Interactions between large-scale modes of climate variability that influence Australian hydroclimatic regimes

by

Kirien Whan

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

Doctor of Philosophy

of the Australian National University

August 2013

To Caz,

For always making me believe that I could.

Candidate's Declaration

This thesis contains no material that has been accepted for the award of any other degree or diploma in any university. To the best of the author's knowledge, it contains no material previously published or written by another person, except where due reference is made in the text.

Kirien Whan Date: 26th August 2013

Acknowledgements

This thesis would not have been possible without the help of many people. First and foremost, I am grateful to my supervisor, Dr Janette Lindesay, who has supported me from my undergraduate studies through to my PhD thesis with patience and guidance. I also wish to extend my sincerest thanks to Dr Bertrand Timbal, Dr Michael Raupach and Dr Emlyn Williams whose guidance and support has been invaluable. I appreciate the time all my supervisors spent helping me grow into the scientist I am.

Thanks to the PhD students, colleagues and staff at both The Fenner School of Environment and Society (ANU) and the Climate Change Research Centre (UNSW) who provided happy, supportive and fun research environments through various stages of my PhD. I am grateful for the support and understanding of Dr Lisa Alexander during my time at UNSW when I was working fulltime and finishing PhD. Thanks also to Dr Peter Briggs and Clive Hillier for data and technical support. Thanks to the South-East Australian Climate Initiative (SEACI) for the financial, technical and scientific support throughout my candidature. This research would not have been possible with out the development of the Australian Water Availability Project (AWAP) data set.

I could not have made it through this PhD-rollercoaster without the love, support, encouragement and kindness of Vanessa Bateman – I am truly thankful and indebted. Thanks to my family for always giving me their unconditional support – thanks Dad, Dot and Tazz. I am grateful to Tazz Whan and Sarlae McAlpine, Bronwen Jones, the Wolks and the Beans for housing me on various trips back to Canberra. “High-fives” all my friends for tolerating PhD-stories and all the excellent distractions.

Abstract

Effective management of water resources, including surface and ground water, is vital and relies on a thorough understanding of climatic and hydrological (or 'hydroclimatic') variability. In Australia hydroclimatic variability is associated with several large-scale climate modes, including remote phenomena such as El Niño – Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD), and more regional climate indices such as the sub-tropical ridge (STR). Individually, the large-scale climate regimes typically associated with rainfall events are well understood. However, less is known about the interactions between, or combinations of, different large-scale conditions that influence Australian hydroclimatic regimes. These interactions are non-linear so traditional statistical frameworks may be unable to adequately characterise these relationships.

Classification and Regression Trees (CART) are well suited to analysing relationships between predictor and response variables, including those based on categorical events, that may be modulated by several predictor variables acting together. By employing a more appropriate and novel statistical method this thesis aims to better understand relationships between large-scale modes of climate variability and Australian hydroclimatic regimes.

In this work, tree-based models were used to classify regional Australian rainfall regimes from indices of ENSO, the IOD and the STR, yielding the following conclusions.

(1) Interactions between tropical (ENSO, IOD) predictor variables and the STR control the strength of the tropical teleconnection and the influence on regional rainfall regimes in southern Australia. When tropical modes and the STR are in the same phase, rainfall regimes are continent-wide and spatially coherent. However, when indices of climate modes are in the opposite phases, i.e. El Niño combined with low STR intensity, the modulation of the tropical teleconnection by the STR is evident, as rainfall anomalies are confined to the northeast of the continent.

(2) The influence of both STR intensity and position on rainfall regimes in southeastern Australia was defined. STR position was crucial for defining two distinct types of "wet" autumns, a "summer-like" ("winter-like") regime when the STR was in a southerly (northerly) position. The summer-like regime occurs at frequencies that have not changed detectably over the instrumental record. However, the frequency of the winter-like regime has declined significantly. In addition, the dry regime defined by high STR intensity has been the most frequent regime in recent years, consistent with the attribution of STR intensity as the main driver of the Millennium Drought.

(3) The predictive persistence of relationships between a suite of predictor variables (indices of ENSO, IOD and the STR) and rainfall, upper-layer and lower-layer soil moisture was explored. The predictability of spring rainfall was similar using both random forests (a bootstrapping implementation of CART) and linear regression, suggesting results are not dependent on method. The key result, of possible use in seasonal forecasting, is that, deep soil moisture in spring and summer exhibits significantly more predictability than rainfall and shallow soil moisture, due to the persistence of tropical climate drivers and the removal of high-frequency variability in deep layers by natural temporal smoothing as soil moisture is transferred to deep soil layers.

Table of Contents

List of Figures

List of Tables

List of Equations

Glossary and list of acronyms and abbreviations

AWAP – The Australian Water Availability Project

CART – The Classification and Regression Tree method. A type of decision tree. Can either refer to a single tree or the method itself.

Climate drivers – The indices and phenomena that influence climate variability, specifically in this thesis: ENSO, the IOD, the STR.

DJF – December-January-February

DMI – Dipole Mode Index

Drainage divisions – These are defined in Chapter 3. I often use 'regions' and 'drainage divisions' interchangeably

ENSO – El Niño-Southern Oscillation. In Chapter 6, ENSO refers to El Niño events.

GDR – Great Diving Range

Hydroclimate variability – Combined hydrological and climatic variability

IOD – Indian Ocean Dipole

ITCZ – Intertropical Convergence Zone

JJA – June-July-August

MAM – March-April-May

MDB – Murray-Darling Basin

SOI – Southern Oscillation Index

SON – September-October-November

SST – Sea-Surface Temperature

STR – Sub-Tropical Ridge

STR-I – Sub-Tropical Ridge Intensity

STR-P – Sub-Tropical Ridge Position

TPI – Tri-pole Index

Tree-based models – Decision tree type statistical models that predicting a response variable from a set of predictor variables, including individual Classification and Regression Trees or random forests models.

Chapter 1 : Introduction

1.1 Motivation

Current and future fresh water resource management is one of the most important environmental issues globally. Effective management of global fresh water resources, including surface and ground water, is vital and relies on a thorough understanding of climatic and hydrological (or 'hydroclimatic') variability, thereby benefiting society, the environment and economies (Campbell 2005). Globally, droughts have been common over at least the last millennium, often related to large-scale phenomena such as El Niño – Southern Oscillation (ENSO) (Dai 2010). In addition, changes in hydroclimatic variability are expected in the future (Dai 2010, Leblanc et al. 2012) which will likely increase pressure on water resource management.

Australian hydroclimate is highly variable, with periods of above and below average rainfall and runoff prevalent throughout the instrumental record (Australian Bureau of Meteorology 2013b). Drying and warming trends in southeastern Australia over the last 60-years (Figure 1.1) have increased pressure on water resources (Leblanc et al. 2012). This is particularly important since Australia has the lowest runoff per unit area of any inhabited continent (Smith 1998). These pressures are compounded by the changes to the hydrological cycle itself including changing the seasonality of rainfall and hydrological variability (CSIRO 2012, Dore 2005).

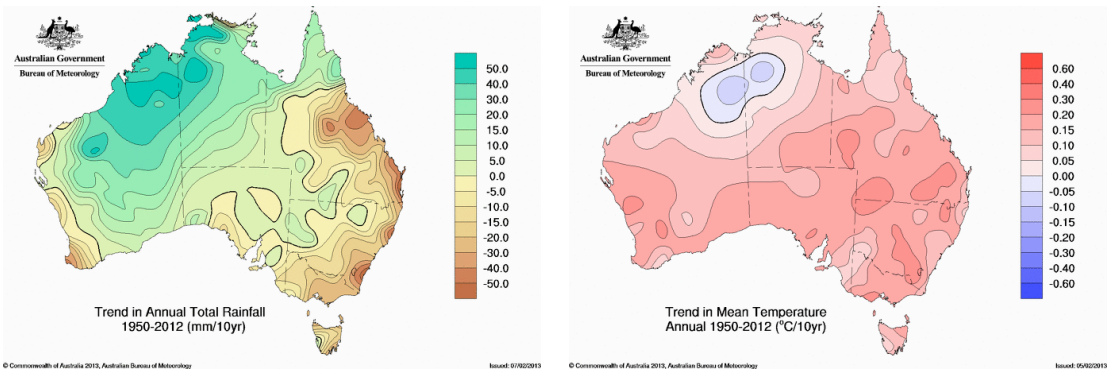

Figure 1.1. Linear trend in Australian annual average total rainfall (left) and mean temperature (right) over the period 1950 – 2012 (Australian Bureau of Meteorology 2013b)

Understanding the large-scale conditions that are associated with hydroclimatic regimes, and the interactions between hydroclimatic variability and 'climate drivers' (i.e. indices of

large-scale modes of climate variability that have relationships with hydroclimates), is critical for natural resource managers in Australia. This is the principal theme of this thesis. Previous research on this topic has focused on southeastern Australia, due to its importance as a food-producing region, heavily reliant on hydrological resources (CSIRO 2012). Here, the focus will be broadened to include other regions in northern and southern Australia so that regional variability in relationships can be assessed.

Numerous climate drivers influence southeastern Australian hydroclimate variability, including ENSO, the Indian Ocean Dipole (IOD) and the sub-tropical ridge (Cai et al. 2011b, CSIRO 2012, Drosdowsky and Chambers 2001, Kirono et al. 2010, McBride and Nicholls 1983, Meyers et al. 2007, Nicholls 1989, Risbey et al. 2009, Timbal and Drosdowsky 2013, Ummenhofer et al. 2009a, Ummenhofer et al. 2011), rendering it a complex region to analyse (Figure 1.2). However, it is critical that we understand the current drivers of hydroclimatic variability to give context for future climate changes and corresponding water resources management. In addition, is it is also vital to understand the i*nteractions* between large-scale modes of climate variability that drive hydroclimatic variability.

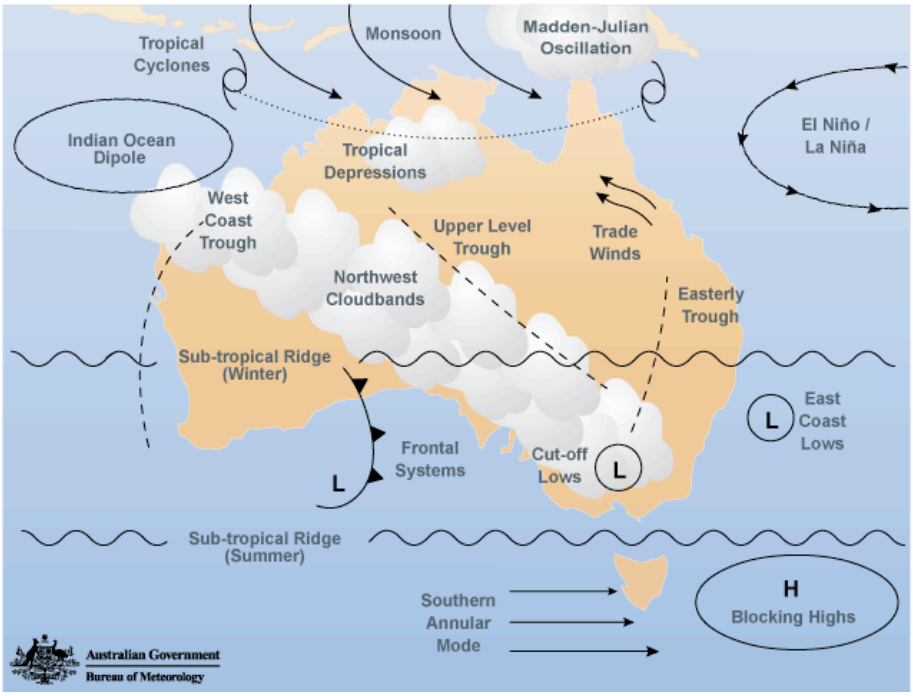

Figure 1.2. Schematic representation of the major Australian climate influences, including El Niño – Southern Oscillation, the Indian Ocean Dipole and the summer and winter positions of the sub-tropical ridge (Australian Bureau of Meteorology 2010) based on Risbey et al., 2009)

Each of the large-scale modes that govern Australia's climate has been individually studied in the scientific literature (see Chapter 2), and, increasingly, interactions between climate drivers is the focus of research (Ummenhofer et al. 2009a, Ummenhofer et al. 2009b). Linear statistics have been the cornerstone of climate research, with statistical techniques such as correlation, regression and principal components analysis forming the foundation of the scientific knowledge base (Drosdowsky and Williams 1991, McBride and Nicholls 1983, Nicholls 1989, Timbal and Drosdowsky 2013, Timbal and Hendon 2011). However, many classical statistical techniques are limited to capturing the linear portion of the variability (Campbell et al. 2000) while research into the interactions between hydroclimatic variability and large-scale modes of climate variability is complicated by non-linear relationships. Research within the classical linear framework has increased scientific understanding of many climate processes that influence Australia, particularly those related to tropical influences (Drosdowsky and Williams 1991, Evans and Allan 1992, Nicholls 1989, Power et al. 1998). Yet important questions remain about the nature of interactions between the large-scale modes of climate variability that influence Australian hydroclimate. This thesis endeavours to answer such questions.

Rainfall, and by extension all hydroclimatic variability, is provisional upon specific local conditions, e.g. sufficient moisture availability, uplift or instability. These local conditions are, in turn, associated with larger-scale regimes, so that rainfall events are much more likely under particular large-scale conditions than others, e.g. La Niña compared to El Niño events. In addition, the large-scale conditions associated with rainfall can be viewed as discrete events that must be combined in particular ways to make rainfall possible. As such, a Classification and Regression Trees (CART) analysis may be more suited to the analysis of discrete events, such as rainfall or the large-scale conditions associated with rainfall.

Statistical models must be able to capture interactions in the system, such as regime-dependent behaviours (Campbell 2005). Tree-based methods, which include CART (see Chapter 3) (Breiman et al. 1984), are a set of statistical techniques that facilitate exploration of discrete climate processes. Tree-based methods therefore have considerable potential for examining the relationships between large-scale climate modes and Australian hydroclimatic variability. It is not suggested that tree-based models should be used to the exclusion of the linear statistical method, but all statistical methods have various strengths and weaknesses. Indeed, one strength of CART is its ability to explore relationships between categorical event based response variables and multiple predictor variables. A classical statistical framework has worked well in the past to describe the linear aspects and lower dimensional problems (i.e. influence of individual drivers on hydroclimatic variability). But as we increasingly consider event-based, non-linear and multi-dimensional problems we need to move to a statistical framework that has more ability to capture this complexity. The economic, social and environmental impacts of changing climate regimes are extensive so there is an obligation to utilise the most appropriate methods to understand the event-based, conditional nature of relationships between large-scale modes of climate variability and Australian hydroclimatic regimes.

.

1.2 Thesis Aim

The aim of the research reported in this thesis is to increase understanding of interactions between large-scale modes of climate variability that influence Australian hydroclimatic regimes. More specifically, a novel statistical tree-based method will be applied to the Australian Water Availability Project data set (AWAP – see Chapter 3) that will enable an assessment of the interactions between large-scale climate modes that are important for Australian hydroclimatic regimes across the continent. This will increase our understanding of interactions between large-scale modes of climate variability across the hydrological cycle, thereby answering the key research questions below:

- What is the current state of understanding on the topic (Chapter 2), including the existing datasets to study the problem (Chapter 3)?
- Are the known relationships between indices of large-scale climate variability and Australian hydroclimatic regimes reflected in the AWAP data set (Chapter 4)?
- What can relationships between Indo-Pacific sea-surface temperature variability and the sub-tropical ridge interactions tell us about regional Australian hydroclimatic regimes (Chapter 5)?
- What additional information about spatial and temporal relationships between Australian hydroclimatic regimes and large-scale modes of climate variability can be gained through a tree-based analysis (Chapter 6)?
- How much predictive information is available in CART-based relationships between Australian hydroclimatic regimes and large-scale modes of climate variability, and can such predictive information be applied to problems of seasonal hydroclimatic forecasting (Chapter 7)?

1.3 Significance

The new developments presented here are:

1. A comprehensive analysis of linear relationships between regional and remote large-scale modes of climate variability and Australian hydroclimate using the AWAP data set (Chapter 4);
2. A demonstration of the importance of tropical – sub-tropical interactions in modulating the spatial extent of ENSO events (Chapter 5);

3. A demonstration of the importance of STR position, in addition to the role of STR intensity, in determining rainfall regimes in southeastern Australia (Chapter 6); and
4. A demonstration of the greater predictability of lower-layer soil moisture over rainfall and upper-layer soil moisture (Chapter 7);

1.4 Structure of this thesis

This thesis combines traditional thesis chapters with chapters in the form of published papers (Table 1.1).

Table 1.1. The structure of the chapters in this thesis and the percentage contribution of each chapter by Kirien Whan.

Chapter	Structure	Contribution
Chapter 1: Introduction	Thesis chapter	100%
Chapter 2: Literature Review	Thesis chapter	100%
Chapter 3: Data and Methods	Thesis chapter	100%
Chapter 4: Linear analysis	Thesis chapter	100%
Chapter 5: Tropical – sub-tropical interactions	Thesis chapter	100%
Chapter 6: STR intensity and position interactions	Paper published	70%
Chapter 7: Predictability of hydroclimatic regimes	Thesis chapter	70%
Chapter 8: Synthesis and Conclusions	Thesis chapter	100%

A higher-level literature review and explanation of data and methods are presented next in Chapters 2 and 3, respectively, to provide an overall context to the thesis. Following this are two results chapters structured as traditional thesis chapters (Chapters 4 and 5). Chapter 4 establishes the linear relationships between indices of large-scale modes of climate variability and Australian hydroclimate, as expressed in the AWAP data set. The first of the tree-based analyses are presented in Chapter 5, which contains an introduction to the tree-based methods and then focuses on tropical – sub-tropical interactions that affect regional Australian rainfall variability. The analysis in Chapter 6 considers the roles of STR intensity and position in determining rainfall regimes in southeastern Australia and has been published in the International Journal of Climatology (Whan et al. 2013). Bertrand Timbal and Janette Lindesay assisted me in framing the problem. I conducted the analysis but had assistance from Bertrand Timbal with the interpretation of results and writing the paper. Chapter 7 explores the predictability of hydroclimatic regimes using linear regression and random forests. I conducted the analysis but had assistance from Michael Raupach (problem framing and interpretation), Janette Lindesay (problem framing) and Emlyn Williams (interpretation). Finally, Chapter 8 summarises the conclusions and novel aspects of this work, and presents suggestions of future work.

Chapter 2 : Literature Review

2.1 Chapter summary and context

This chapter will introduce the major features and current understanding of Australian hydroclimatic variability and trends (Section 2.2). To preface subsequent discussion of how interactions between remote large-scale modes of climate variability influence hydroclimatic regimes in Australia, the current understanding of the mechanisms and impacts of these large-scale modes is discussed in Section 2.3 and relevant research gaps are identified. Finally, the predictive capability of tree-based methods is explored in Chapter 7 so some discussion of the predictability of Australian hydroclimatic regimes will be outlined in Section 2.4.

2.2 Background of Australian climate and hydrology

This section will outline the global climate features that broadly define Australian climate on seasonal timescales along with discussions of spatial and temporal hydroclimatic variability in Australia (Section 2.2.1). This is followed by a summary of interactions between climatic and hydrological variability (Section 2.2.2).

2.2.1 Hydroclimatic variability and trends

The majority of the Australian population resides in the temperate southeast (Campbell and Palmer 2010) with higher rainfall recorded in the north, south and east coasts, while a large portion of the Australian continent is arid or semi-arid (Barros and Bowden 2008). Australia is a large continent stretching from the tropics to the mid-latitudes and under the influence of several features of the global circulation, namely the Hadley and Walker Circulations (Figure 2.1). The Hadley Cell is a meridional circulation driven by differences in surface heating between the equator and the poles. An excess of heat at the equator causes convection and the upward movement of air that is known as the Intertropical Convergence Zone (ITCZ). Air and heat travel poleward in the upper atmosphere, descending in the sub-tropics. This region of subsidence is known as the sub-tropical ridge (STR). The annual meridional movement in the ITCZ, and concurrent shift in the STR, controls the seasonality of Australian rainfall including the Australian Monsoon (Sturman and Tapper 2005) and the position of the mid-latitude storm track (Drosdowsky 2005, Peixoto and Oort 1992, Sturman and Tapper 2005, Timbal and Drosdowsky 2013).

Zonal circulations in all ocean basins modulate the meridional Hadley Cell. The largest and most well known of these zonal features that influences Australian climate variability is the Walker Circulation. Driven by differences in sea surface temperature (SST) between the tropical eastern and western Pacific Ocean, the Walker Circulation (or Southern Oscillation) is the atmospheric branch of the coupled ocean-atmosphere El Niño-Southern Oscillation (for further discussion of ENSO see Section 2.3.1).

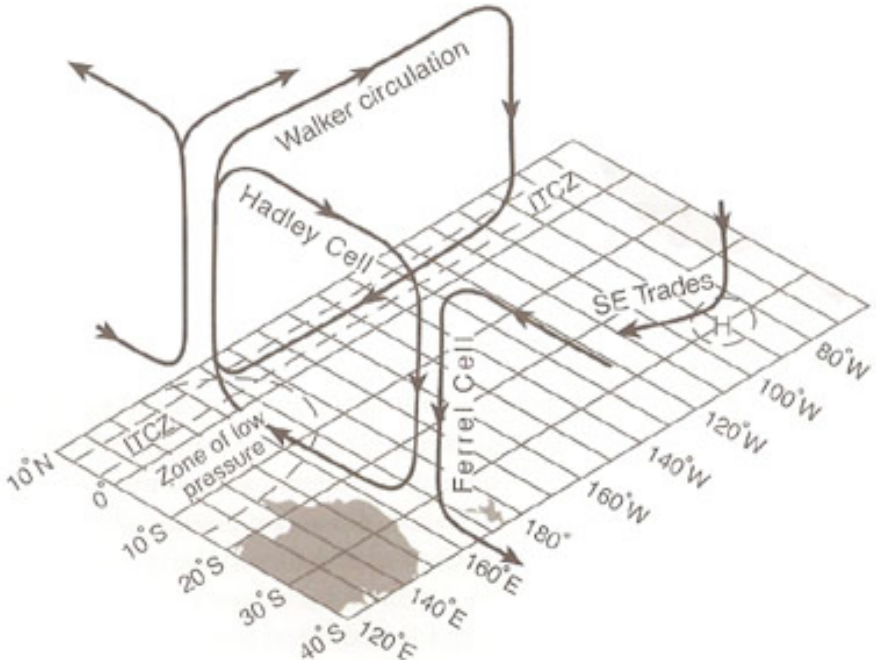

Figure 2.1. The zonal and meridional elements of the global circulation that are relevant to Australia, including the Hadley and Walker Circulations (Sturman and Tapper 2005).

A majority of research has focused on rainfall because of its practical importance, its superior observational network, and the importance of significant cumulative effects such as flood and drought (Pittock 1975). Annual average rainfall is highest in northern and eastern Australia, along the Great Dividing Range (GDR), and in western Tasmania (Figure 2.2). Mean seasonal rainfall varies greatly throughout the year and is controlled on the broad scale by seasonal movements in the Hadley Circulation. As such, summer rainfall dominates in tropical northern Australia as the ITCZ is located over northern Australia during the monsoon season (Sturman and Tapper 2005). In contrast, parts of southern Australia are dominated by winter rainfall as the location of the STR at lower latitudes allows the passage of frontal systems, imbedded in the mid-latitude westerly airstream, over southern Australia. The location of the STR at higher latitudes in summer results in a southerly shift in the mid-latitude storm track that pushes these rain-bearing systems to the south of the continent (Drosdowsky 2005, Peixoto and Oort 1992, Sturman and Tapper 2005, Timbal and Drosdowsky 2013). Rainfall in southern

Australian is also influenced by topography, with increased rainfall along the Great Dividing Range (GDR) in all seasons (Figure 2.2). Southeasterly surface trade winds are associated with the mean Walker Circulation (Figure 2.1) and result in higher rainfall across the east coast, particularly in northern Australia (Figure 2.2).

a)

b)

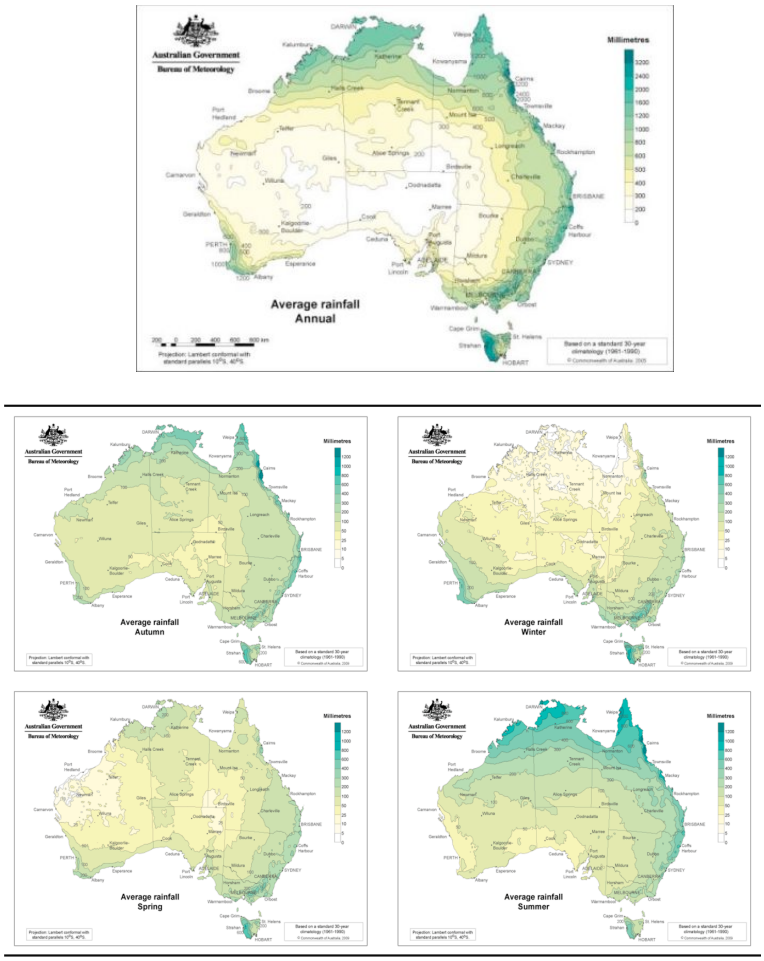

Figure 2.2. a) Annual average Australian rainfall (1961-1900) and b) Mean seasonal Australian rainfall (1961-1990) in autumn (top left), winter (top right), spring (bottom left) and summer (bottom right) (Australian Bureau of Meteorology 2011b).

The variability in Australian rainfall changes seasonally, with most variability for southern Australia focused in the late summer – early autumn period (Figure 2.3). The tropical climate drivers are active from winter to summer when there is less rainfall variability. This suggests that random weather noise during the late summer – autumn period, when the large-scale modes are inactive, is responsible for the largest variability (see Sections 2.3.1 and 2.3.2 for further discussion of Pacific and Indian Ocean climate drivers). There is limited understanding of how

large-scale climate drivers interact with smaller scale weather features (see subsequent sections for some discussion of this). These interactions are the focus of Chapter 5.

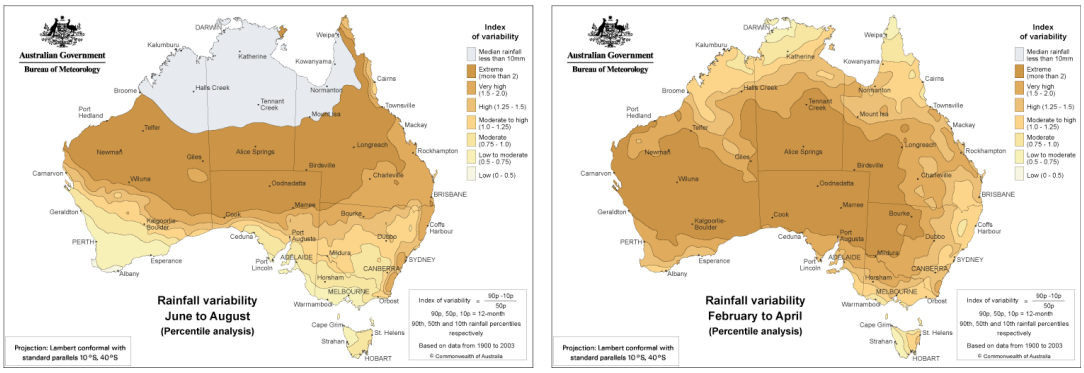

Figure 2.3. Variability of seasonal Australian rainfall in the low variability June-August (left), and the high variability February-April (right), calculated over the period 1900 – 2003. Variability is calculated by dividing the difference between the 90th and 10th percentiles by the 50th percentile (Australian Bureau of Meteorology 2011b).

Trends in rainfall, over the period 1970 – 2010, vary seasonally and spatially over the continent (Figure 2.4). The northwest of Australia has experienced large increases in rainfall, particularly in summer when the bulk of rainfall is delivered to the region, while rainfall declines are evident in autumn in the southeast (Murphy and Timbal 2008). During the Millennium Drought (1997 – 2009), southeast Australia experienced a rainfall deficit with the driest 13-year period on record that was 45% drier than the next driest period (CSIRO 2012, Timbal and Drosdowsky 2013). During this time no stations in northwest Victoria experienced an extremely wet season (Pook et al. 2009) and this has occurred as part of a more widespread decline with decile one rainfall apparent across the entire MDB catchment (Timbal 2009). This rainfall deficit occurred primarily in the cool season, from March – October (Timbal and Drosdowsky 2013), with a majority (60%) of the decline in autumn (Cai and Cowan 2008a, Cai et al. 2009a, Kiem and Verdon-Kidd 2010, Murphy and Timbal 2008). This long dry spell ended in 2010 when a strong La Niña event was associated with heavy rainfall in southeastern Australia (Australian Bureau of Meteorology 2011a).

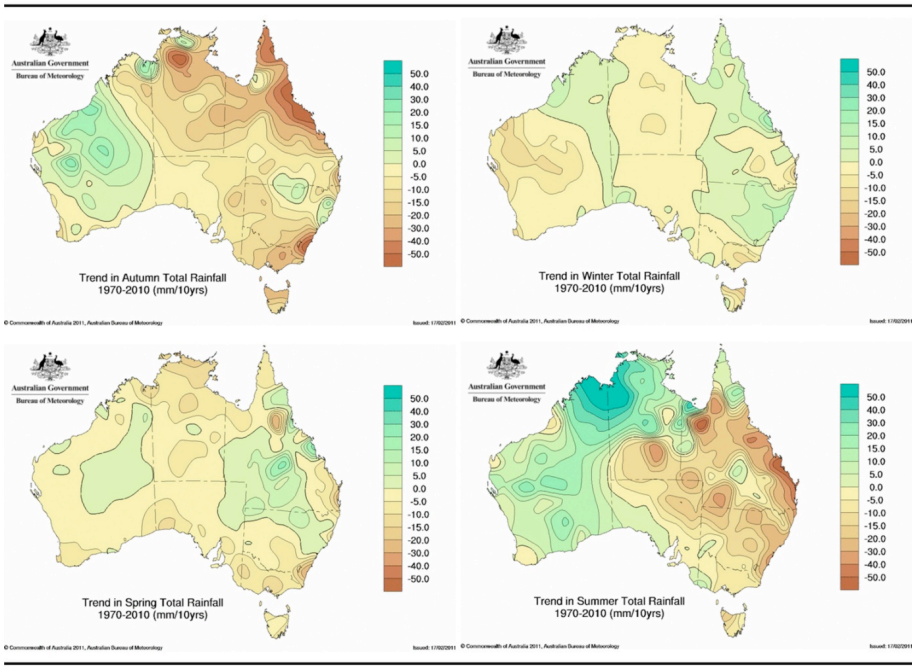

Figure 2.4. Seasonal Australian rainfall trends over the period 1970 – 2010 in autumn (top left), winter (top right), spring (bottom left) and summer (bottom right) (Australian Bureau of Meteorology 2011b).

Trends in rainfall are less robust over different time periods as decadal and interannual variability dominates the rainfall signal (Figure 2.5). Understanding the drivers of this interannual rainfall variability is a key issue for Australia. A great deal of research is focused in this area but significant research gaps remain that are explored throughout this thesis.

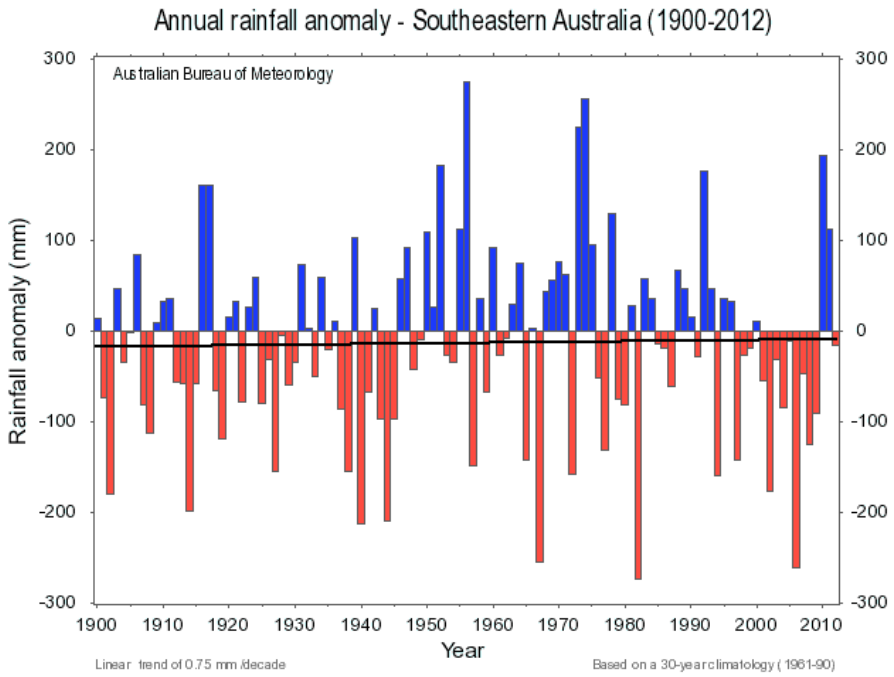

Figure 2.5. Southeastern Australian annual rainfall anomalies (from the 1960-1990 climatology), over the period 1900 – 2012 (Australian Bureau of Meteorology 2011b).

Seasonal and temporal hydrological variability broadly mirrors that of rainfall as it responds directly to changes in rainfall and temperature. Regionally, annual average runoff (shown as a percentage of total Australian runoff in Figure 2.6) varies in a broadly similar way to rainfall (Figure 2.2) with the largest values in northern and eastern Australia. Runoff seasonality also reflects rainfall, with most runoff produced in southeast Australia in winter and spring when rainfall levels are high and temperatures are low (Kirono et al. 2010). Runoff has decreased in Australia (Cai et al. 2009a) and in other large regions globally, even when accounting for human uses, despite little trend in rainfall levels (e.g. Fu et al., 2004).

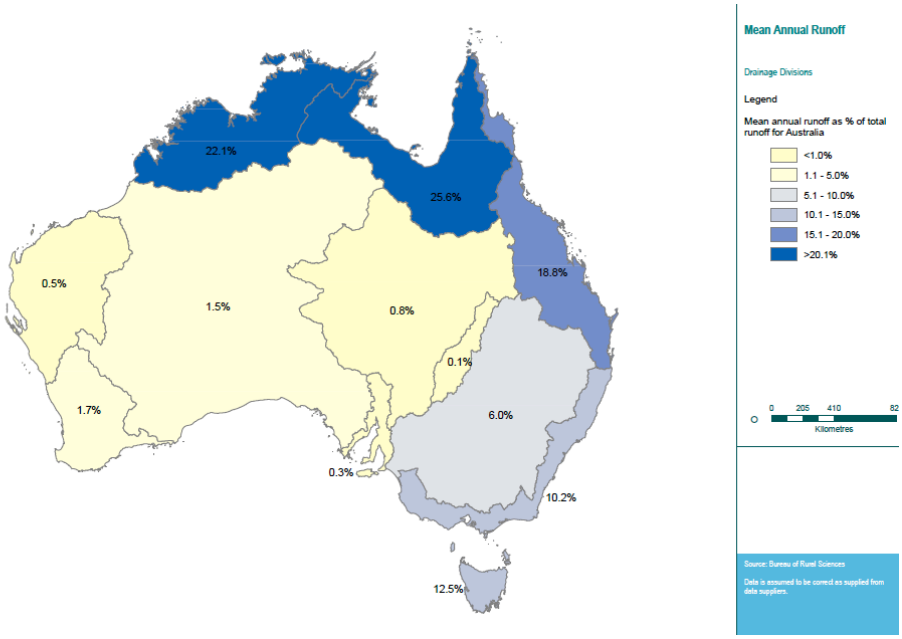

Figure 2.6. Mean annual runoff, as a percentage of totally Australian runoff, in Australian drainage divisions (National Water Commission 2005). See Chapter 3 for the definition of the regions.

The Millennium Drought (Leblanc et al. 2012) highlighted the importance of understanding hydroclimatic variability in a seasonal context. The 59% reduction in runoff during the period 1997-2008 was larger than expected from the 15% rainfall reduction and the long-term rainfall elasticity of runoff which suggested a 39% reduction in runoff (Potter and Chiew 2011). A majority (52%) of the runoff reduction is associated with reduced rainfall (Potter and Chiew 2011). Other sources have been identified, namely the seasonality of the rainfall decline and reduced interannual variability (Cai et al. 2009a, Potter and Chiew 2011, Verdon-Kidd and Kiem 2009a). While early research suggested a large portion of the runoff decline was associated with increased temperatures (Cai et al. 2009a) more recent work suggests only a relatively minor role (5%) of increased temperatures through changes in potential evaporation (Potter and Chiew 2011). Differences between the estimated role of temperature on declining runoff likely stem from the incorporation of seasonality or interannual changes in the temperature residual used by the earlier study (Cai et al. 2009a, Potter and Chiew 2011). Early research suggest that a 1°C rise in temperature is associated with a 15% reduction in annual streamflow (Cai and Cowan 2008b) but this figure was later updated as research suggested that the same temperature rise is associated with a 4% reduction in streamflow (CSIRO 2012). Differences in the estimations using various methods (statistical or dynamical modelling) highlight the complexity of rainfall – runoff relationships. The interactions between climate and hydrology will be discussed further in Section 2.2.2.

Autumn rainfall is essential to establish antecedent soil moisture conditions that are necessary to efficiently convert rainfall to runoff throughout the following seasons (Cai et al. 2009a, Kiem and Verdon-Kidd 2010, Pook et al. 2009). For this reason, the drought was particularly apparent in the parts of southeast Australia that experienced winter rainfall, with a focus on the early part of the season – May, June, July (Timbal and Jones 2008). The Millennium Drought was characterised by the lowest river inflows (up to 44% below average) on record in the MDB, despite the fact that lower-layer soil moisture and rainfall had not been the lowest recorded values.

Until the introduction of The Australian Water Availability Project (AWAP) data set (Raupach et al. 2008) hydrological research had traditionally focused on selective unimpaired gauged catchments (Kirono et al. 2010) as gridded information across the Australian continent was relatively scarce. Runoff and streamflow have been the focus of research due to data availability and the importance of water availability and river levels for management. Even so there have been many issues associated with streamflow data, for example, record length is typically reduced compared to climatic data sets and many streams have since been regulated or undergone major changes affecting flow characteristics (Chiew et al. 1998).

Data availability has left a significant research gap, to be addressed in this thesis, as to the nature of the relationship between other hydrological variables, such as soil moisture, and large-scale climate drivers. Soil moisture is one of the major constraints on agricultural productivity (Mueller et al. 2010) and it moderates other facets of hydrology such as runoff production, streamflow responses to rainfall and evapotranspiration (Perry and Niemann 2007). Accordingly, soil moisture patterns are important for flood forecasting, climate modelling and land management (Mueller et al. 2010). Soil moisture patterns are determined by multiple non-linear physical processes which adds to the complexity and makes estimation difficult (Perry and Niemann 2007). These processes may be locally controlled by spatial variations in soil properties or through vertical fluxes such as infiltration or by evapotranspiration, particularly during dry conditions (Perry and Niemann 2007). Different processes have strong relationships with soil moisture at different scales; evapotranspiration affects larger scales and topography dependant lateral redistribution at smaller scales (see references in Perry and Niemann, 2007). In the past, widespread soil moisture observations have not been widely available. This highlights the value in the AWAP data set and identifies a significant space in the literature that this thesis addresses.

This section has outlined the features that broadly control climatic and hydrological spatial and temporal variability across Australia. The importance of understanding interannual climate variability was highlighted as changes in rainfall are dominated by these shorter-term

fluctuations. Next, in Section 2.2.2, the important interactions between climate and hydrology will be outlined with a discussion of the large-scale modes that affect interannual variability in the Australian water balance to follow (Section 2.3).

2.2.2 Hydroclimatic interactions

The previous section (2.2.1) outlined the spatial and seasonal variability and trends in climate and hydrology and touched on the relationship between hydroclimatic variables. This section will further discuss the complex relationship between hydrology and climatic variables because, on the large-scale that is the focus of this research, hydroclimatic variability is dominated by rainfall variability (Lacey and Grayson 1998). The high degree of climatic variability that southeast Australia experiences is reflected in a highly variable hydrological cycle (Yu et al. 2010), with the sensitivity of streamflow to climatic variability globally recognised (Fu et al. 2004, Seneviratne et al. 2012).

The relationship between rainfall and runoff is influenced by many factors, including the seasonality and interannual variability of rainfall (Potter and Chiew 2011), temperature, losses to ground water, irrigation or dam extractions, changes in fire regimes, land-use or vegetation type, air pollution, sunshine hours or in the seasonality of rainfall (See Kiem and Verdon-Kidd, 2010, for references). Rainfall variability directly controls runoff generation, while temperature variability has a second order effect controlling runoff through changes in evaporation and transpiration (Yu et al. 2010).

The relationship between climate and hydrology is highly non-linear. Examples of this non-linearity include the catchment response to rainfall due to the effects of soil moisture (Murphy and Timbal 2008) and the magnification of a rainfall anomaly in the runoff response (Barros and Bowden 2008). For example, in the MDB, a rainfall deficit of 20% is associated with a 35% reduction in streamflow if there is no change in temperature, or a 45% reduction in streamflow if temperature increases by 0.4°C (Yu et al. 2010). The largest streamflow anomalies in the MDB are related to a specific temperature range, while changes in temperature outside this range had little impact on streamflow. The physical mechanisms for complex non-linear interactions are not clear, but it is suggested that temperature changes may be associated with rainfall changes that directly modify hydrological processes (Yu et al. 2010). In addition, while temperature has little influence on seasonal, de-trended upper level soil moisture the relationship with lower-layer soil moisture is significant, particularly during spring and summer (Cai et al. 2009a). This may be due to lagged effects and the more temporally consistent nature of temperature compared to rainfall and suggests that as temperatures rise under climate change, more water will be required to maintain the same soil moisture levels (Cai et al. 2009a).

However this research is based on partial correlation analysis from the residual of the linear regression of temperature on to rainfall. The complex non-linear interactions between climatic and hydrological processes suggest that linear statistics may not be adequate.

The relationship between rainfall and soil moisture also varies with depth. Over the period 1950 – 2008, the trend in rainfall and upper level soil moisture show a coherent seasonality with declines in autumn (Murphy and Timbal 2008). Lower level soil moisture on the other hand, displays little seasonality due to the inherent 'memory' of the lower-layer. Upper-layer soil moisture mirrors rainfall, as levels decline quickly due to evaporation to the atmosphere and drainage to the lower-layers without additional rainfall recharge. This is reflected in the correlation between the two variables that depletes at any lag. Lower-layer soil moisture levels don't show the seasonality of upper-layer soil moisture and rainfall. Lagged correlation between rainfall and lower level soil moisture persists for several seasons (Cai et al. 2009a)

Further complicating hydroclimatic interactions are the effects with other factors such as geology and vegetation. It has been shown that climate only has a significant effect on base flow (e.g. the slowly varying portion of groundwater that is the sum of groundwater runoff and delayed throughflow) for some geology – vegetation groups; i.e. when a Silurian or Devonian sedimentary and mixed wet/damp forest, which may be due to the saturation of certain soil types which may lead to higher base flow contributions, or to increased bedrock weathering and thus deeper soils (Lacey and Grayson 1998).

This section has outlined the complexities associated with the relationship between hydrological variables (streamflow, runoff, soil moisture and base flow) and climatic variability. The non-linearity of the relationship suggests that linear statistics may not adequately describe all aspects of the variability and that there is a role for novel statistical methods that can capture the variability, such as CART. The following section (2.3) will outline the relationships between large-scale mode of climate variability and Australian hydroclimate.

2.3 Influence of large-scale modes of climate variability on Australian hydroclimate

The previous sections outlined the spatial and seasonal variability in Australian hydroclimate (Section 2.2.1) and discussed the complexity of interactions between climate and hydrology (Section 2.2.2). A general understanding of the spatial and seasonal variability is necessary to give context to the drivers of interannual variability that will be discussed below. This section will discuss the indices used to characterise each phenomena, the impacts and mechanisms, and highlight research gaps, associated with the teleconnections to Australian hydroclimate from the large-scale modes of climate variability used throughout this thesis. It

should be noted that climate indices are a coarse and simplistic way to characterise complex and spatially extensive phenomena (Sharma et al. 2000) but they are widely used which allows comparison with other work.

Southeast Australia is under the influence of multiple climate drivers that have significant and spatially consistent relationships with hydroclimatic variability. However, it should be noted that these climate drivers are not independent of each other (Risbey et al. 2009) and the interactions between the modes are complex and non-linear (Pook et al. 2009). Understanding the relative and combined influences of non-independent, non-linear variables is challenging. In addition, the region is expected to be particularly sensitive to interactions between climate drivers (Kiem and Verdon-Kidd 2010, Risbey et al. 2009). The importance of a holistic approach when examining the drivers of southeast Australian climate variability is highlighted by the small fraction (less than 20%) of rainfall variability that large-scale drivers can explain individually (Risbey et al. 2009). In addition, each of these large-scale modes of climate variability has phases that are associated, individually, with increased or decreased rainfall probabilities. So, like rainfall, these large-scale conditions can be considered discrete events that must combine in particular ways to be conducive to rainfall. Increasing understanding of complex interactions between such variables is difficult within a classical statistical framework. This is a key location in the research space that the current research can add value by analysis with a novel statistical method.

While the drivers of Australian hydroclimate vary spatially and seasonally, rainfall anomalies can be broadly characterised by two major climate modes (Barros and Bowden 2008). The first, related to ENSO, is centred over the eastern half of the continent and associated with SST anomalies in the Pacific Ocean (Drosdowsky 1993). The second pattern is associated with the IOD and is related to northwest cloud-bands developing off the northwest coast of Australia (Nicholls 1989). Generally, most rainfall variability can be explained by the tropical modes of climate variability in spring (Risbey et al. 2009) although the STR also plays an important role in southern Australia in autumn and early winter (Timbal and Drosdowsky 2013).

The remainder of this section discusses each of the major large-scale climate drivers in more detail. Several key drivers of Australian climate (e.g. the Southern Annual Mode) are not discussed or used throughout this work due to their limited predictability and data availability. The omission of these drivers presents an avenue for future work. The tropical modes (ENSO and the IOD) will be discussed first in Sections 2.3.1and 2.3.2, followed by the STR in Section 2.3.3.

2.3.1 El Niño – Southern Oscillation

This section will outline ENSO mechanisms, indices and impacts on Australian hydroclimatic variability. ENSO effects ecosystems and economies globally, as the impacts of droughts and floods can be financially and environmentally devastating (Drosdowsky and Williams 1991, Horel and Wallace 1981, Latif et al. 1998, Nicholls et al. 1997, Ropelewski and Halpert 1987).

2.3.1.a Mechanisms

Exploration into the phenomena, beginning in the 1920s by Sir Gilbert Walker (Ropelewski and Halpert 1987), led to a preliminary understanding of the Southern Oscillation, or Walker Circulation. The zonal overturning of air across the equatorial Pacific Ocean is the atmospheric component of the coupled ocean-atmosphere system with critical feedbacks between the atmosphere and ocean (Vecchi et al. 2006). A fundamental mechanism driving ENSO is the positive feedback between the difference in equatorial SST and easterly equatorial wind stress, known as 'Bjerknes feedback' (Bjerknes 1969).

During La Niña or normal conditions (Figure 2.7), strong easterly trade winds tilt the thermocline and push warm surface water to the western Pacific around the maritime continent region. The temperature gradient caused by a shallow thermocline and warmer water in the western Pacific Ocean strengthens trade winds and maintains the large temperature gradient. The warm (cool) surface water in the western (eastern) Pacific drives a centre of convection (subsidence) that is associated with increased (decreased) rainfall. Conversely, during El Niño events, a weakening or reversal of the trade winds flattens the thermocline and shifts the warm water anomaly and the centre of convective activity to the eastern Pacific, with cooler SST anomalies, subsidence and decreased rainfall in the maritime continent region (Horel and Wallace 1981, Trenberth 1997, Wang and Fiedler 2006).

A sufficient general understanding of these ENSO mechanisms has been developed over recent decades. Yet the mechanisms associated with the specifics of ENSO events, such as their termination (McGregor et al. 2013, McGregor et al. 2012), is still an active area of research. Although there is a well-known climatology of the ENSO cycle, each event is different and significant asymmetries exist between El Niño and La Niña events that are not well understood.

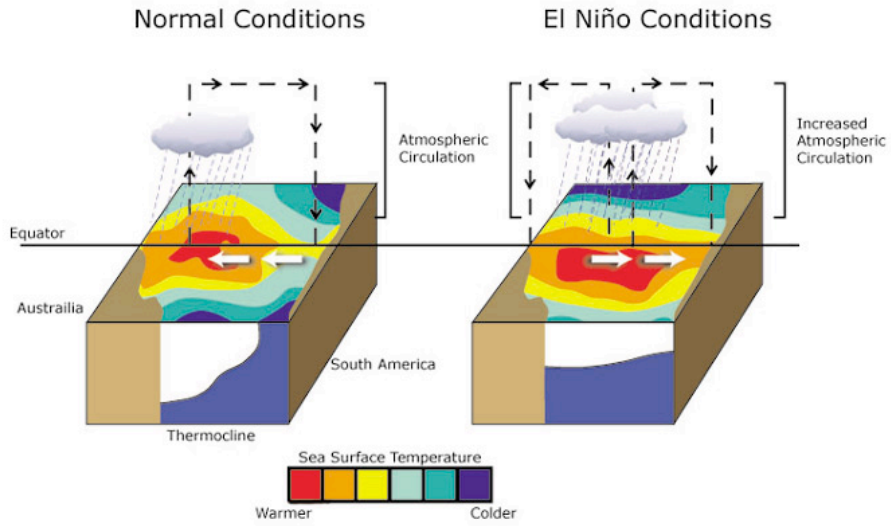

Figure 2.7. Schematic of the typical sea-surface temperature anomalies, atmospheric circulation and cloud cover during the cool (La Niña, left) and warm (El Niño, right) phases of El Niño – Southern Oscillation (Allan et al. 1996, Stevens 2012).

2.3.1.b Indices

This section will outline the main indices used to characterise ENSO variability with a detailed discussion of the specific indices used in this research in Chapter 3.

The Southern Oscillation Index (SOI) is a measure of the atmospheric component of ENSO and is calculated by taking the difference in pressure between Tahiti and Darwin. A positive SOI, or La Niña event, is characterised by lower surface pressure in Darwin. The strongest relationships are found between the SOI and rainfall when both series are averaged over several months (Chiew et al. 1998). The removal of high frequency variability from the series, associated with stochastic weather noise, leaves the ENSO signal.

The SOI has a stronger correlation with Australian rainfall than indices of the oceanic component because the latter are several steps physically removed from the processes that produce rainfall. Yet, the slowly-varying oceanic component is the source of predictability (Risbey et al. 2009). The most common oceanic indices used to capture ENSO variability are simple SST area-averages calculated over three regions in the central (Niño 4), eastern (Niño 3) and central-eastern (Niño 3.4) Pacific Ocean, that define an El Niño event when SST anomalies exceed a certain threshold (Ashok et al. 2007, Trenberth and Stepaniak 2001) . ENSO events have many different "flavours", and are defined by SST anomalies in different locations, including canonical El Niño events, 'dateline' El Niño events, and El Niño Modoki events (Ashok et al. 2007, Kao and Yu 2009, Trenberth and Stepaniak 2001).

The basic mechanisms of ENSO and the most common indices used to characterise the state of the Pacific Ocean ENSO have been explained. A discussion of the impacts of ENSO on Australian climate and hydrology are in the following sections.

2.3.1.c Impacts on climate

The 1982 El Niño event cemented the importance of ENSO in the public eye and focused investment and research, resulting in significant increases in understanding. It was the most severe the El Niño event on record, in terms of the SST anomalies and the associated rainfall deficits (Wang and Fiedler 2006).

ENSO is the leading mode of interannual variability globally and affects many regions through atmospheric teleconnections (Lu et al. 2008). Globally, 17 core ENSO regions were identified from monthly rainfall totals from 1700 stations, mostly located in the vicinity of the Pacific Ocean including the Australian sub-tropics (Ropelewski and Halpert 1987). In Australia, ENSO is the primary driver of Australia hydroclimatic variability, in terms of both spatial coverage and seasonal consistency (Risbey et al. 2009), with El Niño events generally associated with drought conditions (Arblaster et al. 2002, Chiew et al. 1998, Rasmusson and Carpenter 1983, Vecchi et al. 2006). In addition, stations under the influence of ENSO experience higher interannual rainfall variability (by up to 50%) compared to stations not influenced by ENSO (Nicholls 1988).

The spatial characteristics and seasonality of the relationship between ENSO and Australian rainfall are reasonably well established. El Niño (La Niña) events are associated with decreased (increased) rainfall over the north and east of the continent (Figure 2.8). In spring, the SOI explains approximately 35% of the variability in rainfall. This leaves a majority of rainfall variability associated with other drivers or stochastic variability (Drosdowsky and Williams 1991, Risbey et al. 2009). So although ENSO is known as the largest driver of Australian rainfall the low percentage of the rainfall variability explained by ENSO suggests that clearly other factors, such as the IOD and smaller scale weather-type phenomena, are also important. There is a robust linear relationship between the spatial extent of drought in tropical regions and the strength of ENSO events (Lyon 2004), but the severity of a drought is not linearly related to the strength of the ENSO event. Another example of the asymmetry of ENSO is how rainfall increases linearly during La Niña events, but declines non-linearly during El Niño events and the stronger correlation between rainfall and La Niña compared to El Niño (Power et al. 2006).

These non-linearities are inherent to the ENSO system and can limit our understanding and predictability of the system. A vast majority of the research into ENSO impacts is based in linear statistical framework, making use of techniques such as correlation and regression that are

less suited to the analysis of interactions between discrete, event-based phenomena. This is the niche in the literature where novel non-linear methods (e.g. CART) can play a vital role.

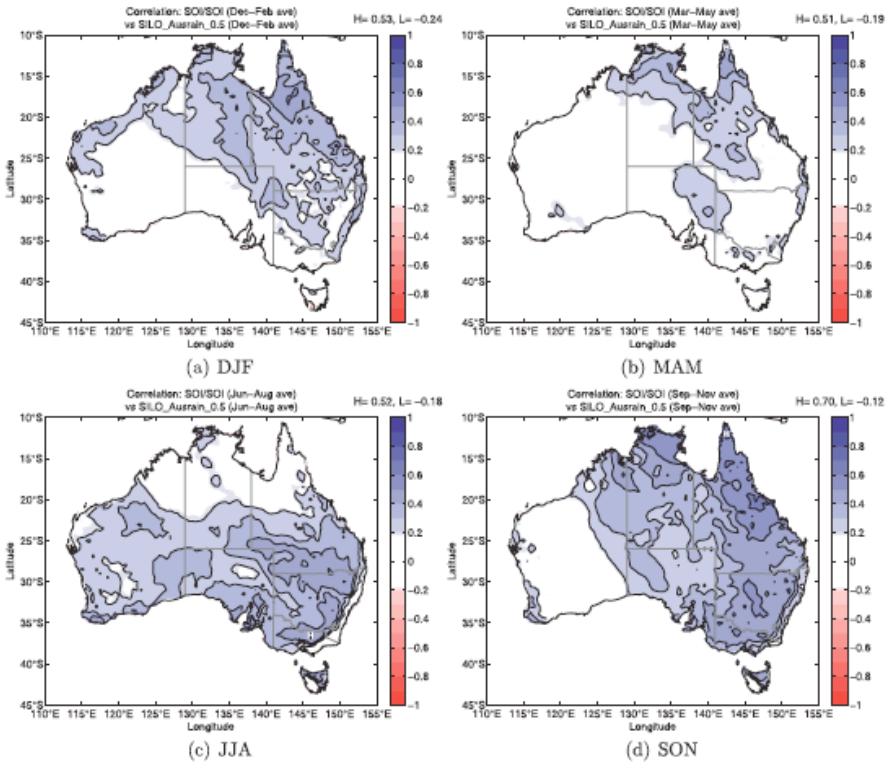

Figure 2.8. Seasonal correlations over the period 1889 – 2006 between Australian a) DJF, b) MAM, c) JJA and d) SON rainfall and the SOI. Correlations significant at the 5% significance level are coloured (Risbey et al. 2009).

The correlation between ENSO and Australian rainfall begins in winter and extends through to the summer months, peaking in spring (Chiew et al. 1998, Drosdowsky and Williams 1991, Pittock 1975, Rakich et al. 2008, Risbey et al. 2009, Shi et al. 2008, Simmonds and Hope 1997). As such, ENSO adheres to a strict seasonal cycle as rainfall deficits begin in the winter of an El Niño year carry on for seven to ten months (Figure 2.9) into the beginning of the following year (Chiew et al. 1998). The relationship between ENSO and Australian rainfall weakens in the late summer and autumn months when El Niño events decay (Drosdowsky and Chambers 2001) with little significant relationship evident in autumn (Murphy and Timbal 2008).

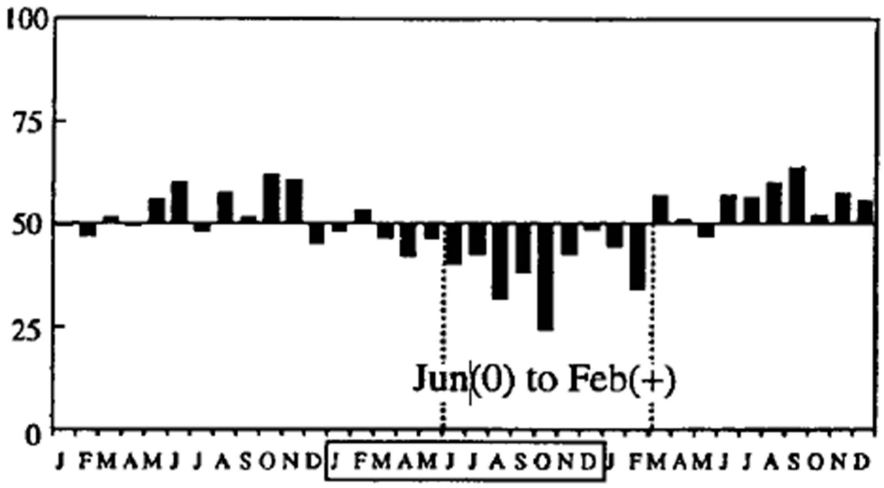

Figure 2.9. Composites of percentile log-normal rainfall distribution in southeast Australia during 14 El Niño events between 1910 - 1988 (Chiew et al. 1998).

Adding complexity is the decadal variability in the relationship between ENSO and Australian rainfall (Figure 2.10). For example, the relationship between the SOI and all-Australian rainfall was weak in the 1930-40s and has been weakening again since the 1970s. Although, the relationship in the winter and spring is more stable over time than that in autumn and summer (Suppiah 2004). The mechanisms for these changes in the warm season (DJF, MAM) are not well understood and may be related to changes in variability. This decadal variability is known as the Inter-decadal Pacific Oscillation (IPO). The IPO modulates ENSO activity and the relationship with hydroclimatic variability (Power et al. 1999, Speer et al. 2009). The IPO has a clear statistical relationship with Australian climate, although some have argued that it is a reflection of unpredictable stochastic ENSO variability related to the frequency of warm and cold events (Speer et al. 2009). The causes of this decadal variability remain not well understood.

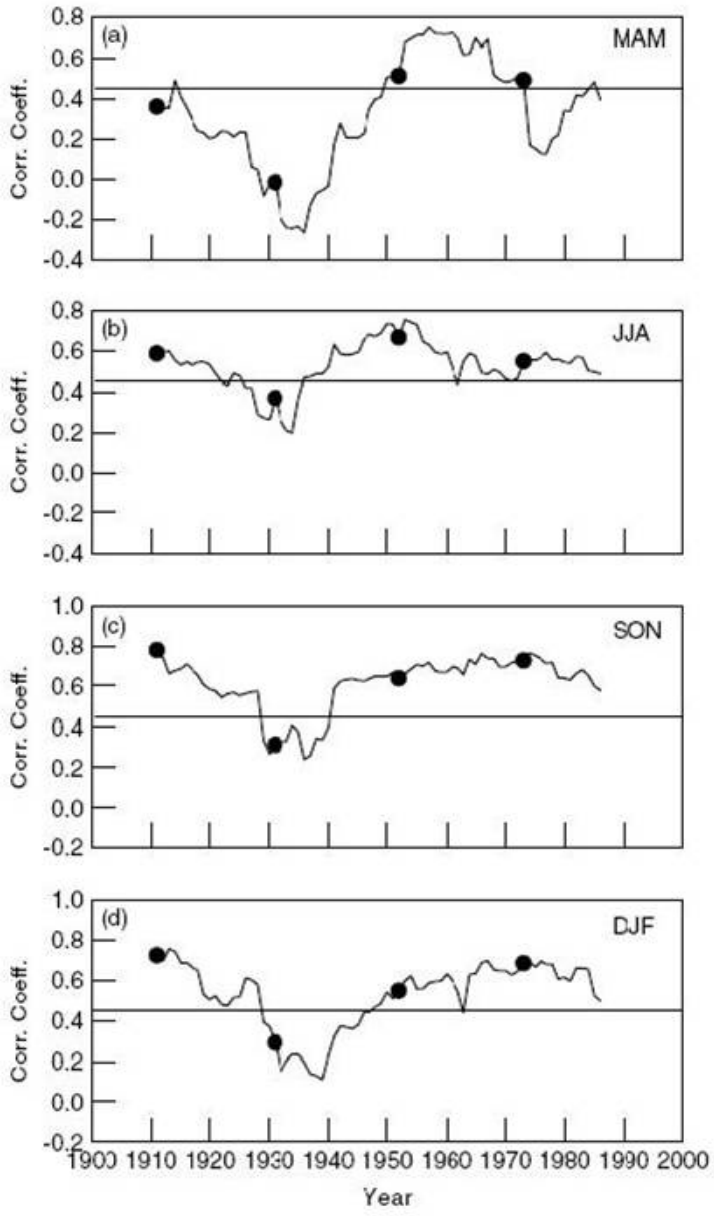

Figure 2.10 Variations in the correlation coefficient over the period 1900 – 1996, with a 21-year sliding window, between seasonal Australian rainfall and the SOI in a) MAM, b) JJA, c) SON and d) DJF. The horizontal lines indicate the 5% significance level, the dots represent independent correlation coefficients for 21-years (Suppiah 2004).

Not all ENSO events have the same impacts; the record strength El Niño in 1997 produced near average rainfall (which is attributed to a large rainfall event near the end of the season, highlighting the role of stochastic weather noise), while the 'weak' event in 2002 was associated with drought conditions in southeastern Australia. Although several competing theories attempt

to explain the differences between ENSO events including zonal changes (Wang and Hendon 2007), internal atmospheric variability (Kumar and Hoerling 1997) and decadal modulation (Power et al. 1999) there is little consensus on the issue. Variability in the strength of the ENSO teleconnection is key and the factors controlling this are not well understood. Interactions between ENSO and smaller scale weather-type phenomena are also not well understood.

