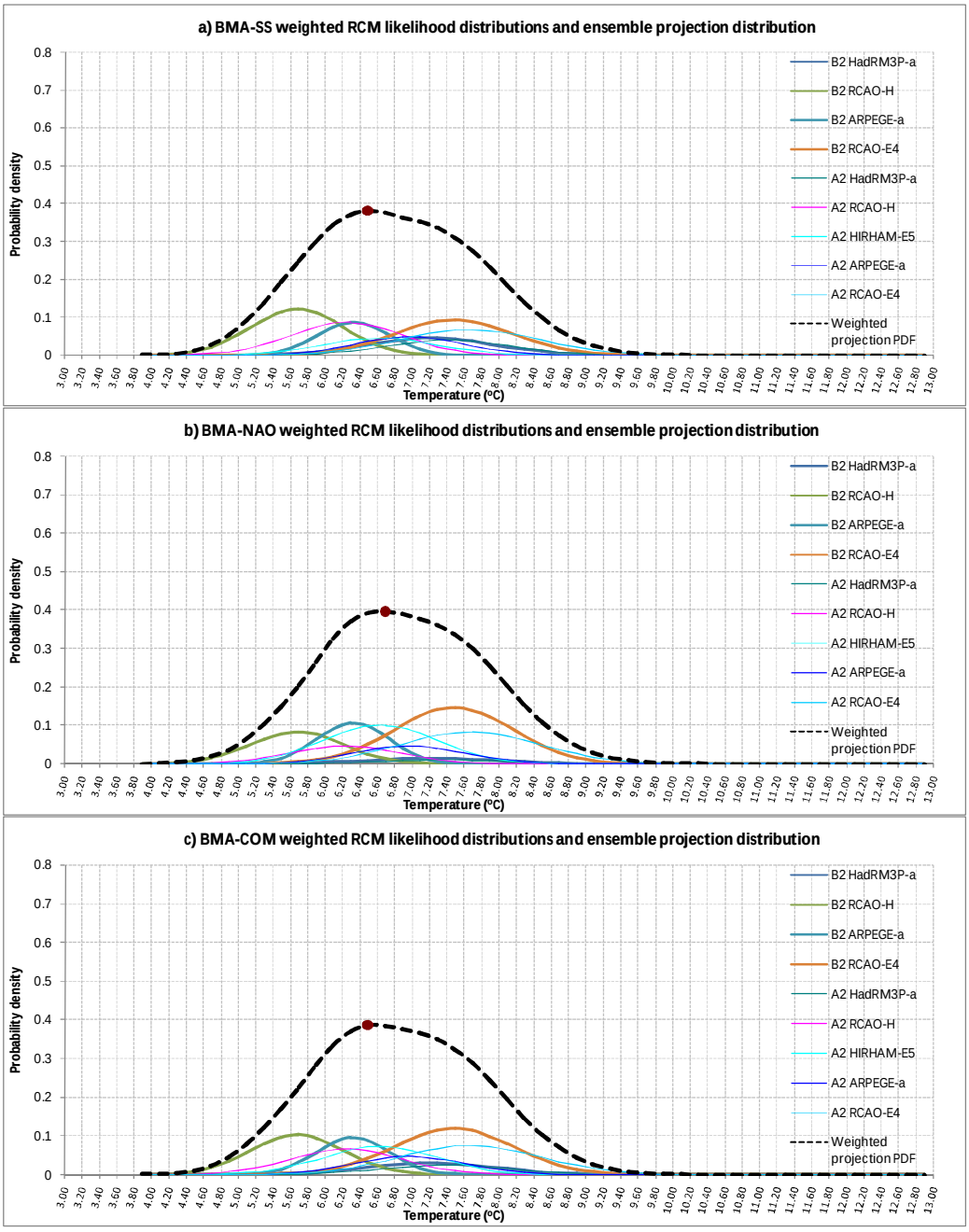


Figure A.1: Projections for winter temperature (2071-2100) using both the A2 and B2 scenarios and 9 simulations. Three different weighting schemes were applied: a) BMA-SS weighting, b) BMA-NAO weighting and c) BMA-COM weighting.

A.2 SUMMER TEMPERATURE (2071-2100)

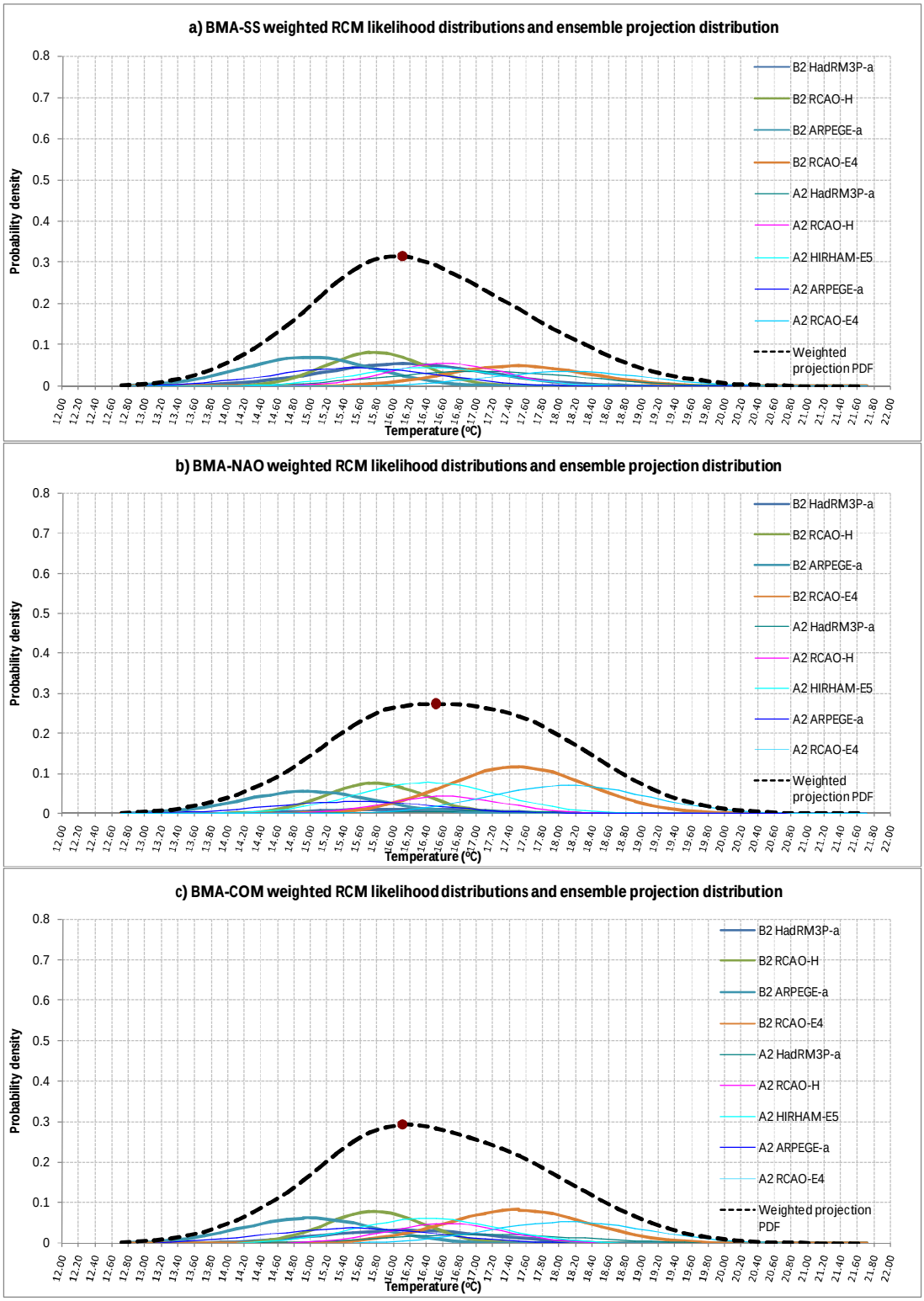


Figure A.2: Projections for summer temperature (2071-2100) using both the A2 and B2 scenarios and 9 simulations. Three different weighting schemes were applied: a) BMA-SS weighting, b) BMA-NAO weighting and c) BMA-COM weighting.

A.3 WINTER PRECIPITATION (2071-2100)

