SPATIO-TEMPORAL HIDDEN MARKOV MODELS FOR INCORPORATING INTER-ANNUAL VARIABILITY IN RAINFALL

Andrew James Frost B.E. (Env.) (Hons.)

A thesis submitted for the degree of Doctor of Philosophy at The University of Newcastle

January 2004

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

_____

Andrew James Frost

Acknowledgements

Firstly, my thanks go to my supervisor George Kuczera, whose encouragement and faith allowed me to follow the winding path I did.

To Mark Thyer, who showed remarkable grace under constant fire during the final stretch of his PhD tenure, my thanks for your patience.

To our DRIP counterparts at the University of Adelaide: Martin Lambert, Shane Jennings (and the Jennings clan) and Andrew Metcalfe, my thanks for the constructive criticism, enjoyable sojourns and Barossa reds (not necessarily in that order).

My thanks go to Kerrie Mengersen from the Department of Statistics here at Newcastle for providing statistical insight and advice on mixture modelling. To Sri Srikanthan of the Bureau of Meteorology for providing the annual rainfall data used in the HMM case studies from the original *Srikanthan et al.* (2002) study.

To my friends and colleagues at the Uni of Newcastle: Stewart Franks, Andrew Krause, Matthew McCabe, Anthony Kiem, Brook Ewers, Dmitri Kavetski, Petras Silinis, Tom Micevski, Adam Wyatt and everyone else; my thanks for your help with my work, and for all the other things that are not encompassed by that.

To my parents Mal and Helen and the rest of my family for your encouragement.

Last but in no means least, to Lana Collison whose miscellaneous support allowed me to complete this thesis.

Table of Contents

Abstract

Two new spatio-temporal hidden Markov models (HMM) are introduced in this thesis, with the purpose of capturing the persistent, spatially non-homogeneous nature of climate influence on annual rainfall series observed in Australia. The models extend the two-state HMM applied by *Thyer* (2001) by relaxing the assumption that all sites are under the same climate control. The Switch HMM (SHMM) allows at-site anomalous states, whilst still maintaining a regional control. The Regional HMM (RHMM), on the other hand, allows sites to be partitioned into different Markovian state regions.

The analyses were conducted using a Bayesian framework to explicitly account for parameter uncertainty and select between competing hypotheses. Bayesian model averaging was used for comparison of the HMM and its generalisations.

The HMM, SHMM and RHMM were applied to four groupings of four sites located on the Eastern coast of Australia, an area that has previously shown evidence of inter-annual persistence. In the majority of case studies, the RHMM variants showed greatest posterior weight, indicating that the data favoured the multiple region RHMM over the single region HMM or the SHMM variants. In no cases does the HMM produce the maximum marginal likelihood when compared to the SHMM and RHMM.

The HMM state series and preferred model variants were sensitive to the parameterisation of the small-scale site-to-site correlation structure. Several parameterisations of the small-scale Gaussian correlation were trialled, namely Fitted Correlation, Exponential Decay Correlation, Empirical and Zero Correlation. Significantly, it was shown that annual rainfall data outliers can have a large effect on inference for a model that uses Gaussian distributions.

The practical value of this modelling is demonstrated by the conditioning of the event based point rainfall model DRIP on the hidden state series of the HMM variants. Short timescale models typically underestimate annual variability because there is no explicit structure to incorporate long-term persistence. The two-state conditioned DRIP model was shown to reproduce the annual variability observed to a greater degree than the single state DRIP.

Notation

Probability Notation

$p(\cdot)$	generalised probability density function
$p(\cdot \mid \cdot)$	generalised conditional probability density function
$P(\cdot)$	generalised event probability function
$\boldsymbol{\theta}$	general term for model parameters vector with support $\boldsymbol{\theta} \in \Theta$
y	general term for the observed data
y_t^d	observed data scalar at time t and site d
$\mathbf{y}_t$	observed data vector at time t over d sites, $\mathbf{y}_t = (y_t^1, ..., y_t^d)$
$\mathbf{Y}_1^T$	general term for the set of observed vector data, $\mathbf{Y}_1^T = (\mathbf{y}_1, ..., \mathbf{y}_T)$
M	general term for model hypothesis
$\overline{y}$	empirically estimated mean of data $Y_1^T = (y_1, ..., y_T)$
$\overline{s}^2$	empirically estimated variance of data $Y_1^T = (y_1, ..., y_T)$

Bayesian modelling notation

$p(\boldsymbol{\theta} \mid M)$	prior distribution of parameters given model hypothesis M
$p(\boldsymbol{\theta} \mid \mathbf{y}, M)$	posterior distribution of parameters given data $\mathbf{y}$ and model hypothesis M
$p(\mathbf{y} \mid \boldsymbol{\theta}, M)$	likelihood of data given parameters $\boldsymbol{\theta}$ and model hypothesis M, alternatively written as a function $f(\mathbf{y} \mid M, \boldsymbol{\theta})$
$p(\mathbf{y} \mid M)$	marginal likelihood of data given model hypothesis M
$p(M \mid \mathbf{y})$	posterior model hypothesis M probability given data $\mathbf{y}$

MCMC Sampling Notation

$\boldsymbol{\theta}^{(i)}$	MCMC parameter sample i
ns	number of samples
nm	number of models
$J_i(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(i-1)})$	proposal/jump distribution of parameters $\boldsymbol{\theta}^*$ given previous sample location $\boldsymbol{\theta}^{(i-1)}$

Hidden State Markov Model Notation

$\mathbf{P}$	Markovian state transition probability matrix
p_{ij}	regional state transition probability that represents the probability of moving from state i to state j where $i, j \in W, D$

W	wet state
D	dry state
r_t	regional hidden state at time t
R_1^T	regional hidden state time series $R_1^T = (r_1, ..., r_T)$
$\mu_{r_t}^{site}$	site mean at time t for given state hidden state r_t
$\sigma_{r_t}^{site}$	site standard deviation at time t for given state hidden state r_t
ρ_{ij}	correlation coefficient between sites i and j
$\boldsymbol{\rho}$	correlation matrix of size $d \times d$, $\boldsymbol{\rho} = \left[\rho_{ij} \right] \quad i, j = 1, ..., d$
$\boldsymbol{\mu}_{r_t}$	site mean vector of size d at time t for given regional state r_t
$\boldsymbol{\Sigma}_{r_t}$	site covariance matrix at time t of size $d \times d$ for given regional state r_t
λ	exponential correlation decay range parameter $\rho_{ij} = \exp(-dist_{ij}/\lambda)$ where $dist_{ij}$ is the distance between sites i and j
τ^2	microscale variance, nugget effect

Switch Hidden State Markov Model Notation

$\mathbf{SP}$	state switch probability matrix
sp_{ij}^{site}	site state switch probability that represents the probability of moving from regional state i to site state j where $i, j \in W, D$ for site $site \in \{1, ..., d\}$, $\mathbf{SP} = \left[\; sp_{ij}^{site} \right]$
s_t^{site}	site hidden state at time t for site $site \in \{1, ..., d\}$
$\mathbf{s}_t$	site hidden state vector at time t over d sites, $\mathbf{s}_t = (s_t^1, s_t^2, ..., s_t^d)$
$\mu_{s_t}^{site}$	site mean at time t for given state hidden state s_t
$\sigma_{s_t}^{site}$	site standard deviation at time t for given state hidden state s_t
$\boldsymbol{\mu}_{S_t}$	site mean vector of size d at time t for given regional state vector S_t
$\boldsymbol{\Sigma}_{S_t}$	site covariance matrix at time t of size $d \times d$ for given regional state vector S_t

Regional Hidden State Markov Model Notation

p_{ij}^{reg}	regional state transition probability that represents the probability of moving from state i to state j where $i, j \in W, D$ for region $reg \in \{1, ..., k\}$ where k is the number of regions. For HMM and SHMM $k = 1$. $\mathbf{P} = \left[\; p_{ij}^{reg} \right]$
r_t^{reg}	regional hidden state at time t for region $reg \in \{1, ..., k\}$
$\mathbf{r}_t$	regional hidden state vector at time t over k regions, $\mathbf{r}_t = (r_t^1, r_t^2, ..., r_t^k)$

reg_{site}	region into which site is partitioned, $reg \in \{1,...,k\}$
H	site-region partition vector over d sites, $H = (reg_1, reg_2, ..., reg_d)$
$\mu_{r_t}^{site}$	site mean at time t for given state hidden state r_t
$\sigma_{r_t}^{site}$	site standard deviation at time t for given state hidden state r_t
$\boldsymbol{\mu}_{\mathbf{r}_t}$	site mean vector of size d at time t for given regional state vector $\mathbf{r}_t$.
$\boldsymbol{\Sigma}_{\mathbf{r}_t}$	site covariance matrix at time t of size $d \times d$ for given regional state vector $\mathbf{r}_t$.

Probability Distribution Notation

$Uniform(0,1)$	uniform probability distribution with limits 0 and 1
$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	multivariate Gaussian distribution with dimensions depending on context, mean vector $\boldsymbol{\mu}$, symmetric positive definite covariance matrix $\boldsymbol{\Sigma}$
$f_N(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	multivariate Gaussian density function for random vector $\mathbf{y}$ with dimensions depending on context, given mean vector $\boldsymbol{\mu}$ and symmetric positive definite covariance matrix $\boldsymbol{\Sigma}$. Also written as $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
$\Phi(\cdot)$	Standard cumulative Gaussian probability
$Inv \sim \chi^2(v, s^2)$	scaled inverse-chi-square distribution with degrees of freedom v and scale s^2
$f_{Inv-\chi^2}(x^2; \nu, s^2)$	scaled inverse-chi-square density for scalar random variable x^2 with degrees of freedom v and scale s^2. Also written as $x^2 \sim Inv-\chi^2(v, s^2)$
$Beta(\alpha, \beta)$	Beta distributions with parameters α and β
$I[\cdot]$	An indicator function with value 1 when the statement contained within is true
$Gamma(\alpha, \beta)$	Gamma distribution with shape α and inverse scale β

Chapter 1 Introduction

The design of dams, floodways and other infrastructure affected by rainfall would be a straightforward task if there existed records long enough to capture the inherent temporal variability caused by climate influence. In flood studies for example, assuming future climate is similar to the past, a rainfall-runoff model using historic rainfall data could be used to simulate the range of flood scenarios that could occur in the future.

However, climate variability is a dynamic process, with large scale effects occurring over years, decades and centuries. Given that the maximum length in most regions for daily rainfall in Australia is around 100 years, and for pluviograph (6 minute) data around 30 years, it is not likely that this source of variability has been adequately sampled. Consequently, predictions and simulations using only this data may overestimate the degree of certainty with which the simulations are made. To overcome the limitations of short historic records stochastic models calibrated to historic data are used to provide insight into the range of events that can occur over planning horizons.

1.1 Modelling the Influence of Climate on Rainfall

The current generation of stochastic point rainfall models operating at daily and sub-daily timescales do not adequately capture the variability present in Australian hydroclimatic time series. In particular, mechanisms to incorporate inter-annual variability and persistence are rarely accounted for. Of the models that do account for this persistence, the spatially varying nature of this long-term persistence has not yet been addressed. The lack of current modelling techniques to explicitly accommodate this spatially non-homogeneous persistence provides the motivation for the work undertaken in this thesis.

Past modelling approaches (see section 2.4) have generally focussed on modelling either short-term spatially homogeneous climate influences or long-term climate influences at a point. However, long-term variation of rainfall induced by a non-homogeneous spatial climate influence has not been modelled explicitly. Although long term climate influences with varying spatial effects (e.g. El Niño Southern Oscillation) have been identified, the current generation of stochastic models does not contain a

conceptual mechanism to incorporate such an effect. This may be considered a shortcoming of current approaches.

Given there is a need to simulate short-timescale rainfall for the design of water-based infrastructure, and that rainfall is influenced by long-term variability of climate, a new spatial framework for downscaling long-term persistence into rainfall simulation down to 6-minute time scales is desired.

1.2 Objectives of Thesis

It is the main goal of this thesis to address the spatially varying long-term effects of climate on rainfall over a range of timescales ranging from sub hourly to decadal timescales.

This overall goal is broken down into three objectives;

- To incorporate a statistically rigorous parameter and model uncertainty framework.
- To develop models that can identify spatially non-homogeneous climate persistence in rainfall on timescales greater than a year, and can then be used to condition smaller timescale models used in design; and
- To condition hierarchically a smaller timescale stochastic rainfall model on the persistence structure identified by the larger timescale persistence model.

1.3 Thesis Outline

Chapter 2 reviews methods of modelling the influence of long-term climate variability on rainfall. Particular emphasis is directed at the two-state hidden Markov model (HMM) applied by *Thyer* (2001). This model has been shown to capture inter-annual persistence in annual rainfall in Australia. However, the current HMM is limited in that it assumes a common climate influence across all sites used in an analysis. Can sites far away from each other be assumed to be under the same climatic control? A generalisation of the HMM is required to accommodate the spatially varying effects of climate.

Because generalisation of the HMM implies a choice between models, an objective model selection method will need to be employed to see if the generalisations are worthwhile. The first objective of this thesis, to use a model framework that coherently incorporates parameter and model uncertainty, is investigated in Chapter 3. The model selection method used in this thesis, Bayesian model selection, allows the simple comparison of model probabilities, whilst also maintaining the advantage of incorporating parameter uncertainty. This general modelling framework has been rarely used in water resource studies. Therefore in Chapter 3 this framework is explained in detail, with a simple case study demonstrating some of the limitations of Bayesian model selection, along with comparison to some other model selection criteria. The Bayesian model selection method is applied in subsequent chapters to select between competing model hypotheses.

Chapter 4 proposes two generalisations of the multiple-site HMM, the Switch HMM (SHMM) and the Regional HMM (RHMM). The Switch HMM allows individual sites to exhibit anomalous behaviour relative to the overall climate control, while the Regional HMM conceptualises annual rainfall as being controlled by a regional climate state with each site being assigned to a particular climate region. These formulations were motivated by the observation that not all sites are affected to the same degree by the same climatic oscillations; as sites become more separated they are less likely to be influenced by the same climate controls. These generalisation address the second objective, by allowing spatially non-homogeneous persistence effects.

Chapter 5 details the application and comparison of the HMM generalisations to four case studies located around Sydney and Brisbane in Eastern Australia. Each case study consists of four sites, with Bayesian model selection being used to determine the most appropriate model structure for each site grouping.

Chapter 6 addresses the third objective of the thesis. A short timescale event based rainfall model, DRIP (Disaggregated Rectangular Intensity Pulse), is conditioned on the state series of the calibrations produced in Chapter 5. The objective is to demonstrate that the conditioning of models like DRIP on climate states persisting over longer time scales can improve their ability to simulate variability of rainfall at time scales of a year

or longer. This method is applicable not only to DRIP, but to any small timescale event based model.

A summary of the conclusions is presented in the final chapter, Chapter 7, following which future directions of research are discussed.

Chapter 2 Modelling Long-term Persistence in Rainfall

2.1 Introduction

Modelling the long-term interaction of climate and rainfall (at short and long time scales) is a relatively new endeavour. Up until the last decade, apart from seasonal variations, short-timescale rainfall was generally modelled as a stationary process. Climate was not modelled as affecting rainfall from year to year explicitly. For the annual rainfall process also, simple models attempting to reproduce key statistics such as auto-correlation were used, whereas the underlying physical processes governing the variability were ignored. However, with increasing computational power, more complex interactions have been investigated, with a greater physical basis. Generally, these models have a hierarchical structure, with longer-term climate variations influencing shorter-term rainfall amounts in some way.

From a modeller's point of view the problem is to identify the largest sources of variability, and try and incorporate them into his/her prospective model. This must be achieved in a way that marries the spatial and temporal variability in a coherent manner.

This chapter introduces the stochastic modelling of the interaction of climate and Australian rainfall. Empirical evidence of inter-annual persistence in hydroclimatic processes is presented first. The remainder of the chapter provides a critical review of the previous attempts at climate-rainfall modelling, and draws some links between the most successful of these. Models calibrated to daily rainfall are surveyed first, followed by models calibrated to annual data.

One particular annual rainfall model, the two-state hidden Markov model (HMM) used by *Thyer* (2001), is discussed in detail. Although this model has close links to most other climate-rainfall models, this model does differ in the respect that it attempts to model long-term persistence in climate rather than the rainfall itself. As demonstrated by *Katz and Zheng* (1999), this annual HMM also has applicability to smaller timescale models. However, there are potentially serious parameter identifiability issues when using single site data. The use of multiple sites can help to overcome these identification

problems. However, multi-site approaches also have drawbacks regarding the assumptions made to substitute space for time. Some extensions to the multi-site HMM are proposed to address these drawbacks. These extensions define the major focus of the remainder of the thesis.

2.2 Long- and Short-term Climate Variability

Climate is defined in the *Oxford English Dictionary* (2000) as the 'Condition (of a region or country) in relation to prevailing atmospheric phenomena, as temperature, dryness or humidity, wind, clearness or dullness of sky, etcetera, especially as these affect human, animal, or vegetable life.'. It is in this sense, namely the integration of all characterising atmospheric phenomena, with which the term climate will be used. Regarding the timescales over which climate variability operates, short-term climate variability, as used in this thesis, will refer to timescales of the order of days to months. Long-term refers to years, decades and possibly longer timescales.

Persistence is defined here as the occurrence of runs of high or low rainfall years relative to the long-term mean that are longer than what would be expected from an independent annual process. This differs to the mathematical definition of persistence provided by *Beran* (1994 p.6-7) related to the sum of correlations over all lags being infinite. Strictly speaking, the models presented within this thesis are not long-range dependent.

Many hydrological studies have focussed on reproducing a statistic related to the cumulative departures from the mean, the rescaled adjusted range (*Hurst*, 1951). This statistic, often used in the definition of long-term persistence, is not used here as it can require very long records for accurate estimation. As we are dealing with records of length around only 100 years, it is unlikely that the Hurst coefficient (see *Bras and Rodriguez-Iturbe*, 1985 p.211) could be accurately estimated.

It is an objective of this thesis to provide a stochastic mechanism for long-term climate persistence within a rainfall model. Before this objective is addressed, some evidence for this long-term hydroclimatic persistence is presented.

2.2.1 Evidence of long-term climate variability and persistence within hydro-climatological time series

'Empirical evidence of many hydro-climatological data shows temporal variability involving trends, oscillatory behaviour, and sudden shifts' (*Sveinsson et al.*, 2003 p.489). In particular, the quasi-cyclic ENSO phenomenon has been well documented as having a strong influence on Australian rainfall and runoff (eg. *McBride and Nicholls*, 1983, *Chiew et al.*, 1998, *Chiew and McMahon*, 2003). The impact of ENSO is not consistent and varies with time (*Nicholls et al.*, 1996, *Kane*, 1997).

Within sea surface temperature indices - the Interdecadal Pacific Oscillation (IPO) and the Pacific decadal Oscillation (PDO) - sudden shifts occur decades apart (*Mantua et al.*, 1997, *Power et al.*, 1999). Indeed, there is some evidence that ENSO frequency and strength are modulated by these indices (*Power et al.*, 1999), with consequences for flood and drought risk (*Kiem et al.*, 2003). Within longer length series, such as ice core measurements in Greenland (*Taylor*, 1999), and coral core records related to runoff from the Burdekin river in Queensland (*Isdale et al.*, 1998), there is evidence of long-term periods of persistence in different climatological modes.

2.2.2 Evidence of persistence within annual rainfall series in Eastern Australia

Specifically for Eastern Australia, the case study area for this thesis, inter-annual persistence has been identified within rainfall records (*Pittock*, 1975, *Cornish*, 1977, *Lough*, 1993). All of these studies identify below average rainfall periods from the early 1900's until the late-1940s, followed by a period of above average rainfall. These studies typically focus on two distinct periods, 1900-1945 and 1946-1980 (depending on the sites used and when the study was produced). The study of *Lough* (1997) identifies some evidence within Queensland (north Eastern Australia) for switching back to a below average rainfall during the period 1975-1995, possibly linked to the increased incidence of El-Niño events during this time period (*Trenberth and Hoar*, 1996). However, there is further empirical evidence of persistence within and outside the two dominant periods.

A plot showing the annual rainfall cumulative departures from the mean for 13 sites on the Eastern Coast of Australia is presented in Figure 2.1. These annual rainfall series are described and used in the case studies presented in Chapter 5. Such a plot is typically

used to empirically diagnose persistence within time series (e.g. *Pittock*, 1975). Consistent periods of negative slope indicate that the rainfall is persistently lower than average, and *vice versa* for positive slope. Within some of these series (e.g. Mount Victoria/Blackheath, Brisbane, Moruya Heads), there appears to be evidence of distinctly wet and dry periods. Within other series (e.g. Cape Moreton, Miles), this *prima facie* evidence is not as persuasive.

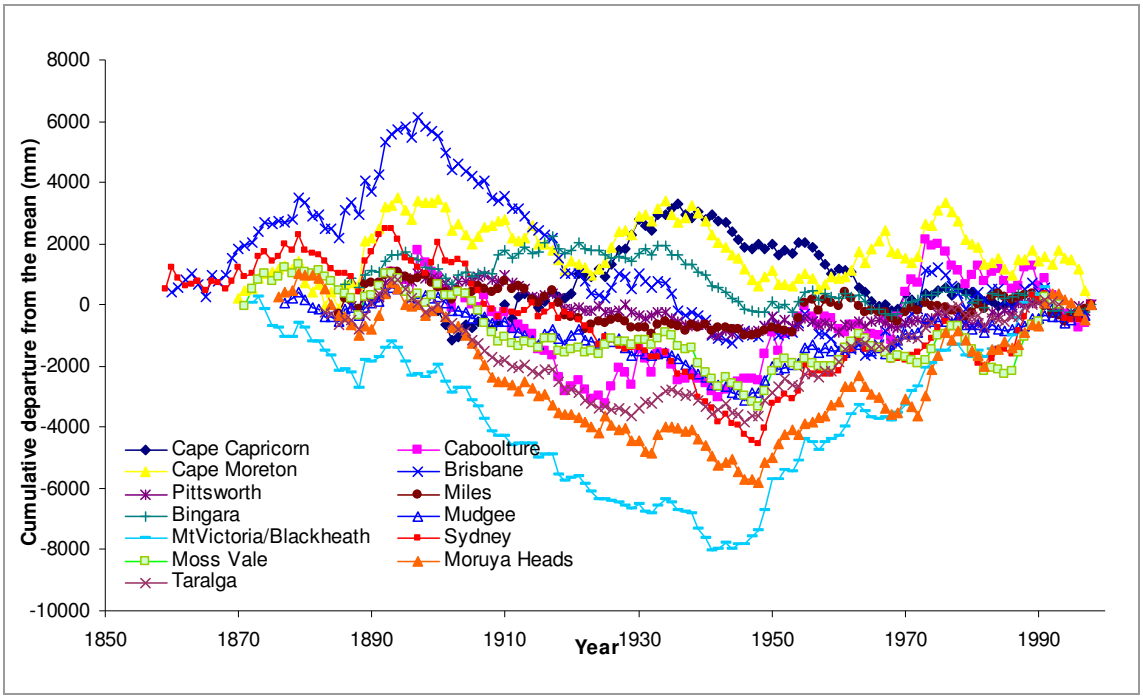


Figure 2.1 Annual rainfall cumulative departures from the mean (May-April water year)

The cumulative departure from mean series show some similarity in the timing when the changes in slope occur, with the slope changing from positive to negative for the majority of series around 1900, and from negative to positive around 1945. This provides some evidence for the hypothesis of a regional controlling climate, with changes in climate affecting rainfall in a region in a similar way.

Overall, from the cumulative departure from the mean plots there appears to be some evidence of annual rainfall being modulated by a controlling climate structure. However, each site is influenced to differing degrees. These observations provide additional evidence of long-term persistence within Eastern Australian rainfall. However, the purpose of this thesis is not to identify evidence of such persistence within Australian rainfall, ample studies have already rendered evidence demonstrating

persistence. The aim of this work is to investigate stochastic mechanisms that can be used to simulate such observed phenomenon.

2.2.3 Implications of persistence

Inter-annual persistence has implications for drought security, water resource management and agriculture. *Franks and Kuczera* (2002) demonstrate that the stationary climate assumption implicit in the estimation of design floods may cause very significant under-/over-estimation of flood magnitude, which in turn means that the associated project could be exposed to a greater risk/cost than desired. The assessment of drought impact is also affected by the assumption of a stationary climate. The mechanisms by which long-term climate variability physically influences rainfall amounts are becoming better understood, with links between climate indices, atmospheric circulation and rainfall being developed (e.g. *Folland et al.*, 2002). However, their influence on the spatial distribution of rainfall is not yet well understood (see *Thyer*, 2001 for a discussion). In the interim period, a stochastic model linking climate and rainfall is required.

Typically at the annual timescale, hydro-climatological processes have not incorporated mechanisms for simulating long-term persistence, despite the growing empirical evidence indicating persistence. However, there have been several attempts at capturing this inter-annual persistence, the focus of the remainder of this chapter.

2.3 Terminology and Modelling Framework

Before the different methods of modelling climate influence on rainfall are described, some terminology, notation and a general modelling framework are introduced under which all models can be discussed and compared.

2.3.1 Data, random processes, probability and stochasticity

We are interested in modelling a spatio-temporal data field denoted $\mathbf{Y}_1^T = (\mathbf{y}_1, ..., \mathbf{y}_T)$, where $\mathbf{y}_t = (y_t^1, ..., y_t^{d_t})$ is the set of observations taken at time t over d_t sites. The T superscript within the data matrix $\mathbf{Y}_1^T$ denotes the total number of time periods, while the subscript denotes the initial timestep that is being considered. This double scripting on $\mathbf{Y}$ is used in model formulation (e.g. Section 3.4.1), where subsections of the overall

data matrix $\mathbf{Y}_1^T$ are required. Most generally, the number of sites d_t can vary from timestep to timestep, however, for the remainder of this thesis the number sites will be equal over all timesteps, denoted d.

If realisations of this data through time are considered random, rather than being purely deterministic, they are modelled as a stochastic process (*Box and Jenkins*, 1976 p.7). Given that the data are considered a random process, the data has an associated probability density function $p(\mathbf{Y}_1^T \mid \boldsymbol{\theta}_M, M)$ given a model hypothesis M and an associated set of unobserved parameters $\boldsymbol{\theta}_M$ (*Bras and Rodriguez-Iturbe*, 1985 p.2). The subscript on the parameter vector is used to emphasise that the parameter vector is dependent on the model M. This subscript is omitted for simplicity in the following text, the exception being where multiple models are being discussed.

2.3.2 Probability Notation

As we are describing and dealing with stochastic processes throughout this thesis, notation is required to denote probability quantities. The notation $p(a \mid b)$ refers to the conditional probability distribution of a given b, while $p(a)$ denotes a marginal distribution of a. The joint probability of observing a, b and c can be calculated according to conditional probability in terms of the factorisation $p(a,b,c) = p(a \mid b,c)\, p(b \mid c)\, p(c)$. As was done in *Gelman et al.* (1995 p.7), the same notation is used for continuous density functions and discrete probability mass functions. In some cases, we use $P(\cdot)$ to denote the probability of an event (for integrated discrete or continuous density functions), as opposed to a continuous density.

2.3.3 Stationary and Non-stationary processes

'A hydrologic time series is stationary if it is free of trends, shifts, or periodicity (cyclicity). This implies that the statistical parameters of the series, such as mean and variance, remain constant in time. Otherwise the series is nonstationary.' (*Salas*, 1993 p.19.5). Strict theoretical definitions of stationarity exist (eg. *Bras and Rodriguez-Iturbe*, 1985 p.5). However, it is in this less formal sense that the terms stationary and non-stationary will be used in this thesis. Models of daily rainfall would typically be non-stationary over the year due to seasonality. The same rainfall model aggregated to

the annual timescale could be stationary in that the annual values vary around a stationary or time invariant mean. Likewise models showing inter-annual shifts (non-stationarity) in mean may be considered stationary over longer periods. Thus most systems can be considered stationary or non-stationary depending on the period over which the process is observed.

2.3.4 Persistence, Overdispersion and Downscaling

It is an objective of this thesis to incorporate long-term persistence into a short-timescale rainfall model DRIP. It is hypothesised that the current underestimation of annual variance within models like DRIP, is due to persistence at longer-term timescales not being considered. This underestimation of annual variance, termed overdispersion (*Katz and Zheng*, 1999) can be addressed by downscaling.

Downscaling refers to the process of overlying a model of larger-spatial scale (that may incorporate long-term persistence) over a local short-timescale model (*Bellone*, 2000). In this thesis downscaling will be implemented by conditioning the DRIP point rainfall model on the output series of the regional two-state HMM. The two-state HMM conditions local rainfall characteristics on a conceptual regional climate influence. Such conditioning is typical of a broad range of models used in the environmental sciences: hierarchical models.

2.3.5 Hierarchical modelling framework

Of the spatio-temporal models used in hydrology and environmental science, many fall within the following broad definition. The model is broken into three stages:

Stage 1.	Data Model:	$p(data \mid process, parameters)$
Stage 2.	Process Model:	$p(process \mid parameters)$
Stage 3.	Parameter Model:	$p(parameters)$

This breakdown has been used by various authors (*Berliner*, 1996, *Wikle*, 2003) in describing hierarchical models. Indeed, breaking a model into a series of conditional models, coherently linked to one another through conditional probability, defines hierarchical modelling (*Wikle*, 2003). The first stage describes the observational process

given the process of interest, and parameters that describe the data model. The second stage describes the process, conditional on other parameters. The final stage accounts for variability in the process model parameters.

The data model generally is used for representing measurement error corrupting the observed process. Of course this data model is problem dependent. In this study the measurement error is judged to be small (as the majority of data was obtained from previous studies indicating its high quality – see Section 5.2) compared with process variability and will thus be disregarded. In the resulting analysis, this choice is challenged – as discussed in Section 5.6. The process model is usually the most critical step in constructing a hierarchical model, with multiple conditional sub-stages usually being employed to describe process dependencies. The process in the context of this study is the interaction of climate and rainfall. The parameter (uncertainty) model will be discussed in the following chapter. For a fuller description of hierarchical models and a general framework for spatio-temporal process models, see *Wikle* (2003).

2.3.6 Process model: Temporal dependency

The ability of rainfall models to reproduce long-term persistence of rainfall will depend on the assumptions made within the model regarding the conditioning of each year's rainfall on previous years. Sometimes it is assumed that each year's rainfall is independent of others yielding:

$$\begin{aligned} p(\mathit{process} \mid \mathit{parameters}) &= p\left(\mathbf{Y}_1^T \mid \boldsymbol{\theta}, M\right) \\ &= \prod_{t=1}^{T} p\left(\mathbf{y}_t \mid \boldsymbol{\theta}, M\right) \end{aligned} \qquad (2.1).$$

Here $\boldsymbol{\theta}$ represents the set of unobserved parameters associated with model M, and $\mathbf{y}_t$ is the rainfall observed in year t. In such a model, there is no mechanism to produce long-term persistence. On other occasions, rainfall is related to that which has occurred in the past. A common method is to condition this year's rainfall on the year immediately preceding it:

$$\begin{aligned} p(\mathit{process} \mid \mathit{parameters}) &= p\left(\mathbf{Y}_1^T \mid \boldsymbol{\theta}, M\right) \\ &= p\left(\mathbf{y}_1 \mid \boldsymbol{\theta}, M\right) \prod_{t=2}^{T} p\left(\mathbf{y}_t \mid \mathbf{y}_{t-1,} \boldsymbol{\theta}, M\right) \end{aligned} \qquad (2.2).$$

The lag one auto-regressive model (*Srikanthan and McMahon*, 1985) is an example of such a model. Such models have been used quite widely as this allows the natural assumption that the immediate past state of an environmental process affects the current state. A generalisation of this model is to introduce a latent variable (or state) r_t, with the rainfall being considered to be a degraded observation of the latent process. This generic family is more widely termed state-space models. Dynamic models (of which the Kalman filter is a member – see *Kalman*, 1960, *Wikle and Cressie*, 1999) is an example of such a model. Hidden Markov models are used to describe such processes where r_t is a discrete random variable, with the individual terms of (2.2) being:

$$\begin{aligned} & p\left(\mathbf{y}_t \mid \mathbf{y}_{t-1,}, \boldsymbol{\theta}, M\right) \\ &= \sum_{r_{t-1} \in R_{t-1}} \left(\sum_{r_t \in R_t} p\left(\mathbf{y}_t, r_t \mid \mathbf{y}_{t-1}, r_{t-1}, \boldsymbol{\theta}, M\right) \right) p\left(r_{t-1} \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}, M\right) \\ &= \sum_{r_{t-1} \in R_{t-1}} \left(\sum_{r_t \in R_t} p\left(\mathbf{y}_t \mid r_t, \mathbf{y}_{t-1}, r_{t-1}, \boldsymbol{\theta}, M\right) p\left(r_t \mid \mathbf{y}_{t-1}, r_{t-1}, \boldsymbol{\theta}, M\right) \right) p\left(r_{t-1} \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}, M\right) \\ &= \sum_{r_{t-1} \in R_{t-1}} \left(\sum_{r_t \in R_t} p\left(\mathbf{y}_t \mid r_t, \boldsymbol{\theta}, M\right) \quad p\left(r_t \mid r_{t-1}, \boldsymbol{\theta}, M\right) \right) \quad p\left(r_{t-1} \mid \mathbf{y}_{t-1}, \boldsymbol{\theta}, M\right) \end{aligned} \qquad (2.3).$$

R_t signifies the possible states r_t can take. This calculation will be discussed in greater detail in the Chapter 3. However, the final line shows the significance of HMM modelling, with the rainfall $\mathbf{y}_t$ only being dependent on the current state r_t. The dependence through time is defined by the $p\left(r_t \mid r_{t-1}, \boldsymbol{\theta}, M\right)$ term.

These models may be generalised further to be dependent on greater lags of rainfall. However, this natural dependency on the previous timestep (lag one) is often assumed in modelling such environmental processes (eg., *Srikanthan et al.*, 2002, *Sveinsson et al.*, 2003, *Wikle*, 2003). The question that arises upon applying these models is whether they are able to reproduce the long-term persistence that is apparent within hydrological data. We use this process model framework to interpret previous attempts at incorporating the influence of climate on rainfall.

2.4 Stochastic Rainfall Models: The challenge of linking short and long timescales

The following section overviews some previous efforts at downscaling the influence of climate on rainfall. As comprehensive reviews of stochastic rainfall (and weather) generation techniques have been given before (see *Wilks and Wilby*, 1999, *Srikanthan and McMahon*, 2001), an exhaustive study is not undertaken. Rather this section focuses on the previously described methods for simulating variability induced by climate on rainfall, at both daily and annual timescales. As noted in both of the previously mentioned reviews, 'conventional weather generation techniques often fail to capture wholly inter-annual variability'. Previous methods are surveyed to assess their applicability to addressing this inter-annual variability.