A composite analysis of cool season rainfall (June – October) showed significantly lower rainfall in El Niño years; 10 out of 12 years experienced rainfall declines, of an average of 11mm month^{-1} in southeast Australia. There was a tendency for La Niña events to show a rainfall increase but the results were not significant (Ummenhofer et al. 2009a). This result is somewhat contrary to previous work that showing the strength of a La Niña event is linearly related to the magnitude of the rainfall increase (Power et al. 2006) and may stem from differences in the types of analysis, i.e. composite analysis compared to examining the whole ENSO cycle, and from comparing La Niña events with positive SOI which includes years that are not classified as La Niña events. The different conclusions from different analysis methods suggest that the results are sensitive to the somewhat arbitrary classification of ENSO events. CART provides an objective way to classify regimes, as will be discussed later.

Ultimately ENSO effects rainfall through changes to synoptic systems and moisture availability. Cut-off lows are a major source of rainfall in southern Australia and are influenced by tropical activity. A case study analysis of 1982 (a "dry canonical" El Niño event), 1997 (a "wet canonical" El Niño event) and 2002 (a "dry dateline" El Niño event) showed much less rainfall per cut-off low events during the dry events (1982, 2002), while average rainfall was achieved in 1997 from a small number of high rainfall events (see Brown et al. 2009 for definition of wet and dry canonical and dateline El Niño events). Back tracking air parcels found that the moisture source for these high rainfall cut-off lows was adjacent to northeastern Australia, while during the dry years air was sourced from the coast of southwestern Australia (Brown et al. 2009). Wet synoptic types associated with the northward movement of the STR and a strong pre-frontal trough are more common in autumn during La Niña years resulting in more reliable autumn rainfall (Verdon-Kidd and Kiem 2009b). These tropical – sub-tropical interactions show how ENSO can influence frontal systems and rainfall in southern Australia. It is important to understand when the effect of one mode depends on the level of a second mode because these interactions are vital to the forcing of southeast Australian hydroclimatic variability, although other research has failed to find a relationship between ENSO and synoptic-scale variability such as east coast lows (Pepler et al. 2013). The Brown et al. (2009) study showed the relationship between ENSO and smaller scale weather-type features that are immediately related to rainfall in southern Australia. While the conclusions are suggestive of an

important interaction between El Niño events and changes in the STR, the study was limited to only three cases. Further examination of this issue and relationships with other drivers such as the IOD, using a more robust statistical approach is required.

This section has outlined the relationship between ENSO and rainfall variability in Australia. Next in Section 2.3.1.d is a discussion of relationships between ENSO and Australian hydrology.

2.3.1.d Impacts on hydrology

The previous section (Section 2.3.1.c) outlined the influence of ENSO on Australian rainfall. This section follows on from the discussion of rainfall to explore the relationship between ENSO and hydrology. As expected from the strong relationship between ENSO and rainfall, ENSO also impacts the near surface hydrology in Australia (Power et al. 1999, Ralph and Hesse 2010).The relationship is well established as El Niño events are associated with decreased water availability. Streamflow responds directly to ENSO variability, consistent with the rainfall response to ENSO, with deficits beginning in June and continuing until the following February. However, the peak drought response to ENSO occurs at a two – three month lag as the rainfall deficit accumulates (Chiew et al. 1998).

River discharge, as measured at a stream gauge, is significantly higher during La Niña years (Ralph and Hesse 2010) and strong flooding has occurred in the MDB during La Niña events, particularly when the IPO is in the negative phase (Figure 2.11)

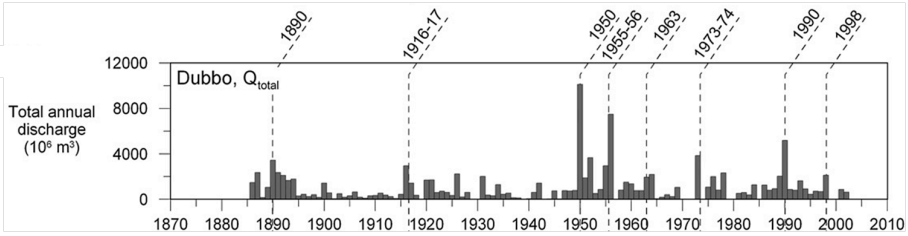

Figure 2.11. Total annual discharge from the Macquarie River at Dubbo with La Niña events marked (Ralph and Hesse 2010).

Upper level soil moisture levels can be assessed with passive microwave observations, which can in turn be used to estimate lower level soil moisture, evapotranspiration and runoff (Liu et al. 2009). The correlation between upper-layer soil moisture and the SOI (Figure 2.12, middle) reaches a maximum in winter and spring over eastern and northeastern Australia (Liu et al. 2009) and is similar to the rainfall relationship (outlined in previous sections and seen in the left column of Figure 2.12),.This shows that upper-layer soil moisture responds directly to changes in rainfall associated with ENSO variability. There is less understanding of how lower-

layer soil moisture relates to ENSO, although the relationship is less widespread (Ummenhofer et al. 2011).

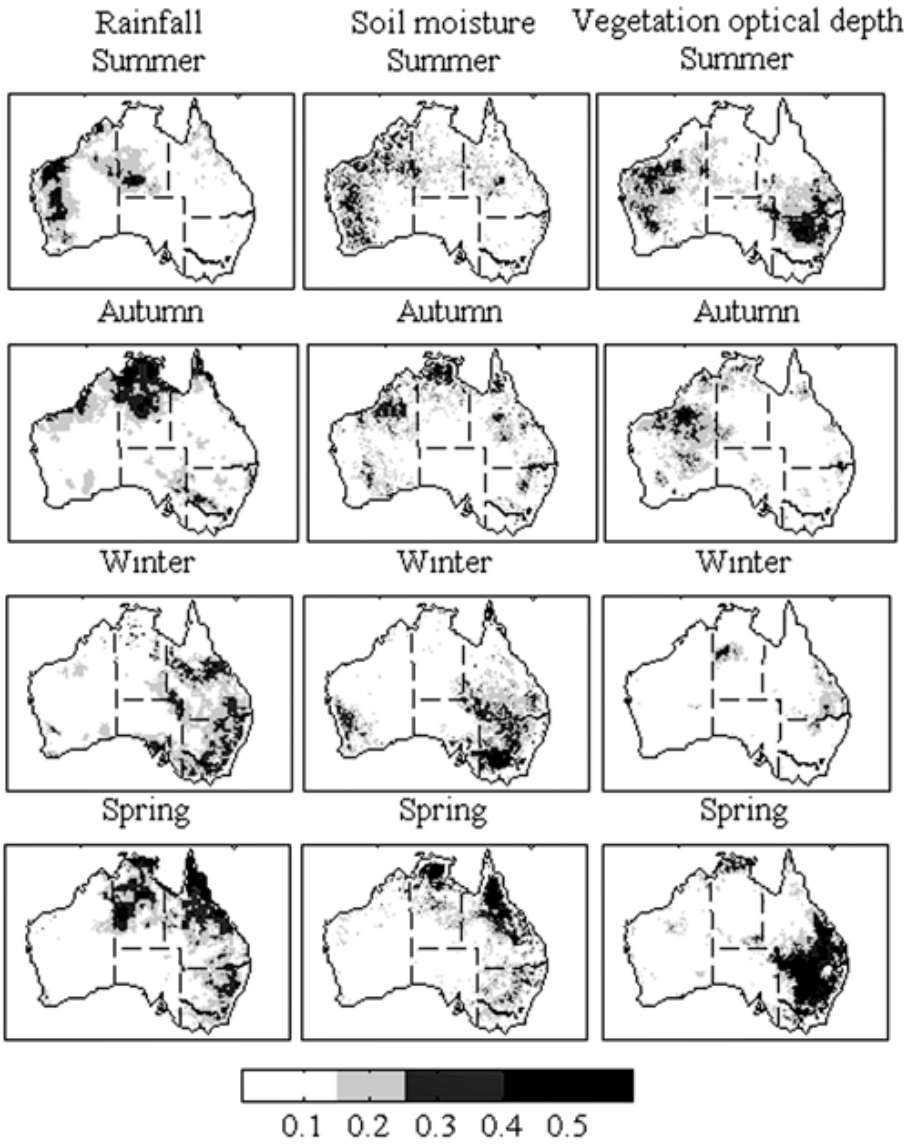

Figure 2.12. Seasonal correlation coefficients between the SOI and rainfall (left), soil moisture (middle) and vegetation optical depth (right), estimated from satellite data. Grey regions are significant at the 5% level and black regions are significant at the 1% level (Liu et al. 2009).

This section (Section 2.3.1.d) has outlined the impacts of ENSO Australian hydroclimate and the complexities associated with intrinsic non-linearities in both the ENSO system and climate-hydrology interactions. The difficulties these non-linearities pose when understanding ENSO impacts is emphasised. There are several key knowledge gaps that this research will fill. Novel statistical methods will likely increase understanding of a complex multi-dimensional system. In addition, while some research has examined the relationship between ENSO and

smaller weather-scale features, there is a great deal of understanding that can be added here. The importance of understanding interactions between large-scale modes has been touched on but will be expanded upon in later sections. The next section will discuss how variability in the Indian Ocean is related to Australian hydroclimate.

2.3.2 Indian Ocean variability

It has been shown how coupled ocean-atmosphere interactions in the Pacific Ocean result in teleconnections to Australian hydroclimate. This section will shift focus to how the Indian Ocean affects hydroclimatic variability in Australia. The mechanisms (Section 2.3.2.a) and indices used to characterise Indian Ocean variability (Section 2.3.2.b) are discussed first, followed by a summary of the impacts on Australian hydroclimate (Section 2.3.2.c) and the interactions with ENSO (Section 2.3.2.d).

2.3.2.a Mechanisms

The IOD is a coupled ocean-atmosphere mode in the Indian Ocean broadly similar to ENSO in the Pacific Ocean, but with smaller interannual variability (Baquero-Bernal et al. 2002, Saji et al. 1999, Verdon-Kidd and Kiem 2009a). There is a great deal of discussion into whether an mode of variability from the Indian Ocean indeed exists that is independent of ENSO and this debate will be discussed in Section 2.3.2.d; here the review is limited to processes associated with Indian Ocean variability.

A positive (negative) IOD event (Figure 2.13) is characterised by warmer (cooler) SST anomalies in the western Indian Ocean and cooler (warmer) SST anomalies in the eastern Indian Ocean, which is associated with reduced (increased) rainfall in southern Australia (Japan Agency for Marine-Earth Science and Technology 2008, Ummenhofer et al. 2009a, Verdon-Kidd and Kiem 2009a). A positive IOD event begins during autumn when as intensified trade winds lead to SST cooling in the eastern Indian Ocean, that combines with warm anomalies in the western Indian Ocean (Behera et al. 2006); large-scale atmospheric circulation anomalies accompany the SST anomalies through ocean-atmosphere coupling (Saji et al. 1999). A positive IOD event can be triggered by a shallow thermocline from remote forcing, strong easterly winds along the Java coast, or a combination of both factors that result in stronger upwelling and cooler SST in the eastern pole of the IOD. The increased SST gradient created by this process strengthens easterly winds, which in turn leads to increased SST cooling in the eastern Indian Ocean. An IOD event ends when monsoon transition stops the easterly wind or remote forcing increases thermocline depth (Meyers et al. 2007). During a positive event, cooler SST anomalies around the maritime continent are associated with a reduction in convection with

the centre of instability shifting to the western Indian Ocean with the positive SST anomalies. The IOD is phased locked to a seasonal cycle that typically peaks in spring but can be seen from May to November (Saji et al. 1999).

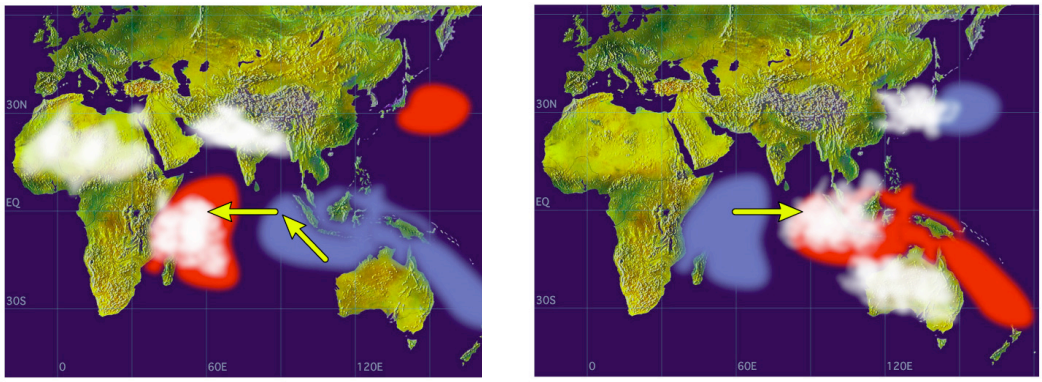

Figure 2.13. Typical SST anomalies and centres of convection for positive (left) and negative (right) phases of the Indian Ocean Dipole (Japan Agency for Marine-Earth Science and Technology 2008).

2.3.2.b Indices

This section will outline the main indices used to characterise Indian Ocean variability, although more detail on each index used in this thesis is given in Chapter 3. The Dipole Mode Index (DMI) is the principle index through which Indian Ocean Dipole (IOD) variability is characterised. The DMI is the difference in SST anomalies between the western and eastern Indian Ocean (Saji et al. 1999). In addition to the DMI, several other indices are used to characterise Indian Ocean variability (Cai and Cowan 2008a, Nicholls 1989, Timbal et al. 2010). It has been shown that the northwest shelf region has significant correlations with minimum temperature and rainfall in winter and spring (Timbal et al. 2010). SST in the Indian Ocean region northwest of Australia are often used instead of the IOD as research has shown this is the area important for the generation of northwest cloud bands which are the primary mechanism through which the Indian Ocean affects Australia (Verdon-Kidd and Kiem 2009a). Research has also pointed to a SST gradient between the maritime continent and the central Indian Ocean that has a significant correlation ($r = 0.75$) with the first EOF of winter Australian rainfall (Nicholls 1989). The Indonesian Throughflow (ITF) region (i.e. "Region C", area-averaged SST over 125° – 130°E and 2° – 7°S) has positive correlations with rainfall in southern and eastern Australia (Cai and Cowan 2008a). SST anomalies in the ITF are linked to ENSO variability and are one pathway through which the influence of ENSO is transmitted to the Indian Ocean (Cai and Cowan 2008a, Meyers 1996). Many more indices are used to characterise Indian Ocean variability compared to Pacific Ocean variability that stems from the

smaller variability in the Indian Ocean. The impacts of various indices of Indian Ocean variability on Australian hydroclimate over the instrumental record are discussed below.

2.3.2.c Impacts

Teleconnections from Indian Ocean variability have widespread impacts, particularly for southern Australia. The Indian Ocean has been implicated in extreme climate periods over the instrumental record, such as during WWII which was characterised by cooler SST around northwest Australia (Verdon-Kidd and Kiem 2009a).

Variability in the Indian Ocean is associated with rainfall in the south and west of the continent (Figure 2.14). The co-variability of ENSO and the IOD in time and space makes separating influences difficult but even with the effects of ENSO removed a broad region in southern Australia has significant correlations with the IOD (Risbey et al. 2009). Partial correlation analysis showed that the influence of the Indian Ocean can be of a similar magnitude to ENSO, although the limitations of partial correlation analysis, discussed below, must be kept in mind.

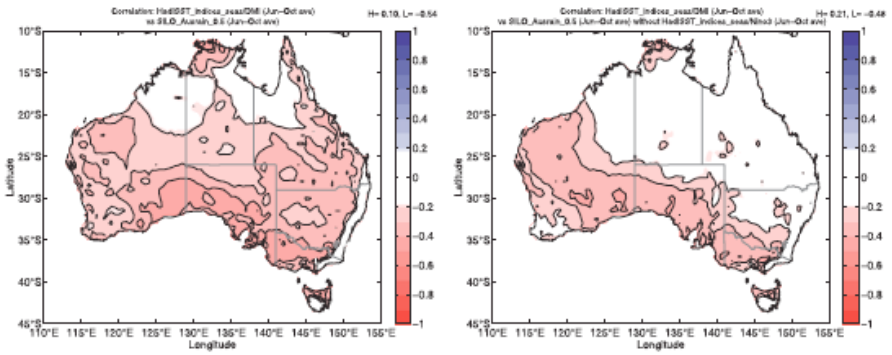

Figure 2.14. Correlation coefficients, over the period 1889 – 2006, between the DMI and Australian rainfall for June – October (left) and the partial correlation when the influence of ENSO has been removed (right). Only correlations significant at the 5% level are coloured (Risbey et al. 2009).

The SST anomalies in the northwest shelf region can alter the orientation of the sub-tropical jet that influences the development, structure and position of synoptic systems over the continent. The SSTs in this region may also play a role in directing moisture into synoptic systems over the continent (Brown et al. 2009, Risbey et al. 2009). Negative IOD events are associated with increased tropical-extratropical interactions that results in enhanced moisture advection to SEA (Ummenhofer et al. 2009a) and increased incidence of northwest cloud bands (Verdon-Kidd and Kiem 2009a).

This section has outlined the impacts of Indian Ocean variability on Australian rainfall, including its seasonality. Research has shown the seasonal relationship between Indian Ocean variability and Australian rainfall, but little has been done to explain interactions between the IOD and smaller scale phenomena. The next section will discuss the interactions with ENSO.

2.3.2.d Interactions

This previous section outlined the various indices of Indian Ocean variability and discussed the impacts that Indian Ocean variability has on Australian rainfall. This section will discuss the complex interactions between the Indian and Pacific Oceans.

The two ocean basins have both statistical and physical relationships between ocean variability, and their impacts on rainfall variability are often coincident (Meyers et al. 2007, Risbey et al. 2009). Two pathways link the ocean basins. Firstly, a system of oceanic currents from the Pacific Ocean to the Indian Ocean makes up the ITF, which is the primary mode of meridional heat transport between the two oceans. The volume transported through the ITF varies with ENSO, with more water transported during La Niña as the easterly trade winds in the Pacific Ocean lead to a build up of warm water in the western Pacific (Meyers 1996). The second mechanism of interaction is the "Tropical Atmospheric Bridge" where atmospheric anomalies in the maritime continent region that result from ENSO forcing then generate SST anomalies in the Indian Ocean through surface heat and radiative fluxes (Lau and Nath 2003). Both the oceanic and atmospheric pathways are likely vital for linking the basins with the relative contributions of each pathway not fully understood.

Research has questioned the existence of the IOD as an independent mode of variability, as the dependence (or independence) of the Indian to the Pacific Ocean has been the subject of a great deal of debate. The debate stems from the use of EOFs to defining the mode as an artificial example showed that a dipole pattern is commonly found in the second EOF due to the orthogonality constraint, even though such a dipole mode does not actually exist in the data set (Dommenget and Latif 2002). The first EOF of Indian Ocean SSTs accounts for 30% of the variance. The strong correlation between this mode and ENSO suggests that it is the Indian Ocean response to Pacific Ocean variability. The second mode, which is by construction uncorrelated with the first, accounts for 12% of the variability and is characterised by the dipole mode (Saji et al. 1999). The key for Australian rainfall variability is identifying which mode is related to rainfall, the first mode that is associated with ENSO or the second mode that is uncorrelated with ENSO.

In addition, it has been suggested that the close relationship between ENSO and the IOD can be attributed to a co-occurrence of a few strong events, as ENSO variance is amplified

during co-occurring events while the IOD variance is attributable to pure IOD events (Saji and Yamagata 2003). Further, the influence of the Indian and Pacific Oceans can be confused as they both peak at the same time and both have large effects on hydroclimatic variability. Partial correlation analysis suggests that both climate modes influence Australian climate, with the influence of the IOD focused on southern and western Australian rainfall (Figure 2.14) and the influence of ENSO focused over eastern and western Australia (Figure 2.15). The partial correlation analysis that removes the variability associated with the IOD (Figure 2.14) shows that the impacts of ENSO are associated with rainfall in northeastern Australia.

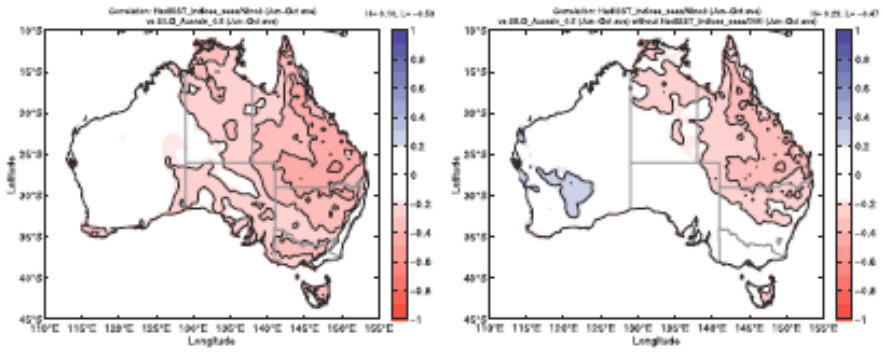

Figure 2.15. Correlation coefficients, over the period 1889 – 2006, between the Niño 3 and Australian rainfall for June – October (left) and the partial correlation when the influence of the DMI has been removed (right). Only correlations significant at the 5% level are coloured (Risbey et al. 2009)

Composite analysis for June – October rainfall shows that both drivers are important, while the combined effect of both indices (Figure 2.16) is more severe than either on its own (Risbey et al. 2009, Ummenhofer et al. 2011). Composite analysis of ENSO and IOD events showed that the largest rainfall anomalies in southeastern Australia were produced by combined events, showing the importance of interactions between the two Ocean basins. Combined positive IOD and warm ENSO events resulted in a rainfall anomaly of -20mm/month, whereas combined negative IOD and cool ENSO events produced a positive rainfall anomaly of +16 mm/month (Ummenhofer et al. 2009a). The rainfall anomalies when ENSO and the IOD are in the same phase are more widespread and spatially homogeneous. The spatial pattern of rainfall anomalies in the El Niño-only composite (Figure 2.16) extends much further south than the ENSO correlation in the partial correlation analysis (Figure 2.14). This discrepancy on the influence of ENSO between methods (composite, partial correlation) suggests further analysis is required.

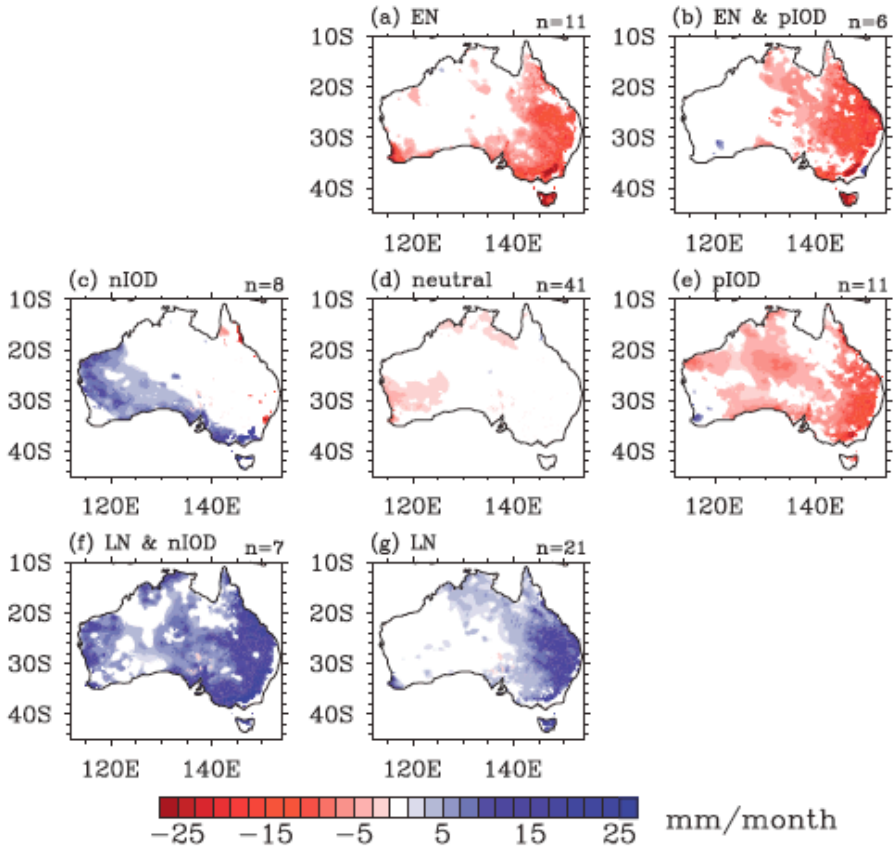

Figure 2.16. Composite June – October rainfall anomalies (mm) in combinations of ENSO and IOD phases for years during the period 1900 – 2006. Anomalies significant at the 20% level are coloured (Ummenhofer et al. 2011).

The mechanism through which the tropical modes of variability influence Australian rainfall varies seasonally (Cai et al. 2011b). In JJA, the impacts of the IOD and ENSO are considered separately (Figure 2.17). El Niño influences northern Australia through the direct response of rainfall to subsiding air motions in the tropics, as part of the western pole of the Southern Oscillation. The impact of the IOD on southern Australia occurs through an equivalent barotropic Rossby wave train (the east Indian Ocean wave train) emanating from the eastern equatorial Indian Ocean, resulting in an equivalent barotropic ridge over southern Australia. In SON (Figure 2.17), the tropical modes of variability cannot be separated but the Pacific South American wave train from the Pacific Ocean and is combined with wave trains emanating from the eastern and western Indian Oceans with a common centre over southern Australia (Cai et al. 2011b).

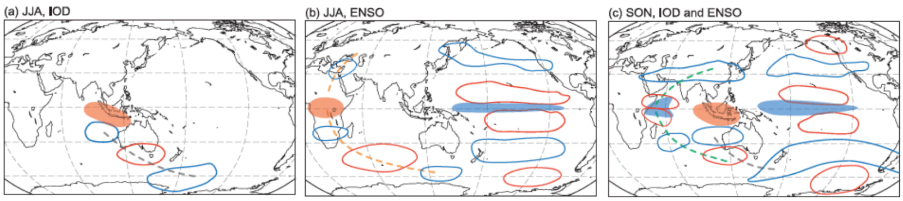

Figure 2.17. Schematic of the typical wave trains associated with a) the IOD and b) ENSO for JJA and c) for the IOD and ENSO combined in SON. These are the typical wave trains during the positive phase of each mode. Shaded blue (red) regions indicate increased (decreased) tropical convection. Blue (red) contours indicate anomalously low (high) upper-level heights. The dashed lines indicate the wave trains: grey is the East Indian Ocean wave train, green is the West Indian Ocean wave train and orange is the Equatorial African wave train (Cai et al. 2011b).

Recent research has taken a more holistic view of tropical influences on Australia hydroclimatic variability (Cai et al. 2011b, Meyers et al. 2007, Timbal and Hendon 2011). The importance of focussing on interactions over the whole Indo-Pacific basin, rather than one basin separately is emphasised as the combined effect of SST anomalies in both ocean basins is greater than either one alone (Figure 2.16). A tri-pole index (TPI) was constructed based on the equatorial SST regions that had the largest correlation with southeast Australian rainfall (Figure 2.18). Both ENSO and IOD variability contributes to the TPI and, due to its definition, correlations between the index and southeast Australian hydroclimatic variability exceeds both traditional indices (Timbal and Hendon 2011), essentially circumventing the IOD independence debate and focusing on variability in the tropical regions that are most relevant to SEA. Southeastern Australian rainfall and the TPI co-vary on interannual and decadal time scales, consistent with the strong relationship between the tropical oceans and Australian hydroclimates (Timbal and Hendon 2011).

Figure 2.18. Standardised 11-year running mean of spring southwest eastern Australian rainfall (green solid), northern MDB rainfall (green dashed) and the TPI (black). Standardised using the mean and standard deviation over the period 1900 – 2009 (Timbal and Hendon 2011).

This section has shown that the Indian Ocean has significant relationships with Australian hydroclimatic variability. The importance of a holistic approach to understanding the hydroclimatic response to tropical forcing was emphasised due to interactions between ENSO and the IOD. Next, in Section 2.3.3, the focus will shift to discussion of the sub-tropical ridge and its impacts on Australia.

2.3.3 Sub-tropical Ridge

Previous sections have shown that the tropical oceans have strong relationships with hydroclimatic variability in Australia. This section will discuss the importance of sub-tropical variability southern Australia. The sub-tropical ridge (STR) is the downward branch of the Hadley Cell and is located at approximately 30°N/S (Peixoto and Oort 1992). The position and intensity of the STR are two facets that influence Australian climate and will be discussed in Sections 2.3.3.a and 2.3.3.b.

2.3.3.a Sub-tropical Ridge Position

The position of the STR refers to the mean meridional location of the STR which is furthest north in August/September (29°S, Figure 2.19). The STR adheres to a seasonal cycle driven by variability in the location of maximum surface heating that controls the Hadley Cell. In summer when the location of maximum surface heating reaches its southerly maximum over northern Australia, the STR also moves south reaching a mean latitude of 40°S in February (Drosdowsky 2005, Timbal and Drosdowsky 2013). This movement of the STR controls

seasonal rainfall variability defining winter rainfall zones in southern Australia. When the STR is at its northerly maximum, a westerly airstream with embedded cold fronts is able to pass over southern Australia bringing rainfall to the region (Drosdowsky 2005). In summer when the STR is at higher latitudes, cold fronts are pushed south and miss continental Australia and there is increased rainfall down the east coast associated with enhanced on-shore, moisture laden winds (Pittock 1975).

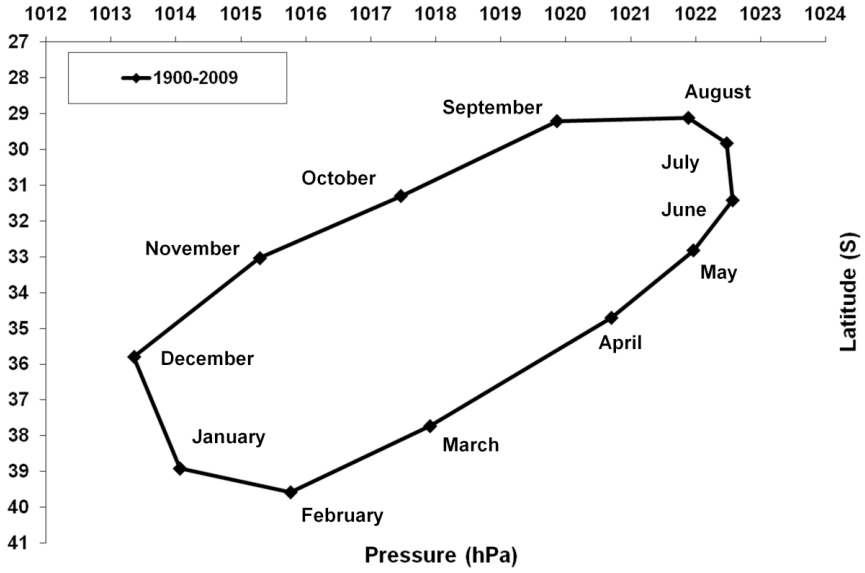

Figure 2.19. Monthly averaged STR intensity (hPa, horizontal axis) and position (degrees south, vertical axis) showing the annual cycle over the period 1900 – 2009 (Timbal and Drosdowsky 2013).

The position of the STR can be calculated from a variety of data sources including extrapolated station data and reanalysis products (Drosdowsky 2005, Williams and Stone 2009). All indices are highly correlated and the annual cycle is robust (Drosdowsky 2005, Timbal and Drosdowsky 2013) but the relationship with rainfall is somewhat sensitive to the choice of the dataset (Williams and Stone 2009), likely due to differences in the mean position of the STR between datasets. STR position has statistically and physically significant relationships with rainfall on a seasonal basis (Drosdowsky 2005, Williams and Stone 2009) as it is a proxy for monthly climatology of the mid-latitude and sub-tropical synoptic-scale features. This is indicated by the relationships between STR position and maximum temperatures, zonal westerly winds, meridional wind and mean air temperature and ozone; and on a seasonal basis with snow depth, zonal westerly winds and coastal rain (see Williams and Stone, 2009 and references within).

Trends in the STR position time series can be up to 2° over the instrumental record but they are seldom significant due to the highly variable record (Timbal and Drosdowsky 2013).

The most significant trends in STR position are a northward trend in summer and June and a southwards trend in May (Williams and Stone 2009), although there is little confidence in reanalysis products in the southern hemisphere before the introduction of satellites in the 1970s, and other research using station data has failed to find a trend in STR position (Drosdowsky 2005). The poleward shift of the latitude of STR in May fits in with the general expansion of the Hadley Cell (Seidel et al. 2007), and also with the rainfall decline in late autumn-early winter.

Spatial correlation patterns between the STR position and Australian rainfall suggest that up to 60% of rainfall variability can be accounted for by STR position variability, as the latitude of the STR effects the position of both westerly and easterly moisture sources (Williams and Stone 2009), but this work doesn't account for the correlation between STR position and intensity (Timbal and Drosdowsky 2013). From March-April-May (MAM) to October-November-December (OND) the correlation between gridded rainfall and STR position is negative across southeast Australia and southwest Western Australia (Figure 2.20) showing that as the STR moves further south rainfall decreases. The orographic effect of the Great Dividing Range (GDR) can be seen in the positive correlations present on the eastern coastal strip in all seasons. This shows that when STR position is further south rainfall on this coastal strip increases as the anti-cyclonic cell increases onshore flow (Timbal and Drosdowsky 2013). The correlation between MAM rainfall and STR position show how the mean position of the STR is important in the season when the STR is transitioning from its southerly to northerly maximum (Figure 2.20). The relationship is positive in central and northern Australia because as STR position moves south, rainfall increases due to increased meridional pressure gradient and tropical interactions. Conversely, the relationship is negative in Victoria because these systems are pushed further south when the STR moves poleward. Understanding the relative impacts of STR position and intensity is difficult with linear statistics as they are highly correlated. For example, while STR position alone is found to have a significant impact on rainfall, no additional impact can be found when it is combined with STR intensity. This suggests that the linear correlation between the two variables causes the significant contribution for STR position alone (Timbal and Drosdowsky 2013). This highlights the difficulty in assessing relationships between highly correlated predictor variables and rainfall.

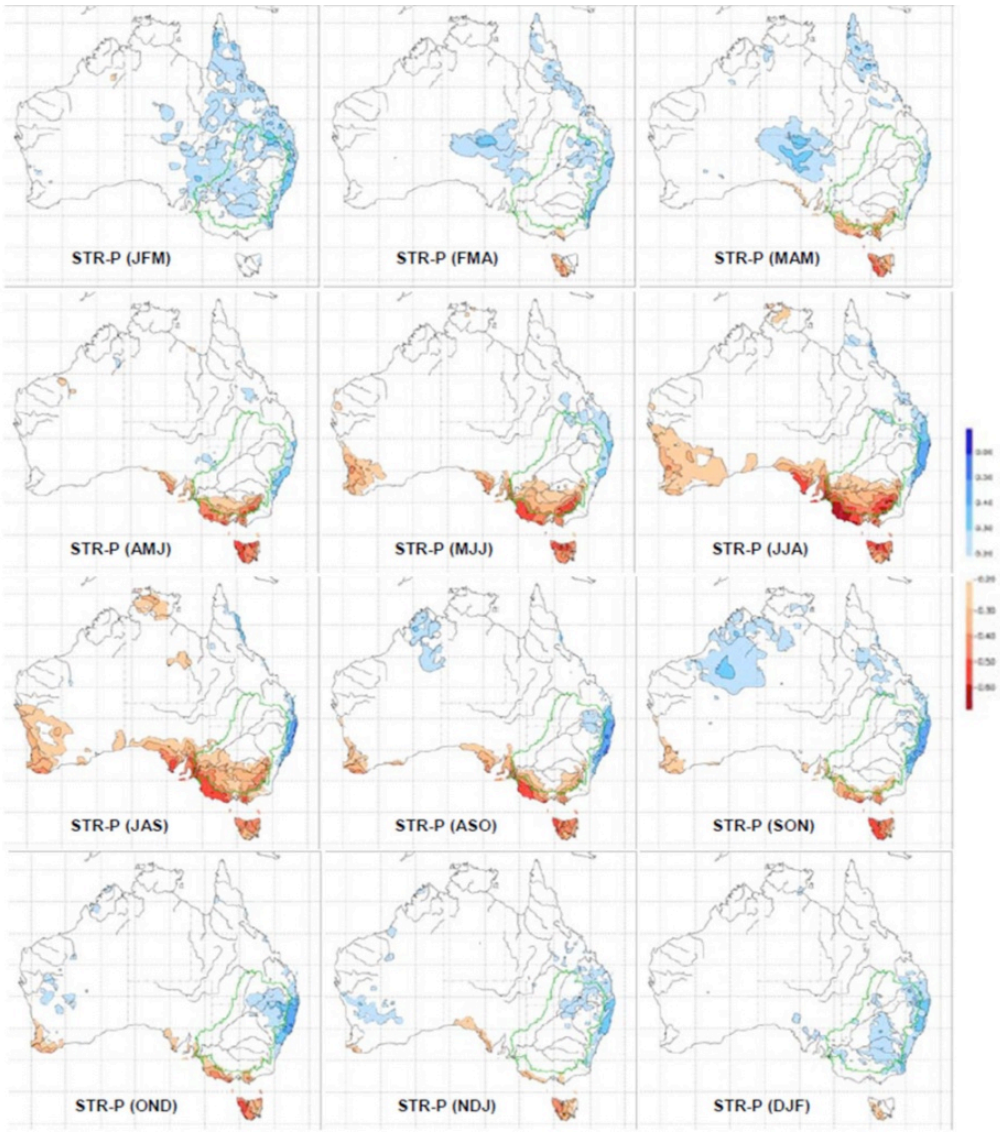

Figure 2.20. The annual cycle of the correlations between STR position and Australian rainfall in the 12 x three-month sliding seasons (e.g. JFM, FMA…DJF), over the period 1900 – 2009 (adapted from Timbal and Drosdowsky, 2013)

The typical synoptic system affecting southeast Australia have changed since 1993. Analysis using self-organising maps shows there are systems representing a stronger STR located further south, preventing cold fronts from passing over the region and a reduction in rain bearing pre-frontal troughs (Kiem and Verdon-Kidd 2010, Verdon-Kidd and Kiem 2009a). STR position is essential in the formation of cut-off lows; a weather system that bring a large proportion of rainfall to southeastern Australia. Preceding this type of wet case was a high-pressure cell off the east coast, directing moist air into the cut-off low. Drier cases occurred when the STR is too far east, west or south, showing the importance of the STR position (Brown et al. 2009).

This section has introduced the concept of STR position and discussed the relationship between STR position and rainfall. Broadly, the relationship between shifts in the Hadley Cell and rainfall is well understood. Yet the lack of relationship once STR intensity is accounted for confounds the issue and shows that significant correlations between predictor variables are not well accounted for using linear statistics. In addition, there is limited research into the relationship between the STR and large-scale climate drivers such as ENSO. The next section will discuss STR intensity.

2.3.3.b Sub-tropical Ridge Intensity

The majority of research has focused on how STR position influences climate in southern Australia. Recent research, however, has shown the importance of STR intensity for rainfall in southern Australia, linking rainfall declines to the weakening in the dominant westerly air stream associated with increased STR intensity (Timbal 2009, Timbal and Drosdowsky 2013). STR intensity adheres to a seasonal cycle, reaching a mean maximum of approximately 1022 hPa in May – August and a minimum of 1013 hPa in December (Figure 2.19). STR intensity is strongest in winter when the STR is located over the Australian continent and it is weakest in summer when the STR is positioned at higher latitudes.

The correlation between STR intensity and climatic variability is negative because high MSLP is associated with a stable subsiding atmosphere that results in lower rainfall and higher temperatures. Negative correlations with rainfall can be seen from April-May-June (AMJ), peaking in May-June-July (Figure 2.21), and decaying into OND (Timbal and Drosdowsky 2013). The relationship between STR intensity and rainfall is more widespread and significant compared to the relationship with STR position.

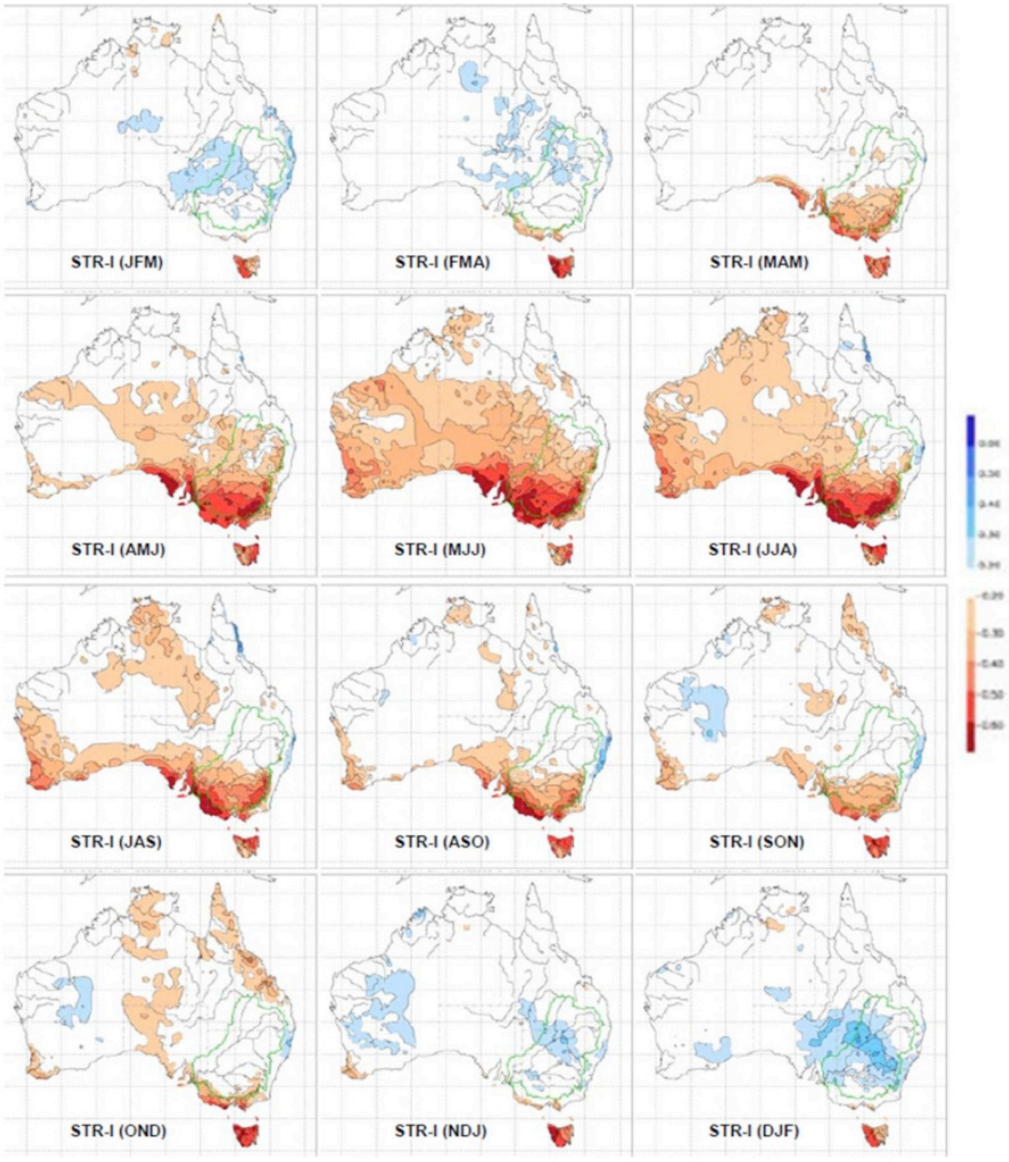

Figure 2.21. The annual cycle of the correlations between STR intensity and Australian rainfall in the 12 x three-month sliding seasons (e.g. JFM, FMA…DJF), over the period 1900 – 2009 (adapted from Timbal and Drosdowsky, 2013).

While trends in the STR position series were unclear, increases in STR intensity are significant (Drosdowsky 2005, Timbal and Drosdowsky 2013). Trends in all monthly and annual mean are positive, between 1 – 2.5 hPa for months when the relationship between STR intensity and southeast Australian rainfall is the strongest (Timbal and Drosdowsky 2013). Research on the relationship between STR variability and southern Australian rainfall has been limited to methods using linear statistics (Timbal and Drosdowsky 2013).

Section (2.3) has outlined the remote and region modes of large-scale climate variability, including the STR and the tropical modes (ENSO and the IOD). The seasonality of the climate drivers and the nature of relationships with Australian hydroclimate were outlined. Strong

relationships between climate drivers inhibit understanding of relative influences. Next is a discussion of the predictability of Australian hydroclimatic variability.

2.4 Predictability of Australian hydroclimatic variability

Section 2.3 summarised the relationship between Australian hydroclimate and large-scale modes of climate variability. This section will discuss sources of potential predictability, as seasonal to interannual prediction of hydroclimatic variability is vital for natural resource management. In particular, hydrological seasonal forecasts could help water managers make decisions related to water allocations while agricultural managers could better manage risk (Kirono et al. 2010, Wang et al. 2009). The main source of Australian seasonal predictability comes from the relationship between ENSO and rainfall variability and this is the basis seasonal forecast issued operationally by the Australian Bureau of Meteorology (Chiew et al. 1998, Doblas-Reyes et al. 2009, Kirono et al. 2010).

Correlations between seasonal Australian hydroclimate and indices of tropical modes of variability from preceding months ('lagged correlations') can be used to assess forecasting potential (Chiew et al. 1998). Lagged correlations stem primarily from the persistence of the large-scale tropical modes of variability from their onset in winter, this is evident in Figure 2.9 for the SOI and the effect is exaggerated in the SST-based indices. Maximum predictability occurs in spring when correlations between rainfall and indices of ENSO, particularly the oceanic indices, at several months lead-time exceed that of persistence (Figure 2.22). The correlation coefficients between station rainfall data and ENSO exceed 0.4 in several locations in eastern Australia up to three months in advance. The relationship between the SOI and rainfall is stronger than the SST index for the one and two month lag, but at the longer lead-time the SST index surpasses the SOI (Chiew et al. 1998). The predictability stems from ENSO's persistence, adherence to a seasonal cycle and from the slowly varying nature of the ocean. This simple relationship is the basis of the majority of seasonal climate forecasts in Australia (Drosdowsky and Chambers 2001).

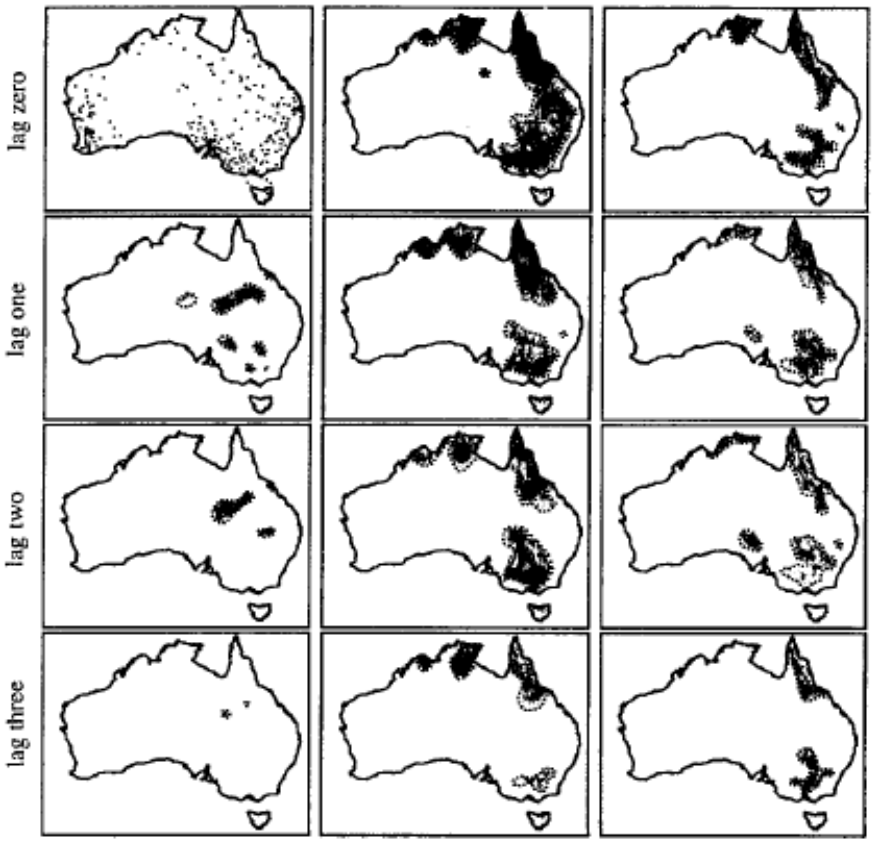

Figure 2.22. Correlation between spring station rainfall (stations shown in top left) and rainfall (left), the SOI (middle) and eastern Pacific SSTs (right) (Chiew et al. 1998).

The lagged relationship between seasonal rainfall and a wide range of predictor variables, averaged over the preceding two months, was examined (Kirono et al. 2010). The strongest relationship was found between spring rainfall with Niño 4 and the thermocline in July – August, although significant relationships were also found with the SOI, Niño 3, an Indonesian Index (120 - 130°E, 10°S – Equator), the western pole of the IOD and the first principal component of SST over both the Indian and Pacific Oceans. The same variables were important for summer rainfall with the exception of the western pole of the IOD and the Indonesian Index. Autumn and winter rainfall was only significantly related to the thermocline (Kirono et al. 2010).

Lagged correlations are also found between ENSO and hydrological variability. The predictor variables that have the strongest relationships with climate data tend to also be the variables that have strong relationships with hydrological variability (Kirono et al. 2010). Lagged correlations between lower level soil moisture and the SOI over the period 1955 – 2006 for various drainage divisions (often referred to as 'regions' in subsequent chapters) in eastern Australia (Figure 2.23) show stronger correlations for lower level soil moisture (r = -0.5) than for rainfall (r = -0.3 – -0.4). The peak correlation occurs when the SOI leads soil moisture by

approximately 3-months while the maximum rainfall correlation was simultaneous (Raupach et al. 2008). This shows that the relationship between ENSO and lower level soil moisture is enhanced compared to rainfall and that there is potential for greater predictability.

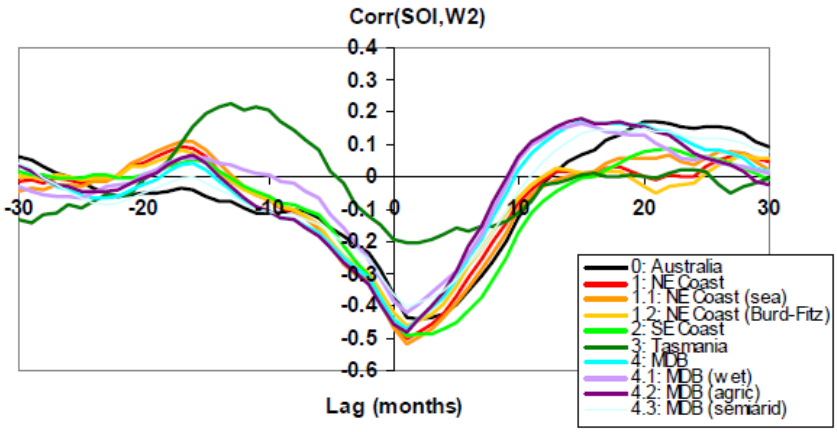

Figure 2.23. Lagged correlations between monthly SOI and lower level soil moisture (soil moisture > 0.2m) over drainage divisions in eastern Australia, over the period 1955 – 2006. See Chapter 3 for definitions of Australian drainage divisions (Raupach et al. 2008).

In southeastern Australia, spring streamflow is strongly correlated with the SOI, while during summer these strong correlations are confined to the east coast (3-month lag) and northeast coast (1-month lag) (Chiew et al. 1998). Streamflow and runoff shows greater persistence than rainfall, with the lagged correlation between streamflow and antecedent streamflow exceeding the correlation with ENSO (Figure 2.24). This results from a memory in the hydrological data due to hydrological processes such as groundwater storage. This shows that spring runoff in southeastern Australia can be forecast with up to 3-months in advance from both the SOI and from streamflow data in previous months (Chiew et al. 1998). The serial correlation between antecedent runoff is strongest with winter and spring runoff and is a better predictor of runoff during these max flow periods than the large-scale modes (Figure 2.24) while climate predictors have a strong relationships with runoff in spring and in summer. Summer is particularly interesting as a relationship is evident between the climate predictors and runoff that is not found with rainfall, showing that the ENSO – rainfall relationship is magnified in the hydrological variables in some seasons. This is likely due to non-linearities in the rainfall – runoff relationship and other hydrological processes such as infiltration, evaporation and storage (Kirono et al. 2010). Examination of longer lead-times suggests that the relationship with large-scale modes and antecedent runoff decreases as the lead-time increases, and that the runoff serial correlation weakens at a greater rate than the correlation with large-scale modes (Kirono et al. 2010).

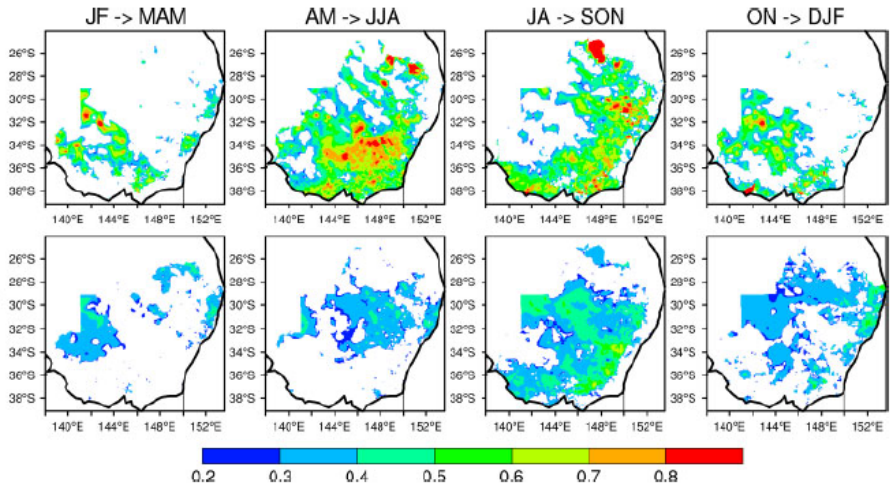

Figure 2.24. The correlation between seasonal runoff (MAM, JJA, SON, DJF) and antecedent runoff from the previous two months (JF, AM, JA, ON) (top) and the best climate predictor (bottom) using absolute values of the correlation coefficient, over the period 1950 – 2006 (Kirono et al. 2010).

Globally, operational meteorological centres make seasonal climate predictions by using statistical or coupled dynamical models (Doblas-Reyes et al. 2009, Drosdowsky and Chambers 2001, Goddard et al. 2001). Uncertainty from dynamical models comes from several sources including those related to initial conditions, model formulation and parameterisation ('tuning' to the current climate) and from the failure of climate models to simulate every aspect of climate dynamics (Doblas-Reyes et al. 2009). Statistical models require long records and a stationary climate for effective forecasting (Chiew et al. 1998). Seasonal forecasts using dynamical models are increasingly utilised, as statistical models will struggle to capture climate variability under future climate change scenarios. Yet statistical models are still a computationally inexpensive method to explore relationships and increase understanding of the predictability of hydroclimatic variability.

Statistical models, based on relationships between large-scale climate modes and hydroclimatic variability are used to make operational hydrological forecasts. A Bayesian joint probability modelling approach to forecasting seasonal streamflow at has been successfully implemented. The joint distribution of future stream flows and predictor variables including ENSO and antecedent rainfall anomalies are modelled. Cross validation results showed that forecasts with this method are of high quality and are free from bias (Wang et al. 2009) Spring streamflow, predicted from antecedent climate (i.e. SOI) and catchment conditions in August, for three gauges in southeastern Australia shows high skill with skill scores over 30%with scores greater than 10% defined as showing good skill (skill is defined with by a linear error in probability space score that has been adapted to probabilistic forecasts) (Wang et al. 2009).

This section has outlined the sources of potential predictability for Australian climate and hydrology and some techniques currently used for forecasting in Australia. Dynamical and statistical techniques are both useful and effective in forecasting.

2.5 Summary and conclusions

This chapter served as an introduction to the major features of Australian hydroclimatic variability and to the climate drivers that will be used in this thesis. It has outlined the current state of Australian hydroclimatic variability and discussed the interactions between climate and hydrology (Section 2.2). The main drivers of Australian climate were outlined and their impacts and interactions discussed (Section 2.3), along with sources of predictability of Australian climate (Section 2.4). Key gaps in the literature were noted throughout the chapter.

Next, in Chapter 3, is a discussion of the data sets and methods used throughout this thesis.

Chapter 3 : Data sets and Methods

3.1 Chapter summary and context

This thesis aims to increase understanding of interactions between large-scale modes of climate variability that influence Australian hydroclimatic regimes. This chapter provides an overview of all the data sets (Section 3.2) and a description of the tree-based models and other methods used throughout this thesis in Section 3.3. In subsequent chapters (Chapters 4 – 7), the specific data sets and methods used for that section of the analysis is briefly outlined.

3.2 Data sets

This thesis aims to increase understanding of interactions between large-scale modes of climate variability that influence Australian hydroclimatic regimes by using correlation and tree-based models. There are several data sets central to this analysis that span the period 1900 – 2009, including a suite of predictor and response variables used to define the relationships between climate drivers and hydroclimatic regimes. In addition, re-analysis products are used to characterise the atmospheric circulation in each of the hydroclimatic regimes.

First, in Section 3.2.1, is a description of the predictor variables used in this research, i.e. the indices of large-scale modes of climate variability (ENSO, IOD, STR). Secondly, in Section 3.2.2, is a description of the rainfall and soil moisture data that are used as response variables. Finally, Section 3.2.3 describes the two reanalysis products used in this thesis.

3.2.1 Predictor variables

The 14 indices of large-scale modes of climate variability used as predictor variables in this thesis are summarised in Table 3.1, and described in further detail below (Sections 3.2.1.a – 3.2.1.h). All 11 oceanic indices (Niño 3, Niño 3.4, Niño 4, the EMI, Region C, the Tri-pole index, Nicholls' SST gradient, the NWS and three indices of the IOD) are calculated by spatially aggregating SST anomalies in specific regions of the Pacific or Indian oceans from the monthly gridded (1° resolution) HadISST data set (Rayner et al. 2003). In addition to these nine oceanic indices, three atmospheric indices (SOI, STR-I and STR-P) are calculated from monthly mean-sea level pressure data sets provided by the Australian Bureau of Meteorology. All monthly indices are then averaged over 12 three-month seasons (e.g. JFM, FMA…DJF), with the focus generally on the traditional seasons (MAM, JJA, SON, DJF) throughout the analysis.

Anomalies are calculated against the full record (1900 – 2009). The linear trend is removed from predictor variables in Chapters 4, 5 and 7.

Table 3.1. Indices of large-scale modes of climate variability used in this thesis.

Name	Group	Reference
Niño 3	Pacific Ocean	(Trenberth 1997)
Niño 3.4	Pacific Ocean	(Trenberth 1997)
Niño 4	Pacific Ocean	(Trenberth 1997)
El Niño Modoki Index (EMI)	Pacific Ocean	(Ashok et al. 2007)
Southern Oscillation Index (SOI)	Pacific Ocean	(Troup 1965)
Tri-pole Index (TPI)	Pacific and Indian Ocean	(Timbal and Hendon 2011)
Region C	Indian Ocean	(Cai and Cowan 2008a)
Nicholls' SST gradient	Indian Ocean	(Nicholls 1989)
Northwest Shelf (NWS)	Indian Ocean	(Timbal and Murphy 2007)
Dipole Mode Index (DMI)	Indian Ocean	(Saji et al. 1999)
DMI – Eastern Pole	Indian Ocean	(Saji et al. 1999)
DMI – Western Pole	Indian Ocean	(Saji et al. 1999)
STR – Intensity	STR	(Drosdowsky 2005, Timbal and Drosdowsky 2013)
STR – Position	STR	(Drosdowsky 2005, Timbal and Drosdowsky 2013)

3.2.1.a Indices of ENSO (Niño 3, Niño 3.4, Niño 4, EMI)

Three oceanic indices of ENSO are defined by area-averaging monthly SST anomalies in regions of the central and eastern Pacific Ocean (Trenberth 1997, Trenberth and Stepaniak 2001) , all spanning latitudes 5°S – 5°N. The Niño 3 index is defined from SST anomalies in the region 150° – 90°W, Niño 4 is defined from SST anomalies in the region 160°E – 150°W and Niño 3.4 straddles the two regions (Figure 3.1).

The El Niño Modoki index represents a relatively new type of ENSO event that has warming centred in the central equatorial Pacific Ocean, rather than in the eastern Pacific during a traditional El Niño event. It is defined as the second empirical orthogonal function (EOF) of tropical Pacific SST anomalies which define 12 % of the variance (Ashok et al. 2007).

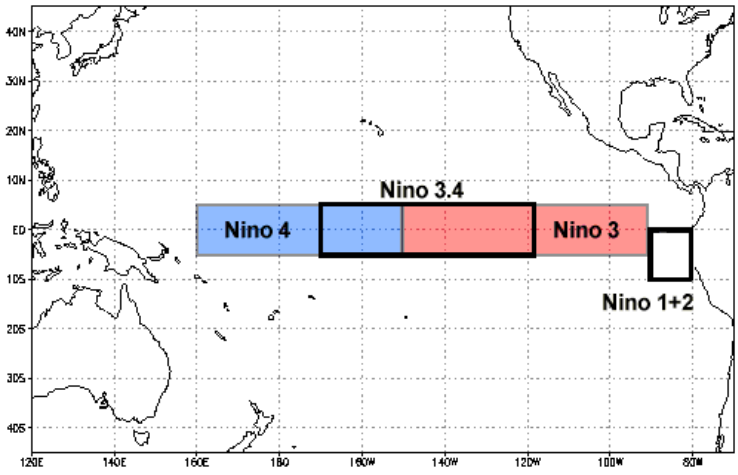

Figure 3.1. Area-averaged SST anomalies in the Nino 3, Nino 3.4 and Nino 4 regions that are used in this thesis to define the oceanic indices of ENSO. Picture from (National Climate Data Centre 2013).

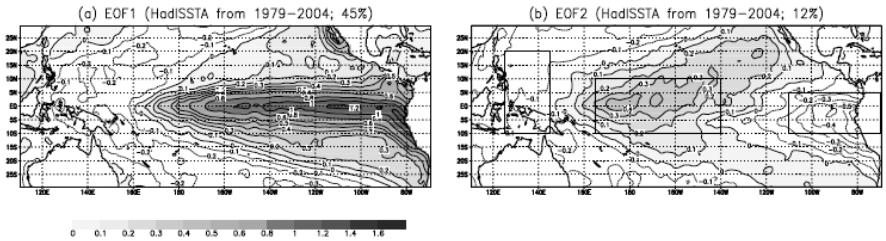

Figure 3.2. The top two EOF modes of tropical Pacific SST anomalies (1979 – 2004) multiplied by their respective standard deviations of principal components (°C), that explain a) 45% and b) 12% of the variance (Ashok et al. 2007).

3.2.1.b The 'Region C' index

'Region C' is a small area-averaged SST box in the Indonesian Throughflow region. It is the spatially aggregated SST anomalies in the region: 125° – 130°E and 2° – 7°S (Figure 3.3). SST anomalies in Indonesian Throughflow region are associated with ENSO and eastern Indian Ocean variability and has significant correlations Victorian rainfall in May (Cai and Cowan 2008a).

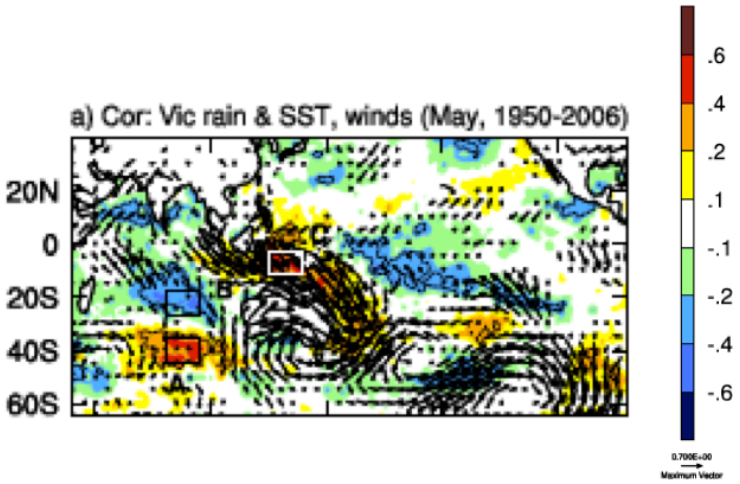

Figure 3.3. The correlation between Victorian rainfall and sea-surface temperature (coloured) and vector wind anomalies in May. Region C is the area-averaged SST anomalies in the white box (125° – 130°E and 2° – 7°S) (Cai and Cowan 2008a).