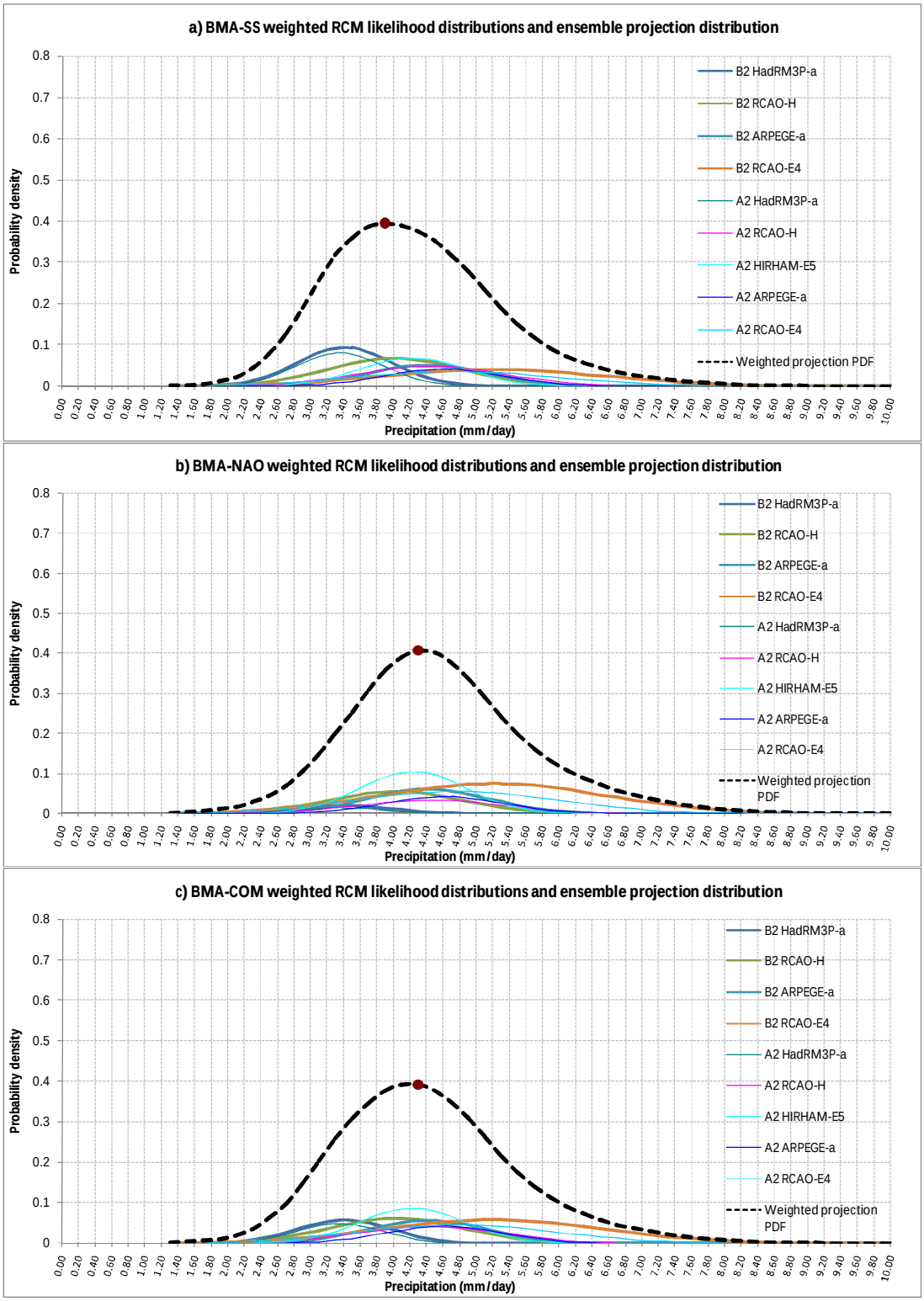


Figure A.3: Projections for winter precipitation (2071-2100) using both the A2 and B2 scenarios and 9 simulations. Three different weighting schemes were applied: a) BMA-SS weighting, b) BMA-NAO weighting and c) BMA-COM weighting.

A.4 SUMMER PRECIPITATION (2071-2100)

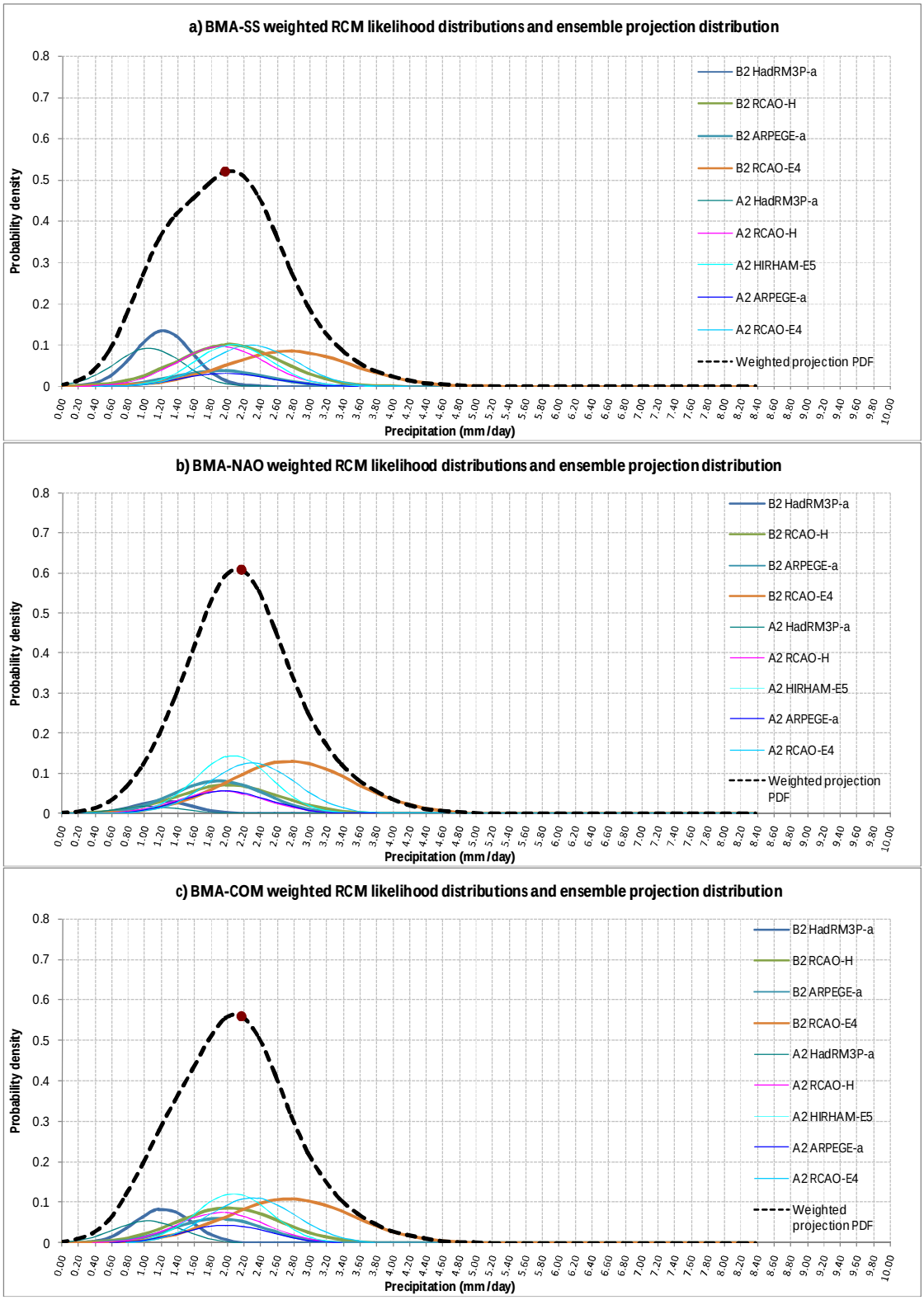


Figure A.4: Projections for summer precipitation (2071-2100) using both the A2 and B2 scenarios and 9 simulations. Three different weighting schemes were applied: a) BMA-SS weighting, b) BMA-NAO weighting and c) BMA-COM weighting.

BIBLIOGRAPHY

Abramowitz, G. (2010) 'Model independence in multi-model ensemble prediction', *Australian Meteorological and Oceanographic Journal*, 59, 3-6.

Adger, W. N., Agrawala, S., Mirza, M. M. Q., Conde, C., O'Brien, K., Pulhin, J., Pulwarty, R., B., S. and K., T. (2007) 'Assessment of adaptation practices, options, constraints and capacity' in Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. and Hanson, C. E., (Eds.), *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, UK: Cambridge University Press, 717-743.

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

Alley, R. B., Meese, D. A., Shuman, C. A., Gow, A. J., Taylor, K. C., Grootes, P. M., White, J. W. C., Ram, M., Waddington, E. D., Mayewski, P. A. and Zielinski, G. A. (1993) 'Abrupt increase in Greenland snow accumulation at the end of the Younger Dryas event', *Nature,* 362(6420), 527-529.

Ambaum, M. H. P., Hoskins, B. J. and Stephenson, D. B. (2001) 'Arctic oscillation or North Atlantic oscillation?', *Journal of Climate,* 14(16), 3495-3507.

Anisimov, O.A. (2007) 'Potential feedback of thawing permafrost to the global climate system through methane emission', *Environmental Research Letters* 4, 045016.

An, S. I., Hsieh, W. W. and Jin, F. F. (2005) 'A Nonlinear analysis of the ENSO cycle and its interdecadal changes', *Journal of Climate,* 18(16), 3229-3239.

Andronova, N. G. and Schlesinger, M. E. (2001) 'Objective estimation of the probability density function for climate sensitivity', *Journal of Geophysical Research-Atmospheres,* 106(D19), 22605-22611.

Antic, S., Laprise, R., Denis, B. and de Elia, R. (2006) 'Testing the downscaling ability of a one-way nested regional climate model in regions of complex topography', *Climate Dynamics,* 26(2-3), 305-325.

Arnell, N.W. and Delaney, K. (2006) 'Adapting to climate change: water supply in England and Wales', *Climatic Change* 78, 227-255.

Baer, P. and Risbey, J.S. (2009) 'Uncertainty and assessment of the issues posed by urgent climate change. An editorial comment', *Climatic Change* 92, 31-36.

Baker, M. B. and Peter, T. (2008) 'Small-scale cloud processes and climate', *Nature,* 451(7176), 299-300.

Barnett, D. N., Brown, S. J., Murphy, J. M., Sexton, D. M. H. and Webb, M. J. (2006) 'Quantifying uncertainty in changes in extreme event frequency in response to doubled CO2 using a large ensemble of GCM simulations', *Climate Dynamics,* 26(5), 489-511.

Barnett, T. P., Hasselmann, K., Chelliah, M., Delworth, T., Hegerl, G., Jones, P., Rasmusson, E., Roeckner, E., Ropelewski, C., Santer, B. and Tett, S. (1999) 'Detection and attribution of recent climate change: A status report', *Bulletin of the American Meteorological Society,* 80(12), 2631-2659.

Barnston, A. G. (1992) 'Correspondence among the correlation, RMSE and Heidke forecast verification measures - Refinement of the Heidke score', *Weather and Forecasting,* 7(4), 699-709.

Barnston, A. G., Glantz, M. H. and He, Y. X. (1999) 'Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997-98 El Nino episode and the 1998 La Nina onset', *Bulletin of the American Meteorological Society,* 80(2), 217-243.

Barth, A., Alvera-Azcarate, A., Rixen, M. and Beckers, J. M. (2005) 'Two-way nested model of mesoscale circulation features in the Ligurian Sea', *Progress in Oceanography,* 66(2-4), 171-189.

Beirlant, J., Glanzel, W., Carbonez, A. and Leemans, H. (2007) 'Scoring research output using statistical quantile plotting', *Journal of Informetrics,* 1(3), 185-192.

Benestad, R. E. (2004) 'Tentative probabilistic temperature scenarios for northern Europe', *Tellus Series a-Dynamic Meteorology and Oceanography,* 56(2), 89-101.

Beniston, M., Stephenson, D. B., Christensen, O. B., Ferro, C. A. T., Frei, C., Goyette, S., Halsnaes, K., Holt, T., Jylha, K., Koffi, B., Palutikof, J., Scholl, R., Semmler, T. and Woth, K. (2007) 'Future extreme events in European climate: an exploration of regional climate model projections', *Climatic Change,* 81, 71-95.