2.4.1 Daily rainfall models

Due to the wide availability of weather data at the daily timescale, and the abundance of impact models driven by daily rainfall input, daily rainfall models are by far the most common model type (*Wilks and Wilby*, 1999). These stochastic models of precipitation have until recently considered precipitation in isolation from the atmospheric processes that drive it. However, methods introduced in the last decade (eg. *Hay et al.*, 1991, *Bardossy and Plate*, 1992, *Katz and Parlange*, 1993, *Hughes and Guttorp*, 1994) have attempted to address this by conditioning on, or correlating to, synoptic atmospheric patterns or indices.

The model formulation is typically Markovian (or state-space), with an atmospheric 'weather state' r_t at time t, having a corresponding rainfall distribution coupled with it. Within these earlier studies the atmospheric state r_t was generally considered a known quantity, given from previous analysis. Such downscaling models require the definition of weather states, which are associated somehow with synoptic atmospheric indices. A problematic issue in the past has been the division of rainfall data into states according to the atmospheric data (*Bellone*, 2000 p.11), with subjective and *ad hoc* measures usually employed to determine the order and occurrence of states.

Hughes et al. (1994) introduced the non-homogeneous Markov model (NHMM) for relating precipitation occurrence at multiple rain-gauge stations to broad scale

atmospheric circulation patterns or states. Rather than requiring *a priori* division of the rainfall into states according to the atmospheric data, this approach enabled the modulation of the Markov model parameters by the atmospheric variables. The order of the model was chosen (via model selection) as a product of model calibration. *Charles et al.* (1999) and *Bellone et al.* (2000) extend the precipitation occurrence NHMM to include precipitation amounts.

Although such downscaling models are potentially quite useful in water resource studies, a drawback of NHMM's is that there is presently no suitable method for simulating long-term atmospheric data (*Thyer*, 2001). The most common method of atmospheric simulation employs General Circulation model's (GCM). Unfortunately, these methods are currently computationally prohibitive. As the NHMM requires input of atmospheric data, this approach is unsuitable for long-term water resource applications. Another drawback is that the level of uncertainty associated with both the calibration (parameter uncertainty) and the GCM itself (model uncertainty) are currently not accounted for within predictions, estimations and other expectations output from the GCM.

Order selection is a fundamental task when dealing with HMM's. Order selection refers here to the choice of number of possible hidden states. As the process is unobservable, it is often unknown *a priori* how many states should be used. Model selection techniques such as the Bayesian Information Criterion (BIC - see Chapter 3) are used to discern between models (*Katz*, 1981). Fewer states are less complex and require less computation time. However depending on the modelling timescale used, and the number of sites used, differing numbers of states can be justified (see *Bellone*, 2000 chapter 3 for a simulated case study demonstrating these effects).

Apart from climate downscaling models that require the input of atmospheric measurements as forcing variables, another very similar modelling technique has been used to incorporate the inherent variability of rainfall over time. These models have essentially the same structure as the downscaling models discussed above, except that there are no atmospheric variables used, only rainfall. This approach reduces the reliance on being able to simulate long-term series of atmospheric variables.

An example of this approach is that of *Katz and Zheng* (1999). They note that stochastic models fitted to time series of daily precipitation commonly do not include a component that explicitly accounts for inter-annual variation. In a method similar to that of *Thyer* (2001), they attempt to address this by overlaying an annual two-state HMM on a daily rainfall model. It is shown through model selection that using an annual two-state HMM is superior to a single state process. They postulate that this is due to the effect of low frequency oscillations on rainfall. However, determining whether the variability is due to low or high frequency oscillations is left as an open question.

Another possible solution modelling this low/high frequency variability is to use Dynamic Linear Models (DLM). These are a continuous state space version of the finite state HMM, and have been used by *Zheng* (1996) and others to model daily temperatures. But as *Katz and Zheng* (1999) note, due to the intermittency of rainfall (at the daily timescale) these models cannot be directly applied. In other words, dealing with persistent dry periods is difficult. *Sanso and Guenni* (2000) sidestep this issue by using 10 day accumulations of Venezuelan rainfall, and apply a DLM. Although the model clearly has a mechanism with which to incorporate non-stationarity and persistence in rainfall (as the underlying dynamic state parameters can persist from year to year around relatively dry or wet modes) the ability of the model to capture inter-annual variability was not examined in that study.

2.4.2 Annual hydro-climatological models

Until recently, the design of annual rainfall models has generally ignored the physical processes causing year-to-year variation, with models being focussed on reproducing statistical characteristics of the data. Several variants of auto-regressive moving average (ARMA) models developed by *Box and Jenkins* (1976) fall into this category. The prevalence (*Salas*, 1993) of these models is a testimony to their robust yet simple nature. One variant, the lag one autoregressive (AR1) model has found wide use for annual rainfall generation throughout Australia (*Srikanthan and McMahon*, 2001).

Srikanthan et al. (2002) conducted a review of annual rainfall models assessing the AR1 model (*Srikanthan and McMahon*, 1985) against the newly introduced HMM over 44 key sites spread across Australia. They found that the AR1 model could not be relied upon to satisfactorily reproduce 2 and 3-year low rainfall sums if parameter uncertainty

was not taken into consideration. Hence, assessment of drought security using the AR1 model may be compromised.

Autoregressive processes are quite abstract in nature, in that it is difficult to conceptualise the process by which autocorrelation occurs. How is this year's rainfall affected by a percentage of last year's rainfall? Other stochastic models have conceptualised the hydrologic process in more physically meaningful ways. Generally, these models conceptualise climate changing in some way, and consequently, influencing the rainfall amounts.

One such example is the change-point model (*Perreault et al.*, 2000a). These models conceptualise the rainfall mean (and/or variance) of the process varying when a change-point has been reached. The locations of the change-points are also calibrated parameters. Closely related are the shifting mean models employed by *Sveinsson et al.* (2003) and *Fortin et al.* (2002, 2003). Rather than an individual change-point being modelled, the mean is allowed to shift multiple times by differing amounts. The frequency of mean shifting is also calibrated. These shifting mean models can also be seen to be a generalisation of the DLM's mentioned in the previous section. In that case, the mean shifted at every time-step. In this case the time between mean shifts is sampled from a geometric distribution.

Another model closely related to the shifting mean model (see *Fortin et al.*, 2002) is the two-state HMM introduced by *Thyer* (2001). This model conceptualises the climate as being in two states, relatively wet, or relatively dry, with different rainfall distributions according to state. Persistence in each state is modelled according to Markovian transitions, with each year's state only being dependent on the previous year. This model has shown promise in identifying regional persistence in climate throughout Australia (*Srikanthan et al.*, 2002). However, as detailed later in this chapter, the HMM framework requires modification if it is to be used routinely.

2.4.3 Point Process models

There is an extensive literature related to point process models (eg. *Waymire et al.*, 1984, *Cowpertwait and O'Connell*, 1997). These are stochastic models with a structure chosen to simulate the observed rainfall process much more closely than the already

described models. These models, like most other rainfall models, typically have not accounted for non-stationarity and persistence other than that induced by seasonal variations. As there have not been any attempts (to the author's knowledge) to incorporate this non-stationarity, these models are not reviewed.

However, one model of this genre is singled out. The Disaggregated Rectangular Intensity Pulse (DRIP) model of *Heneker et al.* (2001) is an event based point process model calibrated to pluviograph data. As with other models of this genre, DRIP did not incorporate inter-annual non-stationarity. It is an objective of this thesis to develop a method for incorporating long-term persistence induced by climate into such a model. The DRIP model will be discussed in greater detail where it is applied in Chapter 6.

2.4.4 Choice of time dependency and inter-annual persistence

The previous section has reviewed various efforts at incorporating the effects of a non-stationary climate on rainfall. Of the daily rainfall modelling approaches, weather state models and dynamic linear models allow the influence of day-to-day dependency to be included with links to atmospheric patterns. Annual models (necessarily) focus on dependency at greater timescales, with HMM, change-point and shifting mean models being used to match empirical characteristics of a changing climate at annual timescales.

Apart from the work of *Katz and Zheng* (1999), there has been little work addressing climate effects of inter-annual persistence on smaller timescale models. This thesis attempts to blend the annual and daily approaches - using a current annual weather state model, the HMM, and overlaying it on a smaller timescale rainfall model DRIP, thereby downscaling the inter-annual persistence modelled by the HMM, into the smaller event based rainfall model. Although conceptually similar to the work of *Katz and Zheng* (1999), this independently developed approach differs in that the state series is derived from annual rainfall records and hence is independent of the short timescale data. The resulting state series of the HMM calibration is used to condition the DRIP model. It is noted that the state series conditioning method used for DRIP represents a general framework which can be used for downscaling to any small timescale model for which a likelihood function can be evaluated.

The separate calibration of the HMM allows regional (multi) site state identification to be addressed. This multi-site approach allows identification of inter-annual persistence, whilst also accounting for spatial variation. Specifically, the generalisations of the HMM model allow different state series to be identified by different sites, thus attempting to match the empirically observed phenomenon of different regions being affected to differing degrees by changes in climate regime.

2.5 Two-State Hidden Markov Model

The HMM (as mentioned within section 2.4.2) has successfully been used to model annual rainfall at a range of sites throughout Australia (*Srikanthan et al.*, 2002) – including 2 and 3-year low rainfall sums. Therefore, the two-state HMM as applied by *Thyer* (2001) is the starting point of this study. *Thyer's* (2001) implementation is based on the algorithm presented by *Chib* (1996) for a HMM, while *Bengio* (1999) provides a general review of various HMM's. The HMM is described here in detail following *Thyer* (2001). Some limitations of this multi-site HMM (in this context) are presented in the following section (2.6), whilst the HMM calibration algorithm (as applied in this thesis) is detailed in section 3.4.

2.5.1 Parametric framework of the HMM

The two-state HMM framework, illustrated in Figure 2.2, assumes the climate is in one of two states: wet or dry. Each state has an independent annual rainfall distribution, assumed to be Gaussian. The persistence in each state varies according to the state transition probabilities. For example, the expected residence time in the dry state is 1/ P(Dry→Wet). This provides an explicit mechanism to simulate the variable length wet and dry rhythms observed in Australian rainfall data.

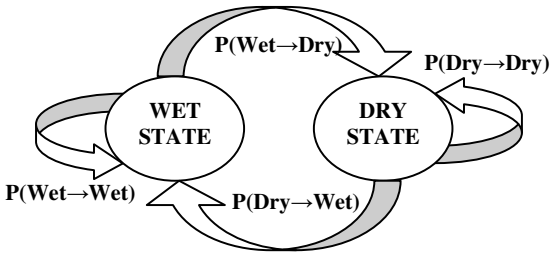


Figure 2.2 Two-state HMM Conceptual Diagram

In more formal terms, the regional climate state at year *t*, r_t, is modelled by a Markovian process:

$$r_t \mid r_{t-1} \sim Markov(\mathbf{P}) \qquad (2.4),$$

where $\mathbf{P}$ is the state transition probability matrix defined by:

$$\mathbf{P} = \left[p_{ij} \right] = p\left(r_t = j \mid r_{t-1} = i\right) \;\; i, j = W, D \qquad (2.5).$$

As we cannot observe which state a particular site is in any year it is necessary to infer the climate state time series, $R_1^T = (r_1, r_2, ..., r_T)$, using the HMM. This series is included as a latent parameter requiring estimation.

It is assumed that these transition probabilities are stationary over time. Depending on the climate state simulated by the Markovian process at time *t*, different Gaussian rainfall distributions are used to simulate a vector, $\mathbf{y_t}$, of rainfall amounts for *d* multiple sites according to:

$$\mathbf{y_t} = \begin{pmatrix} y_t^1 \\ \dots \\ y_t^d \end{pmatrix} \sim \begin{cases} N_d\left(\boldsymbol{\mu}_\mathbf{W}, \boldsymbol{\Sigma}_\mathbf{W}\right) & \text{if } r_t = W \\ N_d\left(\boldsymbol{\mu}_\mathbf{D}, \boldsymbol{\Sigma}_\mathbf{D}\right) & \text{if } r_t = D \end{cases} \qquad (2.6),$$

where $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a multivariate Gaussian distribution in d dimensions with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. It is noted here, that a slightly different parametric framework to that of *Thyer* (2001) is used in this study regarding the covariance matrices $\boldsymbol{\Sigma}$. Gaussian correlation coefficients $\boldsymbol{\rho} = \left[\rho_{ij} \right] : i, j = 1, ..., d$ that are independent of state were fitted with covariance $\boldsymbol{\Sigma}_{r_t} = \left[\rho_{ij} \sigma_{r_t}^i \sigma_{r_t}^j \right]$, $i, j = 1, ..., d$.

Thus, the vector of unknown parameters for the multi-site HMM $\boldsymbol{\theta}$ is composed of the state mean and variance parameters, the correlation coefficient parameters, the state transition probabilities, and the hidden state time series, giving:

$$\boldsymbol{\theta} = \left(\boldsymbol{\mu}_\mathbf{W},\ \boldsymbol{\sigma}_\mathbf{W},\ \boldsymbol{\mu}_\mathbf{D},\ \boldsymbol{\sigma}_\mathbf{D},\ \boldsymbol{\rho}, \mathbf{P},\ R_1^T\right) \qquad (2.7).$$

The modelling assumptions for the multi-site framework are:

1. The distribution of hydrological data is composed of two independent distributions: a wet state and a dry state distribution.

2. Both the wet and dry state distributions are multivariate Gaussian.

3. The climate state at time t, r_t, is purely dependent on the climate state from the previous time step, r_{t-1}.

4. The probability of state transition is assumed to be stationary over time.

5. Every site is assumed to be in the same climate state at every point in time, i.e. the climate state is assumed to be "regional".

Some empirical evidence for the assumption of multiple states has been presented in section 2.2.2, and many studies modeling rainfall have assumed an underlying Gaussian distribution (*Srikanthan et al.*, 2002). The dependence on the previous timestep (through the climate state) is a natural assumption for environmental processes (*Wikle*, 2003) – and is also supported by the apparent persistence identified in section 2.2.2. The assumption of stationarity of state transition probabilities is difficult to verify given the limited amount of data available. However, this model is simpler than a non-homogeneous HMM which relaxes this assumption. The regional state structure is also supported by section 2.2.2, with several sites showing simultaneous apparent changes in rainfall mean.

2.5.2 Hidden State Probability Series

From the calibration of the HMM comes the hidden state probability time series. This series gives the posterior probability (as it is unknown) of any rainfall year used in the calibration being in either a dry (or wet) state. This probability time series is used to condition smaller timescale models, whereas the transition probabilities are used during simulation. An example of a hidden state probability series for Sydney (see Section 2.6 for definition and discussion of sites) is shown in Figure 2.3. A consistently dry period for the first half of the century can be identified by the low probability of the hidden state being wet, whereas wet periods become more common after 1950.

2.6 Limitations of the Multi-site HMM

Srikanthan and McMahon (2001) conducted a review of annual rainfall models, comparing the widely used AR1 model and the two-state HMM. It was found that the two-state HMM assumptions were justified for a range of sites around Australia. However, clear identification of regions where the HMM assumptions held was not possible. The AR1 model (with parameter uncertainty included) was recommended for use as the HMM did not produce a demonstrably better fit to drought statistics such as five year rainfall sums at the sites where the HMM assumptions were justified. In that study, only single site data was used, with the majority of sites having just over 100 years of record. At many sites, the climate states were insufficiently identifiable given this short amount of data.

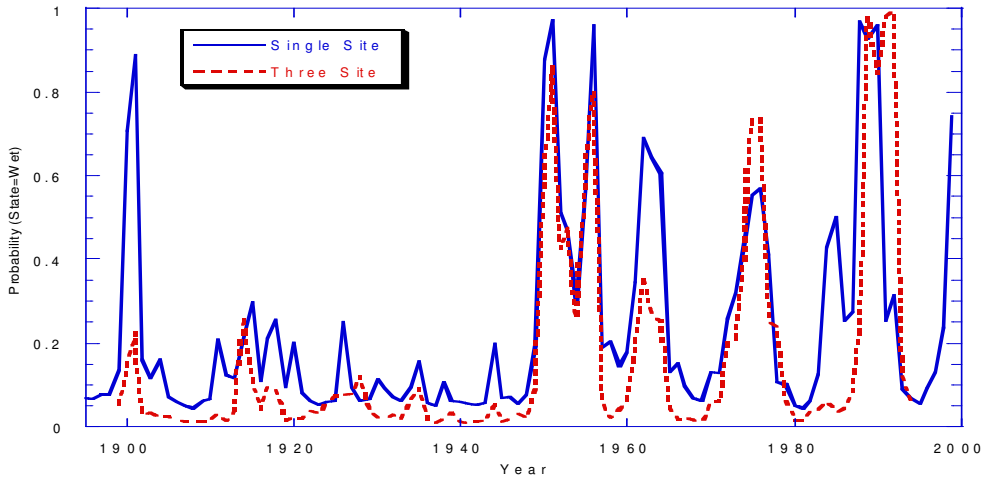


Figure 2.3 Hidden state probability series : Single Site versus Multi Site HMM for Sydney

Extensions of the HMM described in *Thyer and Kuczera* (2003a, 2003b) allowed the use of multiple site annual rainfall data. Assuming that the persistence occurs on a regional scale, rainfall data from multiple sites can be used to strengthen knowledge about actual persistence patterns. However, there is the danger that the inclusion of data from a new site may bias the transition probabilities. This could arise if the climate controls on the new site were different to the controls on the other sites – in other words the new site does not belong to the same persistence region. An example illustrating the effect of including more annual rainfall sites on a state probability series for the Sydney region is shown in Figure 2.3. Using visual inspection the three-site HMM series does not differ from the single site series markedly, possibly indicating that the annual

climate influences are similar over the entire region. The three-site series shows less uncertainty about which state the climate is in, with fewer values being around 0.5 (1973-76 and 1983-84). This may indicate that the climate state has been more clearly identified with the introduction of more sites.

Of course the multi-site HMM is an idealisation of climate-rainfall interaction. In reality, the conceptual regional climate state is actually a continuum of climate influences. It is likely that some sites may experience different climate influences from sites nearby in some years, yet follow the trends of other sites in the majority of years. Inclusion of such a site into a multi-site analysis could bias the estimate of states as the regional state assumption is not true. On the other hand, not including the site in the analysis could result in the state series being insufficiently identified. More generally, the further sites are apart, the less likely they are under the same climate controls.

Methods other than visual inspection are available to differ between the single and three-site results, but there in lies an essential question: how to choose sites to be included in the multi-site HMM analysis arises. That is, which method of grouping sites is appropriate? This problem has also been identified in using change-point models for modelling annual streamflow at multiple sites (*Perreault et al.*, 2000b).

2.6.1 Previous methods for choosing sites in an analysis

Thyer (2001) and *Thyer and Kuczera* (2003a, 2003b) presented a methodology for calibration of the multi-site HMM. However, the development of a rigorous rationale for choosing sites to be included in a multi-site analysis was beyond the scope of their study. Rather those studies focussed on synthetic data calibrations to determine if parameters were identifiable given multi-site data generated from a HMM. Using the Warragamba catchment rainfall data, the model assumptions were checked given all the (five) sites had been used in calibration. This check revolved around an index of separation of the wet and dry means at each site, the wet and dry separation index ($WADSI$) defined as:

$$WADSI = \frac{\mu_W - \mu_D}{\sqrt{\sigma_W^2 + \sigma_D^2}} \qquad (2.8).$$

This measure has links to the probability of a random rainfall variate generated from the wet state y_W being less than that from the dry state y_D, denoted $P(y_W < y_D)$. It is emphasised that the measure refers to marginal wet and dry random variates (hence lack of temporal super/subscripts) given the relevant calibrated state parameters for a single site. $P(y_W < y_D)$ can be rewritten $P(y_W - y_D < 0)$. Given that any linear combination of two Gaussian distributions is also Gaussian, the probability $P(y_W - y_D < 0)$ can be written as:

$$P(y_W - y_D < 0) = \Phi\left(\frac{0-(\mu_W - \mu_D)}{\sqrt{\sigma_W^2 + \sigma_D^2}}\right) = 1 - \Phi\left(\frac{\mu_W - \mu_D}{\sqrt{\sigma_W^2 + \sigma_D^2}}\right) \quad (2.9),$$

where $\Phi(\cdot)$ is the cumulative probability function for the standard Gaussian distribution.

If all sites show a low probability that $P(y_W < y_D)$ over all posterior parameter values, this is a good indication that assumption one of the HMM, is true. In that study, one site (Moss Vale) had the WADSI mode at zero, indicating poor separation of wet and dry parameters.

Another diagnostic was used to check whether using multiple sites in calibration aided parameter identification compared to single site calibration. This check involved viewing the posterior probability plot (parameter uncertainty – see Section 3.2 Bayesian Modelling and Calibration with Application to the HMM) of the single site calibrations versus the multiple site. If the uncertainty decreased with more sites, this was interpreted as an improvement in the identification of the two-state persistence structure. Using the Moss Vale site in the calibration did reduce the uncertainty in the transition probabilities, hence producing a more clearly identified state series (this would be expected if Moss Vale contains information that aids identification of the state series). However, should this site have been included in the analysis given that it may have been violating a model assumption according to the WADSI? Alternatively, is the state series overidentified, with the information on state provided by Moss Vale being superfluous?

Can the fifth assumption of the HMM model, that every site is in same climate state at the same time, be justified?

It was not clear in that study whether using the extra site was justifiable. Indeed *Thyer* (2001) states 'a robust methodology for deciding whether a site belongs to a homogeneous persistence region has yet to be developed'. It is the main objective of this thesis to develop such a methodology – to parsimoniously model the spatially non-homogenous long-term persistence apparent in Australian rainfall.

With this objective in mind, some generalisations of the HMM are proposed. Rather than having to choose which sites should be included in an analysis, the generalised models modify the regional state assumption in some way, thereby alleviating the need to choose sites.

2.7 Possible Generalisations of the HMM

The current multi-site HMM is not equipped to deal with at-site anomalies, nor are there appropriate procedures for choosing sites which should be included in an analysis. Two generalisations of the multi-site HMM are proposed in this thesis, aimed at addressing these issues:

- Switch HMM – Individual sites are permitted to vary from an overall regional climate state by the addition of 'switch' probabilities. That is, given the regional climate state, there is a probability that an individual site state will vary from that. This method was developed to accommodate anomalous sites. Given a regional state control has been identified, the switch parameters allow easy identification of sites which should not have been included in the analysis i.e. if a switch probability is significantly greater than zero at any site.
- Regional HMM – Enables automated partitioning of the HMM into homogeneous climate regions. Each climate region has its own hidden state series. Each climate partitioning is a different model. Some form of model selection is required to choose the optimum partitioning. This method was developed to address choosing which sites to include in analysis: anomalous or otherwise.

2.8 Conclusion

This chapter introduced stochastic modelling of the interaction of climate and rainfall. Empirical evidence of persistence over long timescales in hydroclimatic processes was presented. The current lack of modelling techniques to explicitly accommodate this spatially non-homogeneous long-term persistence provided the motivation for the work undertaken in this thesis.

The remainder of the chapter reviewed previous attempts at incorporating long-term persistence of climate into rainfall models. Links between the most successful of these models were discussed, with the majority conditioning the rainfall on a climate state variable. This climate state variable could be another measured process, or alternatively a latent parameter to be inferred upon calibration.

Of the models surveyed, few attempted to incorporate inter-annual persistence into small timescale rainfall processes. Persistence is typically modelled in two ways: with inter-annual dependency being modelled within annual rainfall (or stream flow) models, or day-to-day dependency within daily rainfall models. The objective of explicitly incorporating inter-annual dependency in short timescale models was introduced to overcome the current underestimation of inter-annual variability within such short timescale models.

The two-state HMM (*Thyer*, 2001) was introduced and compared to other closely related models for downscaling longer-term climate variations on short-duration rainfall. As this model has successfully identified inter-annual persistence in annual rainfall and runoff series at multiple regions throughout Australia, this model has been chosen to condition a short-duration rainfall model DRIP.

The state series of the current multi-site HMM is sensitive to the sites chosen for analysis, with *Thyer's* (2001) multi-site analysis lacking a rigorous rationale for choosing which sites belong to a climate region. Previously used methods for choosing which site to include in analysis were typically *ad hoc*, and the regional state assumption of the HMM could not be tested accurately. Therefore a formal method for choosing which sites to be included in a HMM analysis is required.

These considerations motivated two generalisations of the HMM aimed at modifying the regional state assumption, thereby alleviating the need for site selection. The first generalisation, the Switch HMM allows individual sites to differ from an overall controlling regional Markovian structure. The second, the regional HMM, on the other hand, allows sites to be grouped into different Markovian climate regions.

Chapter 3 Bayesian Modelling and Model Selection

3.1 Introduction

Given that new models are introduced in this thesis, some form of model discrimination/selection is required to choose between competing models or hypothesis. This chapter describes the modelling framework used in this thesis, with the framework being used primarily to discern between model hypotheses, with rigorous allowance for parameter uncertainty. This model selection framework has been rarely applied in the hydrological literature to date, as such a detailed description of this technique, along with a preliminary case study is presented. Although this chapter predominantly consists of a review of statistical literature, the author believes that such a review is required if these techniques are to be applied with understanding by hydrological practitioners. Hence, the understanding gained in the application of this technique is the essence of the contribution of this chapter towards the objectives of the thesis.

In hydrology, calibration methods typically do not account for parameter uncertainty, with an optimisation technique usually employed to find a parameter set that is optimal according to some defined criterion. Reliance on this optimum parameter set overestimates the confidence of output simulations from the model, as parameter uncertainty has not been incorporated. The Bayesian modelling paradigm specifically addresses this parameter uncertainty by making the parameters an object of inference. This chapter introduces the Bayesian modelling framework, along with a parameter sampling technique of Bayesian modelling, Markov Chain Monte Carlo (MCMC).

Two general MCMC sampling techniques have found wide use in Bayesian modelling: the Metropolis-Hastings (MH) sampler and the Gibbs sampler. Application of the MH algorithm to the HMM is presented here. This involves formulation of the model likelihood function, which in turn is used in calculation of the posterior state probability series. This study differs from the previous study of *Thyer* (2001) where the Gibbs sampler was used. The Metropolis-Hastings sampler was chosen for use here due to its simplicity and adaptability to the model selection context. This is demonstrated at the end of the chapter, with the application of a generalised MH sampler, the Metropolised Carlin-Chib (MCC) sampler.

Inherently linked to the modelling framework, is the question of how to choose one model over another. This chapter reviews this task in detail as this thesis relies heavily on model selection techniques to choose the most appropriate generalisation, or otherwise, of the HMM. Just as choosing sites to be included in the HMM involves a choice between models, so does exploring generalisations of the HMM involve a choice between models. Is the uncertainty introduced by the generalisations offset by the information gained? Is the model parsimonious? An objective method of choosing between models, whilst accounting for uncertainty and parsimony needs to be employed.

Choosing the 'right' model, given a set of data, is not usually a trivial task. Indeed, choosing the 'right' model selection method can be just as difficult. Of the many methods available, Bayesian Model Selection (BMS) is attractive in that it directly provides an estimate of the probability of one model compared to another given the data. This chapter provides an introduction to BMS and its use with MCMC sampling techniques. It illustrates BMS by comparing three annual point rainfall models for data collected in Sydney, Australia, using the Bayes factor and other related selection methods (Bayesian Information Criterion, the Posterior Bayes Factor, Akaike's Information Criterion the Likelihood Ratio Test). Particular attention is given to the sensitivity of the Bayes factor to the choice of prior.

3.2 Bayesian Modelling and Calibration with Application to the HMM

The general Bayesian framework is described first, and refined later for the particular models in use. We begin with some definitions – further details can be found in, for example, *Lee* (1989) and *Gelman et al.* (1995). Consider a set of observations $\mathbf{y}$ hypothesized to be a random realization from the probability model M with generalized probability density function $f(\mathbf{y} \mid M, \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a finite-dimensioned parameter vector. The function $f(\mathbf{y} \mid M, \boldsymbol{\theta})$ is given two labels depending on the context. When $f(\mathbf{y} \mid M, \boldsymbol{\theta})$ is used to describe the probability model generating the sample data $\mathbf{y}$ for a given $\boldsymbol{\theta}$ it is called the sampling distribution. However, when inference about the parameter $\boldsymbol{\theta}$ is sought, $f(\mathbf{y} \mid M, \boldsymbol{\theta})$ is called the likelihood function to emphasize that

the data $\mathbf{y}$ is known and the parameter $\boldsymbol{\theta}$ is the object of attention. We use the same notation for the sampling distribution and likelihood function to emphasize its oneness.

In Bayesian inference, the parameter vector $\boldsymbol{\theta}$ is considered a random vector whose probability distribution describes what is known about the true value of $\boldsymbol{\theta}$. Prior to analysing the data $\mathbf{y}$ knowledge about $\boldsymbol{\theta}$ given the probability model M is summarized by the probability distribution $p(\boldsymbol{\theta} \mid M)$ where $p(\cdot)$ is the generalized probability density. This density, referred to as the prior density, can incorporate subjective belief about $\boldsymbol{\theta}$. Bayes theorem is used to process the information contained in the data $\mathbf{y}$ by updating what is known about the true value of $\boldsymbol{\theta}$ as follows:

$$p(\boldsymbol{\theta} \mid \mathbf{y}, M) = \frac{f(\mathbf{y} \mid M, \boldsymbol{\theta}) \quad p(\boldsymbol{\theta} \mid M)}{p(\mathbf{y} \mid M)} \tag{3.1}$$

It is noted here that the collective parameter space Θ where the prior has a non-zero value is termed the support of the parameter $\boldsymbol{\theta}$. This support is symbolized by $\boldsymbol{\theta} \in \Theta$.

The posterior density $p(\boldsymbol{\theta} \mid \mathbf{y}, M)$ describes what is known about the true value of $\boldsymbol{\theta}$ given the data $\mathbf{y}$ and the model hypothesis M. The denominator $p(\mathbf{y} \mid M)$ is the marginal likelihood and is defined as:

$$p(\mathbf{y} \mid M) = \int_{\boldsymbol{\theta} \in \Theta} f(\mathbf{y} \mid M, \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid M)\, d\boldsymbol{\theta} \tag{3.2}$$

3.2.1 Sensitivity to prior specification

The choice of prior is an inherently subjective task. Accordingly the introduction of priors has been the subject of much controversy (*Berger*, 2000). The parameter prior $p(\boldsymbol{\theta} \mid M)$ can represent the modeller's subjective belief - for example an experienced modeller may have a prior belief about the region in which the true parameters lie before fitting a model. The parameter prior allows formal incorporation of this belief. On the other hand, it may be the case that the modeller has little or no idea of where the true parameters lie. In such a case a prior distribution with an equal density over all parameter values could represent the state of prior belief.

Of course the formulation of the model itself is quite subjective. As *Wikle* (2003) states 'One must simply recognize that a strength of the hierarchical (Bayesian) approach is the quantification of such subjective judgement'.

The posterior distribution can be sensitive to parameter prior specification (and hence can affect estimation). As such, much literature regarding the choice of a prior distribution has been directed towards the formulation of non-informative priors, priors which it can be argued that there is 'no information' about the parameter vector $\boldsymbol{\theta}$ (*Gelman et al.*, 1995 p.52-57, *Carlin and Louis*, 2000 p.38-32), implying the resulting analysis is completely objective rather than subjective. *Jeffreys* (1961 p.181) suggests a prior that is invariant under transformation (a desirable property – as the particular parameterisation transformation which is chosen by the modeller is subjective). Subsequent work has moved towards the calculation of reference priors (*Bernardo*, 1979) (for single parameter models) and later modified for multiparameter problems (*Berger and Bernardo*, 1992). However, it is recognised by *Bernardo and Smith* (2000 p.298), that 'every prior has some informative posterior and predictive implications' and 'there is no "objective" prior that represents ignorance'.

Reference priors for the HMM's and subsequent modified models used in this thesis have not been calculated (to the author's knowledge). As in the study of *Stephens* (1997 p.12) we have used proper priors which attempt to be only "weakly informative" – representative of our subjective prior belief. We likewise complete the analysis with the warning that we feel further work is required on the appropriate specification of priors.

All priors used in this thesis are proper, that is, they integral sums to one over the parameter space (*Gelman et al.*, 1995 p.52). Use of proper priors ensures that the posterior distribution is also proper – a requirement of Bayesian inference. Another factor related to weakly informative priors and mixture models is identifiability. This property, along with propriety or the priors, is discussed for the HMM and other newly introduced variants within section 4.5.

Usually (given enough data) the posterior density is not overly sensitive to the choice of prior density. However, as we will discover, choice of prior can influence the marginal

likelihood substantially. This becomes important in BMS, and as such will be discussed in more detail later in this chapter (section 3.7.1).

3.2.2 Posterior distribution

The posterior distribution is the focus of Bayesian modeling calibration, with output simulations and inference being dependent on its shape. Other calibration procedures typically involve a maximization process (e.g. Maximum likelihood, EM algorithm, least squares), with one vector of parameter values being used for simulation purposes. Bayesian modeling differs in that the entire posterior distribution is of interest, not just the maximum. Not including this uncertainty about the true value of $\boldsymbol{\theta}$ could therefore lead to over confident predictions and simulations for any one model.

3.3 MCMC Sampling Techniques

For the models considered in this study it is not possible to derive an analytical expression for the posterior distribution. In such a case it is possible to provide approximations to the posterior using mode finding coupled with multivariate normal or t-distribution approximations about the mode, and then improve upon the approximation using standard Monte Carlo importance sampling techniques (see *Gelman et al.*, 1995 – Chapters 11 and 12). The problem with such sampling techniques and approximations is that they 'may not be adequate for the inferential task at hand' – that is, they may be inefficient and/or inaccurate. A method that can provide efficient estimates of the posterior distribution is Markov Chain Monte Carlo (MCMC). The efficiency of MCMC when compared to independent sampling techniques is due to the Markovian dependency, ensuring that areas of low posterior probability are not sampled unnecessarily. Accuracy of the algorithm is guaranteed as it is proven that samples from MCMC algorithms converge to the posterior distribution.

MCMC calibration methods are employed to draw samples from the posterior distribution. The basic idea of MCMC methods is to simulate a Markov chain iterative sequence, where for each iteration *i*, a sample of the model parameters, $\boldsymbol{\theta}^{(i)}$, is generated according to a jump distribution $J_i\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}^{(i-1)}\right)$ dependent only on the previous sample's position $\boldsymbol{\theta}^{(i-1)}$. With each proposed jump there is also an associated probability

of accepting that jump $\alpha\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{\theta}^{(i-1)}\right)$. A sufficient condition for the simulated Markov chain to provide samples from the posterior distribution is *detailed balance*:

$$\begin{aligned} p\left(\boldsymbol{\theta}^{a} \mid \mathbf{y}, M\right) p\left(\boldsymbol{\theta}^{a}, \boldsymbol{\theta}^{b}\right) &= p\left(\boldsymbol{\theta}^{b} \mid \mathbf{y}, M\right) p\left(\boldsymbol{\theta}^{b}, \boldsymbol{\theta}^{a}\right) \\ p\left(\boldsymbol{\theta}^{a} \mid \mathbf{y}, M\right) J\left(\boldsymbol{\theta}^{b} \mid \boldsymbol{\theta}^{a}\right) \alpha\left(\boldsymbol{\theta}^{b} \mid \boldsymbol{\theta}^{a}\right) &= p\left(\boldsymbol{\theta}^{b} \mid \mathbf{y}, M\right) J\left(\boldsymbol{\theta}^{a} \mid \boldsymbol{\theta}^{b}\right) \alpha\left(\boldsymbol{\theta}^{a} \mid \boldsymbol{\theta}^{b}\right) \end{aligned} \qquad (3.3),$$

where $p\left(\boldsymbol{\theta}^{a}, \boldsymbol{\theta}^{b}\right)$ represents the unconditional transition probability from $\boldsymbol{\theta}^{a}$ to $\boldsymbol{\theta}^{b}$ of the sampling chain. For a jump distribution and associated acceptance probability satisfying this condition (and other mild conditions, see *Gelman et al.*, 1995 p. 325) the sampling chain converges with sufficient samples to a stationary distribution (*Mengersen and Tweedie*, 1996, *Roberts and Tweedie*, 1996), the posterior distribution $p\left(\boldsymbol{\theta} \mid \mathbf{y}, M\right)$. This distribution of parameters is used to evaluate posterior quantities of interest.