3.2.1.c The Tri-pole index

The Tripole Index (TPI, Figure 3.4) captures variability from the maritime continent and the Indian and Pacific Oceans by combining area-averaged SST anomalies in three regions. It captures variability from both ENSO and the IOD. The index subtracts the average of an eastern box from the Pacific Ocean (a blue trapezium in Figure 3.4: from 15°N – 15°S and from 150°E – 140°W at the northern edge and 180°E – 140°W at the southern edge) and a western box from the Indian Ocean (a blue square in Figure 3.4: 10°N – 20°S and 55° – 90°E) from the central box (a red parallelogram in Figure 3.4: extending from 0° – 20°S and from 90° – 140°E at the equator and 110° – 160°E at the southern edge) (Timbal and Hendon 2011) .

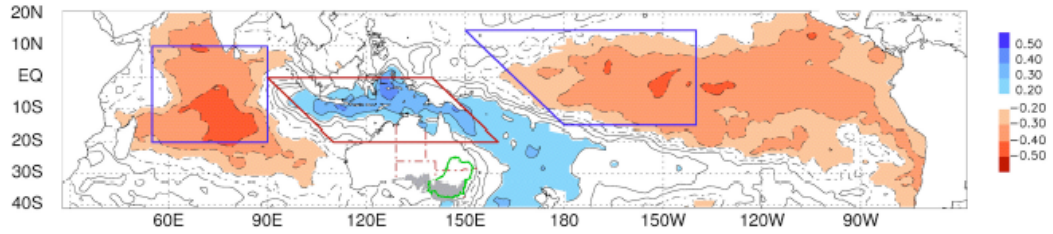

Figure 3.4. Map of the significant correlations between southeast Australian rainfall (grey shaded region) and SST anomalies over the period 1900 – 2009. The boxes used to define the Tripole Index (TPI) from Timbal and Hendon, 2011. The index is the difference between the mean of the SSTs in the central red box minus the average of the mean of the two blue boxes in the Indian and Pacific Oceans (Timbal and Hendon 2011)

3.2.1.d Indices of the Indian Ocean Dipole

The Dipole Mode Index (DMI, Figure 3.5) is the most commonly used index of Indian Ocean variability. It is defined as the difference between SST anomalies in the eastern (90° – 110°E and 10° – 0°S) and western (50° – 70°E and 15°S – 10°N) Indian Ocean (Saji et al. 1999). It has been suggested that variability in the eastern Indian Ocean is more important to the impacts on Australia than the western Indian Ocean (Meyers et al. 2007) and so each pole is used as a predictor variable individually, in addition to the DMI.

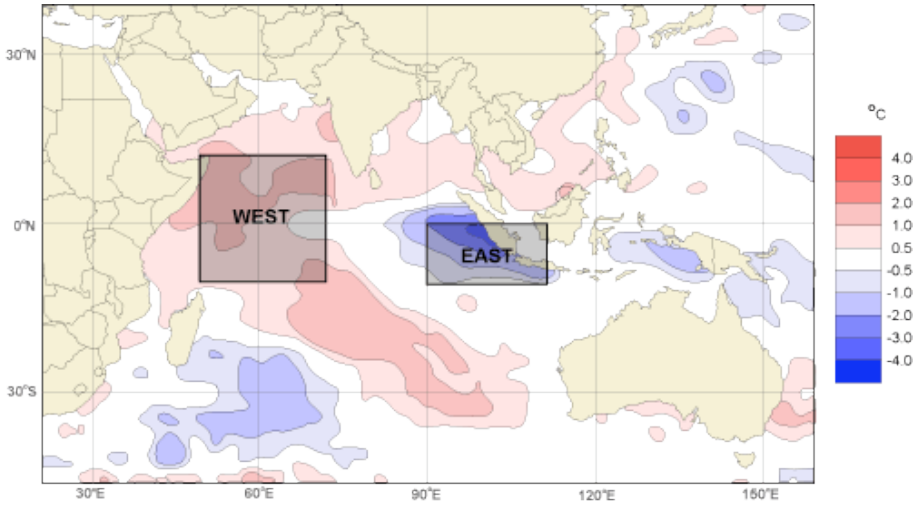

Figure 3.5. The SST anomalies in 1997 (a positive DMI year) with the western and eastern poles of the Indian Ocean Dipole marked (Australian Bureau of Meteorology 2013a).

3.2.1.e The Nicholls' SST gradient

The IOD is the most commonly used index of Indian Ocean variability. However, before the IOD was defined Nicholls (1989) described a gradient between two area-averaged SST regions 1) 120° – 130°E and 10°S – 0°, and 2) 80° – 90°E and 20° – 10°S that captured similar

variability to the IOD. The correlation between the gradient and Australian winter rainfall is not weakened by removing the effect of the SOI, suggesting it is independent of ENSO (Nicholls 1989).

3.2.1.f Northwest Shelf Index

The Northwest Shelf index is an area-averaged SST anomaly index over the region 100°–

130°E and 20°– 5°S. It has been used previously as an index of the eastern pole of the IOD (Timbal and Murphy 2007). The index has been shown to have significant correlations (10% level) with rainfall in western, central and eastern Australia in JJA (Timbal and Murphy 2007).

3.2.1.g The Southern Oscillation Index

The Southern Oscillation index is a measure of the atmospheric component of ENSO. It is the difference between the standardised anomalies of mean-sea level pressure at Darwin and Tahiti (Troup 1965).

3.2.1.h Indices of the sub-tropical ridge

Finally, two indices are used to characterise the sub-tropical ridge (STR). The STR intensity and position indices are calculated as the station-based 'L-index' (Drosdowsky 2005). The indices are based on monthly mean-sea level pressure station data between 10 and 44°S along the eastern Australian coast. The station data are interpolated to a 1° grid and averaged within a 5° longitude band around 150°E. The location of the maximum in pressure within that band gives the position of the STR, as well as the intensity of that maximum.

3.2.2 Response variables

Gridded monthly hydroclimatic data are spatially and temporally aggregated. Seasonal averages are calculated from the gridded monthly data for all 12-three month seasons (e.g. JFM, FMA…DJF) for the Australian continent, with the focus generally on the four traditional seasons (MAM, JJA, SON, DJF). Anomalies are calculated against the full record (1900 – 2009). Detrended data is used in Chapters 4, 5, and 7. Hydroclimatic data from 1900 – 2009 are analysed using the Australian Water Availability Project (AWAP) data set (Jones et al. 2009, Raupach et al. 2008) which was developed to monitor the terrestrial water balance of Australia. Hydrological research is often confined by availability of stream gauge data and climatic research is somewhat bound by the availability and quality of station data (although now several products exist that extrapolate the station data to a gridded product (Jones et al. 2009). The AWAP data set fills an important niche for hydrological data by providing a gridded dataset for

the Australian continent from 1900 to the present, at a relatively high resolution of 0.05° (approximately 5 km), allowing both climatological and hydrological information to be considered across the whole landscape.

The AWAP data include soil moisture and all water fluxes that contribute to soil moisture (i.e. rainfall, transpiration, soil evaporation, surface runoff and deep drainage); meteorological data (such as rainfall and temperature); both upper (0 – 0.2m) and lower-layer (0.2 – 1.5m) soil moisture (derived from information about soil properties used as inputs to a hydrological model); and transpiration, evaporation and local discharge (surface runoff and deep drainage) (Raupach et al. 2008, Timbal et al. 2009). The focus in this thesis is on rainfall and upper and lower-layer soil moisture.

While many analyses presented in this thesis are continental in scale, the focus for interpretation is generally on southern and eastern Australia, as these are important food production regions which are reliant on hydrological resources (Cai et al. 2009a). A further consideration is that data quality in western Australia is often limited (Figure 3.6) due to the sparse observational data network in the region (Ummenhofer et al. 2011). Analysis of the ability of the data set to capture the intensity of rainfall extremes is limited in the data sparse regions (King et al. 2013). This work is focused on terciles of hydroclimatic regimes, rather than on extremes, yet drainage divisions in the data sparse regions are not used (e.g. the Western Plateau (National Land and Water Resources Audit 2001)).

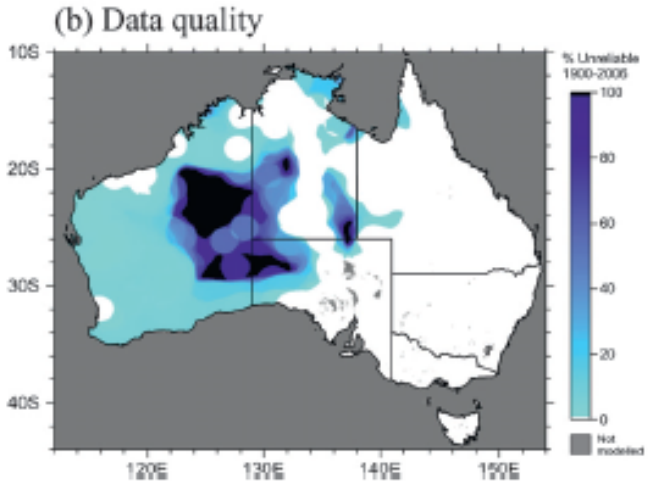

Figure 3.6. AWAP data quality map with black (white) areas being 100% unreliable (reliable with in the bounds of the interpolation scheme), blue areas should be treated with caution (Ummenhofer et al. 2011).

Time series analysis is conducted using gridded rainfall and upper-layer and lower-layer soil moisture data spatially aggregated for 12 regions (Table 3.2, Figure 3.7 and Figure 3.8) of Australian drainage divisions (National Land and Water Resources Audit 2001). However it

should be noted that drainage divisions do not always align with rainfall regimes. For example, the SEC region stretches from Victoria to the Queensland boarder and is likely under the influence of multiple mechanisms. These area-averaged time series were used in both continuous and categorical (converted to terciles) form during various phases of the analysis.

Table 3.2. Names, abbreviations for each of the 12 regions (drainage divisions) used as response variables.

Chapter	Drainage Division (Abbreviation)	Region Number
Chapters 5 and 7	Timor Sea (TiS)	1
	Gulf of Carpentaria (GoC)	2
	Northeast Coast – Sea (NECS)	3a
	Northeast Coast – Burdekin/Fitzroy (NECBF)	3b
	Murray-Darling Basin Agricultural (MDBA)	4a
	Murray-Darling Basin Semi-Arid (MDBSA)	4b
	Murray-Darling Basin Wet (MDBW)	4c
	Southeast Coast (SEC)	5
	Tasmania (Tas)	6
	South-Australian Gulf (SAG)	7
	Southwest Coast (SWC)	8
	Indian Ocean (IO)	9
Chapter 4	Murray-Darling Basin (MDB) – 4a + 4b + 4c	NA
Chapter 6	Southwest eastern Australia (SEA)	NA

A region defined as the southwest of eastern Australia (SEA) is used in the analysis of interactions between sub-tropical ridge intensity and position (Chapter 6). The time series for this region are calculated by averaging values within the region for which STR intensity explains over 20% of the rainfall variability and covers approximately the area to the south-west of a line running from Melbourne to the south of the Flinders ranges and following the end of the Great Dividing Range (GDR) over western Victoria (Figure 3.7, see Timbal et al., 2010 for detail about the construction of the SEA time series).

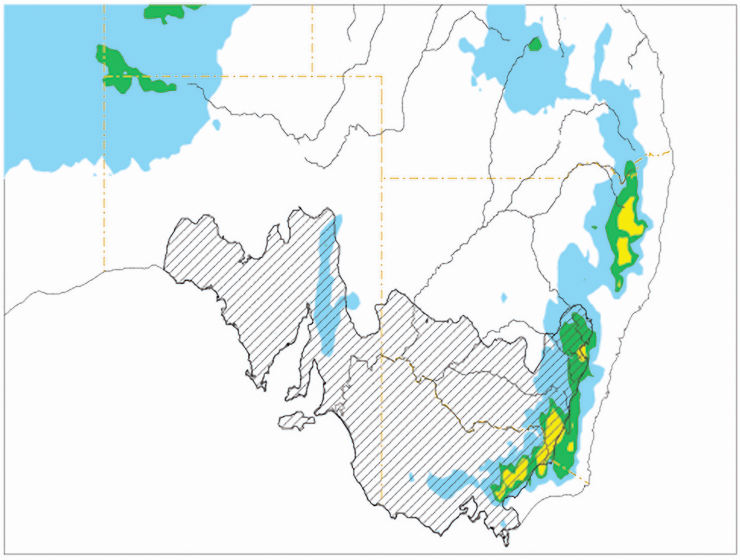

Figure 3.7. The hatched region is southwest eastern Australia. Elevations are shown by colour shading (every 500m). (Timbal and Hendon 2011).

The remaining 13 regions (Table 3.2 and Figure 3.8) used in Chapters 4, 5 and 7 (Region numbers 1 – 9) are based on Australian drainage divisions (often referred to as 'regions' in later chapters) (National Land and Water Resources Audit 2001).

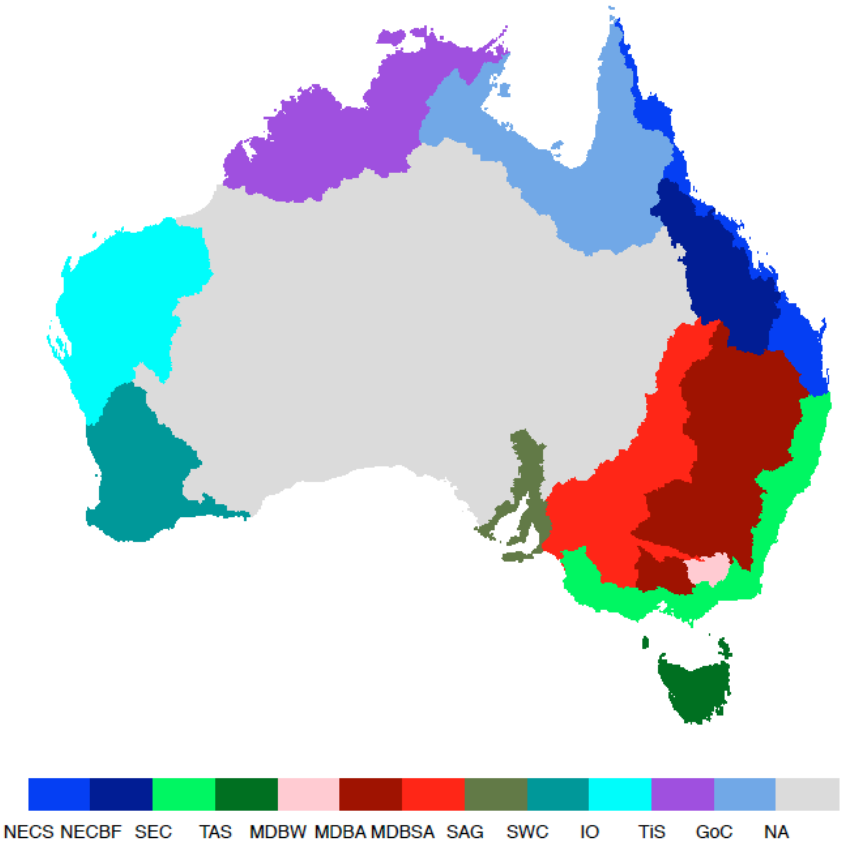

Figure 3.8. The 12 Australian drainage divisions that are area-average and used as response variables in Chapters 4, 5 and 7. The whole MDB region used in Chapter 4 is the sum of MDBW (pink), MDBA (maroon) and MDBSA (red). Table 3.2 for the names of the regions/drainage divisions.

3.2.3 Reanalysis products

Two re-analysis products are also used to assess the atmospheric response in hydroclimatic regimes. Firstly, The National Centers for Environmental Prediction / National Center for Atmospheric Research (NCEP/NCAR) Reanalysis product (Kalnay et al. 1996) was used for the analysis in Chapter 6. A limitation of the NCEP/NCAR reanalysis product has been is that it was only available from 1948; this limitation was overcome in 2011 when the 20th-Century Reanalysis Project (20CRv2) (Compo et al. 2011) was released, providing atmospheric information from 1871. The ensemble mean of the 20CRv2 was used in subsequent phases of the analysis (i.e. results presented in Chapter 5). This reanalysis product relies only on surface observations, that are limited at higher latitudes and in the early part of the record. However, atmospheric circulation in this product is well captured compared to NCEP2 and a gridded sea-level pressure data set (Fan and Liu 2013). The largest discrepancy from observed sea-level pressure is early in the record (< 1920). Although circulation features such as the SH subtropical

high, the Antarctic Oscillation and the Pacific-South-America pattern are well reproduced (Fan and Liu 2013).

3.3 Methods

Section 3.3 provides an overview of the key analytical methods used in Chapters 4 – 7, including an introduction to the CART and Random Forests techniques.

3.3.1 Linear analysis (Chapter 4)

Pearson's correlation coefficients between large-scale modes of climate variability and several hydroclimatic parameters are calculated from the seasonally averaged rainfall and soil moisture data for the Australian continent (0.05° grid). Pearson's correlations statistically significant at the 5% level are coloured.

3.3.2 Classification Tree analysis (Chapters 5 – 6)

This section will discuss some previous research using Classification and Regression Trees (CART) and provide a brief introduction to the method. CART is a statistical method that is widely used in the data mining and machine learning communities (Firth et al. 2005, Prasad et al. 2006). The method classifies response variable data from relationships with predictor variables. There are similarities with a self-organising map (SoM) approach, which is more widely used in climatic research (Alexander et al. 2010, Hope et al. 2006). Both methods create nodes based on similarities between cases. The SoM method uses the shortest Euclidian distance between cases to match each case to a particular node. The relationships between the resultant nodes and large-scale climate drivers can then be assessed with a correlation analysis. Conversely, a CART analysis uses the relationship between the large-scale climate drivers and response variables to classify nodes (often referred to as 'regimes' in this thesis). There are strengths and limitations to each method. For example, a SoM approach is likely to create more homogenous nodes but the relationship between these nodes and climate drivers is less obvious. In addition, while the CART approach ensures each node has a particular relationship with the climate drivers, the homogeneity of the nodes is not ensured.

Tree-based models may be constructed using either continuous (regression trees) or categorical (classification trees) response data. As such, CART is a set of statistical techniques comprising both 'classification trees' and 'regression trees'. There are no restrictions on the format of predictor variables, although in this research all predictor variables are continuous. Throughout this thesis, classification trees (constructed from tercile data) are primarily used, with a comparison of the two types of trees in Chapter 5.

CART is a non-parametric, binary, recursive tree-growing algorithm developed by (Breiman et al. 1984) that presents relationships between predictor and response variables in the form of 'decision trees', so that variability in a single response variable is explained by one or more predictor variables. These decision trees define a set of IF-THEN rules that divide cases of response variables into increasingly homogenous groups based on the corresponding levels of predictor variables (Burrows et al. 1995, Cannon et al. 2002, De'ath and Fabricius 2000, Rodionov and Assel 2000). A tree comprises of several 'leaves' or 'nodes'. A node is 'terminal' when there are no further splits. CART is robust to outliers and extreme values among predictor variables as the technique generally tries to allocate divergent cases to separate nodes (Rodionov et al. 2001).

As an example, a response variable (n = 30) containing three categories (red = A, green = B, blue = C) can be classified on the basis of two predictor variables (Predictor variables 1 and 2). In this example data set (Table 3.3), each value of the response variable (one of A, B or C) corresponds with a value for each predictor variable, so that Predictor variable 1 (2) is less than or greater than x (y) in each of the n-years of data.

Table 3.3. An example data set used in a tree-based analysis. The response (predictor) variables are categorical (continuous or categorical).

n	Response variable	Predictor variable 1	Predictor variable 2
1	A	< x	> y
2	B	> x	< y
…	…	…	…
30	A	< x	< y

In a tree-based model, all n-years of data start in the 'root' or 'parent' node (Node 1 in Figure 3.9). In the example, Node 1 contains ten cases of each of the three categories A, B and C (n = 30) of the response variable. CART searches for a way to divide, or 'split', the response variable into two mutually exclusive groups based on a single predictor variable, so that the purity of the two nodes is increased. In a classification tree, 'node purity' refers to the homogeneity of the node, i.e. if all cases in the node belong to one class (multiple classes) then node purity is high (low).

In the example, the first split divides the response variable so that all cases when 'Predictor variable 1 > x' ('Predictor variable 1 ≤ x') are allocated to the left (right) side of the tree. When the 'splitting criterion' (i.e. 'Predictor variable 1 > x') is true (not true) the corresponding cases of the response variable are always allocated to the left (right). All predictor variables used in this analysis are continuous variables, which means that the criteria for splitting can take any realistic value of the index.

By convention, the first root node is numbered 'Node 1' and the 'child nodes' resulting from the split are numbered so that the left-child node is double the parent node and the right-child node is double the parent node plus one. In the example, the root node is numbered "1", which is split to create "Terminal Node 2" (2 is double 1) on the left when the spitting criterion is true and "Node 3" (3 is double 1 plus 1) on the right when the splitting criterion is not true. Node 3 is then split creating "Terminal Node 6" (6 is double 3) and "Terminal Node 7" (7 is double 3 plus 1).

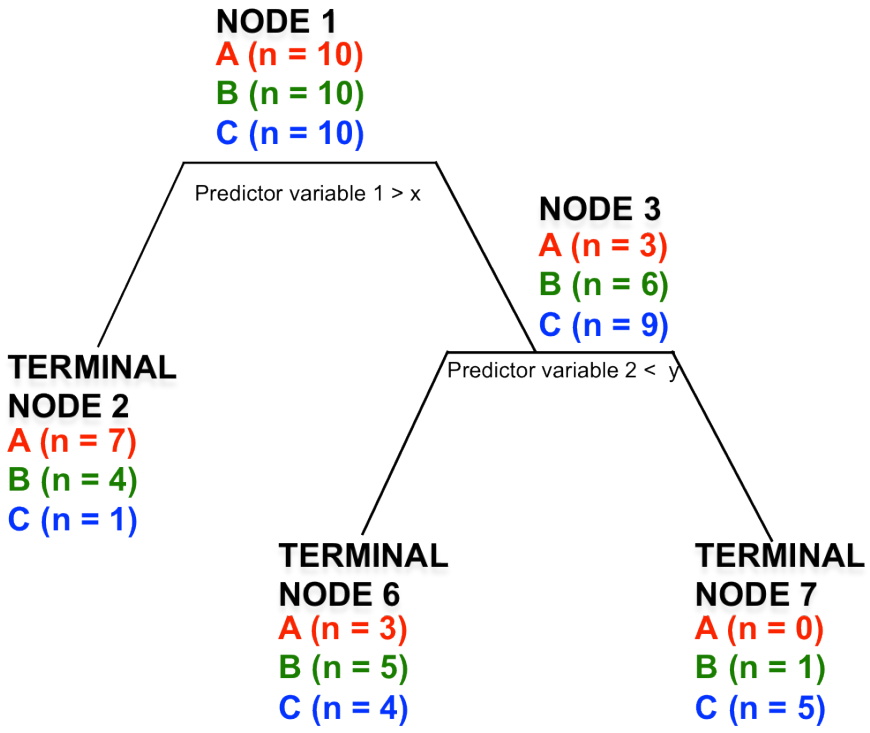

Figure 3.9. An example classification tree that classifies a categorical response variable (three classes each with ten cases, n = 30) from a suite of two predictor variables (Predictor variables 1 and 2).

The 'Gini index' is a measure of node impurity and is calculated based on the number of cases belonging to each class in the node. The index is calculated by subtracting the sum of the squared probability of each class' occurrence in the node (Equation 3.1) from one. This means that when a node contains only one class (i.e. a pure node) the Gini Index is zero, while when there are equal numbers of all classes (i.e. an impure node) the index is 1-1/n (where n = the number of classes). CART identifies the splits that minimise the mean node impurity, as measured by the Gini Index. Splits are chosen based on the Gini Index so that node impurity is decreased. In the example, Node 1 has equal numbers of all classes so the Gini index is at its maximum value of one as node impurity is at its maximum. In the example, Terminal Node 7

has zero cases from 'Group A', one case from 'Group B' and five cases from 'Group C', which results in a Gini index of 0.26 (Equation 3.2), showing that the Gini index is lower (closer to zero) when node purity is higher.

Equation 3.1. A classification tree used the Gini Index to calculate node purity, where j = class, p_j = probability of class j.

$$Gini\ Index = 1 - \sum_j p_j^2$$

Equation 3.2. Calculation of the Gini index of Terminal Node 7, in the example in Figure 3.9.

Gini Index of Terminal Node 7 = 1 – ((1/6)²+ (5/6)²) = 0.2778

Generally in a CART analysis an overly large tree is created so that all possible splits can be assessed and then 'pruned' back to an optimal size (Breiman et al. 1984). Node purity is higher in a larger tree with more terminal nodes, because, if left to run unchecked, the response data would be divided so that the Gini Index of each terminal node was zero, i.e. terminal nodes include only one class of data. This, however, would result in an overly large tree with limited practical significance and little interpretability. So there is a trade off between tree size and node purity. High node purity is easily achievable when tree size is large, but in this case the interpretability of the tree is often compromised. When tree interpretability is high (i.e. in a smaller tree with fewer splits), node purity is often conceded, as is the case when a tree is pruned back to a smaller, more manageable size. Consequently, terminal nodes in classification trees are rarely completely pure (containing only one class of data) and are more regularly comprised of multiple data classes. Cross-validation is used to assess error and prediction accuracy in a CART model. The data set is split into *k*-subsets (default $k = 10$), which are each left out in turn (i.e. the subset is left 'out-of-bag'), and the model is fitted to the remaining subsets (i.e. the 'in-bag' subset). The out-of-bag is then used to assess the predictive capability and error of the model with different tree sizes, so that the tree size with a minimum cross-validated prediction error can be selected (Maindonald and Braun 2010). The size of the tree (i.e. the number of splits) can be controlled directly (as in Chapter 5) or indirectly through the use of a 'complexity parameter' that puts a cost on each additional split (as in Chapters 6 and 7). The increased complexity of a larger tree is balanced against a reduction in node impurity (Maindonald and Braun 2010).

CART uses a 'majority rule' to allocate each terminal node to one of the categories of the response variable data, so that the terminal node is allocated to the class to which a majority of its members belong. In the example, Terminal Node 2 is allocated as 'Class A' as there are 7

cases in class A, 4 cases in class B and 1 case in class C; this node is 58% pure as 7 out of 12 cases belong to the majority class.

An additional feature of CART that makes it useful in climatological studies with incomplete data records is the use of 'surrogate splits'. Instead of discarding those cases where a value of a predictor variable is missing, CART instead classifies the response variable on the basis of a different predictor variable that closely predicts the child nodes classified with the original splitting criterion (Therneau et al. 2010).

The CART analysis in this thesis is conducted using the 'rpart' package (Therneau et al. 2010) in the 'R' statistical environment (R Development Core Team 2012); the package is based on (Breiman et al. 1984). The rpart package requires several arguments to be passed to the function that control aspects of the model fit (Therneau et al. 2010). The default values from the rpart package are outlined in Table 3.4; these are the values used in the following chapters, unless otherwise specified in the methods summary located in each subsequent chapter (Chapters 5 – 7).

Table 3.4. The arguments controlling the 'rpart' function (Therneau et al. 2010) used to create the tree-based models in the R Statistical Computing environment (R Development Core Team 2012).

Argument	Default	Definition
minsplit	20	The minimum number of observations in a node for a split to be attempted.
minbucket	7	The minimum number of observations in a terminal node for a split to be attempted.
cp	0.01	The complexity parameter: any split that does not decrease the overall lack of fit by 'cp' is not attempted.
maxcompete	4	The maximum number of competitor splits retained in the output.
maxsurrogate	5	The number of surrogate splits retained in the output.
usesurrogate	2	How to use surrogate splits; option '2' specifies that if all surrogates are missing then observations are sent in the majority direction (as per Breiman et al., 1984)
xval	10	The number of cross-validations.
surrogatestyle	0	Controls the selection of the best surrogate split; option '0' the total number of correct classifications is used for a potential surrogate variable.
maxdepth	30	Set the maximum depth of the tree (root node is depth 0).

Previously, many statistical methods have been used to understand the complexities of a non-linear and non-stationary climate. Simple linear statistics such as correlation and regression have given basic understanding of relationships between predictor and response variables, such

as ENSO and southeast Australian rainfall. The limitations of a linear statistical framework is clear in a region where multiple climate drivers affect hydroclimatic variability and that interactions between climate drivers are vital for the region. CART is set of statistical techniques that can capture interactions between multiple predictor variables and is well suited to the analysis of discrete and event-bases phenomena, such as hydroclimatic regimes and the large-scale conditions that make such regimes more or less likely. In addition, the method is well suited to event-based phenomena such as rainfall and the conditions that must be present for rainfall to occur. Decision trees are constructed that seek to describe the variability of one response variable by several predictor variables (De'ath and Fabricius 2000), they are particularly useful when you have multiple correlated predictor variables. Although not intended to replace other statistical techniques, CART can add understanding and provide insight into physical mechanisms underlying statistical relationships with easily interpretable results (Breiman et al. 1984, Rodionov et al. 2001). CART has been employed in many fields including ecology and genetics (De'ath and Fabricius 2000, Esther et al. 2010, Prasad et al. 2006, Thuiller et al. 2003) with some applications in climatology (Burrows et al. 1995, Cannon et al. 2002, Carter and Elsner 1997, Degirmendzic and Widbig 2007, Enke et al. 2005, Firth et al. 2005, Kennard et al. 2010, Lawler et al. 2006, Li and Sailor 2000, Rodionov et al. 2001, Rodionov et al. 2005, Wei and Watkins 2011).

The severity of winter regimes in the Great Lakes region in North America was classified from a variety of indices of large-scale modes of climate variability, showing how CART can extract additional information than a linear analysis. Classification trees were able to establish that the Polar/Eurasian index was the best predictor variables of winter severity despite a lower correlation than some other indices. It was suggested that the weak linear relationship was due to an asymmetry in the relationship between the index and winter severity (Rodionov and Assel 2000). In addition, regression trees have been used to skilfully predict streamflow in the Lower Colorado River from relationships with large-scale climate drivers (Wei and Watkins 2011)

This section (3.3.2) has outlined the major features and terminology of the CART models that are used for the analysis in Chapters 5 and 6. Further examples of many of these features, as used in this thesis, are discussed in Chapter 5. Next in Section 3.3.3 is a discussion of the ‘Random Forests’ method that is used in Chapter 7.

3.3.3 Random Forests (Chapter 7)

Random forests are a bagging (bootstrap aggregating) version of CART where multiple independent trees are constructed. The averaging of multiple classification trees has several benefits. Firstly, it can increase the prediction accuracy (Firth et al. 2005). Secondly, it can

reduce the instability and over-fitting associated with individual classification trees (Hastie et al. 2001) where small changes to the response variable can result in large changes to the final tree structure (Firth et al. 2005). While the over-fitting of individual trees is not an issue when the aim is to understand relationships between predictor and response variables, it becomes an issue when the aim is prediction.

A random forest consists of a collection of tree-structured classifiers (i.e. classification or regression trees) that are constructed from identically distributed subsets of the full data set (i.e. the 'in-bag' sample). Each tree in the random forest has a vote as to the final class of each individual case (Breiman 2001). The subset of the sample that is kept 'out-of-bag' during the construction of the random forest can be used to test the accuracy of the individual tree and aggregated across the forest, because this out-of-bag sample has not been presented to the individual tree (Firth et al. 2005).

The random forests analysis used in Chapter 7 of this thesis is implemented using the 'randomForest' package (Liaw and Wiener 2002) in the 'R' statistical environment (R Development Core Team 2012), based on the work of (Breiman 2001).

Although CART has had limited applications in Australian climatic research, random forests have been used to predict the timing of the onset of winter rain in Western Australia from a variety of predictor variables. SST anomalies were found to be the most valuable predictors with the onset correctly classified with 80% accuracy (Firth et al. 2005).

This chapter has outlined the data sets and methods used throughout this thesis to explore interactions between climate drivers that influence hydroclimatic regimes. The suite of predictor variables that were discussed in Chapter 2 have been defined and the regions used as response variables outlined. Each of the subsequent results chapters (Chapters 4 – 7) will summarise the specific data sets and methods used in that section of the analysis.

Chapter 4 : Linear Analysis

4.1 Chapter summary and context

This chapter will outline the linear relationships between hydroclimatic variables and the large-scale modes of climate variability that relevant to the Australian region. Assessment of the linear relationships using these indices and the novel AWAP data set is a vital first step in this research.

The focus in this chapter will be on response variables (rainfall, upper-layer and lower-layer soil moisture) and predictor variables (14 indices of regional and remote large-scale modes of climate variability) that will be used in subsequent chapters (Chapters 5 – 7).

The 'Introduction' (Section 4.2) will discuss previous research to show consistency with the current analysis. Methods and datasets will be described in Section 4.3, with the simultaneous and lagged correlations between response and predictor variables following in Section 4.4. The conclusions of this chapter will be outlined in Section 4.5.

4.2 Introduction

This chapter will outline the linear relationships between hydroclimatic variables and the large-scale modes of climate variability that are relevant to the Australian region. Assessment of the linear relationships using these indices and the novel AWAP data set is a vital first step in this research.

The meteorological products from AWAP (Donohue et al. 2010a, Donohue et al. 2010b, Evans and McCabe 2010, Fawcett 2010, van Dijk and Renzullo 2011) have been widely utilised in both operational and research contexts but the hydrological products have been used less frequently (Cai et al. 2009a, Cai et al. 2009b, Kala et al. 2011, Ummenhofer et al. 2011).

The chapter will focus on each climate driver separately, the sub-tropics in Section 4.4.1, the Pacific Ocean in Section 4.4.2 and the Indian Ocean in Section 4.4.3. Within these sub-sections, the simultaneous and lagged relationships between indices of each large-scale mode of climate variability and three hydroclimatic response variables (rainfall, upper-layer and lower-layer soil moisture) will be discussed.

4.3 Data sets and methods

A full discussion of the data sets and methods can be found in Chapter 3, this section will outline the specific data sets and methods used in this chapter.

4.3.1 Data sets

The linear relationship between the indices large-scale modes of climate variability and Australian hydroclimatic variability is assessed in this chapter. The indices of large-scale modes of climate variability that are used here are formally defined in Chapter 3, and shown below in Table 4.1.

Australian hydroclimatic variability is represented on the continental scale by gridded rainfall, upper-layer soil moisture and lower-layer soil moisture from the AWAP data set (Jones et al. 2009, Raupach et al. 2008). The time series of rainfall in 12 Australian drainage divisions (Figure 4.1 and Table 4.2) is also used throughout this chapter. Rainfall, upper-layer and lower-layer soil moisture in the Murray-Darling Basin (MDB: Regions 4a, 4b and 4c combined) is used to assess the lagged relationships.

Table 4.1. Indices of large-scale modes of climate variability used throughout Chapter 4.

Name	Group	Reference
Niño 3	Pacific Ocean	(Trenberth 1997)
Niño 3.4	Pacific Ocean	(Trenberth 1997)
Niño 4	Pacific Ocean	(Trenberth 1997)
El Niño Modoki Index (EMI)	Pacific Ocean	(Ashok et al. 2007)
Southern Oscillation Index (SOI)	Pacific Ocean	(Troup 1965)
Tri-pole Index (TPI)	Pacific and Indian Ocean	(Timbal and Hendon 2011)
Region C	Indian Ocean	(Cai and Cowan 2008a)
Nicholls' SST gradient	Indian Ocean	(Nicholls 1989)
Northwest Shelf (NWS)	Indian Ocean	(Timbal and Murphy 2007)
Dipole Mode Index (DMI)	Indian Ocean	(Saji et al. 1999)
DMI – Eastern Pole	Indian Ocean	(Saji et al. 1999)
DMI – Western Pole	Indian Ocean	(Saji et al. 1999)
STR – Intensity	STR	(Drosdowsky 2005, Timbal and Drosdowsky 2013)
STR – Position	STR	(Drosdowsky 2005, Timbal and Drosdowsky 2013)

Table 4.2. Names, abbreviations and region numbers for each of the drainage divisions used throughout Chapter 4. Drainage division abbreviations here correspond to the regions shown in Figure 4.1.

Drainage Division (Abbreviation)	Region Number
Timor Sea (TiS)	1
Gulf of Carpentaria (GoC)	2
Northeast Coast – Sea (NECS)	3a
Northeast Coast – Burdekin/Fitzroy (NECBF)	3b
Murray-Darling Basin Agricultural (MDBA)	4a
Murray-Darling Basin Semi-Arid (MDBSA)	4b
Murray-Darling Basin Wet (MDBW)	4c
Southeast Coast (SEC)	5
Tasmania (Tas)	6
South-Australian Gulf (SAG)	7
Southwest Coast (SWC)	8
Indian Ocean (IO)	9

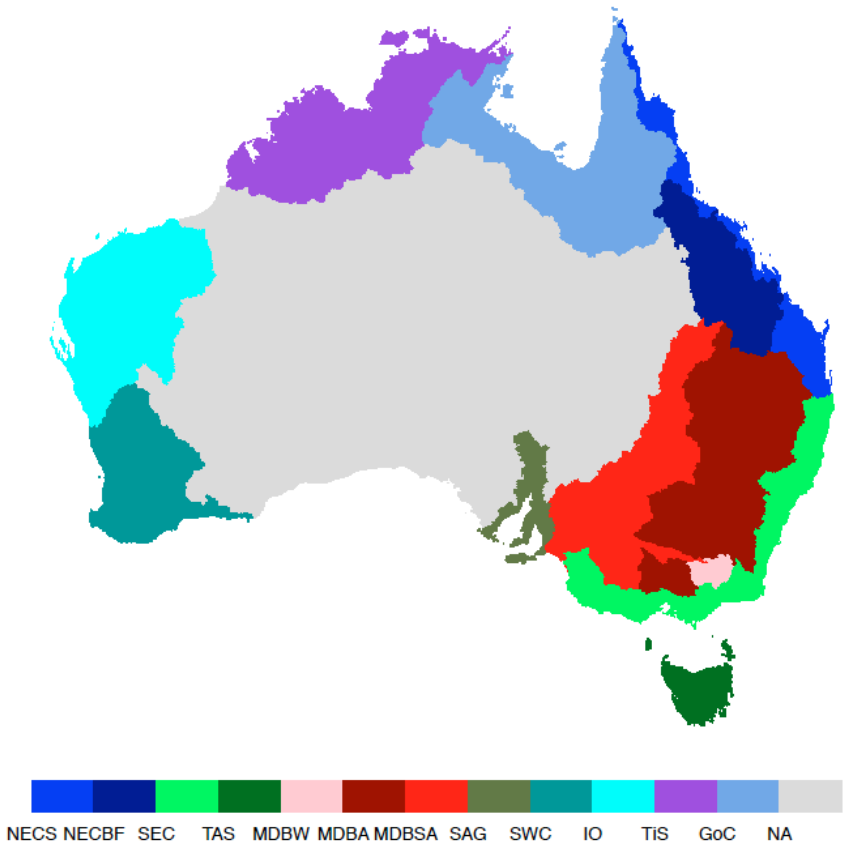

Figure 4.1. The 12 Australian drainage divisions that are area-average and used as response variables in this Chapter. The whole MDB region is the sum of MDBW (pink), MDBA (maroon) and MDBSA (red). Table 4.2 for the names of the regions/drainage divisions.

4.3.2 Methods

The simultaneous linear relationships between the indices of the three climate drivers (ENSO, IOD, STR) and Australian hydroclimatic variability (rainfall, soil moisture) in each of the traditional seasons (MAM, JJA, SON, DJF) are assessed using Pearson's correlation coefficients. Correlations significant at the 5% level are coloured in the continental scale figures and bold in the tables.

The spatial signature of the lagged relationship is assessed by correlating SON (JJA) hydroclimatic variability (rainfall, soil moisture) with indices of ENSO or the IOD (STR) from the preceding season. To assess how quickly the relationship decays, correlations are calculated between SON (JJA) hydroclimatic variability (rainfall, soil moisture) and indices of ENSO and the IOD (STR) from the preceding three-month averaged seasons with lag increasing by one-month.

4.4 Results

4.4.1 Sub-tropical ridge

This section evaluates the linear relationship between facets of the STR (i.e. STR intensity and position) and hydroclimatic variability in Australia. In each sub-section, the relationship between rainfall and the STR intensity and position is discussed first, followed by an analysis of the relationship between STR intensity and position and soil moisture (upper and lower).

4.4.1.a Spatial correlations

The intensity of the STR is negatively correlated with rainfall in southern and eastern Australia (Figure 4.2 and Table 4.3), showing that increased MSLP (a stronger STR) is associated with decreased rainfall. The significant negative correlations between STR intensity and regional MAM rainfall are found in TAS, SAG, SEC and the MDB regions. The strength and extent of the significant correlations between rainfall and STR intensity peaks in JJA, as the whole of southern Australia (TAS, SAG, SEC, MDB regions, IO, SWC) is covered in stronger correlations while a more moderate significant relationship stretches to northern Australia (TiS). The relationship in SON is similar to the other transition season (MAM); negative correlations are confined to southern Australia (TAS, SAG, SWC, MDBW, MDBSA). In SON, the mid-east coast displays a positive correlation on the spatial plots (Figure 4.2) that is not captured by the regional analysis (Table 4.3), this highlights how spatially aggregating rainfall anomalies can mask characteristics of the variability. In SON, increasing STR intensity is associated with increased rainfall as a more larger pressure gradient from the more intense high-pressure cell

increases the mean geostrophic flow and the strength of the moist onshore winds (Rakich et al. 2008). In DJF, as the monsoon trough moves south over the continent, the region of strong negative correlations are confined to TAS, while significant positive correlations cover southern continental Australia (SEC, MDBA, MDBSA, SAG). When the ridge is at its southerly maximum in DJF, a higher MSLP (stronger STR) is associated with increased rainfall, likely as a result of a large pressure gradient and increased tropical interactions over the region. Overall, this analysis is consistent with previous research that found the relationship between STR intensity and SEA rainfall peaked in late autumn / early winter (Timbal and Drosdowsky 2013).

The correlations between STR position and rainfall (Figure 4.2 and Table 4.3) are generally weaker than those with STR intensity; this is a common result across all hydroclimatic variables and agrees with previous research (Timbal and Drosdowsky 2013). In MAM the relationship between southern Australian rainfall (MDBW, TAS) and STR position is negative, suggesting that as the ridge moves further south rainfall decreases; while in northern Australia (NEC) the significant positive correlation suggests that a southerly position is associated with increased rainfall. The relationship peaks in JJA when SEA (MDBW, MDBSA, TAS, SAG, SWC) is covered by strong negative correlations with positive correlations along the eastern coast that are indicated by the non-significant positive correlation in the SEC region. In JJA the STR is located over the continent at about 30°S, which allows frontal systems to move over the southern parts of the continent. Any increases in STR position in JJA will thus move the STR to a more southerly position that results in a southerly shift of the mid-latitude storm track and decreased rainfall for southern Australia. In SON, the spatial pattern of significant correlations between gridded and area-averaged rainfall is similar to MAM with an additional region of negative correlations found in the SWC. In DJF, the relationship between STR position and rainfall is positive in continental Australia (NEC, SEC and MDB regions), suggesting that as the ridge increases its southerly position rainfall in Australia increases. A discussion of upper-layer soil moisture is next.

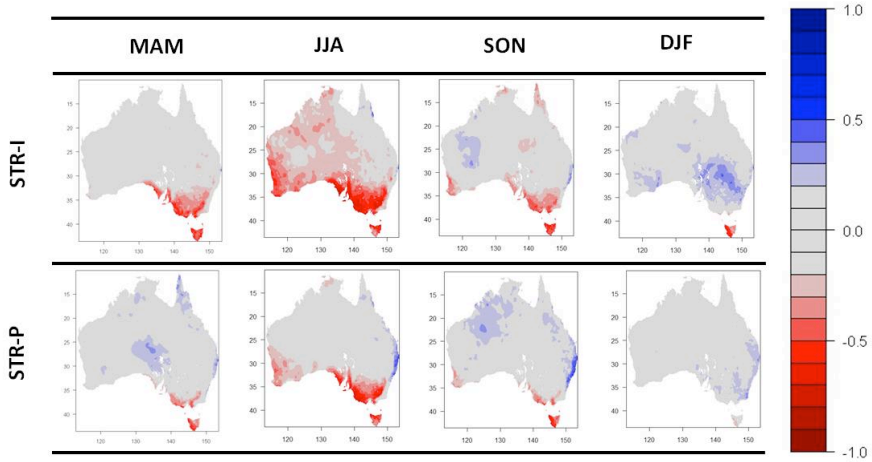

Figure 4.2. Correlation between gridded Australian rainfall and STR intensity (top) and position (bottom) in MAM, JJA, SON and DJF, over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

Table 4.3. Correlations between area-averaged rainfall in 12 drainage divisions and STR intensity ("Int") and position ("Pos") in MAM, JJA, SON and DJF, over the period 1900 – 2009. Bold indicates correlations that are significant at the 5% significance level.

Region	Division	Int: MAM	Pos: MAM	Int: JJA	Pos: JJA	Int: SON	Pos: SON	Int: DJF	Pos: DJF
North	TiS	-0.09	0.10	**-0.29**	-0.05	-0.08	**0.23**	-0.06	0.00
	GoC	-0.11	0.12	-0.05	0.06	-0.16	0.16	-0.04	0.09
Northeast	NEC	-0.01	**0.24**	0.08	0.32	-0.08	**0.24**	0.01	**0.20**
	NECBF	-0.1	0.15	-0.10	0.17	-0.13	0.18	0.01	0.10
Southeast	SEC	**-0.23**	0.10	**-0.37**	0.03	-0.14	0.16	0.19	**0.34**
	MDBW	**-0.46**	**-0.28**	**-0.69**	**-0.60**	**-0.41**	**-0.24**	0.17	**0.24**
	MDBA	**-0.24**	-0.02	**-0.39**	-0.11	-0.17	0.09	**0.23**	**0.25**
	MDBSA	**-0.25**	-0.03	**-0.46**	**-0.22**	**-0.22**	0.00	**0.32**	**0.20**
South	TAS	**-0.57**	**-0.48**	**-0.53**	**-0.51**	**-0.61**	**-0.57**	**-0.55**	-0.16
	SAG	**-0.36**	-0.01	**-0.67**	**-0.47**	**-0.30**	-0.18	**0.25**	0.18
West	SWC	-0.04	0.05	**-0.49**	**-0.34**	**-0.27**	**-0.19**	0.14	0.07
	IO	-0.07	-0.07	-0.34	-0.15	0.12	0.12	0.11	0.02

The relationship with STR variability is evident throughout the hydrological cycle and is spatially and temporally consistent. A remarkably similar pattern to that of rainfall is evident in both the correlations between STR intensity and position with upper-layer soil moisture (Figure 4.3). The relationship between STR intensity and upper-layer soil moisture peaks in JJA with strong (moderate) correlations over SEA (central and Western Australia). The SON and DJF patterns also reflect the rainfall spatial correlation pattern. Although, in MAM the relationship between upper-layer soil moisture and STR intensity is somewhat weaker than for rainfall. The negative correlations suggest that increase STR intensity is associated with decreased upper-layer soil moisture levels. The similarities between the spatial signatures of rainfall and upper-

layer soil moisture imply that rainfall is the climatological process that is the immediate driver of changes in upper-layer soil moisture and that the upper-layer of the soil column responds directly to changes in rainfall without any lag.

The relationship between STR position and upper-layer soil moisture (Figure 4.3) imitates the rainfall pattern in all seasons. The relationship peaks in JJA with negative correlations present across southern Australia. There are some regions of positive correlation along the eastern coast in SON, while DJF and MAM show few significant correlations.

The similarity between the rainfall and upper-layer soil moisture correlation patterns suggest that both variables are responding in the same way to climate drivers. Next is a discussion of the relationship between STR intensity and position with lower-layer soil moisture.

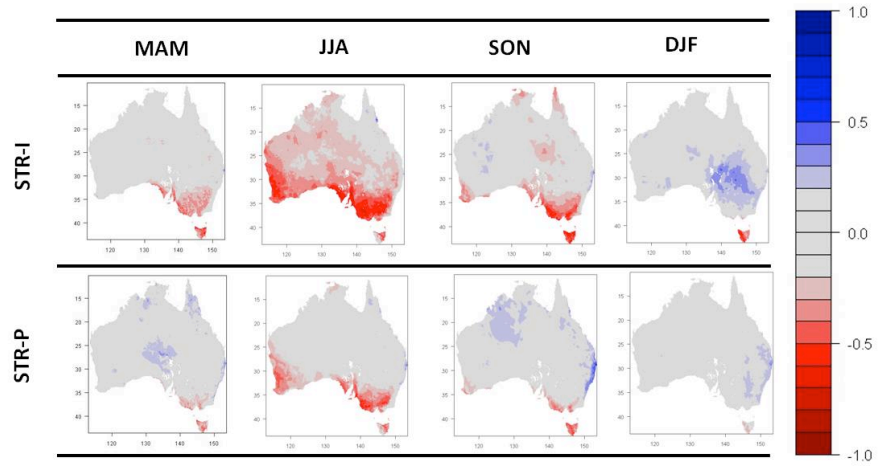

Figure 4.3. Correlation between gridded Australian upper-layer soil moisture and STR intensity (top) and position (bottom) in MAM, JJA, SON and DJF, over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

The simultaneous correlations between lower-layer soil moisture and the STR intensity and position (Figure 4.4) are weaker and have a reduced spatial extent than was seen previously for rainfall and upper-layer soil moisture. In MAM correlations with both STR intensity and position are insignificant over SEA but there are some regions of positive relationship in central Australia. The relationship peaks in JJA, consistent with the annual cycle of the STR, but the extent and strength is greatly reduced. SON and DJF display few significant correlations.

The differences between the correlation patterns for rainfall and upper-layer soil moisture compared to lower-layer soil moisture suggest that rainfall and upper-layer soil moisture respond to different processes than lower-layer soil moisture.

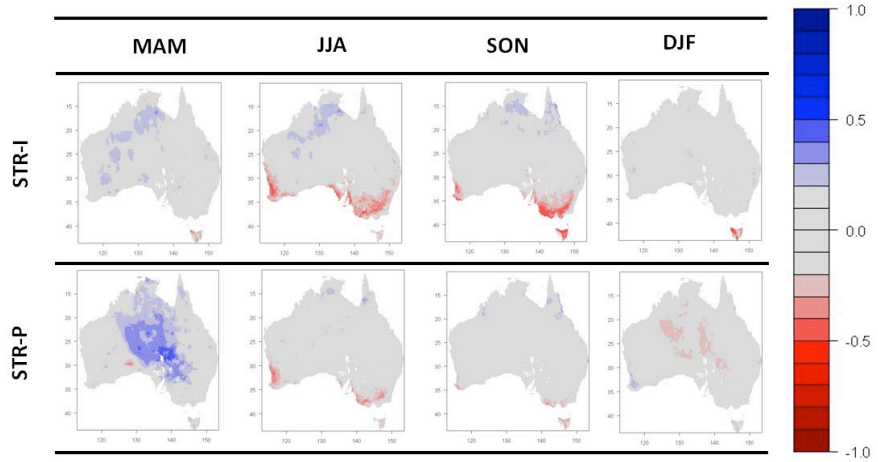

Figure 4.4. Correlation between gridded Australian lower-layer soil moisture and STR intensity (top) and position (bottom) in MAM, JJA, SON and DJF, over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

This section has shown that the simultaneous relationship between STR intensity and rainfall, upper-layer and lower-layer soil moisture is stronger than the relationship with STR position in all seasons. All response variables follow the seasonal cycle of the STR, with correlations peaking in JJA. The following section (4.4.1.b) will discuss the lagged correlations between STR variability and Australian hydroclimate.

4.4.1.b Lagged correlations

It was shown previously (Section 4.4.1.a) that the relationship between gridded and area-averaged rainfall and soil moisture (upper and lower) peaks in JJA. The importance of the MAM transition season (from the DJF to JJA patterns) for the STR has been discussed elsewhere (Timbal and Drosdowsky 2013). As such, the lagged correlations between JJA response variables and MAM STR intensity and position will be discussed here.

Despite large simultaneous correlations in JJA, few significant lagged correlations are evident with either facet of the STR in the preceding season (Figure 4.5). This is consistent with the correlations between MDB area-averaged rainfall and indices of the STR at 12 three-month sliding window lead-times (Figure 4.6). The maximum correlation is between the simultaneous values of STR intensity and position with MDB rainfall, with the linear relationship decaying steadily with increasing lead-time.

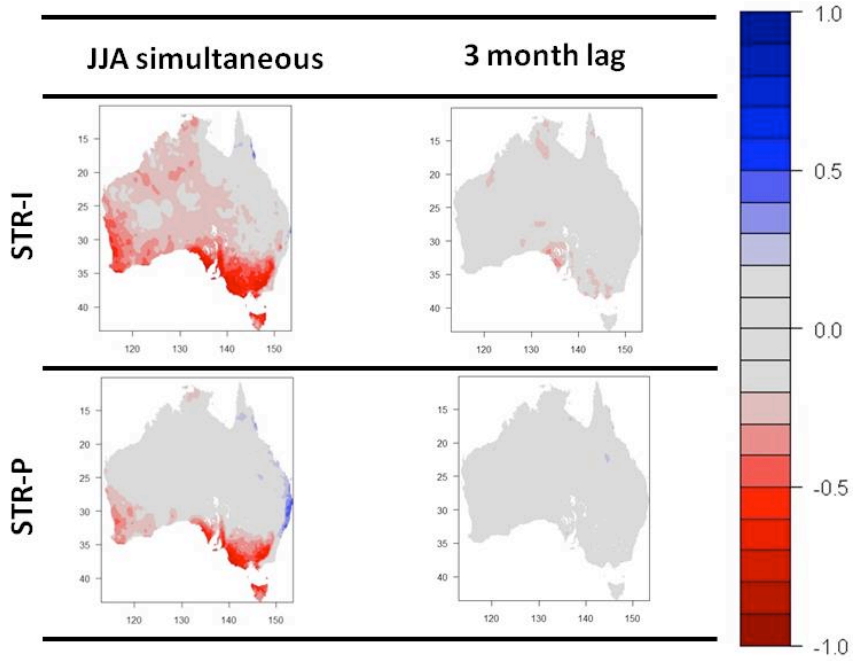

Figure 4.5. The simultaneous correlation between gridded JJA rainfall and STR intensity (top) and position (bottom) in JJA (left) and the 'lagged' correlation between JJA rainfall and the STR indices in MAM (right, three-month lead-time), over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

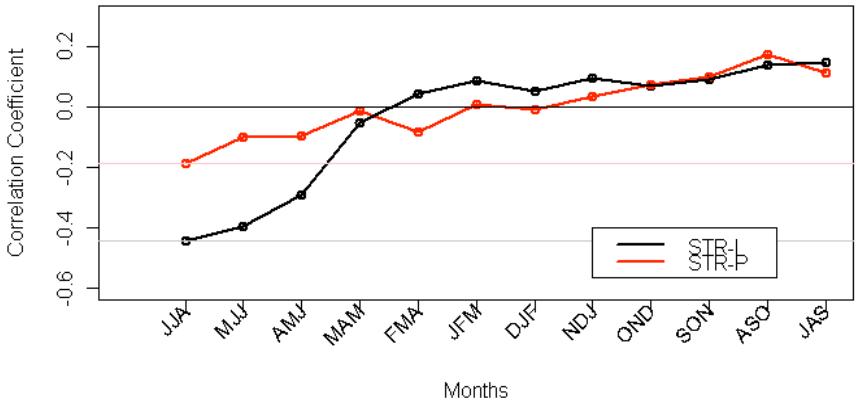

Figure 4.6 The correlation between JJA MDB rainfall and STR intensity (black) and position (red) with increasing lead-time, over the period 1900 – 2009. The simultaneous correlation is shown at JJA and the horizontal coloured lines indicate the simultaneous correlation level. Correlations exceeding +/-0.2 are significant at the 5% significance level.

The relationship between the STR and upper-layer soil moisture decays relatively quickly (Figure 4.7) although there are more significant correlations at the three-month lead-time for this hydrological variable than there was for rainfall. This suggests that while upper-layer soil

moisture responds quickly to changes in rainfall, the upper-layer soil moisture is able to smooth out some of the high-frequency variability in the relationship. At the three-month lead-time, the widespread and strong correlation evident in the simultaneous correlation between STR intensity and JJA rainfall retracts and weakens but is still significant over SEA. The negative correlation between JJA STR position and upper-layer soil moisture is considerably reduced at the three-month lead-time. Area-averaged MDB correlation between STR intensity and upper-layer soil moisture (Figure 4.8) peaks at the 1-month lead-time (MJJ) and with significant correlations continuing until MAM. The correlation between STR position and MDB upper-layer soil moisture decays quickly from the simultaneous maximum and is rarely significant.

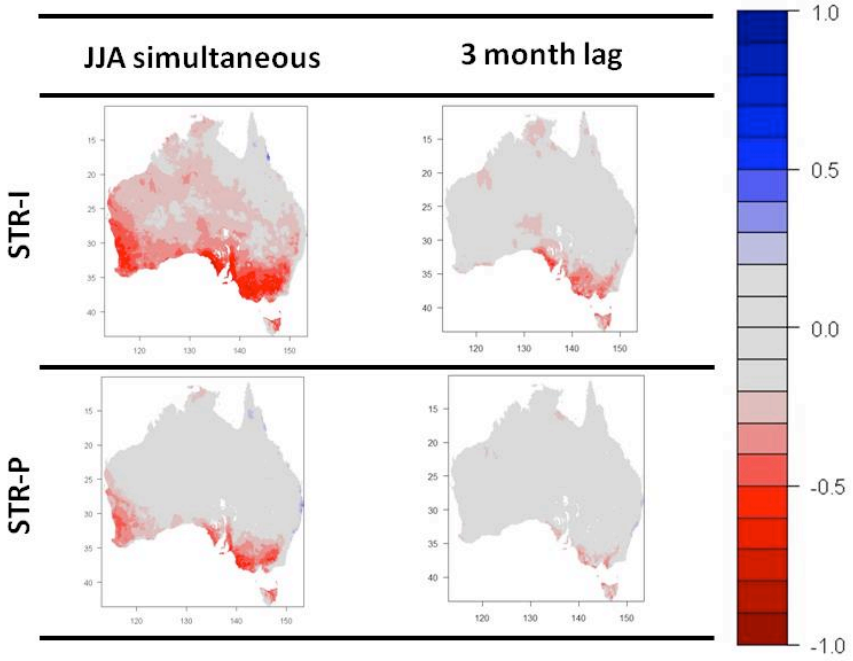

Figure 4.7. The simultaneous correlation between gridded JJA upper-layer soil moisture and STR intensity (top) and position (bottom) in JJA (left) and the 'lagged' correlation between JJA upper-layer soil moisture and the STR indices in MAM (right, three-month lead-time), over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

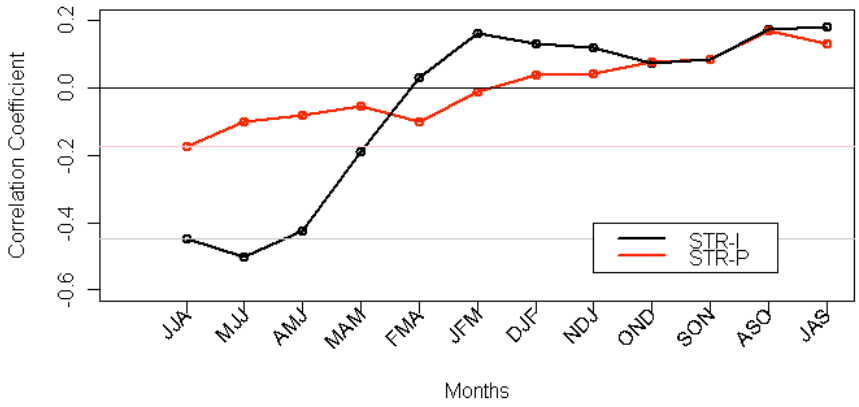

Figure 4.8. The correlation between JJA MDB upper-layer soil moisture and STR intensity (black) and position (red) with increasing lead-time, over the period 1900 – 2009. The simultaneous correlation is shown at JJA and the horizontal coloured lines indicate the simultaneous correlation level. Correlations exceeding +/-0.2 are significant at the 5% significance level.

The relationship between JJA gridded lower-layer soil moisture and MAM STR (Figure 4.9) decays less than the three-month lagged relationship with upper-layer soil moisture (Figure 4.7). The STR intensity correlation decreases in spatial extent but the strength of the negative correlation over SEA is similar. The correlation between MAM STR position and JJA lower-layer soil moisture in SEA also decreases in spatial extent but is not initially large. Interestingly, the three-month lagged relationship between STR position and the lower-layer of soil moisture displays a large, strong region of positive correlation in central Australia. This means that as the STR in MAM moves south, the saturation of the lower soil column in JJA increases in central Australia; although this result should be viewed with caution as data quality in central Australia is poor (Ummenhofer et al. 2011).

The correlation between STR intensity and lower-layer soil moisture (Figure 4.10) displays a strong lag up to two months before (AMJ) as the strength of the correlation increases at this lag. The simultaneous correlation is -0.25, which increases at a maximum of -0.36 at a 1-month lead-time but remains stronger than the simultaneous correlation at a two-month lead-time (r = -0.33). This shows that STR intensity in the preceding months has a stronger relationship with JJA lower-layer soil moisture than JJA STR intensity. This is in line with the lagged relationships we understand between rainfall and lower-layer soil moisture (Cai et al. 2009a).

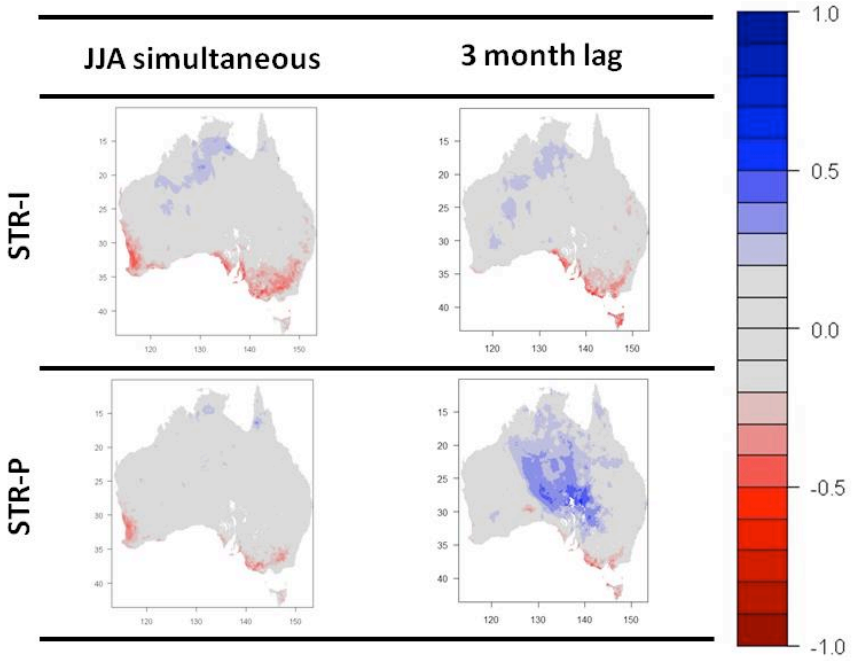

Figure 4.9. The simultaneous correlation between gridded JJA lower-layer soil moisture and STR intensity (top) and position (bottom) in JJA (left) and the 'lagged' correlation between JJA lower-layer soil moisture and the STR indices in MAM (right, three-month lead-time), over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

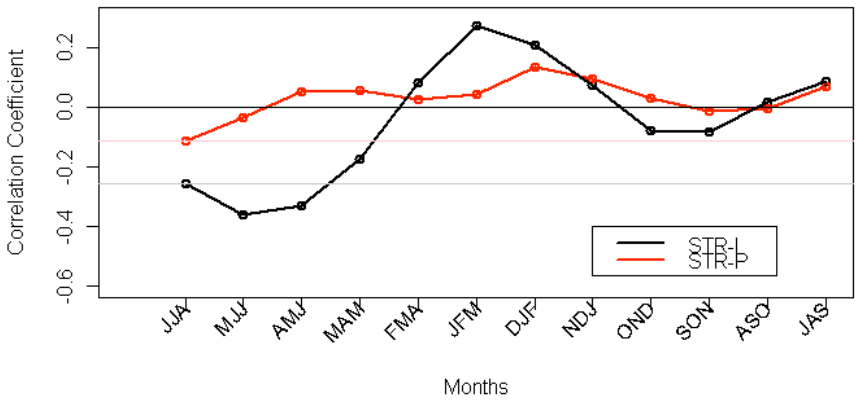

Figure 4.10. The correlation between JJA MDB lower-layer soil moisture and STR intensity (black) and position (red) with increasing lead-time, over the period 1900 – 2009. The simultaneous correlation is shown at JJA and the horizontal coloured lines indicate the simultaneous correlation level. Correlations exceeding +/-0.2 are significant at the 5% significance level.