Berg, W., L'Ecuyer, T. and van den Heever, S. (2008) 'Evidence for the impact of aerosols on the onset and microphysical properties of rainfall from a combination of satellite observations and cloud-resolving model simulations', *Journal of Geophysical Research-Atmospheres* 113.

Berger, J. O., De Oliveira, V. and Sanso, B. (2001) 'Objective Bayesian analysis of spatially correlated data', *Journal of the American Statistical Association,* 96, 1361–1374.

Berliner, L. M., Levine, R. A. and Shea, D. J. (2000) 'Bayesian climate change assessment', *Journal of Climate,* 13(21), 3805-3820.

Betts, A. K., Zhao, M., Dirmeyer, P. A. and Beljaars, A. C. M. (2006) 'Comparison of ERA40 and NCEP/DOE near-surface data sets with other ISLSCP-II data sets', *Journal of Geophysical Research-Atmospheres,* 111(D22), 20.

Bjerknes, V. (1914) 'Meteorology as an exact science', *Monthly Weather Review,* 41(1), 11-14.

Bjerknes, V. (1919) 'Weather forecasting', *Monthly Weather Review,* 47(2), 90-95.

Blyth, C. R. (1972) 'Subjective vs objective methods in statistics', *American Statistician,* 26(3), 20-22.

Boneau, C. A. (1960) 'The effects of violations of assumptions underlying the t-test', *Psychological Bulletin,* 57(1), 49-64.

Bony, S., Dufresne, J. L., Le Treut, H., Morcrette, J. J. and Senior, C. (2004) 'On dynamic and thermodynamic components of cloud changes', *Climate Dynamics,* 22(2-3), 71-86.

Bornstein, R. and Lin, Q. L. (2000) 'Urban heat islands and summertime convective thunderstorms in Atlanta: three case studies', *Atmospheric Environment,* 34(3), 507-516.

Bouwer, L.M., Bubeck, P. and Aerts, J. (2010) 'Changes in future flood risk due to climate and development in a Dutch polder area', *Global Environmental Change,* 20(3), 463-471.

Brown, J. D. (2004) 'Knowledge, uncertainty and physical geography: towards the development of methodologies for questioning belief', *Transactions of the Institute of British Geographers,* 29(3), 367-381.

Brown, M. B. and Forsythe, A. B. (1974) 'Robust tests for equality of variances', *Journal of the American Statistical Association,* 69(346), 364-367.

Brown, S. and Lugo, A. E. (1982) 'The storage and production of organic matter in tropical forests and their role in the global carbon cycle', *Biotropica,* 14(3), 161-187.

Bryan, F. O. (1998) 'Climate drift in a multicentury integration of the NCAR Climate System Model', *Journal of Climate,* 11(6), 1455-1471.

Buell, C. E. (1975) 'The topography of empirical orthogocal functions' in *Preprints, Fourth Conference on Probability and Statistics in the Atmospheric Sciences,* American Meteorological Society, 188-193.

Buesseler, K. O., Andrews, J. E., Pike, S. M. and Charette, M. A. (2004) 'The effects of iron fertilization on carbon sequestration in the Southern Ocean', *Science,* 304(5669), 414-417.

Burby, R. J. (2006) 'Hurricane Katrina and the paradoxes of government disaster policy: Bringing about wise governmental decisions for hazardous areas', *Annals of the American Academy of Political and Social Science,* 604, 171-191.

Buytaert, W., Celleri, R. and Timbe, L. (2009) 'Predicting climate change impacts on water resources in the tropical Andes: Effects of GCM uncertainty', *Geophysical Research Letters,* 36, 5.

Burton, I., Challenger, B., Huq, S., Klein, R. J. T. and Yohe, G. (2001) 'Adaptation to Climate Change in the Context of Sustainable Development and Equity' in McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J. and White, K. S. (Eds.), *Climate Change 2001: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, UK: Cambridge University Press, 879-912.

Busuioc, A., Giorgi, F., Bi, X. and Ionita, M. (2006) 'Comparison of regional climate model and statistical downscaling simulations of different winter precipitation change scenarios over Romania', *Theoretical and Applied Climatology,* 86(1-4), 101-123.

Busuioc, A., von Storch, H. and Schnur, R. (1999) 'Verification of GCM-generated regional seasonal precipitation for current climate and of statistical downscaling estimates under changing climate conditions', *Journal of Climate,* 12(1), 258-272.

Carlson, A. E., Clark, P. U., Haley, B. A., Klinkhammer, G. P., Simmons, K., Brook, E. J. and Meissner, K. J. (2007) 'Geochemical proxies of North American freshwater routing during the Younger Dryas cold event', *Proceedings of the National Academy of Sciences of the United States of America,* 104(16), 6556-6561.

Carter,T.R. and Fronzek, S. (2005). 'Applying probabilistic climate projections to impact models', *Methodological note prepared for the ENSEMBLES RT 6/WP 6.2 meeting, 6-8 June 2005, Exeter, UK*. Unpublished.

Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E. E., Brown, B. G. and Mason, S. (2008) 'Forecast verification: current status and future directions', *Meteorological Applications,* 15(1), 3-18.

Castles, I. and Henderson, D. (2003) 'The IPCC Emission Scenarios: An Economic-Statistical Critique', *Energy and Environment,* 14(2), 159-186.

Castro, M., Fernandez, C. and Gaertner, M. A. (1993) 'Description of a meso-scale atmospheric numerical model' in Diaz, J. I. and Lions, J. L. (Eds.), *Mathematics, Climate and Environment*, Masson, 273.

Charlton, R., Fealy, R., Moore, S., Sweeney, J. and Murphy, C. (2006) 'Assessing the impact of climate change on water supply and flood hazard in Ireland using statistical downscaling and hydrological modelling techniques', *Climatic Change,* 74(4), 475-491.

Chen, R. S., Ersi, K., Yang, J. P., Lu, S. H. and Zhao, W. Z. (2004) 'Validation of five global radiation models with measured daily data in China', *Energy Conversion and Management,* 45(11-12), 1759-1769.

Christensen, J. H., Carter, T. R., Rummukainen, M. and Amanatidis, G. (2007) 'Evaluating the performance and utility of regional climate models: the PRUDENCE project', *Climatic Change,* 81, 1-6.

Christensen, J. H., Bøssing Christensen, O., Lopez, P., van Meijgaard, E. and Botzet, M. (1996) *TheHIRHAM4 Regional Atmospheric Climate Model*. Scientific Report 96-4, Danish Meteorological Institute.

Christensen, J. H. and Kuhry, P. (2000) 'High-resolution regional climate model validation and permafrost simulation for the East European Russian Arctic', *Journal of Geophysical Research-Atmospheres,* 105(D24), 29647-29658.

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

Christensen, O. B., Gaertner, M. A., Prego, J. A. and Polcher, J. (2001) 'Internal variability of regional climate models', *Climate Dynamics,* 17(11), 875-887.

Chu, P. C. (1999) 'Two kinds of predictability in the Lorenz system', *Journal of the Atmospheric Sciences,* 56(10), 1427-1432.

Claussen, M., Mysak, L. A., Weaver, A. J., Crucifix, M., Fichefet, T., Loutre, M. F., Weber, S. L., Alcamo, J., Alexeev, V. A., Berger, A., Calov, R., Ganopolski, A., Goosse, H., Lohmann, G., Lunkeit, F., Mokhov, II, Petoukhov, V., Stone, P. and Wang, Z. (2002) 'Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models', *Climate Dynamics,* 18(7), 579-586.

Cohen, J. and Fletcher, C. (2007) 'Improved skill of Northern Hemisphere winter surface temperature predictions based on land-atmosphere fall anomalies', *Journal of Climate,* 20(16), 4118-4132.

Collins, M. (2007) 'Ensembles and probabilities: a new era in the prediction of climate change', *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences,* 365(1857), 1957-1970.

Collins, M. and Allen, M. R. (2002) 'Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability', *Journal of Climate,* 15(21), 3104-3109.

Collins, M., Tett, S. F. B. and Cooper, C. (2001) 'The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments', *Climate Dynamics*, 17(1), 61-81.

British Irish Council (2003) *Scenarios of Climate Change for Islands within the BIC Region*, Exeter: Met. Office Hadley Centre for Climate Prediction and Research.

Covey, C., Sloan, L. C. and Hoffert, M. I. (1996) 'Paleoclimate data constraints on climate sensitivity: The paleocalibration method', *Climatic Change*, 32(2), 165-184.

Cox, P. M., Betts, R. A., Collins, M., Harris, P. P., Huntingford, C. and Jones, C. D. (2004) 'Amazonian forest dieback under climate-carbon cycle projections for the 21st century', *Theoretical and Applied Climatology*, 78(1-3), 137-156.

Curry, J. A., Schramm, J. L. and Ebert, E. E. (1995) 'Sea-ice albedo climate feedback mechanism', *Journal of Climate*, 8(2), 240-247.

Dang, H. H., Michaelowa, A. and Tuan, D. D. (2003) 'Synergy of adaptation and mitigation strategies in the context of sustainable development: the case of Vietnam', *Climate Policy*, 3, S81-S96.

Davies, J.R., Rowell, D.P. and Folland, C.K. (1997) 'North Atlantic and European seasonal predictability using an ensemble of multidecadal atmospheric GCM simulations', *International Journal of Climatology* 17, 1263-1284.

Davis, J. (1973) *Statistics and Data Analysis in Geology*, New York: Wiley.

Denis, B., Laprise, R. and Caya, D. (2003) 'Sensitivity of a regional climate model to the resolution of the lateral boundary conditions', *Climate Dynamics*, 20(2-3), 107-126.

Deque, M., Marquet, P. and Jones, R. G. (1998) 'Simulation of climate change over Europe using a global variable resolution general circulation model', *Climate Dynamics*, 14(3), 173-189.

Descamps, L. and Talagrand, O. (2007) 'On some aspects of the definition of initial conditions for ensemble prediction', *Monthly Weather Review*, 135(9), 3260-3272.

Dessai, S. and Hulme, M. (2004) 'Does climate adaptation policy need probabilities?', *Climate Policy*, 4(2), 107-128.

Dessai, S., Lu, X. F. and Risbey, J. S. (2005) 'On the role of climate scenarios for adaptation planning', *Global Environmental Change-Human and Policy Dimensions*, 15(2), 87-97.

Devoy, R. J. N. (2008) 'Coastal vulnerability and the implications of sea-level rise for Ireland', *Journal of Coastal Research*, 24(2), 325-+.