3.3.1 Posterior Quantities of Interest

After sampling using MCMC methods, it is often the case that an expectation of some function $\mathrm{G}\left(\boldsymbol{\theta}\right)$ over the posterior distribution will need to be estimated. Such quantities of interest are evaluated according to:

$$\mathrm{E}\left\{\mathrm{G}\left(\boldsymbol{\theta}\right)\right\} = \int \mathrm{G}\left(\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta} \mid \mathbf{y}, M\right) d\boldsymbol{\theta} \qquad (3.4).$$

Given that we have posterior samples from MCMC sampling, such expectations are evaluated according to numerical integration:

$$\mathrm{E}\left\{\mathrm{G}\left(\boldsymbol{\theta}\right)\right\} = \frac{1}{ns} \sum_{i=1}^{ns} \mathrm{G}\left(\boldsymbol{\theta}^{(i)}\right) \quad , \quad \boldsymbol{\theta}^{(i)} \leftarrow p\left(\boldsymbol{\theta} \mid \mathbf{y}, M\right) : i = 1, \ldots, ns \qquad (3.5).$$

One simple example of such a quantity would be where the function $\mathrm{G}(\boldsymbol{\theta})$ is $\boldsymbol{\theta}$ itself. Evaluation of (3.5) would give the posterior mean of the parameter set.

Evaluations of such integrals according to (3.4), and simulations from the model require samples from the posterior distribution. Two very general MCMC sampling techniques have found wide use in providing such samples, they are the Metropolis-Hastings (MH) sampler and the Gibbs sampler.

3.3.2 The Metropolis-Hastings Sampler

The Metropolis-Hastings algorithm (taken from *Gelman et al.*, 1995 p.324) is described below:

1. Draw a starting point $\boldsymbol{\theta}^0$, that has a positive posterior probability.

2. For i=1,2,…:

   a) Sample a candidate point $\boldsymbol{\theta}^*$ from a jump distribution at iteration *i*, $J_i\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(i-1)}\right)$. We will define this jump distribution shortly.

   b) Calculate the ratio of densities:

$$r = \frac{p\left(\boldsymbol{\theta}^* \mid \mathbf{y}, M\right) \Big/ J_i\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(i-1)}\right)}{p\left(\boldsymbol{\theta}^{(i-1)} \mid \mathbf{y}, M\right) \Big/ J_i\left(\boldsymbol{\theta}^{(i-1)} \mid \boldsymbol{\theta}^*\right)} \tag{3.6}$$

   c) Set

$$\boldsymbol{\theta}^{(i)} = \begin{cases} \boldsymbol{\theta}^* & \text{with probability } \min\left(r,1\right) \\ \boldsymbol{\theta}^{(i-1)} & \text{otherwise.} \end{cases} \tag{3.7}$$

   d) Check convergence – If sufficient samples taken, stop. Otherwise continue.

Detailed balance of the Metropolis-Hastings algorithm is verified through substitution of the acceptance probability $\alpha\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(i-1)}\right) = \min\left(r,1\right)$ into (3.3). Samples from this algorithm will converge to the target distribution $p\left(\boldsymbol{\theta} \mid \mathbf{y}, M\right)$ (the posterior distribution) under mild conditions as $i \to \infty$. In practice, the sampling is stopped at a point that approximates the posterior to some degree of satisfaction.

The aim of Bayesian modelling is to infer properties of the posterior. All of the samples given by MCMC sampling are used to summarise the posterior density and to compute quantiles, and other summaries as needed (*Gelman et al.*, 1995). Therefore, monitoring methods are required to give an indication of how representative the samples taken from the MCMC method are, in turn giving a point at which to stop sampling. Multiple chain MCMC is used for this purpose.

Using multiple chains allows testing of the individual chains against all of the samples thereby allowing testing of whether each chain is yielding samples from the same distribution. The measure used in this study, introduced by *Gelman and Rubin* (1992), compares the estimated within chain variance $\text{var}(\boldsymbol{\theta}_{chains} \mid \mathbf{y})$ (for a particular parameter), to the overall variance of samples taken from all chains $\text{var}(\boldsymbol{\theta}_{all} \mid \mathbf{y})$. If the ratio $\text{var}(\boldsymbol{\theta}_{all} \mid \mathbf{y}) / \text{var}(\boldsymbol{\theta}_{chains} \mid \mathbf{y})$ is close to 1, the overall variance is close to each chain's variance. Values significantly greater than 1 indicate that the chain's variance is less than the overall variance, hence the individual sequences have not had time to range over all of the target distribution. This 'scale reduction factor' is given by:

$$\hat{R} = \frac{\text{var}(\boldsymbol{\theta}_{all} \mid \mathbf{y})}{\text{var}(\boldsymbol{\theta}_{chains} \mid \mathbf{y})} \qquad (3.8).$$

For more detail see *Gelman et al.* (1995, p331-332). Values of $\sqrt{\hat{R}}$ below 1.2 for all parameters are recommended as being acceptable. This measure is very popular due to its simplicity, however it focuses only on statistics related to the mean and variance of the distribution, and resulting samples may be inadequate if reproduction of higher order statistics (eg. skew) are important. It is emphasised that there are many other and more complicated diagnostics available, yet none can guarantee convergence as the underlying distribution is unknown – for a review of convergence diagnostics see *Mengersen et al.* (1998).

The MH sampler requires the specification of two distributions, the posterior $p(\boldsymbol{\theta} \mid \mathbf{y}, M)$ and jump $J_i(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(i-1)})$ distributions. The posterior is calculated by simplifying (3.1) to:

$$p(\theta \mid \mathbf{y}, M) \propto f(\mathbf{y} \mid M, \theta) \;\; p(\theta \mid M) \qquad (3.9).$$

The marginal likelihood has been dropped from this equation as it is not a function of parameters, and as such, during sampling it can be considered a constant. As this normalising constant is equal for both $p(\boldsymbol{\theta}^* \mid \mathbf{y}, M)$ and $p(\boldsymbol{\theta}^{(i-1)} \mid \mathbf{y}, M)$ in the acceptance ratio calculation (3.6), only the likelihood and prior require calculation.

The jump distribution $J_i\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(i-1)}\right)$ determines the efficiency of the sampler. Generally, the better the jump distribution approximates the posterior, the more efficient the sampler will be. Properties of the jump distribution - distributional form, location, covariance *etc.* - are user chosen. If the covariance of the jump distribution is too large, jumps will be proposed to areas of low posterior probability, therefore producing low acceptance rates. However, if the covariance of the jump distribution is too small, the sampler will be slow in covering the distribution.

If a symmetric jump distribution is used (eg. Gaussian), the jump distribution densities in the acceptance ratio calculation are equal (the probability of jumping to a point is equal to that of jumping back), and hence drop out of the calculation. This simplification yields the original Metropolis sampler (*Metropolis et al.*, 1953). This algorithm was later generalised by *Hastings* (1970) to enable use of non-symmetric jumping distributions.

Many attempts have been made to improve the convergence efficiency of the MH sampler (eg. *Andrieu and Robert*, 2001). These attempts usually adjust the jump distribution during sampling to better approximate the posterior based on previous samples. Such techniques can cause the chain to lose its Markovian property, and proofs for convergence to the posterior (ergodicity) may not hold. These samplers run the risk of producing samples not from the posterior distribution, and estimates based on sample averages may be biased. Unfortunately, it is difficult to prove whether any one sampler will eventually converge to the posterior. Empirical tests based on synthetic studies are used. The methods used in simulations in this thesis have been widely used in many other studies, and it is expected that they do provide samples from the posterior given sufficient time to converge. These MH sampler modifications are listed below:

- Start all chains at mode of distribution (*Kuczera and Parent*, 1998), thereby reducing the chance of some chains taking a very long time to reach areas of high posterior probability.

- Use a multivariate Gaussian distribution for the jump distribution, with mean located at the current sample location $J_i\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(i-1)}\right) \sim N\left(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\Sigma}_J\right)$. This gives the random-walk Metropolis sampler noted in *Chib and Greenberg* (1995).

- An initial estimate of the posterior covariance, based on the Hessian around the mode, is used for the Gaussian covariance matrix $\boldsymbol{\Sigma}_J$.

- Initial 'warm up' samples are discarded, as using the same starting point for all chains could cause bias. Usually, this number is half of all the samples taken.

- Periodic updating of the jump distribution covariance $\boldsymbol{\Sigma}_J$ based on previous samples (*Kuczera and Parent*, 1998, *Haario et al.*, 1999)

- Periodic scaling of the jump distribution covariance to reproduce an optimal acceptance rate as defined in *Gelman et al.* (1995 p.335).

The sampler used throughout this thesis, with periodic updating of the jump distribution, has later been shown to be biased in some cases by *Haario et al*. (2001). 'As the updating rule depends on previous simulation steps, then the transition probabilities are more complicated than is stated in the Metropolis-Hastings algorithm, and the iterations will not necessarily converge to the target distribution' (*Gelman et al.*, 2004 p.307). That is, detailed balance (a sufficient but not necessary condition for convergence) may not be preserved. Development of adaptive samplers that produced unbiased sampled distributions have been undertaken by various authors (*Andrieu and Robert*, 2001, *Haario et al.*, 2001) and are recommended for future studies.

3.3.3 The Gibbs Sampler

In previous applications of the two-state HMM, the Gibbs sampler was used (*Chib*, 1996, *Thyer*, 2001). The Gibbs sampler (*Geman and Geman*, 1984) is a special case of the MH sampler. The parameter vector $\boldsymbol{\theta}^*$ to be sampled is broken into subvectors $\boldsymbol{\theta}^* = \left(\boldsymbol{\theta}_1^*,...,\boldsymbol{\theta}_d^*\right)$, with each subvector being sampled directly (from an analytically calculated conditional distribution). Sampling directly from these conditional distributions ensures that the acceptance ratio for every jump is one – all samples are accepted. At each iteration, the Gibbs sampler cycles through each of the subvectors of $\boldsymbol{\theta}^*$, drawing each subset conditional on the value of all others – for subvector j $\boldsymbol{\theta}_j^* \sim p\left(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_{-j}^*, \mathbf{y}\right)$ where $\boldsymbol{\theta}_{-j}^* = \left(\boldsymbol{\theta}_1^*,...,\boldsymbol{\theta}_{j-1}^*,\boldsymbol{\theta}_{j+1}^*,...,\boldsymbol{\theta}_d^*\right)$. These conditional distributions can be calculated for most commonly used statistical models.

3.4 Application of the MH algorithm to the HMM

Although the Gibbs sampler can be used for HMM problems, it was not used in this study. The Metropolis sampler was considered to be simpler to implement primarily because there was no need to treat the hidden states as parameters. As a result, the computational trapping states detailed in *Thyer* (2001) are no longer encountered. Although the Gibbs sampler is more efficient in a computational sampling sense (every sample is accepted), the MH sampler avoids programming complicated conditional distributions, with a higher inherent risk of programming error. An additional bonus of using the MH sampler is that it is very easily adaptable to the more general case of sampling from model to model (model selection). This will become apparent when discussed later in this chapter.

3.4.1 Calculation of HMM Likelihood

The Metropolis algorithm requires the calculation of the likelihood (and prior) at each iteration of the sampler. This likelihood calculation requires the application of the forward phase of the Baum-Welch recursive algorithm. For further details of this general algorithm for HMM's see *Bengio* (1999). Noting that each year's rainfall is only dependent on the previous year's rainfall through the (hidden) Markov state, the HMM likelihood can be written as:

$$p\left(\mathbf{Y}_1^T \mid \boldsymbol{\theta}\right) = p\left(\mathbf{y}_1 \mid \boldsymbol{\theta}\right)\prod_{t=2}^{T} p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) \qquad (3.10).$$

This likelihood substitutes for $f(\mathbf{y} \mid M, \boldsymbol{\theta})$ in (3.9) where M is in this case the two-state HMM, and $\boldsymbol{\theta}$ is the particular parameter set. Because (3.9) can be evaluated, MH sampling can be used to provide inference on the posterior parameter distribution. The parameter vector used here is $\boldsymbol{\theta} = \left(\boldsymbol{\mu}_{\mathbf{W}},\ \boldsymbol{\sigma}_{\mathbf{W}},\ \boldsymbol{\mu}_{\mathbf{D}},\ \boldsymbol{\sigma}_{\mathbf{D}}, \boldsymbol{\rho},\ \mathbf{P}\right)$, with wet and dry mean and variance parameters for every site $\left(\boldsymbol{\mu}_{\mathbf{W}},\ \boldsymbol{\sigma}_{\mathbf{W}},\ \boldsymbol{\mu}_{\mathbf{D}},\ \boldsymbol{\sigma}_{\mathbf{D}}\right)$, a correlation coefficient matrix $\boldsymbol{\rho} = \left[\rho_{ij}\right] : i, j = 1,...,d$ that is independent of state, and a single set of regional transition probabilities $\mathbf{P} = \left[\ p_{ij}\ \right] = p\left(r_t = j \mid r_{t-1} = i\right)\ i, j = W, D$. Note that the parameter vector used, does not include the latent regional state series $R_1^T = \left(r_1, r_2, ..., r_T\right)$. This is the important difference when comparing the algorithm to the

Gibbs sampler, the likelihood (3.9) integrates out these regional states, eliminating the need to infer the state series.

To evaluate (3.10), we carry out the following set of calculations repeatedly, starting at the first time-step ($t=1$), for each time-step until the final time-step T. Here the set of data until time t is signified by $\mathbf{Y}_1^t = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t)$. Noting that r_t is the (hidden) regional state at time-step t, the conditional probability $p(r_t | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta})$ can be obtained using total probability:

$$\begin{aligned} p(r_t | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) &= \sum_{r_{t-1}} p(r_t | r_{t-1}, \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) p(r_{t-1} | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) \\ &= \sum_{r_{t-1}} p(r_t | r_{t-1}, \boldsymbol{\theta}) p(r_{t-1} | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) \end{aligned} \quad (3.11).$$

One of the assumptions of the HMM is that the state at each timestep is only conditional on the previous timestep state. Hence the simplification of $p(r_t | r_{t-1}, \mathbf{Y}_1^{t-1}, \boldsymbol{\theta})$ to $p(r_t | r_{t-1}, \boldsymbol{\theta})$, the transition probability. Note the conditioning on the parameter vector throughout the derivation. For time-step t the vector $\mathbf{y}_t$ is drawn from a multivariate normal $\mathbf{y}_t | r_t \sim N(\boldsymbol{\mu}_{r_t}, \boldsymbol{\Sigma}_{r_t})$, where $\boldsymbol{\mu}_{r_t} = \left[\mu_{r_t}^{site} \right] : site = 1, \ldots, d$ and $\boldsymbol{\Sigma}_{r_t} = \left[\rho_{ij} \sigma_{r_t}^i \sigma_{r_t}^j \right] : i, j = 1, \ldots, d$. Using total probability the conditional density $p(\mathbf{y}_t | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta})$ can be obtained:

$$\begin{aligned} p(\mathbf{y}_t | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) &= \sum_{r_t} p(\mathbf{y}_t | r_t, \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) p(r_t | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) \\ &= \sum_{r_t} p(\mathbf{y}_t | r_t, \boldsymbol{\theta}) \, p(r_t | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) \end{aligned} \quad (3.12).$$

The state probability at time-step t given $\mathbf{Y}_1^t$, $p(r_t | \mathbf{Y}_1^t, \boldsymbol{\theta})$, can be obtained using Bayes' theorem yielding:

$$\begin{aligned} p(r_t | \mathbf{Y}_1^t, \boldsymbol{\theta}) &= \frac{p(\mathbf{y}_t | r_t, \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) \, p(r_t | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta})}{p(\mathbf{y}_t | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta})} \\ &= \frac{p(\mathbf{y}_t | r_t, \boldsymbol{\theta}) \, p(r_t | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta})}{p(\mathbf{y}_t | \mathbf{Y}_1^{t-1}, \boldsymbol{\theta})} \end{aligned} \quad (3.13).$$

Now as we can calculate $p\left(r_t \mid \mathbf{Y}_1^t, \boldsymbol{\theta}\right)$ we can evaluate (3.11) for the next time-step, namely $t+1$. Therefore, we can repeat the calculation for the next time-step, and so on. Using the result of (3.12) for each time-step we calculate the overall likelihood by evaluating (3.10).

3.4.2 Posterior Hidden State Series

The posterior hidden state probability series, as introduced in Section 2.5.2, is calculated as a product of the calibration process. Reiterating, this is the posterior probability of a particular state r_t occurring at time-step t, $p\left(r_t \mid \mathbf{Y}_1^T, \boldsymbol{\theta}\right)$, where $t = 1,...,T$. Note that this probability is conditioned on all of the available data, rather than the data until time t as in (3.13). The required probability can be obtained using Bayes theorem:

$$\begin{aligned} p\left(r_t \mid \mathbf{Y}_1^T, \boldsymbol{\theta}\right) &= \frac{p\left(\mathbf{Y}_{t+1}^T \mid r_t, \mathbf{Y}_1^t, \boldsymbol{\theta}\right) p\left(r_t \mid \mathbf{Y}_1^t, \boldsymbol{\theta}\right)}{p\left(\mathbf{Y}_{t+1}^T \mid \mathbf{Y}_1^t, \boldsymbol{\theta}\right)} \\ &= \frac{p\left(\mathbf{Y}_{t+1}^T \mid r_t, \boldsymbol{\theta}\right) p\left(r_t \mid \mathbf{Y}_1^t, \boldsymbol{\theta}\right)}{p\left(\mathbf{Y}_{t+1}^T \mid \mathbf{Y}_1^t, \boldsymbol{\theta}\right)} \end{aligned} \qquad (3.14).$$

The probability $p\left(\mathbf{Y}_{t+1}^T \mid r_t, \mathbf{Y}_1^t, \boldsymbol{\theta}\right)$ simplifies to $p\left(\mathbf{Y}_{t+1}^T \mid r_t, \boldsymbol{\theta}\right)$ due to the Markovian assumption of the HMM. $p\left(r_t \mid \mathbf{Y}_1^t, \boldsymbol{\theta}\right)$ is given in (3.13). $p\left(\mathbf{Y}_{t+1}^T \mid \mathbf{Y}_1^t, \boldsymbol{\theta}\right)$ is a constant independent of r_t and therefore is a normalising constant. The remaining term $p\left(\mathbf{Y}_{t+1}^T \mid r_t, \boldsymbol{\theta}\right)$ requires the 'backward' or smoothing phase of the Baum-Welch recursion. With a similar set of calculations to the likelihood, starting at time-step $t = T$ and working backwards in time we obtain:

$$\begin{aligned} p\left(\mathbf{Y}_t^T \mid r_{t-1}, \boldsymbol{\theta}\right) &= \sum_{r_t} p\left(\mathbf{Y}_t^T, r_t \mid r_{t-1}, \boldsymbol{\theta}\right) \\ &= \sum_{r_t} p\left(\mathbf{Y}_{t+1}^T \mid \mathbf{y}_t, r_t, r_{t-1}, \boldsymbol{\theta}\right) p\left(\mathbf{y}_t, r_t \mid r_{t-1}, \boldsymbol{\theta}\right) \\ &= \sum_{r_t} p\left(\mathbf{Y}_{t+1}^T \mid r_t, \boldsymbol{\theta}\right) p\left(\mathbf{y}_t, r_t \mid r_{t-1}, \boldsymbol{\theta}\right) \end{aligned} \qquad (3.15),$$

where by virtue of Bayes theorem:

$$p(\mathbf{y}_t, r_t \mid r_{t-1}, \boldsymbol{\theta}) = p(\mathbf{y}_t \mid r_t, r_{t-1}, \boldsymbol{\theta})\, p(r_t \mid r_{t-1}, \boldsymbol{\theta}) \\ = p(\mathbf{y}_t \mid r_t, \boldsymbol{\theta})\, p(r_t \mid r_{t-1}, \boldsymbol{\theta}) \qquad (3.16).$$

Note at the initial time-step $t = T$, $p\left(\mathbf{Y}_{t+1}^{T} \mid r_t, \boldsymbol{\theta}\right) = 1$ is inserted into (3.15) so that $p\left(\mathbf{Y}_t^{T} \mid r_{t-1}, \boldsymbol{\theta}\right) = \sum_{r_t} p(\mathbf{y}_t, r_t \mid r_{t-1}, \boldsymbol{\theta})$.

The intuition behind this calculation is that the data observed after the time-step in which the state occurred provides information about the state in question.

3.5 Bayesian Model Selection

Closely related to the likelihood, and sampling from the posterior distribution (in the Bayesian framework), is the question of model selection. The modelling selection technique used in this thesis is Bayesian model selection (BMS).

BMS has seen limited application in the hydrology literature; see *Perreault et al.* (2000a) and *Campbell et al.* (1999) for examples. This may be partly due to the perception that these methods are difficult to implement compared to more traditional methods. However, recent advances make BMS more practicable. Of significance is the fact that MCMC methods enable simple and accurate implementation of BMS procedures. Nonetheless, as we have discovered, BMS demands care in its interpretation.

This section briefly reviews some of the key ideas in BMS. BMS as used herein refers to a specific method of model comparison (through the posterior model probability and related Bayes factors – see the following section). A compact overview of BMS and some other related model comparison methods is appropriate in order to highlight some of the conceptual and practical difficulties that may be encountered when attempting to decide between competing models. It draws principally from the works by *Kass and Raftery* (1995), *Gelfand and Dey* (1994), *Aitkin* (1991) and *Wasserman* (2000).

3.5.1 Marginal Likelihood and the Bayes Factor

We consider the case in which there are nm competing models $\{M_1, ..., M_{nm}\}$. Our interest is to compute the posterior probability $p(M_i \mid \mathbf{y})$, $i = 1, ..., nm$ which describes

the probability that, out of the nm competing models, model M_i generated the observed data $\mathbf{y}$. Application of Bayes theorem yields:

$$p(M_i \mid \mathbf{y}) = \frac{p(\mathbf{y}|M_i) \quad p(M_i)}{\sum_{j=1}^{nm} p(\mathbf{y}|M_j) \quad p(M_j)} \qquad (3.17),$$

where $p(M_i)$ is the prior probability that the model M_i generated the data.

It is fair to express some skepticism about the possibility that the true model is one of the nm competing models (*Wasserman*, 2000). However, the posterior probability $p(M_i \mid \mathbf{y})$ enables us to rank competing models.

For two competing models M_i and M_j, (3.17) can be rearranged to yield the posterior odds that M_i is more likely than M_j given the data $\mathbf{y}$. Thus:

$$\frac{p(M_i \mid \mathbf{y})}{p(M_j \mid \mathbf{y})} = \frac{p(M_i) \quad p(\mathbf{y} \mid M_i)}{p(M_j) \quad p(\mathbf{y} \mid M_j)} = BF(M_i, M_j)\frac{p(M_i)}{p(M_j)} \qquad (3.18).$$

The ratio of the model priors is called the prior odds, while the ratio of the marginal likelihoods is the Bayes factor - defined as $BF(M_i, M_j)$. Because the prior odds is often assigned a value of 1 (for reasons outlined later), the Bayes factor becomes important in BMS and thus has received wide attention in the statistical literature. *Kass and Raftery* (1995) give a thorough overview of Bayes factors, while *Wasserman* (2000) discusses BMS in an overview paper aimed at a non-specialist audience. They both suggest interpretive scales for Bayes factors based on *Jeffreys* (1961) – Kass and Raftery's version is shown in Table 3.1. Note that when the BF is less than one it needs to be inverted with the resulting number interpreted as the evidence in favour of M_j.

Table 3.1 Bayes Factor Interpretive Scale

Bayes Factor BF(M_i,M_j)	Evidence in favour of M_i
1 to 3	Weak
3 to 20	Positive
20 to 150	Strong
>150	Very Strong

3.6 Consistency of the Bayes Factor

Why would we want to use Bayes factors? The short answer is that the Bayes factor will choose the true model given enough data provided that the true model is one of the competing models. This is due to a property of the posterior distribution called consistency, which put roughly states that given enough data, the expected value of the posterior will converge to the true parameter.

3.6.1 Nested model consistency

The consistency of Bayes factors is illustrated for the nested model selection problem where M_0 is the reduced model with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and M_1 is the full model with $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. The Bayes factor (BF), the ratio of the marginal likelihoods, is given by:

$$BF(M_0, M_1) = \frac{f(\mathbf{y}|\boldsymbol{\theta}_0, M_1)}{\int f(\mathbf{y}|\boldsymbol{\theta}, M_1) \quad p(\boldsymbol{\theta} \mid M_1) d\boldsymbol{\theta}} \qquad (3.19).$$

There is no integral term in the numerator because according to the reduced model M_0 there is only a single parameter value $\boldsymbol{\theta}_0$. In the numerator the likelihood is conditioned on the full model M_1 to emphasize that M_0 is a special parametric case of M_1.

It is assumed in the following example that the data was generated from one of the models under consideration; that is, one of the models can be considered the true model.

We consider a simple example involving data generated by an independent and identically distributed Gaussian model with known variance σ^2. We want to test whether the true model is $M_0 : \theta = \theta_0$ or $M_1 : \theta \neq \theta_0$, where θ is the unknown mean of the Gaussian distribution. Given the known variance σ^2, the likelihood function is:

$$\begin{aligned} f(\mathbf{y} \mid \boldsymbol{\theta}, M_1) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^T \prod_{t=1}^{T} \exp\left(-\frac{1}{2\sigma^2}(y_t - \theta)^2\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^T \exp\left(-\frac{(T\text{-}1)\bar{s}^2}{2\sigma^2}\right) \exp\left(-\frac{T}{2\sigma^2}(\bar{y} - \theta)^2\right) \end{aligned} \qquad (3.20),$$

where $\bar{y}$ is the mean of the data and $\bar{s}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(y_t - \bar{y})^2$ is the sample variance. Note here we are considering univariate series with scalar data y_t only at each point *t*.

Substituting (3.20) into (3.19) and then using a uniform prior on θ over the range $[-c, c]$ for model M_1 yields:

$$\begin{aligned} BF(M_0, M_1) &= \frac{2c \exp\left(-\frac{T}{2\sigma^2}(\bar{y} - \theta_0)^2\right)}{\int_{-c}^{c} \exp\left(-\frac{T}{2\sigma^2}(\bar{y} - \theta)^2\right) d\theta} \\ &= \frac{2 \; c\sqrt{T} \; \exp\left(-\frac{T}{2\sigma^2}(\bar{y} - \theta_0)^2\right)}{\sigma\sqrt{2\pi}\left[\Phi\left(\frac{\bar{y} + c}{\sigma/\sqrt{T}}\right) - \Phi\left(\frac{\bar{y} - c}{\sigma/\sqrt{T}}\right)\right]} \end{aligned} \qquad (3.21).$$

Note that the denominator is finite in (3.21). As the number of data T increases, the sample mean $\bar{y}$ converges to the true mean θ_{true}. If, in fact $\theta_{true} = \theta_0$, then the exponential term in the numerator $\to 1$. Therefore, as $T \to \infty$, $BF \to \infty$ which favours M_0, the true model. In contrast, when M_1 is the true model, namely $\theta_{true} \neq \theta_0$, then as $T \to \infty$, the numerator term $\to 0$. Therefore, the $BF \to 0$ which favours M_1, the true model.

3.6.2 General Consistency for Nested and Non-nested models

For finite-dimensional (parametric) models, the posterior distribution can be shown to be consistent under mild regularity conditions (*Schervish*, 1995). This is important because consistency of the posterior guarantees that the true model will be chosen given a sufficiently large number of data. It is accepted that the 'true' model is unlikely to be contained in the models that are being compared. However, it is reassuring to know that if the true model is in the set of competing models, then it will be chosen as the favoured model given sufficient data.

It is emphasised again that even if the true model is not contained in the set that is being compared, the Bayes factor coupled with the model priors provides the ratio of posterior probabilities of one model compared to another given the data. This said, there are pitfalls associated with BMS, the topic of the next section.

3.7 Sensitivity of Bayes Factors to Choice of Prior

In BMS two priors must be specified. First, given a model M, the prior parameter density $p(\boldsymbol{\theta} \mid M)$ requires specification. Second, the prior probability of the model being most consistent with the data $p(M)$ must also be specified. These are inherently subjective tasks and are very important in BMS as the conclusions made can be sensitive to both of these priors.

3.7.1 Parameter Prior

As mentioned in section 3.2.1, priors used in this study were chosen with the aim of being weakly informative. Regarding the HMM in this thesis, priors following *Robert* (1996) and subsequently *Thyer* (2001) were used. These priors were chosen based on mathematical convenience (conjugacy), with empirical Bayes methods used to define location and scale (*Carlin and Louis*, 2000). Use of such priors can be criticised due to their subjectivity (compared to reference priors), but as mentioned previously, reference priors for the models presented in thesis have not been derived.

Often, a uniform parameter prior is employed if there is little prior experience with a model. Unfortunately, a uniform (equal density or flat) priors are not invariant to transformation – and can therefore be quite informative depending on the

transformation used (*Carlin and Louis*, 2000 p.29). As *Carlin and Louis* (2000) note, 'a possible remedy to this problem is to rely on the particular modelling context to provide the most reasonable parameterization and, subsequently, apply the uniform prior on this scale'. Uniform and empirically based priors are used throughout this thesis, and all conclusions made are done so with the caveat that reference priors are not used.

This uniform prior, must however be used with caution in BMS.

3.7.2 Lindley's Paradox

If a uniform prior that ranges to infinity ($c = \infty$) is employed in equation (3.21), the Bayes factor is infinite regardless of whether M_0 is the true model or not. This result, known as Lindley's paradox, has received considerable attention in the literature (e.g., *Bartlett*, 1957, *Lindley*, 1957, *Berger*, 1985). A simple way of sidestepping this is to use bounded priors; that is, require that c have a finite value. Often there are natural bounds on parameters while other times it is left to the modeller to judge the range of parameters within which the true parameter could lie.

What is disturbing is that the Bayes factor in equation (3.21) is effectively proportional to c - and this problem is general to all modelling cases where a uniform unbounded prior is used. This suggests that considerable thought needs to be given to specification of the parameter prior.

3.7.3 Model Prior

The modeller may also have some prior subjective belief about which of the competing models is the best. Such a belief may arise from previously favourable results using a particular model or from knowledge that a particular model better represents the underlying physical characteristics of the process. However, it is usual practice to specify non-informative model priors where all model priors are equal. Because the Bayes factor is independent of model priors, the sensitivity of the posterior odds to model priors can be checked after calculation of the Bayes factor.

3.7.4 Bayes factors versus posterior distributions

Gelman et al. (1995 p.175-177) have questioned the value of BMS in some applications. They argued that in cases where competing models are distinctly and

legitimately different, for example, the two-state HMM and lag-one autoregressive (AR1) models presented in the case study, Bayes factors coupled with proper non-informative model priors provide a worthwhile means for model selection. However, in cases where both competing models are special cases of a more general parametric model, they argue that BMS may be a hindrance because the true model is likely not to be one of the special cases. They recommend that the posterior distribution of the parameters characterizing the different models be studied.

Although the nested model selection case cited in Section 3.6.1 does not fit into *Gelman et al's* latter category we also believe that use of the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y}, M_1)$ may be preferable to BMS. Difficulties may arise in assigning prior probabilities to the reduced model M_0 and to its parametric generalization - the full model M_1.

For example, consider the case of the AR1 model with $\boldsymbol{\theta}$ being the lag-one correlation. In the case study we are interested in comparing the model $M_0 : \boldsymbol{\theta} = 0$ versus its complement $M_1 : \boldsymbol{\theta} \neq 0$. If the data are annual streamflows (not rainfall), there is good prior evidence in favour of the hypothesis that annual streamflows are weakly correlated (see *Yevjevich*, 1963). Therefore, $p(M_1) > p(M_0)$. The fundamental problem is how we assign a probability to $p(M_0 : \boldsymbol{\theta} = 0)$ when the prior evidence is in the form of a probability density about $\boldsymbol{\theta}$ – in such a case the prior probability that $\boldsymbol{\theta} = 0$ is 0. Obviously, this problem would go away if model M_0 hypothesized that $\boldsymbol{\theta}$ lies in a finite interval. However, we believe it is more natural to study the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y}, M_1)$.

3.8 Other Model Selection Criteria

Brief mention is made of some other related model selection methods: the likelihood ratio test, posterior Bayes factors, Akaike's information criterion, and the Schwarz or Bayesian information criterion.

3.8.1 Likelihood Ratio

Neyman and Pearson (see *Marden*, 2000) developed the likelihood ratio test (LRT). Its primary application is for the nested model selection problem where the likelihood ratio LR is:

$$LR(M_0, M_1) = \frac{f(\mathbf{y} \mid \boldsymbol{\theta}_0, M_1)}{f(\mathbf{y} \mid \hat{\boldsymbol{\theta}}, M_1)} \qquad (3.22),$$

with $\hat{\boldsymbol{\theta}}$ being the maximum likelihood estimate of $\boldsymbol{\theta}$ for the full model M_1. If LR is less than some threshold (for large samples the threshold is based on the χ^2 distribution), then the full model M_1 is chosen. The LRT is prone to misinterpretation (*Berger and Sellke*, 1987) and is only applicable to nested model selection problems. Also, as explained in *Gelfand and Dey* (1994), the LRT is inconsistent in that there is always a positive probability of choosing the full model when the reduced model is true, even with a large number of data.

To compensate for the bias by the LRT towards the more complex model, several penalized forms of the LR have been proposed, of which Akaike's Information Criterion (AIC) and the Schwarz Criterion (SC) or Bayesian Information Criterion (BIC) are the best known.