The lack of significant correlations between rainfall and the STR at a three-month lead-time suggests that there is no persistence of either STR intensity or position, which is consistent with the high-frequency variability found in the atmosphere. The significant correlations

between upper-layer and lower-layer soil moisture, however, suggest that these hydrological variables are able to smooth out some high-frequency variability from the rainfall signal, as water takes time to move through the hydrological cycle.

This section has shown the consistency of the relationship between the STR and hydroclimatic variability in Australia. For all response variables, the relationship with STR intensity was stronger than the relationship with STR position, for all response variables, consistent with previous research (Timbal and Drosdowsky 2013). The complexity of the relationship between STR position and Australian hydroclimatic variability was demonstrated as the sign of the relationship depends on the mean position of the STR. The following section will discuss the relationship between ENSO and hydroclimatic variability in Australia.

4.4.2 El Niño – Southern Oscillation (ENSO)

The previous section (4.4.1) discussed the linear relationships between facets of STR variability and Australian hydroclimate. This section will discuss the linear relationship throughout the seasonal cycle between several oceanic and atmospheric indices of ENSO variability with gridded and area-averaged Australian rainfall and soil moisture (upper and lower layers). The relationship between ENSO and Australian hydroclimatic variability is well known (Chiew et al. 1998, Drosdowsky and Williams 1991, Nicholls et al. 1997) and this section will establish that these relationships are reproducible with the current dataset. The seasonality and spatial pattern of the simultaneous relationship will be discussed first (Section 4.4.2.a), followed by a discussion of the lagged relationship (Section 4.4.2.b). In each sub-section, the relationship between rainfall and ENSO is discussed first, followed upper-layer and lower-layer soil moisture.

4.4.2.a Spatial correlations

The relationship between various indices of ENSO (Niño 3, Niño 4, Niño 3.4, SOI) and Australian rainfall follows a distinct seasonal cycle (Figure 4.11 and Table 4.4), consistent with previous research (Risbey et al. 2009). The correlations with the SOI are the strongest in all seasons and this is a common result across response variables. The SOI is calculated from MSLP and is thus more closely related to the processes that generate rainfall, whereas the oceanic measures are physically one step removed from the rainfall generation process and so, accordingly, the correlations here are weaker.

Broadly, the relationship builds in JJA and peaks in SON before decaying in DJF, with little correlation evident in MAM. More specifically, in JJA significant correlations between the oceanic indices are evident over northern and eastern Australia (NEC, NECBF, MDBA,

MDBSA, MDBW, SEC, TAS, SWC), while correlations with the SOI are more extensive in southern Australia (GoC, SAG, IO in addition to the regions listed previously). The sign of the relationships indicate that a positive SOI (oceanic index) is associated with increase (decreased) rainfall. The lack of significant correlations in the northern Australian regions (TiS and GoC) is likely related to the low rainfall received in those divisions in JJA and suggests that rainfall in the northeastern region (NEC, NECBF) stems from different mechanisms. Spatially the SOI has a more extensive relationship with rainfall in JJA, while the oceanic indices are confined to the central eastern coast during this season. This suggests that the atmospheric component of ENSO has a more immediate relationship with rainfall, and that although the SST anomalies begin to build during this season the relationship with rainfall is not evident until SON.

In SON, the eastern third of the continent is covered by significant correlations with both the SOI and the oceanic indices; this is reflected by the significant correlations in all divisions except IO. In DJF, the extent of the significant relationship is confined to northern (TiS, GoC, NEC, NECBF), eastern (SEC, MDBA, MDBSA, MDBW) and southwestern (SWC, IO) Australia, with the magnitude decreasing in most divisions. During DJF the correlations with Niño 3 and Niño 4 are weaker than those found in SON but cover the majority of the eastern coast. The correlations between the SOI and rainfall in DJF are stronger and more widespread than the oceanic indices correlations as they extend west across northern Australia. Finally, in MAM significant correlations are found in northern Australia (TiS, GoC, NEC) and the SEC.

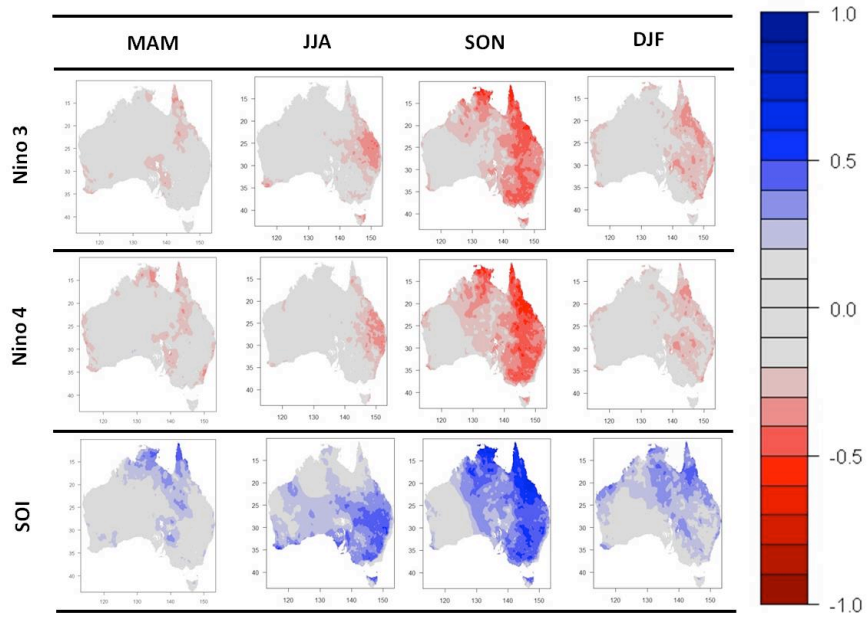

Figure 4.11 Correlation between gridded Australian rainfall and three indices of ENSO - Niño 3 (top), Niño 4 (middle) and the SOI (bottom) - in MAM, JJA, SON and DJF, over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

Table 4.4. Correlations between area-averaged rainfall in 12 drainage divisions and two indices of ENSO - Niño 3.4 (N34) and the SOI (SOI) - in MAM, JJA, SON and DJF, over the period 1900 – 2009. Bold indicates correlations that are significant at the 5% significance level.

Region	Division	N34: MAM	SOI: MAM	N34: JJA	SOI: JJA	N34: SON	SOI: SON	N34: DJF	SOI: DJF
North	TiS	**-0.22**	**0.32**	-0.01	0.16	**-0.53**	**0.59**	**-0.22**	**0.42**
	GoC	**-0.25**	**0.43**	-0.15	**0.19**	**-0.52**	**0.63**	**-0.35**	**0.51**
Northeast	NEC	**-0.25**	**0.35**	**-0.36**	**0.35**	**-0.57**	**0.63**	**-0.39**	**0.50**
	NECBF	-0.12	0.16	**-0.38**	**0.39**	**-0.55**	**0.63**	**-0.35**	**0.42**
Southeast	SEC	**-0.21**	**0.21**	**-0.31**	**0.41**	**-0.45**	**0.49**	**-0.30**	**0.30**
	MDBW	-0.05	0.17	**-0.21**	**0.46**	**-0.48**	**0.49**	**-0.24**	**0.22**
	MDBA	-0.01	0.09	**-0.35**	**0.58**	**-0.49**	**0.57**	**-0.32**	**0.26**
	MDBSA	-0.14	0.20	**-0.28**	**0.50**	**-0.50**	**0.54**	**-0.27**	**0.25**
South	TAS	0.02	0.09	-0.18	**0.32**	**-0.30**	**0.31**	-0.05	0.05
	SAG	-0.18	0.18	-0.17	**0.45**	**-0.32**	**0.32**	-0.08	0.10
West	SWC	**-0.22**	0.16	**-0.21**	**0.36**	**-0.19**	**0.20**	**-0.22**	**0.23**
	IO	-0.10	0.07	-0.11	**0.26**	-0.16	0.11	**-0.30**	**0.35**

The relationship between ENSO and upper-layer soil moisture (Figure 4.12) is similar but somewhat larger in magnitude to that of rainfall, which again suggests that upper-layer soil moisture responds directly to changes in rainfall and the large-scale variability. For the oceanic indices the relationship in JJA is moderate while the SOI has a much stronger and more spatially extensive relationship. There are few differences between the atmospheric and oceanic measures of ENSO in SON as the eastern two thirds of the continent are covered in strong correlations, showing that El Niño events are associated with decreases in upper-layer soil moisture.

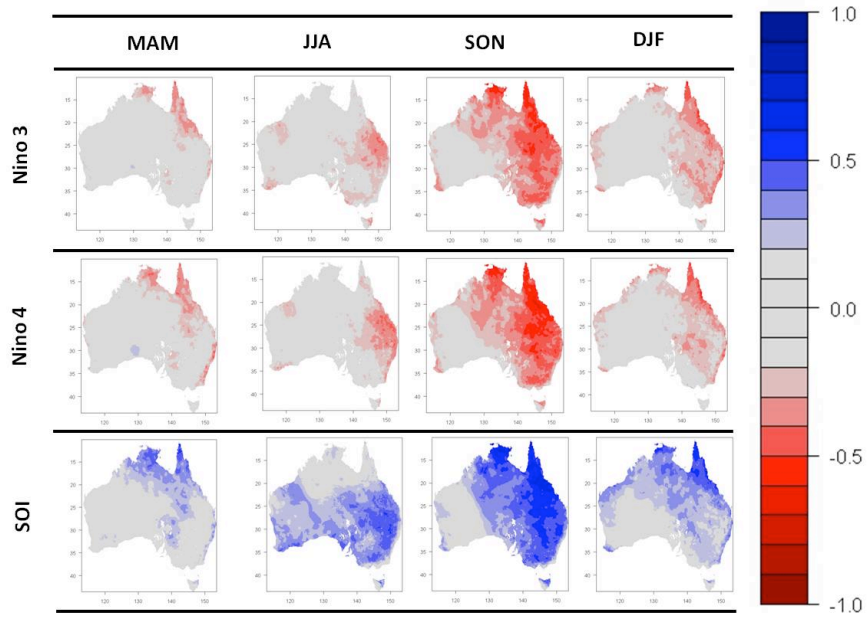

Figure 4.12. Correlation between gridded Australian upper-layer soil moisture and three indices of ENSO - Niño 3 (top), Niño 4 (middle) and the SOI (bottom) - in MAM, JJA, SON and DJF, over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

The relationship between lower-layer soil moisture and ENSO (Figure 4.13) is weaker and has reduced spatial extent than we saw for rainfall or upper-layer soil moisture. Correlations are strongest in MAM, SON and DJF while little relationship is evident in JJA. Previous research has suggested a lag between the climate forcing, a climatic response and a response in the lower-layer of soil moisture (Raupach et al. 2008) and this may explain the correlation patterns we see in Figure 4.13. There is little signal in the lower-layer in JJA as ENSO is inactive in the preceding season. As ENSO develops during JJA we can see a signal in the lower-layer increasing through SON into DJF when the relationship between ENSO and lower-layer soil moisture peaks, which corresponds to the SON peak in ENSO that would be expected. There is a much greater correlation found in MAM than would be expected but this may be due to lagged effects from the DJF ENSO – rainfall relationship.

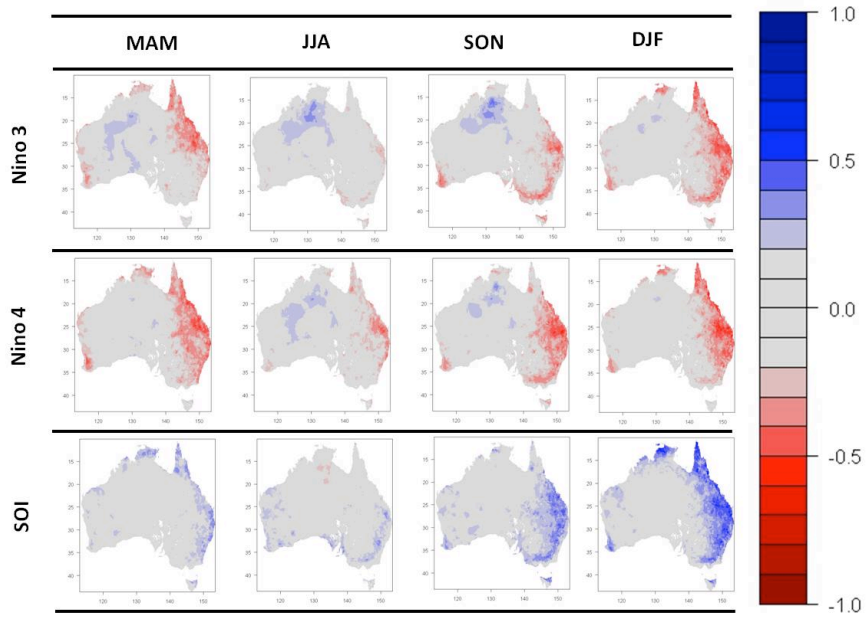

Figure 4.13. Correlation between gridded Australian lower-layer soil moisture and three indices of ENSO - Niño 3 (top), Niño 4 (middle) and the SOI (bottom) - in MAM, JJA, SON and DJF, over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

This section has outlined the spatial extent and magnitude of the relationship between three indices of ENSO and Australian hydroclimatic variability. The relationship between the three indices was consistent, with the SOI having the strongest influence on Australian rainfall and soil moisture. The correlations peaked in SON, with some significant relationships evident in JJA and DJF. The following section (4.4.2.b) will discuss the strength of the correlation between SON hydroclimatic variability and ENSO at various lead-times.

4.4.2.b Lagged correlations

The previous section (4.4.2.a) showed that ENSO related variability is greatest in SON, so this season will be the focus of this section. The significant lagged correlations between rainfall in SON and indices of ENSO in JJA are widespread, covering large parts of eastern Australia (Figure 4.14). Although the significant correlations cover eastern Australia, the strength of the relationship is decreased at a three-month lag, particularly for the SOI. A comparison of simultaneous and lagged spatial patterns suggests that the relationship declines linearly. This is supported by the area averaged MDB lagged correlations (Figure 4.15). The weakest (strongest) simultaneous correlation is with Niño 3 (SOI), r = -0.49 (r = -0.57) but the lagged correlations with the SOI decrease more quickly than the oceanic indices. The lagged relationship with rainfall stems from the persistence of an ENSO event so the stronger lagged correlations with the oceanic indices is consistent with the reduced variability in these slowly evolving indices.

Short-term variations in the atmosphere not associated with the development of ENSO events would weaken the correlation at longer lead-times.

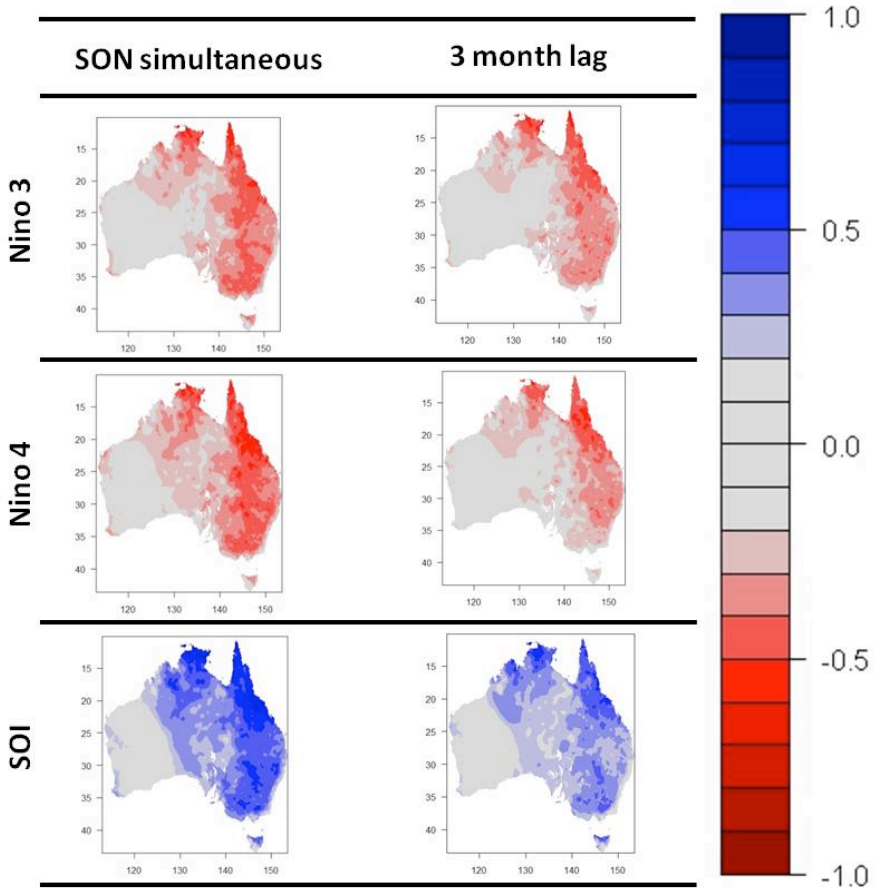

Figure 4.14. The simultaneous correlation between gridded SON rainfall and three indices of ENSO - Niño 3 (top), Niño 4 (middle) and the SOI (bottom) - in SON (left) and the 'lagged' correlation between SON rainfall and the ENSO indices in JJA (right, three-month lead-time), over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

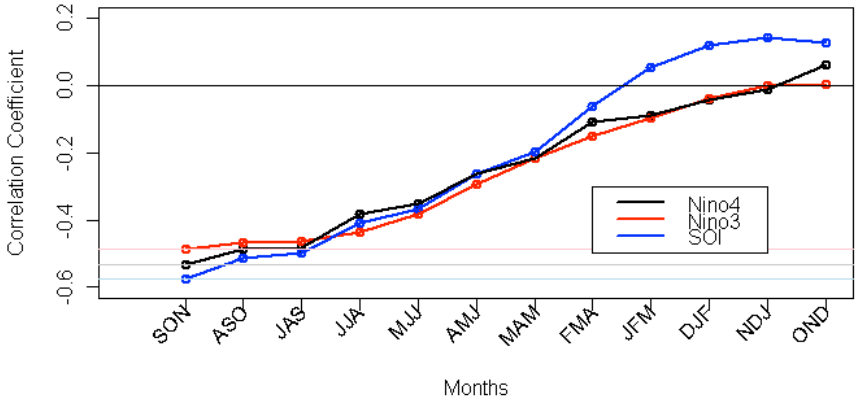

Figure 4.15. The correlation between SON MDB rainfall and three indices of ENSO - Niño 3 (red), Niño 4 (black) and the SOI (blue, sign of the correlation reversed) - with increasing lead-time, over the period 1900 – 2009. The simultaneous correlation is shown at SON and the horizontal coloured lines indicate the simultaneous correlation level. Correlations exceeding +/-0.2 are significant at the 5% significance level.

There is little decay from the simultaneous relationship in correlations between SON upper-layer soil moisture and indices of ENSO at a three-month lead-time, particularly in northeastern Australia (Figure 4.16). The strongest lagged correlations are with the SOI in northern and southeastern Australia. A steady decline in the correlation coefficient with increasing lead-time is evident in the MDB area-averaged lagged correlations (Figure 4.17).

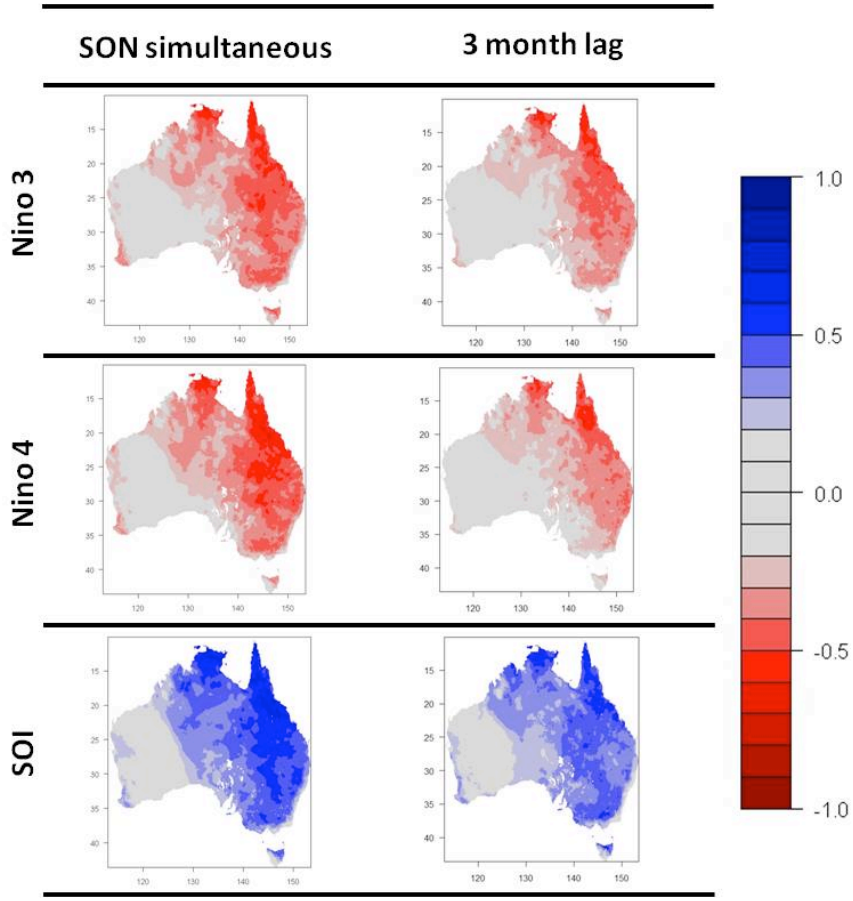

Figure 4.16. The simultaneous correlation between gridded SON upper-layer soil moisture and three indices of ENSO - Niño 3 (top), Niño 4 (middle) and the SOI (bottom) - in SON (left) and the 'lagged' correlation between SON upper-layer soil moisture and the ENSO indices in JJA (right, three-month lead-time), over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

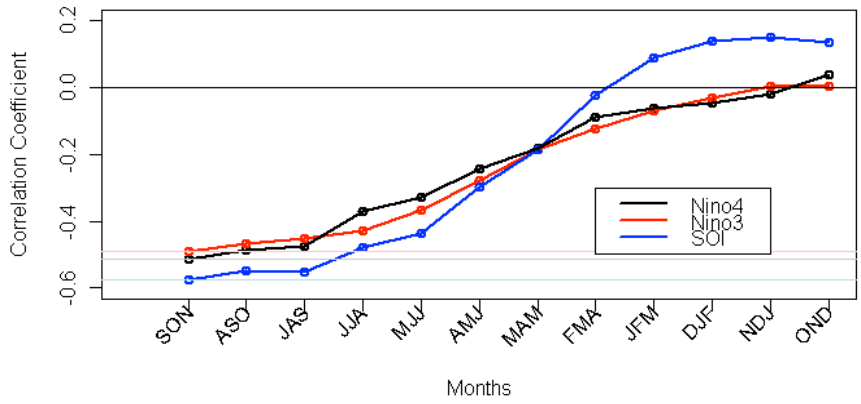

Figure 4.17. The correlation between SON MDB upper-layer soil moisture and three indices of ENSO - Niño 3 (red), Niño 4 (black) and the SOI (blue, sign of the correlation reversed) - with increasing lead-time, over the period 1900 – 2009. The simultaneous correlation is shown at SON and the horizontal coloured lines indicate the simultaneous correlation level. Correlations exceeding +/-0.2 are significant at the 5% significance level.

Similarly to the simultaneous relationship, the lagged relationship between SON lower-layer soil moisture and JJA indices of ENSO is considerably reduced compared to rainfall and upper-layer soil moisture (Figure 4.18). Again, the lagged relationship is strongest with the SOI, particularly along the GDR, the central eastern coast and in Victoria, while Niño 4 has the strongest relationship in southeast Queensland. The MDB area-averaged lower-layer soil moisture correlations with Niño 3 at increasing lead-time (Figure 4.19) decline more gradually than the lagged correlation with upper-layer soil moisture. Indeed, the correlations between lower-layer soil moisture and the oceanic indices of ENSO are significant out to the six-month lead-time and positive until the previous year. The correlation between lower-layer soil moisture and the SOI increases with lead-time. Although no increase in correlation is found with the oceanic indices so this increasing relationship with the SOI may be a statistical artefact caused by sampling variability.

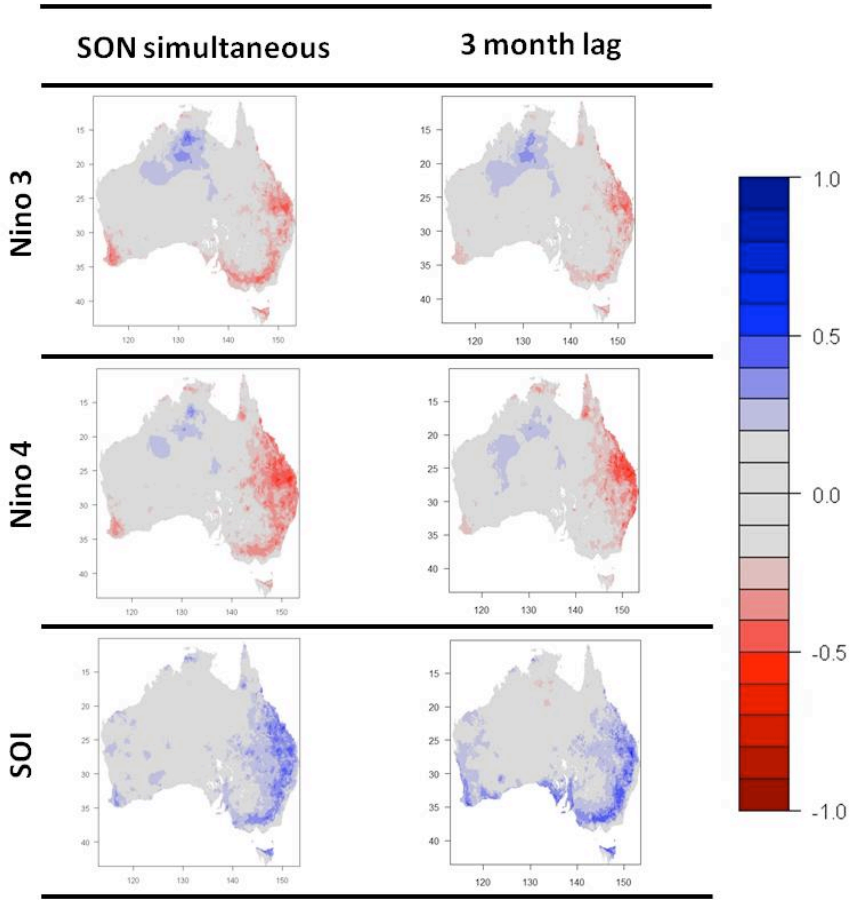

Figure 4.18. The simultaneous correlation between gridded SON lower-layer soil moisture and three indices of ENSO - Niño 3 (top), Niño 4 (middle) and the SOI (bottom) - in SON (left) and the 'lagged' correlation between SON lower-layer soil moisture and the ENSO indices in JJA (right, three-month lead-time), over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

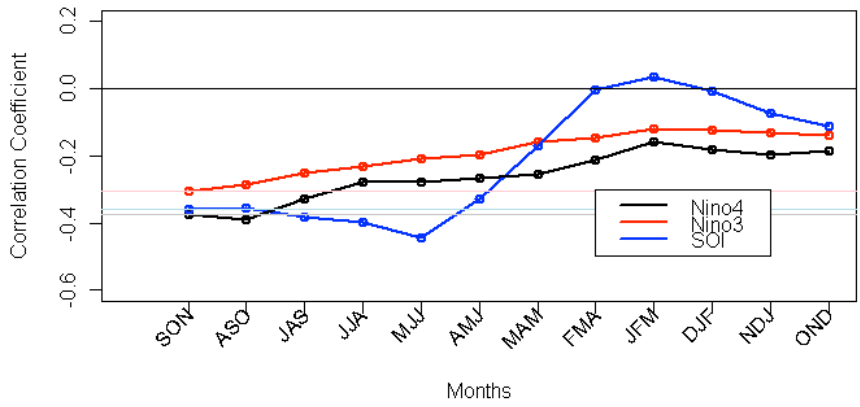

Figure 4.19. The correlation between SON MDB lower-layer soil moisture and three indices of ENSO - Niño 3 (red), Niño 4 (black) and the SOI (blue, sign of the correlation reversed) - with increasing lead-time, over the period 1900 – 2009. The simultaneous correlation is shown at SON and the horizontal coloured lines indicate the simultaneous correlation level. Correlations exceeding +/-0.2 are significant at the 5% significance level.

This section has shown the relationship between ENSO and hydroclimatic variability using the novel AWAP data set. The seasonal relationships between rainfall, upper-layer and lower-layer soil moisture with indices of ENSO variability are consistent with previous research (Kirono et al. 2010, Risbey et al. 2009). Correlations are significant and widespread in JJA, SON and DJF, with the strongest relationships with indices that are closer to Australia (SOI and Niño 4). Significant correlations are present over eastern Australia for all response variables indicating the persistence of ENSO from JJA to SON. Hydrological variables again display less decrease in the strength of the relationship with increasing lead-time, compared to rainfall.

The following section (4.4.3) will look spatial and temporal variability in the relationship between the Indian Ocean and Australian hydroclimatic variability.

4.4.3 Indian Ocean variability

The previous section (4.4.2) showed that relationships between ENSO were significant over eastern Australia, peaking in SON. This section will shift the focus from the Pacific to the Indian Ocean. Indices of Indian Ocean variability have a well-established relationship with Australian hydroclimatic variability (Cai et al. 2009b, Drosdowsky and Chambers 2001, Nicholls 1989, Risbey et al. 2009, Ummenhofer et al. 2009a, Ummenhofer et al. 2011). Here the focus will be on three indices of Indian Ocean variability. 1) The Dipole Mode Index (DMI) is the most commonly used index from the Indian Ocean, capturing variability in both the eastern and western parts of the basin (Saji et al. 1999). 2) Region C is an area-averaged SST anomaly

box near the Indonesian through-flow region (Cai and Cowan 2008a). 3) The Tripole Index (TPI) is comprised of three area-averaged SST anomaly regions; the average of boxes from the Indian Ocean and Pacific Ocean boxes are subtracted from the maritime continent SST anomalies (Timbal and Hendon 2011). This means that the TPI is a measure of both ocean basins combined and is included in this section (following the linear results of ENSO) for comparative purposes.

The focus in the following sections is on the three response variables discussed previously in this chapter (rainfall, upper-layer and lower-layer soil moisture). Next is a discussion of the spatial correlations in the four regular seasons (Section 4.4.3.a), followed an analysis the lagged relationship (Section 4.4.3.b).

4.4.3.a Spatial correlations

Rainfall has a strong relationship with indices from the Indian Ocean in JJA and SON (Figure 4.20), while there are few significant correlations in DJF or MAM, consistent with the development of the IOD (Saji et al. 1999) and previous research (Risbey et al. 2009). In MAM, there is a significant negative correlation between NECBF and the DMI (Table 4.5) reflected in the spatial analysis (Figure 4.20). In DJF, there are some regions of positive correlation between rainfall and the DMI in northeast Australia (GoC, NEC, NECBF). Negative correlations between Region C and area-averaged rainfall extend further south to include SEC, MDBW and MDBA in addition to GoC, NEC and NECBF. While significant correlations in DJF are not widespread both indices change sign in this season compared to JJA and SON (Table 4.5 and Figure 4.20).

The strongest correlations for SEA in both JJA and SON are with the TPI, likely due to the incorporation of both Pacific and Indian Ocean variability in this index. Correlations with Region C are stronger than those with the DMI due to the proximity of Region C to Australia, while the DMI is partly based on remote SST anomalies. The correlation with the TPI is positive, suggesting that as the TPI increases rainfall also increases. Because the TPI is the difference between the maritime continent SST and the average of two SST boxes (from the Indian and Pacific Oceans) a positive index suggests that the maritime continent is warmer than the remote boxes. The DMI is the difference between SST anomalies in boxes in the western and eastern Indian Ocean, so that a positive DMI means the western Indian Ocean is warmer than the eastern part. As such, as cooler SST near Australia are associated with decreased rainfall and a negative correlation. Region C has a positive relationship with rainfall, as warmer SST in the ITF region is associated with increased rainfall. The spatial extent of the TPI is similar to the pure Indian Ocean indices, suggesting that the Indian Ocean is an important

component of the TPI, and the close relationship between the effects of the Indian and Pacific Oceans on Australia through changes to the maritime continent region.

In JJA significant correlations extend over southern Australia. Significant correlations with the DMI are confined to southwest Western Australia and southeast Australia (SEC, MDB regions, TAS, SAG< IO), while the influence of Region C and the TPI extends further into northern Australia. Region C has significant positive correlations with all area-averaged regions in JJA; this suggests that warm SST anomalies in the maritime continent region are associated with increased rainfall over the whole continent, even the regions in southern Australia that received frontal rainfall. The spatial pattern of the significant correlations between rainfall and the Region C index in JJA is similar to the May correlation discussed in (Cai and Cowan 2008a) which suggests that May is more representative of a JJA pattern than a MAM pattern, possibly due to interactions with the STR. This pattern is likely linked to northwest cloud bands (Cai and Cowan 2008a, Nicholls 1989) and the high-moisture flows that veer north eastward from tropics to southeast Australia (Cai and Cowan 2008a)

All three indices affect both eastern and western Australia in JJA, although this is not the case in SON when only the eastern half of the continent shows significant correlations (Figure 4.20). There are significant negative (positive) correlations between upper-layer soil moisture in all (all but IO) area-averaged regions and DMI (Region C, Table 4.5) that are not reflected in the spatial pattern (Figure 4.20).

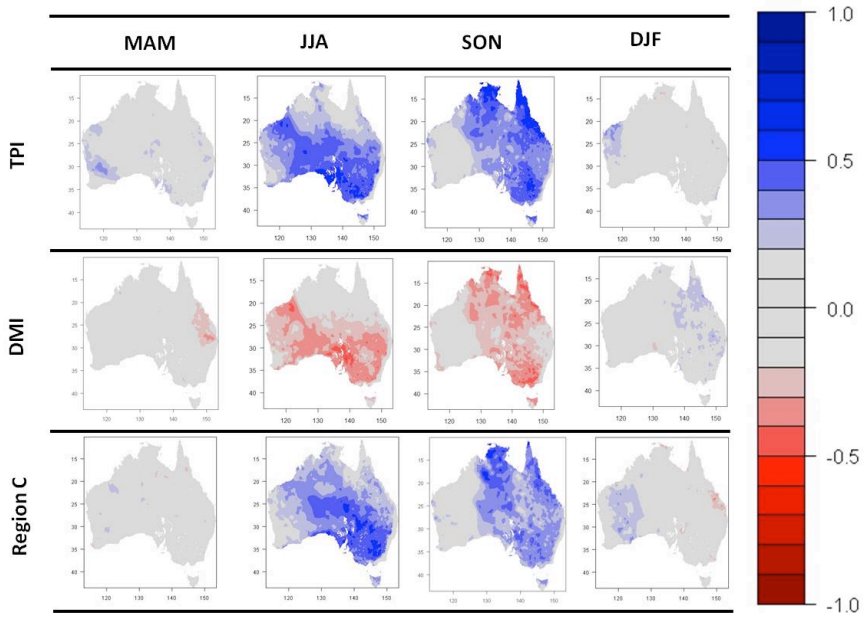

Figure 4.20. Correlation between gridded Australian rainfall and three indices of Indian Ocean variability - Tri-pole index (top), the DMI (middle) and Region C (bottom) - in MAM, JJA, SON and DJF, over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

Table 4.5. Correlations between area-averaged rainfall in 12 drainage divisions and two indices of Indian Ocean variability - the DMI (DMI) and Region C (RegC) - in MAM, JJA, SON and DJF, over the period 1900 – 2009. Bold indicates correlations that are significant at the 5% significance level.

Region	Division	DMI: MAM	RegC: MAM	DMI: JJA	RegC: JJA	DMI: SON	RegC: SON	DMI: DJF	RegC: DJF
North	TiS	0.06	-0.16	-0.03	**0.38**	**-0.44**	**0.47**	0.10	-0.13
	GoC	0.00	-0.17	0.02	**0.31**	**-0.44**	**0.50**	**0.24**	**-0.21**
Northeast	NEC	-0.18	-0.09	-0.10	**0.29**	**-0.39**	**0.49**	**0.21**	**-0.19**
	NECBF	**-0.25**	-0.03	-0.12	**0.37**	**-0.36**	**0.48**	**0.21**	**-0.19**
Southeast	SEC	-0.09	-0.01	**-0.19**	**0.28**	**-0.30**	**0.34**	0.10	**-0.22**
	MDBW	-0.1	0.04	**-0.33**	**0.52**	**-0.46**	**0.44**	0.09	**-0.20**
	MDBA	-0.18	0.06	**-0.33**	**0.59**	**-0.34**	**0.43**	0.14	**-0.20**
	MDBSA	-0.10	-0.01	**-0.35**	**0.60**	**-0.39**	**0.45**	0.14	-0.10
South	TAS	0.00	0.01	-0.15	**0.31**	**-0.30**	**0.34**	0.02	-0.08
	SAG	-0.09	0.04	**-0.39**	**0.56**	**-0.35**	**0.39**	0.09	-0.11
West	SWC	0.00	-0.04	0.00	**0.24**	**-0.22**	**0.22**	0.02	0.05
	IO	-0.01	0.07	**-0.26**	**0.33**	-0.15	0.05	-0.05	0.04

The seasonal spatial patterns of significant correlations between indices of Indian Ocean variability and rainfall are also evident in the upper-layer soil moisture seasonal cycle (Figure 4.21). Similarly to rainfall, the relationship between Indian Ocean variability and upper-layer soil moisture is much stronger and more widespread in the extended cool season (JJA and SON) compared to the warm season (DJF and MAM). However, warm season correlations with upper-layer soil moisture are more widespread than they were for rainfall, particularly in coastal regions.

The relationship is strongest in JJA in the southern half of the continent, particularly for the TPI. Significant correlations for the TPI stretch further north, while those for Region C are centred on SEA and aligned in a northwesterly orientation. In SON the relationship is again confined to the eastern half of the continent. The correlations are stronger with the TPI than the Indian Ocean indices, showing the importance of combined Indo-Pacific forcings for southeast Australian hydroclimatic variability. In many places the relationship in SON is stronger than the relationship with rainfall, for example, the DMI shows more significant correlations for upper-layer soil moisture on the east coast.

The change in the sign of the relationship over eastern Australia in DJF is again evident, although for upper-layer soil moisture the correlations with the DMI are more extensive than they were for rainfall.

The simultaneous correlation here is stronger than the simultaneous correlation for rainfall, showing that variability in the Indian Ocean has a greater relationship with upper-layer soil moisture than it does with rainfall, even though the two patterns have a similar spatial extent.

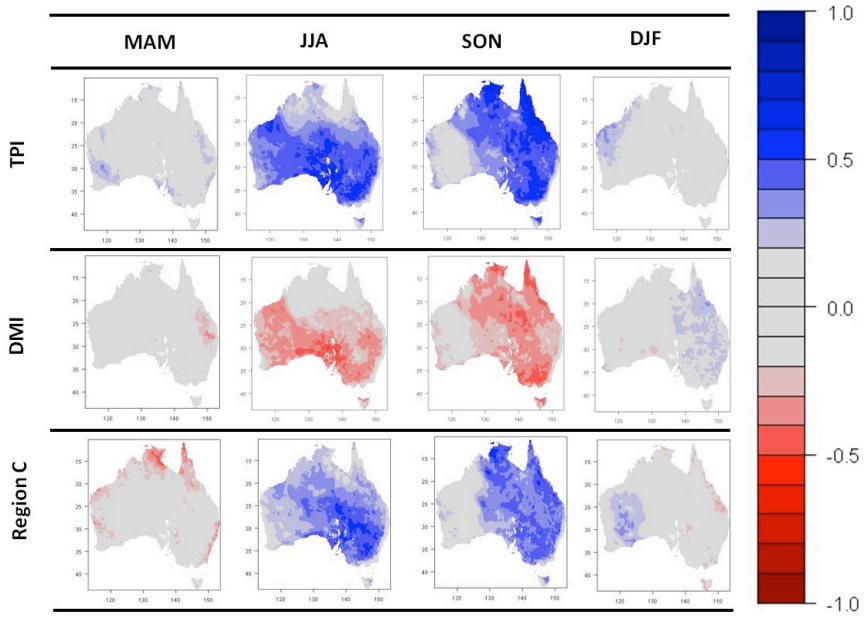

Figure 4.21. Correlation between gridded Australian upper-layer soil moisture and three indices of Indian Ocean variability - Tri-pole index (top), the DMI (middle) and Region C (bottom) - in MAM, JJA, SON and DJF, over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

The relationship between lower-layer soil moisture and Indian Ocean variability is less spatially extensive than was seen previously for rainfall or upper-layer soil moisture (Figure 4.22). Again the relationship is most prominent in JJA and SON, although there are significant correlations reaching into DJF, particularly for the DMI although they are of the same sign as

the cool season correlations. Significant correlations for the TPI and the DMI are centred on southeast Australia where a positive DMI (TPI) is associated with decreased (increased) soil moisture. There are significant correlations in the southwest of Western Australia for both the DMI and the TPI but a much greater affect is seen in the west with Region C. In all seasons the strong correlations in central Western Australia that suggests increasing SST anomalies in Region C is associated with increased lower-layer soil moisture in Western Australia; however it should be kept in mind that data quality in this region is low as there are few observations in this area which can cause clustering around observation points.

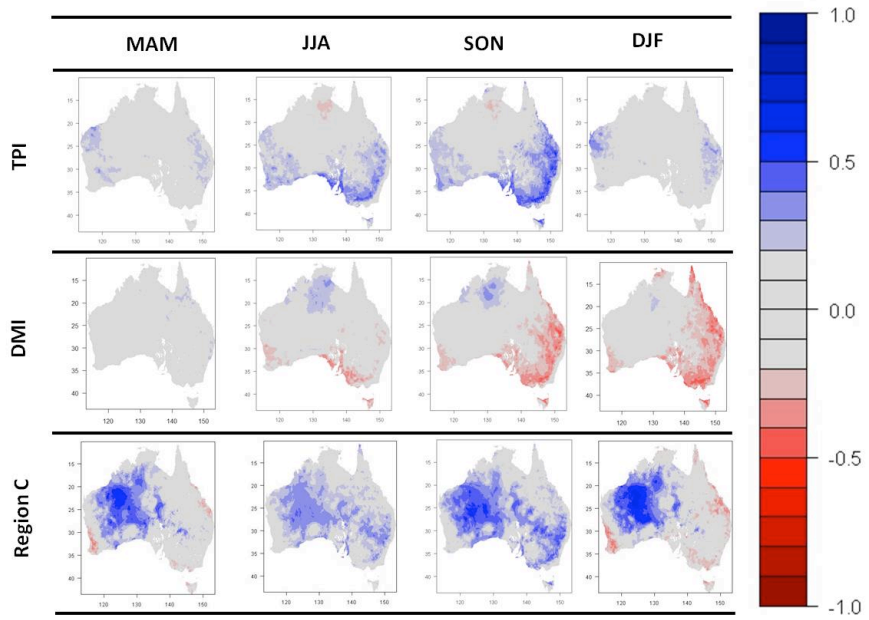

Figure 4.22. Correlation between gridded Australian lower-layer soil moisture and three indices of Indian Ocean variability - Tri-pole index (top), the DMI (middle) and Region C (bottom) - in MAM, JJA, SON and DJF, over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

This section has outlined the relationship between indices of Indian Ocean variability with gridded and regional Australian rainfall. Differences in the magnitude, spatial extent and sign of the relationship are evident between the cool (JJA and SON) and warm (DJF and MAM) season for rainfall, upper-layer and lower-layer soil moisture. In particular, rainfall and upper-layer soil moisture change sign between SON and DJF while lower-layer soil moisture is of the same sign. This may suggest persistence in lower-layer soil moisture that is not evident in the other response variables. This issue will be discussed further below (Section 4.4.3.b).

4.4.3.b Lagged correlations

The previous section (4.4.3.a) showed that the influence of the Indian Ocean was strong in both SON and JJA. This section will outline the correlations between SON response variables and JJA indices of Indian Ocean variability.

The correlations between rainfall and the TPI and Region C are weaker at the one-season lead-time but still widely significant, while the DMI correlations decay markedly (Figure 4.23). The area-averaged MDB correlations (Figure 4.24) reiterate the strength of the TPI – rainfall relationship, particularly compared to the DMI. These two indices display a similar pattern of lagged relationship, although the TPI is considerably stronger, both decaying quickly from the simultaneous maximum. The correlation with Region C decays more slowly for the preceding two months and it is not until JJA that it drops away quickly. As rainfall is expected to respond immediately to tropical SST variations, as was seen for ENSO, the lack of lagged relationship with the IOD suggests limited seasonal persistence. The significant relationship with Region C out to the two-month lead-time suggests that when SST anomalies in the maritime continent region develop in JJA they generally persist through to SON.

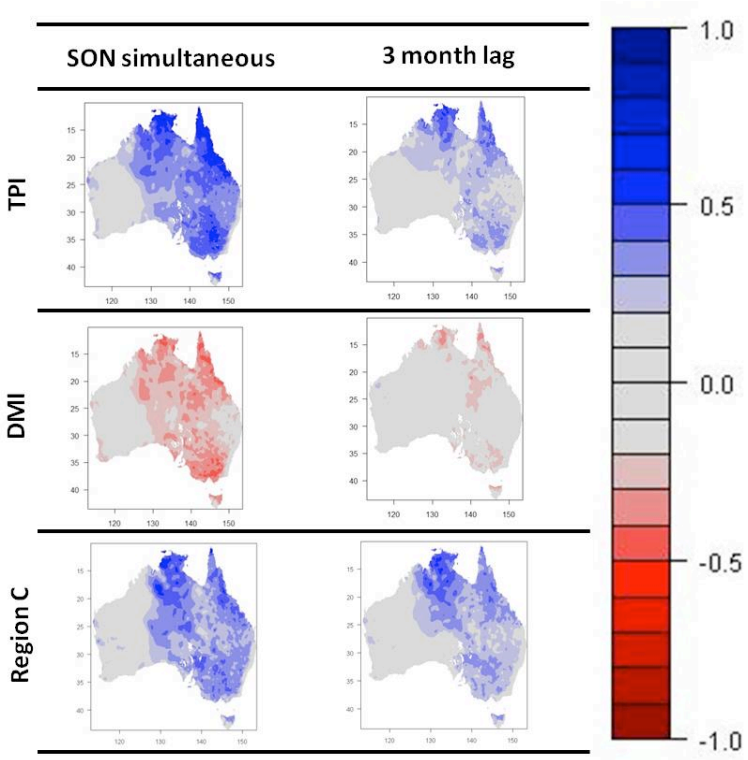

Figure 4.23. The simultaneous correlation between gridded SON rainfall and three indices of Indian Ocean variability - Tri-pole index (top), the DMI (middle) and Region C (bottom) - in SON (left) and the 'lagged' correlation between SON rainfall and the Indian Ocean indices in JJA (right, three-month lead-time), over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

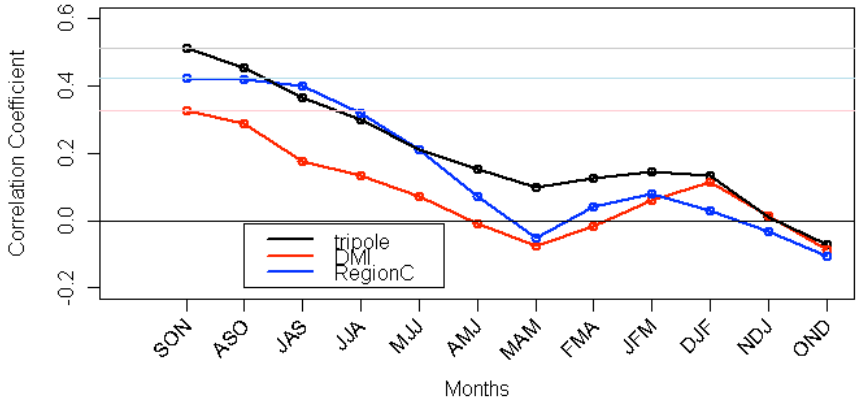

Figure 4.24. The correlation between SON MDB rainfall and three indices of Indian Ocean variability - Tri-pole index (black), the DMI (red, sign of the correlation reversed) and Region C (blue) - with increasing lead-time, over the period 1900 – 2009. The simultaneous correlation is shown at SON and the horizontal coloured lines indicate the simultaneous correlation level. Correlations exceeding +/-0.2 are significant at the 5% significance level.

The lagged correlations between upper-layer soil moisture and Indian Ocean variability are stronger than those for rainfall (Figure 4.25). The lagged correlations are strongest with the TPI

and Region C, particularly in northern Australia and Victoria, although the strength of the most significant correlations is reduced from the simultaneous level. The DMI again shows little lagged relationship with some significant correlations over southeast Australia, although the lagged relationship is stronger with upper-layer soil moisture than it was with rainfall. The MDB area-averaged (Figure 4.26) correlations again show the TPI and the DMI decaying at a steady rate until MAM. Region C shows some increases in correlation over the preceding two months (ASO and JAS) before decaying to zero in MAM like the DMI.

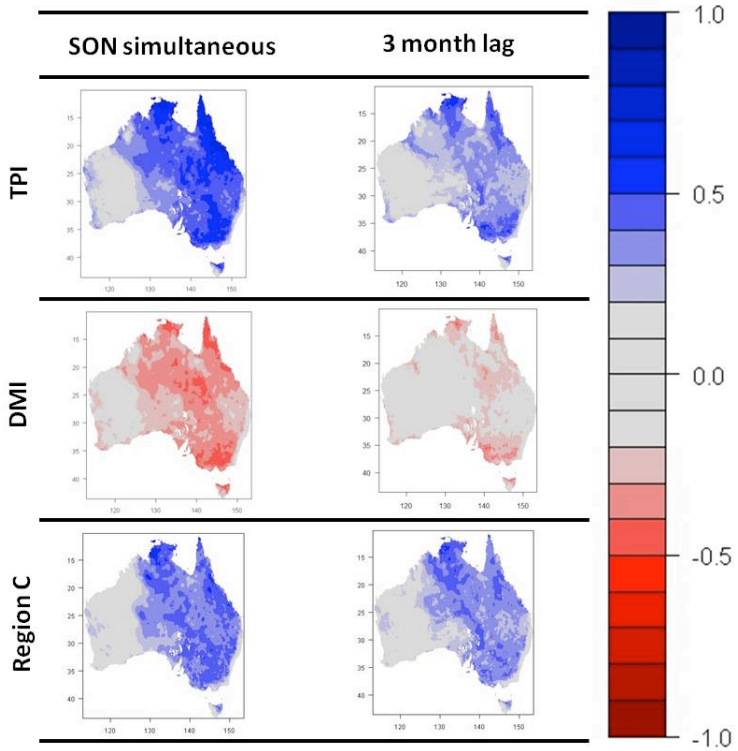

Figure 4.25. The simultaneous correlation between gridded SON upper-layer soil moisture and three indices of Indian Ocean variability - Tri-pole index (top), the DMI (middle) and Region C (bottom) - in SON (left) and the 'lagged' correlation between SON upper-layer soil moisture and the Indian Ocean indices in JJA (right, three-month lead-time), over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

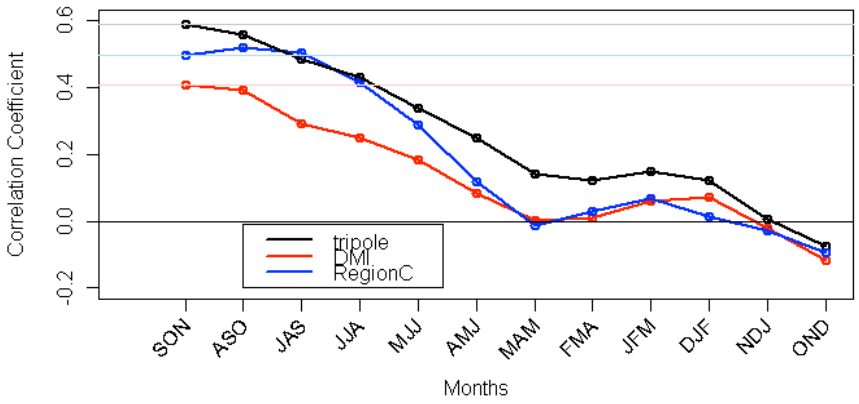

Figure 4.26. The correlation between SON MDB upper-layer soil moisture and three indices of Indian Ocean variability - Tri-pole index (black), the DMI (red, sign of the correlation reversed) and Region C (blue) - with increasing lead-time, over the period 1900 – 2009. The simultaneous correlation is shown at SON and the horizontal coloured lines indicate the simultaneous correlation level. Correlations exceeding +/-0.2 are significant at the 5% significance level.

The gridded and area-average correlation for lower-layer soil moisture (Figure 4.27 and Figure 4.28) show that the season lagged correlations for the TPI are of a similar spatial extent to the simultaneous correlation but appear to be of increased magnitude, particularly in southeast Australia. Similarly, while the DMI decreases spatial extent it appears to increase in magnitude across southern Australia. The correlations with Region C in southeast Australia have increased in both extent and magnitude. For all three indices of Indian Ocean variability, there is no relationship with the eastern coastal strip. The area-averaged MDB lagged correlations highlight the strong lagged relationships between the TPI and Region C. Correlations with both indices increase in strength from the simultaneous correlation until MJJ. The simultaneous correlation with lower-layer soil moisture in the MDB with the TPI is r = 0.48, this increases to a maximum in JJA (r = 0.53). Region C increases from the simultaneous correlation (r = 0.41) to the maximum of r = 0.51 in JAS. The DMI does not increases from its simultaneous maximum correlation of r = 0.40 and although it remains reasonably steady for the preceding season (ASO) it decays quickly. This shows that the TPI and Region C have the strongest lagged relationships with lower-layer soil moisture both across eastern Australia and in the MDB specifically.

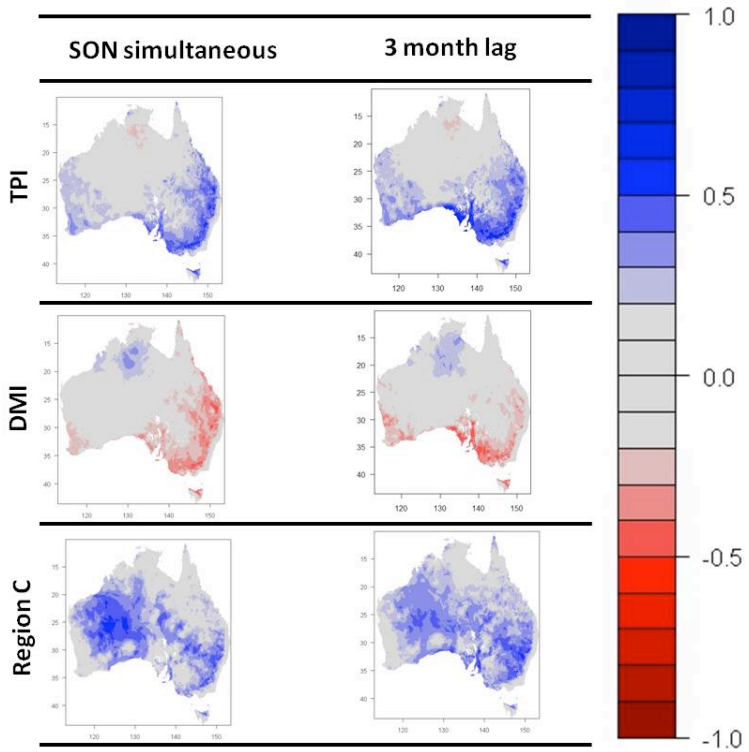

Figure 4.27. The simultaneous correlation between gridded SON lower-layer soil moisture and three indices of Indian Ocean variability - Tri-pole index (top), the DMI (middle) and Region C (bottom) - in SON (left) and the 'lagged' correlation between SON lower-layer soil moisture and the Indian Ocean indices in JJA (right, three-month lead-time), over the period 1900 – 2009. Coloured regions indicate correlations that are significant at the 5% significance level.

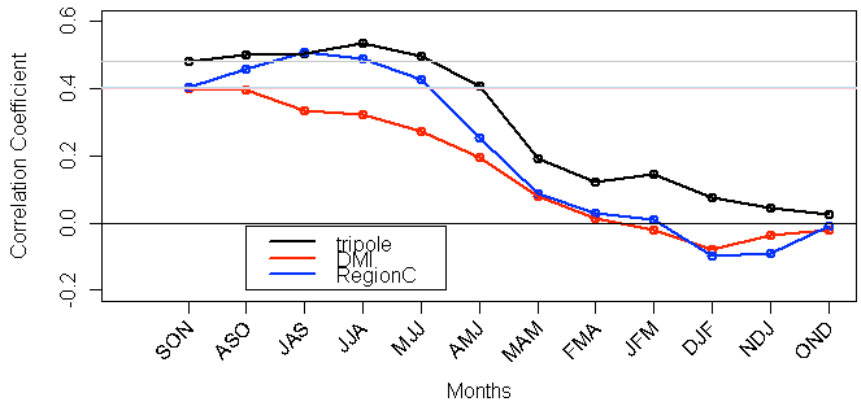

Figure 4.28. The correlation between SON MDB lower-layer soil moisture and three indices of Indian Ocean variability - Tri-pole index (black), the DMI (red, sign of the correlation reversed) and Region C (blue) - with increasing lead-time, over the period 1900 – 2009. The simultaneous correlation is shown at SON and the horizontal coloured lines indicate the simultaneous correlation level. Correlations exceeding +/-0.2 are significant at the 5% significance level.

This section has outlined the seasonal variability in the relationship with Indian Ocean indices and hydroclimatic variability in Australia. In particular, differences between in the magnitude, spatial extent and sign of the relationship were noted between the warm and cool seasons for rainfall and upper-layer soil moisture, while the relationship with lower-layer soil moisture remained in the same sign. Lagged correlations were present for rainfall and soil moisture, likely stemming from the persistence of Indian Ocean variability from JJA to SON. The next section (4.4.4) will discuss linear relationships between indices of large-scale variability.

4.4.4 Linear relationships between predictor variables

The relationships between indices of large-scale modes of climate variability vary seasonally (Figure 4.29). As expected the highest correlations are found between indices of ENSO in all seasons; indices of Indian Ocean variability are also highly correlated, particularly the DMI with each of the poles (IOD-E and IOD-W) and IOD-E with the NWS region due to their proximity. The TPI is more strongly correlated with the DMI than with indices of ENSO. The facets of the STR are highly correlated in JJA and SON, with some relationship in MAM and little in DJF. Indian and Pacific Ocean indices are mostly highly correlated in SON with less relationship in JJA.

This section has shown that there are strong linear relationships between indices used to describe all large-scale climate drivers in most seasons. These strong linear relationships between the indices that are used to characterise large-scale modes of climate variability can

obscure more subtle interactions. The following section (4.5) will summarise the conclusions from this chapter.

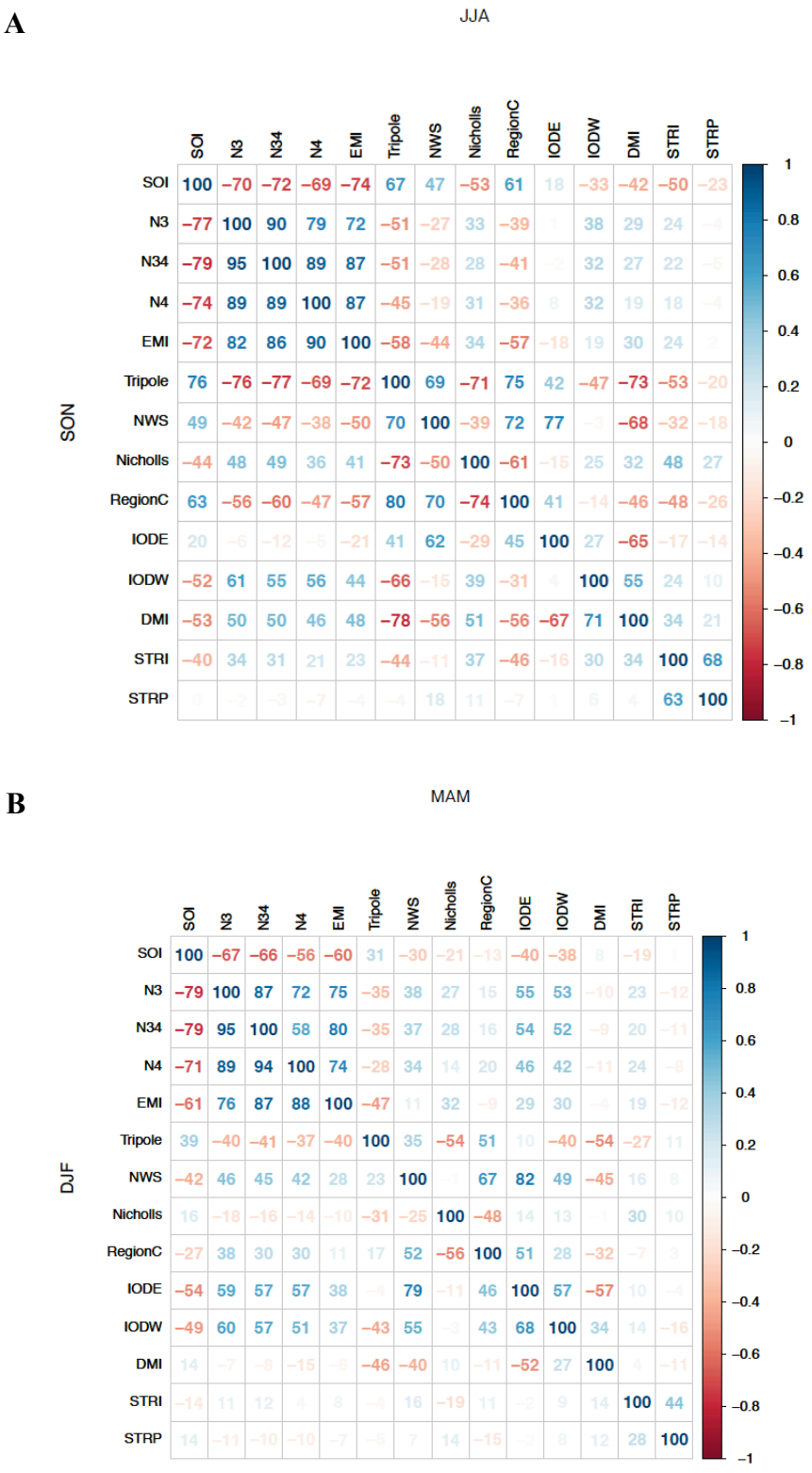

Figure 4.29. The correlation coefficients as percentages (r * 100) between indices of large-scale modes of climate variability (see Table 4.1 in JJA (A – upper triangle), SON (A – lower triangle), MAM (B – upper triangle) and DJF (B – lower triangle), over the period 1900 – 2009.

4.5 Conclusions

This chapter has outlined the linear relationships between indices of remote (i.e. ENSO and IOD) and regional (i.e. STR) large-scale variability and Australian hydroclimates. Although all relationships are consistent with previous research (Chowdhury and Beecham 2013, Drosdowsky and Chambers 2001, McBride and Nicholls 1983, Nicholls 1989, Risbey et al. 2009, Timbal and Drosdowsky 2013, Timbal and Hendon 2011, Ummenhofer et al. 2011), it is important to establish the nature of the linear relationship using the relatively new AWAP data set and the indices as defined in Chapter 1. This validated the novel AWAP data set and provided a foundation for more complex analyses in later chapters.