Ding, Y. H., Shi, X. L., Liu, Y. M., Liu, Y., Li, Q. Q., Qian, F. F., Miao, Q. Q., Zhai, Q. Q. and Gao, K. (2006) 'Multi-year simulations and experimental seasonal predictions for rainy seasons in China by using a nested regional climate model (RegCM_NCC). part I: Sensitivity study', *Advances in Atmospheric Sciences,* 23(3), 323-341.

Diniz Filho, J. A. F., Bini, L. M., Rangel, T. F., Loyola, R. D., Hof, C., Nogues-Bravo, D. and Araujo, M. B. (2009) 'Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change', *Ecography,* 32(6), 897-906.

Dirmeyer, P. A. (2001) 'Climate drift in a coupled land-atmosphere model', *Journal of Hydrometeorology,* 2(1), 89-100.

Doblas-Reyes, F. J., Hagedorn, R. and Palmer, T. N. (2005) 'The rationale behind the success of multi-model ensembles in seasonal forecasting - II. Calibration and combination', *Tellus Series a-Dynamic Meteorology and Oceanography,* 57(3), 234-252.

Dommenget, D. and Latif, M. (2002) 'A cautionary note on the interpretation of EOFs', *Journal of Climate,* 15(2), 216-225.

Doms, G. and Schattler, U. (2002) *A Description of the Nonhydrostatic Regional Model LM Part I :Dynamics and Numerics*, Consortium for Small-Scale Modelling Report, Offenbach: Deutscher Wetterdienst.

Doscher, R., Willen, U., Jones, C., Rutgersson, A., Meier, H. E. M., Hansson, U. and Graham, L. P. (2002) 'The development of the regional coupled ocean-atmosphere model RCAO', *Boreal Environment Research,* 7(3), 183-192.

Dunne, S., Hanafin, J., Lynch, P., McGrath, R., Nishimura, E., Nolan, P., Ratnam, J. V., Semmler, T., Sweeney, C., Varghese, S. and Wang, S. (2008) *Ireland in a warmer world: Scientific predictions of the Irish climate,* McGrath, R. and Lynch, P. (Eds.) Dublin: Community Climate Change Consortium for Ireland.

Dyer, J., Haase-Wittler, P. and Washburn, S. (2003) 'Structuring agricultural education research using conceptual and theoretical frameworks', *Journal of Agricultural Education,* 44(2), 61-74.

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R. amd Mearns, L. O. (2000) 'Climate extremes: Observations, modelling, and impacts', *Science,* 289(5487), 2068-2074.

Edwards, T. L., Crucifix, M. and Harrison, S. P. (2007) 'Using the past to constrain the future: how the palaeorecord can improve estimates of global warming', *Progress in Physical Geography,* 31(5), 481-500.

Efthymiadis, D., Jones, P. D., Briffa, K. R., Bohm, R. and Maugeri, M. (2007) 'Influence of large-scale atmospheric circulation on climate variability in the

Greater Alpine Region of Europe', *Journal of Geophysical Research-Atmospheres,* 112(D12).

Emanuel, K. A., Neelin, J. D. and Bretherton, C. S. (1994) 'On large-scale circulations in convecting atmospheres', *Quarterly Journal of the Royal Meteorological Society,* 120(519), 1111-1143.

Evans, J. P., Smith, R. B. and Oglesby, R. J. (2004) 'Middle East climate simulation and dominant precipitation processes', *International Journal of Climatology,* 24(13), 1671-1694.

Fankhauser, S., Smith, J. B. and Tol, R. S. J. (1999) 'Weathering climate change: some simple rules to guide adaptation decisions', *Ecological Economics,* 30(1), 67-78.

Fealy, R. (2010) *An Assessment of Uncertainties in Climate Modelling at the Regional Scale: The Development of Probabilistic Based Climate Scenarios for Ireland*, STRIVE Report Series No.48, Dublin: Environmental Protection Agency.

Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R. and Webster, M. D. (2002) 'Quantifying uncertainties in climate system properties with the use of recent climate observations', *Science,* 295(5552), 113-117.

Frame, D. J., Faull, N. E., Joshi, M. M. and Allen, M. R. (2007) 'Probabilistic climate forecasts and inductive problems', *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences,* 365(1857), 1971-1992.

Fried, J. S., Torn, M. S. and Mills, E. (2004) 'The impact of climate change on wildfire severity: A regional forecast for northern California', *Climatic Change,* 64(1-2), 169-191.

Fronzek, S., Carter, T.R., Räisänen, J., Ruokolainen, L., and Luoto, M. (2010) 'Applying probabilistic projections of climate change with impact models: a case study for sub-arctic palsa mires in Fennoscandia', *Climatic Change,* 99(3-4), 515-534.

Fyfe, J. C., Boer, G. J. and Flato, G. M. (1999) 'The Arctic and Antarctic oscillations and their projected changes under global warming', *Geophysical Research Letters,* 26(11), 1601-1604.

Gao, X. J., Pal, J. S. and Giorgi, F. (2006) 'Projected changes in mean and extreme precipitation over the Mediterranean region from a high resolution double nested RCM simulation', *Geophysical Research Letters,* 33(3).

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., Fiorino, M., Gleckler, P. J., Hnilo, J. J., Marlais, S. M., Phillips, T. J., Potter, G. L., Santer, B. D., Sperber, K. R., Taylor, K. E. and Williams, D. N. (1999) 'An overview of the results of the Atmospheric Model Intercomparison

Project (AMIP I)', *Bulletin of the American Meteorological Society*, 80(1), 29-55.

Genthon, C., Krinner, G. and Castebrunet, H. (2009) 'Antarctic precipitation and climate-change predictions: horizontal resolution and margin vs plateau issues', *Annals of Glaciology*, 50(50), 55-60.

Gershon, N. (1998) 'Visualization of an imperfect world', *Ieee Computer Graphics and Applications*, 18(4), 43-45.

Gillett, N. P., Graf, H. F. and Osborn, T. J. (2003) 'Climate Change and the North Atlantic Ocean', *Geophysical Monograph*, 134, 193-209.

Giorgi, F. (1990) 'Simulation of regional climate using a limited area model nested in a general-circulation model', *Journal of Climate*, 3(9), 941-963.

Giorgi, F., Bi, X. Q. and Pal, J. (2004) 'Mean, interannual variability and trends in a regional climate change experiment over Europe. II: climate change scenarios (2071-2100)', *Climate Dynamics*, 23(7-8), 839-858.

Giorgi, F., Marinucci, M. R. and Bates, G. T. (1993) 'Development of a 2nd generation regional climate model (REGCM2) 1. Boundary-layer and radiative-transfer processes', *Monthly Weather Review*, 121(10), 2794-2813.

Giorgi, F. and Mearns, L. O. (2003) 'Probability of regional climate change based on the Reliability Ensemble Averaging (REA) method', *Geophysical Research Letters*, 30(12), 4.

Giorgi, F. (2006) 'Regional climate modelling: Status and perspectives', *Journal De Physique Iv*, 139, 101-118.

Goerss, J. S. (2000) 'Tropical cyclone track forecasts using an ensemble of dynamical models', *Monthly Weather Review*, 128(4), 1187-1193.

Good, P., Lowe, J. A. and Rowell, D. P. (2009) 'Understanding uncertainty in future projections for the tropical Atlantic: relationships with the unforced climate', *Climate Dynamics*, 32(2-3), 205-218.

Goodess, C. M. and Palutikof, J. P. (1998) 'Development of daily rainfall scenarios for southeast Spain using a circulation-type approach to downscaling', *International Journal of Climatology*, 18(10), 1051-1083.

Graham, N. E. (1994) 'Decadal-scale climate variability in the Tropical and North Pacific during the 1970s and 1980s - Observations and model results', *Climate Dynamics*, 10(3), 135-162.

Griffies, S. M. and Bryan, K. (1997) 'A predictability study of simulated North Atlantic multidecadal variability', *Climate Dynamics*, 13(7-8), 459-487.

Grimit, E. P. and Mass, C. F. (2002) 'Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest', *Weather and Forecasting,* 17(2), 192-205.

Grotjahn, R. (2008) 'Different data, different general circulations? A comparison of selected fields in NCEP/DOE AMIP-II and ECMWF ERA-40 reanalyses', *Dynamics of Atmospheres and Oceans,* 44(3-4), 108-142.

Gustafsson, N., Kallen, E. and Thorsteinsson, S. (1998) 'Sensitivity of forecast errors to initial and lateral boundary conditions', *Tellus Series a-Dynamic Meteorology and Oceanography,* 50(2), 167-185.

Haack, S. (1993) *Evidence and Inquiry: Towards reconstruction in epistemology,* Oxford: Blackwell.

Hagedorn, R., Doblas-Reyes, F. J. and Palmer, T. N. (2005) 'The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept', *Tellus Series a-Dynamic Meteorology and Oceanography,* 57(3), 219-233.

Hamill, T. M. (2001) 'Interpretation of rank histograms for verifying ensemble forecasts', *Monthly Weather Review,* 129(3), 550-560.

Hamilton, J. M. and Tol, R. S. J. (2007) 'The impact of climate change on tourism in Germany, the UK and Ireland: a simulation study', *Regional Environmental Change,* 7(3), 161-172.

Hannachi, A., Jolliffe, I. T. and Stephenson, D. B. (2007) 'Empirical orthogonal functions and related techniques in atmospheric science: A review', *International Journal of Climatology,* 27(9), 1119-1152.

Hansen, J., Ruedy, R., Glascoe, J. and Sato, M. (1999) 'GISS analysis of surface temperature change', *Journal of Geophysical Research-Atmospheres,* 104(D24), 30997-31022.

Hansen, J., Ruedy, R., Sato, M. and Lo, K. (2010) 'Global Surface Temperature Change',

Hansen, J. E. (2006) 'Can we still avoid dangerous human-made climate change?', *Social Research,* 73(3), 949-971.

Hansen, J. E. and Sato, M. (2001) 'Trends of measured climate forcing agents', *Proceedings of the National Academy of Sciences of the United States of America,* 98(26), 14778-14783.

Hanson, C. E., Palutikof, J. P., Livermore, M. T. J., Barring, L., Bindi, M., Corte-Real, J., Durao, R., Giannakopoulos, C., Good, P., Holt, T., Kundzewicz, Z., Leckebusch, G. C., Moriondo, M., Radziejewski, M., Santos, J., Schlyter, P., Schwarb, M., Stjernquist, I. and Ulbrich, U. (2007) 'Modelling the impact of climate extremes: an overview of the MICE project', *Climatic Change,* 81, 163-177.