3.8.2 AIC

Akaike (1973) introduced a penalized maximized likelihood method for choosing between models. For model M_i, the Akaike's Information Criterion (in a slightly modified form) is:

$$AIC(M_i) = \log f\left(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_i, M_i\right) - k_i \qquad (3.23),$$

where $\hat{\boldsymbol{\theta}}_i$ is the maximum likelihood estimate of $\boldsymbol{\theta}$ and k_i is the number of parameters estimated in model M_i. The model with the largest AIC is the preferred model. Model complexity is penalized on the grounds that the complex model should achieve a higher likelihood than a simpler model (meaning a model with fewer fitted parameters). Thus if two models produce the same likelihood, then the simpler model will be favoured. The AIC was designed to choose the model that best reproduced the data in terms of the

predictive distribution using each model's maximum likelihood parameter estimate. Thus, only the optimal parameter set in terms of likelihood is used to judge the models performance. The Bayesian perspective of the AIC, as discussed in *Kass and Raftery* (1995) is that 'such a predictive distribution is incorrect because it does not incorporate the uncertainty about parameter values and model form'.

3.8.3 Schwarz Criterion

Wasserman (2000) notes that the log marginal likelihood can be approximated by:

$$\log\left(p\left(\mathbf{y}\mid M\right)\right) = \left(\log\left(f\left(\mathbf{y}\mid\hat{\boldsymbol{\theta}},M\right)\right) - \frac{k}{2}\log T\right) + \log\left(1+O(1)\right) \qquad (3.24),$$

where $O(1)$ refers to an error independent of T. This leads to an approximation of the log BF known as the Schwarz criterion (see also the BIC in *Kass and Raftery*, 1995) :

$$\log BF\left(M_i,M_j\right) \approx SC\left(M_i,M\right)_j = \log\left(\frac{f\left(\mathbf{y}\mid\hat{\boldsymbol{\theta}}_i,M_i\right)}{f\left(\mathbf{y}\mid\hat{\boldsymbol{\theta}}_j,M_j\right)}\right)+\frac{\left(k_j-k_i\right)}{2}\log T \qquad (3.25).$$

Use of SC was proved to be consistent by *Schwarz* (1978) for linear exponential models. A proof for general model types has not yet been produced, but as *Kass and Raftery* (1995) explain, the approximation appears to hold much more generally.

As *Kass and Wasserman* (1995 p.928) state, 'at least for some priors, $\exp(SC)$ will be a poor approximation to the Bayes factor and thus a dubious quantification of evidence in favour of a model'. As we will discover in the case study, this error can compromise model selection using the interpretive scale in Table 3.1 especially if there is insufficient data to unequivocally show one model is superior to another. On the other hand, priors can be chosen such that the error has the size $O(T^{-1/2})$ (see *Wasserman*, 2000) - the unit information prior, a prior based on the Fisher information (*Kass and Wasserman*, 1995) is an example of such a prior. Unfortunately, the requirement of using certain priors limits the flexibility of prior specification such that even in cases where the prior distribution is 'known' to some degree, it cannot be used.

3.8.4 Posterior Bayes Factor

Aitkin (1991) introduced the posterior Bayes factor (PBF) to reduce the sensitivity of the Bayes factor on the choice of parameter prior, particularly regarding Lindley's paradox. Instead of averaging the likelihood over the parameter prior as for the marginal likelihood in the BF, *Aitkin* suggested averaging over the parameter posterior to yield:

$$PBF(M_i, M_j) = \frac{\int f\left(\mathbf{y} \mid \boldsymbol{\theta}, M_i\right) p\left(\boldsymbol{\theta} \mid \mathbf{y}, M_i\right) d\boldsymbol{\theta}_i}{\int f\left(\mathbf{y} \mid \boldsymbol{\theta}, M_j\right) p\left(\boldsymbol{\theta} \mid \mathbf{y}, M_j\right) d\boldsymbol{\theta}_j} \qquad (3.26).$$

The PBF is intuitively attractive in that it averages the likelihood over the posterior, or fitted distribution. This procedure uses the posterior in judging a model as do the widely practiced posterior predictive tests (see *Gelman et al.*, 1995 p.167). Hence, we can judge the model's 'postdictive' performance (see commentary of *Aitkin*, 1997 by Dempster). However, the PBF has not been recommended by *Kass and Raftery* (1995) as formally speaking 'the procedure has little Bayesian justification' – the PBF cannot be used in place of BF in (3.18) hence denying use of posterior odds. Nonetheless, it is particularly simple to evaluate using MCMC methods. *Gelfand and Dey* (1994) show that an asymptotic approximation to the PBF is:

$$\log PBF\left(M_i, M_j\right) \approx \log\left(\frac{f\left(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_i, M_i\right)}{f\left(\mathbf{y} \mid \hat{\boldsymbol{\theta}}_j, M_j\right)}\right) + \frac{\left(k_j - k_i\right)}{2} \log 2 \qquad (3.27).$$

This makes clear that the PBF is another penalized form of the likelihood ratio – see *Gelfand and Dey* (1994) for further details.

3.8.5 Asymptotic consistency

The BF along with the SC is appealing in that given a large number of samples it will choose the true model; a property that can not be claimed for the LRT (p-tests), PBF or the AIC (*Aitkin*, 1997). Conversely, when a large number of samples are not available there are cases where the BF chooses the wrong model while other methods such as PBF or AIC choose the correct model (*Zhang*, 1993). As *Wasserman* (2000) observes, for small samples there has been no systematic study to test the ability of each method of model choice.

3.8.6 Other Methods

No attempt has been made here to exhaustively describe model selection methods related to the formal Bayesian method (for an overview of Bayesian model selection variants and computational methods see *Ntzoufras*, 1999). Rather, those which seemed most easily applied given that we were already using MCMC methods for posterior sampling were considered. Recent work has focussed on cross validation techniques (see *Gelfand and Dey*, 1994), in particular methods related to the reference approach such as intrinsic Bayes factors (*Berger and Pericchi*, 1996) and fractional Bayes factors (*O'Hagan*, 1995). *Bernardo and Rueda* (2002), whilst maintaining a reference approach, develop a Bayesian hypothesis testing method in a formal decision setting. An alternate idea that uses the posterior distribution of the likelihood is outlined in *Aitkin* (1997). This method is appealing in that conclusions consistent with the posterior distribution of parameters can be made. More recently, the Deviance Information Criterion (DIC) has been proposed by *Spiegelhalter et al.* (2002) for use in hierarchical modelling where the number of parameters is not clearly defined. Related to the cross validation measures mentioned previously are the widely practiced posterior predictive tests (see *Gelman et al.*, 1995 p.167).

3.9 MCMC Estimation of the Bayes factor

Whereas calculation of the asymptotic approximation to the BF using the SC is straightforward, there is concern that it may not adequately approximate the BF. We have seen that the SC may have an error of order O(1) in approximating the numerator and denominator terms in the log BF. A more accurate estimate of BF may be desired, particularly if data samples are not large. Given our use of MCMC methods to evaluate the posterior distribution, this section focuses on the estimation of the BF using MCMC methods.

An MCMC method generates samples from the posterior distribution. At the i^{th} iteration of the MCMC algorithm a parameter $\boldsymbol{\theta}^{(i)}$ is drawn from the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y}, M)$. To minimise additional computation we would like to compute the marginal likelihood using the information produced by the MCMC method. We consider several such estimators.

3.9.1 Newton-Raftery Approximation

Newton and Raftery (1994) proposed a simple estimator based on the harmonic mean of likelihood function. Starting with the identity:

$$1 = \int p(\boldsymbol{\theta} \mid M) \; d\boldsymbol{\theta} \qquad (3.28),$$

and using (3.1) to substitute for $p(\boldsymbol{\theta} \mid M)$ yields:

$$\frac{1}{p(\mathbf{y} \mid M)} = \int \frac{1}{f(\mathbf{y} \mid \boldsymbol{\theta}, M)} \; p(\boldsymbol{\theta} \mid \mathbf{y}, M) \; d\boldsymbol{\theta} \qquad (3.29).$$

Noting that $1/p(\mathbf{y} \mid M)$ is the posterior expected value of $1/f(\mathbf{y} \mid \boldsymbol{\theta}, M)$ the expectation can be approximated by the arithmetic mean to yield the Newton-Raftery estimator:

$$p_{NR}(\mathbf{y} \mid M) = \left\{ \frac{1}{ns} \sum_{i=1}^{ns} \frac{1}{f(\mathbf{y} \mid \boldsymbol{\theta}^{(i)}, M)} \right\}^{-1} \quad \boldsymbol{\theta}^{(i)} \leftarrow p(\boldsymbol{\theta} \mid \mathbf{y}, M), \quad i = 1,..,ns \qquad (3.30),$$

where $\boldsymbol{\theta}^{(i)} \leftarrow p(\boldsymbol{\theta} \mid \mathbf{y}, M)$ denotes samples are drawn from the posterior distribution. *Kass and Raftery* (1995) note that, although this estimator is consistent, it is unstable – the occasional small likelihood can have a marked effect on the estimator.

3.9.2 Gelfand-Dey Estimate

Gelfand and Dey (1994) proposed a potentially more stable estimator than the Newton-Raftery estimator. For any proper density $\tau(\boldsymbol{\theta})$ we have:

$$1 = \int \tau(\boldsymbol{\theta}) \; d\boldsymbol{\theta} \qquad (3.31).$$

Using the identity (3.1) yields:

$$1 = \int \tau(\boldsymbol{\theta}) \frac{p(\boldsymbol{\theta} \mid \mathbf{y}, M)\, p(\mathbf{y} \mid M)}{f(\mathbf{y} \mid \boldsymbol{\theta}, M)\, p(\boldsymbol{\theta} \mid M)} \; \mathrm{d}\boldsymbol{\theta} \qquad (3.32),$$

which upon rearrangement yields:

$$\frac{1}{p(\mathbf{y} \mid M)} = \int \frac{\tau(\boldsymbol{\theta})}{f(\mathbf{y} \mid \boldsymbol{\theta}, M)\, p(\boldsymbol{\theta} \mid M)} \; p(\boldsymbol{\theta} \mid \mathbf{y}, M) \; \mathrm{d}\boldsymbol{\theta} \qquad (3.33),$$

from which follows the Gelfand-Dey estimator:

$$p_{GD}(\mathbf{y} \mid M) = \left\{ \frac{1}{ns} \sum_{i=1}^{ns} \frac{\tau(\boldsymbol{\theta}^{(i)})}{f(\mathbf{y} \mid \boldsymbol{\theta}^{(i)}, M)\ p(\boldsymbol{\theta}^{(i)} \mid M)} \right\}^{-1}, \boldsymbol{\theta}^{(i)} \leftarrow p(\boldsymbol{\theta} \mid \mathbf{y}, M) \qquad (3.34).$$

Kass and Raftery (1995) observe that provided the tails of $\tau(\boldsymbol{\theta})$ are sufficiently thin, this estimator is stable. This can be intuitively understood by noting that if $\tau(\boldsymbol{\theta}) = p(\boldsymbol{\theta} \mid M)$ (assuming $p(\boldsymbol{\theta} \mid M)$ is proper) the Gelfand-Dey estimator reduces to the Newton-Raftery estimator. The tails of $\tau(\boldsymbol{\theta})$ must be sufficiently thin to ensure that the occasional small likelihood does not exert excessive influence on the estimate. In this study the density chosen for $\tau(\boldsymbol{\theta})$ was the multivariate normal with mean and covariance determined by the posterior samples $\{\boldsymbol{\theta}^{(i)} : i = 1, ..., ns\}$. To help ensure that the tails were thin enough the sample covariance was scaled by factors ranging from 0.1 to 1.0. The marginal likelihood calculated did not vary significantly over this range, indicating that results were not affected by the tails, and therefore that the tails were sufficiently thin. For the results shown a scaling factor of 1.0 was used.

3.9.3 Other Methods

Chib (1995) and *Chib and Jeliazkov* (2001) give an alternate stable and consistent method for situations where Gibbs and Metropolis-Hastings MCMC sampling methods are used respectively. Some other methods combining approximations and simulation are surveyed within *DiCiccio et al.* (1997).

3.9.4 Posterior Bayes factor

The PBF can be simply estimated using:

$$p_{PBF}(\mathbf{y} \mid M) = \frac{1}{ns} \sum_{i=1}^{ns} f(\mathbf{y} \mid \boldsymbol{\theta}^{(i)}, M), \quad \boldsymbol{\theta}^{(i)} \leftarrow p(\boldsymbol{\theta} \mid \mathbf{y}, M) \qquad (3.35).$$

3.9.5 Prior Estimator

Given the instability of the Newton-Raftery estimator and the potential instability of the Gelfand-Dey estimator, we estimate the marginal likelihood directly as a check.

Provided that samples can be drawn from the prior, the prior estimator of the marginal likelihood, sometimes referred to as the golden method, is:

$$p_{prior}(\mathbf{y} \mid M) = \frac{1}{ns}\sum_{i=1}^{ns} f\left(\mathbf{y} \mid \boldsymbol{\theta}^{(i)}, M\right), \quad \boldsymbol{\theta}^{(i)} \leftarrow p(\boldsymbol{\theta} \mid M) \qquad (3.36).$$

Although this estimator is stable, a large number of samples may be required to obtain accuracy, with more samples required as more parameters are introduced or as the prior is made more diffuse.

3.10 Case Study: Modelling Persistence in Annual Rainfall

This case study applies BMS to rank three models of inter-annual persistence in annual rainfall. For each of these models the Metropolis algorithm is used to sample from the posterior distribution. Accordingly, the estimators described in Section 3.9 were used to estimate the marginal likelihood $p(\mathbf{y} \mid M)$. The objective is to critically evaluate competing model selection methods (in particular BMS). This provides the interpretive framework for Chapter 5 where different models of regional persistence are assessed.

3.10.1 Data

The case study data consisted of 137 years of continuous daily point rainfall, extending from January 1859 to April 1997 and recorded at the official Australian Bureau of Meteorology site, Observatory Hill Sydney. The daily data was aggregated up to the annual scale using a September-August water year. September was used as the start of the water year as this allowed best identification of the HMM parameters (*Thyer*, 2001). Of the 137 years of daily data, there were less than eight days flagged as being corrected values. This is considered to have insignificant effect on annual totals.

3.10.2 Competing Models

Several model selection techniques will be applied to discriminate between three models of annual rainfall.

Independent transformed normal model IND

The independent transformed normal model (IND) is the simplest model considered. It was selected as a candidate model because *a priori* it was expected that long-term

persistence in annual rainfall is likely to be weak. The IND model assumes the annual rainfall series $\mathbf{y} = (y_1,..,y_n)$ is independently and identically distributed. A transformed normal distribution is assumed to describe the marginal distribution. After transforming the annual rainfall y_t using the Box-Cox transformation:

$$z_t = \begin{cases} \dfrac{y_t^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log(y_t) & \lambda = 0 \end{cases} \qquad (3.37),$$

the transformed rainfall z_t is assumed to follow a truncated normal distribution. The parameter vector for the IND model is $\boldsymbol{\theta}' = (\mu_y, \sigma_y, \lambda)$ where μ_y and σ_y are the mean and standard deviation of the rainfall y.

Autoregressive lag one model AR1

The lag-one autoregressive model, a generalization of the IND model, has been the model of choice for use in annual hydrologic time series in Australia (*Grayson et al.*, 1996, *Srikanthan and McMahon*, 2001). After transforming the rainfall using (3.37) the model has the form:

$$z_t = \mu + \phi(z_{t-1} - \mu) + \varepsilon_t \qquad (3.38),$$

where μ is the mean of the transformed rainfall z, ϕ is the serial correlation and ε_t is an independent truncated normal random variable. The parameter vector is $\boldsymbol{\theta}' = (\mu_y, \sigma_y, \phi, \lambda)$. If the autocorrelation parameter ϕ is set to zero, the model reduces to the IND model. This is an example of a nested model as discussed in Section 3.6.

Two State HMM

As a discussion of the HMM was given in the previous chapter it will not be reiterated here. We are dealing in this case with the single site simplification of that model.

The two state HMM can degenerate to a two state Gaussian mixture model. To demonstrate this relationship, the parameterisation of the mixture model must be shown to have a functional relationship with the parameters of the HMM. A mixture model does not have any temporal Markovian dependency. That is, the probability of being in

a particular state does not alter from one time-step to the next, and does not depend on previous time-steps. Hence, for the two state HMM to act as a mixture model, the following relationship must hold:

$$p(r_t \mid r_{t-1}) = p(r_t) \tag{3.39}.$$

The right hand side of this equation is the marginal probability of being in a particular state. For the stationary HMM, this probability does not change in time, thus:

$$p(r_t) = p(r_{t-1}) \tag{3.40}.$$

The marginal probability of being in a dry state is given by total probability:

$$\begin{aligned} p(r_t = D) = {} & p(r_t = D \mid r_{t-1} = W)\, p(r_{t-1} = W) \\ & + p(r_t = D \mid r_{t-1} = D)\, p(r_{t-1} = D) \end{aligned} \tag{3.41}.$$

Using (3.40) and the fact that $p(r_t = D) + p(r_t = W) = 1$ yields the marginal probability:

$$p(r_t = D) = \frac{p(r_t = D \mid r_{t-1} = W)}{p(r_t = W \mid r_{t-1} = D) + p(r_t = D \mid r_{t-1} = W)} \tag{3.42},$$

with an analogous result for the wet state probability $p(r_t = W)$. Noting that (3.39) requires that $p(r_t = D) = p(r_t = D \mid r_{t-1} = W)$ it immediately follows that a two state HMM degenerates to a mixture model if:

$$p(r_t = W \mid r_{t-1} = D) + p(r_t = D \mid r_{t-1} = W) = 1 \tag{3.43}.$$

An identical result is obtained using $p(r_t = W)$. Thus, we have the simple relation that if the transition probabilities sum to one, the HMM degenerates to a mixture model.

3.10.3 Parameter Priors

Proper but disperse parameter priors were chosen so as to allow the data to dominate the posterior distribution. Nonetheless, it was ensured that the priors were proper in order to avoid the marginal likelihood becoming infinite. Table 3.2 presents a summary of the priors. The parameters defining the shape of the priors, referred to as hyperparameters, were varied to test the sensitivity of the results to prior specification.

Table 3.2 lists firstly the parameter of interest, then the model(s) that the parameter applies to. The prior distribution used for the parameter is also defined. Also listed in Table 3.2 are the lower and upper bounds for the prior distributions. In some cases parameter bounds and hyperparameter values were varied to check the sensitivity of the BMS results to prior specification. The range over which each bound or hyperparameter was varied is given where such sensitivity was tested.

Table 3.2 Parameter prior distributions and bounds

Parameter	Model(s)	Prior distribution	Lower bound	Upper bound	Hyper-parameters
μ_y	IND, AR1, HMM	$\mu_k \sim N\left(\mu_0, \sigma_y^2/\kappa\right)$	0	∞	$\mu_0 = \bar{y}$, $\kappa = 1$
σ_y	IND, AR1, HMM	$\sigma_y^2 \sim Inv-\chi^2\left(\nu_0, \sigma_0^2\right)$	0	∞	$\sigma_0^2 = \bar{s}^2$, $\nu_0 = 2$
λ	IND, AR1	$\lambda \sim N\left(\lambda_0, \sigma_\lambda^2\right)$	Max $(0,\lambda_o-2\sigma_\lambda)$	$\lambda_o+2\sigma_\lambda$	λ_o=0.5, σ_λ =[0.25,1.0]
ϕ	AR1	$\phi \sim N\left(\phi_0, \sigma_\phi^2\right)$	-1.0	1.0	ϕ_o=0.07, σ_ϕ =[0.15,0.5]
P_{ij}	HMM	$P_{ij} \sim Uniform$	[0.0,0.05]	[0.5,1.0]	not applicable

Each of the prior distributions are briefly discussed.

Mean μ_y and variance σ^2_y

The same priors for the mean μ_y and variance σ_y^2 were used for all models so as to not favour one model *a priori* over another. An empirical-Bayes approach was adopted to ensure that, in expectation, the priors were consistent with the gross characteristics of the data. Accordingly the sample mean $\bar{y}$ and sample variance $\bar{s}^2$ (see Section 3.6.1) were used to define the prior means. The prior degrees of freedom (κ and ν_0) were kept to a minimum to ensure that the priors remained diffuse. All means were truncated below zero, while for the IND and AR1 models there was no upper bound. The HMM wet mean had the dry mean as a lower bound, while the HMM dry mean had the wet mean as an upper bound. For a full explanation of this prior see *Thyer* (2001).

Box-Cox transform parameter λ

For the AR1 and IND models, the mean of the Box-Cox transformation parameter λ was set to 0.5. This represents the midpoint point between the normal and log-normal distributions. It was considered that most data sets would lie between these two extremes. The standard deviation σ_λ was sensitivity tested with values ranging from 0.25 to 1.00. The resulting normal distribution was truncated at both ends at two standard deviations from the mean λ_0.

Serial correlation ϕ

In a study of 40 high quality annual rainfall data sets spread throughout Australia by *Srikanthan et al.* (2001), the mean annual serial correlation was found to be 0.07. This was considered to be an appropriate value for the prior mean ϕ_o. Standard deviation of ϕ was varied from the relatively vague value of 0.5 to the moderately sharp value of 0.15. The ϕ parameter has bounds of $[-1,1]$.

HMM Transition Probabilities

A uniform prior was used for the transition probabilities. The least informative prior was a uniform distribution over the interval $[0,1]$. This prior was judged to be unreasonable in the context of this study because it allows the HMM to operate as a single-state model. For example, if a transition probability of 0.01 were assigned to a state then the average residence time in that state is 100 years. Given that our data is 140 years in length, this would likely result in one state dominating the entire series. Effectively the HMM has become an IND model without the benefit of the Box-Cox transformation. Because the HMM was designed to simulate long-term persistence, it is reasonable to *a priori* exclude transition probability combinations which produce single state behaviour. *Thyer* (2001) suggests state residence times could range from 2 years under the influence of the El Niño Southern Oscillation to several decades under the influence of the Inter-decadal Pacific Oscillation. Accordingly, the most informative prior on the transition probabilities was a uniform distribution over the interval $[0.05, 0.5]$ which corresponds to expected state residence times lying between 2 and 20 years.

3.10.4 Model Priors

All model priors were set equal in this study, and no variation was undertaken. For non-nested models, we believe that the interpretive scales given for the BF provide sufficient leniency for the subjectivity of prior model probability specification. In nested model comparison scenarios, the assumption of equal priors may not be as justifiable – why give equal weight to one parameter set versus an infinite selection of others?

3.10.5 Results

Because our primary focus is on formal BMS, results relating to Bayes Factors will be presented first. Comparisons with other model selection methods are then presented.

As many combinations of the prior specifications (summarized in Table 3.2) were tested, not all of the calculated Bayes Factors are presented. For ease of interpretation Table 3.1 only presents the maximum and minimum values.

Table 3.3 Maximum and minimum Bayes Factors and interpretation

Models	Max BF	Strength*	Min BF	Strength
IND/AR1	1.3	W - IND	1 / 4.5	P- AR1
IND/HMM	1.7	W- IND	1 / 3.4	P- HMM
AR1/HMM	5.1	P- AR1	1 / 3.0	W- HMM

*Note: W=weak, P=positive

The maximum BF can be interpreted as the maximum support in favour of the first model; while the minimum BF gives the maximum support for the second model. In all the cases tested, the maximum BF values occurred where the first model had a small prior range or variance and the second model had its largest tested range or variance.

No strong conclusions from the pairwise comparisons can be made, with the BF favouring the first model when it has sharp priors, and the second model when it has flatter priors. The IND model fairs the worst in that both the AR1 and HMM have positive strength minimum BF's, while having weak maximum BF's. However, for the comparison between the AR1 and the HMM, the maximum BF for the AR1 has positive strength, while the minimum BF is almost classed positive for HMM also. This

demonstrates the effects that prior specification can have on the conclusions made, especially in cases where large samples are not available.

Overall, it was concluded that both the AR1 and HMM are superior to the IND model. However, it could not be concluded which of the models, AR1 or HMM, was superior. The relatively strong result in favour of AR1 for the sharp AR1 prior versus the broad range prior on HMM is explainable in that the HMM is quite complex compared to the AR1. Many HMM parameter combinations represent situations that could not be identified given the amount of data, and therefore parameter bounds were tightened to cut out these situations (why model situations that are very unlikely to occur). Also, as a uniform prior distribution was used for the HMM parameters, and sharper Gaussian priors were used for the AR1, the priors tend to favour AR1 slightly *a priori*.

Table 3.4 shows a comparison of the three Bayes Factor estimation methods used. The Gelfand-Dey (GD) estimates agreed with values calculated using the prior estimator (the golden method). The Newton-Raftery (NR) estimate of the BF was not as accurate as the Gelfand-Dey method. Although not shown here, the Gelfand-Dey estimates of the marginal likelihood were consistent from one simulation to the next. As such the Gelfand-Dey estimates were used for Table 3.3 and subsequent comparison to other model choice methods. The Newton-Raftery method was not used as it was found to have a high degree of variability from simulation to simulation.

Table 3.4 BF Calculation Methods

	Bayes Factor					
Models	**Max**			**Min**		
	Prior	**GD**	**NR**	**Prior**	**GD**	**NR**
IND/AR1	1.4	1.3	1.0	1/4.9	1/4.5	1/8.8
IND/HMM	1.7	1.7	1.8	1/3.4	1/3.4	1/ 11.5
AR1/HMM	5.0	5.1	5.1	1/ 3.0	1/ 3.0	1/ 3.6

The PBF, AIC and SC were calculated for the three models and are compared to the BF in Table 3.5. It should be noted that the AIC was not designed to be judged using Jeffrey's interpretive scale like the BF and SC. Rather, a ratio greater than one indicates

that the first model is preferred. Likewise the PBF cannot strictly speaking be interpreted using Jeffrey's scale.

Table 3.5 Related Model Selection Methods

Models	PBF	AIC*	SC*	BF	
				Max	Min
IND / AR1	1/3.3	1/1.7	2.6	1.3	1/4.5
IND / HMM	1/8.3	1/1.1	70.1	1.7	1/3.4
AR1/ HMM	1/2.6	1.5	27.1	5.1	1/3.0

*Note: The AIC and SC as defined in equations (3.23) and (3.25) have been exponentiated to give values in comparable terms to the BF.

Changes in specification of the prior did not affect the penalized likelihood measures as the maximum likelihood estimate was within the parameter bounds in all cases. The PBF was also unaffected, presumably due to the majority of the posterior distribution also lying within these bounds.

The PBF distinctly favours the HMM followed by AR1 and then the IND model. The AIC ranks the models as AR1, HMM and then IND. The SC strongly favours the IND, AR1 and then HMM. Compare this to the BF which put the AR1 and HMM on about level terms, with IND lagging behind. Which one of these methods is the correct one and what conclusions, if any, can we draw from such a range of results?

As an additional note, in the case of a nested model, the posterior distribution of parameters can be used for model selection as mentioned in Section 3.7.4. Figure 3.1 shows a histogram of the posterior distribution for the ϕ parameter of the AR1 model. It can be easily seen from this plot that it is unlikely that the data came from the IND model (ϕ=0), as only 3% of the posterior probability is less than zero. Calculation of the posterior distribution does not require model prior specification. Hence, the assumption we have used regarding BF, that each model has equal prior probability, is not needed. This method can also be applied to the HMM in determining whether the HMM hypothesis is justified compared to a special case of the HMM, the two state mixture. As was discussed in Section 3.10.2, the HMM degenerates to a simple mixture model if the sum of the transition probabilities are 1. The posterior distribution of the transition

probabilities along with the transition probability sums is shown in Figure 3.2, with the solid line indicating the line along which the transition probabilities sum to one. The majority of the posterior cloud lies well away from the mixture line, justifying the Markovian assumption.

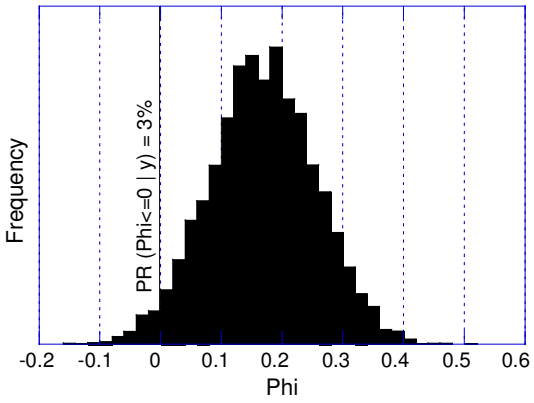


Figure 3.1 Posterior distribution of Φ for the AR1 model

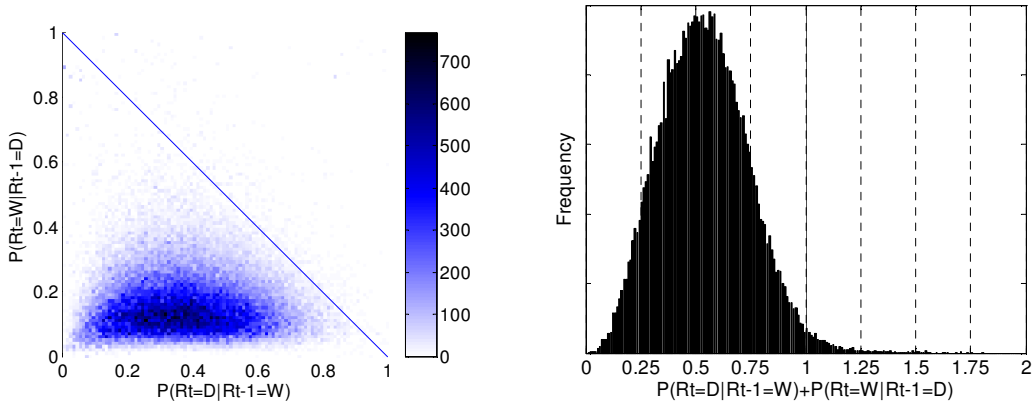


Figure 3.2 Posterior transition probability (a) distribution and (b) sum histogram for the HMM.

3.10.6 Discussion

In a Bayesian framework the Bayes factor is the accepted formal model selection method. However, the BF can produce a range of results depending on the prior used. In this case study it is not clear whether the AR1 or the HMM is the superior model. This result is presumably influenced by the lack of data, with more data required to unequivocally identify the better model (whatever the selection method).

The Gelfand-Dey estimator of the BF was found to be more accurate than the Newton-Raftery estimator and consistently repeatable. Although slightly more complex to calculate than the Newton-Raftery estimator, the accuracy of the Gelfand-Dey estimator justifies the effort.

A variety of conclusions could be made from the comparison of the model selection measures. The PBF favoured the more complex model, but can not be justified theoretically from a Bayesian viewpoint. The AIC produces results very similar to the BF, but, again from a Bayesian perspective, this method cannot be justified because it disregards parameter uncertainty.

The SC is a large sample approximation to the BF, but as demonstrated can suffer from gross errors. These errors are explained by (3.24) which reveals the asymptotic approximation which underpins the SC has errors of constant order in the log marginal likelihood. It is speculated that significant errors in the SC have arisen from use of limited data and comparison of two structurally very different models, AR1 and HMM.

As Bayes factors are fundamental to BMS, our conclusions are based on the BF results in Table 3.3. The AR1 and the HMM are considered to model the data equally well.

In nested-model comparisons, tests involving the posterior distribution of parameters appear to offer a more rational approach. Use of the posterior distribution avoids the need to assign equal prior probabilities to $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and the infinite set $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. However, in cases where many (nested and non-nested) models are compared, comparing posterior distributions of nested versus non-nested variants is not possible. Providing care is taken in elicitation of priors, and resulting inference, BMS offers a simple method of ranking models according to posterior model probability.

3.10.7 Conclusion

The Bayes factor represents the formal approach to BMS and is compared to related less justifiable approaches. Some of the typical perils of BMS such as Lindley's paradox are identified, and simple steps to avoid them are described. The case study demonstrates that the BF is sensitive to prior parameter specification. This sensitivity introduces an

indeterminacy in BMS so that only models with distinctly superior performance would be unequivocally identified as the superior model.

The Newton-Raftery estimate of the BF is found to be less reliable than both the Gelfand-Dey estimate and the golden method. The Schwarz criterion, although consistent, was found to be a poor approximation to the BF in this case study where only a small number of data samples were available and distinctly different models compared. This highlights the danger of relying on asymptotic approximations. In the remainder of the thesis the BF will be calculated using the Gelfand-Dey estimator rather than the simpler but less reliable SC. Of course, the stability and accuracy of the various estimates will be problem dependent, for example problems are possible with the Gelfand-Dey estimator with multimodal, skewed and thin tailed posteriors. However, the case study serves to provide a practical test of the estimators available for problems of moderate dimension where approximate normality around exists.

The specification of model priors, in nested model testing, is identified as a possible difficulty associated with BF. The often assumed equal model priors for the nested model and its generalization can rightly be questioned. A posterior plot of the nested parameter can provide simple and clear evidence for model choice. In the case study, the posterior parameter plots showed that it was very likely that the AR1 model was preferable to the IND model, and also likely that the Markovian assumption of the HMM was justified compared to a non-Markovian mixture model.

The AR1 and HMMs were found to be approximately on equal footing for the Sydney annual rainfall data. This result concurs with the results of *Thyer* (2001), who was unable to choose between the two using other informal methods such as posterior predictive checking of test repetitions and parameter identification.

3.11 Bayesian Model Averaging

So far in this chapter, model selection has been emphasised in a Bayesian modelling framework - choosing one model from a set $\{M_1,...,M_{nm}\}$ of candidate models where *nm* are the number of models. A natural extension to Bayesian model selection is model averaging. Model averaging refers to the process of estimating some quantity under each model M_j and then averaging the estimates according to how likely each model is

(*Wasserman*, 2000). This is the more general case of estimating a posterior quantity of interest as discussed in 3.3.1, with not only parameter uncertainty incorporated, but also model uncertainty:

$$\begin{aligned} \mathrm{E}\{\mathrm{G}(\boldsymbol{\theta},M)\} &= \sum_{j=1}^{nm}\int \mathrm{G}(\boldsymbol{\theta},M_j)\,p(\boldsymbol{\theta},M_j \mid \mathbf{y})\,d\boldsymbol{\theta} \\ &= \sum_{j=1}^{nm}\left(\int \mathrm{G}(\boldsymbol{\theta},M_j)\,p(\boldsymbol{\theta} \mid \mathbf{y},M_j)\,d\boldsymbol{\theta}\right)p(M_j \mid \mathbf{y}) \end{aligned} \quad (3.44),$$

where $\mathrm{G}(\boldsymbol{\theta},M)$ is the quantity of interest. For Bayesian model averaging (BMA), we are no longer choosing the 'best model', but weighting the predictions depending on the posterior model probabilities. This is the strength of model averaging- there is no need to rely on vague rules such as Jeffrey's Bayes factor interpretive scale (Table 3.1) to accept or reject a particular model.

As there were many model variants tested in the following chapters, Bayesian model averaging rather than BMS was chosen for presentation of results. It is believed that this method is more justifiable, in that model uncertainty is also accounted for within prediction. Somewhat analogously to the argument for incorporating parameter uncertainty, posterior model probability fully accounts for uncertainty, whereas methods choosing a single model may produce over confident predictions.

A quantity of particular interest in this study is the hidden state series defined in Section 3.4.2. BMA is used in the following chapters to produce the state series upon which the small time-scale rainfall model DRIP will be conditioned.