The relationship between facets of the STR (i.e. position and intensity) and Australian rainfall follows a distinctive seasonal cycle, peaking in JJA. In all seasons, the correlation between hydroclimatic variability and STR intensity is stronger than the correlation with STR position. Increased STR intensity is generally associated with below average rainfall, but there are some regions and seasons when the relationship changes sign. This is also true for the relationship with STR position, where the sign of the relationship depends on the mean position of the STR. The mean position of the STR changes seasonally and is thought to have changed over the instrumental record as the tropics have broadened (Seidel et al. 2007). There is very little persistence in correlations between rainfall and facets of the STR at increasing lead-time, likely due to the high-frequency nature of the atmospheric variability. In contrast, there are significant relationships between soil moisture (upper and lower) and indices of STR variability at a one-season lead-time. The lack of significant relationships between rainfall and the STR at one-season lead-time, suggests that hydrological processes are responsible for the significant correlations with soil moisture and lagged indices of the STR.

The relationship between the tropical modes of variability and Australian hydroclimates also adheres to a seasonal cycle. The relationship with ENSO peaks in SON but is also significant and widespread over eastern Australia in JJA and DJF. El Niño events are associated with reduced rainfall and soil moisture across eastern and northern Australia. The persistence of ENSO from onset in late MAM or early JJA translates into significant correlations at the one-season lead-time for SON rainfall, upper-layer and lower-layer soil moisture. Indian Ocean variability behaves in a similar way, with the relationship with rainfall and soil moisture peaking in JJA and SON, with the significant correlations focused in southern (eastern) Australia in JJA (SON). A positive IOD event, or cooler SST anomalies in the maritime continent region is associated with reduced rainfall in southern Australia. The strength of the

relationship between indices of Indian Ocean variability and soil moisture increases with increasing lead-time.

These strong linear relationships are the foundation of our understanding of Australian hydroclimatic variability (Drosdowsky and Chambers 2001, McBride and Nicholls 1983, Risbey et al. 2009) and they have formed the basis of seasonal forecasts in Australia. The persistence of the tropical modes from onset underpins the lagged relationships that are exploited in seasonal forecasting (Drosdowsky and Chambers 2001).This chapter has shown that lower-layer soil moisture has significant linear relationships at lead-times of up to several months, often exceeding the lagged linear relationship with rainfall. The potential increased predictability of lower-layer soil moisture over rainfall will be discussed further in Chapter 7.

These strong linear relationships have been valuable for increasing understanding of the drivers of Australian climate yet they don't tell the whole story. Strong correlations between large-scale modes of climate variability and Australian rainfall, in the range of $r \approx 0.5 - 0.7$, are reported here and elsewhere in the literature (McBride and Nicholls 1983), but these linear relationships only explain 30 – 40% of the variability. This leaves a majority of the variability unexplained, and while stochastic noise is an important factor, it is likely that non-linear interaction between various climate drivers plays a large role. In addition, it is likely that the strength of the correlations between large-scale modes of climate variability mask subtleties in the relationships between Australian hydroclimatic variability and large-scale climate drivers.

This chapter has outlined the seasonal and spatial linear relationships between regional and remote large-scale modes of climate variability and Australian hydroclimates, which subsequent chapters will build upon. The following three chapters (Chapters 5 – 7) will use tree-based methods to further explore nuances in the drivers of Australian climate and hydrology.

Chapter 5 : Interactions between tropical and sub-tropical climate modes of climate variability

5.1 Chapter summary and context

The previous chapters have provided an introduction to Australian climate variability and the large-scale modes of climate variability (or 'climate drivers') examined throughout this thesis (Chapter 2), followed by an outline of the methods (Chapter 1) and a discussion of the linear relationships between large-scale modes of climate variability and Australian hydroclimatic variability (Chapter 4). The purpose of Chapter 5 is twofold. Firstly, the chapter will introduce the reader to the novel statistical method, Classification and Regression Trees (CART). Several key features of the CART method will be discussed, including the differences between trees constructed from categorical and continuous data (Section 5.3.2.a), node purity (Section 5.3.2.b), the impact of tree size (Section 5.3.2.c), interactions between predictor variables (Section 5.3.2.d) and how composite analysis is used in this thesis (Section 5.3.2.e).

Secondly, the spatial and seasonal variability of relationships between tropical and sub-tropical predictor variables and regional rainfall is discussed and the most important predictor variables used to classify rainfall regimes are identified (Section 5.4.1). Following this broad outline, a case study methodology is adopted to explore interactions between predictor variables that influence regional Australian hydroclimatic regimes. The tree-based models are able to give more nuanced information about the discrete events that is not available from a linear analysis. This additional insight shows that the strength of the tropical teleconnection is key to transmitting the impacts of tropical variability to Australian rainfall and that interactions between tropical modes of variability and the STR can modulate the impacts of tropical variability. This chapter introduces the importance of STR variability, with the focus here on STR intensity. Further discussion of interactions between facets of STR variability (i.e. position and intensity) is found in Chapter 6.

5.2 Introduction

This chapter will introduce CART with a description of the method and a thorough examination of the relationship between indices of tropical and sub-tropical modes of climate variability and hydroclimatic regimes across Australia. This is a multi-dimensional and conceptually difficult problem. The relationship between tropical climate drivers (such as

ENSO and the IOD) and regional phenomena with Australian hydroclimatic variability was outlined in Chapter 2. There is limited research into the interactions between tropical and sub-tropical climate drivers that influence regional Australian hydroclimatic regimes, although it has been identified that phases changes between ENSO, the IOD and the Southern Annual Mode are associated with changing frequency of wet and dry synoptic types (Verdon-Kidd and Kiem 2009b).

This chapter begins broadly with an introduction to the CART method. Firstly in Section 5.3.2 examples are presented that show various features of the tree-based method; a comparison of Classification and Regression Trees is presented along with definitions of CART terminology (node purity, tree structure, terminal nodes, split variables and predictor variable importance) and a discussion on how tree size influences response variable regimes, interactions between predictor variables composite analysis.

Next, in Section 5.4.1, is an analysis of rainfall in 12 spatially aggregated regions across the Australian continent (see Chapter 1 for a discussion of the regions) in the four traditional seasons (MAM, JJA, SON, DJF). The spatial and temporal variability in tree-based models classified from all predictor variables are discussed.

Finally, in Section 5.4.2, a pairwise analysis of predictor variables is presented to discuss interactions between predictor variables that influence Australian rainfall regimes. A case study approach is taken to explore the additional information that can be gained by an analysis with tree-based models. The focus is on interactions between tropical (ENSO and IOD) and sub-tropical (STR) climate drivers in JJA and SON.

Tropical – sub-tropical interactions are the focus of this chapter, and Chapter 2 outlined the large-scale features that influence Australian climate variability, i.e. ENSO, the IOD and the STR. The strong linear relationships between these climate drivers (discussed in Chapters 2 and 4) make separating the relative influences and relevant interactions between climate drivers difficult.

The remainder of the chapter is structured in the following way; Section 5.3 is a discussion of the data sets and methods used in this chapter, including an example based introduction to CART terminology and methodology. This is followed by a results section that examines the tree-based models defining regional Australian rainfall regimes from all predictor variables (Section 5.4.1) and the pairwise analysis of interactions between tropical – sub-tropical climate drivers that influence rainfall regimes (Section 5.4.2). Finally, discussion and conclusions are presented in Section 5.5.

5.3 Data sets and methods

A detailed discussion of the data sets and methods used throughout the thesis can be found in Chapter 3. This section will outline the specific data sets and methods used in this chapter.

5.3.1 Data sets

Section 5.4.1 classifies rainfall in all 12 regions (Figure 5.1 and Table 5.1) from all 14 predictor variables defined in Chapter 3, and shown below in Table 5.2.

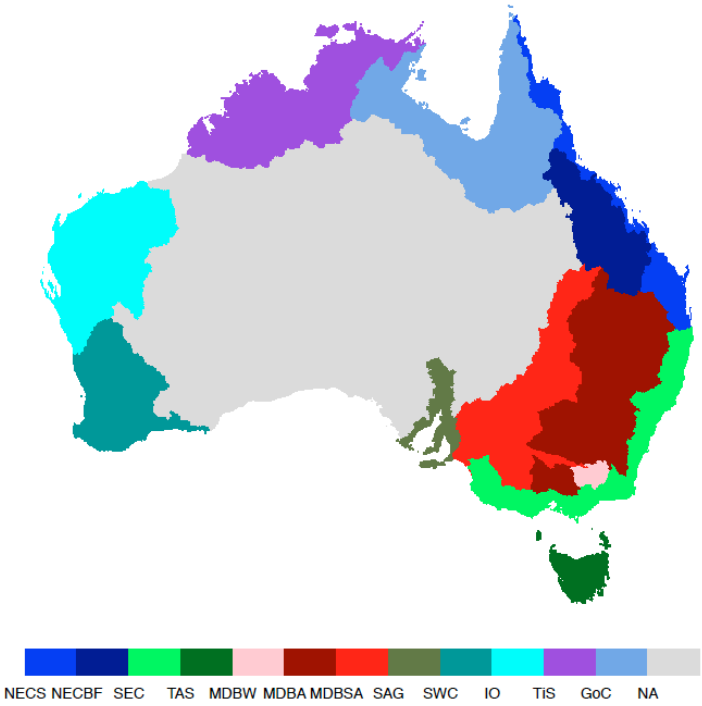

Figure 5.1. The 12 Australian drainage divisions (regions) that are area-average and used as response variables in this Chapter. See Table 5.1 for region names.

Table 5.1. Names, abbreviations for each of the 12 regions (drainage divisions) used as response variables. Region abbreviations here correspond to the regions shown in Figure 5.1.

ion (Abbreviation)	Region Number
S)	1
ntaria (GoC)	2
st – Sea (NECS)	3a
st – Burdekin/Fitzroy (NECBF)	3b
g Basin Agricultural (MDBA)	4a
g Basin Semi-Arid (MDBSA)	4b
g Basin Wet (MDBW)	4c
st (SEC)	5
s)	6
ian Gulf (SAG)	7
ast (SWC)	8
(IO)	9

Table 5.2. Indices of large-scale modes of climate variability used as predictor variables in Chapter 5. See Chapter 3 for definitions of each index.

ame	Group	Reference
Niño 3	Pacific Ocean	(Trenberth 1997)
Niño 3.4	Pacific Ocean	(Trenberth 1997)
Niño 4	Pacific Ocean	(Trenberth 1997)
El Niño Modoki Index (EMI)	Pacific Ocean	(Ashok et al. 2007)
Southern Oscillation Index (SOI)	Pacific Ocean	(Troup 1965)
Tri-pole Index (TPI)	Pacific and Indian Ocean	(Timbal and Hendon 2011)
Region C	Indian Ocean	(Cai and Cowan 2008a)
Nicholls' SST gradient	Indian Ocean	(Nicholls 1989)
Northwest Shelf (NWS)	Indian Ocean	(Timbal and Murphy 2007)
Dipole Mode Index (DMI)	Indian Ocean	(Saji et al. 1999)
DMI – Eastern Pole	Indian Ocean	(Saji et al. 1999)
DMI – Western Pole	Indian Ocean	(Saji et al. 1999)
STR – Intensity	STR	(Drosdowsky 2005, Timbal and Drosdowsky 2013)
STR – Position	STR	(Drosdowsky 2005, Timbal and Drosdowsky 2013)

In Section 5.4.2 a subset of predictor variables (i.e. Niño 3.4, DMI and STR-I), representing the major climate drivers (ENSO, IOD and STR, respectively), are used to explore interactions between tropical (ENSO, IOD) and sub-tropical (STR) climate drivers and regional rainfall regimes. Fewer response variables are also used, with the focus on rainfall in MDBSA in JJA, SAG in JJA and SEC in SON.

In Section 5.3.2, the major features of the tree-based models are illustrated with JJA rainfall in the SEC, where possible. However, additional response variables are selected from the 288 tree-based models constructed; 3 response variables (rainfall, upper-layer and lower-layer soil moisture) x 4 seasons (MAM, JJA, SON, DJF) x 12 regions (Figure 5.1 and Table 5.1) x 2 tree structures (classification and regression trees). See Appendix A and B on the accompanying CD for additional tree-based models and composite anomaly patterns, respectively. The JJA season is chosen to highlight influences between tropical and sub-tropical climate drivers, because this is the season when both tropical modes and the STR influence Australian rainfall (Risbey et al. 2009, Timbal and Drosdowsky 2013, Whan et al. 2013).

In this chapter the reanalysis product is used to examine the atmospheric response for climate regimes is the 20th Century Reanalysis Product (Compo et al. 2011). Various atmospheric fields were averaged across the years classified into each regime, including 850 and 200 hPa geopotential height anomalies, 500 hPa vertical velocity and 1000 hPa wind speed and direction.

5.3.2 Methods

Classification trees are used in all sections of this analysis, except for a comparison with regression trees in Section 5.3.2.a. All default control arguments for the 'rpart' function outlined in Chapter 3 (Table 3.3.2) are used except for the 'minsplit' and 'maxdepth' arguments. Minsplit controls the minimum number of observations in a node for a split to be attempted. This is changed from the default value (i.e. 20) to 10. This means that additional splits will be attempted, compared to the default value, if the node has between 10 and 20 cases. In addition, the 'maxdepth' argument controls the size of the tree. The use of cross-validation to select the optimum tree size is a feature of CART that is generally not implemented in this chapter as tree-size is set to maintain consistency and interpretability of trees. Tree-size is varied in Section 5.3.2.c to explore the influence of different tree sizes. Maxdepth is generally set at three so that interpretability of interactions between predictor variables is maintained. 'Variable importance' is defined from the first split, as the predictor variable with the strongest relationship with the response variable is chosen for the first split.

The remainder of this section (Sections 5.3.2.a - 5.3.2.e) is comprised of a discussion of the major features of CART. Firstly, this section will outline the differences and similarities between classification and regression trees, justifying the selection of classification trees for further analysis. Second, selected tree-based models are used to discuss the major features of the method. Features of the CART method include node purity, tree structure and how it relates to linear relationships, the information that can be gained from the years classified into each terminal node and both hydroclimatic and atmospheric composites of the regimes.

5.3.2.a Tree structure – classification and regression trees

CART is comprised of two tree-based methods, classification and regression trees, that are used when the response variable is categorical or continuous, respectively. This section outlines some of the similarities and differences between regression trees constructed from continuous area-averaged rainfall and classification trees constructed from terciles of the same rainfall time series. JJA rainfall from the SEC will primarily be used in this section, with some additional examples from MDBA and the SWC.

The degree of similarity between corresponding classification and regression trees varies greatly. In some cases the classification and regression trees are very similar, e.g. upper-layer soil moisture regimes in MDBA in JJA are both classified from variability in the Tripole index (Figure 5.2). In others they are very different, for example, the classification tree for MDBSA in JJA requires information from both the Tripole index and the SOI, while the regression uses Region C and STR-I to define rainfall regimes (Figure 5.2). A full description of classification trees is given in Section 3.3.2. In Figure 5.2a, the colours (red, green, blue) of the nodes represent the regime ('dry', 'average' or 'wet', respectively) that the node has been assigned to by the 'majority rule'. For example, the 'wet' regime in this tree contains 1 dry, 3 average and 18 wet years so the terminal node (i.e. a node with no further splits) is classified as a 'wet' regime.

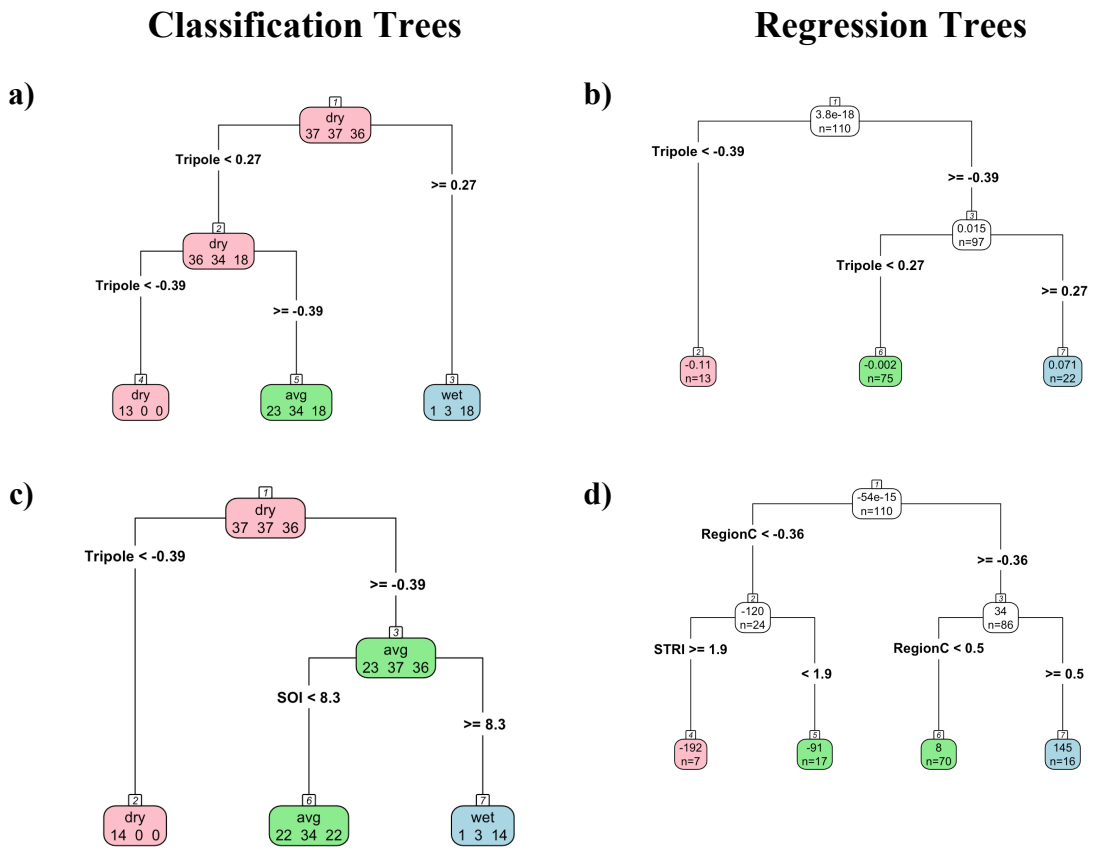

Figure 5.2. The classification (a and c) and regression (b and d) trees defining JJA MDBA upper-layer soil moisture regimes (a and b) and JJA MDBSA rainfall regimes (c and d), from a suite of 14 predictor variables (Table 5.2), over the period 1900 – 2009. The number of years classified into the dry (red), average (green) and wet (blue) regimes is marked (classification trees only). The number of years classified into the regime and mean of these years are given in the leaves (regression trees only).

Differences between the classification and regression trees highlight the instability of the tree-based methods (Hastie et al. 2001) associated with their propensity to overfit data. Differences between the classification and regression trees stem solely from the aggregation of continuous response variables into terciles with the top, middle and bottom thirds categorised as wet, average and dry years, respectively (Figure 5.3). The degree of similarity between the two trees is largely dependent on intrinsic characteristics of the time-series.

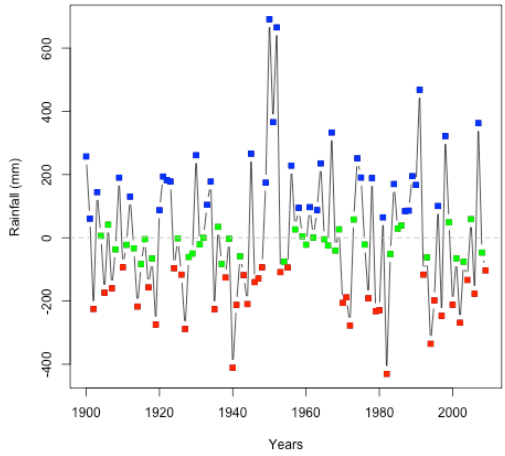

Figure 5.3. The time series of JJA SEC rainfall anomalies over the period 1900 – 2009, with years assigned to the dry (red), average (green) and wet (blue) categories marked.

Despite the differences in the predictor variables chosen to define regimes, the years in each node are often similar. This shows that while the trees may choose to define hydroclimatic regimes from different predictor variables, the strong relationships between indices of the same or different climate drivers means that similar teleconnections are being drawn upon. For example, the classification tree defining SEC JJA rainfall regimes (Figure 5.4) requires information from STR-I and Niño 4. On the other hand, the regression tree (Figure 5.4) defines rainfall regimes from variability in the SOI and STR-I. The "terminal nodes" (i.e. the nodes of the tree that have no further divisions) of both trees that are classified as "dry" (the red terminal nodes in Figure 5.4) are defined during strong El Niño conditions (Niño 4 > 0.47°C for the classification tree and SOI < -6.7 for the regression tree) that occur when the STR is neutral or strong (STR-I > -1.5 for the classification tree or STR-I > 1.9 for the regression tree). The

regression tree uses neutral or positive SOI and weak STR-I to define a wet terminal node (the regression tree blue terminal node in Figure 5.4), while the classification tree uses only weak STR-I to define the wet regime (the classification tree blue terminal node in Figure 5.4). Despite differences in the details of the tree structures, similar years are classified from similar relationships with indices of large-scale variability. Both trees built to predict categorical and continuous JJA rainfall in SEC have wet and dry regimes with similar years due to the broadly similar tree structures; in fact the wet terminal nodes include the same 19 years.

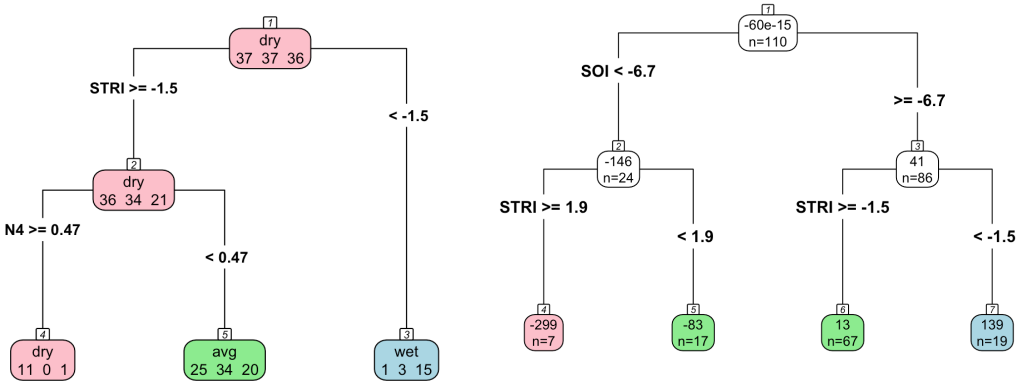

Figure 5.4 The classification (left) and regression (right) trees defining JJA SEC rainfall regimes, from a suite of 14 predictor variables (Table 5.2), over the period 1900 – 2009. The number of years classified into the dry (red), average (green) and wet (blue) regimes is marked (classification tree only). The number of years classified into the regime and mean of these years are given in the leaves (regression trees only).

Terminal nodes in classification trees are classified into one of the original terciles of the response variable (i.e. dry, average, wet) according to the 'majority rule' that assigns terminal nodes to the group with the most members present (Firth et al. 2005). Terminal nodes in regression trees are assigned to the mean of all years classified into the node. Here the terminal nodes with the largest (smallest) rainfall anomalies are called the "wet" ("dry") regimes. In Figure 5.4 the regression tree classifies 19 years into the wet regime and these 19 years have a mean rainfall anomaly of 139 mm. Information about the root mean square error (RMSE) of each terminal node can be extracted that shows the range of scores is smallest in the dry regime and largest in one of the average regimes. The assignment of terminal nodes into predetermined classes, in a classification tree, facilitates model evaluation as 'skill' is defined by how successfully the tree has separated homogeneous regimes.

This section has discussed the differences between classification and regression trees (i.e. response variable structure) and shown that the two methods can give similar results despite differences in the details of trees. Previous research has established that tree-based methods can

be unstable and are sensitive to small changes in the data set (Hastie et al. 2001); here we show that differences in data format (i.e. continuous and categorical) can result in different tree structures, but often with similar years assigned to terminal nodes. This section has shown that changing the format of the response variable does affect the characteristics of the final tree, but that the trees can be broadly similar despite these differences. The choice of tree structure also depends on the purpose of modelling. If the model is attempting to predict the amount of precipitation then a regression tree needs to be used. Yet classification trees can be used to increase understanding of rainfall regimes (i.e. dry, medium or wet). Classification trees are chosen to explore interactions between indices of large-scale variability and Australian hydroclimates because they are well suited to the event-based nature of rainfall and the large-scale conditions that make rainfall more or less probable.

5.3.2.b Node purity

'Node purity' it is the criterion against which 'splits' (or divisions in the data) are evaluated, either though the Gini Index in classification trees or through RMSE for regression trees. The definition of node purity, particularly for classification trees, is reasonably straightforward and this section will discuss how it is used in this thesis and the type of information that can be drawn from it.

CART identifies the splits that minimise the mean node impurity, as measured by the Gini Index. Splits that are chosen based on the Gini Index so that node impurity is decreased, but this does not always result in "useful" terminal nodes (i.e. nodes with no further splits) that are able add insight about hydroclimatic regimes. However, we can assess how useful the chosen splits are by looking at the node purity. For example, DJF rainfall in the NECBF region (Figure 5.5) is classified solely on indices of SST anomaly from the eastern (IODE) and western (IODW) Indian Ocean. Terminal Node 3 is a wet regime defined solely on eastern Indian Ocean SST variability (IODE < -0.34°C) with 82% purity –14 out of 17 years are classified as 'wet'. The second split on SST anomalies in the western Indian Ocean (IODW < 0.44°C) increases the purity of the tree by classifying a highly pure average terminal node, while also defining a dry terminal node with low purity (41% - 35 dry, 28 average and 22 wet years). It is unlikely that composite analysis of the atmospheric response in heterogeneous terminal nodes will be able to provide additional insight when the signal to noise ratio is so small.

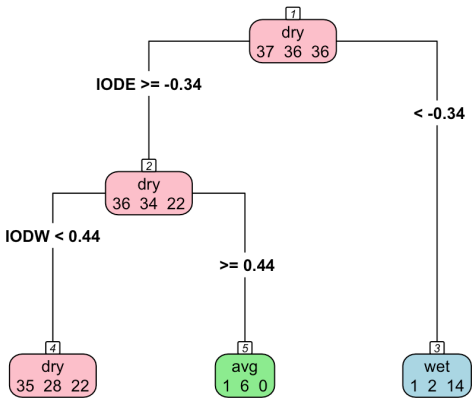

Figure 5.5. The tree classifying DJF rainfall regimes in the NECBF, from a suite of 14 predictor variables, over the period 1900 – 2009. The number of years classified into the dry (red), average (green) and wet (blue) regimes is marked.

This example shows that CART does not automatically classify homogeneous hydroclimatic regimes that are useful for further analysis of the associated causal mechanisms. Throughout this analysis node purity is considered high, moderate or low when nodes are > 80%, between 50 and 80% and < 50% pure, respectively. CART is a statistical method that follows certain rules; it can be a useful tool but should not be used as a "black box".

5.3.2.c Tree structure – tree size

Here we compare the effect of tree size on the classification of SEC rainfall regimes in JJA. "Tree size" refers jointly to the number of splits and the number of terminal nodes in a tree. The issue of tree size is closely related to the issue of when to assign "terminal" status to a node, i.e. no further splits. Several factors may lead to the assignment of a terminal node, including when the node comprises only one class of response variable (i.e. no splits make it more pure), if the node has reached the minimum size so that further splitting would result in nodes that were too small for analysis (the minimum size is 10 cases in this analysis), if there are no splits that increase node purity or if additional splits obscure information that can be drawn from the tree (Firth et al. 2005).

CART models are generally constructed through the growth of an overly-large tree that is then pruned back to the desired size (De'ath and Fabricius 2000). Generally, dividing until no further splits are possible grows a 'maximal tree'. The maximal tree (given the control arguments used in the analysis, see Table 3.4) for the SEC rainfall in JJA example is shown in Figure 5.6 (with the minsplit criterion set to the default value). CART uses cross-validation and a 'complexity parameter' to estimate the optimal tree size that minimises the relative cross-

validated error (Figure 5.7). In this example, the optimum tree size is three (as this tree size has the minimum cross validated error), with the tree pruned to this size shown in Figure 5.6.

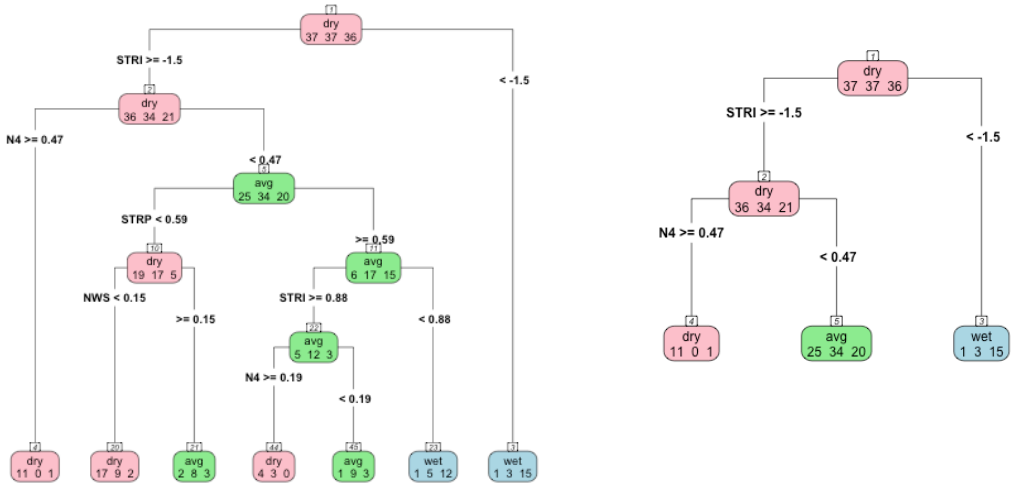

Figure 5.6. Two trees classifying JJA rainfall regimes in the SEC, from a suite of 14 predictor variables, over the period 1900 – 2009. The maximum tree (given the function control arguments outlined in Chapter 3) is shown on the left and the pruned tree is on the right. The number of years classified into the dry (red), average (green) and wet (blue) regimes is marked.

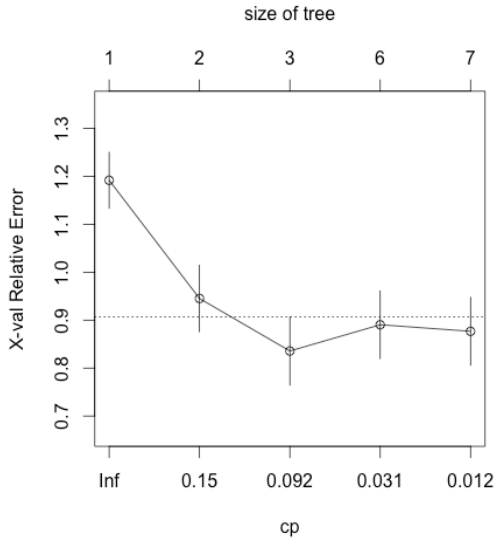

Figure 5.7. The relative cross-validated error for the trees of various sizes (1 to 7 split levels) classifying JJA SEC rainfall regimes, from a suite of 14 predictor variables, over the period 1900 – 2009. The complexity parameter ('cp') for each tree size is marked.

Tree size can also be controlled directly. This section compares trees that classify SEC rainfall in JJA from all 14 predictor variables; the small classification tree in Figure 5.4 with only three terminal nodes is compared with two larger trees that have four and nine terminal nodes (Figure 5.8). In this example, increasing tree size does not alter the extreme wet and dry

regimes classified in the smallest tree; no further classification of these regimes occurs in the larger trees because additional division of the years in these highly pure nodes does not result in a large increase in node purity. The larger trees, however, have additional division of the impure average terminal node that contained 25 dry, 34 average and 20 wet years (Figure 5.4). In this example, the smallest tree with three terminal nodes (Figure 5.4) and the tree with four terminal nodes (Figure 5.8) are very easily interpretable, as they define rainfall regimes from two or three different predictor variables. This example demonstrates that tree interpretability can be maintained when tree size is increased. However, the largest tree (with nine terminal nodes) shows that increasing tree size can inhibit interpretation of climatic regimes, while adding few insights into physical mechanisms. Additionally, it is important to avoid building trees that are large and over-fitting the data as this limits the generalisability of conclusions, in addition to inhibiting model interpretation. Overly complicated trees with terminal nodes that either remain relatively impure or have too few observations add little to our understanding of climatic regimes. Information about subsequent splits and the further refinement of response variable regimes is lost when using a smaller, simpler tree, but interpretability and the capacity to explore the physical mechanisms between predictor variables are gained. This is why the smaller tree is favoured in the remainder of this thesis.

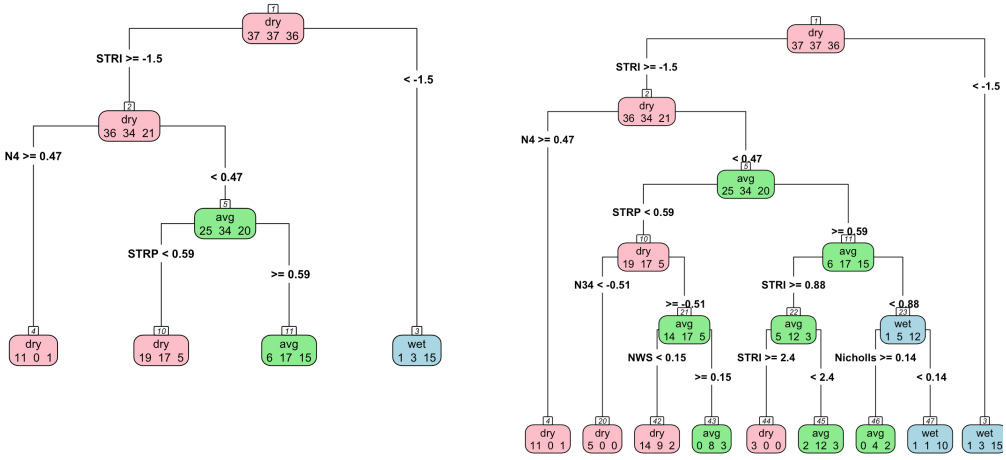

Figure 5.8. Two trees with three (left) and five (right) split levels, classifying JJA rainfall regimes in the SEC, from a suite of 14 predictor variables, over the period 1900 – 2009. The number of years classified into the dry (red), average (green) and wet (blue) regimes is marked.

This section has outlined two methods that can be used for controlling tree size, firstly, through use of a complexity parameter that minimises relative cross-validated error, and secondly, by directly limiting the size of the tree. For the remainder of this chapter, tree size is controlled directly and the smaller trees are favoured for the advantages highlighted above.

5.3.2.d Tree structure – interactions between predictor variables

This section discusses some characteristics of the predictor variables chosen to separate response variable regimes. Firstly, the predictor variable chosen for the first or primary split indicates which predictor variable has the strongest relationship with the response variable; this variable often, but not always, has the strongest correlation with the response data. This shows that CART can draw from the linear component of the signal and validates the use of a classical statistical framework to explore such relationships. CART is a "greedy algorithm" because it selects the split to minimise the mean node impurity at each step, with no regard for subsequent splits (Breiman et al. 1984). This means that subsequent splits are not dependent on preceding splits and that they in turn indicate predictor variable importance for their particular parent-regime.

Secondly, the recursive binary-partitioning nature of CART results in a somewhat clumsy handling of linear relationships as multiple splits on single predictor variables represent linear relationships. Figure 5.9 (left) is an example of a tree that classifies SON rainfall regimes in northeastern Australia (NECBF) from linear splits on the SOI. None of the 11 other predictor variables were selected to divide the response variable, suggesting that other predictor variables were not able to decrease node impurity as much as the SOI. In this example, the linear splits on the SOI reflect the known relationship with ENSO; when the SOI is negative (< -2.5) a moderately pure dry regime is classified (64% purity). A highly pure wet regime is defined when the SOI > 7 (80% purity). The tree structure and the high node purity when a more extreme split point is chosen (i.e. SOI > 7, compared to SOI < -2.5) is consistent with a linear relationship between ENSO and SON rainfall in NECBF, that is supported by a strong correlation ($r = 0.63$). This suggests a relatively large portion of variability in the response variable is explained by the linear relationship.

Finally, the combination of predictor variables that defines response variable regimes can potentially give insight into important interactions between those predictor variables. For

example, high values of STR-I define a dry regime for SON rainfall in TAS (Figure 5.9), consistent with the linear relationship (r = -0.61). If splitting ceases when criteria determining terminal node status have not been met (i.e. the tree has not reached maximum depth (controlled directly through the CART function), or there are more than the minimum number of observations in the node), this suggests that the node is terminal because no other splits increase mean node purity, as is the case for the dry regime for TAS rainfall (Figure 5.9). The second split draws information from the DMI and separates a wet regime and a second dry regime with low and moderate purity, respectively. Here we see that when STR-I intensity is weak (< 0.94 hPa) information about Indian Ocean variability is required to differentiate wet and dry rainfall regimes. It highlights the importance of the DMI, compared to ENSO, for this region in southern Australia. This is a relationship that is obscured by the linear correlation analysis as the two modes of variability have similar strength relationship with rainfall (DMI r = -0.30, SOI r = 0.31).

This section has outlined some typical behaviour of CART and their interpretation. It has shown that tree-based models can extract information that is suggestive of a the non-linear component of the observed variability, in addition to the linear component captured by the correlation analysis.

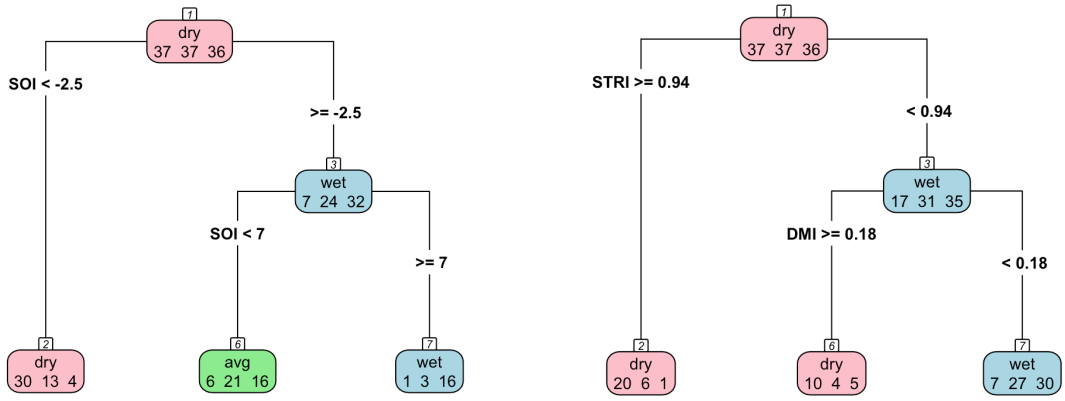

Figure 5.9. The trees classifying SON rainfall regimes in the NECBF (left) and TAS (right), from a suite of 14 predictor variables, over the period 1900 – 2009. The number of years classified into the dry (red), average (green) and wet (blue) regimes is marked.

5.3.2.e Composite analysis

Composite analysis of response variable regimes and the associated atmospheric conditions are a powerful part of a CART analysis. The years classified into each terminal node are averaged to characterise the hydroclimatic and atmospheric response to combinations of

predictor variables. Seasonally averaged atmospheric variables (anomalous 850 hPa and 200 hPa geopotential height, 500 hPa vertical velocity and 1000 hPa vector wind), extracted from the 20th Century Reanalysis Project (Compo et al. 2011), are averaged to indicate the likely causal mechanisms associated with response variable regimes.

Section 5.3.2 has outlined the major features of CART and shown how the method is implemented in this chapter. The following sections will discuss the tree-based models in further detail.

5.4 Results

The results of the classification trees that define rainfall regimes from all predictor variables are presented first (Section 5.4.1). This is followed by discussion of a series of case studies defining rainfall regimes from combinations of predictor variables to explore the influence of tropical – sub-tropical interactions on rainfall variability (Section 5.4.2).

5.4.1 Spatial and temporal variability of Australian rainfall regimes classified from all predictor variables

This section outlines the temporal and spatial features of rainfall regimes classified from a suite of 14 predictor variables. Seasonally and area-averaged rainfall regimes in each of the 12 regions were classified from a suite of 14 seasonally averaged indices of Pacific Ocean, Indian Ocean and sub-tropical ridge variability. The CART analysis shows many interesting features that confirm and extend our understanding of the relationships between indices of large-scale modes of climate variability and Australian rainfall, validating the use of the tree-based methods. The details of the trees for each region and season are shown in (Table 5.3). The structure of the table is related to the tree structures shown in the previous section. In the table, the first split (or primary split) is coloured green and labelled "1st". The subsequent second left and right splits in the trees are coloured black and labelled "L" and "R", respectively.

Firstly, the tree-based analysis shows that teleconnections with large-scale modes of variability are phase locked to the seasonal cycle, consistent with a linear analysis. The most important set of predictor variables in MAM are related to ENSO, as 7 of 12 regions draw on this mode for the primary split. The relationship between Australian rainfall and ENSO is not strong during MAM, indeed the average correlation between regional rainfall and indices of ENSO ranges from -0.12 (EMI) to 0.21 (SOI). While the relationship with ENSO is weak, it is stronger than with other predictor variables as splits based on ENSO result in the largest increase in mean node purity. The importance of the STR compared to ENSO increases during MAM with four regions splitting first on a facet of subtropical pressure variability. In JJA, when

the mean position of the STR is over the continent, the STR is the most important predictor variable: it is selected for the primary split for 6 of 12 regions located in southern Australia. ENSO dominates during SON with 9 of 12 regions splitting on an index of Pacific Ocean variability. In DJF, as the influence on ENSO is reduced and the STR is located to the south of Australia, no single mode of variability dominates. Indices related to ENSO and Indian Ocean variability are each relevant to 5 of 12 regions, with the STR chosen in the two remaining regions.

Secondly, the spatial distribution of variable importance is also consistent with known relationships with large-scale modes of variability. Indices of ENSO variability dominate in the northeast across all seasons. In the southeast indices of ENSO and STR variability are used to define rainfall regimes, while in the southwest indices of Indian Ocean variability appear in combination with ENSO or the STR. This summary has shown that spatially and temporally CART draws on known relationships between large-scale climate modes of variability and regional Australian rainfall. A further examination of the tree characteristics in each season follows.

Table 5.3. The trees classifying rainfall regimes in 12 Australian drainage divisions from a suite of 14 predictor variables, over the period 1900 – 2009, in MAM, JJA, SON, and DJF. The first predictor variable (labelled "1st", coloured green, preceding the backslash) is the primary split, with the left and right secondary splits following (labelled "L" or "R", black, bold). "x" indicates no further splits on that node.

Region	Division	MAM (1st: *first split* / L: *left split* – R: *right split*)	JJA (1st: *first split* / L: *left split* – R: *right split*)	SON (1st: *first split* / L: *left split* – R: *right split*)	DJF (1st: *first split* / L: *left split* – R: *right split*)
North	TiS	1st: DMI / L: Tripole – R: Nicholls	1st: Tripole / x - R: SOI	1st: SOI / L: Tripole - R:DMI	1st: SOI / L: EMI - R: SOI
	GoC	1st: SOI / L: RegionC – x	1st: RegionC / L: STRP - R: IODE	1st: SOI / x - SOI	1st: N4 / L: Nicholls - x
Northeast	NEC	1st: STRP / L: N4 - R: SOI	1st: SOI / L: STRP - R: STRP	1st: SOI / L: Nicholls - R: STRP	1st: NWS / L: SOI - x
	NECBF	1st: SOI / L: STRP – x	1st: EMI / L: RegionC - x	1st: SOI / x - R: SOI	1st: IODE / L: IODW - x
Southeast	SEC	1st: SOI / L: STRP – R: RegionC	1st: STRI / N4 - x	1st: SOI / x - R: STRP	1st: IODW / L: STRP - x
	MDBW	1st: STRP / L: DMI – x	1st: STRI / x - R: Nicholls	1st: RegionC / L: IODW - R: N4	1st: IODE / L: NWS - R: Tripole
	MDBA	1st: SOI / L: IODW – R: N3	1st: SOI / x - R: RegionC	1st: SOI / L: N4 - R: N3	1st: IODE / L: STRP - x
	MDBSA	1st: N34 / L: NWS – R: SOI	1st: Tripole / x - R: SOI	1st: Tripole / L: N4 - R: N3	1st: N34 / L: NWS - R: STRI
South	TAS	1st: STRI / L: N3 – R: N4	1st: STRI / x - R: RegionC	1st: STRI / x - R: DMI	1st: STRI / L: SOI - R: STRI
	SAG	1st: N34 / L: N3 – x	1st: STRI / L: NWS - R: RegionC	1st: Nicholls / L: Nicholls - R: IODW	1st: STRI / x - R: STRI
West	SWC	1st: N34 / L: IODW - R: IODW	1st: STRI / x - R: DMI	1st: N34 / x - R: N34	1st: N34 / x - R: NWS
	IO	1st: STRI / x – R: IODE	1st: STRI / L: STRI - R: RegionC	1st: Tripole / L: STRP - R: STRI	1st: SOI / L: N4 - x

Further examination of tree characteristics in SON (Table 5.3) reveals that the SOI is the most important of the ENSO indices, as it is chosen for 6 of 9 regions located in northern (TiS, GoC, NECS, NECBF) and southeastern (SEC, MDBA) Australia. This is consistent with the understanding using linear statistics, as the SOI is the most highly correlated predictor variable ($r \geq 0.49$) in these regions. Additionally, in SON, regions in southern Australia require information from the Indian Ocean or STR variability to define rainfall regimes; the primary split in SAG, TAS and MDBW is the Nicholls SST gradient, STR-I and Region C, respectively. The linear correlation may give an indication of the predictor variables chosen for splits, but CART draws upon additional information. For example, in MBDW there is a stronger correlation with SOI than with the predictor variable chosen for the primary split (Region C).

In JJA, the importance of STR variability in defining Australian rainfall regimes increases. STR-I is selected as the primary split in 6 of 12 regions in southern Australia (SEC, MDBW, TAS, SAG, SWC, IO). The ability of CART to extract additional information compared to the linear regression is again highlighted in the SEC. This region has a stronger correlation with ENSO indices than with STR-I, but low STR-I (<= -1.5 hPa) defines a wet regime with moderate purity (78%). The influence of ENSO on the SEC is incorporated in the second split, where Niño 4 is used to define a high purity (91%) dry node. Southern Australia (e.g. SAG, TAS) is known to have strong relationships with both the STR (Timbal and Drosdowsky 2013) and Indian Ocean (Nicholls 1989) variability in JJA, which is highlighted as indices of both climate drivers are required to define rainfall regimes in southern Australia. In both TAS and SAG, information about STR-I and Indian Ocean variability is needed to adequately classify rainfall regimes. The interaction between STR-I and indices of Indian Ocean variability can be seen in the SAG tree (Figure 5.10).

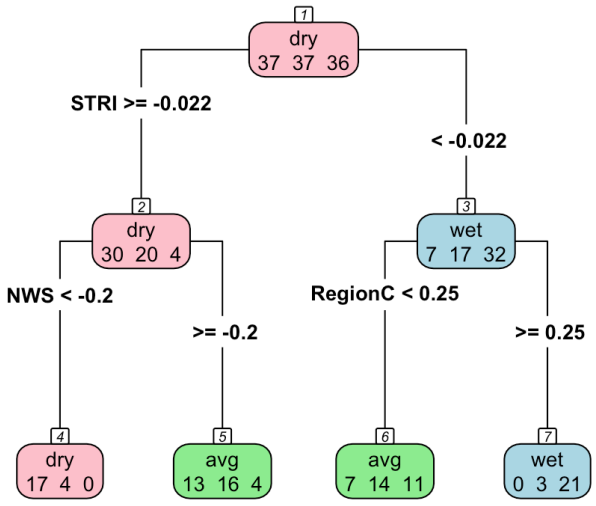

Figure 5.10. The tree classifying JJA rainfall regimes in the SAG, from a suite of 14 predictor variables, over the period 1900 – 2009. The number of years classified into the dry (red), average (green) and wet (blue) regimes is marked.

The interaction between sub-tropical and tropical modes of variability can be a useful predictor of wet and dry regimes. Figure 5.10 shows that a wet regime (88% purity) is defined when the intensity of the STR is low (< -0.02 hPa) and warm SST anomalies are located in the maritime continent (Region C >= 0.25°C). Conversely a strong STR (STR-I >= -0.02 hPa) and cool SST anomalies in the maritime continent region (NWS < 0.2°C) are associated with dry conditions (80% purity). Comparison of rainfall composites for the driest and the wettest regimes suggest non-linearities in the rainfall response over the southeast Australia and the coastal region (Figure 5.11). The negative rainfall anomaly in the dry regime is stronger than the corresponding positive anomaly in the wet regime despite the fact both regimes are defined from similar sized SST anomalies in the maritime continent, which could indicate that a stronger rainfall response is associated with more widespread SST anomalies in the NWS region. The positive (negative) 850 hPa geopotential height anomalies over southern Australia in dry (wet) regime are associated with a warmer (cooler) atmosphere and higher (lower) atmospheric pressure. Both regimes have geopotential height anomalies that extend through the atmosphere but are confined to southern Australia in the upper atmosphere (Figure 5.11).

Dry regime: Terminal Node 4	Average regime: Terminal Node 5	Average regime: Terminal Node 6	Wet regime: Terminal Node 7

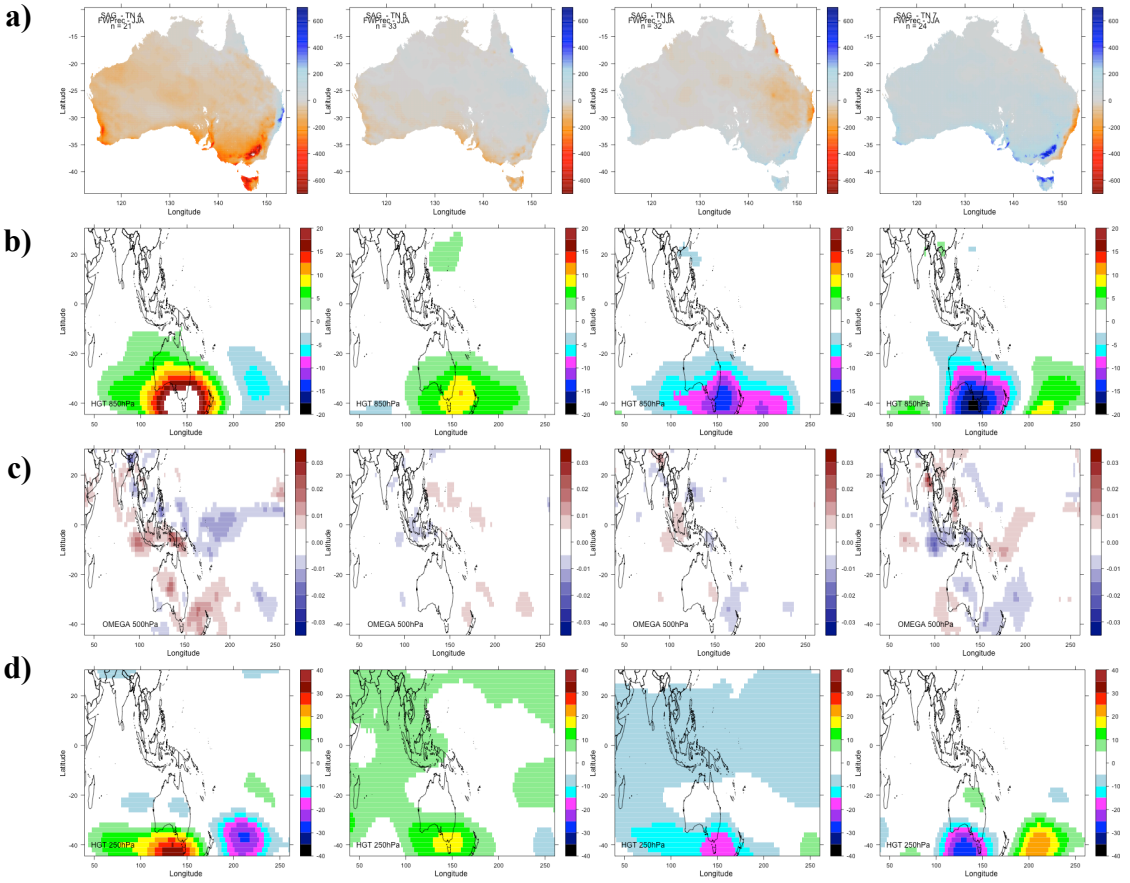

Figure 5.11. Spatial distribution of composite a) rainfall, b) 850 hPa geopotential height, c) 500 hPa vertical velocity and d) 200 hPa geopotential height anomalies in the four regimes defined when JJA SAG rainfall is classified from a suite of 14 predictor variables, over the period 1900 – 2009.

In DJF a wider range of predictor variables have relationships with rainfall as the dominance of ENSO in SON recedes; this is reflected in the larger number of indices that are selected for the first split. In northern Australia, indices of ENSO or maritime continent variability dominate, as the first split in TiS, GoC, NECS, NCEBF are SOI, Niño 4, NWS and IOD-E, respectively. In southeastern Australia, indices of Indian Ocean variability dominate (MDBW, MDBA = IOD-E, SEC = IOD-W) with only MDBSA classified from a primary split on Niño 3.4.

Correlations between predictor variables and regional rainfall in MAM are considerably lower with the strongest linear relationship in GoC and NEC with the SOI (r=0.43 and 0.35, respectively). In MAM, there is a less consistent signal from any particular mode of variability

as various indices of ENSO are selected in 7 of 12 regions. Despite little linear correlation, the CART method is able to produce skilful trees that predict wet and dry regimes with accuracy. However, the issue of over-fitting should be kept in mind (Hastie et al. 2001), as tree-based models that are fit to the stochastic variability are likely when there are limited relationships with predictor variables.

The tree-based models presented here give an indication of the maximum skill possible (i.e. the most homogeneous terminal nodes) with this data set and model parameters; subsequent analysis (Section 5.4.2) using a limited number of predictor variables will have decreased node purity. It has shown that predictor variable importance varies spatially and seasonally consistently with previous research and using traditional linear methods (Risbey et al. 2009, Timbal and Drosdowsky 2013). This section showed that although CART is generally consistent with the linear analysis that there are instances when the tree-based classification method behaves differently than the linear statistical methods, suggesting that CART extracts information stemming from the non-linear portion of the variability. This section has validated the use of tree-based in the classification of Australian hydroclimatic regimes, as the models are drawing on known relationships and capturing known sources of variability without the restriction to a linear statistical framework. From this foundation, a pairwise analysis of various combinations of predictor variables follows that will shed further light on the interactions and relationships between indices of large-scale climate variability that influence Australian hydroclimate.

5.4.2 Tropical – sub-tropical interactions

Following from the validation of the CART method in Section 5.4.1, this section focuses on several case studies to explore interactions between tropical (i.e. ENSO and the IOD) and sub-tropical (i.e. STR) large-scale climate drivers that influence regional rainfall regimes.

Section 5.4.1 showed that indices of ENSO were vital for the classification of rainfall variability, particularly in the north and southeast of the continent. Generally, the SOI was selected preferentially over Niño 3.4, but the oceanic index is used here because of the greater predictability and greater persistence of SST anomalies. Indian Ocean variability is represented by the DMI, which was crucial for the classification of rainfall regimes in southern and western Australia (Section 5.4.1). Here the STR is characterised by its intensity (STR-I) rather than the position, because STR-I has a stronger correlation with rainfall variability in southern Australia. The positive correlation between STR intensity and position should be kept in mind, as it suggests that as the STR intensifies as it moves further south (Timbal and Drosdowsky 2013).

Indices of remote and regional climate drivers are significantly correlated with each other and with rainfall, which can make characterising influences difficult. Moreover, these strong linear correlations can mask more subtle aspects of the relationship that may be important in defining climate regimes. The relationship between STR intensity and rainfall in JJA is strong in the south and east of the continent (SEC, MDBW, MDBA, MDBSA, SAG, TAS) with correlations ranging between r = -0.37 (SEC) to r = -0.69 (MDBW), indicating that increased STR intensity is associated with lower rainfall in southern Australia (Peixoto and Oort 1992). In addition, STR intensity is strongly correlated with indices of Indian and Pacific Ocean variability in this season through changes to the Hadley Circulation.

A comparison of classification trees constructed from two pairs of predictor variables, the ENSO and IOD models, in JJA and SON is presented (Table 5.4). The analysis includes an examination of classification trees together with composite rainfall anomalies and atmospheric variables such as upper and lower troposphere geopotential height anomalies, mid-troposphere vertical velocity and lower troposphere vector wind anomalies.

Table 5.4. The predictor variables used in the ENSO and IOD models to classify regional rainfall regimes in JJA and SON.

Model	Predictor variables
"ENSO model"	Niño 3.4 and STR-I
"IOD model"	DMI and STR-I

The tree-based models classifying JJA rainfall regimes in 12 regions from the ENSO and IOD models are shown in Table 5.5. These models are spatially consistent with previous research (Nicholls 1989, Risbey et al. 2009, Timbal and Drosdowsky 2013) and the analysis using all predictor variables (Section 5.4.1), as the STR tends to be preferentially selected in southern Australia while Niño 3.4 or the DMI dominate in the north. This confirms that the ENSO and IOD models adequately capture the known sources of variability. It should be noted that node purity will likely be lower in these models compared to those constructed from all predictor variables (Section 5.4.1) for two reasons. Firstly, including fewer predictor variable swill limit over-fitting in the model, and secondly, as the ENSO and IOD models are forced to classify regimes from fewer predictor variables that may have weaker relationships with the response variable. Diagnostics similar to the Akaike Information Criteria can be used to balance closeness of fit with a penalty term to guard against over-fitting. This has not been implemented here as the tree size is maintained to maximise interpretability. The value of this analysis comes

from examining the combinations of predictor variables that define particular regimes, and analysis of the composite rainfall and atmospheric anomalies.

The selection of the DMI as the primary split in the northern regions (TiS, GoC, NECS) may reflect dependence of the IOD on ENSO (Meyers et al. 2007) or it may be the "better of two bad choices" for the region; node purity is low which suggests the latter. Given the strongly data-dependent nature of classification trees, the similarities between the ENSO and IOD models in classifying JJA rainfall regimes is striking; SEC and MDBW split only on STR-I in both models, while SAG splits first on STR-I with additional splits from the tropical predictor variable.

The remainder of this section will discuss case-studies to explore interactions between large-scale modes of climate variability that are relevant for Australian hydroclimatic regimes and highlight the value in a tree-based analysis. Case studies from four regions are discussed in greater detail, the Murray-Darling Basin semi-arid region in JJA (MDBSA, Section 5.4.2.a), the South Australian gulf region in JJA (SAG, Section 5.4.2.b) and the southeast coast region in SON (SEC, Section 5.4.2.c)

Table 5.5. The trees classifying rainfall regimes in 12 Australian drainage divisions in JJA from the ENSO model (left) and the IOD model (right), over the period 1900 – 2009. The first index preceding the backslash (green) is the predictor variable chosen for the first split, with the left and right secondary splits following (black, labelled "L" and "R"). "x" indicates no further splits on that node.

Region	Division	N34 + STRI (JJA) (1st: *first split* / L: *left split* – R: *right split*)	DMI + STRI (JJA) (1st: *first split* / L: *left split* – R: *right split*)
North	TiS	**1st: STRI** / x - **R: STRI**	**1st: DMI** / x - **R: STRI**
	GoC	**1st: N34** / **L: N34** - x	**1st: DMI** / x - x
Northeast	NECS	**1st: N34** /x - **R: STRI**	**1st: DMI** / x - x
	NECBF	**1st: N34** / **L: STRI** - x	**1st: STRI** / **L: DMI** - **R: DMI**
Southeast	SEC	**1st: STRI** / **L: STRI** - x	**1st: STRI** / **L: STRI** - x
	MDBW	**1st: STRI** / x - **R: STRI**	**1st: STRI** / x - **R: STRI**
	MDBA	**1st: STRI** / x - **R: N34**	**1st: STRI** / x - **R: STRI**
	MDBSA	**1st: STRI** / **L: N34** - **R: N34**	**1st: DMI** / x - **R: DMI**
South	TAS	**1st: STRI** / x - **R: STRI**	**1st: STRI** / **L: DMI** - **R: DMI**
	SAG	**1st: STRI** / **L: STRI** - **R: N34**	**1st: STRI** / **L: DMI** - **R: STRI**
West	SWC	**1st: STRI** / x - **R: STRI**	**1st: STRI** / x - **R: DMI**
	IO	**1st: STRI** / **L: STRI** - x	**1st: STRI** / **L: STRI** - **R: DMI**

5.4.2.a Murray-Darling Basin Semi-Arid (MDBSA)

This section examines the interactions between indices of large-scale modes of climate variability that influence MDBSA rainfall in JJA. Significant correlations exist between all predictor variables (Niño 3.4, DMI and STR-I) and MDBSA rainfall (r = -0.28, -0.35 and -0.46, respectively). Interestingly, the IOD model does not require information about STR-I to classify rainfall regimes, despite a stronger correlation between STR-I and rainfall compared to the DMI and rainfall (Figure 5.12). This tree structure is in contrast to the wetter regions of the MDB (i.e. MDB Wet and MDB Agricultural) that define rainfall regimes solely from STR-I (Table 5.5). The differences between these tree structures suggest that different mechanisms are associated

with rainfall variability in different regions of the MDB. The more coastal regions (MDBW and MDBA) have a stronger relationship with the STR, likely through changes in moist onshore easterly winds. In contrast, rainfall in the inland MDBSA region has a stronger relationship with Indian Ocean SST variability that brings rainfall to the region through northwest cloud bands. In addition, the IOD model classifies years as average more than would be expected, i.e. 56 times compared to an expected 37 years.

ENSO and the STR are also positively correlated in JJA (r = 0.22) and both indices are required to adequately define rainfall regimes. As such, the ENSO model will be the focus for the remainder of this section (Figure 5.12).

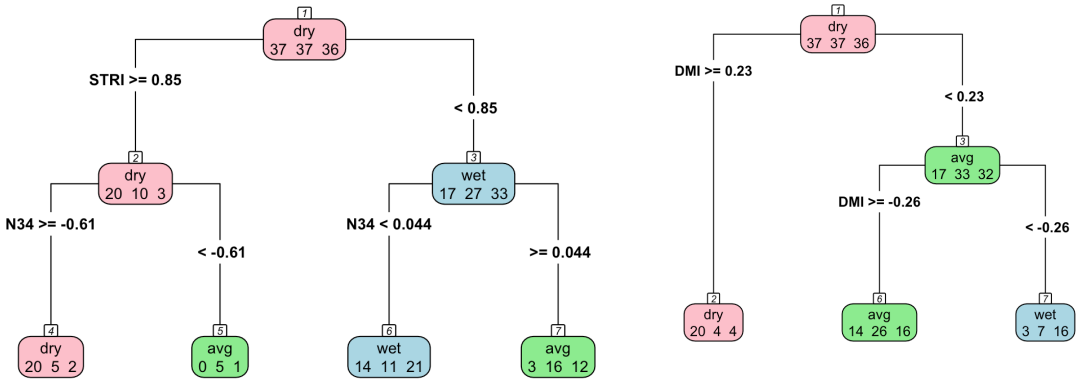

Figure 5.12. The trees classifying JJA rainfall regimes in the MDBSA, from the ENSO model (left) and the IOD model (right), over the period 1900 – 2009. The number of years classified into the dry (red), average (green) and wet (blue) regimes is marked.

The ENSO model classifies four distinct rainfall regimes (Figure 5.13) with both phases of ENSO (El Niño and La Niña) and the STR-I (high and low STR-I) each with two regimes. This tree structure allows the examination of combined tropical – sub-tropical climate drivers.