Hartmann, D. L., Wallace, J. M., Limpasuvan, V., Thompson, D. W. J. and Holton, J. R. (2000) 'Can ozone depletion and global warming interact to produce rapid climate change?', *Proceedings of the National Academy of Sciences of the United States of America,* 97(4), 1412-1417.

Hawkins, E. and Sutton, R. (2009) 'The potential to narrow uncertainty in regional climate predictions', *Bulletin of the American Meteorological Society* 90, 1095-1107.

Hay, L. E., Wilby, R. J. L. and Leavesley, G. H. (2000) 'A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States', *Journal of the American Water Resources Association,* 36(2), 387-397.

Haylock, M. R., Cawley, G. C., Harpham, C., Wilby, R. L. and Goodess, C. M. (2006) 'Downscaling heavy precipitation over the United Kingdom: A comparison of dynamical and statistical methods and their future scenarios', *International Journal of Climatology,* 26(10), 1397-1415.

Hellstrom, C., Chen, D. L., Achberger, C. and Raisanen, J. (2001) 'Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation', *Climate Research,* 19(1), 45-55.

Henderson-Sellers, A. (1993) 'An antipodean climate of uncertainty', *Climatic Change,* 25(3-4), 203-224.

Henderson-Sellers, A. (1993) 'An antipodean climate of uncertainty', *Climatic Change,* 25(3-4), 203-224.

Hingray, B., Mezghani, A. and Buishand, T. A. (2007) 'Development of probability distributions for regional climate change from uncertain global mean warming and an uncertain scaling relationship', *Hydrology and Earth System Sciences,* 11(3), 1097-1114.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999) 'Bayesian model averaging: A tutorial', *Statistical Science,* 14(4), 382-401.

Hoffert, M. I. and Covey, C. (1992) 'Deriving global climate sensitivity from paleoclimate reconstructions', *Nature,* 360(6404), 573-576.

Holden, N. M. and Brereton, A. J. (2003) 'Potential impacts of climate change on maize production and the introduction of soybean in Ireland', *Irish Journal of Agricultural and Food Research,* 42(1), 1-15.

Holden, N. M., Brereton, A. J., Fealy, R. and Sweeney, J. (2003) 'Possible change in Irish climate and its impact on barley and potato yields', *Agricultural and Forest Meteorology*, 116(3-4), 181-196.

Holland, M. M., Serreze, M. C. and Stroeve, J. (2010) 'The sea ice mass budget of the Arctic and its future change as simulated by coupled climate models', *Climate Dynamics,* 34(2-3), 185-200.

Houghton, R. A. (1995) 'Land-use change and the carbon-cycle', *Global Change Biology,* 1(4), 275-287.

Hsieh, W. W. and Tang, B. Y. (1998) 'Applying neural network models to prediction and data analysis in meteorology and oceanography', *Bulletin of the American Meteorological Society,* 79(9), 1855-1870.

Hubbard, D. (2007) *How to Measure Anything: Finding the Value of Intangibles in Business*, New York: Wiley.

Hulme, M. (2008) 'Geographical work at the boundaries of climate change', *Transactions of the Institute of British Geographers,* 33(1), 5-11.

Hulme, M. and Carter, T. R. (1999) 'Representing Uncertainty in Climate Change Scenarios and Impact Studies', *ECLAT-2 Red Workshop Report*, T. Carter, M. Hulme and D. Viner (Eds.), Norwich: Climatic Research Unit.

Hurrell, J. W. (1995) 'Decadal trends in the North-Atlantic Oscillation - Regional temperatures and precipitation', *Science,* 269(5224), 676-679.

Hurrell, J. W. and van Loon, H. (1997) 'Decadal variations in climate associated with the north Atlantic oscillation', *Climatic Change,* 36(3-4), 301-326.

Im, E. S., Park, E. H., Kwon, W. T. and Giorgi, F. (2006) 'Present climate simulation over Korea with a regional climate model using a one-way double-nested system', *Theoretical and Applied Climatology,* 86(1-4), 187-200.

IPCC (2007) *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change,* Cambridge, UK and New York, USA: Cambridge University Press.

Irwin, R. (2010) 'Climate Change and Heidegger's Philosophy of Science', *Climate Ethics,* 11(1), 1-12.

Jacob, D., Barring, L., Christensen, O. B., Christensen, J. H., de Castro, M., Deque, M., Giorgi, F., Hagemann, S., Lenderink, G., Rockel, B., Sanchez, E., Schar, C., Seneviratne, S. I., Somot, S., van Ulden, A. and van den Hurk, B. (2007) 'An inter-comparison of regional climate models for Europe: model performance in present-day climate', *Climatic Change,* 81, 31-52.

Jacob, D. and Podzun, R. (1997) 'Sensitivity studies with the regional climate model REMO', *Meteorology and Atmospheric Physics,* 63(1-2), 119-129.

Jenkins, G. (2003) *Handling uncertainties in the UKCIP02 scenarios of climate change*, Hadley Centre technical note 44, Exeter: Met Office.

Jenkinson, A. F. and Collison, F. P. (1977) *An initial climatology of gales over the North sea*, Synoptic Climatology Branch Memorandum 62, London: Meteorological Office.

Jones, G. S., Gregory, J. M., Stott, P. A., Tett, S. F. B. and Thorpe, R. B. (2005) 'An AOGCM simulation of the climate response to a volcanic super-eruption', *Climate Dynamics,* 25(7-8), 725-738.

Jones, P. D., Hulme, M. and Briffa, K. R. (1993) 'A comparison of Lamb circulation types with an objective classification scheme', *International Journal of Climatology*, 13(6), 655-663.

Jones, P. D. and Reid, P. A. (2001) 'Assessing future changes in extreme precipitation over Britain using regional climate model integrations', *International Journal of Climatology,* 21(11), 1337-1356.

Jones, R. G., Murphy, J. M. and Noguer, M. (1995) 'Simulation of climate change over Europe using a nested regional climate model. 1. Assessment of control climate, including sensitivity to location of lateral boundaries', *Quarterly Journal of the Royal Meteorological Society*, 121(526), 1413-1449.

Jones, R. N. (2000a) 'Analysing the risk of climate change using an irrigation demand model', *Climate Research*, 14(2), 89-100.

Jones, R.N. (2000b) 'Managing uncertainty in climate change projections - Issues for impact assessment - An editorial comment', *Climatic Change* 45, 403-419.

Ju, L. X., Wang, H. K. and Jiang, D. B. (2007) 'Simulation of the Last Glacial Maximum climate over East Asia with a regional climate model nested in a general circulation model', *Palaeogeography Palaeoclimatology Palaeoecology*, 248(3-4), 376-390.

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

Kay, A. L., Davies, H. N., Bell, V. A. and Jones, R. G. (2009) 'Comparison of uncertainty sources for climate change impacts: flood frequency in England', *Climatic Change,* 92(1-2), 41-63.

Keeling, C. D., Bacastow, R. B., Bainbridge, A. E., Ekdahl, C. A., Guenther, P. R., Waterman, L. S. and Chin, J. F. S. (1976) 'Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii', *Tellus,* 28(6), 538-551.

Kennedy, M., Mrofka, D. and von der Borch, C. (2008) 'Snowball Earth termination by destabilization of equatorial permafrost methane clathrate', *Nature* 453, 642-645,

Khain, A., Rosenfeld, D. and Pokrovsky, A. (2005) 'Aerosol impact on the dynamics and microphysics of deep convective clouds', *Quarterly Journal of the Royal Meteorological Society*, 131(611), 2639-2663.

Kiely, G. (1999) 'Climate change in Ireland from precipitation and streamflow observations', *Advances in Water Resources*, 23(2), 141-151.

Kim, K. Y. and North, G. R. (1993) 'EOF analysis of surface temperature field in a stochastic climate model', *Journal of Climate*, 6(9), 1681-1690.

King, D. A. (2004) 'Environment - Climate change science: Adapt, mitigate, or ignore?', *Science*, 303(5655), 176-177.

Kjellstrom, E. and Giorgi, F. (2010) 'Introduction', *Climate Research*, 44, 117–119.

Knight, F. H. (1921) *Risk, Uncertainty, and Profit*, Boston: Hart, Schaffner & Marx.

Knutti, R. and Stocker, T. F. (2002) 'Limited predictability of the future thermohaline circulation close to an instability threshold', *Journal of Climate*, 15(2), 179-186.

Knutti, R., Stocker, T. F., Joos, F. and Plattner, G. K. (2002) 'Constraints on radiative forcing and future climate change from observations and climate model ensembles', *Nature*, 416(6882), 719-723.

Kriegler, E. (2005) *Imprecise probability analysis for integrated assessment of climate change*, unpublished thesis, Potsdam University.

Kvale, S. (1996) *Interviews: An introduction to qualitative research interviewing*, London, England: SAGE Publications.

Lal, M., McGregor, J. L. and Nguyen, K. C. (2008) 'Very high-resolution climate simulation over Fiji using a global variable-resolution model', *Climate Dynamics*, 30(2-3), 293-305.

Lamb, P. J. and Peppler, R. A. (1987) 'North-Atlantic Oscillation - Concept and application', *Bulletin of the American Meteorological Society*, 68(10), 1218-1225.

Latif, M., Anderson, D., Barnett, T., Cane, M., Kleeman, R., Leetmaa, A., O'Brien, J., Rosati, A. and Schneider, E. (1998) 'A review of the predictability and prediction of ENSO', *Journal of Geophysical Research-Oceans*, 103(C7), 14375-14393.

Latif, M. and Keenlyside, N. S. (2009) 'El Nino/Southern Oscillation response to global warming', *Proceedings of the National Academy of Sciences of the United States of America*, 106(49), 20578-20583.

Laurance, W. F. and Williamson, G. B. (2001) 'Positive feedbacks among forest fragmentation, drought, and climate change in the Amazon', *Conservation Biology*, 15(6), 1529-1535.

Laut, P. (2003) 'Solar activity and terrestrial climate: an analysis of some purported correlations', *Journal of Atmospheric and Solar-Terrestrial Physics*, 65(7), 801-812.