3.12 Model-Parameter Product Space Sampling Techniques

In the applications to come in later chapters, model choice/averaging is to be undertaken between many models. An MCMC technique that is more efficient than sampling from individual models, one at a time to produce posterior quantities of interest, such as model weight, would be advantageous. Model-Parameter product space samplers provide an alternative to sampling from individual models, and can provide a more efficient method of evaluating posterior model weight than evaluating marginal likelihoods individually. MCMC Model-Parameter product space samplers describe a sampling technique where not only the parameter set is a sampled, but the model is

sampled also. Of the product space searches available, the Metropolised Carlin-Chib (MCC) sampler (*Godsill*, 1998, *Dellaportas et al.*, 2002) was trialled due to it being a simple generalisation of the Metropolis-Hastings algorithm. Therefore, there was minimal coding involved moving from the MH sampler to MCC. Although, not used in the final production of results due to implementation difficulties, it deserved discussion as a method of MCMC model sampling. Its relationship to the MH algorithm is now discussed.

Model-Parameter space product samplers such as the reversible jump (RJMCMC) techniques of *Green* (1995) or the MCC sampler could have been applied to the test problem in the previous section. Because the product space methods sample from model to model as well as within parameter space, it is easier in cases involving a limited number of models to implement individual estimates of the marginal likelihood rather than one mega-simulation as detailed in *Han and Carlin* (2000).

3.12.1 Metropolised Carlin-Chib sampler

A model-parameter product space MH sampler was proposed (independently) in *Dellaportas et al.* (2002), *Godsill* (1998) and *Besag* (1997). These samplers present essentially the same idea - the discussion and derivation below follows *Godsill* (1998). To see how this algorithm works let us first define the support over which the sampler operates. Consider the 'pool' of parameters $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_{M_1},...,\boldsymbol{\theta}_{M_{nm}}\right)$ such that $\boldsymbol{\theta}_M \in \Theta_M$ is the parameter space pertaining to model M with $M \in \mathbf{M}$, where $\mathbf{M} = \left(M_1,...,M_{nm}\right)$ and *nm* are the number of models. The support for the overall product space $(M,\boldsymbol{\theta}) \in \Omega$ is given by:

$$\Omega \;=\; \mathbf{M} \times \prod_{M \in \mathbf{M}} \Theta_M \qquad (3.45).$$

The posterior distribution for the full composite space (with the parameters used in model M, signified by $\boldsymbol{\theta}_M$, and parameters not used in model M signified by $\boldsymbol{\theta}_{-M}$) is expressed as:

$$\begin{aligned} p(M,\boldsymbol{\theta}\,|\,\mathbf{y}) &= p(M,\boldsymbol{\theta}_M,\boldsymbol{\theta}_{-M}\,|\,\mathbf{y}) \\ &= \frac{p(\mathbf{y}\,|\,\boldsymbol{\theta}_M,\boldsymbol{\theta}_{-M},M)\,p(\boldsymbol{\theta}_M,\boldsymbol{\theta}_{-M},M)}{p(\mathbf{y})} \\ &= \frac{p(\mathbf{y}\,|\,\boldsymbol{\theta}_M,M)\quad p(\boldsymbol{\theta}_M,\boldsymbol{\theta}_{-M},M)}{p(\mathbf{y})} \\ &= \frac{p(\mathbf{y}\,|\,\boldsymbol{\theta}_M,M)\,p(\boldsymbol{\theta}_M\,|\,M)\,p(\boldsymbol{\theta}_{-M}\,|\,\boldsymbol{\theta}_M,M)\,p(M)}{p(\mathbf{y})} \end{aligned} \qquad (3.46).$$

As we have now defined a full composite space over which the Metropolis-Hastings sampler can operate (the dimension of Ω does not change), we use an acceptance ratio of the same form as (3.6), replacing the posterior identity with (3.46) to give:

$$r = \frac{p(\mathbf{y}\,|\,\boldsymbol{\theta}^*_{M^*},M^*)\,p(\boldsymbol{\theta}^*_{M^*}\,|\,M^*)\,p(\boldsymbol{\theta}^*_{-M^*}\,|\,\boldsymbol{\theta}^*_{M^*},M^*)\,p(M^*)}{p(\mathbf{y}\,|\,\boldsymbol{\theta}_M,M)\,p(\boldsymbol{\theta}_M\,|\,M)\,p(\boldsymbol{\theta}_{-M}\,|\,\boldsymbol{\theta}_M,M)\,p(M)} \times \frac{J(\boldsymbol{\theta},M\,|\boldsymbol{\theta}^*,M^*)}{J(\boldsymbol{\theta}^*,M^*|\boldsymbol{\theta},M)} \qquad (3.47).$$

Now, defining the product space jump distribution from $(M,\boldsymbol{\theta})$ to $(M^*,\boldsymbol{\theta}^*)$ as:

$$\begin{aligned} J(M^*,\boldsymbol{\theta}^*\,|\,M,\boldsymbol{\theta}) &= J(\boldsymbol{\theta}^*_{-M^*}\,|\,\boldsymbol{\theta}^*_{M^*},M^*,M,\boldsymbol{\theta})\,J(\boldsymbol{\theta}^*_{M^*}\,|\,M^*,M,\boldsymbol{\theta})\,J(M^*\,|\,M,\boldsymbol{\theta}) \\ &= J(\boldsymbol{\theta}^*_{-M^*}\,|\,\boldsymbol{\theta}^*_{M^*},M^*)\quad J(\boldsymbol{\theta}^*_{M^*}\,|\,\boldsymbol{\theta}_M,\boldsymbol{\theta}_{-M})\quad J(M^*\,|\,M) \\ &= J(\boldsymbol{\theta}^*_{-M^*}\,|\,\boldsymbol{\theta}^*_{M^*},M^*)\quad J(\boldsymbol{\theta}^*_{M^*}\,|\,\boldsymbol{\theta}_M)\quad J(M^*\,|\,M) \end{aligned} \qquad (3.48).$$

Such simplifications are possible as we are free here to choose any jump distribution we desire that has coverage over $(M,\boldsymbol{\theta})$; that is, any function which has positive probability of jumping to all points in Ω. Inserting this identity into (3.47), and letting $J(\boldsymbol{\theta}_{-M}\,|\,\boldsymbol{\theta}_M,M) = p(\boldsymbol{\theta}_{-M}\,|\,\boldsymbol{\theta}_M,M)$ for any $(M,\boldsymbol{\theta})$ gives:

$$r = \frac{p(\mathbf{y}\,|\,\boldsymbol{\theta}^*_{M^*},M^*)\,p(\boldsymbol{\theta}^*_{M^*}\,|\,M^*)\,p(M^*)}{p(\mathbf{y}\,|\,\boldsymbol{\theta}_M,M)\,p(\boldsymbol{\theta}_M\,|\,M)\,p(M)} \times \frac{J(M\,|M^*)\,J(\boldsymbol{\theta}_M\,|\boldsymbol{\theta}^*_{M^*})}{J(M^*|M)\,J(\boldsymbol{\theta}^*_{M^*}|\boldsymbol{\theta}_M)} \qquad (3.49).$$

Using this acceptance ratio in the following algorithm yields the Metropolised Carlin-Chib sampler:

1. Draw a starting model $M^{(0)}$, and an associated parameter vector $\boldsymbol{\theta}^{(0)}_M$ that has a positive posterior probability.

2. For i=1,2,…:

a) Sample candidate model M^* from model jump distribution $J\left(M^* \mid M^{(i-1)}\right)$

b) Conditional on the proposed model M^*, sample a candidate point $\boldsymbol{\theta}^*_{M^*}$ from a jumping distribution $J\left(\boldsymbol{\theta}^*_{M^*} \mid \boldsymbol{\theta}^{(i-1)}_{M}\right)$.

c) Calculate the ratio of densities according to (3.49).

e) Set

$$\left(\boldsymbol{\theta}^{(i)}, M^{(i)}\right) = \begin{cases} \left(\boldsymbol{\theta}^*, M^*\right) & \text{with probability } \min\left(r,1\right) \\ \left(\boldsymbol{\theta}^{(i-1)}, M^{(i-1)}\right) & \text{otherwise.} \end{cases} \qquad (3.50).$$

f) Check convergence – If sufficient samples taken, stop. Otherwise continue.

Note here that we now have a conditional proposal step: first we propose a new model, and then dependent on this we propose a parameter set. However, the acceptance ratio takes into account both jumps. This is an example of a Gibbs within Metropolis sampler (the model parameters are proposed conditional on the proposed model).

Notice also in the jump distributions and the acceptance ratio that we are only interested in parameters associated with the current model $\boldsymbol{\theta}^{(i-1)}_{M}$, and the proposed model $\boldsymbol{\theta}^*_{M^*}$. The complementing parameters in the 'pool' $\boldsymbol{\theta}^{(i-1)}_{-M}$ and $\boldsymbol{\theta}^*_{-M^*}$ are not generated or stored. Thus $\left(\boldsymbol{\theta}^{(i)}, M^{(i)}\right)$ is effectively equal to $\left(\boldsymbol{\theta}^{(i)}_{M}, M^{(i)}\right)$.

3.12.2 Relationship between MCC and other model space samplers

The derivation given here is a simplification of that in *Godsill* (1998). In that paper, parameters were allowed to be shared between models. This particular algorithm was coined the Metropolised Carlin-Chib sampler due to it being the general (Metropolised) case of a model-parameter sampler employed by *Carlin and Chib* (1995). The Carlin-Chib sampler uses Gibbs steps to generate the model parameters $\boldsymbol{\theta}_M$ and M separately. Unfortunately, breaking the sampling into full conditional steps means that we also need to generate parameters not used in the proposed model $\boldsymbol{\theta}^*_{-M^*}$. These parameters are generated from the pseudoprior $J\left(\boldsymbol{\theta}_{-M} \mid \boldsymbol{\theta}_M, M\right)$, which now requires specification. This term cancelled out of the MCC algorithm. However, in the Carlin-Chib sampler we

must generate both the $\boldsymbol{\theta}_M$ and $\boldsymbol{\theta}_{-M}$ parameter vectors at every step. In cases where many models are being sampled, this specification and generation becomes impractical (*Godsill*, 1998, *Han and Carlin*, 2000).

The reversible jump sampler introduced by *Green* (1995) is a special case of the product space sampler where there are deterministic moves proposed between parameter spaces. *Dellaportas et al.* (2002) note that this flexibility, allowing 'the model parameters of the proposed model to depend on the current model in a totally general way', is a great strength of reversible jump. However, this generality can also be a burden, with the jump functions requiring specification. The MCC method appeared simpler to implement and apply, especially considering minimal coding changes are required from sampling using MH.

3.12.3 Specification of jump distributions

In the studies undertaken in this thesis the model jump distribution $J(M^* \mid M)$ was set uniform:

$$\begin{aligned} J\left(M^* \mid M\right) &= J(M^*) \\ &= \frac{1}{nm} \end{aligned} \tag{3.51},$$

while the parameter jump vector $J(\boldsymbol{\theta}^*_{M^*} \mid \boldsymbol{\theta}_M)$ was made independent of the previous models parameter vector $\boldsymbol{\theta}_M$ according to:

$$\begin{aligned} J\left(\boldsymbol{\theta}^*_{M^*} \mid \boldsymbol{\theta}_M\right) &= J\left(\boldsymbol{\theta}^*_{M^*}\right) \\ &\sim N\left(\hat{\boldsymbol{\theta}}_{M^*}, c^2\Sigma_{M^*}\right) \end{aligned} \tag{3.52},$$

with a normal distribution located at the modal parameter set for model M^* (found previous to sampling). The covariance Σ_{M^*} is (as before) an estimate taken from using an inversion of the Hessian taken at the modal parameter set. A scaling parameter c is used to expand the covariance to ensure the estimated covariance has larger scale than the posterior distribution. We use a scaling value of $c = \sqrt{1.5}$.

The jump distributions used here are equivalent to an independence sampler in model space. *Dellaportas et al.* (2002) state that using such an approach works best when

$J\left(M^*\right)$ is a reasonable approximation to $p\left(M^* \mid \mathbf{y}\right)$, and 'perhaps more importantly' when $J\left(\boldsymbol{\theta}^*_{M^*}\right)$ is a reasonable approximation to $p\left(\boldsymbol{\theta}^*_{M^*} \mid \mathbf{y}, M^*\right)$ for every M^*. In other words, the algorithm above will work well if the scaled normal estimate of the parameter distribution is a reasonable estimate of the posterior distribution.

The approach above is viable when the computational time required to calculate $\hat{\boldsymbol{\theta}}_M$ and Σ_M for all models is not prohibitive, and the resulting approximation of the covariance is approximately correct. If the approximation of the covariance is poor, given that there is no adaptation of covariance according to previous samples, the sampler will be inefficient. In high dimensional constrained parameter problems such as that encountered in this thesis, finding a good approximation to the posterior covariance is often difficult. The method of Hessian evaluation used in this thesis relied on finite differences to evaluate individual components of gradient. Given all the individually evaluated Hessian components, a Choleski decomposition was then used for inversion. Often the Hessian was found to be non-positive definite/ill conditioned, thus not permitting inversion. More advanced covariance estimation techniques may provide closer approximation to the posterior covariance (provided the posterior is approximately Gaussian). As the sampler was found to be quite inefficient due to these poor covariance estimates compared to individual model MH sampling, MH sampling was used for the results in following chapters. Apparently, the adaptation of the MH sampler covariance provides the main advantage, allowing the jump distribution to better approximate the posterior surface as sampling continues.

Considerable refinement of the algorithm used here is possible, especially regarding covariance estimation. Also, refinement of the model to model jump probability could be investigated to promote efficiency. A possible avenue of work would be generalising the parameter jump distribution adaptation method of *Haario et al.* (2001) so as to also include a model indicator. As the *Haario et al.* (2001) sampler is ergodic, it could be applied with confidence, rather than using *ad hoc* adaptation rules which may or may not be sampling from the posterior. This method would therefore adapt the model jump probability in tandem with the parameter jump probability. With such adaptation, it is

expected that much more efficient samplers would result. This subject is left for future work.

3.12.4 Posterior model probabilities from product space samplers

To estimate the posterior probability of model M, $p(M \mid \mathbf{y})$, using posterior samples from $(M, \boldsymbol{\theta})$ provided by this acceptance ratio the following integral needs to be evaluated:

$$p(M \mid y) = \int_{\Omega} p(\boldsymbol{\theta}, M \mid y) d\Omega \qquad (3.53).$$

Using numerical integration we have:

$$p(M_j \mid y) = \frac{1}{ns} \sum_{i=1}^{ns} I\left[M^{(i)} = M_j\right] \quad , \quad \Omega^{(i)} \leftarrow p(\boldsymbol{\theta}, M \mid y) \qquad (3.54),$$

where $I\left[M^{(i)} = M_j\right]$ is an indicator function with value 1 when $\Omega^{(i)}$ sampled model M_j, and value 0 otherwise.

3.12.5 Discussion of Model-Parameter product space samplers

In testing not presented in this thesis, the MCC sampler described in this section performed less efficiently than sampling from within individual models. Therefore the individual model sampler (using the Gelfand-Dey estimate of marginal likelihood) was applied in the case studies of Chapter 5. Although not used in producing the results of this thesis, the MCC sampler with refinement could provide a more efficient method of evaluating posterior model probabilities than sampling from individual models. Alternatively, Reversible Jump may be more efficient than the MCC, due to the allowance for specification of jumping rules between parameter spaces. The final word is left to *Han and Carlin* (2000) who caution that all of the methods of calculating marginal likelihoods that they discuss (including RJMCMC and MCC) 'require substantial time and effort (both human and computer) for a rather modest payoff, namely a collection of posterior model probability estimates, possibly augmented with associated standard error estimates'.

3.13 Conclusion

This chapter outlined the Bayesian modelling calibration framework used in this study. An MCMC sampling technique, the random walk (adaptive) MH sampler was introduced and compared to the Gibbs sampler, the model calibration technique used in the study of *Thyer* (2001). The MH sampler was chosen in this study due to its comparative simplicity. Use of the MH sampler obviated the need to simulate the hidden state series as part of the sampling process, thus reducing the chance of occurrence of 'trapping state' encountered for the Gibbs sampler for HMM problems. This simplification arose because the MH sampler could employ the Baum-Welch likelihood formulation for the HMM which integrated out the hidden state.

The closely related topic of Bayesian Model Selection was introduced and discussed in detail. A test study demonstrated some of the perils involved in BMS, along with giving comparisons to other widely used model selection methods. Significantly, in cases where there is little data or highly dissimilar models, the Schwarz Criterion (or BIC) can give a poor approximation to the Bayes Factor. The method of Bayes factor estimation found to be most accurate was the Gelfand-Dey estimator.

A mega-model (model-parameter product space) sampling technique, the Metropolised Carlin-Chib algorithm was also discussed for use in situations where there are many models to choose from. However, the MCC applied was found to be less efficient than using the MH sampler on individual models. Thus, the MH and the associated Gelfand-Dey estimate of the marginal likelihood were chosen for model selection/averaging in the remainder of this thesis.

Bayesian model averaging, the extension of Bayesian principles to model space, was discussed. This method is preferred to model selection as it removes the need to select particular models according to some (usually *ad hoc*) criteria. Methods of generating quantities of interest, such as the HMM hidden state probability series, were presented in terms of generating from one model, or averaging over all models tested. This model averaging technique is to be used in Chapter 5 where it is applied to the HMM and its generalisations to be described in the next chapter.

Chapter 4 Extensions to the HMM : The Switch and Regional HMM

4.1 Introduction

The HMM assumes a common (or regional) climate state occurs simultaneously across all sites used in an analysis. Can this assumption be justified for an arbitrarily chosen set of sites? Local meteorological anomalies can mask the influence of large scale climate phenomena. Also, as sites become further apart it is less likely they are affected by the same climate controls. The current HMM methodology advanced by *Thyer* (2001) cannot assess the soundness of the regional climate state assumption satisfactorily. Two generalisations of the HMM are proposed to specifically address this assumption: the Switch HMM and the Regional HMM.

A Switch HMM is introduced as a new approach to modelling the effects of regional and local interactions. As before, the regional climate state follows a Markov process thereby providing the means for simulating long-term persistence. However, at individual sites within the region local meteorological anomalies may affect the hidden regional state. Therefore, a second layer is added to the HMM to simulate at-site anomalies from year to year, thus giving each site the opportunity to 'switch' from the overall regional climate state.

The Regional HMM, rather than allowing at-site anomalies to occur, relaxes the assumption of a single controlling regional climate state by allowing multiple climate regions within the study region. Each of these climate regions has individual climate state series with an associated set of transition probabilities. The primary challenge is to identify the most appropriate partitioning of regions.

This chapter describes each of these generalisations in detail. Testing which of the two generalisations is more applicable for a set of sites is a problem of model selection. As such, it will be left for the next chapter where the results of the application of the models are compared. However, the ways in which the models differ should be kept in mind to allow interpretation of the results presented in the next chapter.

In addition to the introduction of the HMM generalisations, the important issue of correlation structure is discussed. A distinction is made between small- and large-scale correlation. Correlation is interpreted in terms of the modelling the dependence of rainfall between sites. Several parameterisations of the correlation matrix are put forward including zero correlations, empirical correlations, exponentially decaying functional correlations and fitted correlations.

4.2 Switch HMM

The Switch HMM (SHMM) is a generalisation of the multiple-site HMM and conceptualises annual rainfall as being controlled by a regional climate state with each site being allowed to 'escape' from this overall control. This formulation was born from the observation that not all sites are affected to the same degree by large scale climate controls such as El Niño with single sites sometimes showing totally opposite or anomalous responses when compared to surrounding sites.

The SHMM structure is illustrated in Figure 4.1. The Switch HMM differs from the HMM due to the inclusion of the 'switch' probability parameters $p(s_t^m = W \,|\, r_t = D)$ and $p(s_t^m = D \,|\, r_t = W)$ for each site. As before, r_t signifies the regional state at time t with s_t^m corresponding to the state at site m. D and W represent dry and wet states respectively. The switch probabilities determine how closely the site state series follows the regional state series.

4.2.1 Calculation of Switch HMM Likelihood

Although the SHMM likelihood is similar to the ordinary HMM likelihood, derived in Section 3.4.1, a full derivation will be given here for clarity. As before, the overall likelihood is calculated using:

$$p\left(\mathbf{Y}_1^T \,|\, \boldsymbol{\theta}\right) = p\left(\mathbf{y}_1 \,|\, \boldsymbol{\theta}\right) \prod_{t=2}^{T} p\left(\mathbf{y}_t \,|\, \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) \qquad (4.1).$$

Note that the parameter vector used here is $\boldsymbol{\theta} = \left(\boldsymbol{\mu}_{\mathbf{W}}, \boldsymbol{\sigma}_{\mathbf{W}}, \boldsymbol{\mu}_{\mathbf{D}}, \boldsymbol{\sigma}_{\mathbf{D}}, \boldsymbol{\rho}, \mathbf{P}, \mathbf{SP}\right)$ with wet and dry mean and variance parameters for every site $\left(\boldsymbol{\mu}_{\mathbf{W}}, \boldsymbol{\sigma}_{\mathbf{W}}, \boldsymbol{\mu}_{\mathbf{D}}, \boldsymbol{\sigma}_{\mathbf{D}}\right)$, a correlation coefficient matrix $\boldsymbol{\rho} = \left[\rho_{ij}\right], \quad i, j = 1, ..., d$ that is independent of state, a single set of

regional transition probabilities $\mathbf{P} = \left[\ p_{ij}\ \right] = p\left(r_t = j \mid r_{t-1} = i\right),\ i, j = W, D$, and a set of switch probabilities $\mathbf{SP} = \left[\ sp_{ij}^{site}\ \right] = p\left(s_t^{site} = j \mid r_t = i\right),\ i, j = W, D,\ site = 1,...,d$.

The evaluation of a typical term $p(r_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta})$ is now considered. The first step is to calculate the regional state probability according to:

$$p(r_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) = \sum_{r_{t-1}} p(r_t \mid r_{t-1}, \boldsymbol{\theta}) p(r_{t-1} \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) \tag{4.2}.$$

Given this regional probability, we can now calculate the probability of the individual sites being in a particular state s_t^m according to:

$$p(s_t^m \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) = \sum_{r_t} p(s_t^m \mid r_t, \boldsymbol{\theta}) p(r_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}) \tag{4.3},$$

where $\sum_{s_t^m} p(s_t^m \mid r_t, \boldsymbol{\theta}) = 1$. The probability of being in a particular state can vary from site to site. The larger the switch probability $p(s_t^m \neq r_t \mid r_t, \boldsymbol{\theta})$, the less likely the site state series will correspond to the regional state series. The Switch HMM degenerates to the HMM when the switch probabilities are zero.

The SHMM structure produces conditionally independent states at each site. This independence is used to calculate the site permutation probability:

$$p\left(\mathbf{s}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) = \prod_{site=1}^{d} p\left(s_t^{site} \in \mathbf{s}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) \tag{4.4},$$

where $\mathbf{s}_t = \left(s_t^1, s_t^2, ..., s_t^d\right)$ and d is the total number of sites. The probability of the site state permutation $\mathbf{s}_t$, can be used to weight the Gaussian rainfall distribution as follows:

$$\begin{aligned} p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) &= \sum_{\mathbf{s}_t} p\left(\mathbf{y}_t \mid \mathbf{s}_t, \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) p\left(\mathbf{s}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) \\ &= \sum_{\mathbf{s}_t} p\left(\mathbf{y}_t \mid \mathbf{s}_t, \boldsymbol{\theta}\right) p\left(\mathbf{s}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) \end{aligned} \tag{4.5}.$$

Parameters

Regional Transition Probs

$$\mathbf{P} = \begin{bmatrix} 1-p_{DW} & p_{DW} \\ p_{WD} & 1-p_{WD} \end{bmatrix}$$

Given previous state r_{t-1}, *generate regional state* $\rightarrow r_t$

Site Switch Probs

$$\mathbf{SP}^{site} = \begin{bmatrix} 1-sp_{DW}^{site} & sp_{DW}^{site} \\ sp_{WD}^{site} & 1-sp_{WD}^{site} \end{bmatrix},$$

Given regional state r_t, *generate site state*$\rightarrow s_t^{site}$

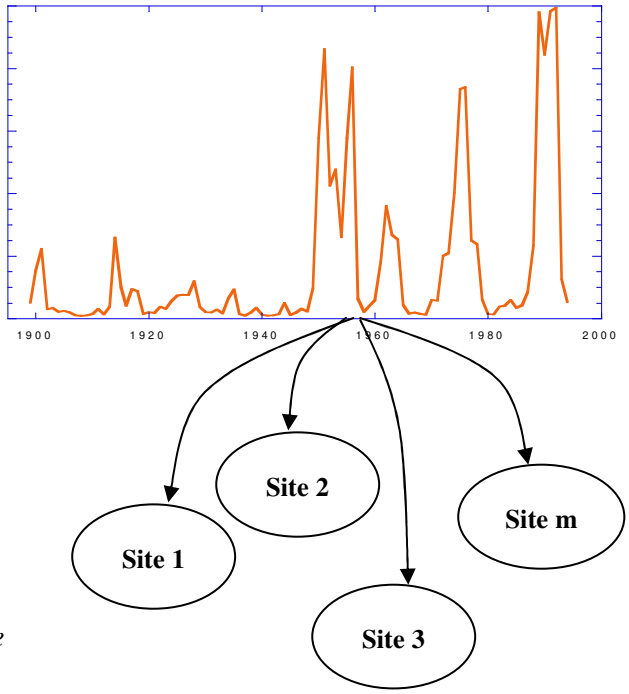
Means and Variance of Rainfall

$$\begin{bmatrix} \mu_D^{site} \\ \mu_W^{site} \end{bmatrix} \quad \begin{bmatrix} \sigma_D^{site} \\ \sigma_W^{site} \end{bmatrix}$$

Coefficient of correlation between sites

$$\begin{bmatrix} 1 & & & \\ \rho_{2,1} & 1 & & \\ \vdots & \cdots & \ddots & \\ \rho_{d-1,1} & \rho_{d-1,2} & \vdots & 1 \\ \rho_{d,1} & \rho_{d,2} & \cdots & \rho_{d,d-1} \quad 1 \end{bmatrix}$$

Regional State Series



Given site state set permutation $\mathbf{s}_t = \left(s_t^1, s_t^2, ..., s_t^d\right)$ *sample* $\mathbf{y}_t = \left(y_t^1, \cdots, y_t^d\right)$ *from multinormal distribution* $\mathbf{y}_t \sim N\left(\boldsymbol{\mu}_{\mathbf{s}_t}, \boldsymbol{\Sigma}_{\mathbf{s}_t}\right)$ *with mean and variances inserted according to* S_t.

Figure 4.1 Switch Hidden State Markov Model Conceptual Diagram

In simple terms, the wet state rainfall distribution is most likely in years where there is a high wet state probability. A multivariate normal is used for the density function $\mathbf{y}_t \left| \mathbf{s}_t \right. \sim N(\boldsymbol{\mu}_{\mathbf{s}_t}, \boldsymbol{\Sigma}_{\mathbf{s}_t})$ where $\boldsymbol{\mu}_{\mathbf{s}_t} = \left[\mu_{s_t^{site}}^{site}\right]$, $site = 1,...,d$ and $\boldsymbol{\Sigma}_{\mathbf{s}_t} = \left[\rho_{ij}\sigma_{s_t^i}^i \sigma_{s_t^j}^j\right]$, $i, j = 1,...,d$.

The regional state probability is updated using:

$$
\begin{aligned}
p\left(r_t \mid \mathbf{Y}_1^t, \boldsymbol{\theta}\right) &= \frac{p\left(\mathbf{y}_t \mid r_t, \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) p\left(r_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)}{p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)} \\
&= \frac{p\left(\mathbf{y}_t \mid r_t, \boldsymbol{\theta}\right) p\left(r_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)}{p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)} \\
&= \frac{\left(\sum_{\mathbf{s}_t} p\left(\mathbf{y}_t \mid \mathbf{s}_t, \boldsymbol{\theta}\right) p\left(\mathbf{s}_t \mid r_t, \boldsymbol{\theta}\right)\right) p\left(r_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)}{p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)} \\
&= \frac{\left(\sum_{\mathbf{s}_t} p\left(\mathbf{y}_t \mid \mathbf{s}_t, \boldsymbol{\theta}\right) \left(\prod_{site=1}^{d} p\left(s_t^{site} \in \mathbf{s}_t \mid r_t, \boldsymbol{\theta}\right)\right)\right) p\left(r_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)}{p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)}
\end{aligned}
\qquad (4.6).
$$

This calculation is repeated until $t = T$. Once the forward phase has been completed, $p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)$ terms can be inserted into the overall likelihood equation (4.1).

4.2.2 Interpretation of Switch Probabilities

The switch probabilities provide an assessment of whether or not each fitted site behaves consistently with the regional state series. If all switch probabilities for a set of sites are identified to lie around zero, each site is following closely the regional state probability series. A regional climate control is dominating all sites.

If, for example, only one of a set of sites identifies switch parameters significantly greater than zero, say 0.5, with the other sites having switch probabilities near zero, it is unlikely that the site is affected by the same regional controls, yet a regional state series has been identified. This could either be due to local effects (such as topography) or otherwise due to the added site being outside the region of influence of the regional climate effect.

If the majority of sites identify switch probabilities away from zero, it is unlikely that the regional state series will be identified. This is hence an indicator that local site effects dominate the rainfall variability.

The switch probabilities thus provide a quick check of whether an introduced site is affected by the same regional controls as other sites. This can provide a way of identifying regions where the hidden state Markov structure assumptions are justifiable

4.3 Regional HMM

The Regional HMM differs from the Switch and ordinary HMM in that more than one regional climate control exists. Sites are partitioned into different regions with the number of regions being specified by the user, along with what sites are partitioned into each region. This formulation was designed to test the homogeneous climate region assumption explicitly by testing the single region hypothesis versus the many possible groupings of sites.

The Regional HMM structure is illustrated in Figure 4.2. Rather than a single regional state r_t being used, there is now a vector of regional states $\mathbf{r}_t = \left(r_t^1, r_t^2, ..., r_t^k\right)$, with k being the number of regions. The regional states are conditionally independent of one another given the previous regional state. That is, the regional state only depends on the previous state in the same region. Each site must be assigned to a region from $\{1,...,k\}$ with all regions containing at least one site.

4.3.1 Calculation of Regional HMM Likelihood

As before, the overall likelihood is calculated using:

$$p\left(\mathbf{Y}_1^T \mid \boldsymbol{\theta}\right) = p\left(\mathbf{y}_1 \mid \boldsymbol{\theta}\right)\prod_{t=2}^{T} p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) \qquad (4.7).$$

The parameter vector used for the RHMM is $\boldsymbol{\theta} = \left(\boldsymbol{\mu}_{\mathbf{W}}, \boldsymbol{\sigma}_{\mathbf{W}}, \boldsymbol{\mu}_{\mathbf{D}}, \boldsymbol{\sigma}_{\mathbf{D}}, \boldsymbol{\rho}, \mathbf{P}\right)$, with wet and dry mean and variance parameters for every site $\left(\boldsymbol{\mu}_{\mathbf{W}}, \boldsymbol{\sigma}_{\mathbf{W}}, \boldsymbol{\mu}_{\mathbf{D}}, \boldsymbol{\sigma}_{\mathbf{D}}\right)$, a correlation coefficient matrix $\boldsymbol{\rho} = \left[\rho_{ij}\right] : i, j = 1,...,d$ that is independent of state, and a set of regional transition probabilities $\mathbf{P} = \left[\ p_{ij}^{reg}\ \right] = p\left(r_t^{reg} = j \mid\ r_{t-1}^{reg} = i\right)$, where $i, j = W, D$ $reg = 1,..,k$. Notice here there are now k regional sets of transition probabilities.

Parameters

Regional Transition Probs

$$\mathbf{P}^{reg} = \begin{bmatrix} 1-p_{DW}^{reg} & p_{DW}^{reg} \\ p_{WD}^{reg} & 1-p_{WD}^{reg} \end{bmatrix}$$

Given previous state r_{t-1}^{reg},

generate regional state $\rightarrow$

r_t^{reg} *for all regions*

⇩

Map regional states to sites according site partitioning

$H = (reg_1, reg_2, ...reg_d)$ *where*

⇩

Means and Variance of Rainfall

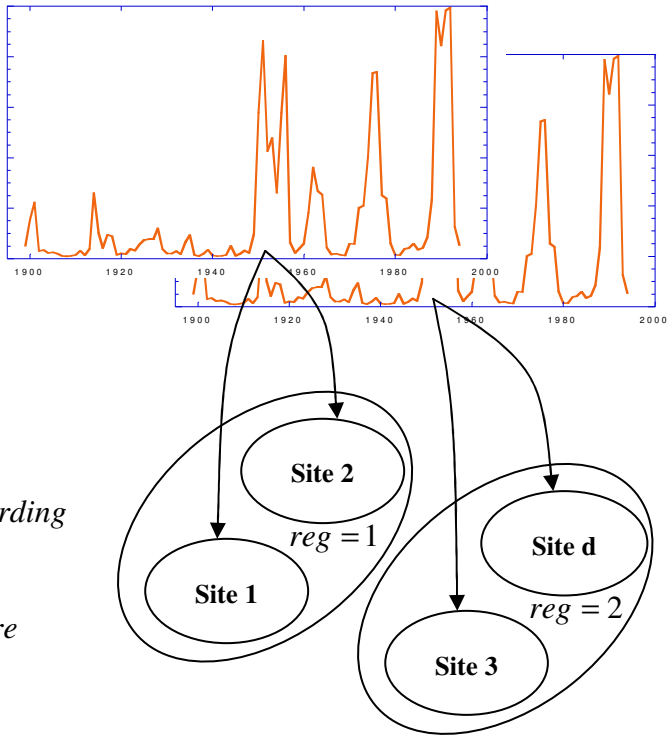
$$\begin{bmatrix} \mu_D^{site} \\ \mu_W^{site} \end{bmatrix} \quad \begin{bmatrix} \sigma_D^{site} \\ \sigma_W^{site} \end{bmatrix}$$

⇨

Coefficient of correlation between sites

$$\begin{bmatrix} 1 & & & \\ \rho_{2,1} & 1 & & \\ \vdots & \cdots & \ddots & \\ \rho_{d-1,1} & \rho_{d-1,2} & \vdots & 1 \\ \rho_{d,1} & \rho_{d,2} & \cdots & \rho_{d,d-1} \quad 1 \end{bmatrix}$$

Regional State Series



Given regional state permutation $\mathbf{r}_t = (r_t^1, r_t^2, ..., r_t^k)$ *sample* $\mathbf{y_t} = (y_t^1, \cdots, y_t^d)$ *from multinormal distribution* $\mathbf{y_t} \sim N(\boldsymbol{\mu}_{\mathbf{r}_t}, \boldsymbol{\Sigma}_{\mathbf{r}_t})$ *with mean and variances inserted according to* $\mathbf{r}_t$ *mapped to each site via* $H = (reg_1, reg_2, ...reg_d)$.