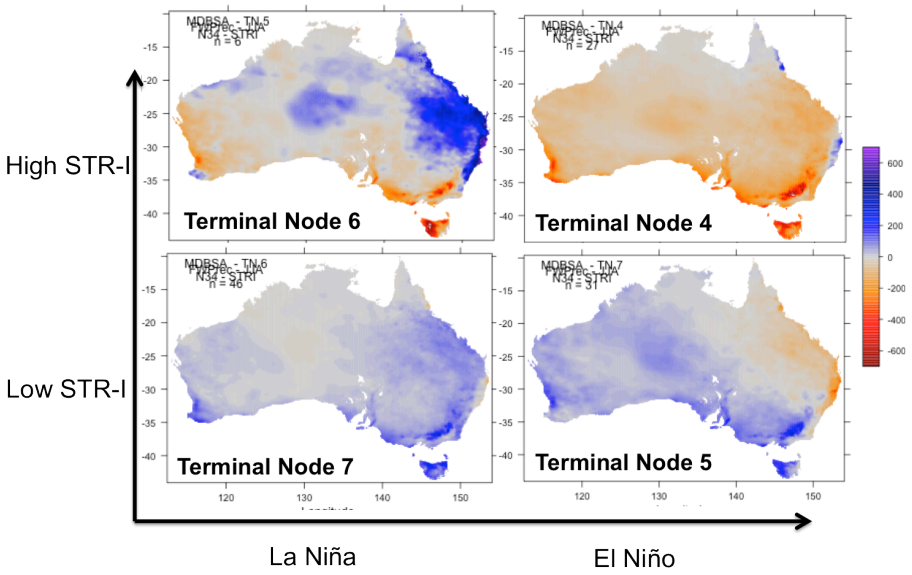

Figure 5.13. The JJA rainfall regimes from the tree classifying JJA MDBSA, from the ENSO model, over the period 1900 – 2009. The combinations of high and low STR intensity and La Niña and El Niño conditions in each regime are shown.

The 'dry' and 'wet' regimes are defined when ENSO and the STR are in the same phase, so that a combination of high STR-I and El Niño conditions results in the classification of a 'dry' regime (Terminal Node 4) while the opposite combination (i.e. low STR-I and La Niña conditions) result in a 'wet' regime (Terminal Node 6). Rainfall and atmospheric composites in these regimes are mirror images (Figure 5.13 and Figure 5.14). For example, in the 'dry' regime the negative rainfall anomalies are spatially consistent across the continent, with vertical motion, vector wind and geopotential height anomalies all indicating that El Niño is well developed in the Pacific Ocean and that southern Australia is under the influence of anomalously high MSLP.

In contrast, the two 'average' regimes show the interactions between the STR-I and ENSO that are important for southeastern Australian rainfall. For example, a comparison of the two rainfall regimes defined during the La Niña phase but with high or low STR-I (Terminal Nodes 5 and 6, respectively, in Figure 5.13 and Figure 5.14) shows how the STR can modulate the teleconnection of ENSO influence over southern Australia. In the 'average' regime defined by high STR-I and La Niña conditions (Terminal Node 5: STR-I $\geq$ 0.85 hPa and Niño 3.4 < -0.61°C) the high pressure anomaly over southern Australia confines the positive rainfall anomalies that occur during La Niña to north-eastern Australia. Indeed, the vector wind anomalies in this regime suggests that the strong positive rainfall anomalies on the east coast are likely caused by the interaction between increased trade winds associated with La Niña conditions and enhanced onshore flow associated with a strong STR (Figure 5.14).

In contrast, the vector wind anomalies in the 'average' regime defined by El Niño conditions and low STR-I suggests that increased rainfall in southern Australia is associated with increased westerly winds while northeastern Australia is influenced by decreased trade winds and warmer SST anomalies in the central Pacific Ocean. In this regime, the negative rainfall anomalies generally associated with El Niño conditions are confined to the northeast by the lower MSLP anomalies associated with low values of the STR-I.

The following section (5.4.2.b) discusses rainfall regimes in a neighbouring region, the South Australian Gulf (SAG).

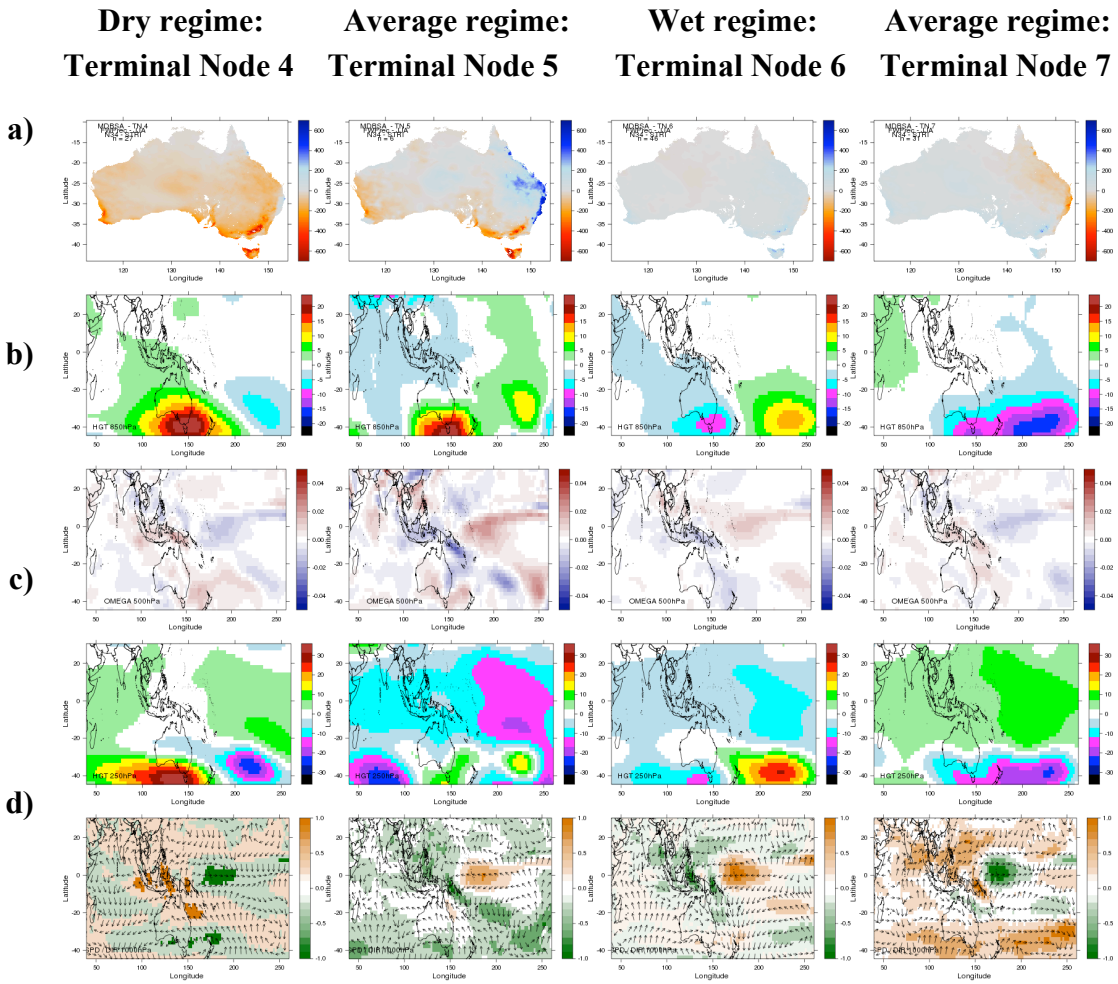

Figure 5.14. Spatial distribution of composite a) rainfall, b) 850 hPa geopotential height, c) 500 hPa vertical velocity and d) 200 hPa geopotential height anomalies in the four regimes defined when JJA MDBSA rainfall is classified from the ENSO model, over the period 1900 – 2009.

5.4.2.b South Australian Gulf (SAG)

The previous section showed the importance of interactions between ENSO and the STR in defining rainfall regimes in the MDB. This section will further explore these interactions, using a case study from the SAG, defining rainfall regimes in JJA with the ENSO and IOD models.

In JJA, when the mean position of the STR is over the continent, rainfall in southern Australian is dominated by STR variability. This is demonstrated by the classification of two rainfall regimes in both models from only the STR-I, consistent with the strong correlation between the STR-I and SAG rainfall (r = -0.67). In both models, a 'dry' regime (Terminal Node 4) is defined when STR-I is high (STR-I ≥ 0.68 hPa) while an 'average' regime (Terminal Node 5) is classified from median positive values of STR-I (STR-I between -0.022 – 0.68 hPa). 1997, a strong El Niño and positive IOD year, was classified into the 'dry' regime by both models, based only on STR-I. The classification of regimes using only the STR highlights the linearity of the relationship between STR-I and SAG rainfall (Figure 5.15). In these rainfall regimes the state of the tropical oceans has little effect, as a La Niña or negative IOD year can be dry provided the STR-I is high (Figure 5.15).

Despite the similarities in defining the 'dry' and 'average' rainfall regimes, both models then require information on the state of the tropical modes of variability to define the 'wet' regimes that are the focus of the following sections (5.4.2.b.i and 5.4.2.b.ii).

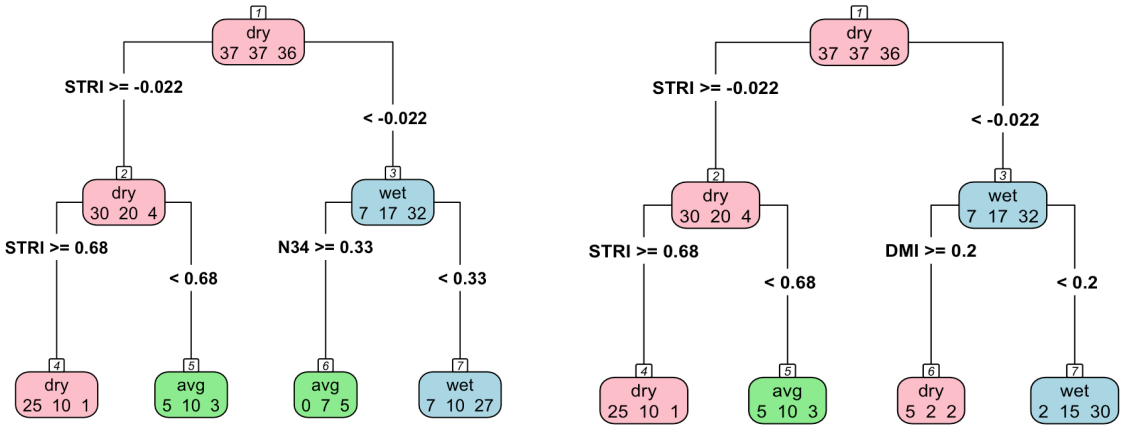

Figure 5.15. The trees classifying JJA rainfall regimes in the SAG, from the ENSO model (left) and the IOD model (right), over the period 1900 – 2009. The number of years classified into the dry (red), average (green) and wet (blue) regimes is marked.

5.4.2.b.i SAG ENSO model (JJA)

Section 5.4.2.b outlined the rainfall regimes classified from only the STR-I, common to both models. This section discusses the rainfall regimes in SAG that were defined from both the STR-I and Niño 3.4. The inclusion of ENSO variability in defining SAG rainfall regimes shows that the tree-based models are able to extract additional information compared to linear methods, as the correlation between Niño 3.4 and SAG rainfall is insignificant (r = -0.17 n/s). The ENSO model defines a 'wet' regime (Terminal Node 7) when a weak STR-I is combined with the neutral or negative phase of ENSO (Figure 5.15, STR-I < -0.022 and Niño 3.4 ≤ 0.33). Many La Niña years are classified into this 'wet' regime (Table 5.6 and Figure 5.15).

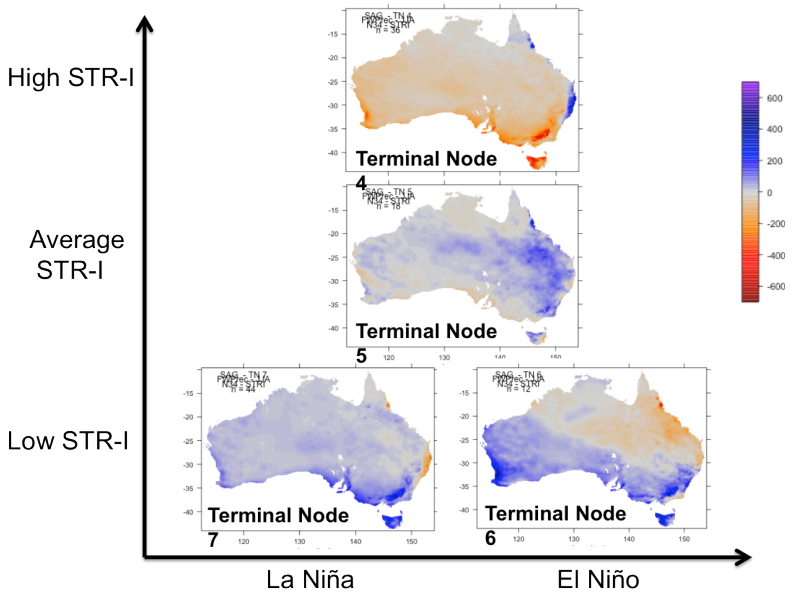

Figure 5.16. The JJA rainfall regimes from the tree classifying JJA SAG, from the ENSO model, over the period 1900 – 2009. The combinations of high and low STR intensity and La Niña and El Niño conditions in each regime are shown.

Positive rainfall anomalies in the 'wet' regime (Terminal Node 7, Figure 5.16 and Figure 5.17) are widespread and spatially homogeneous across Australia, with only the southeast coast displaying negative anomalies (Figure 5.16 and Figure 5.17). The composite atmospheric response in this regime indicates that La Niña is active. The Australian region and Indo-Pacific basin are dominated by upper and lower tropospheric circulation anomalies that indicate a cooler and wetter atmosphere. Anomalous cyclonic circulation is located over southern Australian bringing southerly air and stronger westerly winds over the continent (Figure 5.16 and Figure 5.17).

Table 5.6. The years classified into the average (Terminal Node 6) and wet (Terminal Node 7) rainfall regimes in SAG from the ENSO model, over the period 1900 – 2009. Red (blue) indicates the positive (negative) phase of ENSO.

Model	Node	Years
ENSO model	Average regime (Terminal Node 6)	1900, 1904, 1920, 1926, **1930**, 1931, 1958, **1963**, **1972**, 1991, 1992, 2004
	Wet regime (Terminal Node 7)	1901, **1903**, 1907, **1909**, **1910**, 1915, **1916**, **1917**, **1922**, 1923, 1927, 1929, 1932, 1935, 1939, **1942**, 1943, 1946, 1951, 1952, 1953, **1955**, **1956**, 1961, **1964**, 1968, **1970**, 1971, **1973**, 1974, **1975**, **1978**, **1981**, **1984**, 1985, **1986**, 1989, 1990, **1996**, **2000**, 2001, 2003, **2007**[1], **2009**[2]

[1] 2007 is not classified in Ummenhofer et al (2011), but is considered a La Niña and positive IOD year by BoM (2013).

[2] 2009 is not classified in Ummenhofer et al (2011), but is considered an El Niño by BoM (2013).

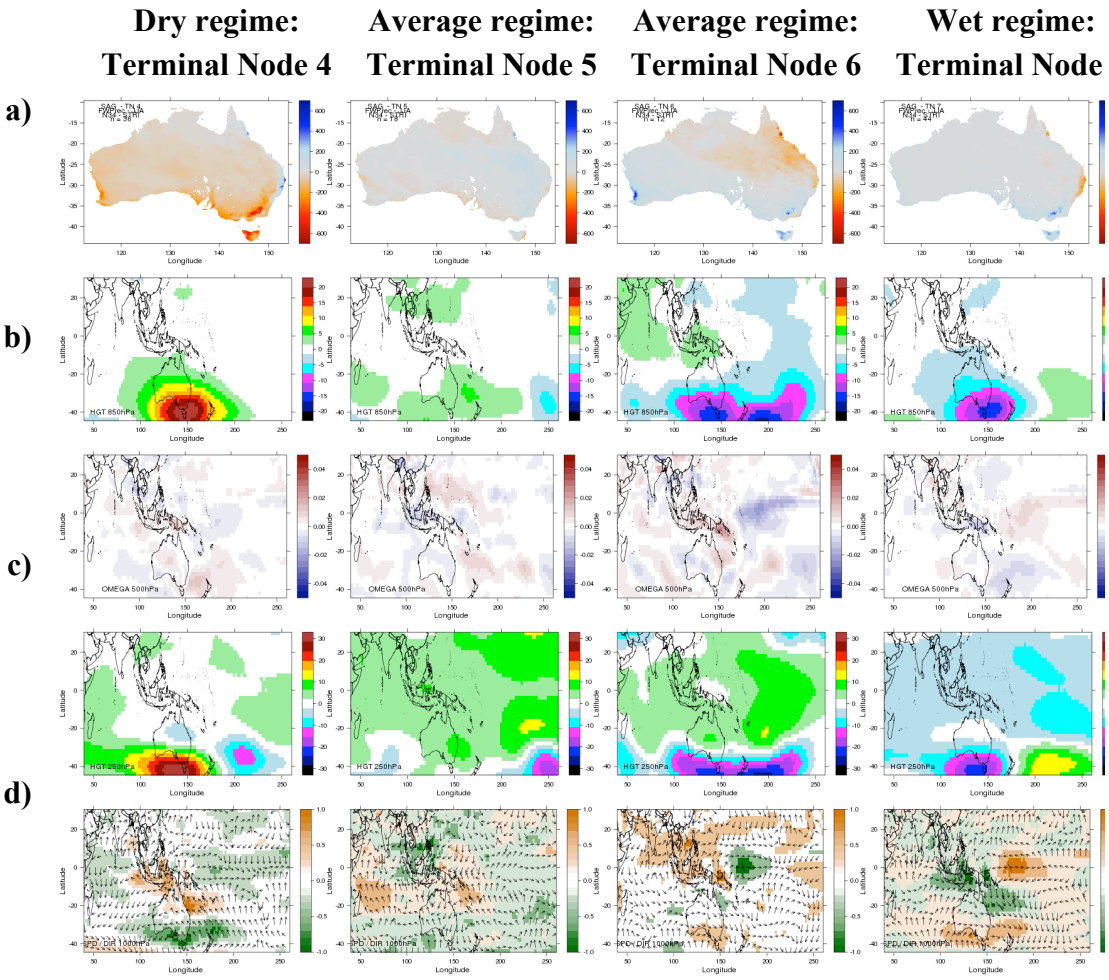

Figure 5.17. Spatial distribution of composite a) rainfall, b) 850 hPa geopotential height, c) 500 hPa vertical velocity and d) 200 hPa geopotential height anomalies in the four regimes defined when JJA SAG rainfall is classified from the ENSO model, over the period 1900 – 2009.

An 'average' rainfall regime (Terminal Node 6, Figure 5.17) is defined when a weak STR is combined with the positive phase of ENSO (STR-I < -0.022 hPa and Niño 3.4 > 0.33°C), resulting in regions of positive (negative) rainfall anomalies in southern (northeastern) Australia. This spatial signature is reflected in the atmospheric circulation composites as a cooler atmosphere is found over southern Australia in the lower and upper atmosphere. In addition, the vector wind anomalies over southern Australia are reminiscent of those found in the 'wet' regime, while the vertical velocity anomalies in the central equatorial Pacific indicate an El Niño-like state (Figure 5.17). Many of the years in this regime are considered to be neutral ENSO years but three El Niño years (1930, 1963, 1972) have SAG rainfall in the middle or upper tercile (Table 5.6). In 1930, 1963, and 1972 the anomalous JJA values for Niño 3.4 (STR-I) were 0.46, 0.46 and 0.79°C (-0.41, -1.85, -0.19 hPa), respectively; these values are consistent with the positive linear relationship between the predictor variables. Rainfall anomalies (Figure

5.18) for these individual years are more spatially heterogeneous than the rainfall regime composites as they lack smoothing that stems from averaging years. In 1972, average or positive rainfall anomalies in southern Australia (SAG region and Tasmania) are combined with widespread negative rainfall anomalies over eastern Australia that reflects the strong SST anomalies in the eastern Pacific Ocean (Niño 3.4 = 0.79°C) and moderate STR-I (STR-I = -0.19 hPa). The Pacific Ocean is in a similar state in 1930 and 1963 (Niño 3.4 = 0.46°C) but STR-I is much lower in 1963 (STR-I = -1.85 hPa). This is reflected in the spatial distribution of rainfall with more positive rainfall anomalies over southern Australia in 1963 compared to 1930.

The comparison of these three years supports the analysis of the regimes, which showed how interactions between tropical and sub-tropical climate drivers control the spatial distribution of the rainfall anomalies. ENSO and STR-I are positively correlated but the STR has relationships with other large-scale climate drivers. The impact of ENSO cannot be felt in southern Australia if the signal cannot be teleconnected to the region through changes to the STR. If increased SST anomalies in Niño 3.4 are not combined with a more intense STR, then the teleconnection can break down and rainfall anomalies do not decline as expected during an El Niño event. STR position likely plays an additional role but it is not explored in this chapter. The following section examines the interactions between Indian Ocean and STR variability the influence SAG rainfall in JJA.

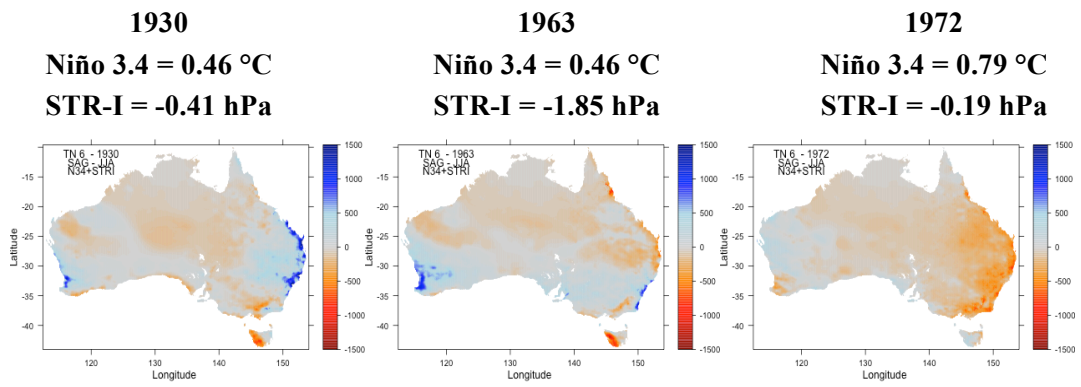

Figure 5.18. JJA rainfall anomalies for three El Niño years (1930, 1963 and 1972 – left to right) classified into the average rainfall regimes from the ENSO model, with the seasonally averaged Niño 3.4 and STR-I values indicated for each year.

5.4.2.b.ii SAG IOD model (JJA)

Similarly to the ENSO model, the IOD model only requires input from the tropical climate driver to classify rainfall regimes when STR-I is low (Figure 5.16 and Figure 5.19). The

inclusion of the DMI in the classification tree is consistent with the significant correlation between the DMI and rainfall in the SAG (r = -0.39).

When the STR-I is low, the phase of the DMI is essential for controlling the sign of the rainfall anomaly over southern Australia, as the IOD model classifies a 'wet and a 'dry' regime, depending on the state of the Indian Ocean (Figure 5.19 and Figure 5.20). This is in contrast to the ENSO model, that defined 'wet' and 'average' rainfall regimes when STR-I was low, depending on ENSO phase, (Figure 5.15). Weak STR-I in combination with the positive phase of the IOD (STR-I < -0.022 hPa and DMI > 0.2°C) defines the 'dry' regime (Terminal Node 6, Figure 5.15 and Figure 5.19). The definition of a 'dry' regime in southern Australia in the IOD model compared to an 'average' regime in the ENSO model suggests that Indian Ocean variability has a stronger influence over southern Australian rainfall than ENSO, which is consistent with our current understanding of the linear relationships (Risbey et al. 2009).

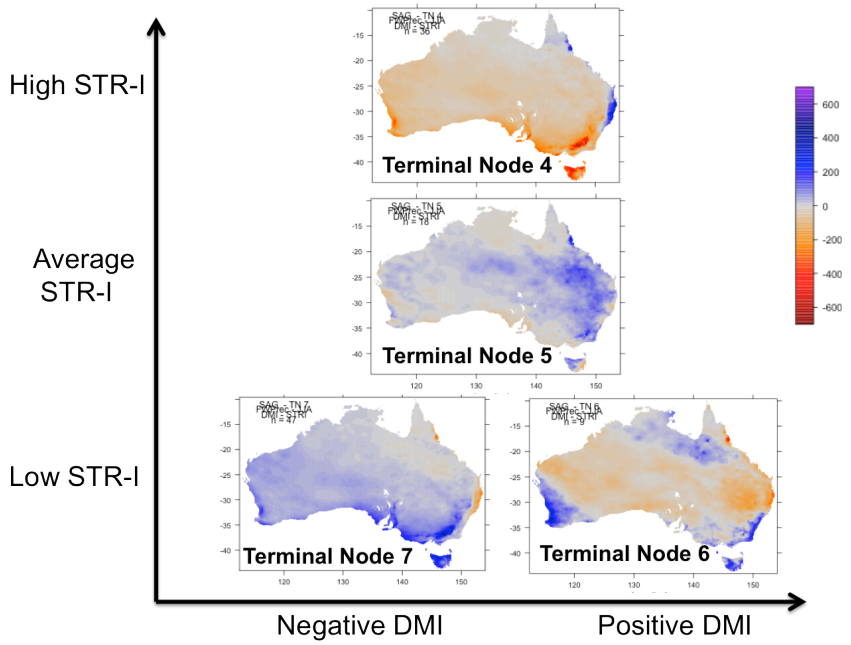

Figure 5.19. The JJA rainfall regimes from the tree classifying JJA SAG, from the IOD model, over the period 1900 – 2009. The combinations of high and low STR intensity and the negative and positive IOD phases in each regime are shown.

The positive rainfall anomalies in this 'dry' regime (Terminal Node 6) are confined to southwest Western Australia, southwest eastern Australia, Tasmania and northern Australia, while negative anomalies occur in the central third of the continent (Figure 5.19 and Figure 5.20). This is in contrast to the 'dry' regime classified solely from high STR-I (Terminal Node 4) that has a very different rainfall and atmospheric signature. The positive rainfall anomalies in

southern Australia are likely associated with the weak STR-I, evidenced by the strengthened westerly winds and negative geopotential height anomalies across southern Australia (Figure 5.20). The influence of the IOD on the central third of Australia is seen in the anomalous subsiding air stretching from the eastern Indian Ocean across Australia to the mid-east coast. The positive phase of the IOD results in negative rainfall anomalies in southern Australia even when MSLP is low (i.e. STR-I is weak), which suggests that the IOD influences rainfall via other processes such as changing moisture transport, rather than through circulation changes. This shows that the phase of the IOD is more effective in modulating the impacts of a weak STR on rainfall in southern Australia, compared to teleconnections from the Pacific Ocean (i.e. ENSO).

Table 5.7. The years classified into the dry (Terminal Node 6) and wet (Terminal Node 7) rainfall regimes in SAG from the IOD model, over the period 1900 – 2009. Red (blue) indicates the positive (negative) phase of IOD.

Model	Node	Years
IOD model	Dry regime (Terminal Node 6)	**1923**, **1926**, **1935**, **1946**, **1961**, **1963**, 1972, **1975**, **2007**[1]
	Wet regime (Terminal Node 7)	1900, 1901, 1903, 1904, 1907, **1909**, 1910, **1915, 1916, 1917**, 1920, 1922, 1927, 1929, **1930**, 1931, 1932, 1939, **1942**, 1943, 1951, 1952, 1953, 1955, 1956, **1958**, 1964, **1968**, 1970, 1971, 1973, **1974**, 1978, 1981, 1984, **1985**, 1986, **1989**, 1990, **1991**, **1992**, 1996, 2000, 2001, 2003, **2004**, 2009[2]

[1] 2007 is not classified in Ummenhofer et al (2011), but is considered a La Niña and positive IOD year by BoM (2013).

[2] 2009 is not classified in Ummenhofer et al (2011), but is considered an El Niño by BoM (2013).

In addition, this analysis highlights the importance of phase locking between tropical and regional modes of variability in defining widespread rainfall anomalies. For example, weak STR-I combined with the neutral or negative phase of the IOD (STR-I < -0.022 hPa and DMI ≤ 0.2°C) defines a 'wet' regime. This regime has spatially homogeneous positive rainfall anomalies across eastern Australia (Figure 5.19 and Figure 5.20).

This section (Section 5.4.2.b) has compared the ENSO and IOD models classifying JJA rainfall in the SAG to explore interactions between remote Indian and Pacific Ocean modes of variability with the intensity of the STR.

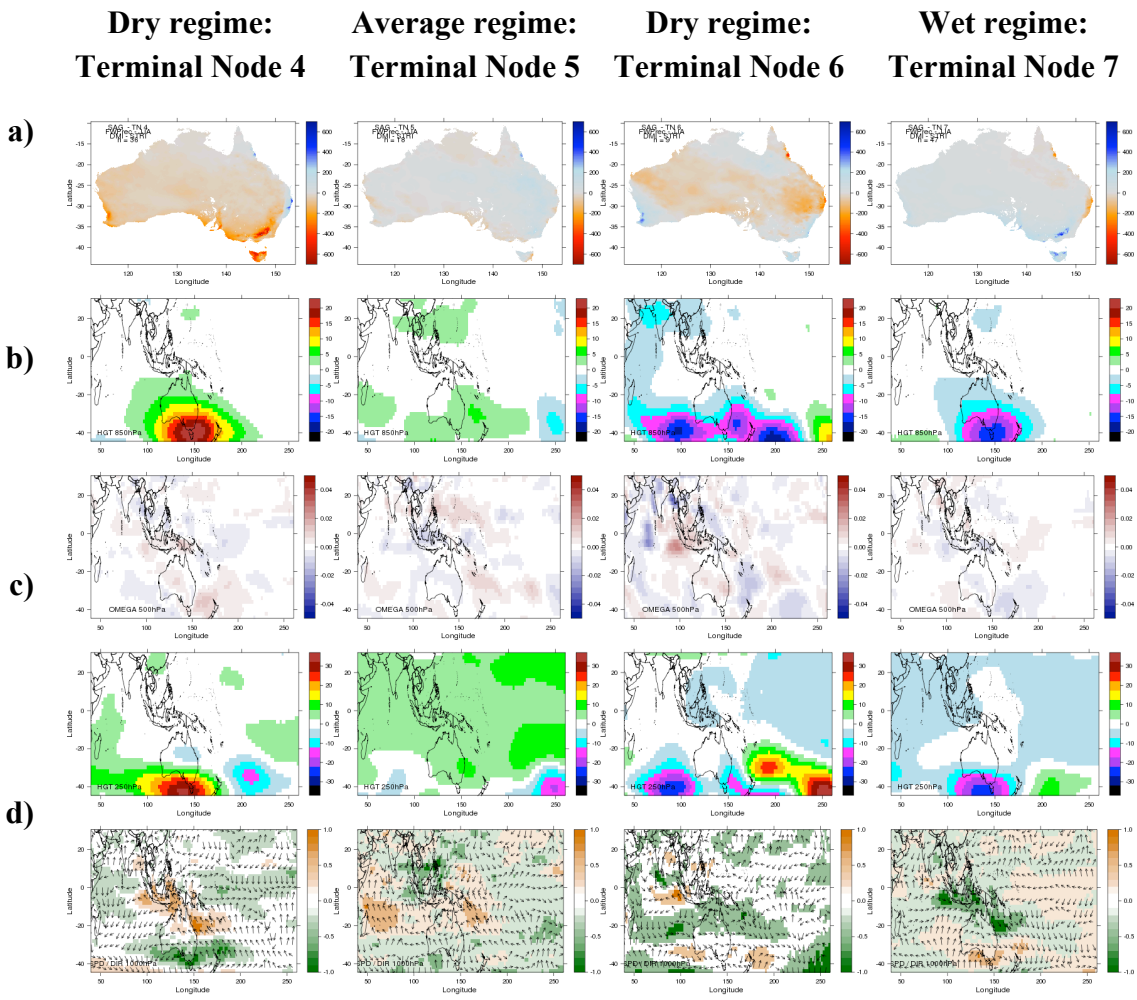

Figure 5.20. Spatial distribution of composite a) rainfall, b) 850 hPa geopotential height, c) 500 hPa vertical velocity and d) 200 hPa geopotential height anomalies in the four regimes defined when JJA SAG rainfall is classified from the IOD model, over the period 1900 – 2009.

5.4.2.c Southeast Coast ENSO model (SON)

This section discusses the combined influence of STR-I and Pacific Ocean variability on SON rainfall regimes in the southeast coast (SEC) region. Consistent with the correlation analysis, the classification tree suggests there is a stronger relationship between ENSO variability and rainfall, compared to STR-I and rainfall, as Niño 3.4 is selected for the first split (Figure 5.21). Only 18 years are classified into the 'average' regime than the expected 37. This lowers confidence in the model and suggests too many years are being classified into the 'dry' regime.

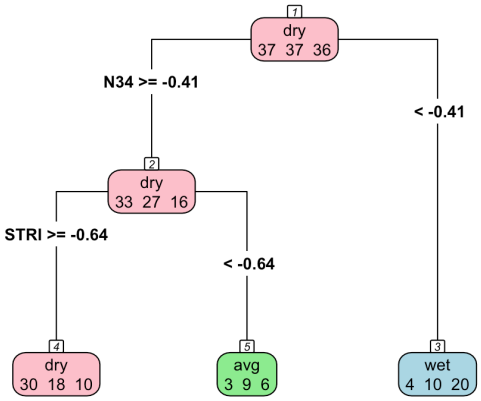

Figure 5.21. The tree classifying SON rainfall regimes in the SEC, from the ENSO model, over the period 1900 – 2009. The number of years classified into the dry (red), average (green) and wet (blue) regimes is marked.

A 'wet' regime is defined by the La Niña phase of ENSO (Niño 3.4 < -0.41), which is associated with positive rainfall anomalies across the eastern third of the continent (Terminal Node 3, Figure 5.22 and Figure 5.23). The intensity of the STR is not required to classify the wet regime. The impact of the STR-I is only required when SST anomalies in the Pacific Ocean are warm (Niño 3.4 ≥ -0.41). A 'dry' regime (Terminal Node 4, Figure 5.22 and Figure 5.23) with negative rainfall anomalies over eastern Australia is defined when both ENSO and the STR-I are in the dry phases (i.e. Niño 3.4 ≥ -0.41 and STR-I ≥ -0.64). The negative rainfall anomalies in the dry regime are associated with a warmer troposphere and vertical velocity anomalies that are generally associated with El Niño, i.e. negative (positive) anomalies indicating rising (subsiding) air over the equatorial Pacific (maritime continent) and anti-cyclonic wind anomalies over southern Australia. This is evident despite the classification of some neutral ENSO and average STR-I years into the regime (Figure 5.22 and Figure 5.23).

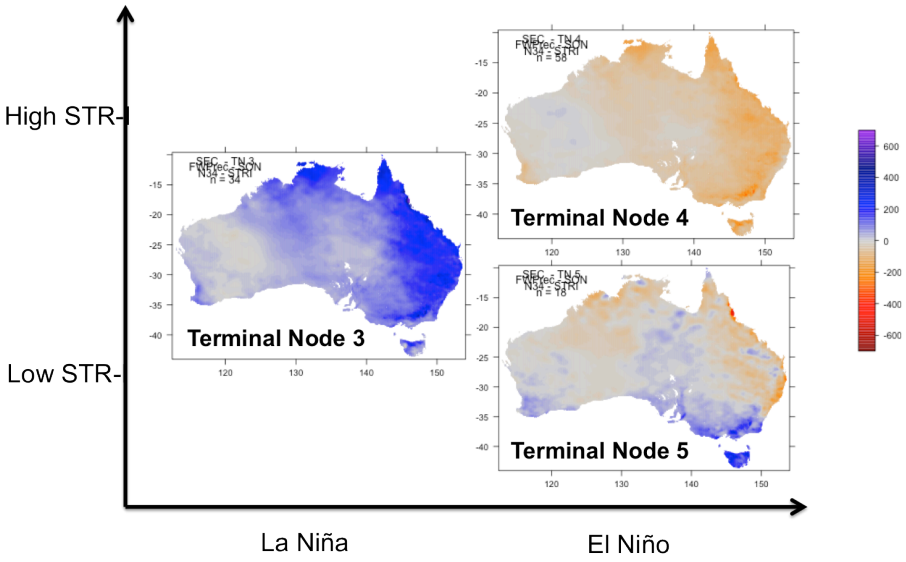

Figure 5.22. The SON rainfall regimes from the tree classifying SON SEC, from the ENSO model, over the period 1900 – 2009. The combinations of high and low STR intensity and the negative and positive ENSO phases in each regime are shown.

An 'average' rainfall regime (Terminal Node 5, Figure 5.22 and Figure 5.23) is defined when the dry phase of ENSO is combined with a weak STR (i.e. Niño 3.4 ≥ -0.41 and STR-I < -0.64). Spatially, this regime is characterised by negative rainfall anomalies in northeastern Australia and positive anomalies in the southeast, dividing the SEC region and likely causing the classification as an 'average' rainfall regime. The vertical velocity anomalies in this regime suggest that the Pacific Ocean is characterised by El Niño-like variability. However, the teleconnection from the tropical Pacific Ocean does not reach southern Australia as a cooler lower troposphere and vector wind anomalies that are associated with a weak STR dominate in the south (Figure 5.23). The Pacific-South American pattern can be seen in the upper troposphere geopotential height anomalies in the 'wet' and 'dry' regimes is evident but less developed in the average regime, consistent with the state of the Pacific Ocean. In the dry and wet regimes the warm and cool upper tropospheric anomalies stretch across the Pacific and Indian Oceans, while in the average regime the warm anomalies are confined to the Pacific Ocean. This suggests a possible role for the Indian Ocean, through the relationship with the STR.

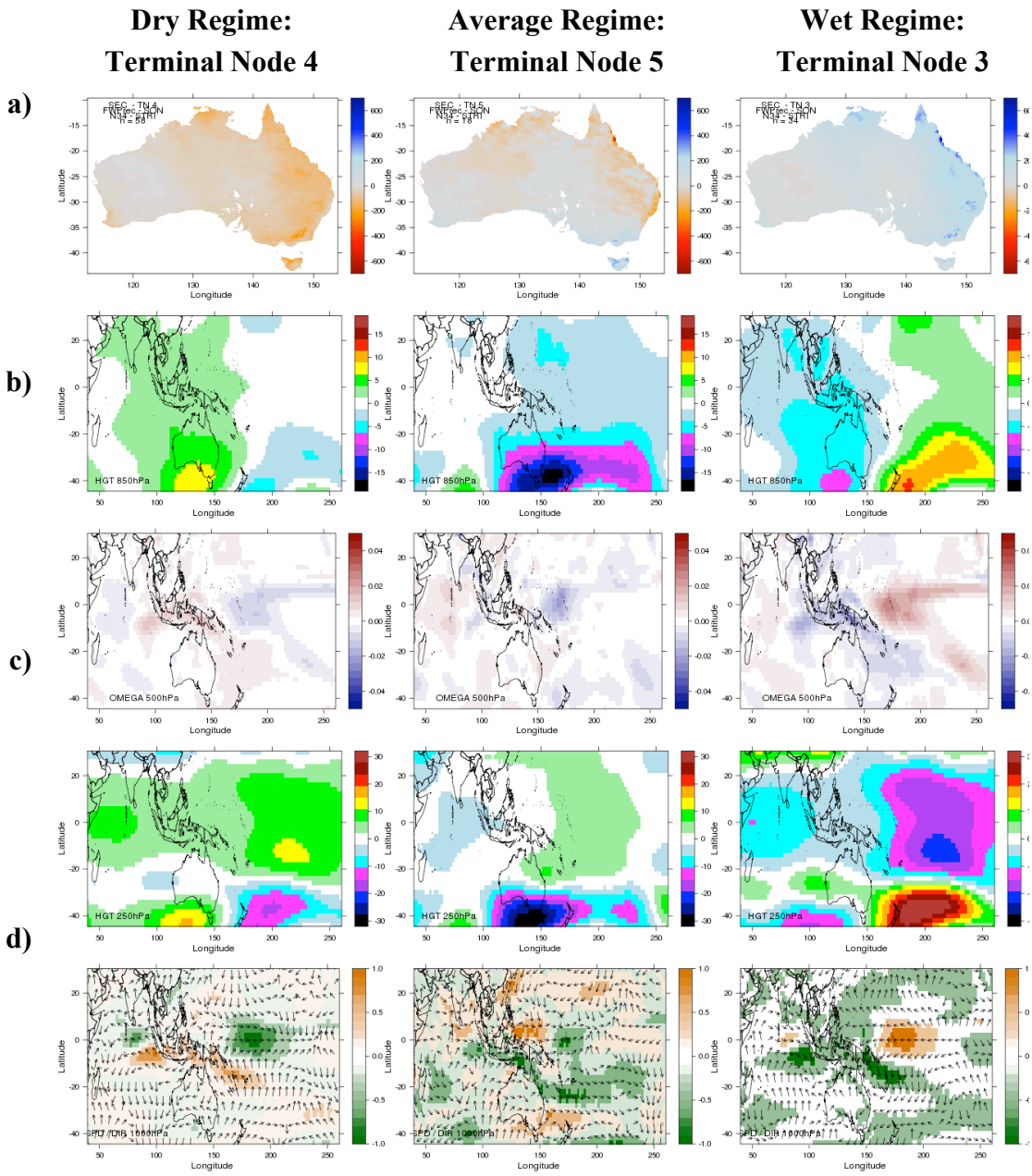

Figure 5.23. Spatial distribution of composite a) rainfall, b) 850 hPa geopotential height, c) 500 hPa vertical velocity and d) 200 hPa geopotential height anomalies in the four regimes defined when SON SEC rainfall is classified from the ENSO model, over the period 1900 – 2009.

5.5 Discussion and conclusions

This chapter was broadly structured in two parts. The first part (Section 5.3.2) played a structural role in the thesis, introducing the methodology and outlining the ways tree-based models are used throughout the thesis. Several key features of CART were introduced,

including concepts related to tree structure (type of tree, size of tree and predictor variables), node purity and the composite analysis. Following from this example based general introduction, Section 5.4.1 presented tree-based models that classified rainfall regimes in 12 regions in each of the traditional four seasons from all predictor variables. The section showed that the seasonal and spatial variability in the predictor variables selected to define rainfall regimes in the tree-based analysis is consistent with previous research. This served several purposes. Firstly, the use of the tree-based method is validated, as the results are broadly similar to those obtained from a classical statistical framework. Secondly, it adds confidence to the use of linear methods by showing that such methods that are limited to capturing the linear portion of the variability can adequately describe most rainfall regimes. Finally, the maximum node purity that can be achieved with this set of predictor variables was presented to give context to the subsequent analysis where the number of predictor variables is limited. The inclusion of all predictor variables allowed the tree-based models to search many more possible split points to divide the response variable to results in the largest node purity.

The second part of this chapter (Section 5.4.2) presented a pairwise comparison of rainfall regimes defined from a subset of predictor variables (Niño 3.4, DMI, STR-I). This section showed how tree-based models can provide additional insights about interactions between climate drivers that are important for Australian rainfall regimes. The tree-based models can give increased knowledge about interactions between the regional and remote modes of variability associated with Australian climate regimes. For example, the MDBSA case study suggested that different mechanisms were associated with rainfall in the coastal and more inland parts of the basin (MDBW, MDBA), which is consistent with rainfall associated with moist onshore easterly winds (Rakich et al. 2008). This section has again validated the use of the novel method and shown the value that tree-based models can bring to an analysis of discrete events such as rainfall.

The models predicting rainfall regimes from interactions between the STR and either ENSO or the IOD (the ENSO and IOD models) captured the major sources of variability, consistent with trees using all predictor variables (Section 5.4.1) and with previous research (McBride and Nicholls 1983, Nicholls 1989, Risbey et al. 2009). The STR dominates over the tropical modes of climate variability (i.e. ENSO and the IOD) in southern Australia, as it is chosen for the primary split in many regions. This is consistent with the correlation analysis and physically expected due to the close relationship between MSLP and rainfall. The tree-based models are best able to extract additional information in regions where multiple predictor variables have strong relationships and one predictor variable isn't dominant.

The interrelated nature of the climate system and the strong correlations between remote and regional modes of climate variability can mask more subtle, but important, relationships. The different spatial patterns of rainfall anomalies in regimes with similar ENSO phase and differing STR-I shows how the tropical teleconnections can be modulated by other drivers. For example, in the MDBSA, rainfall anomalies are positive and widespread across Australia when La Niña is combined with low STR-I. Phase locking of tropical modes and sub-tropical drivers consistently resulted in the strongest and most widespread rainfall anomalies that affected southern and northern Australia. Conversely, positive rainfall anomalies are confined to the northeast while southern Australia experiences rainfall deficits when La Niña is combined with high STR-I. The confinement of ENSO impacts to the northeast was also found in the SAG ENSO model. The inclusion of ENSO variability in defining SAG rainfall regimes at all is interesting, given the insignificant correlation between the variables. This shows that the tree-based models are able to add more nuanced information about interactions between tropical and sub-tropical climate drivers.

The SAG and MDBSA ENSO models confined negative rainfall anomalies to northeast Australia when El Niño was combined with low STR-I. This pattern is spatially consistent with the partial correlation between ENSO and Australian rainfall once the influence of the IOD is removed (Risbey et al. 2009). This suggests that the low STR-I effectively removes the influence of the IOD, and shows that the STR is a mechanism for the interaction between the IOD and ENSO.

The IOD model classifying SAG rainfall highlighted the importance of interactions between the IOD and the STR for rainfall regimes in southern Australia. When the STR-I is weak the IOD is able to modulate the impact of low MSLP. Low STR-I is generally associated with stronger mean westerly winds and positive rainfall anomalies in southern Australia. When STR-I is low and the IOD is in the positive phase the region of positive rainfall anomalies is confined to the most southern parts of continental Australia while negative rainfall anomalies associated with the positive phase of the IOD are found over much of the continent. It is not suggested that the IOD and STR independently influence southern Australian rainfall, as the variables are positively correlated and the STR is associated with the 'East Indian Ocean' wave train that teleconnects the IOD influence to southern Australia (Cai et al. 2011b).

Various mechanisms have been put forward to explain variability between El Niño events, including decadal variability (Power et al. 1999), shifts in the Walker Circulation (Wang and Hendon 2007) and internal atmospheric variability (Kumar and Hoerling 1997). This analysis suggests that STR variability plays a significant role. The intensity of the STR modulates the spatial extent of ENSO impacts on Australia, so that when STR-I is low (high) the impacts of an

El Niño (La Niña) event are confined to northeastern Australia. Multiple large-scale modes of climate variability affect the STR, including ENSO through the strength of the Hadley Circulation, the IOD through the East Indian Ocean wave train and the Southern Annual Mode through zonal pressure changes. The positive linear relationship between both tropical indices and the STR suggests that the positive phases of the tropical oceans tend to be associated with increased STR-I. Yet the relationship is not perfect and the strength of the teleconnection is not uniform between years, likely due to interactions between large-scale modes of climate variability. This suggests that the strength of the teleconnection is as important as the state of the tropical oceans in defining the spatial characteristics of rainfall regimes. This highlights the importance of tropical – sub-tropical interactions for rainfall variability in Australia. The tree-based method is able to give additional insight about differences between ENSO and IOD events (e.g. "inter-El Niño" variability), as the interaction with the STR is critical. The impacts of ENSO or IOD events are not felt in southern Australia if the teleconnection is weak and the influence of the modes is not transmitted to changes in the STR.

This analysis showed that CART can identify optimal ways of 'binning' response variable data into smaller groups that may initially be non-obvious. It is common for ENSO-IOD interactions to be discussed by separating years into each of the nine ENSO and IOD phases, i.e. positive, negative, neutral ENSO and IOD, and their combinations (Ummenhofer et al. 2011). Such classifications are often based on SST anomalies, alone or in combination with atmospheric and oceanic indicators (Meyers et al. 2007, Ummenhofer et al. 2011). Regardless of the method, such classifications of ENSO and IOD years are somewhat arbitrary and may not be useful in the future when the characteristics of tropical variability may change. As such, CART provides an objective way to separate years into homogeneous nodes.

The importance of facets of STR variability (intensity and position) for southern Australia will be further explored in the following chapter (Chapter 6).

Chapter 6 : Linear and nonlinear statistical analysis of the impact of sub-tropical ridge intensity and position on southeast Australian rainfall.

Kirien Whan[1], Bertrand Timbal[2] and Janette Lindesay[1]

[1] The Fenner School and Environment and Society, The Australian National University.

[2] The Centre of Climate and Weather Research, The Australian Bureau of Meteorology.

Manuscript for the International Journal of Climatology

Status: Submitted in 2010, published in 2013 – DOI: 10.1002/joc.3689

6.1 Chapter summary and context

The previous chapter served as an introduction to the tree-based methods and demonstrated how they are used in this thesis (Chapter 5). In Chapter 5, the focus was on interactions between tropical and sub-tropical modes of variability that have known relationships with Australian climate variability (Risbey et al. 2009, Timbal and Drosdowsky 2013, Timbal and Hendon 2011). Several case studies were selected to emphasise the validation of this novel statistical method, and showed how added insight can be gained from this type of analysis compared to that within a classical statistical framework. It was shown that variability between the impacts of various ENSO (or IOD) events could be explained, in part, by the varying strength of the STR intensity and the modulating role it can on tropical teleconnections.

Following on from that work, this chapter looks at interactions between facets of the STR (i.e. intensity and position) in greater detail. The importance of STR intensity for southern Australian rainfall was highlighted in Chapter 5, and so the focus in this chapter is on south-west eastern Australia; a region where STR intensity explains a large portion of the variability

This chapter answers the question "Can a CART analysis of the impact of sub-tropical ridge intensity and position on rainfall provide additional understanding of the relationship between the STR and precipitation in SEA?" We show that indeed a tree-based analysis is able to add insight as to the nature of the interactions between STR intensity and position that are important in defining southeast Australian rainfall regimes. The intensity and position of the sub-tropical ridge (STR) have strong relationships with rainfall variability in southern Australia. The combined effect of intensity and position in March-April-May (MAM) and June-July-August (JJA) is the focus of this research. Linear statistics were used first: area-averaged and Australia-wide spatial correlations of STR intensity and position with precipitation in south-west eastern Australia reveal that STR intensity has a much stronger and more widespread relationship with precipitation in both seasons. Over time, these relationships vary in magnitude and spatial extent with the sign of the correlation changing between two 50-year epochs. These nonlinearities were investigated further using classification trees. Area-averaged precipitation data (terciles) for southwest eastern Australia was classified on the basis of STR intensity and position. In both seasons the classification trees identify STR intensity as the primary partition defining the dry group, supporting the linear analysis. In the transition season of MAM, the time of year when the mean position of the STR is more southerly, STR position is important in distinguishing between a 'winter-like' and a 'summer-like' wet groups, providing STR intensity is low. Vector wind analyses were computed to explain the composite seasonal precipitation anomaly results in terms of different circulation patterns associated with these two wet groups.

The frequency of wet and dry cases in each group was examined with changes evident over the recent years. The research confirms that STR intensity is more important than STR position in explaining inter-annual rainfall variability across southern Australia but also demonstrates the additional role of STR position in MAM. These results explain the low correlation between rainfall and STR position and why this relationship has evolved during the 20th century as the mean location of the STR has shifted south in MAM. This shows tree-based methods can add additional insight compared to a linear analysis.

The remainder of this chapter is extracted (with some alterations from the original text to correct typographic errors and make referencing consistent) from the publication Whan et al. (2013) from the International Journal of Climatology (Whan et al. 2013).

6.2 Introduction

Since 1997 south-eastern Australia has experienced record low rainfall, exacerbated by record high temperatures (Murphy and Timbal 2008). Several large-scale modes of climate variability have relationships with south-east Australian climate and may therefore have contributed to the recent rainfall deficit, including tropical modes of variability; such as El Niño-Southern Oscillation (ENSO) (Nicholls et al. 1997, Suppiah and Hennessy 1998, Trenberth 1997) and the Indian Ocean (Saji et al. 1999, Simmonds and Rocha 1991, Ummenhofer et al. 2011); high-latitude modes such as the SAM (Marshall 2003, Meneghini et al. 2007) and regional mean sea level pressure indices such as the sub-tropical ridge (STR) (Drosdowsky 2005, Larsen and Nicholls 2009, Timbal and Drosdowsky 2013, Williams and Stone 2009). The spatial and temporal signature of the recent rainfall deficit makes the STR particularly relevant for two reasons. Firstly, the majority of the deficit is evident in MAM (March-April-May) when the tropical modes have little influence (Timbal 2009) and secondly, the spatial extent of the deficit matches well with the area under the influence of the STR (Timbal and Drosdowsky 2013).

The STR is the downward branch of the Hadley Cell; a band of high-pressure that circles the globe at approximately 30° latitude in both hemispheres. The subsiding air associated with this band of high pressure is responsible for arid regions on several continents, including Australia (Peixoto and Oort 1992). Increasing global temperature is associated with changes to the general circulation of the atmosphere including changes to the Hadley Cell. Generally the Hadley Cell has been broadening, leading to a poleward expansion of the subtropical dry zone (Seidel et al. 2007, Zhou et al. 2011). However, there is little scientific consensus about changes to the Hadley Cell, as disagreement exists between reanalysis products and global climate models. A study using precipitation observations supports the results from reanalysis products and points to a poleward shift in the sub-tropical dry zone (Zhou et al. 2011). In addition, regional variability exists in these trends; in particular, the sub-tropical dry zones associated with the downward branch of the Hadley Cell have expanded in regions under the influence of the Asian monsoon. Hemispheric differences are also evident as larger significant trends are found in the Northern Hemisphere, while the Southern Hemisphere displays insignificant trends towards Hadley Cell expansion (Zhou et al. 2011). Changes in the Hadley Cell are associated with changes to the STR; it is therefore imperative that we understand how changes in the characteristics of the STR may affect the hydrometeorology of southeast Australia.

The STR is calculated over eastern Australia (Drosdowsky 2005, Timbal and Drosdowsky 2013) and is closely linked to precipitation variability in south-east Australia (Timbal and Drosdowsky 2013). The seasonal cycle of the STR (Figure 6.1) shows that ridge position is furthest south in late summer (39.6°S in February) and furthest north in late winter (29.1◦ S in August). Climatologically, STR intensity is highest from April to September, reaching a peak of 1022.5hPa in June. Summer experiences the weakest STR intensity with a minimum of 1013.3hPa reached in December. The STR moves north during MAM, with the largest northward shift from March to April as it moves an average of almost 3° over the month. The STR is north of southeast Australia in June-July-August (JJA) and September-October-November (SON) and south of southeast Australia in December-January-February (DJF) and MAM. Thus, the STR is more intense and at lower latitudes during JJA compared with that in MAM. Some differences are evident between the seasonal cycle presented here and that used in other research because different periods are used (Cai et al. 2011a, Drosdowsky 2005).

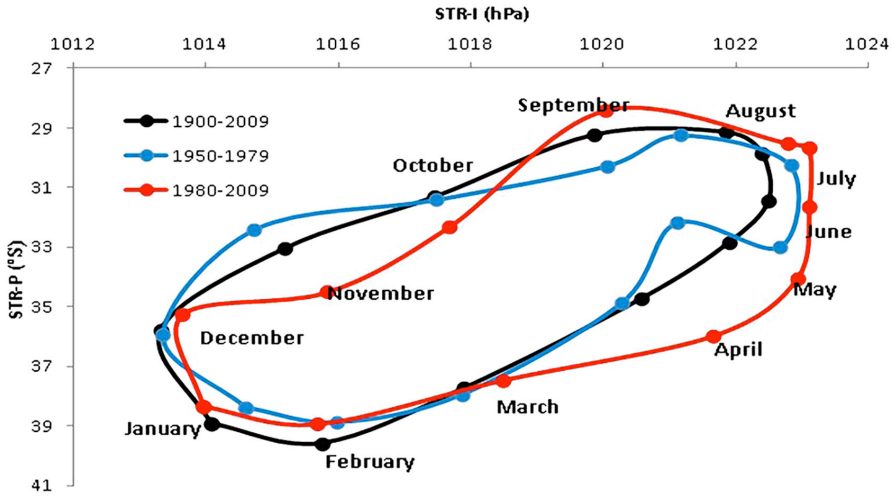

Figure 6.1. The monthly averages showing the annual cycle of mean STR position and intensity. Black – months averaged from 1900 – 2009, blue – months averaged from 1950 -1979, red – months averaged from 1980-2009.

The seasonal north–south movement of the STR plays an important role in determining the climate of southern Australia through changes to the mid-latitude storm track (Murphy and Timbal 2008). When the STR is at higher latitudes in DJF, westerly winds, with embedded frontal activity, are pushed southwards missing continental Australia, while tropical influences may bring precipitation to the region. When the STR is at lower latitudes in JJA, cold fronts are able to pass over southeastern Australia bringing rain to the region. As southeastern Australia receives the majority of its precipitation in JJA, the importance of the meridional movement of the STR is obvious, and research has shown that up to 60% of inter-annual variability in JJA

rainfall is related to movement of the STR (Williams and Stone, 2009). The intensity of the STR is also related to rainfall, as increasing pressure is associated with decreasing rainfall in the sub-tropics (Ansell et al. 2000). This study will expand our understanding of how inter-annual variability in STR intensity and position combine to affect rainfall.

Historically, research has focused on how the variability in STR position influences climate, particularly in the east of Australia (Drosdowsky 2005, Pittock 1973, Williams and Stone 2009). Research into the recent rainfall deficit shows that a majority (62%) of the deficit can be linked to the intensity of the STR (Timbal and Drosdowsky 2013); using simple linear statistics, Timbal and Drosdowsky (2013) were unable to find an effect of STR position on rainfall in addition to the effect of STR intensity. They noted that overall the relationships between the two aspects of the STR are fairly stable over time (as are the relationships between the two facets of the STR and rainfall, in particular for the STR intensity and rainfall relationship). However, other researchers have found changes in the relationship over time (Cai et al. 2011a, Drosdowsky 2005). Cai et al. (2011a) showed that the STR intensity and position relationship has weakened during the last 30 years, particularly in MAM. Increasing understanding on whether the MAM relationship between STR intensity and position is changing, and if so increasing understanding about the nature of that change, is part of the motivation for this paper. An additional motivation is to understand why STR position does not have an effect on rainfall in addition to that of intensity, despite the fact that STR position on its own can be linked to a small amount of the rainfall deficit. The lack of an additional effect attributable to the STR position is even more puzzling when considering that the delay in the northward shift of the STR in the MAM, evident since 1975, is the largest shift, and it is the season when most of the rainfall deficit has been observed record (Drosdowsky 2005, Timbal and Drosdowsky 2013).

Globally, as many climatological parameters display a significant climate shift from the mid-1970s (Trenberth 1990) it is interesting to also see this shift in the STR record. Figure 6.1 shows the shift in STR intensity and position, calculated from gridded monthly mean-sea level pressure data, from the early (1950–1979) to the later (1980–2009) part of the record, with the largest changes found in the transition seasons (MAM, SON). In the latter period the STR is more intense and further south in April, May, June and November, while STR intensity is greater in July and August, compared to the earlier period.

Many authors have raised concerns about the use of linear statistics to describe climate processes that are intrinsically nonlinear (Campbell et al. 2000, Firth et al. 2005, Zorita et al. 1995). Linear statistics are a useful tool to help us understand the climate system but are one tool of many that researchers should employ and have a limited ability to deal with the

complexity of the climate system. Nonlinear techniques should be used in conjunction with traditional methods to help us fully understand relationships and interactions between complex variables, such as precipitation, that have multiple nonlinear associations with highly correlated predictor variables. Classification and regression trees (CART) are one nonlinear strategy that appears to be appropriate for the problem described. CART is able to provide insight into the physical mechanisms underlying statistical relationships with easily interpretable results and is shown to have the greatest advantage over linear techniques when applied to data that is nonlinear (Rodionov et al. 2001).

The importance of the STR for southern Australian rainfall is clear. Changes in STR intensity have a verified effect on precipitation (Ansell et al. 2000, Larsen and Nicholls 2009, Murphy and Timbal 2008, Timbal and Drosdowsky 2013), yet the role of the position is less well understood. This study will explore what effect inter-annual variability in STR position has, on southeast Australian rainfall (in addition to that of intensity), using both linear and nonlinear statistical techniques. The next section will describe the datasets and methods used, including an introduction to the CART method. The following sections will present the results from both the linear and CART analyses. Then, composites of precipitation and vector wind are used to describe the atmospheric circulation associated with particular phases of the STR, based on results of the CART analysis. Finally, results are discussed further and conclusions are drawn on how STR intensity and position are associated with precipitation in Australia.

6.3 Data and Methods

6.3.1 Data sets

This study relies on Australia-wide precipitation data sourced from the Australian Water Availability (AWAP) dataset (Jones et al. 2009) from 1900 to 2009. The AWAP dataset interpolates station data and is available on a 0.05° grid for the Australian continent. A precipitation time series for southwest eastern Australia (SEA) was calculated by seasonally averaging values within the hatched region in Figure 6.2. This region is the area where the STR influence is the greatest as STR intensity explains over 20% of the rainfall variability. SEA approximately covers the area to the south-west of a line running from Melbourne to the south of the Flinders ranges and following the end of the Great Dividing Range (GDR) over western Victoria (see (Timbal et al. 2010) for detail about the construction of the SEA time series). The 110-year SEA precipitation time series was divided into three groups, with 37 cases in the dry and average groups and 36 cases in the wet group.

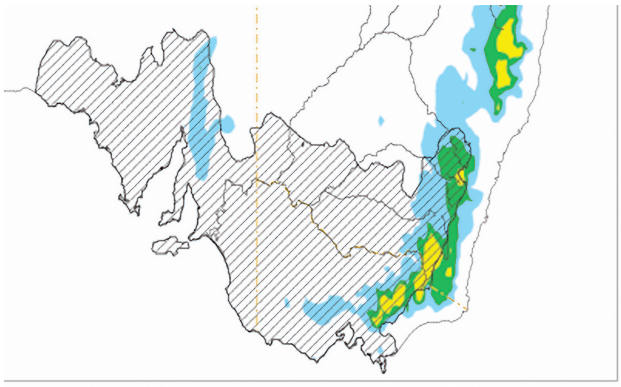

Figure 6.2. Map of southeast Australia. The hatched area defines the southwest of eastern Australia (SEA) that is area-averaged and used in this study. Colour shading indicates topography, with the highest peaks in yellow.

The STR intensity and position indices are calculated as per the station-based 'L-index' in Drosdowsky (2005). The indices are based on monthly mean-sea level pressure station data between 10 and 44°S along the eastern Australian coast. The station data is interpolated to 1° grid and averaged within a 5° longitude band around 150°E. The location of the maximum in pressure within that band gives the position of the STR (with 1° accuracy), as well as the intensity of that maximum. This index is simple to calculate, robust and consistent over time (Drosdowsky 2005). The use of station data means that the index can be calculated as far back as 1890; however, in this study we used data from 1900 to 2009 to match the available gridded precipitation data. ENSO years are obtained from the Bureau of Meteorology list of 25 ENSO years (Table 6.1) that have occurred since 1900 (Australian Bureau of Meteorology 2010).

Table 6.1. ENSO years (from Bureau of Meteorology, 2010).

ENSO years
1902–1903, 1905–1906, 1911–1912, 1913–1914, 1914–1915, 1919–1920, 1925–1926, 1940–1941, 1941–1942, 1946–1947, 1951–1952, 1957–1958, 1963–1964, 1965–1966, 1972–1973, 1977–1978, 1982–1983, 1987–1988, 1991–1992, 1993–1994, 1994–1995, 1997–1998, 2002–2003, 2006–2007, 2009–2010

6.3.2 Methods

The first analysis is a linear examination of SEA precipitation and the two facets of the STR. Correlations (significant at the 5 percent level with a two-tailed Student's t-test) were conducted between the SEA area-averaged time series and indices of STR intensity and position, followed by spatial correlation analysis between continental gridded Australian rainfall (0.05°) with both facets of the STR. Statistical significance is defined throughout this research

at the 5 percent level, which based on 110-years of data means correlations that exceed 0.2 are significant (and are thus coloured in Figure 6.4 and Figure 6.5). A decrease in the degrees of freedom stemming from persistence in Australian rainfall (Simmonds and Hope 1997) has not been included in the significance testing.