Lempert, R., Nakicenovic, N., Sarewitz, D. and Schlesinger, M. (2004) 'Characterizing climate-change uncertainties for decision-makers - An editorial essay', *Climatic Change* 65, 1-9.

Le Quere, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., Conway, T. J., Doney, S. C., Feely, R. A., Foster, P., Friedlingstein, P., Gurney, K., Houghton, R. A., House, J. I., Huntingford, C., Levy, P. E., Lomas, M. R., Majkut, J., Metzl, N., Ometto, J. P., Peters, G. P., Prentice, I. C., Randerson, J. T., Running, S. W., Sarmiento, J. L., Schuster, U., Sitch, S., Takahashi, T., Viovy, N., van der Werf, G. R. and Woodward, F. I. (2009) 'Trends in the sources and sinks of carbon dioxide', *Nature Geoscience*, 2(12), 831-836.

Leung, L. R. and Ghan, S. J. (1999) 'Pacific Northwest climate sensitivity simulated by a regional climate model driven by a GCM. Part I: Control simulations', *Journal of Climate*, 12(7), 2010-2030.

Leung, L. R., Mearns, L. O., Giorgi, F., Wilby, R. L. and Bm (2003) 'Regional climate research - Needs and opportunities', *Bulletin of the American Meteorological Society*, 84(1), 89-95.

Lightbody, B. (2006) 'Virtue Foundherentism', *Kriterion*, 20, 14-21.

Linderson, M. L. (2001) 'Objective classification of atmospheric circulation over southern Scandinavia', *International Journal of Climatology*, 21(2), 155-169.

Lohmann, U. (2008) 'Global anthropogenic aerosol effects on convective clouds in ECHAM5-HAM', *Atmospheric Chemistry and Physics*, 8(7), 2115-2131.

Lopez A., Fung F., New, M., Watts G., Weston A. and Wilby R.L. (2009) 'From climate model ensembles to climate change impacts: A case study of water resource management in the South West of England', *Water Resources Research*, 45, W08419.

Lorenz, P. and Jacob, D. (2005) 'Influence of regional scale information on the global circulation: A two-way nesting climate simulation', *Geophysical Research Letters*, 32(18).

Lu, M.L., Feingold, G., Jonsson, H.H., Chuang, P.Y., Gates, H., Flagan, R.C. and Seinfeld, J.H. (2008) 'Aerosol-cloud relationships in continental shallow cumulus', *Journal of Geophysical Research-Atmospheres* 113.

Lucarini, V., Speranza, A. and Vitolo, R. (2007) 'Parametric smoothness and self-scaling of the statistical properties of a minimal climate model: What beyond the mean field theories?', *Physica D-Nonlinear Phenomena,* 234(2), 105-123.

Lucas-Picher, P., Caya, D., de Elia, R. and Laprise, R. (2008) 'Investigation of regional climate models' internal variability with a ten-member ensemble of 10-year simulations over a large domain', *Climate Dynamics,* 31(7-8), 927-940.

Lynn, B. H., Healy, R. and Druyan, L. M. (2009) 'Quantifying the sensitivity of simulated climate change to model configuration', *Climatic Change,* 92(3-4), 275-298.

Ma, L. J., Zhang, T. J., Li, Q. X., Frauenfeld, O. W. and Qin, D. (2008) 'Evaluation of ERA-40, NCEP-1, and NCEP-2 reanalysis air temperatures with ground-based measurements in China', *Journal of Geophysical Research-Atmospheres,* 113(D15), 15.

Mahasenan, N., Smith, S., Humphreys, K. and Kaya, Y. (2003) '*Greenhouse Gas Control Technologies - 6th International Conference*' in, Oxford: Pergamon, 995–1000.

Maier-Reimer, E. (1987) 'Transport and storage of CO2 in the ocean - An inorganic ocean-circulation carbon cycle model', *Climate Dynamics,* 2(2), 63-90.

Marcus, M. G. and Brazel, S. W. B. (1984) 'Climate Changes in Arizona's Future', *Arizona State Climate Publication No. 1,*

May, W. (2007) 'The simulation of the variability and extremes of daily precipitation over Europe by the HIRHAM regional climate model', *Global and Planetary Change,* 57(1-2), 59-82.

McCormick, M.P., Thomason, L.W. and Trepte, C.R. (1995) 'Atmospheric effects of the Mt. Pinatubo eruption', *Nature* 373, 399-404.

McCuen, R. H., Knight, Z. and Cutter, A. G. (2006) 'Evaluation of the Nash-Sutcliffe efficiency index', *Journal of Hydrologic Engineering,* 11(6), 597-602.

Mearns, L. O., Giorgi, F., McDaniel, L. and Shields, C. (2003) 'Climate scenarios for the southeastern US based on GCM and regional model simulations', *Climatic Change,* 60(1-2), 7-35.

Meehl, G. A., Zwiers, F., Evans, J., Knutson, T., Mearns, L. and Whetton, P. (2000) 'Trends in extreme weather and climate events: Issues related to modelling extremes in projections of future climate change', *Bulletin of the American Meteorological Society,* 81(3), 427-436.

Miller, R. L., Schmidt, G. A. and Shindell, D. T. (2006) 'Forced annular variations in the 20th century intergovernmental panel on climate change fourth

assessment report models', *Journal of Geophysical Research-Atmospheres*, 111(D18).

Mitchell, T. (2006) *Adapting to Climate Change: Challenges and opportunities for the development community*, Middlesex: Tearfund.

Mitchell, T. D. and Hulme, M. (1999) 'Predicting regional climate change: living with uncertainty', *Progress in Physical Geography*, 23(1), 57-78.

Moberg, A. and Jones, P. D. (2004) 'Regional climate model simulations of daily maximum and minimum near-surface temperatures across Europe compared with observed station data 1961-1990', *Climate Dynamics*, 23(7-8), 695-715.

Mohapatra, M., Mohanty, U. C. and Behera, S. (2003) 'Spatial variability of daily rainfall over Orissa, India, during the southwest summer monsoon season', *International Journal of Climatology*, 23(15), 1867-1887.

Mooney, P. A., Mulligan, F. J. and Fealy, R. (2010) 'Comparison of ERA-40, ERA-Interim and NCEP/NCAR reanalysis data with observed surface air temperatures over Ireland', *International Journal of Climatology* [in press].

Murphy, A. H. (1993) 'What is a good forecast - An essay on the nature of goodness in weather forecasting', *Weather and Forecasting*, 8(2), 281-293.

Murphy, A. H. and Wilks, D. S. (1998) 'A case study of the use of statistical models in forecast verification: Precipitation probability forecasts', *Weather and Forecasting*, 13(3), 795-810.

Murphy, J. (2000) 'Predictions of climate change over Europe using statistical and dynamical downscaling techniques', *International Journal of Climatology*, 20(5), 489-501.

Murphy, J. (1999) 'An evaluation of statistical and dynamical techniques for downscaling local climate', *Journal of Climate*, 12(8), 2256-2284.

Murphy, J. M., Booth, B. B. B., Collins, M., Harris, G. R., Sexton, D. M. H. and Webb, M. J. (2007) 'A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles', *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, 365(1857), 1993-2028.

Murphy, S. J. and Washington, R. (2001) 'United Kingdom and Ireland precipitation variability and the North Atlantic sea-level pressure field', *International Journal of Climatology*, 21(8), 939-959.

Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S., Gregory, K., Grubler, A., Jung, T. Y., Kram, T., La Rovere, E. L., Michaelis, L., Mori, S., Morita, T., Pepper, W., Pitcher, H. M., Price, L., Riahi, K., Roehrl, A., Rogner, H.-H., Sankovski, A., Schlesinger, M., Shukla, P., Smith, S. J., Swart, R., van Rooijen, S., Victor, N. and Dadi, Z. (2000) *Special*

Report on Emissions Scenarios : a special report of Working Group III of the Intergovernmental Panel on Climate Change, New York: Cambridge University Press.

Nash, J. E. and J.V. Sutcliffe, J. V. (1970) 'River flow forecasting through conceptual models part I — A discussion of principles', *Journal of Hydrology*, 10(3), 282-290.

New M., Lopez A., Dessai S. and Wilby R. (2007) 'Challenges in using probabilistic climate change information for impact assessments: an example from the water sector', *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, 365, 2117–2131.

Nott, M. P., Desante, D. F., Siegel, R. B. and Pyle, P. (2002) 'Influences of the El Nino/Southern Oscillation and the North Atlantic Oscillation on avian productivity in forests of the Pacific Northwest of North America', *Global Ecology and Biogeography*, 11(4), 333-342.

Oreskes, N., Shraderfrechette, K. and Belitz, K. (1994) 'Verification, validation and confirmation of numerical models in the Earth sciences', *Science*, 263(5147), 641-646.

Ormerod, R. (2006) 'The history and ideas of pragmatism', *Journal of the Operational Research Society*, 57(8), 892-909.

Osborn, T. J., Briffa, K. R., Tett, S. F. B., Jones, P. D. and Trigo, R. M. (1999) 'Evaluation of the North Atlantic Oscillation as simulated by a coupled climate model', *Climate Dynamics*, 15(9), 685-702.

Overland, J. E. and Preisendorfer, R. W. (1982) 'A significance test for principal components applied to a cyclone climatology', *Monthly Weather Review*, 110(1), 1-4.

Paeth, H., Hense, A., Glowienka-Hense, R., Voss, R. and Cubasch, U. (1999a) 'The North Atlantic Oscillation as an indicator for greenhouse-gas induced regional climate change', *Climate Dynamics*, 15(12), 953-960.

Paeth, H., Hense, A., Glowienka-Hense, R., Voss, R. and Cubasch, U. (1999b) 'The North Atlantic Oscillation as an indicator for greenhouse-gas induced regional climate change', *Climate Dynamics*, 15(12), 953-960.

Palmer, T. N., Shutts, G. J., Hagedorn, R., Doblas-Reyes, E., Jung, T. and Leutbecher, M. (2005) 'Representing model uncertainty in weather and climate prediction', *Annual Review of Earth and Planetary Sciences*, 33, 163-193.

Palutikof, J. P., Winkler, J. A., Goodess, C. M. and Andresen, J. A. (1997) 'The simulation of daily temperature time series from GCM output .1. Comparison of model data with observations', *Journal of Climate*, 10(10), 2497-2513.

Pan, Z., Christensen, J. H., Arritt, R. W., Gutowski, W. J., Takle, E. S. and Otieno, F. (2001) 'Evaluation of uncertainties in regional climate change simulations', *Journal of Geophysical Research-Atmospheres,* 106(D16), 17735-17751.