Figure 4.2 Regional HMM Conceptual Diagram

The evaluation of a typical term $p(r_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta})$ is now considered. The Regional HMM likelihood revolves around updating of the regional state permutation probability:

$$
\begin{aligned}
p\left(\mathbf{r}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) &= \sum_{\mathbf{r}_{t-1}} p\left(\mathbf{r}_t \mid \mathbf{r}_{t-1}, \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) p\left(\mathbf{r}_{t-1} \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) \\
&= \sum_{\mathbf{r}_{t-1}} p\left(\mathbf{r}_t \mid \mathbf{r}_{t-1}, \boldsymbol{\theta}\right) p\left(\mathbf{r}_{t-1} \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) \\
&= \sum_{\mathbf{r}_{t-1}} \left(\prod_{reg=1}^{k} p\left(r_t^{reg} \in \mathbf{r}_t \mid \mathbf{r}_{t-1}, \boldsymbol{\theta}\right) \right) p\left(\mathbf{r}_{t-1} \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) \\
&= \sum_{\mathbf{r}_{t-1}} \left(\prod_{reg=1}^{k} p\left(r_t^{reg} \in \mathbf{r}_t \mid r_{t-1}^{reg} \in \mathbf{r}_{t-1}, \boldsymbol{\theta}\right) \right) p\left(\mathbf{r}_{t-1} \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)
\end{aligned}
\tag{4.8}.
$$

Note there are $nstate^k$ regional permutations, with $p\left(r_t^{reg} \in \mathbf{r}_t \mid r_{t-1}^{reg} \in \mathbf{r}_{t-1}, \boldsymbol{\theta}\right)$ being the transition probabilities in each region. A user chosen indicator is required to partition the sites into regions $H = \left(reg_1, reg_2, ..., reg_d\right)$ where $reg \in \{1,...,k\}$ and d is the number of sites and thus mapping the regional state permutation set to each site. The probability of the regional state permutation $\mathbf{r}_t = \left(r_t^1, r_t^2, ..., r_t^k\right)$, can be used to weight the Gaussian rainfall distributions as follows:

$$
\begin{aligned}
p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) &= \sum_{\mathbf{r}_t} p\left(\mathbf{y}_t \mid \mathbf{r}_t, \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) p\left(\mathbf{r}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) \\
&= \sum_{\mathbf{r}_t} p\left(\mathbf{y}_t \mid \mathbf{r}_t, \boldsymbol{\theta}\right) p\left(\mathbf{r}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)
\end{aligned}
\tag{4.9}.
$$

A multivariate normal is used for the density function, $\mathbf{y}_t \left| \mathbf{r}_t \sim N(\mu_{\mathbf{r}_t}, \Sigma_{\mathbf{r}_t}) \right.$ where $\boldsymbol{\mu}_{\mathbf{r}_t} = \left[\mu_{r_t^{reg_{site}}}^{site} \right]$, $site = 1,...,d$ and $\boldsymbol{\Sigma}_{\mathbf{r}_t} = \left[\rho_{ij} \sigma^{i}_{r_t^{reg_i}} \sigma^{j}_{r_t^{reg_j}} \right]$, $i, j = 1,...,d$. Finally, the regional permutation probability is updated:

$$
\begin{aligned}
p\left(\mathbf{r}_t \mid \mathbf{Y}_1^{t}, \boldsymbol{\theta}\right) &= \frac{p\left(\mathbf{y}_t \mid \mathbf{r}_t, \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right) p\left(\mathbf{r}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)}{p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)} \\
&= \frac{p\left(\mathbf{y}_t \mid \mathbf{r}_t, \boldsymbol{\theta}\right) p\left(\mathbf{r}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)}{p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)}
\end{aligned}
\tag{4.10}.
$$

Once updated, this set of calculations are repeated until $t = T$. Once the forward phase has been completed, $p\left(\mathbf{y}_t \mid \mathbf{Y}_1^{t-1}, \boldsymbol{\theta}\right)$ terms can be inserted into the overall likelihood equation (4.7).

4.4 Spatial Dependence: small-scale correlation

4.4.1 Introduction

We now turn to the correlation coefficients of the Gaussian distribution. Taking a step back, we examine the reasons for incorporating these correlations – in short, for spatial dependency.

The modelling of spatial dependency of data is well established with many ways having been proposed to incorporate spatial dependence between data collected at a number of geographical sites (termed a lattice). A comprehensive review of spatial statistical methods can be found in *Cressie* (1991). As stated in *Cressie* (1991 p.113), the following model can be useful is to model spatial data $Z(\mathbf{s})$ where $\mathbf{s}$ is a set of spatial locations on $\mathbb{R}^2$:

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + W(\mathbf{s}) + \eta(\mathbf{s}) + \boldsymbol{\varepsilon}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^2 \qquad (4.11).$$

Here $\mu(\cdot)$ is the (sometimes deterministic) mean structure, called the large-scale variation. $W(\cdot)$ is a zero mean process/distribution, called the smooth small-scale variation. $\eta(\cdot)$ is a zero mean process/distribution, independent of W, called the microscale variation. $\boldsymbol{\varepsilon}(\cdot)$ is a zero-mean white-noise process, independent of W and η, used to capture measurement error. This model is useful as it breaks down the overall variability of the data into components which have intuitive meaning. This additivity assumption for decomposition of the variability within the data is important in enabling this breakdown. Significantly, this decomposition is non-unique as different modellers can use different processes/distributions for each component. Thus, conclusions will vary dependent on the model applied.

Using this framework we examine the variability structure of the models used in this thesis. Of course the models used in this thesis differ from that given in (4.11) in that there is temporal dependence. However, (4.11) is useful in analysing the spatial dependence structure of the model. The overall correlation observed (spatially) within the lattice data $\mathbf{Y}_1^t = (\mathbf{y}_1,..,\mathbf{y}_t)$ is modelled here by two processes which respectively exhibit large- and small-scale variability.

4.4.2 Large- and small-scale variability

A Markovian state occurring concurrently across sites induces some degree of correlation. If two sites share the same state series and have well separated wet and dry distributions, there will be significant correlation between the sites. The ellipses plotted in Figure 4.3a represents the bivariate Gaussian distribution for such a case, where the parameter vector of influence is $\boldsymbol{\theta} = \left(\mu_D^i, \mu_W^i, \mu_D^j, \mu_W^j, \sigma_D^i, \sigma_W^i, \sigma_D^j, \sigma_W^j, \rho_{ij}\right)$ for sites i and j. Indeed as the wet and dry variances go to zero as shown in Figure 4.3b, the correlation will approach one. Conversely, if two sites do not necessarily share the same state, the sites will be less correlated, possibly visiting the four site state combinations $\{DD, DW, WD, WW\}$ as demonstrated in Figure 4.3c. This variation caused by the state series is coined the large-scale correlation. It is intended to model dependency between sites that is independent of distance - that is, correlation due to large-scale climate effects. A positive correlation will result if sites are grouped into the same climate region.

Another means to describe the dependence of annual rainfall between sites is the Gaussian correlation coefficients. Reiterating, the covariance matrix is parameterised $\boldsymbol{\Sigma}_{\mathbf{s}_t} = \left[\rho_{ij}\sigma^i_{s_t^i}\sigma^j_{s_t^j}\right]$, $i, j = 1, ..., d$ where s_t^i is the hidden state at site i (for the switch model). The correlation coefficients are constrained such that the overall correlation coefficient matrix $\boldsymbol{\rho}$ must be positive definite. This is labelled the small-scale correlation, intended at capturing dependency between sites due to topographical features and events occurring over sub-regional scales.

Although the small-scale variability is intended to capture variability at sub-regional scales it is possible that spatial dependency over large scales may be identified by the small-scale structure. Likewise, the large-scale variability included by the Markovian regions of the HMM and its variants can act as small-scale correlation structure. How much of the spatial dependency should be attributed to the small and large scale structures is not known *a priori*. It is possible, indeed likely, that the calibration tries to accommodate the proposed model by fitting the best combination of small and large-scale correlation in terms of likelihood, whether or not this combination of small and large-correlation is what the modeller intended. This issue is linked to the nonunique

nature of the decomposition (4.11), and is explored further upon application of the models in Section 5.5.

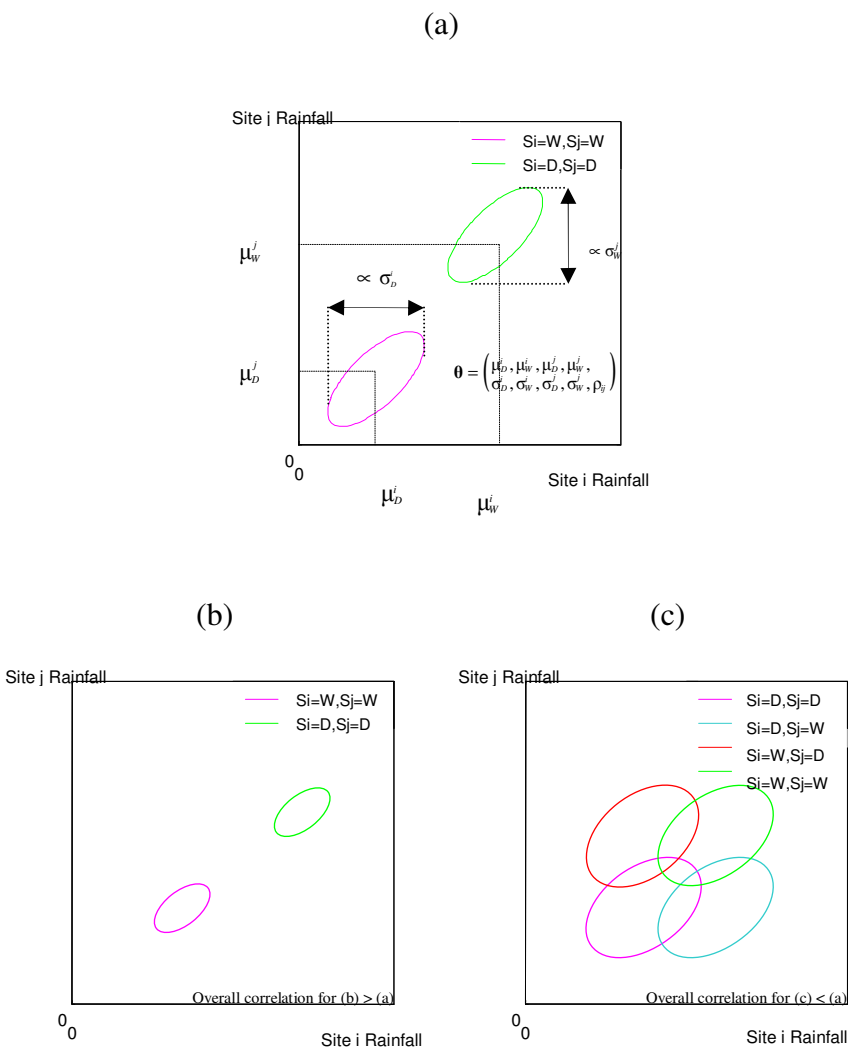


Figure 4.3 Gaussian ellipses for (a) sites in the same region, (b) lower identified state variance and (c) sites not in the same region.

The large-scale mean structure here is determined by the Markovian hidden states (and regional groupings). The small-scale variation $W(\cdot)$ is modelled as a zero mean Gaussian distribution, with covariance determined according to site state variances, and the correlation coefficient matrix. This structure differs from decomposition of (4.11) in that there are no terms explicitly modelling the microscale variation $\eta(\cdot)$ or measurement error $\boldsymbol{\varepsilon}(\cdot)$. However, due to the additive structure of the (4.11), the

Gaussian $W(\cdot)$ distribution could be considered to include these effects. Typically, a Gaussian distribution with constant variance across all sites (and zero covariance) is used to model $\eta(\cdot)+\boldsymbol{\varepsilon}(\cdot) \sim N(0,\tau^2\mathbf{I})$, where $\mathbf{I}$ is the identity matrix and τ^2 is a variance parameter to be inferred. The microscale variation term, often called the nugget effect (*Matheron*, 1962, *Cressie*, 1991 p.59), is included to reproduce unexplained variability occurring over scales smaller than the distance between sites. For example, consider two closely spaced sites. It is possible in a climate which produces significant rainfall from small-scale systems that rainfall at the two sites may be affected by different events. Put another way, a significant meso- or micro-scale storm cell may pass over one site but miss the other thereby weakening the correlation one would expect between closely spaced sites.

It is noted that the necessity for such a nugget term modelling unexplained variation may be reduced by the inclusion of key covariates in the model – for example other atmospheric indices (temperature, air pressure, number of cyclone crossings etc).

4.4.3 Parameterisation of the correlation matrix

Many parameterisations of the correlation coefficient matrix are possible. One alternative is to use zero correlation, thus leaving the Markovian states to account for inter-site dependency. For the full RHMM, with every site in its own region, this zero correlation model is essentially the single site HMM of *Thyer* (2001). Another simple method would be to use correlation coefficients estimated empirically (from the data) – an empirical Bayes approach (*Morris*, 1983). This was the first method used when formulating the models applied in this thesis. However, this method is not in keeping with the Bayesian ideal of allowing uncertainty within parameters – that is, there is inadequate accommodation of uncertainty. Also, forcing these low level parameters to assume the value of the overall estimated correlation runs the risk of forcing an inappropriate structure. That is, the overall correlation observed in the data may not be produced by this small-scale correlation alone.

Another approach is to consider the correlation matrix as completely unknown, and fit every individual correlation. Indeed this is the main approach taken in the remainder of this thesis. This method was chosen as it allows a better understanding of the

relationship between these correlations and variability induced by the Markovian state series.

Correlation functional relationships based on distances are typically used in daily rainfall studies to capture short-range dependence between sites (eg. *Sanso and Guenni*, 2000). As annual rainfall accumulations are used in this study, this short-range dependence along with large-scale climate dependence is likely to be the cause of the observed correlation. These techniques are useful as they dramatically reduce the number of parameters requiring inference. One such functional (from the Matérn class - *Handcock and Wallis*, 1994) correlation structure is the exponentially decaying correlation structure:

$$\rho_{ij} = \exp\left(\frac{-dist_{ij}}{\lambda}\right), \lambda > 0 \qquad (4.12),$$

where $dist_{ij}$ is the distance between sites $i, j = 1,...,d$, and λ is a range parameter to be inferred.

Although not used in the main case studies described in Chapter 5, a functional relationship such as the exponential correlation decay is required for application of the RHMM model to a larger number of sites simply to avoid an unmanageable number of parameters. Some initial testing of the exponential correlation structure is undertaken, and comparisons are made to results using individually fitted correlations.

The microscale variation (nugget effect term) was not applied in the case studies of Chapter 5. Considering the additive nature of (4.11), the flexibility of fitting individual correlations (and site variances) and the small number of sites used, this was not considered a limitation. The intuition behind this consideration is due to the additive nature of Gaussian distributions, the nugget variance is not formally identifiable. That is, every possible covariance structure available when using the nugget term is also possible for the cases where individual fitted correlations and no nugget term are used. However, every possible overall covariance for the exponential decay combined with the nugget effect is not possible using the exponential decay alone. Therefore, upon application to a larger number of sites, with a less flexible correlation structure (such as the exponential decay) it is expected that use of a nugget effect term would be justified.

Again, some initial tests applying the nugget effect term are presented in conjunction with the exponential correlation decay testing.

The parameterisations of the correlation coefficient (and hence covariance) matrix differ from the work of *Thyer* (2001), where all elements of the covariance matrices (wet and dry) were considered unknown. This essentially means that correlation parameters were fitted for both the wet and dry states. Here however, the correlation matrix is modelled independently of state. Table 4.1 compares the different number of parameters for each of the HMM, Switch HMM and Regional HMM correlation parameterisations. Note the quadratic rise in correlation/covariance parameters with the number of sites d.

Table 4.1 Number of parameters for the Ordinary, Switch and Regional HMM

Model	Correlation parameterisation	Trans Prob	Switch Prob	Mean & Variance	Covariance or Correlation	Total Parameters
Thyer's HMM	Fitted Covariance	2		$4d$	$d(d-1)$	$\left(2d^2+6d+4\right)/2$
HMM	Zero/Empirical	2		$4d$		$\left(8d+4\right)/2$
	Exponential decay	2		$4d$	1	$\left(8d+6\right)/2$
	Fitted Correlation	2		$4d$	$d(d-1)/2$	$\left(d^2+7d+4\right)/2$
Full Switch HMM	Zero/Empirical	2	$2d$	$4d$		$\left(12d+4\right)/2$
	Exponential decay	2	$2d$	$4d$	1	$\left(12d+6\right)/2$
	Fitted Correlation	2	$2d$	$4d$	$d(d-1)/2$	$\left(d^2+11d+4\right)/2$
Full Regional HMM	Zero/Empirical	$2d$		$4d$		$\left(12d\right)/2$
	Exponential decay	$2d$		$4d$	1	$\left(12d+2\right)/2$
	Fitted Correlation	$2d$		$4d$	$d(d-1)/2$	$\left(d^2+11d\right)/2$

4.5 Parameter Priors

Within the Bayesian framework, it is necessary to specify the prior distribution of parameters $p(\boldsymbol{\theta})$. An empirical Bayes approach is adopted in this study. Empirical Bayes describes a modelling outlook where the data, as far as possible, is used for inference on parameters (*Berger*, 2000, *Carlin and Louis*, 2000). In line with the

empirical Bayes approach, priors were chosen with the intention of being weakly informative (see discussion in sections 3.2.1 and 3.7.1). Also priors on parameters shared in all models were not altered from one model to the next so as to not favour one model over another *a priori*.

Table 4.2 presents a summary of the priors, listing firstly the parameter of interest, then the model(s) that the parameter applies to. The prior distributions used for the parameters along with parameter bounds are given. In sections 4.5.1 and 4.5.2 some practical considerations involving application of these priors and models are discussed, while the following sections justify the choice of priors for each of the models.

Table 4.2 Parameter prior distributions and bounds

Parameter	Model(s)	Prior distribution	Lower bound	Upper bound	Hyper-parameters
$\mu_D^{site}, \mu_W^{site}$	ALL	$N\left(\mu_0, \sigma_y^{\,2}/\kappa\right)$	0	10,000	$\mu_0 = \overline{y}_{site}$, $\kappa = 1$
$\sigma_D^{site}, \sigma_W^{site}$	ALL	$Inv-\chi^2\left(\nu_0, \sigma_0^2\right)$	0	∞	$\sigma_0^2 = \overline{s}_{site}^{\,2}$, $\nu_0 = 2$
p_{DW}^{reg}, p_{WD}^{reg}	ALL	*Uniform*	0	1	not applicable
$p(s_t^{site} = W \mid r_t = D)$ $p(s_t^{site} = D \mid r_t = W)$	SHMM	$Beta(\alpha, \beta)$	0	1	$\alpha = 1$, $\beta = 2$
ρ_{ij}	ALL	*Uniform*	0	0.95	

4.5.1 Parameter non-identifiability and label switching

Before the individual parameter priors are discussed, the issue of parameter non-identifiability warrants discussion. An inherent feature of mixture modelling (we are using a mixture of wet and dry Gaussian distributions) is a property known as non-identifiability (*Celeux et al.*, 2000, *Stephens*, 2000, *Fruhwirth-Schnatter*, 2001). Parameter non-identifiability describes the invariance of the likelihood (and the posterior for flat and symmetric priors) to permutations of the labeling of parameters.

For the two-state HMM, every parameter set $\boldsymbol{\theta}$ has a corresponding parameter set $\boldsymbol{\theta}^*$ that fits the data equally well. For example, for the single site HMM there are six parameters, the two transition probabilities p_{DW} and p_{WD} and the state specific

parameters, the wet and dry means and standard deviations $(\mu_W, \sigma_W, \mu_D, \sigma_D)$. Let us say we have a set of these parameters which we think are the true parameters. If the labelling on these parameters is switched, that is, if for all the parameters we switch the labelling of *W* and *D*, an identical fit to the data is found. The model itself cannot differentiate between the two different labellings. This means that for any state specific parameter (either the mean or the variance), there will be a symmetry in the posterior distributions (with the posterior distributions for each state being identical) as shown in schematic form in Figure 4.4. Theoretically, the number of samples taken in each labeling subspace using unconstrained MCMC sampling should be equal. In practice the sampler is only run for a finite amount of iterations, and may not jump between the mirrored distributions in a balanced way. This may occur when the distributions are well separated such as in Figure 4.4a, with the sampler possibly not jumping between these subspaces at all, occupying only one subspace. A parameter constraint is usually applied in such cases forcing the model to occupy only one subspace within the sampling distribution (eg. as discussed in *Stephens*, 1997 p.43). If the results are going to be interpreted in a physical sense we also would want posterior distributions to differ for the wet and dry means. Given that an intended use of this model is to condition a smaller timescale model using the posterior state series, it is a necessary requirement that such a unique labelling is used.

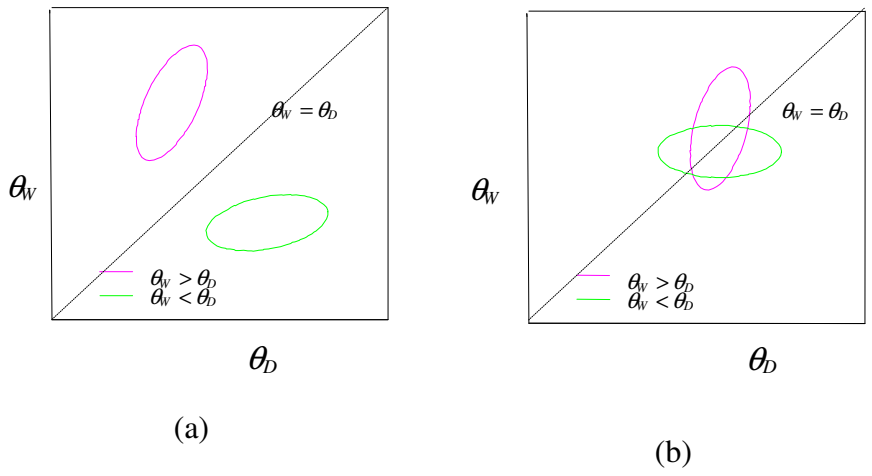


Figure 4.4 Posterior parameter schematic for parameter showing label switching (a) a well separated posterior and (b) overlapping posterior.

In the case of the two state HMM, imposing the parameter constraint on every site mean rainfall $\boldsymbol{\mu_W > \mu_D}$ during MCMC sampling is usually able to address this problem (*Richardson and Green*, 1997, *Thyer*, 2001). However, with the extra layer of switch parameters in the Switch HMM comes a new opportunity for label switching. Another simple parameter constraint is used to overcome this problem. The transition probabilities are bounded by the constraint $p_{DW} < p_{WD}$ which in terms of the model prior is an expectation that the dry year residence times are expected to be longer than the wet year residence times.

Theoretically, the marginal likelihood produced from a mixture model under a labeling constraint (with the prior normalized for the unoccupied space), and the marginal likelihood of an unconstrained model should be equal. However, in some mixture models, it is possible for label switching to cause a downwards bias in calculation of the marginal likelihood according to the Gelfand-Dey reciprocal importance estimator (*Fruhwirth-Schnatter*, 2002) even where parameter constraints have been employed. This bias occurs when the posterior distribution approaches a labeling constraint. If the unconstrained posterior is overlapping like that shown in Figure 4.4b, and a sampler is applied with constraint $\theta_W > \theta_D$, label switching can still occur due to the reflection of the distribution below the constraint (the magenta ellipse). For the HMM, SHMM and RHMM the $\boldsymbol{\mu_W > \mu_D}$ constraint was applied. Label switching as described above would require the mean distributions to be poorly separated $\mu_W \approx \mu_D$. For the HMM and SHMM it is not expected that this label switching will affect results as the label switching requires all sites under the influence of a regional climate state to show poor separation. Previous testing using the HMM has shown that significant mean separation occurs for at least one of the sites used in each analysis within Chapter 5. For the RHMM, a greater possibility of label switching arises as there are transition probabilities associated with individual sites. Therefore, as there are more groupings of sites under regional state controls, it is more likely that one of these regions identifies poor mean separation for all sites within. However, it was considered that a downward bias in the marginal likelihood for poorly separated regions was not of significance in this study as such a bias favours grouping of sites such that greater state separation occurs.

It is reiterated that the scale of the label switching problem depends on the aim of the analysis: it is indeed a problem if the aim is model identification, but is less important if the aim is prediction from a given model. Predictions/simulations from each model should have the same statistical characteristics whether or not label switching has been addressed. It is again a problem if model averaged predictions are used (if estimates of model probability are biased), as resulting simulations will be biased.

Alternate methods to address the label switching problem without the imposition of parameter constraints during sampling (*Celeux et al.*, 2000, *Stephens*, 2000, *Fruhwirth-Schnatter*, 2001) were investigated. Essentially these methods force balanced switching to occur such that an equal number of samples are taken from all labeling subspaces. However, the simplicity with which the parameter constraints can be applied was favoured over these more complicated sampling schemes. The possibility of a biased marginal likelihood is a weakness of the current sampling implementation, and results must be interpreted with it being understood that models with poorly separated means may have been discriminated against. Future work will need to formally address this label switching problem with the application of one of the aforementioned methods.

4.5.2 Parameter constraints: normalizing prior distributions

The parameter constraints related to label switching are actually employed within the parameter priors. The unconstrained prior $p(\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \boldsymbol{\Theta})$ defined over support $\boldsymbol{\Theta}$ requires normalisation, as many parameter combinations are no longer feasible. The following identity is used to define the normalised prior over the constrained support $\boldsymbol{\Theta}' \subset \boldsymbol{\Theta}$:

$$p(\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \boldsymbol{\Theta}') = \frac{p(\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \boldsymbol{\Theta}) I[\boldsymbol{\theta} \in \boldsymbol{\Theta}']}{P(\boldsymbol{\theta} \in \boldsymbol{\Theta}')} = \frac{p(\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \boldsymbol{\Theta}) I[\boldsymbol{\theta} \in \boldsymbol{\Theta}']}{\int_{\boldsymbol{\Theta}'} p(\boldsymbol{\theta} \mid \boldsymbol{\theta} \in \boldsymbol{\Theta}) d\boldsymbol{\theta}} \quad (4.13),$$

where $I[\cdot]$ is an indicator function with value 1 if the statement is true, or 0 if the statement is false. A MCMC proposed parameter set which does not satisfy the constraints is given a prior value of zero, hence making it impossible for the sampler to visit the sampled value.

It is noted here that if the normalising constant relates to parameters which are common to the models in the selection set and have the same priors, it is not necessary to normalise the associated prior distribution. Such normalising constants cancel when calculation of model weights is performed using (3.18). In this study, all of the normalising constants were calculated (analytically or by numerical integration). These calculations were performed so as to allow simple comparison in the future between these models and other models which have not been tested here and which do not share the same normalising constants. Normalising constants related to label switching (required as a prior on the mean) and the positive definite requirement for the coefficient of correlation matrix are presented in the following section discussing choice of priors.

4.5.3 Shared Parameters: Ordinary, Switch and Regional HMM

As many parameters are used in the HMM and the generalisations tested in this thesis, the same priors were placed on these shared parameters so as to not favour one model over another *a priori*.

Means $\mu_{D,W}^{site}$ and variances $\sigma_{D,W}^{site\ 2}$

The same priors as those used in the single site test case in Section 3.10 were used for all models for the means $\mu_{D,W}^{site}$ and variances $\sigma_{D,W}^{site\ 2}$. The means and variances at each site are independent of one another, with identical priors on the wet and dry distributions. The mean and variance are modelled jointly according to the following relationship:

$$\begin{aligned} p\left(\boldsymbol{\mu},\boldsymbol{\sigma} \mid \mu_D^{site} < \mu_W^{site} : site = 1,...,d\right) &= \prod_{site=1}^{d} \left(p\left(\mu_D^{site},\sigma_D^{site},\mu_W^{site},\sigma_W^{site} \mid \mu_D^{site} < \mu_W^{site}\right)\right) \\ &= \prod_{site=1}^{d} \left(\prod_{i=D,W} p\left(\mu_i^{site},\sigma_i^{site} \mid \mu_D^{site} < \mu_W^{site}\right)\right) \\ &= \prod_{site=1}^{d} \left(\frac{\left(\prod_{i=D,W} p\left(\mu_i^{site},\sigma_i^{site}\right)\right) I\left[\mu_D^{site} < \mu_W^{site}\right]}{P\left(\mu_D^{site} < \mu_W^{site}\right)} \right) \end{aligned} \quad (4.14).$$

The outer product over the sites follows the prior independence among sites assumption, while the inner product is a consequence of the independence between states. The probability $P\left(\mu_D^{site} < \mu_W^{site}\right)$ normalises the joint distribution according the parameter constraint, and is equivalent to the $P\left(\boldsymbol{\theta} \in \boldsymbol{\Theta}'\right)$ term in (4.13). Because the prior on μ_D^{site} and μ_W^{site} are assumed equal, the prior is (marginally) symmetric about the line $\mu_D^{site} = \mu_W^{site}$. Accordingly $P\left(\mu_D^{site} < \mu_W^{site}\right) = 1/2$.

Given that the priors on the mean and variance are specified by:

$$\begin{aligned} \mu_i^{site} \mid \sigma_i^{site^2} &\sim N\left(\mu_0, \sigma_y{}^2/\kappa\right) \\ \sigma_i^{site^2} &\sim Inv-\chi^2\left(\nu_0, \sigma_0^2\right) \end{aligned} \qquad (4.15),$$

the joint probability of the state means and variance is defined as:

$$\begin{aligned} p\left(\mu_i^{site}, \sigma_i^{site^2}\right) &= p\left(\mu_i^{site} \mid \sigma_i^{site}, lbnd < \mu_i^{site} < ubnd\right) p\left(\sigma_i^{site^2}\right) \\ &= \frac{f_N\left(\mu_i^{site}; \mu_0, \left(\sigma_i^{site}\right)^2/\kappa\right) I\left[lbnd < \mu_i^{site} < ubnd\right]}{P\left(lbnd < \mu_i^{site} < ubnd\right)} f_{Inv-\chi^2}\left(\sigma_i^{site^2}; \nu_0, \sigma_0^2\right) \\ &= \frac{f_N\left(\mu_i^{site}; \mu_0, \left(\sigma_i^{site}\right)^2/\kappa\right) I\left[lbnd < \mu_i^{site} < ubnd\right] f_{Inv-\chi^2}\left(\sigma_i^{site^2}; \nu_0, \sigma_0^2\right)}{\int_{\mu_i^{site}=lbnd}^{ubnd} \int_{\sigma_i^{site^2}=0}^{\infty} f_N\left(\mu_i^{site}; \mu_0, \left(\sigma_i^{site}\right)^2/\kappa\right) f_{Inv-\chi^2}\left(\sigma_i^{site^2}; \nu_0, \sigma_0^2\right) d\sigma_i^{site^2} d\mu_i^{site}} \end{aligned} \qquad (4.16),$$

where f_N and $f_{Inv-\chi^2}$ are the density functions for the Normal and Inverse-χ^2 distributions respectively. The indicator function imposes the upper (*ubnd*) and lower (*lbnd*) bound restrictions on the mean. Calculation of the Bayes factor and MCC sampling both require exact calculation of the prior (rather than to a constant of proportionality). As the marginal joint distribution has been truncated, the denominator term is inserted here to normalise this distribution to sum to one. The analytic calculation of the denominator is provided in Appendix A.

The application of priors on the variance parameters independently of correlation parameters differs from the work of *Thyer* (2001), where a prior was applied to all elements of the covariance matrices (wet and dry) jointly. In that study, an Inverse-Wishart distribution was used for the prior on the covariance matrices. An Inverse-

Wishart distribution is the multivariate generalisation of the Inverse-χ^2 distribution. In this study, an Inverse-χ^2 distribution was used for the diagonals of the covariance matrices (the variances), and hence has the same prior on the variance parameters as *Thyer* (2001).

Following *Thyer* (2001) the site means $\overline{y}_{site}$ and variance $\overline{s}_{site}^{\,2}$ (see Section 3.6.1) were used to define the prior mean and variance. The prior degrees of freedom (κ and ν_0) were kept to a minimum to ensure that the priors remained diffuse. All means were truncated below zero, with an upper bound of 10,000.

Correlations ρ_{ij} and other small-scale variation parameters

The priors on parameters related to small-scale and microscale variability are summarised in Table 4.3.

Table 4.3 Small-scale variability parameter prior distributions and bounds

Parameter	Small Scale Variability Model(s)	Prior distribution	Lower bound	Upper bound	Hyper-parameters
ρ_{ij}	Zero Correlation				$\rho_{ij} = 0$
	Empirical Correlation				$\rho_{ij} = \overline{\rho}_{ij}$
	Fitted Correlation	*Uniform*	0	0.95	
λ	Exponential Correlation	$Gamma(\alpha_\lambda, \beta_\lambda)$	0	∞	$\alpha_\lambda = 5$, $\beta_\lambda = 100$
τ	Nugget Effect (Microscale)	*Uniform*	0	200	

Within the majority of the case studies undertaken in this thesis, individual (site-to-site) correlations were inferred i.e. the fitted correlation model in Table 4.3. Correspondingly, a prior is required on all correlation parameters ρ_{ij}. Uniform priors were placed on these correlations with bounds $[0, 0.95]$. The 0.95 upper bound was applied as no pair of sites has previously produced empirical correlations as high as 0.95. Within the case studies this correlation structure is referred to as Fitted Correlations. It is noted that not every combination of correlations produce a positive-definite covariance matrix. Consequently, for comparison of this model using marginal

likelihood against others which are positive definite for all parameter values (such as the exponential decay), normalisation of the joint prior on all correlation coefficients is required as per Section 4.5.2. For the four site case studies in the following chapter, this normalising constant was calculated using numerical integration. This involved sampling correlations from the prior, and testing whether or not the correlation matrix was positive definite. The ratio of accepted correlation matrices were then used to normalise the marginal likelihood according to (4.13).

Zero Correlations and Empirical Correlations (estimated from data) are applied in some case studies. In terms of prior, this is essentially equivalent to putting the entire prior distribution at the appropriate parameter value.

The alternate formulation of the correlation matrix, the Exponential Correlation decay structure, was also trialled for some initial testing and comparison to the Fitted Correlation model. Thus a prior is required on the λ range parameter. As in the study of *Sanso and Guenni* (2000), the prior on λ was a Gamma distribution:

$$\lambda \sim Gamma(\alpha_\lambda, \beta_\lambda) \tag{4.17}.$$

The hyperparameters chosen, $\alpha_\lambda = 5$ and $\beta_\lambda = 100$, were based on the plot of empirically estimated correlations from the 13 sites used in the case studies, and is shown in Figure 4.5.

As Figure 4.5 demonstrates, the prior on λ was chosen so as to accommodate the empirical correlations for the majority of sites, and also be reasonably vague. Sites within 400km were considered more important than those outside this radius, as the exponential correlation decay represents an attempt to model small-scale variability.