To explore the non-stationarity of the relationship between the STR and rainfall, and their spatial features, spatial correlation maps were produced for two 50-year epochs (Epoch 1: 1900–1949 and Epoch 2: 1960–2009) to contrast how these relationships have evolved during the 20th century but using long enough periods to ensure results are statistically significant. The same significance level is also used in the epoch analysis, despite the smaller sample size (which would mean that correlations significant at the 5 percent level exceed 0.28, rather than the 0.20 shown here). A common significance level is used here for both phases of the analysis because the purpose is to show changes in the relationships over time; hence to emphasize these temporal changes, the significance level is kept constant throughout the whole research. As per the previous discussion, some autocorrelation may exist in the JJA series so there may be some reduction in their significance. However, the focus is on MAM and the relative changes in the epochs.

Previous research has shown the stability of STR-rainfall relationships in SEA over the instrumental record (Timbal and Drosdowsky 2013), which gives confidence in the correlations shown in this region, although the same confidence cannot be given for those correlations outside the region (e.g. Queensland).

Next is the CART analysis, which uses area-averaged SEA rainfall time series as a categorical response variable (the three groups are dry, average and wet), with STR position and intensity as predictor variables. The cases that comprise each terminal node are used in a composite anomaly analysis of the AWAP precipitation data with anomalies calculated against the full record. The 20 most recent cases in each terminal node that occur after 1948 are also used in a composite anomaly of vector wind using the NCEP/NCAR Reanalysis product (Kalnay et al. 1996).

CART is a binary recursive partitioning technique first developed in the 1980s (Breiman et al. 1984). CART has been employed in many fields including ecology and genetics (De'ath and Fabricius 2000, Esther et al. 2010, Thuiller et al. 2003), with some applications in climatology (Burrows et al. 1995, Firth et al. 2005, Rodionov et al. 2001, Rodionov et al. 2005). Decision trees are constructed that seek to describe the variability of one response variable by several predictor variables (De'ath and Fabricius 2000). They are particularly useful when there are multiple correlated predictor variables, such as is the case with STR intensity and position in both MAM and JJA ($r = 0.49$ and 0.70, respectively). Trees may be constructed using either

continuous (regression trees) or categorical (classification trees) response data; in this research classification trees are used (illustrative example in Figure 6.3). In the example, a response variable (n = 30) containing three classes (A, B and C) is classified on the basis of two predictor variables (1 and 2). This can be compared to the current research in which the response variable used is SEA precipitation (n = 110), which is divided into terciles (dry, average and wet groups), and classified on the basis of two predictor variables (STR intensity and position). The first parent node (Node 1), in the example, contains 30 cases of the three classes of response variable each with ten members. The daughter node to the left is named by doubling the parent node (i.e. Node 2), while the right-hand daughter node is named by doubling the parent node and adding one (i.e. Node 3). Splits are selected on the basis of predictor variables and continue until a terminal node (a node that has no further splits). This is can be seen in the example where the original parent node is 'Node 1' which is split into two daughter nodes (Terminal Nodes 2 and 3) on the basis of 'Predictor variable 1'. When the splitting criterion is met (i.e. 'Predictor variable 1' is > x) the cases of the response variable are put into the left node. If the criterion is not met (i.e. 'Predictor variable 1' is ≤ x) the cases are put into the right node.

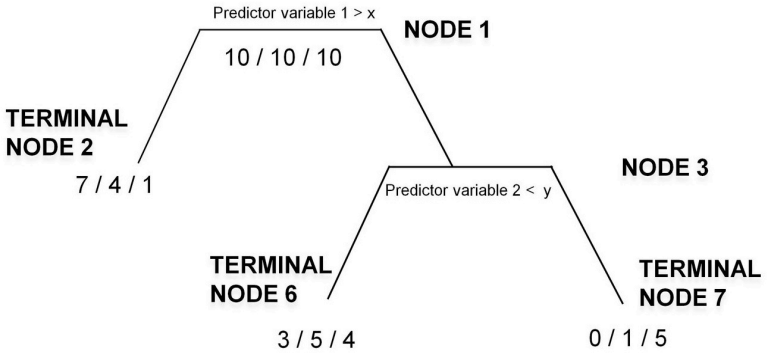

Figure 6.3. Classification tree example. Predictand data = three groups each with ten members (n = 30), which is classified on the basis of two predictor variables (predictor variables 1 and 2).

Increasingly homogenous groups are created as parent nodes are split into daughter nodes on the basis of specific values of predictor variables. Splits are selected to maximize the increase in node purity, as measured by the Gini index. The Gini index measures node impurity and is calculated from of the number of cases belonging to each class in the node. The index is calculated by subtracting the sum of the squared probability of each class' occurrence in the node from one (Equation (1)). The Gini index, where j is the class and p_j is the probability of class j:

$$\text{Gini index} = 1 - \sum_j p_j^2 \qquad (1)$$

This means that when a node contains only one class the Gini index is zero, while when there are equal numbers of all classes the index is one. In the example, Node 1 has equal numbers of all classes so the Gini index is at its maximum value of one as node impurity is at its maximum. In Terminal Node 7 the Gini index is 0.26 showing that it is lower when node purity is higher.

Complex trees can be grown that maximize node purity by allocating every case to a separate terminal node, but such an overfit tree is unlikely to be replicable with new datasets, limits interpretability and has little practical significance. The result then is a trade-off between node purity and tree size or interpretability. An overly large tree is created and then pruned back based on an arbitrary decision, although in this research tree size is selected to minimize the cross-validated error and maintain physical consistency. In the cross-validation process, the dataset is randomly divided into subsets. To each subset of data a training/test sample approach is applied to estimate misclassification rates in sub-trees. From this, a complexity parameter is chosen that minimizes misclassification error and is used to select the optimal tree size. This enables tree construction that is both statistically robust and physically consistent. It is important to recognize that when looking at SEA rainfall using only STR intensity and position as predictor variables the highest values of node purity are not expected as this would fail to recognize the other modes of climate variability that are at play and are not being captured using only these two predictors. High node purity is easily achievable when tree size is large, but in this case the interpretability of the tree is often compromised. When tree interpretability is high, node purity is often conceded as the tree is pruned back to a smaller size more manageable size. Consequently, terminal nodes in classification trees are rarely completely pure (containing only one class of data) and are more regularly comprised of multiple data classes, as we would expect in climatic research using only two predictor variables to describe a complex variable such as precipitation. CART uses a ‘majority rule’ to allocate each terminal node to a class, so that the terminal node is allocated to the class to which a majority of its member belong. For example, in Terminal Node 2 there are seven cases in class A, four cases in class B and one case in class C. This means that this terminal node will be allocated as ‘class A’ and would be 58% pure as 7/12 cases belong to the class employed by the majority rule. CART is robust to outliers and extreme values among predictor variables as the technique tries to put divergent cases in separate nodes (Rodionov et al. 2001).

In the current study decision trees were constructed using the statistical package R (R Development Core Team 2012), with the library 'rpart' (Therneau et al. 2010).

6.4 Results

6.4.1 Results from linear statistics

Over the full record, negative correlations between SEA precipitation and STR intensity and position (Table 6.2) show that as the intensity of the STR increases and as the STR moves further south, precipitation in SEA decreases.

Table 6.2. Correlation between precipitation in SEA and STR intensity and position in MAM and JJA.

	STRI	STRP
MAM	−0.46	−0.19
JJA	−0.69	−0.53

The correlation with STR intensity is stronger and more significant than that with STR position, particularly in MAM. This confirms previous linear research that found intensity is the most important aspect of the STR to explain SEA rainfall variability (Timbal and Drosdowsky 2013), and is in general agreement with the known relationship between pressure and rainfall in the subtropics (Ansell et al. 2000). The spatial correlation analysis of STR intensity and position with gridded precipitation is shown in Figure 6.4.

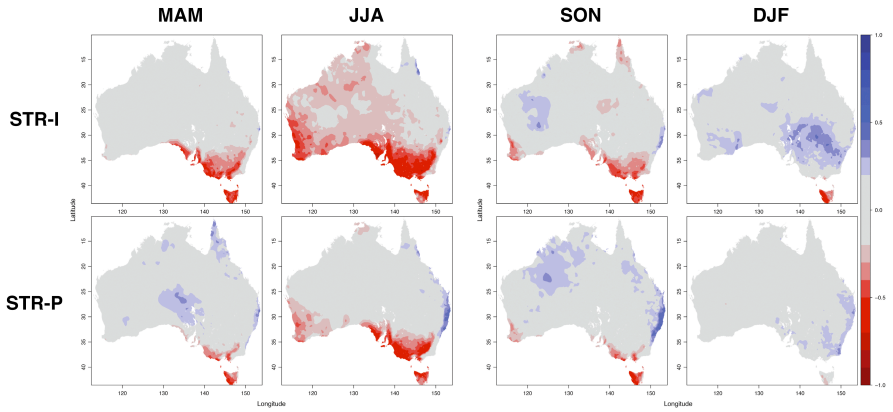

Figure 6.4. Seasonal correlation coefficients between STR intensity (top row) and position (bottom row) and Australian rainfall in MAM, JJA, SON and DJF over the period 1900 to 2009. Shaded areas show significant correlations at the 95% confidence level. Refer to Figure 6.2 for the SEA region.

The seasonal cycle of both facets of the STR and SEA rainfall are closely linked. Correlations strengthen during MAM and reach a maximum in JJA, before decaying throughout

SON to a minimum in DJF. In MAM, the maximum correlation between rainfall and intensity (position) is -0.65 (-0.60). As the relationship develops into the JJA months, the maximum correlation between rainfall and intensity (position) is -0.74 (-0.68), again showing the dominance of intensity over position. The spatial extent of significant correlations for both intensity and position is the greatest in JJA, and larger for intensity than position. In MAM, all significant correlations with intensity are negative, so increases in STR intensity is associated with decreased precipitation (or no impact when the correlations are not significant). STR position, however, has a more spatially variable relationship with precipitation; negative correlations are evident in SEA but positive correlations are found further north. This shows that the relationship with position depends on the mean location of the STR, which in MAM varies between 37°S and 32°S, in between the areas with significant positive and negative correlations. When the mean position of the STR is to the north (south) of the area considered, a shift south in STR position will be associated with decreased (increased) rainfall. The mean location of the STR in relation to SEA changes with the annual cycle (Figure 6.1), and also on a multi-decadal timescale (see later this section), as such, the complexities of the STR position –rainfall relationship are difficult to grasp with linear statistics. This is particularly relevant in MAM when the mean position of the STR is located above SEA and is changing rapidly from a southerly DJF position to a northerly position in JJA. The high variability found in the STR position record adds an additional layer of complexity to the issue and results in few significant trends (Timbal and Drosdowsky 2013). However, this mechanism shows that subtle variations in the mean position of the STR are important for precipitation despite a lack of statistical significance found with linear statistics.

Temporal changes are found in the relationship between STR position and Australian rainfall over the instrumental record (Drosdowsky 2005, Timbal and Drosdowsky 2013). Timbal and Drosdowsky (2013) found variation in correlation coefficients (calculated over 30-year periods between 1929 and 2009) for relationships between SEA rainfall and both STR intensity and position. They found less multi-decadal variability in the correlation between SEA rainfall and STR intensity, with the largest changes evident in May. The relationship between STR position and rainfall, however, shows more variability between the two 30-year epochs (Timbal and Drosdowsky 2013). Timbal and Drosdowsky (2013) found that in July correlations increase from -0.2 in the 1940s to -0.6 to -0.8 in the most recent period. In SEA, March and April had negative correlations during the early period, while the end of the record shows positive correlations in these months. These results outline the complexity of the relationship between the STR position and precipitation, particularly in MAM, the vital transition season between DJF and JJA patterns. Variability of the STR position is larger in JJA compared with

that in MAM (standard deviation = 2.3° of latitude compared to 1.8° of latitude). Yet the mean position of the STR in MAM is key, as the STR can be located either across or north or south of SEA in this season, with the nature of the relationship depending on the position of the STR. In the current research the epoch analysis (Figure 6.5) was conducted here to extend the understanding of STR non-stationarity. Looking firstly at the relationship between rainfall and STR intensity, the negative correlation is stronger and more extensive in Epoch 2, and when correlations are significant in both epochs, magnitudes are also greater in the later epoch. Specifically, in Epoch 2 significant correlations between STR intensity and rainfall are evident across all of New South Wales (NSW), the majority of Queensland and into the Northern Territory; compared with those in Epoch 1, where the significant correlations are confined to SEA and Tasmania. The multi-decadal differences associated with STR intensity in MAM include a reversal in the sign of the relationship as positive correlations present during Epoch 1 in Queensland are replaced by negative correlations in Epoch 2.

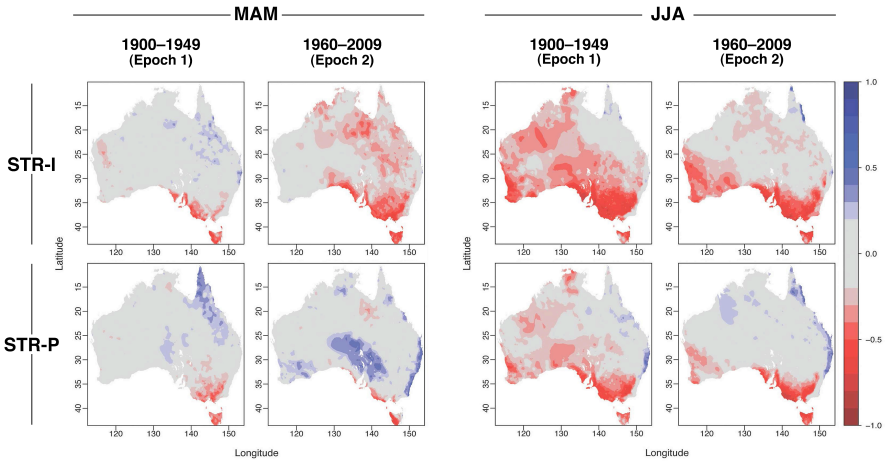

Figure 6.5. Correlation coefficients between STR intensity (top row) and position (bottom row) with rainfall in MAM (left two columns) and JJA (right two columns) over two 50-year epochs: 1900 – 1949 (Epoch 1) and 1960 – 2009 (Epoch 2). Shaded areas show significant correlations with 5% significance level. Refer to Figure 6.2 for the SEA region.

While the 20th century has seen an increase in the spatial extent and magnitude of the correlation between STR intensity and rainfall in MAM, the opposite is true in JJA. This season has experienced a reduction in the spatial extent of the correlation between STR intensity and precipitation from Epoch 1 to Epoch 2. However, the spatial reduction is not associated with a reduction in magnitude of the correlation over southern Australia where the relationship remains significant (in contrast to the increases in correlation over SEA in MAM). Epoch 1 displays a significant negative relationship over the majority of Australia (excluding Queensland), whereas in Epoch 2 the significant correlations are confined to southern Australia. This may be related to

a small long-term southward trend in the position of the STR over the 20th century (Timbal and Drosdowsky 2013). Although Timbal and Drosdowsky (2013) found the strength of the STR intensity correlation in SEA to be relatively steady over the 20th century, this analysis shows that there are spatial differences in the extent of the relationship (away from SEA) in both MAM and JJA.

Secondly, the relationship between STR position and precipitation has also changed between the two epochs. SEA exhibits a reversal in the sign of the correlation between position and rainfall over the 20th century in some months (Timbal and Drosdowsky 2013), which is evident in the spatial correlation maps. In MAM during Epoch 1, there is a region of positive correlation in north Queensland with negative correlation covering most of SEA. The region of positive correlation shifted south and stretches east from South Australia down the east coast of NSW, while the areas with negative correlation shrink and are confined to a small fringe along the coast from South Australia to Victoria and western Tasmania in Epoch 2. During JJA, the spatial extent of the correlation decreases over the 20th century, while the magnitude and sign of the correlation over SEA are broadly unchanged. The key features are a negative relationship dominating the SEA region and extending along the coast to Western Australia and a positive relationship on the east coast.

This section of the analysis highlights the importance of utilizing long records in research, as even here with 50-year epochs the non-stationarity of the series is evident. Different conclusions can be drawn when a smaller or longer subset of the record is selected for the analysis, although this can highlight how relationships change over time. Cai et al. (2011a) concluded that MAM rainfall cannot be linked to the STR as only insignificant correlations (-0.35 and 0.22 for intensity and position respectively) can be found in this season. This difference stems from using a smaller subset of the dataset, although when longer time periods are examined the relationship is significant (Table 6.2). The positive correlation (Cai et al. 2011a) found between STR position and SEA rainfall from 1979 to 2008 reflects the shifting sign of the relationship found by Timbal and Drosdowsky (2013) and shown spatially in Figure 6.5.

It is worth noting that for SEA overall in JJA the evolution of the relationship between STR intensity and position with rainfall between Epoch 1 and Epoch 2 is broadly similar, with the biggest changes evident further north over inland Australia with little change across SEA. This suggests that while the STR has shifted south (Timbal and Drosdowsky 2013), it has not shifted south enough to change the relationship with rainfall over the southern parts of the continent because of the northerly mean position of the STR in this season. Most changes are evident further north where a shift south of the STR has had sizeable impact. More interesting

are the changes in the relationships, outlined previously, between rainfall and both facets of the STR in MAM; most notably, a strengthening of the relationship with intensity over most of eastern Australia including SEA and a contraction of the negative influence of the STR position to a very small part of SEA in conjunction with the increase in the area with a positive relationship. These results illustrate that the relationship between the position of the STR and rainfall is highly complex and changes over time depending on the mean location of the STR.

The association between the STR and SEA climatic variability is therefore evidently nonlinear with the impact dependent on the background climatology (i.e. the difference in the mean between MAM and JJA). The current linear analysis confirms and extends recent research and gives a good understanding of how the STR varies throughout the year and how it has changed over the instrumental record. Linear analysis has confirmed the dominance of STR intensity (compared with position) in controlling SEA precipitation but does not provide a clear picture about the interaction between the two aspects of the STR. Therefore, while linear statistics are useful in basic description and analysis of such nonlinear processes, it is essential that more complex statistics are employed to fully understand STR interactions. We will now use CART to explore the nonlinear interactions between intensity, position and SEA rainfall.

6.4.2 Results from CART analysis

6.4.2.a MAM

In MAM, the CART procedure identifies four terminal nodes, with both predictor variables required to adequately classify the data (Figure 6.6). The first split uses STR intensity to separate 40/110 cases, confirming the predominance of STR intensity from the linear analysis. This split defines the dry Terminal Node 2, when high STR intensity ≥ 1021hPa. Node purity is 60% as high pressure is associated with low precipitation in nearly 2/3 of the cases, a high percentage when considering the sample size (24 out of 37 dry years are identified solely on STR intensity being above 1021hPa). Composite rainfall anomaly analysis of Terminal Node 2 (Figure 6.7) shows the spatial impact on precipitation as this node is characterized by negative precipitation anomalies over much of SEA, with the largest negative anomalies are located in the areas of largest mean rainfall (western Tasmania and on the Southern Alps). High pressure is associated with a reduction in the mean westerly flow (Timbal and Drosdowsky 2013) and the effects of this are clearly seen in the spatial pattern emerging from the composite anomaly map.

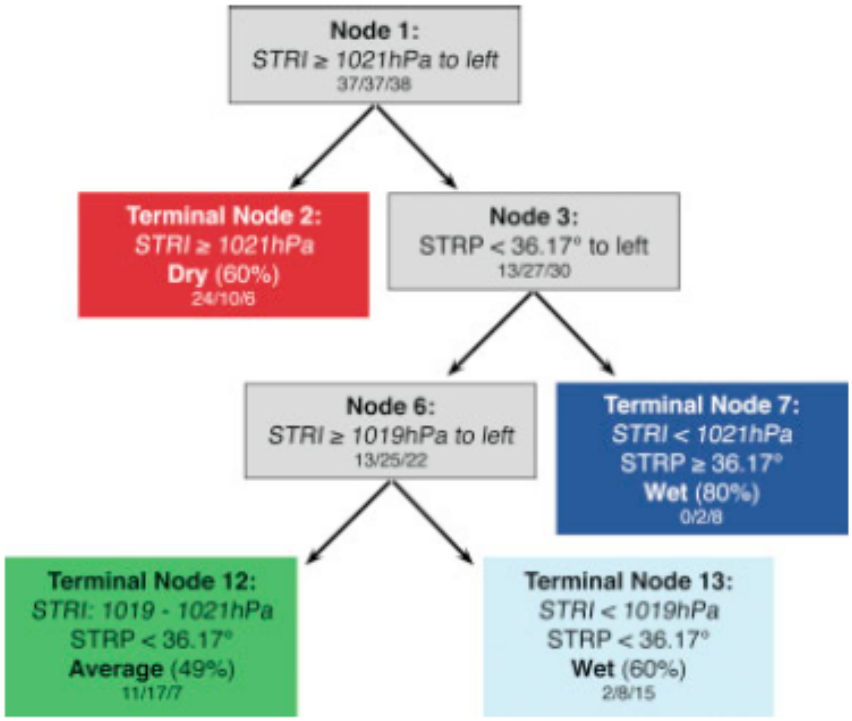

Figure 6.6.Decision tree from the CART analysis in MAM after pruning to four terminal nodes.

The first split supports the proposition that STR intensity is more important than position in determining drought in SEA (Timbal and Drosdowsky 2013); however, the position of the STR is necessary to understand the full range of precipitation variability as the second split (for cases below 1021hPa) is made on position. When intensity is below 1021hPa and the position of the STR is further south (≥ 36.17°) the CART technique classifies Terminal Node 7 as a wet node with very high purity; ten cases during 110 years are classified in this terminal node with 80% belonging to the wet class and no dry cases classified into this group. The composite analysis (Figure 6.7) of Terminal Node 7 displays positive rainfall anomalies over the whole of SEA including the northern inland regions with very high anomalies on the eastern coast. However, this very wet node displays a strip of negative rainfall anomaly over the southwest coast of eastern Australia and Tasmania with the largest negative anomalies on the western Tasmanian coast. This pattern of negative and positive rainfall anomalies is consistent with tropical penetration of moisture from northern Australia and is reminiscent of a 'summer-like' rainfall pattern, as opposed to the 'winter-like' rainfall anomalies where one would expect to see largest rainfall anomalies along the southern edge of the continent.

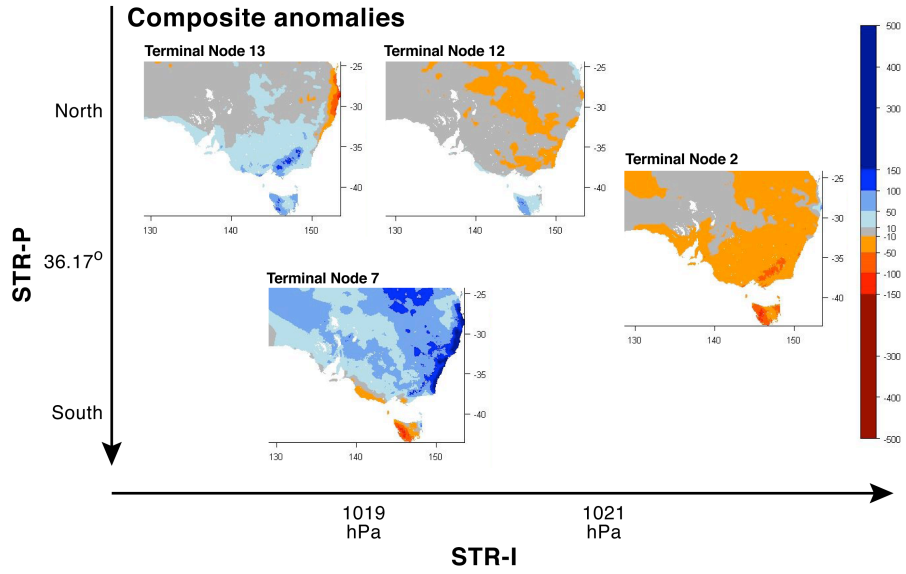

Figure 6.7. Seasonal precipitation composite anomaly maps (mm/year) for a large part of the Australian continent encompassing SEA in MAM. Composites are based on seasonal means from 1900-2009.

When the position of the STR is in a more northerly position (< 36.17°) the third split is based on intensity. When the STR intensity is low (< 1019hPa) and the STR is in a northerly position a wet MAM occurs (Terminal Node 13). The composite seasonal precipitation anomaly for this terminal node (Figure 6.7) shows that positive precipitation anomalies confined to SEA, with the largest positive anomalies along the Southern Alps and in western Tasmania while negative anomalies are present in this node along the eastern coast of NSW. Spatially rainfall is consistent with an increase westerly flow and is therefore more 'winter-like'. Although Terminal Nodes 7 and 13 are both classified as wet nodes, spatially the precipitation between these two nodes are very different. In Terminal Node 7 wet conditions occur when the STR is in a southerly position (≥ 36.17°) even when STR intensity is above average (up to 1021hPa while the MAM mean is below 1020hPa), while in Terminal Node13 wet 'winter-like' conditions occur only when STR intensity is below average and the STR is in a more northerly position (< 36.17°). These two terminal nodes show two of the possible positions of the STR in the transition season.

Finally, when the STR is at low latitudes and STR intensity is at median levels (between 1019 – 1021hPa) an average terminal node is defined. Terminal Node 12 is a highly heterogeneous node with 11 dry cases, 17 average cases and 8 wet cases. The low node purity clearly indicates that this final node is poorly determined by both STR intensity and position, and hence in these cases another variable is likely to be the primary driver of the rainfall variability. This is logical because it corresponds to years with near average STR intensity and position. As expected, the rainfall composite map for Terminal Node 12 does not show spatially

consistent rainfall anomalies across the SEA region although further south, in Tasmania, positive anomalies are similar to the rainfall regime associated with Terminal Node 13. The low latitude of the STR in this node is only evident in these Tasmanian anomalies while across the continent rainfall anomalies are small and patchy.

The years identified in each terminal node can give additional information about variability in rainfall regimes. Figure 6.8 shows the time series for SEA rainfall in MAM (vertical bars), the terciles that the data was split into for the CART analysis (horizontal lines) and the terminal nodes each year was classified into (colours of vertical bars). It is clear that the recent MAM-dominated rainfall deficit is associated with STR intensity. Eleven of fourteen years in the last two decades classified in Terminal Node 2 are dry years. This supports the research that found increased STR intensity to be the primary driver of the recent rainfall deficit, particularly in MAM (Timbal and Drosdowsky 2013) as high pressure is associated with lower rainfall owing to increased stability of the air column and a weakening of the mean westerly airstream.

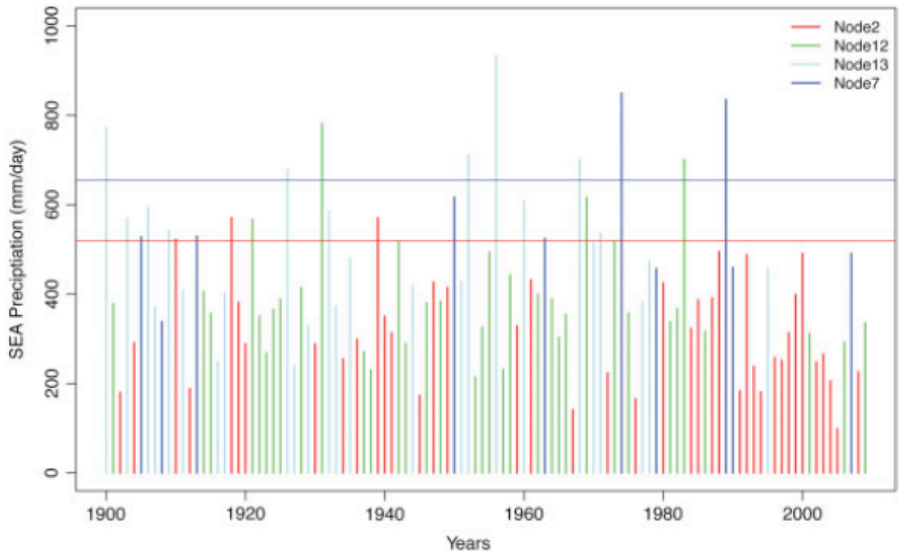

Figure 6.8.SEA MAM precipitation time series. Colour of the vertical bars indicates the terminal nodes the year has been classified into. The red and blue horizontal lines indicate the cut-off points for the original classification (dry, average and wet), giving an indication of misclassification when the horizontal and vertical classifications do not match.

The 'summer-like' rainfall pattern associated with Terminal Node 7 has been a rare event over the instrumental record, with only ten cases recorded. The low number prevents an in-depth analysis but it can be noted that while only four cases were recorded by the first half of the record (1900–1955), six were recorded in the second half, with cases observed in every decade and hence are a continuing occurrence. On the contrary, while Terminal Node 13 has been more common over the entire record (25 cases recorded), in the last 30 years has only seen

this ‘winter-like’ pattern occur once in 1995, while the ‘summer-like’ wet pattern was observed three times (1989, 1990 and 2007).

Overall, increasing STR intensity over the latter part of the 20th century has seen an increase in the occurrence of the dry Terminal Node 2, a near disappearance of the ‘winter-like’ wet Terminal Node 13 in the last 30 years and a small increase of the rare occurrences of ‘summer-like’ wet MAM.

6.4.2.b JJA

In JJA, the first split in the CART model is on intensity (Figure 6.9), separating a dry node (Terminal Node 2) with 49/100 years (with 65% purity) when intensity is high (≥ 1022hPa); this is very similar to MAM. The higher threshold value in JJA compared with that in MAM is consistent with the higher climatological value as shown in the annual cycle (Figure 6.1). The increased node purity in JJA’s Terminal Node 2 compared with MAM is consistent with the higher correlation between STR intensity and rainfall in JJA compared with that in MAM. The composite rainfall anomaly analysis (Figure 6.10) for this high intensity group (Terminal Node 2) displays a broad region of negative precipitation anomalies over SEA, with the largest anomalies on the mountain regions and western Tasmania, combined with a small region along the eastern seaboard that displays a positive precipitation anomaly. Again, for SEA, this is very similar to MAM. The main difference between the two seasons is that in JJA the second split is also based on intensity (<1020hPa), with position counting only in a third split between Terminal Nodes 12 and 13 (Figure 6.9). Terminal Node 7 is classified as wet and is 100% pure but is a rare occurrence (6/110 years). The composite anomaly pattern of this low intensity wet terminal node has a spatial pattern that is reminiscent of the ‘winter-like’ Terminal Node 13 in MAM with the largest positive anomalies along the Alps and in the west coast of Tasmania and the mainland with negative anomaly found on the eastern seaboard, i.e. east of the GDR. The rainfall patterns of Terminal Nodes 2 and 7, which are mirror images suggest a linear relationship between intensity and precipitation as high or low pressure (either ≥1022 or <1020hPa respectively) is associated with high and low precipitation extremes.

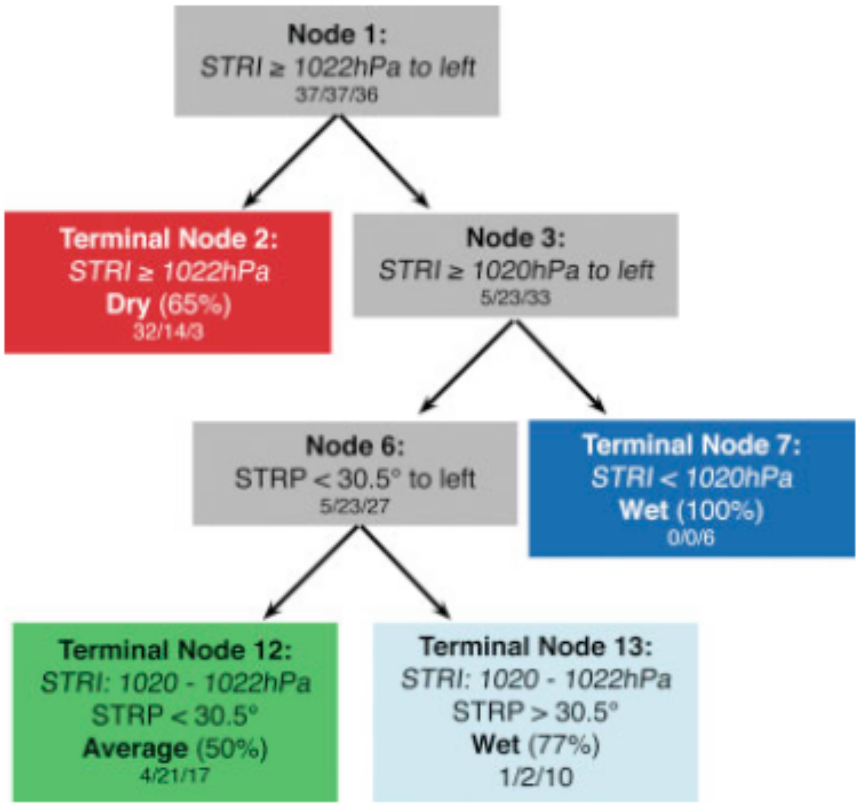

Figure 6.9. Decision tree from the CART analysis in JJA after pruning to four terminal nodes.

The rainfall pattern emerging from Terminal node 12, highly heterogeneous modes with many average years as was the case in MAM and more wet years than dry years and a rainfall pattern tending towards wetter across SEA but half way between wet Terminal Node 7 and dry Terminal Node 2.

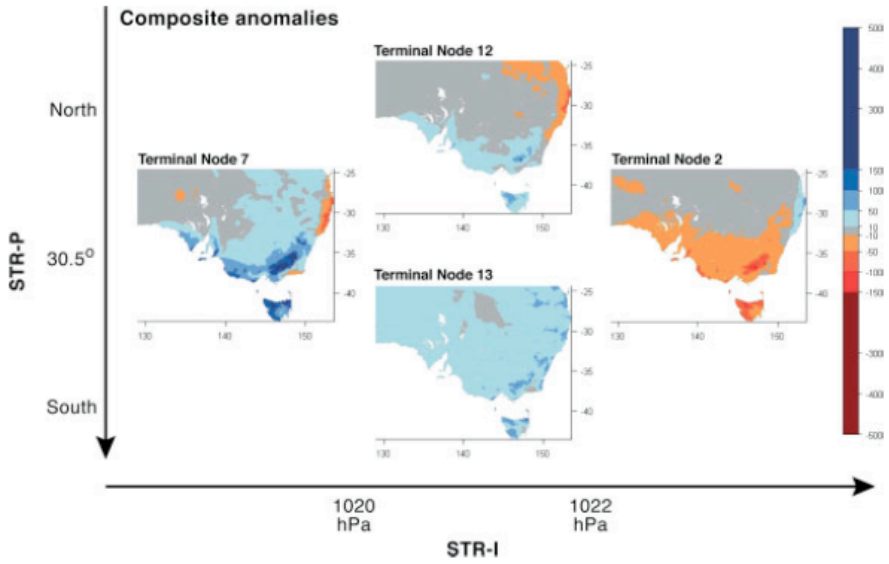

Figure 6.10. Seasonal precipitation composite anomaly maps (mm/year) for a large part of the Australian continent encompassing SEA in JJA. Composites are based on seasonal means from 1900-2009.

An interesting result of the JJA CART model is that when STR intensity is fairly neutral (between 1020 and 1022hPa), the position of the STR is identified as being important. When the position of the STR is at further south (latitudes > 30.5°) the result is Terminal Node 13, a wet node with high purity (77%). Although, the CART model was constructed using anomalies across SEA and the rainfall anomalies across SEA are very similar between Terminal Nodes 12 and 13, the most visible differences on the composite anomaly maps are located further north. Negative rainfall anomalies are evident across eastern Australia when the STR is further north (Terminal Node 12), but positive anomalies extend across all eastern Australia when the STR is located further south (Terminal Node 13). This difference between the two nodes illustrates that the two facets of the STR (intensity and position) are not independent; a ridge of moderate intensity will prevent rainfall bearing westerly front to penetrate inland is located north, but this effect disappears when the STR is displaced further south. Although it is also very likely that in those years with neutral STR intensity, neither STR position nor intensity are the major forcing and other large-scale modes of climate variability (possibly due to ENSO as the effects start to be felt in JJA) are important in these terminal nodes.

The years classified in each terminal node in JJA are noteworthy (Figure 6.11). The enhanced 'winter-like' group (Terminal Node 7) is rare throughout the instrumental record and has not occurred since 1981, making the last 28 years the longest stretch without an occurrence of this node. Also of interest is the disappearance of the wet Terminal Node 13, last observed in 1978. There is no other 30-year period, during the instrumental record, when this wet node has not occurred.

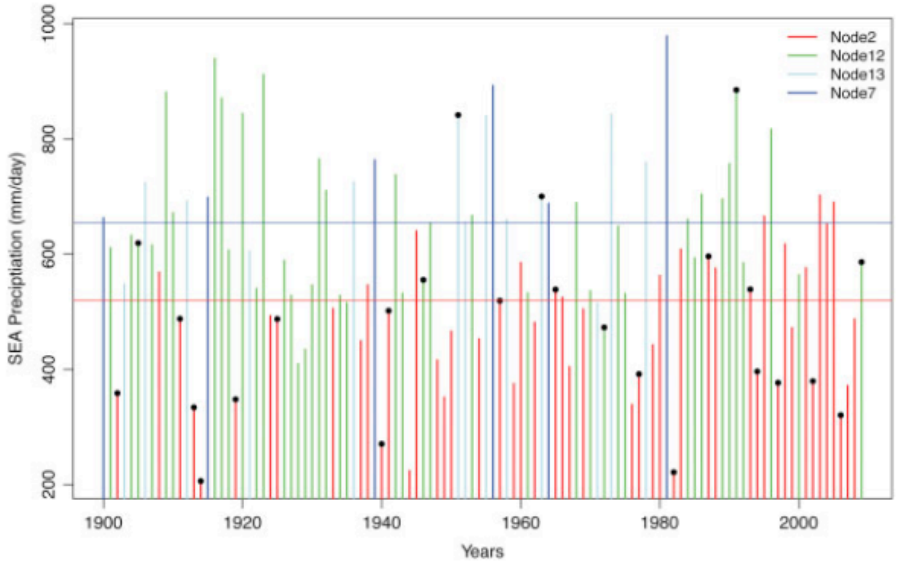

Figure 6.11. SEA JJA precipitation time series with ENSO years marked by dots. Colour of the vertical bars indicates the terminal nodes the year has been classified into. The red and blue horizontal lines indicate the cut-off points for the original classification (dry, average and wet), giving an indication of misclassification when the horizontal and vertical classifications do not match.

The signal of the recent drought is found primarily in MAM (Murphy and Timbal 2008), but there is evidence of a growing importance of the JJA rainfall deficit (Timbal 2009). In the last two decades 7 years are classed as dry, with all of these dry years classified on the basis of high STR intensity into Terminal Node 2. However, this is not unprecedented as many periods in the past have large numbers of dry years classified on the basis of high STR intensity. Although the node purity in Terminal Node 2 is very high, three wet years appear in the dry terminal node. Interestingly, for over 90 years no wet years were found when STR intensity was high yet the three wet instances have occurred in the last 15 years – 1995, 2003 and 2005. This could be purely random as it is a very small number, but it is also interesting to note an increase in the occurrence of average rainfall years when STR intensity is high (Terminal Node 2) from 3 cases in the first half (1900–1955) of the record to 11 cases in the second half. Precipitation in SEA during these 3 years is much higher than other cases in the same terminal node (1.82 mm/day compared with 0.92 mm/day across the remainder of Terminal Node 2), which may be because of a warming of sea surface temperature (SST) surrounding Australia (Casey and Cornillon, 2001). The rainfall for these three wet years in Terminal Node 2 was analysed further. The spatial analysis reveals that the positive precipitation anomalies for the three wet years in SEA are widespread, with negative anomalies along the eastern coastal strip. The intensity of the STR is slightly lower on average in these wet cases (1023.8hPa rather than 1024.1hPa for Terminal Node 2); however, the STR is 2.4° further north during these 3 years

(29.3°S compared with 31.7°S for Terminal Node 2). An analysis of the moisture transport (not shown) points to the possibility that during these years, despite the northerly location of the STR, anomalous moisture flux originating from the Coral Sea (150° – 170°E, 5°S–25°N) reaches SEA from the north and may contribute to reduce easterly flow along the NSW coast (in agreement with the rainfall anomalies). This exploratory analysis is far from definitive and further research is required to confirm this hypothesis.

6.4.3 Interaction with ENSO

As ENSO events develop during JJA we may expect to see some interaction between ENSO and the STR in this season that is not evident in MAM. In the MAM analysis ENSO events are evenly spaced throughout the four terminal nodes as expected as there is little ENSO signal during this season (not shown). During JJA, however, El Nño events are much more tightly grouped with 18/25 events occurring in Terminal Node 2 (dots in Figure 6.11). Of the 18 El Niño events classified into this dry terminal node, all show average (3 years) or below average (15 years) rainfall. Interestingly, the remainder of the ENSO years classified into other terminal nodes are more evenly distributed with a counter-intuitive tendency towards being wet: one dry, three average and three wet El Niño years. This analysis suggests a strong interplay between ENSO and the STR in JJA. This supports the conclusions of Cai et al. (2011a) that found a strong relationship between tropical variability (expressed through the Indian Ocean Dipole) and the STR in this season. Indeed ENSO variability contributes to STR variability as it can be seen that most of the El Niño years end up having above average STR intensity (Terminal Node 2). In addition, STR variability appears to be able to modulate El Niño impact on SEA rainfall: El Niño years classified in other nodes are a mix of wet, dry and normal years.

6.4.4 Vector wind analysis

Further analysis of the climatology of the terminal nodes is performed by looking at lower tropospheric wind in both MAM and JJA. The composite vector wind analysis for the years between 1948 and 2009 identified in each terminal node identified in the CART analysis can be seen in Figure 6.12 for MAM and Figure 6.13 for JJA. Over the period 1948–2009, the mean flow over the study region is westerly airflow of between 2 and 6m/s in MAM, picking up to 4–8 m/s in JJA, and intruding further over the Australian continent in JJA compared with MAM. In the dry Terminal Nodes of both MAM and JJA (both Terminal Node 2), the anomalous flow over the study region is south-easterly (MAM) to easterly (JJA) as part of a broad anti-cyclonic regime over the study region (with corresponding high rainfall in JJA along the east coast at the

location of the strongest easterly anomalies as the GDR acts to uplift the moist oceanic air). The strongest anomalies are in JJA.

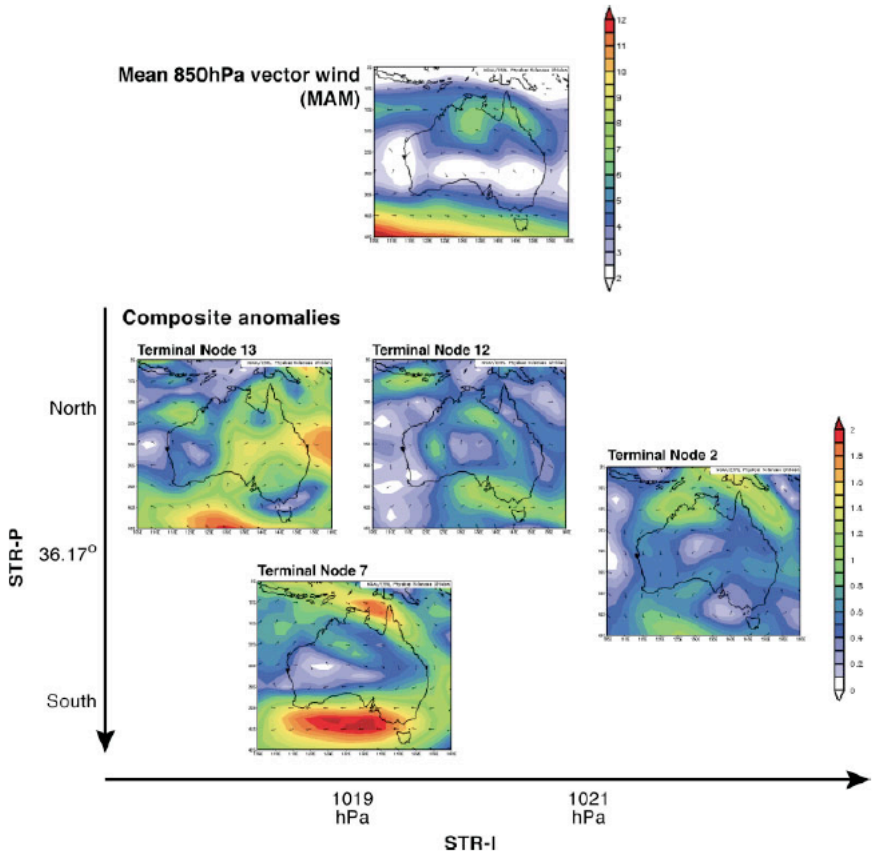

Figure 6.12. The 850hPa vector wind MAM mean (top diagram) and composite anomalies (for the four terminal nodes, below). Units are m/s. Computed from re-analysis from 1948-2009.

The anomalous wind flow in MAM's Terminal Node 7 and JJA's terminal Node 13 shows the southerly position of the STR, with easterly anomalies (i.e. weakening of the westerly flow) on the southern side of the STR. In the case of MAM, where Terminal mode 7 was linked to a 'summer-like' rainfall node, wind anomalies can be seen to originate from the tropics (Figure 6.12). Strong easterlies over north Queensland are reduced (1.5–2 m/s) and instead northeasterly anomalies affect SEA. The predominant westerly anomalies on the western and southern flank of the STR are strongly reduced (up to 2m/s) and shifting south.

Finally, the anomalous vector wind composites for Terminal Nodes 12 and 13 in MAM (Figure 6.12) and Terminal Nodes 7 and 12 in JJA (Figure 6.13) show the different flavour of westerly mean flow being pushed further north and hence affecting SEA, and the strength of the anomalies and the location of the westerly maximum differ from case to case. The most obvious case is the wind anomalies for JJA's 'winter-like' Terminal Node 7; it displays the large westerly wind anomaly in excess of 2m/s over inland Australia, suggesting a strong push of the

westerlies further north. In the case of MAM's Terminal Node 12, wind anomalies are harder to reconcile with the rainfall anomaly maps than in the other previous terminal nodes. They show consistent westerly anomalies over SEA, yet a very limited rainfall signature apart from a small region on the Victorian coast and Tasmania.

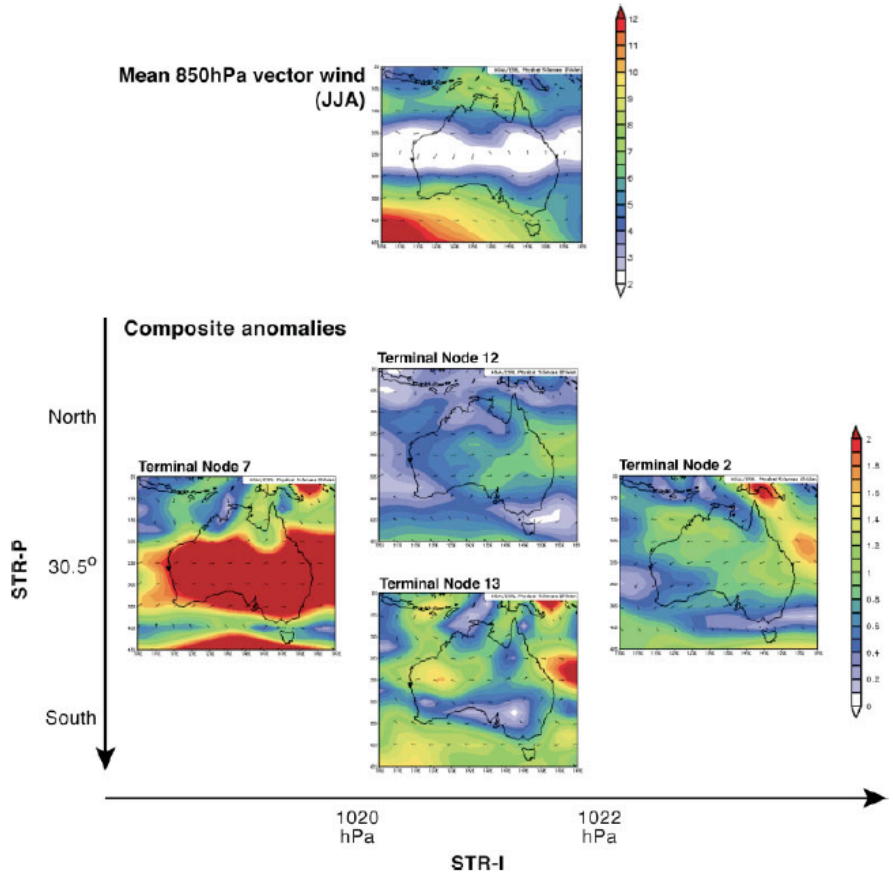

Figure 6.13. The 850hPa vector wind JJA mean (top diagram) and composite anomalies (for the four terminal nodes, below). Units are m/s. Computed from re-analysis from 1948-2009.

6.5 Discussion and conclusions

This is a pioneering study introducing the use of CART to explore complex relationships within the climate system, and the large-scale influences of regional rainfall in SEA. The focus in this study is on the interplay between STR changes in both intensity and position, and SEA precipitation in MAM and JJA. CART analysis is used to complement results obtained with linear statistics.

The linear analysis confirms recent research showing the relationship between STR intensity and position with SEA rainfall in all seasons (Williams and Stone 2009). The dominant role played by changes in intensity over position for SEA rainfall variability is highlighted, particularly in light of the current rainfall deficit in SEA, which is mainly a MAM phenomenon

with an additional contribution in JJA (Timbal and Drosdowsky 2013). Rainfall in both seasons displayed a stronger relationship with STR intensity than with position and followed a seasonal cycle with correlations reaching maximum values (spatial extent and magnitude) in JJA. However, a split in two 50-year epochs showed significant changes in the relationship between the STR and precipitation over time. Most notable was a change in the sign of the relationship with the position of the STR in MAM across SEA, which highlighted the importance of the mean position of the STR, particularly in a non-stationary climate. Correlation and regression analysis are traditionally used heavily to investigate relationships within the climate system(Timbal and Drosdowsky 2013, Williams and Stone 2009) so we turned to more complex nonlinear statistics to confirm and extend the current understanding. This is important because correlation analysis, particularly when sample size is limited to about hundred years, does not provide a robust understanding of more complex issues such as the interplay between position and intensity, i.e. linear regression does not indicate an additional role for changes in position beside the STR intensity (Timbal and Drosdowsky 2013). The nonlinearity of the relationship hinted at by computing correlations over different epochs, and the complex interplay between changes in position and intensity was investigated further by using classification trees. Confirming our understanding of the system gained from the linear analysis, the CART analysis showed that intensity was the most important variable in both seasons for classifying dry years but also played a role in classifying wet years. Most importantly, the CART analysis was able to tease out the complexities of this interaction and how the two aspects of the STR (intensity and position) interact.

The CART analysis confirms that STR intensity is the most important variable for classifying dry conditions in both MAM and JJA; in both seasons, the first split identified dry terminal nodes with negative precipitation anomalies over SEA classified on the basis of high STR intensity. An analysis of vector wind anomaly composites for these nodes showed a reduction in the mean westerly flow across SEA. This is consistent with the spatial rainfall composite for these years, with negative precipitation anomalies across the entire region under the influence of frontal activity in both seasons. The years in each of the dry terminal nodes confirmed that the current rainfall deficit is linked to increasing STR intensity (Timbal 2009). Indeed, in the last two decades a majority (11/13) of dry MAMs are classified in the dry terminal node (on the basis of high STR intensity) and while the rainfall deficit in JJA is not as severe, one can observe that all dry years in the last two decades are classified into the dry terminal node on the basis of high STR intensity.

But the real value of this approach is that the complexities of the STR position–rainfall relationship that were masked by the coarse linear analysis appear more clearly within the

CART analysis. Indeed, the STR position is needed in both seasons to classify the remaining years. In MAM, position played a more prominent role when intensity is close to or below average, the secondary split in the CART analysis is based on position. This greater role for position in MAM is consistent with the large changes in position observed in this transition season. In JJA, position was only necessary for a third split between terminal nodes when intensity was neither significantly above or below average. Indeed in JJA, the relationship between rainfall in SEA and STR intensity appears more linear than in MAM with extreme values of intensity associated with the driest and wettest cases. In this season, the position of the STR is only playing an important role when intensity is near average. Overall, in both seasons, the fact that the role of the STR position depends on the intensity of the STR may explain why multiple linear regression statistics cannot find an additional role for the STR position independently of the STR intensity (Timbal and Drosdowsky 2013). It is only once particular intensity conditions are met that position becomes important to determine the sign and spatial signature of precipitation anomalies over SEA.

One of the most revealing findings is the role that STR position has in MAM on the atmospheric circulation. The CART analysis revealed two wet regimes in MAM that are very different in their nature; a 'summer-like' rainfall node (Terminal Node 7) and a 'winter-like' rainfall node (Terminal Node 13). The 'summer-like' rainfall node shows that when intensity is low and the STR is in a southerly position wet conditions are evident over much of the east coast as tropical interactions bring moist air to southern Australia. This combines with a reduction of the mean westerly flow which pushes frontal activity south resulting in the negative precipitation anomaly on the Victorian coast, indicative of the southerly position of the STR. The highest rainfall anomalies are apparent on the eastern coast as the mean anti-cyclonic winds associated with the high-pressure cell bring a moist onshore airflow over the continent. The positioning of the STR at higher latitudes is consistent with a widening of the tropics and Hadley Circulation (Seidel et al. 2007), although the rare occurrence of this node does not enable us to draw very firm conclusions apart to note that it has occurred in the most recent decades. More importantly, the existence of this 'summer-like' node highlights that natural variability in the width of the tropical zone is associated with rainfall anomalies over eastern Australia as far south as SEA, despite the lack of relationship between rainfall in SEA and tropical modes of variability in MAM (Timbal et al., 2010). This 'summer-like' rainfall mode differs markedly from the 'winter-like' node (Terminal Node 13) and highlights the role of the STR position. When the STR is located in a northerly position ($\leq 36.17°$) the precipitation pattern is indicative of precipitation caused by cold fronts imbedded in a moist westerly airflow;

positive anomalies confined to the SEA region with the largest values on the Alps and the west coast of Tasmania, suggesting an increased mean westerly flow.

Overall, the CART analysis shows that in JJA the effects of position are not well defined. This is consistent with the mean position of the STR, north of SEA during this season and fairly stable across the JJA months while in MAM it was noted to change dramatically from month to month from a position south of SEA to the north (Figure 6.1). Overall the CART analysis confirmed many known features identified using simple correlations. The increased node purity in all terminal nodes in JJA compared with that in MAM is consistent with the higher correlation between both STR intensity and position with rainfall in JJA. The threshold values upon which splits were based support the seasonal cycle, i.e. higher threshold values for STR intensity in JJA is a result of higher STR intensity during the JJA months compared with the MAM months. The precipitation patterns of JJA's dry and the wet terminal nodes (Terminal Nodes 2 and 7) are mirror images of each other (see also the vector wind analysis) and are associated with high and low STR intensity because of the linear nature of the relationship between STR intensity and rainfall in this season. Overall, the CART analysis confirms that the STR intensity–rainfall relationship is much more linear than in MAM and that the nonlinear relationship with the STR position play a minor role for rainfall as the STR is located further north. Terminal Nodes 12 and 13, however, are classified on the basis of STR position, and the composite analyses (precipitation and vector wind) of these nodes do not show many differences in SEA, although more sizeable differences are seen further north.

On the basis of the analysis provided here, one can conclude that while the position of the STR plays a clear role in MAM it is not the case in JJA and this is likely owing to the mean position of the STR in both seasons. In JJA, the mean position of the STR is to the north of the study region, and thus variation in the position has little effect on precipitation in SEA, while variation in STR intensity is still able to play an important role during this season. Variability in STR position has a more important association with precipitation in SEA in MAM, the transition season from a southerly to a northerly position, as the mean position of the STR is further south combined with its annual meridional movement.

A noted feature of the JJA CART analysis was a small number of wet years, three amongst the predominantly dry Terminal Node 2, which occurred in the most recent period (1995, 2003 and 2005). These three cases experienced much higher rainfall while the intensity of the STR remained high, although the mean position of the STR was further south in these 3 years compared to the full record. No wet cases were recorded during the first 94 years of the instrumental record yet three wet cases were recorded in the last 15 years. The anomalous vector wind analysis of these years showed that a moist northerly airflow dominated SEA. Analysis of

the Coral Sea SST showed that all 3 years had positive SST anomalies in this region; in addition 2003 and 2005 were in the top 10th percentile. This shows that in rare occasions, despite high STR intensity, a small shift southward of the STR position, in combination with warm SST anomalies in the Coral Sea region, may affect precipitation in SEA. It was suggested that perhaps these occurrences have only recently appeared in the record because of increased regional SST (Casey and Cornillon 2001). Although logical, this hypothesis would require more in-depth analysis to be demonstrated. Less controversial were the results showing interplay between El Niño events and STR intensity in JJA, i.e. El Niño events predominantly being observed in conjunction with above-average STR intensity and resulting predominantly in dry years, while other El Niño years not associated with above average STR intensity were predominantly average or wet years. This suggests that both ENSO variability impacts STR variability, and that independent STR variability can also modulate the impact of ENSO on SEA rainfall.

It is clear that in recent years STR intensity has been important for precipitation in SEA (Timbal and Drosdowsky 2013) but that the role of position should not be underestimated, particularly in MAM, despite the lack of additional effect reported in that same study (Timbal and Drosdowsky 2013). Indeed it has been shown here that MAM precipitation variability in SEA cannot be described solely by STR intensity and that wet years in particular require the role of position to be understood. The mean STR intensity in MAM is 1020.2hPa, and we see in the classification tree that over the instrumental record when the intensity is ≥ 1021hPa conditions are dry 60% of the time. The implications for this relationship in a climate change regime are interesting; while the magnitude is unclear at this stage, changes to the Hadley Circulation under a climate change regime are likely (Seidel et al. 2007). Owing to its geographic position, SEA is highly sensitive to changes in the STR position. Consequently, the subtle effect position has precipitation in SEA, while not as important as the impact of rising intensity, should not be discarded.

6.6 Acknowledgements

This study was supported, in part, by the South Eastern Australian Climate Initiative. We thank Carolyn Whan and Clive Hilliker for assistance with graphics. We are grateful to two anonymous reviewers whose suggestions improved the quality of this article.

Chapter 7 : The potential predictability of hydroclimatic regimes with random forest and linear regression models

7.1 Chapter summary and context

Previous chapters have established the significant linear relationships between the set of predictor variables used throughout the thesis, including indices of Pacific and Indian Ocean variability and indices capturing facets of the STR, and Australian hydroclimatic variability (Chapter 4). In addition the value of a tree-based approach in giving insight to interactions between tropical and sub-tropical climate drivers that are important for Australian rainfall regimes has been demonstrated (Chapters 5 and 6). This chapter builds on the previous work and explores the potential predictability of seasonal hydroclimatic regimes (rainfall and soil moisture), defined from tropical and sub-tropical predictor variables, using a classical linear framework and the tree-based methods. The full discussion of background literature and methods can be found in Chapters 2 and 3.

Agricultural, environmental and water system management require seasonal hydroclimatic forecasts. We investigate two statistical techniques (multiple linear regression and tree-based models) for assessing the potential predictability of rainfall and soil moisture in Australia, at lead-times of 0-11 months. Predictor variables include indices for Pacific- and Indian-Ocean climate modes and the sub-tropical ridge. We find that: (1) linear regression and tree-based models have comparable predictive skill; (2) model skill is greatest when the impact of tropical climate modes is largest in austral winter, spring and summer for rainfall and soil moisture; and (3) greater skill can be achieved in predicting lower-layer soil moisture (below 0.2 m) than for rainfall and upper-layer soil moisture, because of the natural smoothing in lower-layer soil moisture. The last result could improve seasonal forecasting of water availability and hydroclimatic regimes, especially for agricultural applications.

7.2 Introduction

Increasing understanding of the predictive persistence of relationships between predictor and response variables at the catchment scale is needed for efficient water resource management and informed risk based decision-making in regions of high hydroclimatic variability (Leblanc et al. 2009). The focus throughout this thesis has been primarily on rainfall, but soil moisture is an important hydroclimatic variable because it interacts with runoff (Chiew et al. 2011), has

regional-scale feedbacks with rainfall (GarcÌa-GarcÌa et al. 2011) and temperature (Cai and Cowan 2008b).

This research is set in a seasonal forecasting context, but the aim is to explore of predictive persistence of relationships between a response variable and a set of predictor variables using the tree-based models, rather than to specifically create improved seasonal forecasts for Australia. Two approaches are widely applied to seasonal forecasting. Computationally intensive dynamical models that do not suffer from stationarity constraints (Lim et al. 2011) (but are often 'tuned' to the current climate) are moving to the forefront of climatic research. In comparison, statistical models have the advantage of increasing understanding of physical processes at little computational cost. As such, although statistical models are limited to understanding variability during the instrumental record, they remain a powerful tool in this type of application, which is focused on understanding relationships between predictor and response variables.

The predictability of Australian hydroclimate primarily stems from lagged relationships with persisting tropical modes of climate variability (Simmonds and Hope 1997), including the El Niño – Southern Oscillation (ENSO) (Drosdowsky and Williams 1991, Kirono et al. 2010, Nicholls et al. 1997) and the Indian Ocean Dipole Mode (Nicholls 1989, Saji et al. 1999). The sub-tropical ridge (STR) also has an important relationship with southern Australian rainfall (Drosdowsky 2005, Timbal and Drosdowsky 2013, Whan et al. 2013). This should not imply that other large-scale climate drivers are not equally important, but the focus in this research is on interannual interactions between tropical and sub-tropical climate drivers that influence Australian hydroclimatic regimes.

Classification and regression trees (CART) (Breiman et al. 1984) are a set of statistical techniques that can clarify non-linear aspects of relationships between event-based predictor and response variables (Firth et al. 2005). These methods have been used to diagnose relationships between southeast Australian rainfall and facets of STR behaviour (Whan et al. 2013), to predict hydrological regimes from indices of sea surface temperature (SST) and antecedent streamflow (Wei and Watkins 2011) and to predict the onset of climate regimes from SST indices in Western Australia (Firth et al. 2005). Random forests are a bootstrapping version of CART that protects against the over-fitting problems that tend to affect other tree-based methods (Breiman 2001).

Here we compare random forests with traditional multiple linear regression, with particular emphasis on the predictability of soil moisture and rainfall.

7.3 Data and methods

Spatially aggregated terciles of gridded rainfall, upper-layer soil moisture (USM, soil depths typically < 0.2 m) and lower-layer soil moisture (LSM, soil depths typically > 0.2 m) for the 110-year period 1900 – 2009 are used as response variables. These quantities were derived from the Australian Water Availability Project data set (Raupach et al. 2008). All quantities used here are area-averaged over 12 Australian drainage divisions (National Land and Water Resources Audit 2001) (Table 7.1).

Table 7.1. Names, abbreviations and annual average rainfall for each of the drainage divisions used as response variables.

Drainage Division	Region Number	Average annual rainfall (mm year^{-1})
Timor Sea (TiS)	1	879.95
Gulf of Carpentaria (GoC)	2	786.67
Northeast Coast – Sea (NECS)	3a	1145.66
Northeast Coast – Burdekin/Fitzroy (NECBF)	3b	654.89
Murray-Darling Basin Agricultural (MDBA)	4a	539.44
Murray-Darling Basin Semi-Arid (MDBSA)	4b	326.00
Murray-Darling Basin Wet (MDBW)	4c	1056.35
Southeast Coast (SEC)	5	862.76
Tasmania (Tas)	6	1428.03
South-Australian Gulf (SAG)	7	354.49
Southwest Coast (SWC)	8	452.84
Indian Ocean (IO)	9	274.38

Terciles were calculated for each of the four standard seasons (MAM, JJA, SON, and DJF). Predictor variables include 14 seasonally averaged indices characterising large-scale modes of climate variability in the Pacific and Indian oceans, together with the position and intensity of the STR (Table 7.2).

Table 7.2. Indices of large-scale modes of climate variability used as predictor variables in the linear and tree-based statistical analysis.