Parry, M. L., Rosenzweig, C., Iglesias, A., Livermore, M. and Fischer, G. (2004) 'Effects of climate change on global food production under SRES emissions and socio-economic scenarios', *Global Environmental Change-Human and Policy Dimensions,* 14(1), 53-67.

Patt, A. and Dessai, S. (2005) 'Communicating uncertainty: lessons learned and suggestions for climate change assessment', *Comptes Rendus Geoscience,* 337(4), 425-441.

Pearson, A. (2009) 'Why storms are good news for fishermen', *New Scientist,* 2689.

Peirce, C. S. (1868) 'Some consequences of four incapacities', *Journal of Speculative Philosophy,* 2, 140-157.

Peixoto, J. and Oort, A. (1992) *Physics of Climate,* New York: American Institute of Physics.

Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J. M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V. M., Legrand, M., Lipenkov, V. Y., Lorius, C., Pepin, L., Ritz, C., Saltzman, E. and Stievenard, M. (1999) 'Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica', *Nature,* 399(6735), 429-436.

Piani, C., Frame, D.J., Stainforth, D.A. and Allen, M.R. (2005) 'Constraints on climate change from a multi-thousand member ensemble of simulations', *Geophysical Research Letters* 32.

Phillips, V. T. J., Donner, L. J. and Garner, S. T. (2007) 'Nucleation processes in deep convection simulated by a cloud-system-resolving model with double-moment bulk microphysics', *Journal of the Atmospheric Sciences,* 64(3), 738-761.

Poliakoff, M. and Licence, P. (2007) 'Sustainable technology - Green chemistry', *Nature,* 450(7171), 810-812.

Pope, V. D., Gallani, M. L., Rowntree, P. R. and Stratton, R. A. (2000) 'The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3', *Climate Dynamics,* 16(2-3), 123-146.

Popper, S.W., Lempert, R.J. and Bankes, S.C. (2005) 'Shaping the future', *Scientific American* 292, 66-71.

Quadrelli, R., Lazzeri, M., Cacciamani, C. and Tibaldi, S. (2001) 'Observed winter Alpine precipitation variability and links with large-scale circulation patterns', *Climate Research,* 17(3), 275-284.

Rabier, F., Klinker, E., Courtier, P. and Hollingsworth, A. (1996) 'Sensitivity of forecast errors to initial conditions', *Quarterly Journal of the Royal Meteorological Society*, 122(529), 121-150.

Raisanen, J., Hansson, U., Ullerstig, A., Doscher, R., Graham, L. P., Jones, C., Meier, H. E. M., Samuelsson, P. and Willen, U. (2004) 'European climate in the late twenty-first century: regional simulations with two driving global models and two forcing scenarios', *Climate Dynamics*, 22(1), 13-31.

Raisanen, J. and Joelsson, R. (2001) 'Changes in average and extreme precipitation in two regional climate model experiments', *Tellus Series a-Dynamic Meteorology and Oceanography*, 53(5), 547-566.

Raisanen, J. and Palmer, T. N. (2001) 'A probability and decision-model analysis of a multimodel ensemble of climate change simulations', *Journal of Climate*, 14(15), 3212-3226.

Raisanen, J., Rummukainen, M. and Ullerstig, A. (2001) 'Downscaling of greenhouse gas induced climate change in two GCMs with the Rossby Centre regional climate model for northern Europe', *Tellus Series a-Dynamic Meteorology and Oceanography*, 53(2), 168-191.

Reichler, T. and Kim, J. (2008) 'How well do coupled models simulate today's climate?', *Bulletin of the American Meteorological Society*, 89(3), 303.

Richman, M. B. (1993) 'Comments on - The effect of domain shape on Principal Components Analysis', *International Journal of Climatology*, 13(2), 203-218.

Rinke, A., Dethloff, K., Cassano, J. J., Christensen, J. H., Curry, J. A., Du, P., Girard, E., Haugen, J. E., Jacob, D., Jones, C. G., Koltzow, M., Laprise, R., Lynch, A. H., Pfeifer, S., Serreze, M. C., Shaw, M. J., Tjernstrom, M., Wyser, K. and Zagar, M. (2006) 'Evaluation of an ensemble of Arctic regional climate models: spatiotemporal fields during the SHEBA year', *Climate Dynamics*, 26(5), 459-472.

Robertson, D. P. and Hull, R. B. (2001) 'Beyond biology: toward a more public ecology for conservation', *Conservation Biology*, 15(4), 970-979.

Robinson, P. J. and Finkelstein, P. L. (1991) 'The develpement of impact-oriented climate scenarios', *Bulletin of the American Meteorological Society*, 72(4), 481-490.

Rockel, B. and Woth, K. (2007) 'Extremes of near-surface wind speed over Europe and their future changes as estimated from an ensemble of RCM simulations', *Climatic Change*, 81, 267-280.

Rodwell, M. J., Rowell, D. P. and Folland, C. K. (1999) 'Oceanic forcing of the wintertime North Atlantic Oscillation and European climate', *Nature*, 398(6725), 320-323.

Roeckner, E., Arpe, K., Bengtsson, L., Christoph, M., Claussen, M., Dümenil, L., Esch, M., Giorgetta, M., Schlese, U. and Schulzweida, U. (1996) *The atmospheric general circulation model ECHAM-4: model description and simulation of present day climate*, Hamburg: Max Planck Institute for Meteorology.

Rotmans, J. and van Asselt, M. B. A. (2001) 'Uncertainty management in integrated assessment modelling: Towards a pluralistic approach', *Environmental Monitoring and Assessment*, 69(2), 101-130.

Rummukainen, M. (2010) 'State-of-the-art with regional climate models', *Wiley Interdisciplinary Reviews: Climate Change*, 1(1), 82–96.

Rykiel, E. J. (1996) 'Testing ecological models: The meaning of validation', *Ecological Modelling*, 90(3), 229-244.

Saether, B. E., Tufto, J., Engen, S., Jerstad, K., Rostad, O. W. and Skatan, J. E. (2000) 'Population dynamical consequences of climate change for a small temperate songbird', *Science*, 287(5454), 854-856.

Sanchez, E., Romera, R., Gaertner, M. A., Gallardo, C. and Castro, M. (2009) 'A weighting proposal for an ensemble of regional climate models over Europe driven by 1961-2000 ERA40 based on monthly precipitation probability density functions', *Atmospheric Science Letters*, 10(4), 241-248.

Sanderson, B. M., Piani, C., Ingram, W. J., Stone, D. A. and Allen, M. R. (2008) 'Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations', *Climate Dynamics*, 30(2-3), 175-190.

Sato, T., Kimura, F. and Kitoh, A. (2007) 'Projection of global warming onto regional precipitation over Mongolia using a regional climate model', *Journal of Hydrology*, 333(1), 144-154.

Saunders, M. A., Qian, B. D. and Lloyd-Hughes, B. (2003) 'Summer snow extent heralding of the winter North Atlantic Oscillation', *Geophysical Research Letters*, 30(7).

Sawilowsky, S. S. and Blair, R. C. (1992) 'A more realistic look at the robustness and type-II error properties of the t-test to departures from population normality', *Psychological Bulletin*, 111(2), 352-360.

Scheraga, J. D. and Grambsch, A. E. (1998) 'Risks, opportunities, and adaptation to climate change', *Climate Research*, 11(1), 85-95.

Scherrer, S. C., Appenzeller, C., Liniger, M. A. and Schar, C. (2005) 'European temperature distribution changes in observations and climate change scenarios', *Geophysical Research Letters*, 32(19), 5.

Schmidli, J., Goodess, C. M., Frei, C., Haylock, M. R., Hundecha, Y., Ribalaygua, J. and Schmith, T. (2007) 'Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps', *Journal of Geophysical Research-Atmospheres,* 112(D4).

Schmittner, A., Latif, M. and Schneider, B. (2005) 'Model projections of the North Atlantic thermohaline circulation for the 21st century assessed by observations', *Geophysical Research Letters,* 32(23).

Schneider, S. H. (1983) 'CO2, Climate and Society: A Brief Overview' in Chen, R. S., Boulding, E. and Schneider, S. H. (Eds.), *Social Science Research and Climate Change: In Interdisciplinary Appraisal,* Boston: D. Reidel.

Schoof, J. T. and Pryor, S. C. (2001) 'Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks', *International Journal of Climatology,* 21(7), 773-790.

Schwartz, S. E. (2008) 'Uncertainty in climate sensitivity: Causes, consequences, challenges', *Energy & Environmental Science,* 1(4), 430-453.

Schwierz, C., Appenzeller, C., Davies, H. C., Liniger, M. A., Muller, W., Stocker, T. F. and Yoshimori, M. (2006) 'Challenges posed by and approaches to the study of seasonal-to-decadal climate variability', *Climatic Change,* 79(1-2), 31-63.

Scott, D. B., Collins, E. S., Gayes, P. T. and Wright, E. (2003) 'Records of prehistoric hurricanes on the South Carolina coast based on micropaleontological and sedimentological evidence, with comparison to other Atlantic Coast records', *Geological Society of America Bulletin,* 115(9), 1027-1039.

Senior, C. A. (1999) 'Comparison of mechanisms of cloud-climate feedbacks in GCMs', *Journal of Climate,* 12(5), 1480-1489.

Shapiro, S. S. and Wilk, M. B. (1965) 'An analysis of variance test for normality (complete samples)', *Biometrika,* 52, 591.

Smit, B. and Wandel, J. (2006) 'Adaptation, adaptive capacity and vulnerability', *Global Environmental Change-Human and Policy Dimensions,* 16(3), 282-292.

Smith, P. (2004) 'Carbon sequestration in croplands: the potential in Europe and the global context', *European Journal of Agronomy,* 20(3), 229-236.

Spak, S., Holloway, T., Lynn, B. and Goldberg, R. (2007) 'A comparison of statistical and dynamical downscaling for surface temperature in North America', *Journal of Geophysical Research-Atmospheres,* 112(D8).

Stainforth, D.A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D.J., Kettleborough, J.A., Knight, S., Martin, A., Murphy, J.M., Piani, C., Sexton,

D., Smith, L.A., Spicer, R.A., Thorpe, A.J. and Allen, M.R. (2005) 'Uncertainty in predictions of the climate response to rising levels of greenhouse gases', *Nature*, 433, 403–406.