In addition to the small scale variance modelling, some initial tests were undertaken with a nugget effect (microscale) variation term added as describe in Section 4.4.2. This Gaussian distribution requires a variance parameter τ^2 to be inferred. Thus a prior is required on τ. A uniform prior was used on τ with bounds $[0,200]$. The upper bound was chosen as it is greater than the empirically estimated standard deviations for the majority of sites tested.

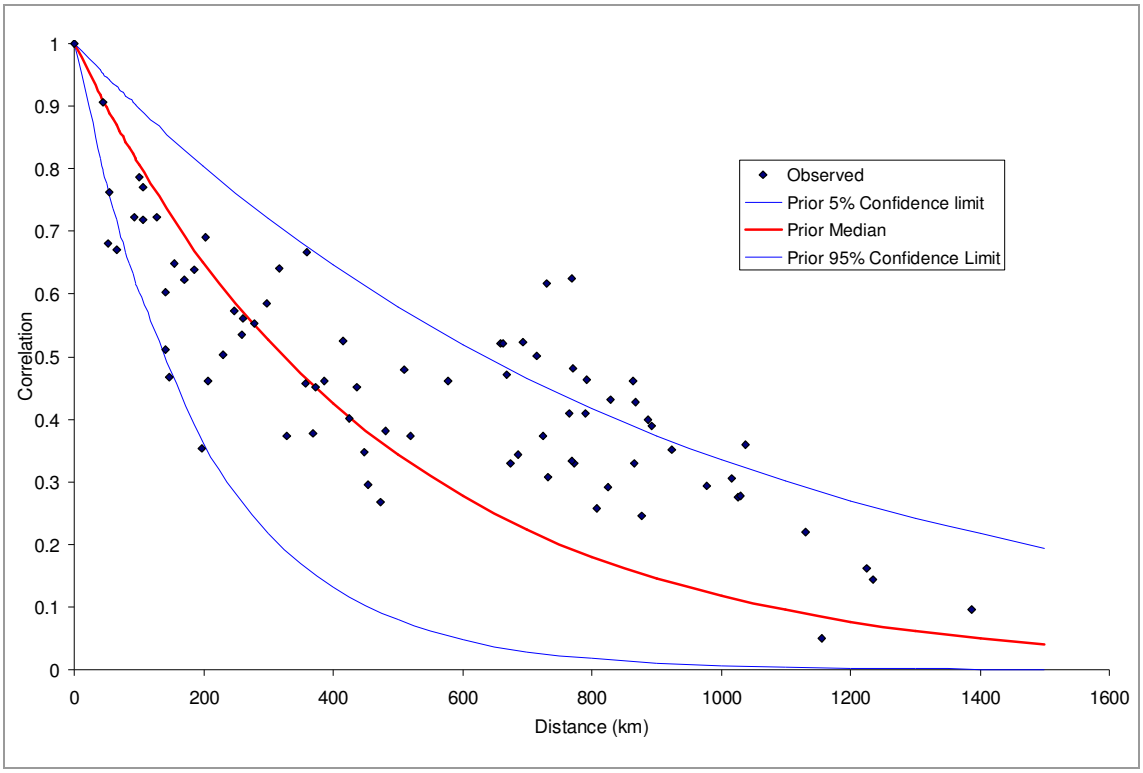


Figure 4.5 Empirically estimated correlation coefficients using 13 sites spanning 1900-1986 versus λ prior distribution (converted to correlation)

Transition Probabilities p_{DW} and p_{WD}

Uniform priors over the range $[0,1]$ were used on all transition parameters throughout this study.

4.5.4 Switch HMM

Transition Probabilities Constraint Normalisation

As mentioned in Section 4.5.1, an additional constraint was imposed on the transition probabilities within the Switch HMM to uniquely identify the parameters. As was done for the parameter means, normalization of the priors must be performed. The probability that the switch constraint is satisfied is:

$$P(p_{DW} < p_{WD}) = \tfrac{1}{2} \qquad (4.18).$$

The need for guiding priors

If the switch HMM is to be useful in practice, it needs to be able to identify the regional two state Markov structure given around 100 years of data. The simplest two state Markov structure within the scope of the SHMM is the HMM itself. With the HMM, there are no switch probabilities to confuse the identification of the regional persistence structure. To ensure that the SHMM was able to correctly identify this simplest case, 100 years of HMM data were generated, and the SHMM was calibrated to this data. As the state series is known for generated data, the posterior state series can be compared to the true state series. Differences between these two series can indicate a lack of identifiability arising from the level of complexity of the SHMM. In accordance with the empirical Bayes approach, a flat prior distribution was used for the newly added switch parameters.

The 100 years of generated data was based on the parameters shown in Table 4.4. These parameters reflect the greatest separation of states observed through previous calibrations of the HMM. Therefore, these parameters represent the most easily identifiable HMM structure. Figure 4.6 shows for a single site and two site simulation the result of calibration of the switch HMM. The true state series is indicated by the dotted line.

Table 4.4 Generated Data Parameters

Parameter		Value
p_{DW}		0.05
p_{WD}		0.25
All sites	μ_D^{site}	1000
	μ_W^{site}	1400
	$\sigma_{D,W}^{site}$	100
	ρ_{ij}	0.8

For both the single site and two state calibrations, the posterior regional state probability series does not correspond to the generated state series particularly well. The single site probability series lingers around a value of 0.5, not identifying the persistence structure. The two site identification is slightly better (as there is more regional state information

contained in two sites). However, given that this is the easiest Markov structure to identify (no switching) and also that this is the best state separation we are likely to encounter, this level of identification is considered inadequate. Priors on parameter distributions will be needed to combat the amount of uncertainty introduced by the switch parameters.

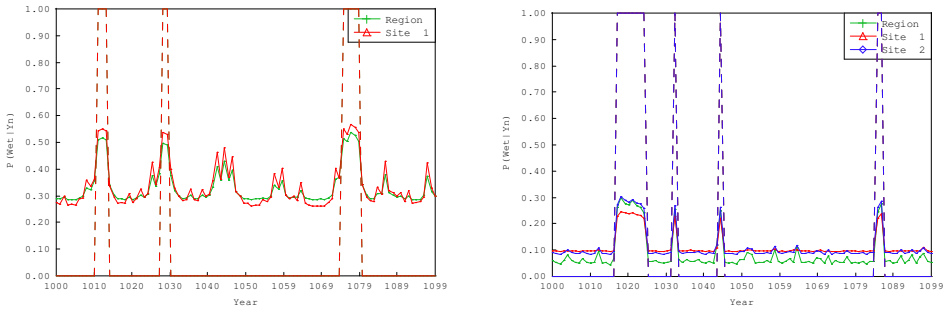


Figure 4.6 Generated vs. posterior state series (a) single site and (b) two site for a flat prior

A Beta distribution ($p(s_t \neq r_t \mid r_t) \sim Beta(1,2)$) was used for the prior distribution of switch probabilities and is shown in Figure 4.7. This prior favors switch probabilities near zero, and represents a belief that the sites are likely to be controlled by the regional climate state. The contribution to identification of this prior is evident in Figure 4.8 when compared to Figure 4.6. Although not perfect for the single site, the state probability series for the two site case is near perfect. As this model is not intended to be used for single sites, the prior was considered adequate for identification. There is a danger when using priors that too strongly favor a particular parameter set. However, as this was the best possible opportunity for identification (the data was generated by a HMM), and the model still was not clearly identified for the single site, the prior was considered to be a good balance between the empirical Bayes perspective and the need for identification.

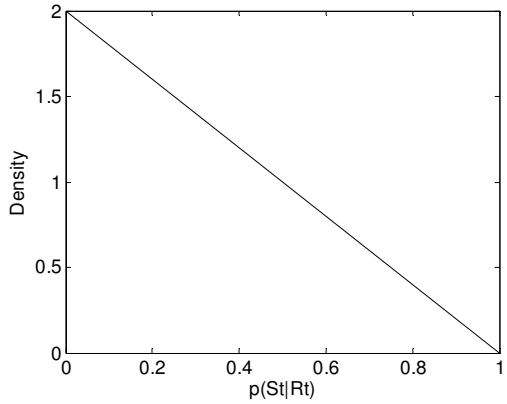


Figure 4.7 Switch parameter $Beta(1,2)$ prior density

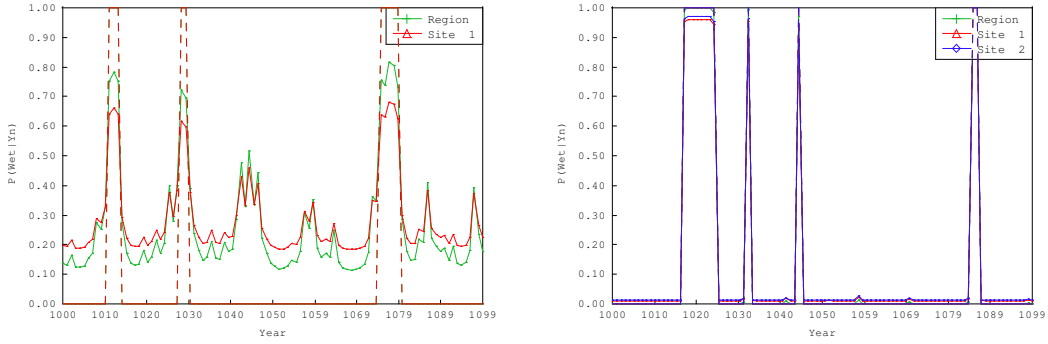


Figure 4.8 Generated vs. posterior state series (a) single site and (b) two site for a beta prior

4.5.5 Regional HMM

Transition Probabilities

As in the ordinary HMM and the Switch HMM, uniform priors over the range $[0,1]$ are used on each of the individual transition probabilities $\mathbf{P} = \left[p_{ij}^{reg} \right] = p\left(r_t^{reg} = j \mid r_{t-1}^{reg} = i \right)$, where $i, j = W, D,\ i \neq j,\ reg = 1,...,k$ within the Regional HMM.

4.6 Model Priors

Within the Switch HMM and the Regional HMM frameworks there are many possible models.

4.6.1 Switch HMM variants

The full Switch HMM with switch probabilities at every site is one possibility, with the HMM at the other extreme (no sites with switch probabilities). In between there are many other combinations possible, with some sites having switch parameters, and others having none. Overall, there are 2^{nsite} possible Switch HMM variants.

4.6.2 Regional HMM variants

The Regional HMM possibilities are more complex. The problem is to list the total number of ways d rainfall gauge sites can be partitioned into k climate regions, where $k \le d$. In mathematical parlance, we wish to generate partitions of a d*-set* into k non-empty *blocks*. Table 4.5 lists the possible ways to partition four objects, with each number corresponding to the region that that site is grouped. For example, the block 1,1,1,1 groups all sites into region 1, whereas 1,2,3,4 partitions each site into its own region. This discussion of set partitions, and the algorithms used in generating the set partitions follows *Stanton and White* (1986 p.18).

Table 4.5 Possible Partitions of a four site set

1,1,1,1	1,2,1,1	1,2,2,3
1,1,1,2	1,2,1,2	1,2,3,1
1,1,2,1	1,2,1,3	1,2,3,2
1,1,2,2	1,2,2,1	1,2,3,3
1,1,2,3	1,2,2,2	1,2,3,4

The number of set partitions of $[d]$ with k blocks is called the *Stirling number of the second kind* $S(d,k)$, and is calculated according to the recursive sum:

$$S(d,k) = S(d-1,k-1) + k \cdot S(d-1,k) \quad , \; where\, S(1,k) = 1 \qquad (4.19).$$

The total number of set partitions of $[d]$ is called the *Bell number*, and is simply the sum of the *Stirling numbers* over k. As the Regional HMM is intended to be used on groups containing at least four sites, Table 4.6 lists the Stirling and Bell Numbers against a few demonstrative number of sites.

Table 4.6 Possible blocks of *d* sites into *k* blocks

Number of Sites *d*	Number of blocks *k* (*Stirling number*)								Total number of partitions
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	
1	1								1
2	1	1							2
4	1	7	6	1					15
8	1	127	966	1701	1050	266	28	1	4140

Clearly, the total number of possible models (partitions) grows faster for the Regional HMM than the Switch HMM ($2^8 = 256$). Should all of these models be investigated, calibrated and compared? From a computation viewpoint, the fewer models the better. As we have chosen here to optimise every model, and sample from every model using MCMC, too many models could mean prohibitive computation time. That is, for a given number of sites, computation time rises at least linearly with the number of models. Hence, when the number of sites increases, it may not be possible to model all combinations. More importantly, is allowing each and every possible partitioning feasible in a physical sense? A methodology of subjectively choosing plausible sets of partitions to work with is needed.

A method of culling models is required with culled models effectively being allocated a prior model weight $p(M) = 0$. Such models are designated *a priori* impossible. A culling method based on topographical location was introduced. A map is drawn with edges joining certain (neighbouring) designated pairs of sites that can be partitioned into a block. If for a particular partitioning, there is no edge connecting a site with any other sites within the same block, that partitioning is given zero weight. This mapping produces climate regions that are more physically realistic than partitionings with sites well separated from one another being in the same region and in-between sites in a different region. To demonstrate this mapping a simple four site example is shown in Figure 4.9, with ellipses representing the vertices (sites) and lines representing edges (neighbours).

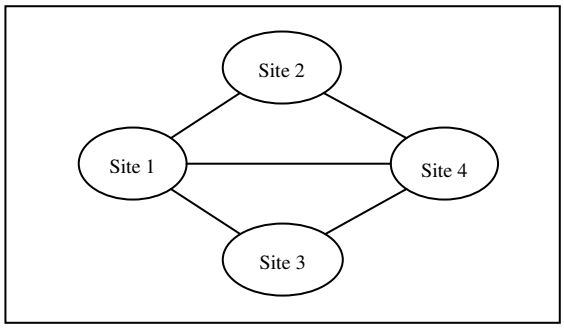


Figure 4.9 Four Site Partition Map

Figure 4.9 shows all four sites being neighbours of one another with the single exception of Site 2 and Site 3. This omission of an edge, means that any partitioning which groups Sites 2 and 3, whilst not being partitioned with Sites 1 or 4 has a zero model prior. In terms of the 15 models/partitions given in Table 4.5, two models are given zero prior weight, models 1,2,2,1 and 1,2,2,3. Although in this simple 4 site case, this culling only reduces the number of sites by 2, for cases above 4 sites, this process exponentially culls partitions. Of course, it may be desired that all partitions be modelled, and this can be achieved by maintaining all sites as neighbours. Neighbouring sites are designated by the user, and hence are arbitrary. However, it is believed that basing site groupings on site locations that are nearby is a reasonable method of designating models that are physically plausible.

4.6.3 Uniform model prior

Uniform model priors were used for all models tested in the next chapter. That is, the HMM, the individual Switch HMM and Regional HMM variants that were not culled were all assigned equal weighting *a priori*.

4.7 Conclusion

This chapter introduced the two generalisations of the HMM, the Switch HMM and the Regional HMM. Both of these frameworks were designed to relax the assumption of a single regional climate state exerting control over all sites. The Switch HMM allows individual sites to exhibit anomalous behaviour with regard to the regional state. The Regional HMM, on the other hand, partitions the sites into various climate regions, each with their own climate state series.

Calculation of the likelihood along with associated parameter priors for both the models was detailed. Many models are possible within these new frameworks. A uniform model prior was used over all possible HMM, Switch HMM and Regional HMM variants. A culling technique based on neighbouring sites was used to reduce the number of Regional HMM variants to avoid searching through an unmanageably large model space.

The spatial dependency structure of the models was examined. It was noted that the regional climate state induces a large-scale correlation independent of distance within a region, whereas the multi-normal distribution accounts for small-scale correlation. In particular, several parameterisations of the small-scale Gaussian correlation were proposed for testing in the following chapter. The most flexible correlation structure, treating the correlation coefficients as completely unknown, and fitting individual site-to-site correlations was chosen for use in the majority of the case studies. It is believed this will provide insight into using other (less parameterised) functional correlation relationships such as the exponential decay.

This chapter has laid the foundation for the next chapter in which Bayesian model selection will be used to guide the identification of regional climate controls for two Australian case study regions, one in NSW and the other in Queensland.

Chapter 5 Switch and Regional HMM Case Studies

5.1 Introduction

In this chapter the HMM, Switch HMM (SHMM) and Regional HMM (RHMM) are applied to several groups of sites located in Eastern Australia. The purpose of this exercise is to determine which of the model structures is in some sense the 'better' model. The main tool used to assess each model is Bayesian model selection (BMS).

The data used is introduced, with two sets of four sites (Close and Far), located around Sydney and Brisbane being the foci of the case studies. Sydney and Brisbane were chosen as the centres of these regions due to the previous identification of inter-annual persistence in these areas.

Each of the full models are applied to the data sets. These initial tests on the HMM, the full SHMM (all sites with switch probabilities) and the full RHMM (all sites in their own climate region) are used to gain an appreciation of the parameter identification issues associated with each model. This provides an interpretive framework when all of the model variants are tested.

BMS is applied to all model variants of Switch and Regional HMM, from which model averaged state series are produced. These state series are used in calibration of the short timescale rainfall model DRIP in the following chapter.

Finally, modelling the small-scale correlation structure is considered in more detail with comparison of the Zero, Empirical, Exponential Decay and Fitted correlation coefficient structures undertaken. In particular, the Exponential Decay structure is examined as a possible successor to the Fitted Correlation approach. A nugget effect term is also introduced to account for microscale variation observed within the data.

5.2 Data

There were several factors influencing the choice of sites for this study. The main factors were:

- Areas have previously exhibited evidence of inter-annual persistence.
- Areas have a sufficient amount of high quality annual rainfall data sets.

- Sites have previously been calibrated using the HMM by *Thyer* (2001) or *Srikanthan et al.* (2001) thus allowing comparison of results.
- Each site has an annual rainfall length which spans at least the length of the 6-min pluviograph records used to calibrate the DRIP model.

5.2.1 Key Sites: Sydney and Brisbane

In the studies of *Thyer* (2001) and *Srikanthan et al.* (2001), Sydney and Brisbane have shown evidence of long-term persistence, supporting the two-state hypothesis of the HMM model. Given that there exist long-term 6-min pluviograph records at these sites and that the long-term variability of these records is of interest (being large metropolitan areas), Sydney and Brisbane were chosen as the key sites in this study. The term 'key site' is used here to denote that they are the sites of most interest, with other sites surrounding the key sites being included essentially to identify the state series at that point more clearly. This definition of key sites would be irrelevant if the purpose of the study were solely to determine which sites should be grouped into the same climate region.

The models presented in this thesis have not yet been generalised to accommodate data commencing and finishing in different years. Hence the time-span of data used will be from the beginning of the latest starting time series, and the end of the earliest finishing. It is desired that the inferred state series at least span the pluviograph data to be used within the DRIP calibration. The length of 6-min pluviograph data for each key site is given in Table 5.1.

Table 5.1 6-min Pluviograph data used in this study

Site No.	Name	Start	Finish
66062	Sydney RO	Jan 1913	Nov 1991
40214	Brisbane RO	Jan 1908	Dec 1991

5.2.2 Surrounding Sites: Close and Far

Sites surrounding the two key sites must now be chosen. It is not clear on what basis sites should be chosen in a HMM analysis. Indeed, this is a major motivating factor of this work. Choice of sites too far from one another could mean they do not belong to the same climate region, whereas, sites close to one another having high rainfall correlations may make it harder for extra information to be gained on the climate state.

Although, the two generalisations to the HMM are designed to give some flexibility in the choice of sites, a balance between these two extremes is desired.

Two groups of sites per key site were tested in this thesis: a group of sites within 200km of the key site (Close) and a group at greater than 200km (Far). This testing will thus permit assessment of the effect of distance on inference. The position of these sites is shown in Figure 5.1. Four sites in each grouping were used (as opposed to many sites) as it is believed that small test cases are required first to gain some understanding of the models introduced. Another factor is that the computation time for optimisation and sampling rises exponentially with the number of sites; the previous chapter detailed the possible rise in number of models and parameters with the number of sites. Note that RHMM neighbouring sites are indicated in Figure 5.1b, and were determined following the culling protocol defined in Section 4.6.2.

The annual rainfall data used in *Thyer* (2001) and *Srikanthan et al.* (2001) formed the majority of sites chosen in this study. Table 5.2 lists the sites chosen, along with their length and site grouping. Two columns indicate whether or not the data was used in the *Srikanthan et al.* (2001) or *Thyer* (2001) studies, while the final column indicates whether that site was listed as part of the high quality rainfall data set identified by *Lavery et al.* (1997). All data sets were used in these previous studies or identified in the *Lavery et al.* (1997) set with the exception of Caboolture (40038). Caboolture was included in the Brisbane Close set as there were insufficient sites used in *Thyer* (2001) or *Srikanthan et al.* (2001) that were sufficiently close to the Brisbane key site. The Caboolture record was derived from aggregated daily records spanning 1892-1997, with data quality control flags indicating there were no missing daily rainfalls in this period, with the exception of accumulated readings taken over periods of four days or less. These accumulated readings were particularly prevalent over weekends. As the data used here is an annual total from the daily data, this weekend accumulation is not expected to have any effect on inference. The resulting contiguous data length of each site grouping is reported in Table 5.3.

Of the surrounding sites chosen, Cape Capricorn LH (39023) is the only site with a shorter span than the associated key site pluviograph. This in turn means that not all of the available pluviograph data will be used in the DRIP Brisbane Far site calibration

next chapter. As it is only five years shorter, this was not considered as having sufficient impact on the DRIP calibration to warrant a more extended search for sites.

Table 5.2 Annual rainfall data used in this study

Site No.	Name	Start	Finish	Site Group*	Srikanthan Set	Thyer Set	Lavery Set
39023	Cape Capricorn	1900	1986	BF	Yes	No	No
40038	Caboolture	1892	1997	BC	No	No	No
40043	Cape Moreton	1870	1998	BC	Yes	No	Yes
40214	Brisbane	1860	1992	BC,BF	Yes	Yes	No
41082	Pittsworth	1887	1998	BC	Yes	No	Yes
42023	Miles	1885	1998	BF	Yes	No	Yes
54004	Bingara	1886	1998	SF,BF	Yes	No	Yes
62021	Mudgee	1877	1998	SF	Yes	No	Yes
63056	MtVic/Blackheath	1872	1993	SC	No	Yes	No
66062	Sydney	1859	1998	SC,SF	Yes	Yes	No
68045	Moss Vale	1871	1993	SC	No	Yes	No
69081	Moruya Heads	1876	1998	SF	Yes	No	Yes
70080	Taralga	1883	1993	SC	No	Yes	No

*Note : SC/F-Sydney Close/Far BC/F-Brisbane Close/Far

Table 5.3 Contiguous data lengths for Sydney and Brisbane Close/Far site groupings

Key Site	Group	Start	Finish	Length
Sydney	Close	1883	1993	111
	Far	1886	1998	113
Brisbane	Close	1892	1992	101
	Far	1900	1986	87

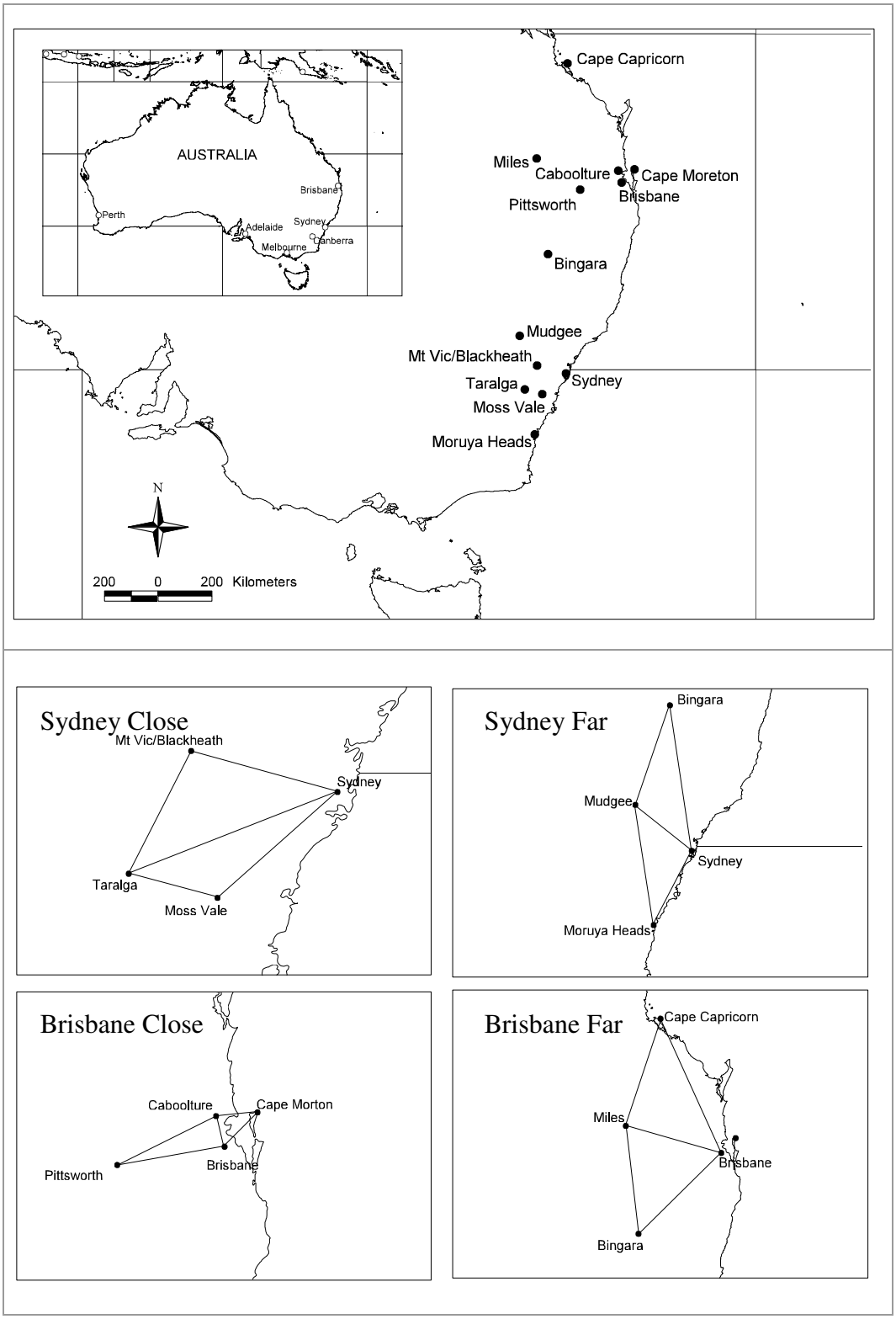


Figure 5.1 Case study (a) site positions and (b) site groupings with RHMM neighbouring sites indicated.

5.2.3 Water Year: Choice of starting month

The annual data used here is derived from aggregated monthly data at each site. A question arises: what month to use as the start of the water year? *Thyer* (2001) used an index, the SSI, to indicate *post* calibration which of the 12 months as the start of the water year gave the strongest state identification. This index, along with another measure of state separation, the *WADSI* was used to determine if changes in the water year had any influence on persistence identification. *Thyer* (2001) notes that for his multiple site analysis, the starting month did not have a significant effect on the persistence structure. In the single site analysis over 40 sites spread across Australia, *Srikanthan et al.* (2001) could not identify any 'noticeable pattern in the starting months'. With little constraint on the water year starting month, the May-April water year was selected on the grounds that ENSO events tend to break by the end of Austral autumn. The start and finish times for the data series in Table 5.2 correspond to this water year, with a series spanning 1900-1986 starting in May 1900 and ending in April 1987.

5.3 Application of the HMM, Switch HMM and Regional HMM

The HMM, full SHMM and full RHMM are now applied to the four sets of data. Within each grouping, posterior parameter plots, posterior state series and posterior model weight distributions are compared. The aim of this analysis is to gain an understanding of each model. This understanding will then be used to interpret the calibration of all model variants.

For all results in this chapter, 200000 posterior samples (5 MCMC paths × 40000 samples) were generated for each model. This number of samples was determined to be adequate based on the 'scale reduction' convergence diagnostic of *Gelman and Rubin* (1992) presented in section 3.3.2 being below 1.2 for all parameters. The multiple number of paths allowed use of this diagnostic, along with standard visual inspection of chain mixing. An equal number of burn in samples were also simulated. The burn in samples were also required to show a scale reduction score being below 1.2 to ensure that starting effects were eliminated.

5.3.1 Sydney Close

The Sydney Close data set consists of MtVic/Blackheath, Sydney, Moss Vale and Taralga. The posterior transition and switch probability distributions are plotted for the HMM, full SHMM and full RHMM in Figure 5.2. The applicable parameters are organized into three columns, each pertaining to a particular model. A single set of regional transition probabilities is given in the first column for the HMM. The regional transition probabilities and four sets of switch probabilities (associated with each site) are in the second column for the SHMM. Column three compares four sets of regional transition probabilities (associated with each site) for the RHMM. These plots are bivariate histograms derived from the MCMC samples, with the shading denoting the bin count. The shading scale gives the range of bin counts for that particular plot, with differing maximum levels depending on the spread of the parameters. Figure 5.3 presents the posterior state series for all three models.

We firstly compare the regional transition probabilities identified by each model. The SHMM shows the least uncertainty, with a dense mass located near the origin. Note the parameter constraint $p(r_t = D \mid r_{t-1} = W) > p(r_t = W \mid r_{t-1} = D)$ was imposed for the SHMM. The ordinary HMM identifies a less dense cloud, with much greater variation shown on $p(r_t = D \mid r_{t-1} = W)$. There is some overlap for the SHMM and HMM transition probability distributions, however, the mean $p(r_t = D \mid r_{t-1} = W)$ locations differ markedly, with a greater mean for the HMM. These transition probabilities can be interpreted in terms of the state series they produce. Focusing on the HMM and SHMM regional state series, denoted by a thick red line, there is generally a high probability of being in a dry state for the period 1900-1940 and a higher probability of being in a wet state for the period 1940-1985 for both models. However, the HMM shows a greater frequency of changing state. This frequency is controlled by the transition probabilities, hence the greater $p(r_t = D \mid r_{t-1} = W)$ values identified for the HMM model.

Now we consider the site specific state series and transition probabilities of the RHMM. The MtVic/Blackheath and Moss Vale transition probability plots for the RHMM model give similar distributions to the HMM model. Sydney shows a greater degree of variability, with the $p(r_t = D \mid r_{t-1} = W)$ parameter mean being greater than for the

HMM or SHMM, yet it overlaps those MtVic/Blackheath and Moss Vale. Taralga has the most poorly identified transition probabilities, with a centre of mass well away from the origin, and significant spread across the entire parameter space. Indeed the evidence suggests that at Taralga the HMM degenerates to a mixture model. The site specific state series of the RHMM echo these transition probabilities. The Taralga series dithers around the 0.5 probability mark, without any clear identification of periods being in one state or another. In contrast, MtVic/Blackheath and Moss Vale identify state series clearly, with little dithering, producing dense posterior transition probability distribution clouds. The Sydney state series, lies between the two extremes of identification, with some general trends identifiable, yet also containing a large degree of dithering. Of significance is that although the transition probability clouds of the MtVic/Blackheath and Moss Vale series are very similar, the state series are quite dissimilar. If state series are similar, then transition probabilities will be similar. However, the reverse does not necessarily apply. The reason is that an identified state series is influenced by the temporal occurrence of the rainfall, whereas the transition probabilities remain constant over time. To demonstrate, consider two sets of data $\left\{ y_t^1, y_{t+1}^2 : t = 1,...,T \right\}$, the second set being an identical copy of the first, yet occurring one time-step after the equivalent first set. Each set fitted individually with the HMM would produce identical transition probabilities, yet the state series would be offset by one year.

The site specific switch probabilities of the SHMM define the degree to which the individual site state series follow the overall controlling regional Markov series. The regional state structure identified by the model is followed closely by the MtVic/Blackheath composite during dry regional periods. The Moss Vale state series does not follow the regional series closely, especially when the regional series approaches a high probability of being in the wet state. Conversely, Sydney does not correspond to a dry regional state well, although during wet regional times it follows the regional state most closely. These differences are explained by the switch probabilities. The MtVic/Blackheath $p(s_t = W \,|\, r_t = D)$ distribution is closest to the origin, while $p(s_t = D \,|\, r_t = W)$ shows more variation and is further away from the origin. The posterior cloud furthest from the origin occurs for Moss Vale, with both switch

probabilities being significantly greater than zero. Sydney identifies a low probability of switching from the wet state, and the highest probability of switching from the dry state.

The application of the SHMM to the Sydney Close data demonstrates its ability to identify an overlying temporal Markov structure like the HMM, yet allows sites to vary from this overall control. Of the four sites, Moss Vale is found to differ from the regional state control to the greatest degree in years where there is a high probability that the regional state is wet, whilst also displaying a poorly identified state series during dry regional years.

Interpretation of these results can be strengthened when all state series and parameter plots are compared, with the RHMM state series shedding some light on the HMM and SHMM results. The HMM regional state series and associated transition probabilities are well identified, showing the greatest similarity to the MtVic\Blackheath RHMM series. Naturally, the individual RHMM series show more variability than the HMM, as there is no link between sites at the Markovian state level of the hierarchy. The HMM averages out differences in state variability resulting in a smoother regional state series with sites that cannot identify RHMM state series (e.g. Taralga) having little effect on the overall HMM calibration, whilst the strongly identified MtVic/Blackheath and Moss Vale state series influence the HMM series to the greatest degree. The SHMM results suggest that the MtVic/Blackheath record is the major influence on the regional state series (small switch probabilities), and also that the Sydney and Taralga state series are more aligned with the MtVic/Blackheath series overall. This implies that the MtVic/Blackheath series shows the strongest evidence of a two state Markovian structure.

The *WADSI* introduced in Section 2.6.1 is a measure of how well separated are the wet and dry Gaussian rainfall distributions for each site. A *WADSI* with little posterior probability at zero denotes a well separated distribution. Based on synthetic data studies, *Thyer* (2001) recommends that generated data with *WADSI* values of 1 or greater is required for identification of a two-state HMM persistence structure, given 140 years of data, using the single site HMM.

The *WADSI* distributions for the three models are shown in Figure 5.4. All site *WADSI* distributions for the HMM are significantly greater than zero, indicating well separated distributions. The SHMM results are well separated from zero, with the exception of Sydney. RHMM sites with strongly identified individual state series (MtVic/Blackheath, Moss Vale) show the greatest separation from zero. Overall, the HMM has greater mean values than SHMM and RHMM, whilst also having the greatest parameter variance. To some extent, these results agree with the results in *Thyer* (2001) who applied the multiple-site HMM to MtVic/Blackheath, Moss Vale and Taralga (along with two other sites not used here – Yarra composite and Cataract Dam). In that study, Moss Vale was found to have the least probability of exhibiting a two state structure according to the *WADSI* . Within this study, the RHMM Moss Vale state series identifies a strong state structure, however it differs significantly from the dominant MtVic/Blackheath state series. The strongly identified MtVic/Blackheath state series swamps the common climate state across all sites of the HMM, causing Moss Vale data to be misclassified as being wet or dry. This in turn results in the state separation decreasing for Moss Vale, with a *WADSI* distribution with significant probability near zero.