Name	Group	Reference
Niño 3	Pacific Ocean	(Trenberth 1997)
Niño 3.4	Pacific Ocean	(Trenberth 1997)
Niño 4	Pacific Ocean	(Trenberth 1997)
El Niño Modoki Index (EMI)	Pacific Ocean	(Ashok et al. 2007)
Southern Oscillation Index (SOI)	Pacific Ocean	(Troup 1965)
Tri-pole Index (TPI)	Pacific and Indian Ocean	(Timbal and Hendon 2011)
Region C	Indian Ocean	(Cai and Cowan 2008a)
Nicholls' SST gradient	Indian Ocean	(Nicholls 1989)
Northwest Shelf (NWS)	Indian Ocean	(Timbal and Murphy 2007)
Dipole Mode Index (DMI)	Indian Ocean	(Saji et al. 1999)
DMI – Eastern Pole	Indian Ocean	(Saji et al. 1999)
DMI – Western Pole	Indian Ocean	(Saji et al. 1999)
STR – Intensity	STR	(Drosdowsky 2005, Timbal and Drosdowsky 2013)
STR – Position	STR	(Drosdowsky 2005, Timbal and Drosdowsky 2013)

From the complete data set, 10 ensembles were created with each member consisting of a 'training' data set (80 randomly selected years) and a 'test' data set (the remaining 30 years)

Random forest (each containing 100 individual trees) and stepwise selected multiple linear regression models were constructed from the training data in each ensemble and were evaluated against corresponding test data. Model skill is defined by the correlation between predicted and observed terciles in the test data sets.

First, the random forest and linear regression techniques are compared by assessing their skill at predicting Murray-Darling Basin rainfall in SON from the preceding nine overlapping three-month seasons (e.g. SON, ASO, JAS etc.). The spread of skill scores from the random ensembles indicates the range of skill score uncertainty. The separation in skill levels required for significance is arbitrarily assessed at 0.1. The random forest technique is then applied to other regions (Table 1), seasons and response variables (rainfall, USM and LSM) (Figures 7.2 and 7.3).

7.4 Results and discussion

7.4.1 Model comparison

The linear regression and random forest models show similar skill at predicting SON rainfall in the Murray-Darling Basin (Figure 7.1). Skill is nearly steady (within statistical variability) at 0.4 to lead-times of 4-months, when model skill decreases and statistical noise increases. At longer lead-times the skill of both methods declines, consistent with the autocorrelation time scale of ENSO. Random forest models have the additional benefits of

reducing over-fitting and requiring no assumptions about the distribution of data (Hastie et al., 2001). This comparison provides confidence that the following results, all obtained with random forest, are unlikely to be dependent on method.

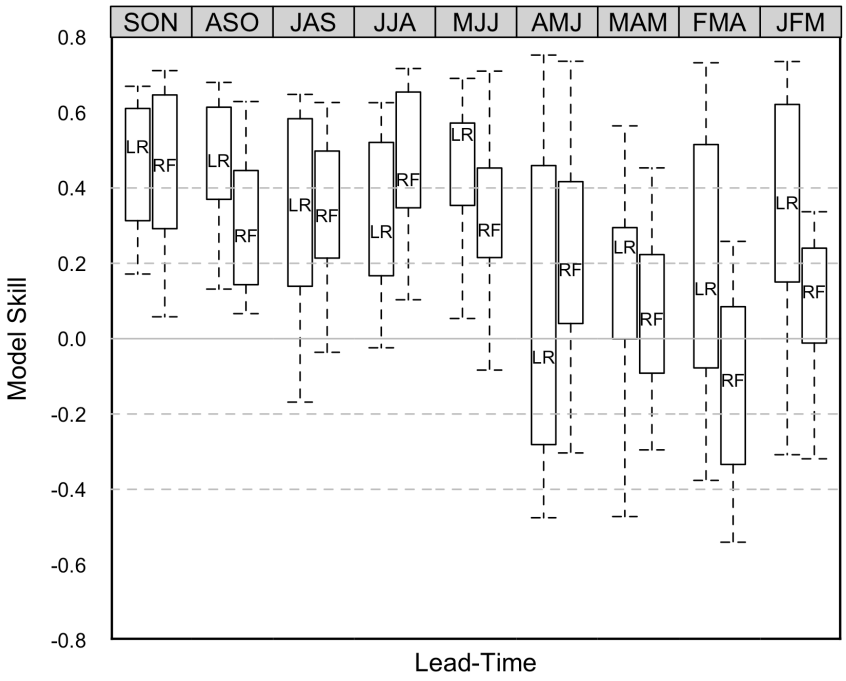

Figure 7.1. Model skill for the stepwise selected linear regression models ("LR") and the random forest ("RF") analysis predicting all-MDB (Regions 4a, 4b, 4c combined) SON rainfall from lagged predictor variables. Boxes indicate the 25th – 75th percentile; top panels indicate months used to predict SON rainfall. The box and whisker plots show the range of skill scores of the 10 random ensembles.

7.4.2 Prediction of hydroclimatic regimes

The skill of random forest models at predicting seasonal rainfall, USM and LSM for three representative regions in the southeast, northeast and southwest of the continent (Figure 7.2) shows that predictability of rainfall and USM is generally greatest in regions (4b - northeast and 3a - southeast) and seasons (JJA, SON, DJF) where the relationships with large-scale modes of climate variability are strongest. This shows the statistical models are capturing known sources of predictability (Drosdowsky and Chambers, 2001; Kirono et al.; 2010, Risbey et al.; 2009). In addition, we see skill in regions (southeast, southwest) and seasons (MAM, DJF) not identified in previous work (Drosdowsky and Chambers, 2001), because additional predictor variables are included here. The key addition is the STR, which has been found previously to be an important predictor of rainfall in southern Australia in MAM and JJA (Drosdowsky, 2005).

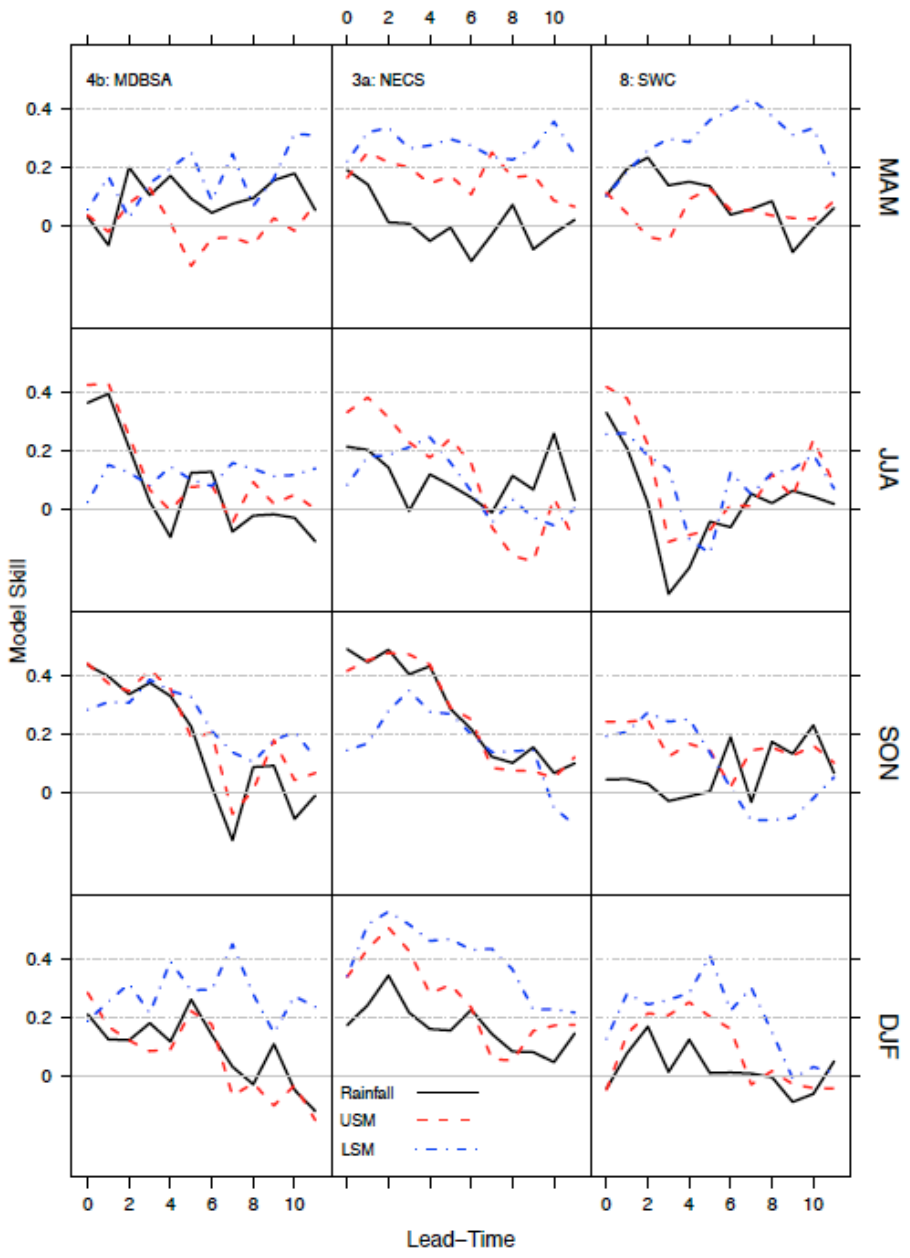

Figure 7.2. Skill of random forest models at predicting rainfall (black solid), upper-layer soil moisture (red dashed) and lower-layer soil moisture (blue dot-dashed) in SON with increasing lead-time (months). The spatial distribution of regions is shown in Figure 7.3. Region numbers correspond to regions outlined in Table 7.1.

The remainder of this discussion focuses on regions and lead-times with 'useful' predictability (which is qualitatively assessed as skill levels greater than 0.4). Firstly, in Region 4b (southeast) rainfall and USM are skilfully predicted in JJA and SON at short lead-times only. Skill levels in JJA and SON are affected by the ENSO 'autumn predictability barrier' (Balmaseda et al. 1995). The short-lead predictability of JJA rainfall and USM is related to their near-instantaneous response to ENSO. LSM is not predictable in ENSO-driven regions in JJA at

short lead times because available surface water requires time to penetrate into deeper soil layers. Secondly, in Region 3a (northeast) USM and rainfall are skilfully predicted in SON, and predictions of USM and LSM are skilful at longer lead-times (~6 months) in DJF. Finally, in Region 8 (southwest) rainfall and USM are skilfully predicted at short-lead times in JJA. This rainfall signal appears persistent, as LSM has strong skill scores at a 7-month (5-month) lead-time in MAM (DJF). Theses results show that indices of large-scale climate variability are able to skilfully predict rainfall and soil moisture in JJA, SON and DJF in the northeast (Region 3a) and southeast (Region 4b) of Australia (Figure 7.2).

In Figure 7.3 we explore the predictability of SON and DJF rainfall, USM and LSM in other regions. In SON (Figure 7.3a) maximum predictability occurs in eastern and northern Australia, where skill levels for rainfall and USM reach 0.5 at lead-times up to 5-months. In the southeast (Regions 4a–7) LSM skill levels are maintained for an additional 1-2 months. In Region 4b, in the agriculturally important Murray-Darling Basin predictions of SON catchment wetness from two seasons ahead are as skilful as the simultaneous model as skill is maintained to 6-month lead-time. In the north (Regions 1, 2, 3b) there is little skill at predicting LSM up to the 5-month lead-time. At longer lead-times (> 5-months), LSM predictability increases in Region 2, likely as a consequence of smoothing of shorter-term month-to-month variability. The difference in LSM predictability between Regions 1, 2 and 3b (with distinct wet and dry seasons) and Region 3a (wet throughout the year) shows the additional predictability in LSM that stems from the slow movement of water through the hydrological cycle, in addition to the effect of persistence of large-scale climate modes on rainfall, USM and LSM. In southeast Australia (Regions 4c, 5 and 6) both USM and LSM have greater predictability than rainfall to a 6-month lead-time; thereafter all three variables follow a similar skill decay pattern.

A larger increase of LSM predictability over rainfall and USM occurs in DJF (Figure 7.3b), compared to SON, in the southern and eastern Australia. Predictability for USM often mirrors that for rainfall (e.g. Regions 3a–8), albeit with slightly higher skill levels as USM smooths some high frequency variability. Predictability is highest for LSM, with a maximum at the 4-5 month lead-time (e.g. Region 4a). In Region 5, LSM is predictable to a lead-time of 8 months. In Region 7 the predictability of LSM exceeds 0.4 at the 4-month lead-time despite little skill at zero lag.

The skill in predicting LSM several seasons in advance stems from three factors: (1) the persistence of ENSO for 9-10 months from onset; (2) its typical adherence to the annual cycle and (3) the time moisture takes to move through the system. Soil moisture filters some high frequency variability in high-rainfall drainage divisions and the natural smoothing of LSM results in greater predictability, as stochastic noise in the rainfall record is not present.

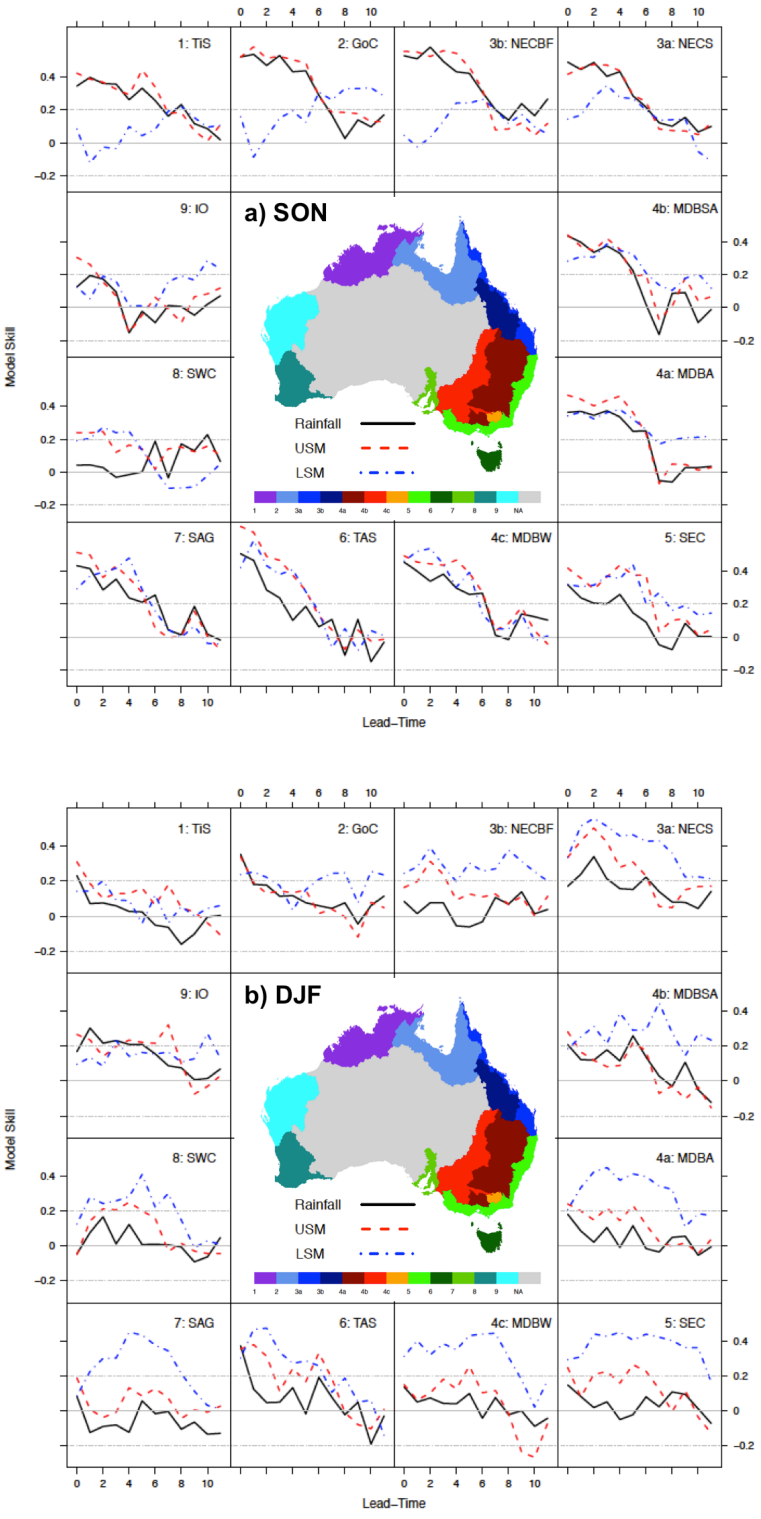

Figure 7.3. Skill of random forest models at predicting rainfall (black solid), upper-layer soil moisture (red dashed) and lower-layer soil moisture (blue dot-dashed) in a) SON and b) DJF. The spatial distribution of regions is shown on the map. Region numbers correspond to regions outlined in Table 7.1.

7.5 Conclusions

Here the predictive persistence of seasonal rainfall and soil moisture in Australian drainage divisions from a suite of oceanic and atmospheric indices of large-scale modes of climate variability that have relationships with Australian rainfall (Risbey et al. 2009) has been assessed. Random forest and linear regression show comparable skill in predicting SON rainfall regimes in the Murray-Darling Basin; demonstrating that our results are unlikely to be dependant on statistical method.

Predictability of hydroclimatic regimes varies consistently with known relationships. The novel finding of this research is that the predictability of rainfall and USM behaves differently to that of LSM. For example, maximum skill at predicting northern and eastern Australian rainfall and USM is tied to the seasonal cycle and persistence of the tropical modes of variability. However, LSM can be skilfully predicted for an additional 1-2 months lead-time in southeastern Australia. Maximum skill levels for rainfall and USM generally occur at zero lag, as they respond immediately to variability in large-scale climate modes. Previous research demonstrated that soil moisture has a stronger relationship with antecedent soil moisture compared to large-scale drivers of climate variability (Kirono et al. 2010). Here we extend this finding by showing that LSM can be skilfully predicted for a longer lead-time than rainfall or USM, from large-scale climate drivers. The predictability of rainfall and USM stems from the persistence of large-scale modes of climate variability from onset, in contrast, the predictability of LSM has two sources. First, water takes time to travel through the hydrological cycle. A hydrological signal propagates from the beginning of the ENSO cycle in response to the influence on rainfall and affects LSM for some months. LSM smooths the signal and removes some of the noise inherent in the rainfall records. Secondly, this effect is enhanced by the persistence of ENSO, as ongoing impacts on rainfall propagate through the hydrological cycle to LSM. For rainfall and USM the predictability stems only from the tendency of ENSO to remain in phase from onset, while for LSM the persistence of ENSO acts to reinforce the initial signal. LSM is therefore more predictable than rainfall or USM in DJF and, at some lead-times, in SON. This suggests that greater confidence can be attached to seasonal forecasts of LSM compared to rainfall, which is valuable for industries such as agriculture. We have shown that tree-based statistical methods, recognised as a powerful data-mining technique, can be useful in climate research and have potential as a forecast tool. There are two limitations of this study that provide avenues for future research. Firstly, the only predictor variables included in this study are indices of ENSO, the IOD and the STR. The inclusion of additional predictor variables, such as indices of the Southern Annular Mode or antecedent hydroclimatic conditions, would likely

increase predictability. The additional predictability gained by the inclusion of these indices using tree-based methods is one avenue for future research. Secondly, the relationship between ENSO and Australian rainfall is known to vary on multi-decadal (Power et al. 1999) and so the predictability of Australian hydroclimatic regimes also varies on a multi-decadal scale. Throughout this thesis the focus has been on the drivers of interannual hydroclimatic variability, but the relationships between predictor variables that vary on a multi-decadal scale remains an important questions. The impact of this multi-decadal variability could be the subject of future research. Random forest can incorporate multiple highly correlated predictor variables, be applied where little is known *a priori* about the data, requires no assumptions about data structure and distribution and contains increased protections against model over-fitting through the implementation of in-bag and out-of-bag samples. The random forest technique also removes some of the instability inherent in highly data dependent individual tree-based methods.

7.6 Acknowledgements.

This work was supported, in part, by the South-East Australian Climate Initiative (SEACI). We thank Wendy Merritt for useful comments.

.

Chapter 8 : Synthesis and Conclusions

8.1 Wider context and innovation

High interannual variability and low water availability, combined with social and economic pressures, makes water resource management a key issue for Australia. These pressures are likely to increase in the future when it is predicted that the most densely populated region of Australia, the southeast, will become warmer and drier.

The relationship between large-scale modes of climate variability and Australian rainfall has been well established over recent decades, and there is extensive evidence that interactions between these climate drivers are critical in explaining hydroclimatic variability. This thesis aimed to increase understanding of how interactions between large-scale climate drivers influence Australian hydroclimatic regimes.

Four research questions were developed within this theme and answered through Chapters 4 – 7. The complexity of the research questions increased throughout the thesis, beginning with an examination of the linear relationships between indices of large-scale variability and Australian hydroclimatic variability with a relatively new data set (Chapter 4). Next was a thorough introduction to the tree-based methods and the associated terminology, and an exploration of tropical and sub-tropical interactions that are important for regional rainfall regimes in Chapter 5. Chapter 6 investigated interactions between facets of STR variability that are important in defining southeast Australian rainfall regimes. Finally, in Chapter 7, the predictability of Australian hydroclimates using tree-based methods was discussed.

The unique aspects of this work are listed below, followed by a detailed discussion of each contribution individually (Sections 8.2 - 8.5) and finishing with a synthesis and concluding remarks (Section 8.6). Methodologically, this thesis has demonstrated the value of tree-based methods to extend understanding of interactions between climate drivers. A tree-based analysis is appropriate for studying discrete, event-based phenomena, such as rainfall or the large-scale conditions (or combinations of conditions) that make rainfall more or less likely. As such, additional insights and more nuanced information about interactions between indices of large-scale climate variability are gained from a tree-based analysis.

The first scientific contribution of this thesis was a complete analysis of the simultaneous and lagged linear relationships between large-scale modes of climate variability and Australian hydroclimate (rainfall and two layers of soil moisture), validating the dataset used throughout the thesis (Section 8.2). Analysis of tree-based models using all predictor variables (indices of

ENSO, IOD and the STR) confirmed that the novel statistical technique was consistent with the correlation analysis. In addition, the models defining rainfall regimes from a subset of predictor variables demonstrated additional insights as to the nature of tropical – sub-tropical interactions that are important for regional Australian rainfall regimes (Section 8.3). Next, the role of STR position in defining 'winter-like' and 'summer-like' rainfall regimes was demonstrated (Section 8.4). The final contribution of the research was to show the increased predictability of lower-layer soil moisture compared to upper-layer soil moisture and rainfall (Section 8.5).

8.2 Linear relationships (Chapter 4)

The linear relationships between indices of remote (i.e. ENSO and IOD) and regional (i.e. STR) large-scale variability and Australian hydroclimatic variability (rainfall and soil moisture) were analysed in Chapter 4. It was important to establish the nature of the linear relationship between the large-scale climate drivers and Australian hydroclimate using the indices as defined in Chapter 3 and the AWAP dataset. Relationships were spatially and temporally consistent with previous research, as indices of the STR, Pacific and Indian Ocean variability all influenced hydroclimatic variability in southeastern Australia, particularly in JJA.

Both tropical climate drivers (ENSO and the IOD) have significant relationships with hydroclimatic variability in eastern and southern Australia from JJA to DJF, with the relationship peaking in SON. El Niño (positive IOD) events are associated with decreased rainfall in eastern (southern) Australia. The persistence of the tropical modes of variability translates into significant predictability of SON rainfall and soil moisture from oceanic indices in preceding months. Spatial and seasonal co-variability between the influences of the climate drivers (i.e. the IOD, ENSO or the STR) can make separating the relative influences difficult.

The relationship between facets of STR variability (i.e. the position and intensity of the STR) and Australian rainfall also follows a defined seasonal cycle, with the correlations peaking in JJA. In all seasons, there is a stronger relationship between rainfall and soil moisture with STR intensity, compared to position. In various seasons, the changing sign of correlations between rainfall and both STR intensity and position suggests that a linear analysis may not adequately capture the relationships, as small departures from the long-term mean can change the sign of the relationship. Generally, in southern Australia increasing STR intensity is associated with decreased rainfall, but for some regions (e.g. the east coast) and seasons (e.g. DJF) increasing STR intensity is associated with increased rainfall, likely associated with an intensification of the geostrophic flow from an increased pressure gradient. The relationship with STR position is more complex and depends on the mean position of the ridge, which varies seasonally and is thought to have changed over the instrumental record. The strong correlation between facets of the STR and the changing nature of the relationship suggests that a statistical technique capable of capturing non-linear interactions may be better able to extend understanding of the relationships between the STR and Australian hydroclimatic variability.

This section validated the dataset and provided a foundation for more complex analyses in subsequent sections. In addition, regions where there may be significant value in moving to

more novel methods, that are suited to discrete predictor and response variables and can capture non-linearities, were identified.

8.3 Tropical – sub-tropical interactions that influence Australian hydroclimatic regimes (Chapter 5)

Following from the validation of the datasets, Chapter 5 explored the relationships between Australian hydroclimate and large-scale modes of climate variability using novel tree-based methods (i.e. Classification and Regression Trees, CART).

In a validation of the novel statistical technique, tree-based models using all predictor variables defined hydroclimatic regimes in 12 regions of Australia, that were seasonally and spatially consistently with the conventional linear analysis. For example, classification trees showed that indices of ENSO generally dominated in northern and eastern Australia, while indices of Indian Ocean variability and the STR typically dominating in the south. Furthermore, the tree-based models are able to capture seasonal variability in the relationships between predictor and response variables. When the influence of the tropical modes is weak in MAM, facets of STR variability are selected as the most important predictor variables in a majority of regions. This shows that the tree-based models capture the same relationships and sources of variability as the linear analysis.

The most important predictor variable defined in the tree-based analysis was often also the most highly linearly correlated variable. Where strong linear relationships existed between predictor and response variables, the tree-based models tended to be dominated by this linear relationship with multiple splits on a single predictor variable. This suggests that an analysis with a linear statistical framework adequately captures those relationships. However, the most highly correlated predictor variable was not always selected as the most important variable, which shows that the tree-based method can extract different information about the relationships between predictor and response variables.

A pairwise analysis using a subset of predictor variables (i.e. ENSO, IOD and STR) demonstrated the additional insights a tree-based analysis could provide as to the nature of tropical – sub-tropical interactions. Case studies from regions in north- and southeastern Australia showed the importance of interactions between tropical – sub-tropical climate drivers in defining rainfall regimes. Consistently, across different regions, the most spatially homogenous rainfall regimes were defined when the tropical modes of variability and the STR were in the same phase, showing the importance of phase locking for rainfall regimes; i.e. a wet regime with continent wide positive rainfall anomalies is defined when a weak STR combines with a La Niña event.

For example, during JJA in the semi-arid portion of the Murray-Darling Basin, the "in-phase" regimes classified from El Niño (La Niña) conditions and high (low) STR intensity were characterised by continent-wide negative (positive) rainfall anomalies. Conversely, when El Niño (La Niña) conditions were combined with low (high) STR intensity the influence of ENSO is felt only in northeast Australia while the STR dominates in the south. This shows how the STR can modulate the ENSO teleconnection to southern Australia. This suggests that these interactions may be able to account for some differences in the impacts of El Niño events. The modulation of the ENSO teleconnection over southern Australia by the STR is an important aspect of this work and leads to further examination of the role of the STR in southeastern Australia.

8.4 Interactions between STR intensity and position that influence southeast Australian rainfall (Chapter 6)

The linear relationships between southeast Australian rainfall and STR intensity and position were shown to follow the seasonal cycle, developing during MAM and peaking in JJA. Southeastern Australian rainfall anomalies in MAM and JJA were the focus of this chapter and it was shown that in both seasons rainfall has stronger linear relationships with STR intensity than with its position. Analysis of spatial correlations in two 50-year epochs showed significant changes in relationships over time. In particular, the sign of the relationship between STR position and rainfall in southeast Australia changes between epochs which suggests a non-linear relationship and shows the importance of the mean position of the STR in a non-stationary climate.

The interactions between STR position and intensity that influence SEA rainfall were explored using classification trees. Consistent with the linear analysis, the tree-based methods showed that STR intensity was the most important predictor variable in both seasons, defining dry regimes when MSLP was high. Composite rainfall and vector wind anomalies showed a reduction in the mean western flow in these dry regimes suggesting a reduction in frontal activity across southern Australia when STR intensity is high. Analysis of the years in the dry terminal nodes suggested that the Millennium Drought was linked to increasing STR intensity, as 11 out of a total of 13 dry MAMs and all dry JJAs since 1990 and were classified into this dry regime.

The tree-based analysis demonstrated complexities in the STR – rainfall relationship that were masked by the linear analysis. In MAM, STR position was required to classify rainfall regimes when STR intensity is below average. In JJA, STR position was only required to classify rainfall regimes when the intensity was neither high nor low. This suggests that the relationship between STR intensity and rainfall is much more linear in JJA compared to MAM.

In MAM, when STR intensity was low, two distinct rainfall regimes where defined depending on the position of the STR. A ‘summer-like’ rainfall regime was defined when the STR is in a southerly position. In this regime, positive rainfall anomalies stretch across the east coast and stemmed from tropical interactions. This ‘summer-like’ regime occurred infrequently across the instrumental record but with no distinct trend. In contrast, the ‘winter-like’ regime was defined when the STR was at lower latitudes and differs markedly. Positive rainfall anomalies are confined to the region in southern Australia that is influenced by frontal activity,

consistent with an increase in the mean westerly flow. The 'winter-like' regime was much more common over the instrumental record, but has only been seen once since 1990. In JJA, the role of STR position was less well defined, consistent with the mean position of the ridge (north of southeast Australia) in this season.

This analysis highlighted the importance of STR position in the transition seasons (e.g. MAM) when the STR is rapidly moving from the southerly summer position to the northerly winter position. In addition, the role of the STR in the Millennium Drought was demonstrated.

8.5 The potential predictability of hydroclimatic regimes with random forest and linear regression models (Chapter 7)

The potential predictability of rainfall and soil moisture regimes using tree-based methods was assessed in Chapter 7. Although this work was set in a seasonal forecasting context, the aim was to examine the relative predictability of hydroclimatic regimes. Area-averaged seasonal rainfall, upper-layer and lower-layer soil moisture in 12 Australian drainage divisions were predicted from a suite of 14 remote and regional, oceanic and atmospheric indices of large-scale climate variability. Initially, it was shown that stepwise selected linear regression models and random forests (a bootstrapping implementation of CART) have similar skill in predicting Murray-Darling Basin SON rainfall, suggesting that the following results are not dependent on statistical method.

The variability in predictability of hydroclimatic regimes is spatially and temporally consistent with linear analysis. The novel contribution of this work was to demonstrate that the predictability, from large-scale modes of climate variability, of rainfall and upper-layer soil moisture behaves differently than lower-layer soil moisture. Confirming previous work, it was shown that the maximum skill at predicting rainfall and upper-layer soil moisture occurs in SON at short lead-times (< 5-months) in northern and eastern Australia, with predictability tied to the seasonal cycle and the persistence of the tropical modes of variability. In other seasons, rainfall and upper-layer soil moisture respond directly to predictor variables and so maximum skill occurs at the shortest lead-times. The source of rainfall and upper-layer soil moisture predictability is the persistence of the large-scale modes of climate variability from their onset (e.g. ENSO), so predictability is limited by the tropical modes' adherence to the seasonal cycle.

In SON, lower-layer soil moisture can be skilfully predicted for an additional 1-2 months in southeastern Australia. However, the predictability of south, east and northern Australian lower-layer soil moisture peaks in DJF due to the additional time from the onset of drivers such as ENSO. The predictability of lower-layer soil moisture in has two sources; the time water takes to move through the hydrological system and the persistence of the tropical modes of variability.

8.6 Concluding remarks and future work

The importance of effective water resource management is currently a critical issue in Australia and will increase in the future if southeast Australia becomes drier and hotter as a result of climate change, as predicted by most climate models. Year-to-year variability in large-scale climate drivers can have severe consequences on water intensive primary industries such as agriculture and mining. Additional information about the influence of predictable large-scale modes of climate variability on hydroclimatic regimes is valuable as it allows more informed decision-making. Increasing understanding about the factors controlling interannual climate variability will increase resilience to the effects of climate change. The aim of this research was to enhance understanding about interactions between large-scale modes of climate variability that influenced Australian hydroclimatic regimes. The key findings of this research add valuable insight about drivers of Australian hydroclimatic variability. The ability of tree-based methods to provide additional understanding and insights about a complex, multivariate system was demonstrated. Thus tree-based methods can be applied in other complex systems to increase understanding about interactions between strongly correlated predictor variables. Specifically, tree-based models added insight about the nature of interactions between drivers of Australian hydroclimatic regimes and demonstrate the increased predictability of lower-layer soil moisture over rainfall and upper-layer soil moisture, however, future work should be aimed at quantifying the significance of splits. The impacts of hydroclimatic extremes can be economically and environmentally costly, so increased understanding about interactions between climate drivers that influence hydroclimatic regimes will enable better planning by resource managers.

There are many possible application of this approach in the climate systems science, where understanding complex interactions in non-linear systems is key. Numerous large-scale modes of climate variability influence hydroclimatic variability in all continents. Often linear relationships are well recognised but interactions between multiple climate drivers are less well understood, as they are more difficult to explore with in a classical linear framework. Tree-based methods can easily be employed in this type of analysis with little computational cost. In addition, while the aims of this thesis have been achieved, there are some limitations to this research that provides avenues for future work. The focus here has been on tropical and sub-tropical drivers of Australian climate; future work could explore interactions with high latitude modes of variability, e.g. the Southern Annular Mode (SAM), or the relationship with antecedent conditions. The focus in this work has been on the interannual variability; as such

the role of multi-decadal variability (i.e. the IPO) remains an important question. Furthermore, the value of tree-based methods for examining interactions between climate drivers as simulated in regional or global climate models should be explored in the future.

References

Alexander, L.V., P. Uotila, N. Nicholls and A. Lynch 2010. A New Daily Pressure Dataset for Australia and Its Application to the Assessment of Changes in Synoptic Patterns during the Last Century. *Journal of Climate,* 23**,** 1111-1126.

Allan, R., J. Lindesay and D. Parker 1996. *El Niño Southern Oscillation and Climatic Variability,* Collingwood, CSIRO Publishing.

Ansell, T., C. Reason, I. Smith and K. Keay 2000. Evidence for decadal variability in southern Australian rainfall and relationships with regional pressure and sea surface temperature. *International Journal of Climatology,* 20**,** 1113-1129.

Arblaster, J., G. Meehl and A. Moore 2002. Interdecadal modulation of Australian rainfall. *Climate Dynamics,* 18**,** 519-531.

Ashok, K., S. Behera, S. Rao, H. Weng and T. Yamagata 2007. El Nino Modoki and its possible teleconnections. *Journal of Geophysical Research,* 112**,** C11007.

Australian Bureau of Meteorology. 2010. *Climate* [Online]. [Accessed 15 October 2010].

Australian Bureau of Meteorology. 2011a. *Annual Australian Climate Statement 2010* [Online]. Available: http://www.bom.gov.au/announcements/media_releases/climate/change/20110105.shtml [Accessed 15 May 2013].

Australian Bureau of Meteorology. 2011b. *Climate* [Online]. [Accessed 4 April 2011].

Australian Bureau of Meteorology. 2013a. *About the Indian Ocean Dipole* [Online]. Available: http://www.bom.gov.au/climate/IOD/about_IOD.shtml [Accessed 13 May 2013].

Australian Bureau of Meteorology. 2013b. *Climate change and variability* [Online]. Melbourne.

Balmaseda, M., M. Davey and D. Anderson 1995. Decadal and Seasonal Dependence of ENSO predicition skill. *Journal of Climate,* 8**,** 2705-2715.

Baquero-Bernal, A., M. Latif and S. Legutke 2002. On Dipolelike Variability of Sea Surface Temerature in the Tropical Indian Ocean. *Journal of Climate,* 15**,** 1358-1368.

Barros, A. and G. Bowden 2008. Toward long-lead operational forecasts of drought: An experimental study in the Murray-Darling Basin. *journal of Hydrology,* 357**,** 349-367.

Behera, S., J. Luo, S. Masson, S. Rao and H. Sakuma 2006. A CGCM Study on the Interaction between IOD and ENSO. *Journal of Climate,* 19**,** 1688-1705.

Bjerknes, J. 1969. Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Review,* 97**,** 163-172.

Breiman, L. 2001. Random Forests. *Machine Learning,* 45**,** 5-35.

Breiman, L., J. Friedman, R. Olshen and C. Stone 1984. *Classification and Regression Trees*, Wadsworth Advanced Books and Software.

Brown, J., P. McIntosh, M. Pook and J. Risbey 2009. An Investigation of the Links between ENSO Flavors and Rainfall Processes in Southeastern Australia. *Monthly Weather Review,* 137**,** 3786-3795.

Burrows, W., M. Benjamin, S. Beauchamp, E. Lord, D. McCollor and B. Thompson 1995. CART Decision-Tree Statistical Analysis and Prediction of Summer Season Maximum Surface Ozone for the Vancouver, Montreal and Atlantic Regions of Canada. *Journal of Applied Meteorology,* 34**,** 1848-1862.

Cai, W. and T. Cowan 2008a. Dynamics of late autumn rainfall reduction over southeastern Australia. *Geophysical Research Letters,* 35**,** L09708.

Cai, W. and T. Cowan 2008b. Evidence of impacts from rising temperature on inflows to the Murray-Darling Basin. *Geophysical Research Letters,* 35**,** L07701.

Cai, W., T. Cowan, P. Briggs and M. Raupach 2009a. Rising temperature depletes soil moisture and exacerbates severe drought conditions across southeast Australia. *Geophysical Research Letters,* 36**,** L21709.

Cai, W., T. Cowan and M. Raupach 2009b. Positive Indian Ocean Dipole events precondition southeast Australia bushfires. *Geophysical Research Letters,* 36**,** L19710.

Cai, W., P. van Rensch and T. Cowan 2011a. Influence of global-scale variability on the subtropical ridge over southeast Australia. *Journal of Climate*.

Cai, W., P. van Rensch, T. Cowan and H.H. Hendon 2011b. Teleconnection Pathways of ENSO and the IOD and the Mechanisms for Impacts on Australian Rainfall. *Journal of Climate,* 24**,** 3910-3923.

Campbell, E. 2005. Statistical modeling in nonlinear systems. *Journal of Climate,* 18**,** 3388-3399.

Campbell, E., B. Bates and S. Charles 2000. Nonlinear Statistical Methods for Climate Forecasting. *In:* Indian Ocean Climate Initiative (ed.) *Towards understanding climate variability in south western Australia - Research reports on the first phase of the Indian Ocean Climate Initiative.* East Perth, Australia.

Campbell, E. and M. Palmer 2010. Modeling and forecasting climate variables using a physical-statistical approach. *Journal of Geophysical Research,* 115**,** D10114.

Cannon, A., P. Whitfield and E. Lord 2002. Synoptic Map-Pattern Classifcation Using Recursinve Partitioning and Principal Component Analysis. *Monthly Weather Review,* 130**,** 1187-1206.

Carter, M. and J. Elsner 1997. A Statistical Method for Forecasting Rainfall over Puerto Rico. *Weather and Forecasting,* 12**,** 515-525.

Casey, K. and P. Cornillon 2001. Global and Regional Sea Surface Temperature Trends. *Journal of Climate,* 14**,** 3801-3818.

Chiew, F., T. Piechota, J. Dracup and T. McMahon 1998. El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *journal of Hydrology,* 204**,** 138-149.

Chiew, F., W. Young, W. Cai and J. Teng 2011. Current drought and future hydroclimate projections in southeast Australia and implications for water resources management. *Stochastic Environmental Research and Risk Assessment,* 25**,** 601-612.

Chowdhury, R.K. and S. Beecham 2013. Influence of SOI, DMI and Niño3.4 on South Australian rainfall. *Stochastic Environmental Research and Risk Assessment*.

Compo, G.P., J.S. Whitaker, P.D. Sardeshmukh, N. Matsui, R.J. Allan, X. Yin, B.E. Gleason, R.S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R.I. Crouthamel, A.N. Grant, P.Y. Groisman, P.D. Jones, M.C. Kruk, A.C. Kruger, G.J. Marshall, M. Maugeri, H.Y. Mok, Ø. Nordli, T.F. Ross, R.M. Trigo, X.L. Wang, S.D. Woodruff and S.J. Worley 2011. The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society,* 137**,** 1-28.

CSIRO 2012. Climate and water availability in south-eastern Australia: A synthesis of findings from Phase 2 of the South Eastern Australian Climate Initiative (SEACI). Australia.

Dai, A. 2010. Drought under global warming: a review. *Wiley Interdisciplinary Reviews: Climate Change,* 2**,** 45-65.

De'ath, G. and K. Fabricius 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology,* 81**,** 3178-3192.

Degirmendzic, J. and J. Widbig 2007. Jet strem patterns over Europe in the period 1950-2001 - classification and basic statistical properties. *Theoretical and Applied Climatology,* 88**,** 149-167.

Doblas-Reyes, F., A. Weisheimer, M. Deque, N. Keenlyside, M. McVean, J. Murphy, P. Rogel, D. Smith and T. Palmer 2009. Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts. *Quarterly Journal of the Royal Meterological Socity,* 135**,** 1538-1559.

Dommenget, D. and M. Latif 2002. A Cautionary Note on the Interpretation of EOFs. *Journal of Climate,* 15**,** 216-225.

Donohue, R., T. McVicar, L. Li and M. Roderick 2010a. A data resource for analysing dynamics in Australian ecohydrological conditions. *Austral Ecology,* 35**,** 593-594.

Donohue, R., M. Roderick and T. McVicar 2010b. Can dynamical vegetation information improve the accuracy of Budyko's hydrological model? *journal of Hydrology,* 390**,** 23-34.

Dore, M.H. 2005. Climate change and changes in global precipitation patterns: what do we know? *Environ Int,* 31**,** 1167-81.

Drosdowsky, W. 1993. An analysis of Austraian seasonal rainfall anomalies : 1950 - 87. II: Temporal variability and teleconnection patterns. *International Journal of Climatology,* 13**,** 111-149.

Drosdowsky, W. 2005. The Latitude of the Subtropical Ridge over Eastern Australia: The L Index revisited. *International Journal of Climatology,* 25**,** 1291-1299.

Drosdowsky, W. and L. Chambers 2001. Near-Global Sea Surface Temerature Anomalies as Predictors of Australian Seasonal Rainfall. *Journal of Climate,* 14**,** 1677-1688.

Drosdowsky, W. and M. Williams 1991. The Southern Oscillation in the Australian Region. Part 1: Anomalies at the Extremes of the Oscillation. *Journal of Climate,* 4**,** 619-638.

Enke, W., F. Schneider and T. Deutschlander 2005. A novel scheme to derive optimized circulation pattern classifications for downscaling and forecast purposes. *Theoretical and Applied Climatology,* 82**,** 51-63.

Esther, A., J. Groeneveld, N. Enright, B. Miller, B. Lamont, G. Perry, B. Blank and F. Jeltsch 2010. Sensitivity of plant functional types to climate change: classification tree analysis of a simulation model. *Journal of Vegetation Science,* 21**,** 447-461.

Evans, J. and R. Allan 1992. El Nino/Southern Oscillation modification to the structure of the monsoon and tropical cyclone activity in the Australiasian region. *International Journal of Climatology,* 12**,** 611-623.

Evans, J. and M. McCabe 2010. Regional climate simulation over Australia's Murray-Darling basin: A multitemporal assessment. *Journal of Geophysical Research,* 115**,** D14114.

Fan, K. and H. Liu 2013. Evaluation of Atmospheric Circulation in the Southern Hemisphere in 20Crv2. *Atmospheric and Oceanic Science Letters,* 6**,** 337-342.

Fawcett, R. 2010. Seasonal climate summary southern hemisphere (spring 2009): rapid intensification of El Nino, drier than average in northern and eastern Australia and warmer than average throughout. *Australian Meteorological and Oceanographic Journal,* 60**,** 127-137.

Firth, L., M. Hazelton and E. Campbell 2005. Predicting the Onset of Australian Winter Rainfall by Nonlinear Classification. *Journal of Climate,* 18**,** 772-781.

Fu, G., S. Chen, C. Liu and D. Shepard 2004. Hydro-climate trends of the Yellow River Basin for the last 50 years. *Climatic Change,* 65**,** 149-178.

GarcÌa-GarcÌa, D., C. Ummenhofer and V. Zlotnicki 2011. Australian water mass variations from GRACE data linked to Indo-Pacific climate variability. *Remote Sensing of Environment,* 115**,** 2175-2183.

Goddard, L., S.J. Mason, S.E. Zebiak, C.F. Ropelewski, R. Basher and M.A. Cane 2001. Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology,* 21**,** 1111-1152.

Hastie, T., R. Tibshirani and J. Friedman 2001. *The Elements of Statistical Learning,* New York, Springer-Verlag.

Hope, P.K., W. Drosdowsky and N. Nicholls 2006. Shifts in the synoptic systems influencing southwest Western Australia. *Climate Dynamics,* 26**,** 751-764.

Horel, J.D. and J.M. Wallace 1981. Planetary-Scale Atmospheric Phenomena Associated with the Southern Oscillation. *Monthly Weather Review,* 109**,** 813-829.

Japan Agency for Marine-Earth Science and Technology. 2008. *Indian Ocean Dipole (IOD)* [Online]. http://www.jamstec.go.jp/frsgc/research/d1/iod/iod_home.html.en. [Accessed 06/10/12].

Jones, D., W. Wang and R. Fawcett 2009. High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal,* 58**,** 233-248.

Kala, J., T. Lyons and U. Nair 2011. Numerical Simulations of the Impacts of Land-Cover Change on Cold Fronts in South-West Western Australia. *Boundary-Layer Meteorology,* 139**,** 121-138.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne and D. Joseph 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society,* 77**,** 437-471.

Kao, H. and J. Yu 2009. Contrasting Eastern-Pacific and Central-Pacific Types of ENSO. *Journal of Climate,* 22**,** 615-632.

Kennard, M., B. Pusey, J. Olden, S. Mackay, J. Stein and N. Marsh 2010. Classification of natural flow regimes in Australia to support environmental flow management. *Freshwater Biology,* 55**,** 171-193.

Kiem, A. and D. Verdon-Kidd 2010. Towards understanding hydroclimatic change in Victoria, Australia - preliminary insights into the "Big Dry". *Hydrological Earth Systems Science,* 14**,** 433-445.

King, A.D., L.V. Alexander and M.G. Donat 2013. The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia. *International Journal of Climatology,* 33**,** 2376-2387.

Kirono, D., F. Chiew and D. Kent 2010. Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. *Hydrological Processes,* 24**,** 1237-1247.

Kumar, A. and M. Hoerling 1997. Interpretation and Implications of the Observed Inter-El Nino Varibility. *Journal of Climate,* 10**,** 83-91.

Lacey, G. and R. Grayson 1998. Relating baseflow to catchment properties in south-eastern Australia. *journal of Hydrology,* 204**,** 231-250.

Larsen, S.H. and N. Nicholls 2009. Southern Australian rainfall and the subtropical ridge: Variations, interrelationships, and trends. *Geophysical Research Letters,* 36.

Latif, M., D. Anderson, T. Barnett, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brian, A. Rosati and E. Schneider 1998. A review of the predictability and prediction of ENSO. *Journal of Geophysical Research,* 103**,** 14375-14393.

Lau, N. and M. Nath 2003. Atmosphere-Ocean Variations in the Indo-Pacific Sector during ENSO episodes. *Journal of Climate,* 16**,** 3-20.

Lawler, J., D. White, R. Neilson and A. Blaustein 2006. Predicting climate-induced shifts: model differences and model reliability. *Global change biology,* 12**,** 1568-1584.

Leblanc, M., P. Tregoning, G. Ramillien, S. Tweed and A. Fakes 2009. Basin-scale integrated observations of the early 21st century multiyear drought in southeast Australia. *Water Resources Research,* 45**,** W04408.

Leblanc, M., S. Tweed, A. Van Dijk and B. Timbal 2012. A review of historic and future hydrological changes in the Murray-Darling Basin. *Global and Planetary Change,* 80-81**,** 226-246.

Li, X.S. and D. Sailor 2000. Application of tree-structured regression for regional precipitation prediction using general circulation model output. *Climate Research,* 16**,** 17-30.

Liaw, A. and M. Wiener 2002. Classificaton and Regression by randomForest. *R News,* 2**,** 18-22.

Lim, E., H. Hendon, D. Anderson, A. Charles and O. Alves 2011. Dynamical, Statistical-Dynamical and Multimodel Ensemble Forecasts of Australian Spring Season Rainfall. *Monthly Weather Review,* 139**,** 958-975.

Liu, Y., A. van Dijk, R. de Jeu and T. Holmes 2009. An analysis of spatiotemporal variations of soil and vegetation moisture from a 29-year satellite-derived data set over mainland Australia. *Water Resources Research,* 45**,** WR007187.

Lu, J., G. Chen and D. Frierson 2008. Responce of the Zonal Mean Atmospheric Circulation to El Nino versus Global Warming. *Journal of Climate,* 21**,** 5835-5851.

Lyon, B. 2004. The strength of El Niño and the spatial extent of tropical drought. *Geophysical Research Letters,* 31**,** L21204.

Maindonald, J. and J. Braun 2010. *Data Analysis and Graphics Using R - An Example-Based Approach,* United Kingdom, Cambridge University Press.

Marshall, G. 2003. Trends in the Southern Annular Mode from Observation and Reanalysis. *Journal of Climate,* 16**,** 4134-4143.

McBride, J. and N. Nicholls 1983. Seasonal Relationships between Australian Rainfall and the Southern Oscillation. *Monthly Weather Review,* 111**,** 1998-2004.

McGregor, S., N. Ramesh, P. Spence, M.H. England, M.J. McPhaden and A. Santoso 2013. Meridional movement of wind anomalies during ENSO events and their role in event termination. *Geophysical Research Letters,* 40**,** 749-754.

McGregor, S., A. Timmermann, N. Schneider, M.F. Stuecker and M.H. England 2012. The Effect of the South Pacific Convergence Zone on the Termination of El Niño Events and the Meridional Asymmetry of ENSO*. *Journal of Climate,* 25**,** 5566-5586.

Meneghini, B., I. Simmonds and I. Smith 2007. Association between Australian rainfall and the Southern Annular Mode. *International Journal of Climatology,* 27**,** 109-121.

Meyers, G. 1996. Variation of Indonesian throughflow and the El Niño-Southern Oscillation. *Journal of Geophysical Research,* 101**,** 1225-12263.

Meyers, G., P. McIntosh, L. Pigot and M. Pook 2007. The Years of El Niño, La Niña and Interactions with the Tropical Indian Ocean. *Journal of Climate,* 20**,** 2872-2880.

Mueller, L., U. Schindler, W. Mirschel, T. Shepherd, B. Ball, K. Helming, J. Rogasik, F. Eulenstein and H. Wiggering 2010. Assessing the productivity function of soils. A review. *Agronomy for Sustainable Development,* 30**,** 601-614.

Murphy, B. and B. Timbal 2008. A review of recent climate variability and climate change in southeastern Australia. *International Journal of Climatology,* 28**,** 859-879.

National Climate Data Centre. 2013. *Sea Surface Temperatures (SST)* [Online]. National Oceanic and Atmospheric Administration National Climatic Data Center. Available: http://www.ncdc.noaa.gov/teleconnections/enso/indicators/sst.php [Accessed 13 May 2013].

National Land and Water Resources Audit 2001. Australian Water Resources Assessment 2000. National Land and Water Resources Audit. Canberra, Australia.

National Water Commission. 2005. *What is our total water resource?* [Online]. Available: http://www.water.gov.au/WaterAvailability/Whatisourtotalwaterresource/Runoff/index.aspx?Menu=Level1_3_1_5 [Accessed 15 May 2013.

Nicholls, N. 1988. El Nino-Southern Oscillation and Rainfall Variability. *Journal of Climate,* 1**,** 418-421.

Nicholls, N. 1989. Sea Surface Temperatures and Australian Winter Rainfall. *Journal of Climate,* 2**,** 965-973.

Nicholls, N., W. Drosdowsky and B. Lavery 1997. Australian rainfall variability and change. *Weather,* 52**,** 66-72.

Peixoto, J. and A. Oort 1992. *Physics of Climate*, Springer.

Pepler, A., A. Coutts-Smith and B. Timbal 2013. The role of East Coast Lows on rainfall patterns and inter-annual variability across the East Coast of Australia. *International Journal of Climatology***,** n/a-n/a.

Perry, M. and J. Niemann 2007. Analysis and estimation of soil moisture at the catchment scale using EOFs. *journal of Hydrology,* 334**,** 388-404.

Pittock, A. 1973. Global meridional interactions in stratosphere and troposphere. *Quarterly Journal of the Royal Meterological Socity,* 99**,** 424-437.

Pittock, A. 1975. Climatic Change and the Patterns of Variation in Australian Rainfall. *Search,* 6**,** 498-504.

Pook, M., S. Lisson, J. Risbey, C. Ummenhofer, P. McIntosh and M. Rebbeck 2009. The autumn break for cropping in southeast Australia: trends, synoptic influences and impacts on wheat yield. *International Journal of Climatology,* 29**,** 2012-2026.

Potter, N.J. and F.H.S. Chiew 2011. An investigation into changes in climate characteristics causing the recent very low runoff in the southern Murray-Darling Basin using rainfall-runoff models. *Water Resources Research,* 47**,** n/a-n/a.

Power, S., T. Casey, C. Folland, A. Colman and V. Mehta 1999. Inter-decadal modulation of the impact of ENSO on Australia. *Climate Dynamics,* 15**,** 319-324.

Power, S., M. Haylock, R. Colman and X. Wang 2006. The predictability of interdecadal changes in ENSO activity and ENSO teleconnections. *Journal of Climate,* 19**,** 4755-4771.

Power, S., F. Tseitkin, S. Torok, B. Lavery, R. Dahni and B. McAvaney 1998. Australian temperature, Australian rainfall and the Southern Oscillation, 1910-1992: coherent variability and recent changes. *Australian Meteorological Magazine,* 47**,** 85-101.

Prasad, A.M., L.R. Iverson and A. Liaw 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems,* 9**,** 181-199.

R Development Core Team 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: http://www.R-project.org/.

Rakich, C., N. Holbrook and B. Timbal 2008. A pressure gradient metric capturing planetary-scale influences on eastern Australian rainfall. *Geophysical Research Letters,* 35**,** L08713.

Ralph, T. and P. Hesse 2010. Downstream hydrogeomorphic changes along the Macquarie River, southeastern Australia, leading to channel breakdown and floodplain wetlands. *Geomorphology,* 118**,** 48-64.

Rasmusson, E. and T. Carpenter 1983. The Relationship Between Easter Equatorial Pacifc Sea Surface Temperatures and Rainfall over India and Sri Lanka. *Monthly Weather Review,* 111**,** 517-528.

Raupach, M., P. Briggs, V. Haverd, E. King, M. Paget and C. Trudinger 2008. Australian Water Availability Project (AWAP). CSIRO Marine and Atmospheric Research Component: Final Report for Phase 3. Canberra: CSIRO Marine and Atmospheric Research.

Rayner, N.A., D. Parker, E. Horton, C. Folland, L. Alexander and D. Rowell 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research,* 108**,** D144407.

Risbey, J., M. Pook, P. McIntosh, M. Wheeler and H. Hendon 2009. On the remote drivers of rainfall variability in Australia. *Monthly Weather Review,* 137**,** 3233-3253.

Rodionov, S. and R. Assel 2000. Atmospheric teleconnection patterns and severity of winters in the Laurentian Great Lakes Basin. *Atmosphere-Ocean,* 38**,** 601-635.

Rodionov, S., R. Assel and L. Herche 2001. Tree-Structured Modelling of the Relationship Between Great Lakes Ice Cover and Atmospheric Circulation Patterns. *International Association of Great Lakes Research,* 27**,** 486-502.

Rodionov, S., J. Overland and N. Bond 2005. The Aleutian Low and Winter Climatic Conditions in the Bering Sea. Part 1: Classification *Journal of Climate,* 18**,** 160-177.

Ropelewski, C. and M. Halpert 1987. Global and Regional Scale Precipitation Patterns Associated with the El Niño/Southern Oscillation. *Monthly Weather Review,* 115**,** 1606-1626.

Saji, N., B. Goswami, P. Vinayachandran and T. Yamagata 1999. A dipole mode in the tropical Indian Ocean. *Nature,* 401**,** 360-363.

Saji, N. and T. Yamagata 2003. Structure of SST and Surface Wind Variability during Indian Ocean Dipole Mode Events: COADS Observations. *Journal of Climate,* 16**,** 2735-2751.

Seidel, D., Q. Fu, W. Randel and T. Reichler 2007. Widening of the tropical belt in a changing climate. *Nature Geoscience,* 1**,** 21-24.

Seneviratne, S.I., N. Nicholls, D. Easterling, C.M. Goodess, S. Kanae, J. Kossin, J. Luo, J. Marengo, K. McInnes, M. Rahimi, M. Reichstein, A. Sorteberg, C. Vera and X. Zhang 2012. Change in climate extremes and their impacts on the natural physical environment. *In:* Field, C., Barros, V., Stocker, T., Qin, D., Dokken, D., Ebi, K.,

Mastrandrea, M., Mach, K., Plattner, G., Allen, M. R., Tignor, M. & Midgley, P. (eds.) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation.* A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change: Cambridge University Press, Cambridge, UK and New York, NY, USA.

Sharma, A., K. Luk, I. Cordery and U. Lall 2000. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 2 - Predictor identification of quarterly rainfall using ocean-atmosphere information. *journal of Hydrology,* 239**,** 240-248.

Shi, G., J. Ribbe, W. Cai and T. Cowan 2008. An interpretation of Australian rainfall projections. *Geophysical Research Letters,* 35**,** L02702.

Simmonds, I. and P. Hope 1997. Persistence Characteristics of Australian Rainfall Anomalies. *International Journal of Climatology,* 17**,** 597-613.

Simmonds, I. and A. Rocha 1991. The Association with Australian Winter Climate with Ocean Temperatures to the West. *Journal of Applied Meteorology,* 4**,** 1147-1161.

Smith, D.I. 1998. *Water in Australia. Resources and Management,* South Melbourne, Oxford University Press.

Speer, M., L. Leslie and A. Fierro 2009. Australian east coast rainfall decline related to large scale climate drivers. *Climate Dynamics*.

Stevens, A. 2012. Introduction to the basic drivers of climate. *Nature Education Knowedge,* 3**,** 10.

Sturman, A. and N. Tapper 2005. *The Weather and Climate of Australia and New Zealand. Second Edition.,* Melbourne, Oxford University Press.

Suppiah, R. 2004. Trends in the southern oscillation phenomenon and Australian rainfall and changes in their relationship. *International Journal of Climatology,* 24**,** 269-290.

Suppiah, R. and K. Hennessy 1998. Trends in total rainfall, heavy rain events and number of dry day in Australia, 1910-1990. *International Journal of Climatology,* 10**,** 1141-1164.

Therneau, T., B. Atkinson and R port by Brian Ripley 2010. rpart: Recursive Partitioning. R package version 3.1-46: http://CRAN.R-project.org/package=rpart.

Thuiller, W., M. Araujo and S. Lavorel 2003. Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science,* 14**,** 669-680.

Timbal, B. 2009. The continuing decline in South-East Australian rainfall - Update to May 2009. *CAWCR Research Letters,* Issue 2 4-12.

Timbal, B., J. Arblaster, K. Braganza, E. Fernandez, H. Hendon, B. Murphy, M. Raupach, C. Rakich, I. Smith, K. Whan and M. Wheeler 2010. Understanding the anthropogenic nature of the observed rainfall decline across South Eastern Australia. *The CAWCR - Bureau of Meteorology Contribution to SEACI -1 Theme 1 From January 2006 to June 2009.* Melbourne: SEACI.

Timbal, B. and W. Drosdowsky 2013. The relationship between the decline of South Eastern Australia rainfall and the strengthening of the sub-tropical ridge. *International Journal of Climatology,* 33**,** 1021-1034.

Timbal, B. and H. Hendon 2011. The role of tropical modes of variability in recent rainfall deficits across the Murray-Darling Basin. *Water Resources Research,* 47**,** 16.

Timbal, B. and D. Jones 2008. Future projections of winter rainfall in southeast Australia using a statistical downscaling technique. *Climatic Change,* 86**,** 165-187.

Timbal, B. and B. Murphy 2007. Observed climate change in South-East of Australia and its relation to large-scale modes of variability. *BRMC Research Letters,* 6**,** 6-11.

Timbal, B., K. Whan and M. Raupach 2009. Climate change influence of changes in evapotranspiration, runoff and drainage across SEA, through both physical and ecological processes. *Final Report for Project 1.3.1P.* South Eastern Australian Climate initiative.

Trenberth, K. 1990. Recent Observed Interdecadal Climate Changes in the Northern Hemisphere. *Bulletin of the American Meteorological Society,* 71**,** 988-993.

Trenberth, K. 1997. The Definition of El Niño. *Bulletin of the American Meteorological Society,* 78**,** 2771-2777.

Trenberth, K. and D. Stepaniak 2001. Indices of El Nino Evolution. *Journal of Climate,* 14**,** 1697-1701.

Troup, A. 1965. The Southern Oscillation. *Quarterly Journal of the Royal Meterological Socity,* 91**,** 73-76.

Ummenhofer, C., M. England, P. McIntosh, G. Meyers, M. Pook, J. Risbey, A. Gupta and A. Taschetto 2009a. What causes southeast Australia's worst droughts? *Geophysical Research Letters,* 36**,** L04706.

Ummenhofer, C., A. Sen Gupta, P. Briggs, M. England, P. McIntosh, G. Meyers, M. Pook, M. Raupach and J. Risbey 2009b. Indian and Pacific Ocean Influences on Southeast Australian Drought and Hydrology. *Journal of Climate,* submitted.

Ummenhofer, C., A. Sen Gupta, P. Briggs, M. England, P. McIntosh, G. Meyers, M. Pook, M. Raupach and J. Risbey 2011. Indian and Pacific Ocean Influences on Southeast Australian Drought and Soil Moisture. *Journal of Climate,* 24**,** 1313-1366.

van Dijk, A. and L. Renzullo 2011. Water resource monitoring systems and the role of satellite observations. *Hydrological Earth Systems Science,* 15**,** 39-55.

Vecchi, G., B. Soden, A. Wittenberg, I. Held, A. Leetmaa and M. Harrison 2006. Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing. *Nature,* 441**,** 73-76.

Verdon-Kidd, D. and A. Kiem 2009a. Nature and causes of protracted droughts in southeast Australia: Comparison between the Federation, WWII and Big Dry droughts. *Geophysical Research Letters,* 36**,** L22707.

Verdon-Kidd, D. and A. Kiem 2009b. On the relationship between large-scale climate modes and regional synoptic patterns that drive Victorian rainfall. *Hydrological Earth Systems Science,* 13**,** 467-479.

Wang, C. and P. Fiedler 2006. ENSO variability and the eastern tropical Pacific: A review. *Progress in Oceanography,* 69**,** 239-266.

Wang, G. and H. Hendon 2007. Sensitivity of Australian Rainfall to Inter-El Niño Variations. *Journal of Climate,* 20**,** 4211-4226.

Wang, Q.J., D.E. Robertson and F.H.S. Chiew 2009. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research,* 45**,** W05407.

Wei, W. and D.W. Watkins 2011. Data mining methods for hydroclimatic forecasting. *Advances in Water Resources,* 34**,** 1390-1400.

Whan, K., B. Timbal and J. Lindesay 2013. A CART analysis of the joint impact of the sub-tropical ridge intensity and position on South Eastern Australia rainfall. *International Journal of Climatology,* In Press.

Williams, A. and R. Stone 2009. An assessment of relationships between the Australian subtropical ridge, rainfall variability and high-latitude circulation patterns. *International Journal of Climatology,* 29**,** 691-709.

Yu, J., G. Fu, W. Cai and T. Cowan 2010. Impacts of precipitation and temperature chnages on annual streamflow in the Murray-Darling Basin. *Water International,* 35**,** 313-323.

Zhou, Y.P., K. Xu, Y.C. Sud and A.K. Betts 2011. Recent trends of the tropical hydrological cycle inferred from Global Precipitation Climatology Project and International Satellite Cloud Climatology Project data. *Journal of Geophysical Research,* 116**,** D015197.

Zorita, E., J.P. Hughes, D.P. Lettemaier and H. von Storch 1995. Stochastic Characterization of Regional Circulation Patterns for Climate Model Diagnosis and Estimation of Local Precipitation. *Journal of Climate,* 8**,** 1023-1042.