Stainforth, D. A., Downing, T. E., Washington, R., Lopez, A. and New, M. (2007) 'Issues in the interpretation of climate model ensembles to inform decisions', *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, 365(1857), 2163-2177.

Stefanova, L. and Krishnamurti, T. N. (2002) 'Interpretation of seasonal climate forecast using Brier skill score, the Florida State University superensemble, and the AMIP-I dataset', *Journal of Climate*, 15(5), 537-544.

Stephenson, D. B., Pavan, V., Collins, M., Junge, M. M., Quadrelli, R. and Participating, C. M. G. (2006) 'North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment', *Climate Dynamics*, 27(4), 401-420.

Stone, D. A., Allen, M. R., Selten, F., Kliphuis, M. and Stott, P. A. (2007) 'The detection and attribution of climate change using an ensemble of opportunity', *Journal of Climate*, 20(3), 504-516.

Streets, D. G. and Glantz, M. H. (2000) 'Exploring the concept of climate surprise', *Global Environmental Change-Human and Policy Dimensions*, 10(2), 97-107.

Sugiyama, T. (2005) 'Scenario Analyses for the Future Climate Regime', *International Environmental Agreements: Politics, Law and Economics*, 5(1), 1-3.

Sutton, R.T. and Hodson, D.L.R. (2005) 'Atlantic Ocean forcing of North American and European summer climate', *Science* 309, 115-118.

Sweeney, J. (1989) 'The Climatic Influence of the Irish Sea' in Sweeney, J. (Ed.) *The Irish Sea: Resource at risk*, Dublin: Geographical Society of Ireland, 47-57.

Sweeney, J. and Fealy, R. (2003) *Establishing reference climate scenarios for Ireland*, Dublin: Environmental Protection Agency.

Sweeney, J. C. (1985) 'The changing synoptic origins of Irish precipitation', *Transactions of the Institute of British Geographers*, 10(4), 467-480.

Tanaka, K., Raddatz, T., O'Neill, BC. and Reick, C.H. (2009) 'Insufficient forcing uncertainty underestimates the risk of high climate sensitivity' *Geophysical Research Letters* 36, L16709

Tannert, C., Elvers, H. D. and Jandrig, B. (2007) 'The ethics of uncertainty - In the light of possible dangers, research becomes a moral duty', *Embo Reports*, 8(10), 892-896.

Tanser, F. C., Sharp, B. and le Sueur, D. (2003) 'Potential effect of climate change on malaria transmission in Africa', *Lancet,* 362(9398), 1792-1798.

Tebaldi, C. and Knutti, R. (2007) 'The use of the multi-model ensemble in probabilistic climate projections', *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences,* 365(1857), 2053-2075.

Tebaldi, C., Mearns, L. O., Nychka, D., Smith, R. L. and By (2004) 'Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations', *Geophysical Research Letters,* 31(24), 5.

Tebaldi, C., Smith, R. L., Nychka, D. and Mearns, L. O. (2005) 'Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles', *Journal of Climate,* 18(10), 1524-1540.

Thomas, J. A., Telfer, M. G., Roy, D. B., Preston, C. D., Greenwood, J. J. D., Asher, J., Fox, R., Clarke, R. T. and Lawton, J. H. (2004) 'Comparative losses of British butterflies, birds, and plants and the global extinction crisis', *Science,* 303(5665), 1879-1881.

Tol, R. S. J. (2005) 'Adaptation and mitigation: trade-offs in substance and methods', *Environmental Science & Policy,* 8(6), 572-578.

Tracton, M. S. and Kalnay, E. (1993) 'Operational ensemble prediction at the National Meteorological Center - Practical aspects', *Weather and Forecasting,* 8(3), 379-398.

Trigo, R. M. and Palutikof, J. P. (2001) 'Precipitation scenarios over Iberia: A comparison between direct GCM output and different downscaling techniques', *Journal of Climate,* 14(23), 4422-4446.

Tukey, J. (1962) 'The future of data analysis', *Annals of Mathematical Statistics,* 33(1).

Vallis, G. K. and Gerber, E. P. (2008) 'Local and hemispheric dynamics of the North Atlantic Oscillation, annular patterns and the zonal index', *Dynamics of Atmospheres and Oceans,* 44(3-4), 184-212.

van den Dool, H. (2007) *Empirical methods in short-term climate prediction,* Oxford: Oxford University Press.

van der Keur, P., Henriksen, H. J., Refsgaard, J. C., Brugnach, M., Pahl-Wostl, C., Dewulf, A. and Buiteveld, H. (2008) 'Identification of major sources of uncertainty in current IWRM practice. Illustrated for the Rhine basin', *Water Resources Management,* 22(11), 1677-1708.

van Ulden, A., Lenderink, G., van den Hurk, B. and van Meijgaard, E. (2007) 'Circulation statistics and climate change in Central Europe: PRUDENCE simulations and observations', *Climatic Change,* 81, 179-192.

Veizer, J. (2005) 'Celestial climate driver: A perspective from four billion years of the carbon cycle', *Geoscience Canada,* 32(1), 13-28.

Vidale, P. L., Luthi, D., Frei, C., Seneviratne, S. I. and Schar, C. (2003) 'Predictability and uncertainty in a regional climate model', *Journal of Geophysical Research-Atmospheres,* 108(D18), 23.

Vidale, P. L., Luthi, D., Wegmann, R. and Schar, C. (2007) 'European summer climate variability in a heterogeneous multi-model ensemble', *Climatic Change,* 81, 209-232.

Vimont, D. J. (2005) 'The contribution of the interannual ENSO cycle to the spatial pattern of decadal ENSO-like variability', *Journal of Climate,* 18(12), 2080-2092.

Vizcaino, M., Mikolajewicz, U., Jungclaus, J. and Schurgers, G. (2010) 'Climate modification by future ice sheet changes and consequences for ice sheet mass balance', *Climate Dynamics,* 34(2-3), 301-324.

Wang, G. L. and Schimel, D. (2003) 'Climate change, climate modes, and climate impacts', *Annual Review of Environment and Resources,* 28, 1-28.

Wang, S. Y., McGrath, R., Semmler, T. and Sweeney, C. (2006) 'Validation of simulated precipitation patterns over Ireland for the period 1961-2000', *International Journal of Climatology,* 26(2), 251-266.

Wang, W. and Seaman, N. L. (1997) 'A comparison study of convective parameterization schemes in a mesoscale model', *Monthly Weather Review,* 125(2), 252-278.

Watson, A. J. and Lovelock, J. E. (1983) 'Biological homeostasis of the global environment - The parable of Daisyworld', *Tellus Series B-Chemical and Physical Meteorology,* 35(4), 284-289.

Weigel, A. P., Liniger, M. A. and Appenzeller, C. (2008) 'Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?', *Quarterly Journal of the Royal Meteorological Society,* 134(630), 241-260.

Weisheimer, A., Doblas-Reyes, F.J., Palmer, T.N., Alessandri, A., Arribas, A., Déqué, M., Keenlyside, N., MacVean, M., Navarra, A. and Rogel, P. (2009) 'ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions — Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs', *Geophysical Research Letters,* 36, L21711.

Wetherald, R. T., Stouffer, R. J. and Dixon, K. W. (2001) 'Committed warming and its implications for climate change', *Geophysical Research Letters*, 28(8), 1535-1538.

Whitaker, J. S. and Loughe, A. F. (1998) 'The relationship between ensemble spread and ensemble mean skill', *Monthly Weather Review*, 126(12), 3292-3302.

Wibig, J. (1999) 'Precipitation in Europe in relation to circulation patterns at the 500 hPa level', *International Journal of Climatology*, 19(3), 253-269.

Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A. and Snyder, M. A. (2009) *Niches, models, and climate change: assessing the assumptions and uncertainties*, translated by National Academy of Sciences, 19729-19736.

Wigley, T. M. L. and Raper, S. C. B. (2001) 'Interpretation of high projections for global-mean warming', *Science*, 293(5529), 451-454.

Wilby, R. L., Dawson, C. W. and Barrow, E. M. (2002) 'SDSM - a decision support tool for the assessment of regional climate change impacts', *Environmental Modelling & Software*, 17(2), 147-159.

Wilby, R. L. and Dessai, S. (2010) 'Robust adaptation to climate change', *Weather*, 65(7), 180-185.

Wilks, D. S. (2006) *Statistical Methods in the Atmospheric Sciences*, London: Academic Press.

Wu, W. L., Lynch, A. H. and Rivers, A. (2005) 'Estimating the uncertainty in a regional climate model related to initial and lateral boundary conditions', *Journal of Climate*, 18(7), 917-933.

Yang, Z. W. and Arritt, R. W. (2002) 'Tests of a perturbed physics ensemble approach for regional climate modelling', *Journal of Climate*, 15(20), 2881-2896.

Yarnal, B. (1993) Synoptic *climatology in environmental analysis: A primer*, London: Belhaven Press.

Yeh, S. W. and Kirtman, B. P. (2007) 'ENSO amplitude changes due to climate change projections in different coupled models', *Journal of Climate*, 20(2), 203-217.

Yokohata, T., Emori, S., Nozawa, T., Ogura, T., Kawamiya, M., Tsushima, Y., Suzuki, T., Yukimoto, S., Abe-Ouchi, A., Hasumi, H., Sumi, A. and Kimoto, M. (2008) 'Comparison of equilibrium and transient responses to CO2 increase in eight state-of-the-art climate models', *Tellus Series a-Dynamic Meteorology and Oceanography* 60, 946-961.

Yun, W. T., Stefanova, L., Mitra, A. K., Kumar, T., Dewar, W. and Krishnamurti, T. N. (2005) 'A multi-model superensemble algorithm for seasonal climate

prediction using DEMETER forecasts', *Tellus Series a-Dynamic Meteorology and Oceanography*, 57(3), 280-289.

Zhang, X. B., Zwiers, F. W., Hegerl, G. C., Lambert, F. H., Gillett, N. P., Solomon, S., Stott, P. A. and Nozawa, T. (2007) 'Detection of human influence on twentieth-century precipitation trends', *Nature*, 448(7152), 461-U4.

Zheng, X. G. and Frederiksen, C. S. (1999) 'Validating interannual variability in an ensemble of AGCM simulations', *Journal of Climate*, 12(8), 2386-2396.

Zorita, E. and von Storch, H. (1999) 'The analog method as a simple statistical downscaling technique: Comparison with more complicated methods', *Journal of Climate*, 12(8), 2474-2489.