Finally, the models are compared using posterior model probabilities in Table 5.4. The HMM has the greatest posterior weight, being approximately twice the weight of the SHMM, with the RHMM having a marginally lower weight than the SHMM. The introduced complexity of the HMM generalizations is apparently not justified in this case, with the simplest model being sufficient to capture the interannual persistence. This result suggests that an overlying regional structure is present, as opposed to conditionally independent Markov state series at each site. The well identified transition probabilities, and state series of the HMM go some way to explaining this result. The RHMM does not identify the Sydney and Taralga state series well, with the state series resembling noise, and poor identification of the transition probabilities ineffective. Recall that if the transition probabilities satisfy (3.43), $1 = p\left(r_t = W \mid r_{t-1} = D\right) + p\left(r_t = D \mid r_{t-1} = W\right)$, the HMM model degenerates to the two-state independent mixture model. Taralga shows the most significant probability mass about this line, while the majority of the Sydney distribution lies below this line. Similarly to the RHMM, the extra complexity of the SHMM is not justified here. A

possible reason for this SHMM result is that the majority of sites are not significantly different from the regional state series to justify these extra switch parameters. This concurs with the RHMM result that an overall regional control is influencing all sites.

These results do not imply that less complicated SHMM and RHMM variants are not justified compared to the HMM. Further comparison of the individual variants will determine this. This testing does however indicate strong evidence that Markov persistence can be identified over this region, in particular in the MtVic/Blackheath series. The RHMM and SHMM results indicate that not all sites are influenced to the same degree each year. However, BMS comparison reveals that the simplicity of modeling all sites under one regional controlling series outweighs the complexity introduced by these models.

This section has discussed in detail the Sydney Close results. This detail was primarily aimed at demonstrating the link between posterior transition and switch probability parameters, posterior state series and posterior model probabilities. The remaining data sets will not be discussed in as much detail. However, a comparison between sites and an overall discussion of this analysis will be given at the end of this section.

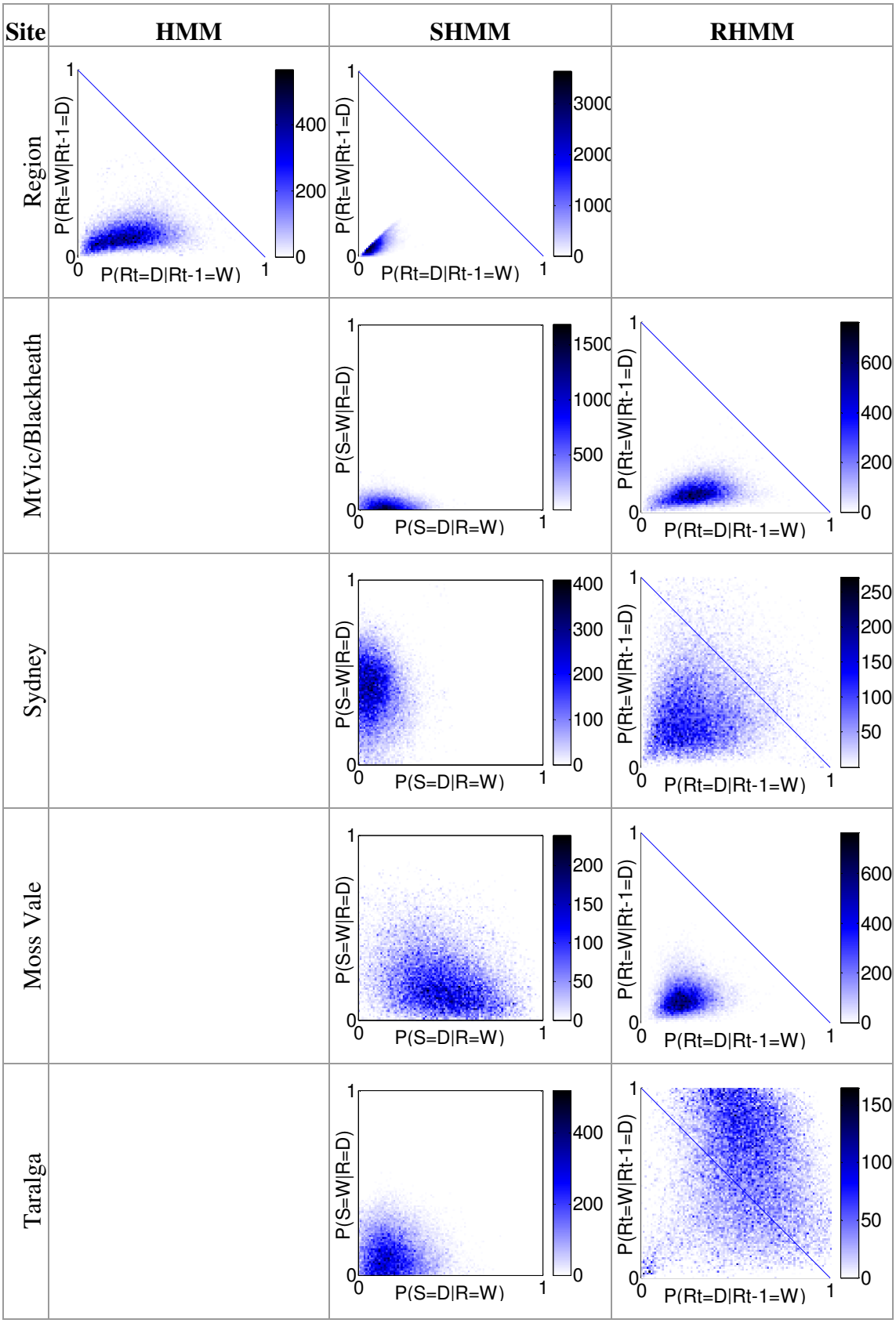


Figure 5.2 Sydney Close Sampled Posterior distribution of Transition probabilities and Switch probabilities for (a) HMM, (b) Switch HMM and (c) Regional HMM.

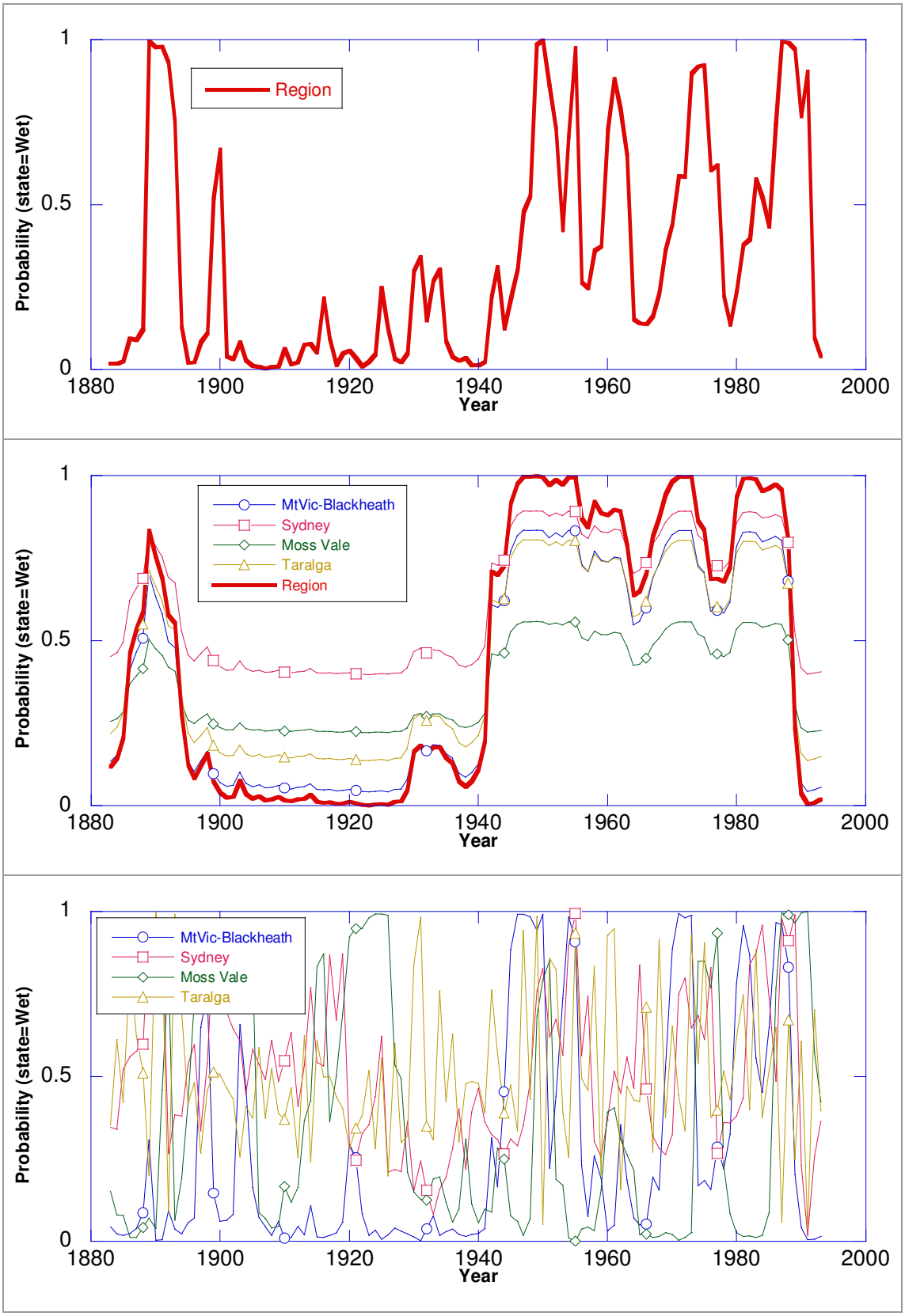


Figure 5.3 Sydney Close Posterior State Series probabilities for (a) HMM, (b) Switch HMM and (c) Regional HMM.

Table 5.4 Sydney Close posterior model probabilities

Model	Posterior Weight
HMM	0.5333
Switch HMM	0.2884
Regional HMM	0.1783

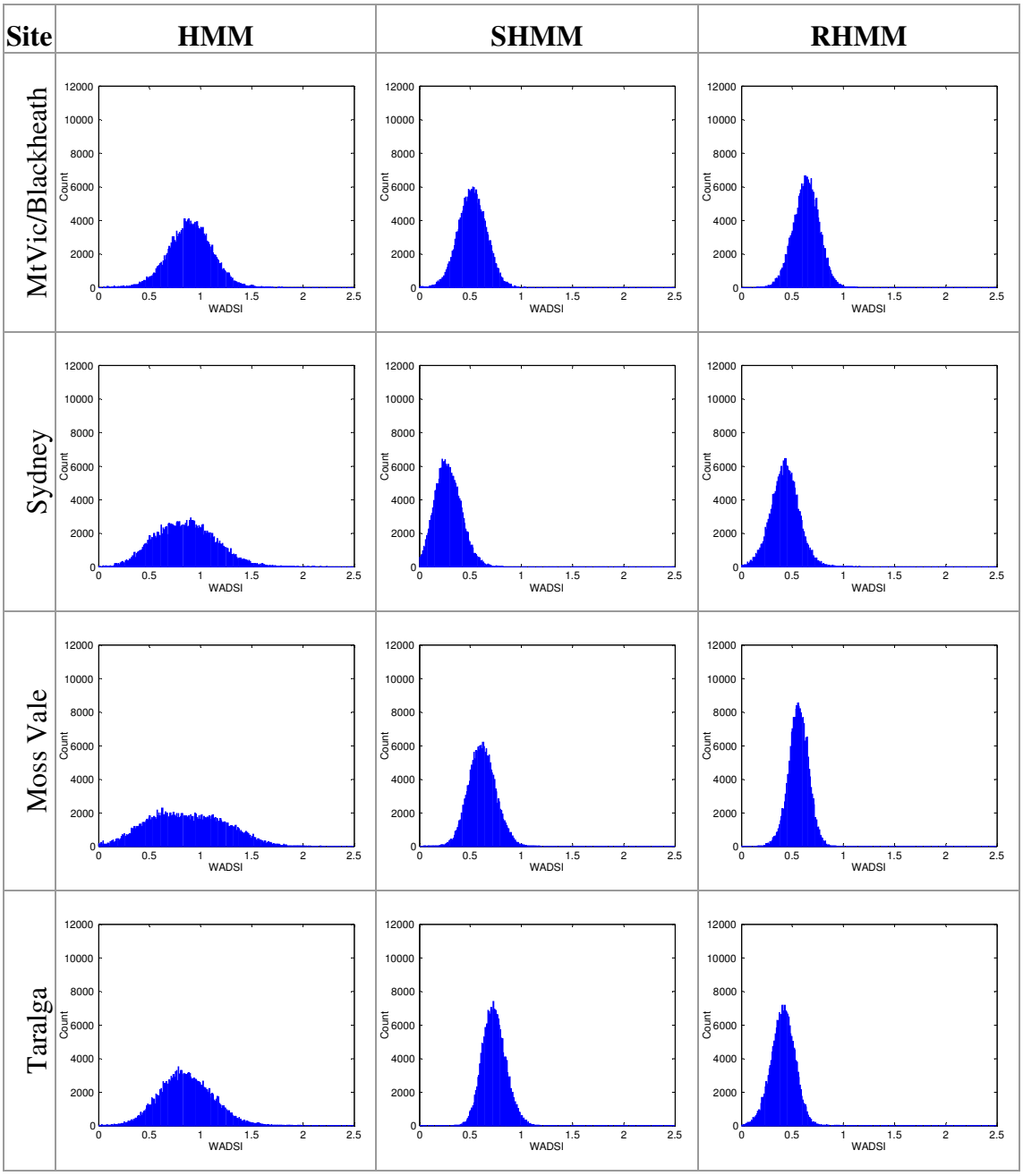


Figure 5.4 Sydney Close Sampled Posterior WADSI distribution for (a) HMM, (b) Switch HMM and (c) Regional HMM.

5.3.2 Sydney Far

The Sydney Far sites consist of Bingara, Mudgee, Sydney and Moruya. Bivariate histograms and plots of state series are shown in Figure 5.5 and Figure 5.6 respectively.

The HMM identifies a symmetrical transition probability cloud, located around the $p(r_t = D \mid r_{t-1} = W) = p(r_t = W \mid r_{t-1} = D)$ line. The HMM state series is reasonably well identified (little dithering around 0.5), with approximately equal time spent in each state. There is less persistence identified here than in the Sydney Close HMM series. Nonetheless, the overall 1900-1940 dry and 1940-1985 wet trend identified in the Sydney Close calibration remains discernible.

The SHMM calibration produces more variable transition probabilities than Sydney Close, with the imposition of the parameter constraint influencing the resulting state series. The posterior state series is similar to the HMM, however the states are not identified with the same strength. The switch probabilities generally have density clouds centred around the origin, yet are well spread over the parameter space. The individual state series do not show any significant trends, with all series (with the exception of Bingara) around the 0.5 mark. Bingara follows the regional state series more closely during dry regional state due to the $p(s_t = W \mid r_t = D)$ parameter distribution being located near zero. The SHMM had difficulty identifying an overall controlling Markovian structure showing dithering around 0.5. The switch parameters for nearly all sites indicate that each of the sites need to differ from the identified regional control occasionally, however these switch parameters are well spread reflecting the poor identification of the overall controlling regional series.

The individual state series of the RHMM are extremely variable, with no noticeable correlation between states at sites. Of the four sites, Bingara dithers closest to the 0.5 probability mark. Mudgee identifies a relatively wet state series, while Sydney and Moruya show frequent switching between the two probability extremes of 0 and 1. Consequently, Mudgee shows the most well identified transition probabilities, with Sydney and Moruya identifying clouds centred around (0.5,0.5). Such parameter clouds, lying around the line $p(r_t = D \mid r_{t-1} = W) = 1 - p(r_t = W \mid r_{t-1} = D)$ are effectively allowing the Markovian structure to degenerate to a simple mixture model, do not

justify the Markovian assumption. This differs from the HMM which did show some evidence to justify the Markovian assumption. Again, it seems the complexity of the full RHMM introduces a greater degree of uncertainty than the additional information gained on climate structure.

The *WADSI* distributions for the three models are shown in Figure 5.7. The distributions are quite similar for corresponding sites across all models, with the mean values greater than 0.5, and low posterior probability at 0. The exceptions lie with the HMM, with Bingara showing significant posterior probability at zero, and Moruya with a large mean value.

The BMS model comparison in Table 5.5 lists the SHMM as being around three times more likely than HMM or RHMM. This result can be interpreted as indicating that although a Markov structure can be identified (as in the HMM series), there is some additional variability not explained by a common climate state or by individual climate series. The climate structure is somewhere between these two extremes, here using the SHMM to allow some at site anomalies, yet following the overall (albeit weakly identified) regional control.

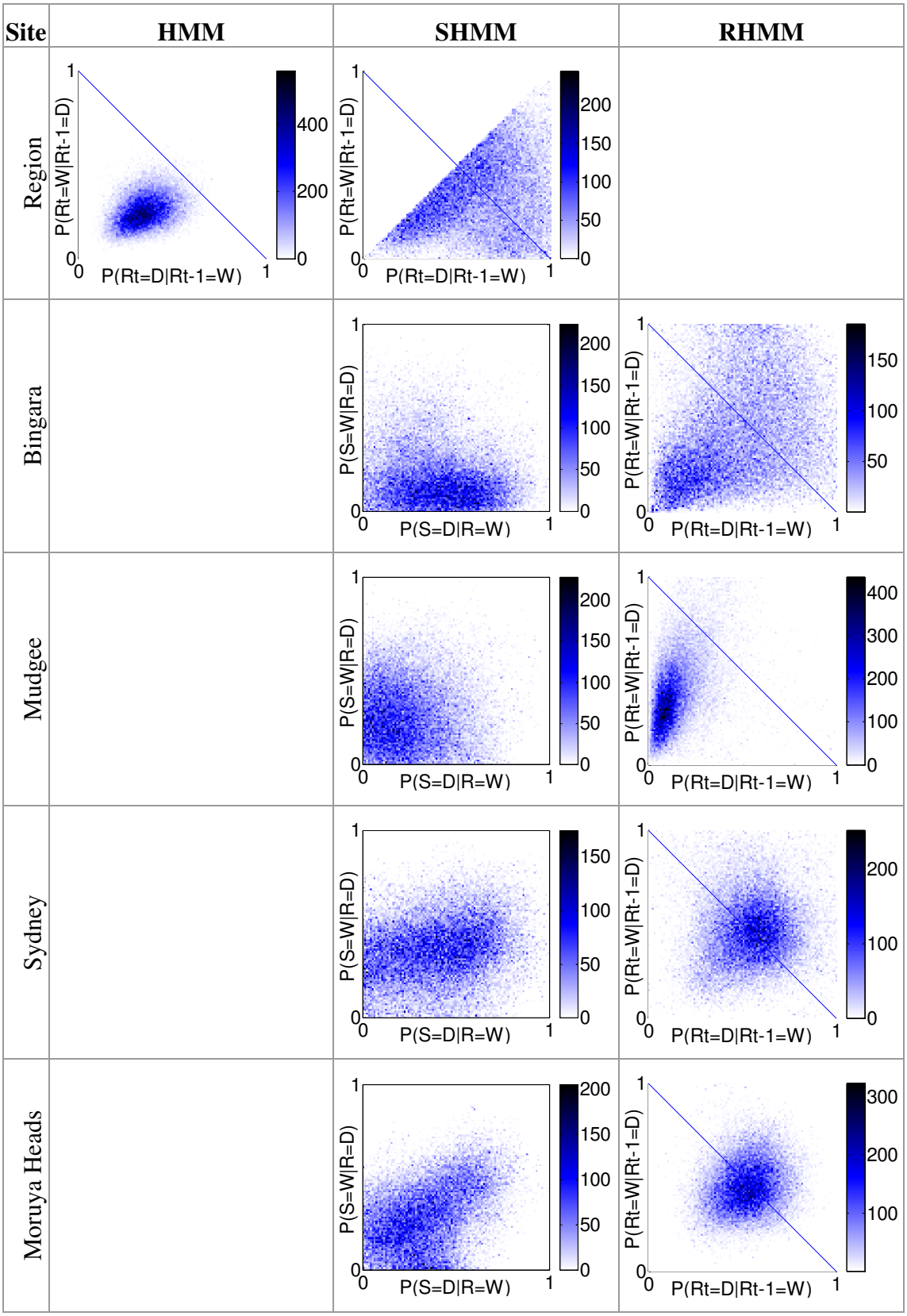


Figure 5.5 Sydney Far Sampled Posterior distribution of Transition probabilities and Switch probabilities for (a) HMM, (b) Switch HMM and (c) Regional HMM.

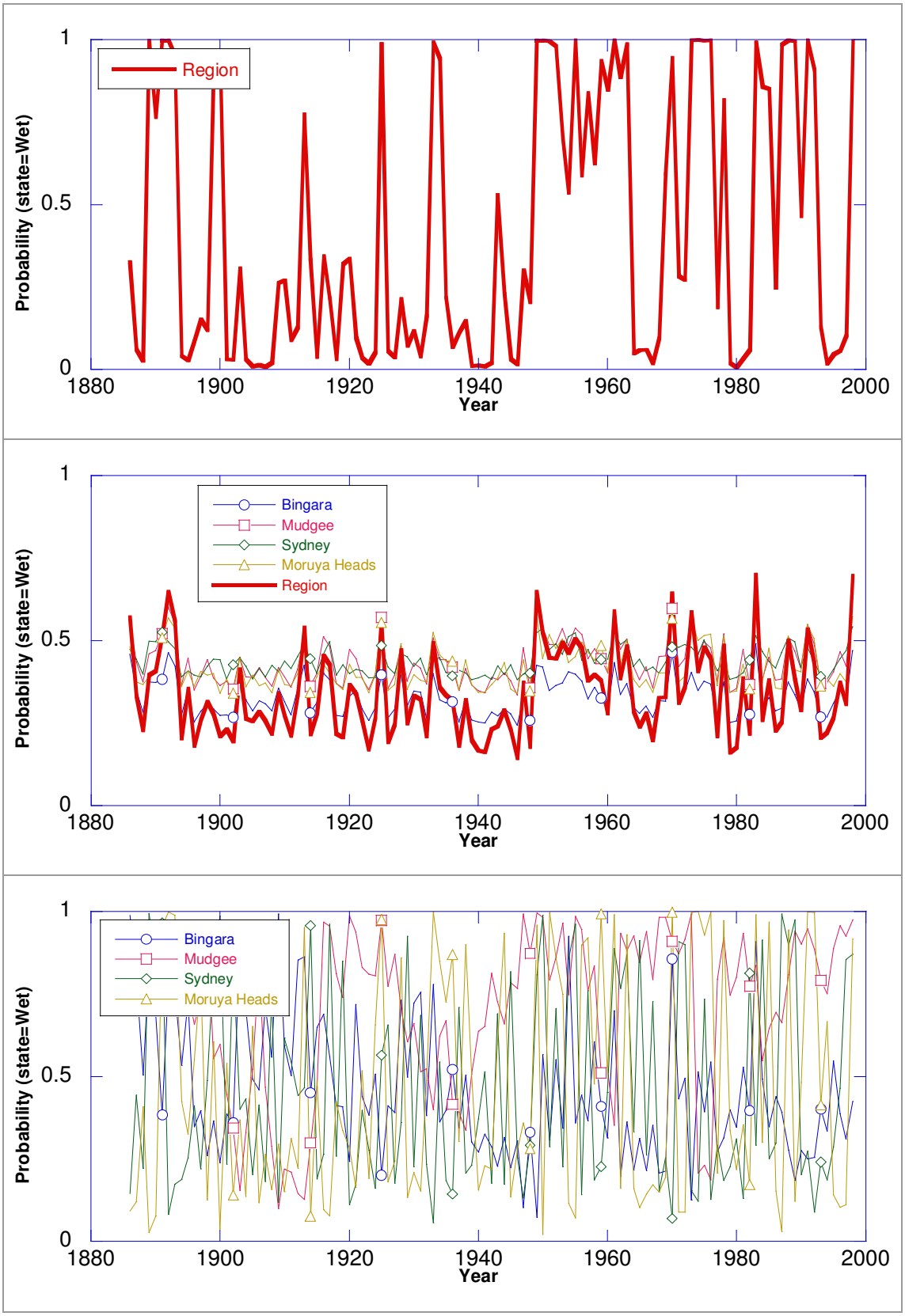


Figure 5.6 Sydney Far Posterior State Series probabilities for (a) HMM, (b) Switch HMM and (c) Regional HMM.

Table 5.5 Sydney Far posterior model probabilities

Model	Posterior Weight
HMM	0.2085
Switch HMM	0.6090
Regional HMM	0.1825

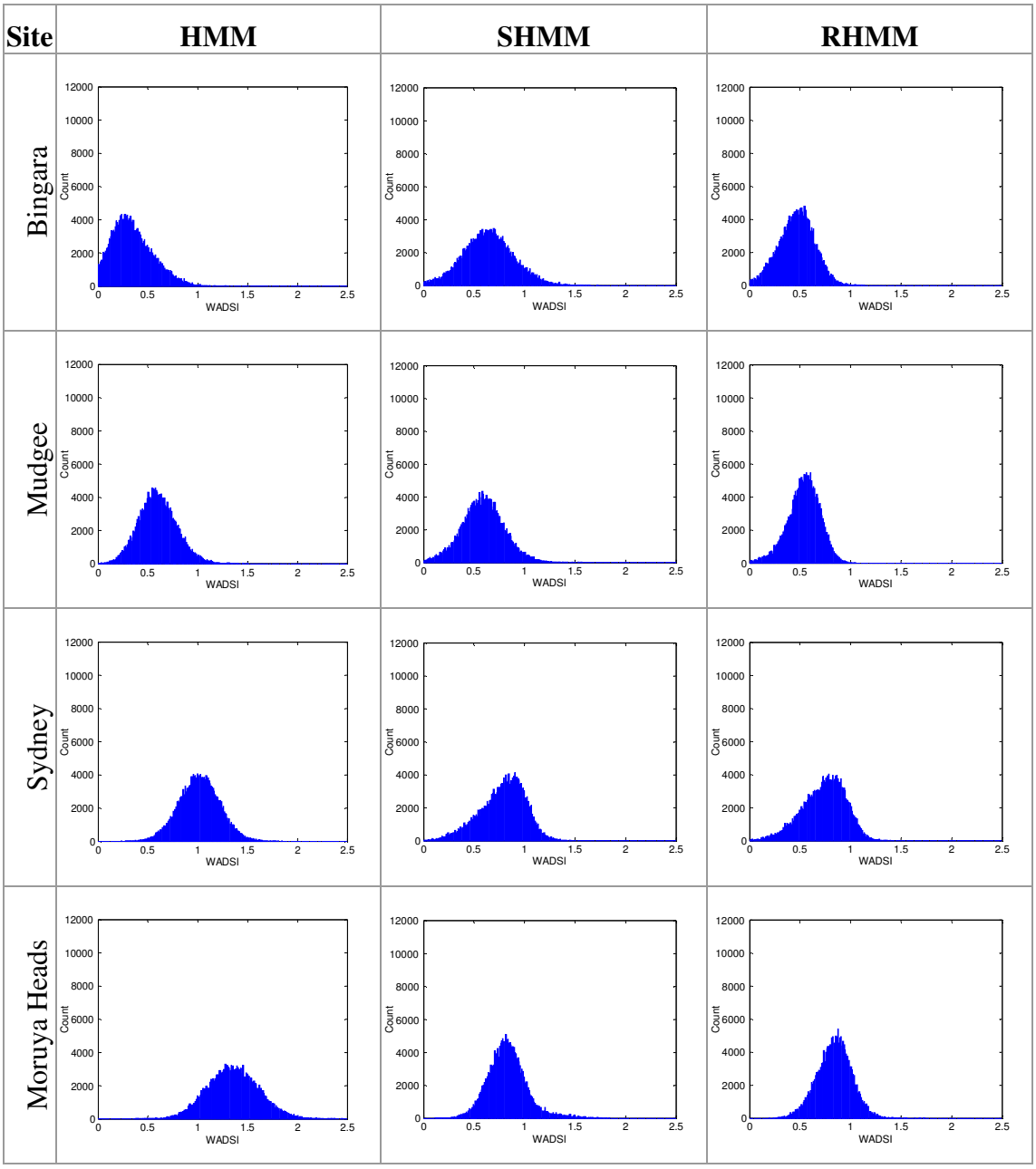


Figure 5.7 Sydney Far Sampled Posterior WADSI distribution for (a) HMM, (b) Switch HMM and (c) Regional HMM.

5.3.3 Brisbane Close

The Brisbane Close sites consist of Caboolture, Cape Moreton, Brisbane and Pittsworth. Bivariate histograms and plots of state series are shown in Figure 5.8 and Figure 5.9 respectively.

The HMM identifies a variable state series, with a predominantly dry series interspersed with short wet periods. The majority of the wet periods lie inside the 1940-1985 wet period identified in the Sydney Close and Far calibrations.

The SHMM state series shows a 1900-1960 regional dry period, with wetter periods either side. The SHMM individual state series for Brisbane follows the regional control reasonably closely, while Caboolture, Cape Moreton and Pittsworth remain closer to the 0.5 probability mark.

The individual RHMM series do not bear much resemblance to either of the HMM and SHMM regional series, with the exception of Brisbane. Brisbane appears to be showing the strongest evidence of persistence, with RHMM transition probabilities most similar to the HMM, and SHMM switch probabilities located near the origin. Pittsworth and Caboolture are poorly identified, while Cape Moreton is essentially wet except for three dry periods: 1940-1955, 1968-71 and 1977-1986. Cape Moreton's RHMM behaviour is anomalous when compared to the SHMM site series, which followed the predominantly dry regional state series reasonably closely. Curiously, though Caboolture is located close to Brisbane and Cape Moreton, it does not display a similar SHMM or RHMM site state series. It is thought this result may be due to anomalous interactions of the RHMM when high rainfall correlations are identified between sites. This topic is discussed in Section 5.5 where different methods of modelling spatial correlation are compared.

Figure 5.10 illustrates the *WADSI* distributions for the three models. Caboolture and Brisbane show zero posterior probability at zero for all models, while Cape Moreton and Pittsworth show poor separation generally. The best separation in terms of models is given by the RHMM, with *WADSI* distributions well separated from zero for all sites.

Table 5.6 gives the BMS weights for the three models, with the probabilities comprehensively favouring the RHMM. This result indicates that the data overwhelmingly prefers the RHMM, with different regions for each site. An explanation for this behaviour is that there is extra variability present that cannot be accounted for by the two other models. The RHMM has no inherent correlation of Markov states between sites, whereas HMM and SHMM do, thus allowing greater flexibility in producing variability in time series. The choice of the RHMM is due to small-scale spatial variation (modelled here by fitting individual correlation coefficients) being more dominant in the overall variation than the spatial correlation induced by sites being in the same state at the same time.

These sites differ from the other sites tested in that they are located closer to one another, especially Brisbane, Caboolture and Cape Moreton. It is expected this closeness (and similar distance from the coast) implies high correlations between sites, thus causing small-scale (and possible microscale) variation to have a higher proportionate influence on variability than in the previous data sets. This result is discussed further when compared against other methods for modelling small-scale variation in Section 5.5.

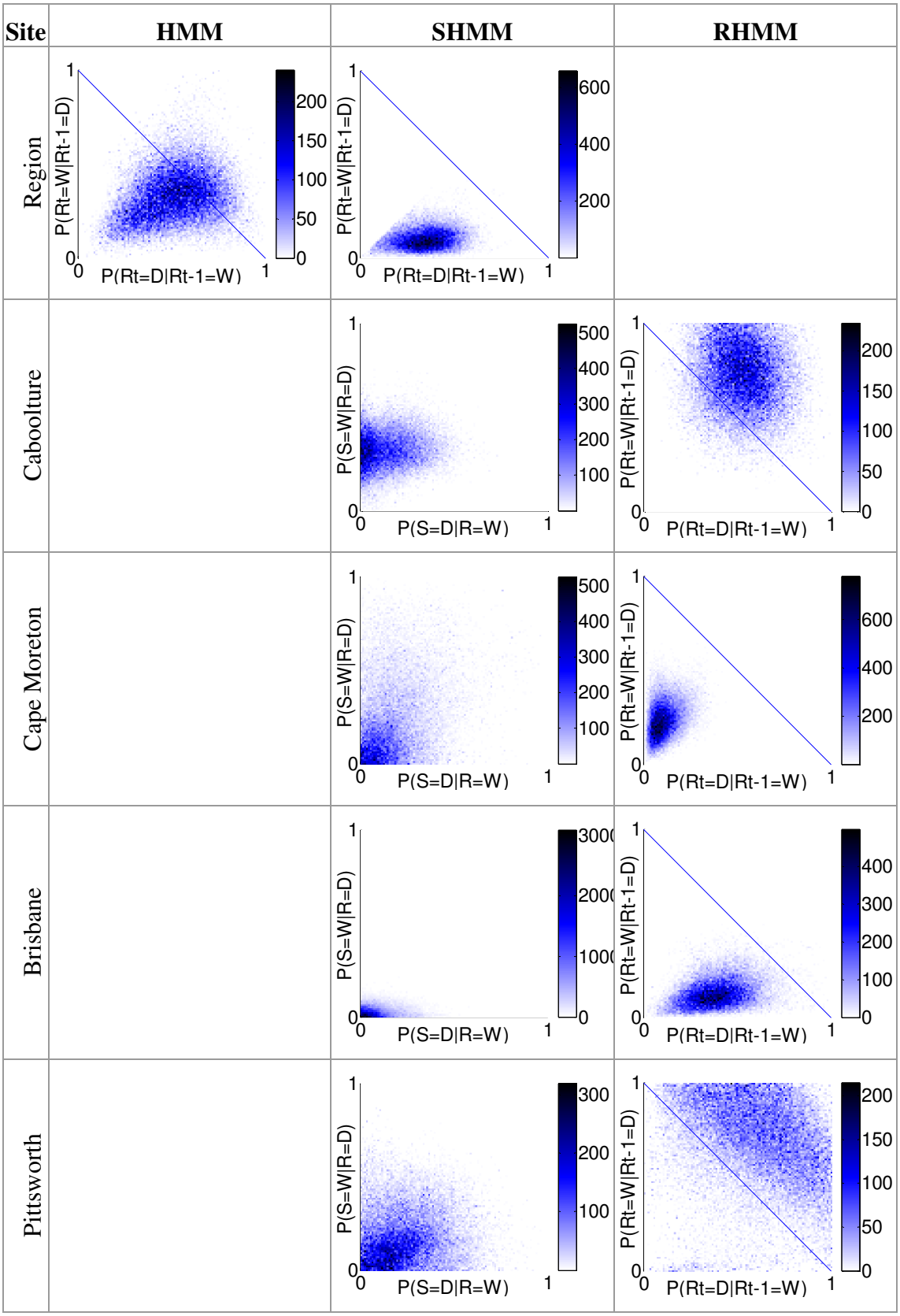


Figure 5.8 Brisbane Close Posterior distribution of Transition probabilities and Switch probabilities for (a) HMM, (b) Switch HMM and (c) Regional HMM.

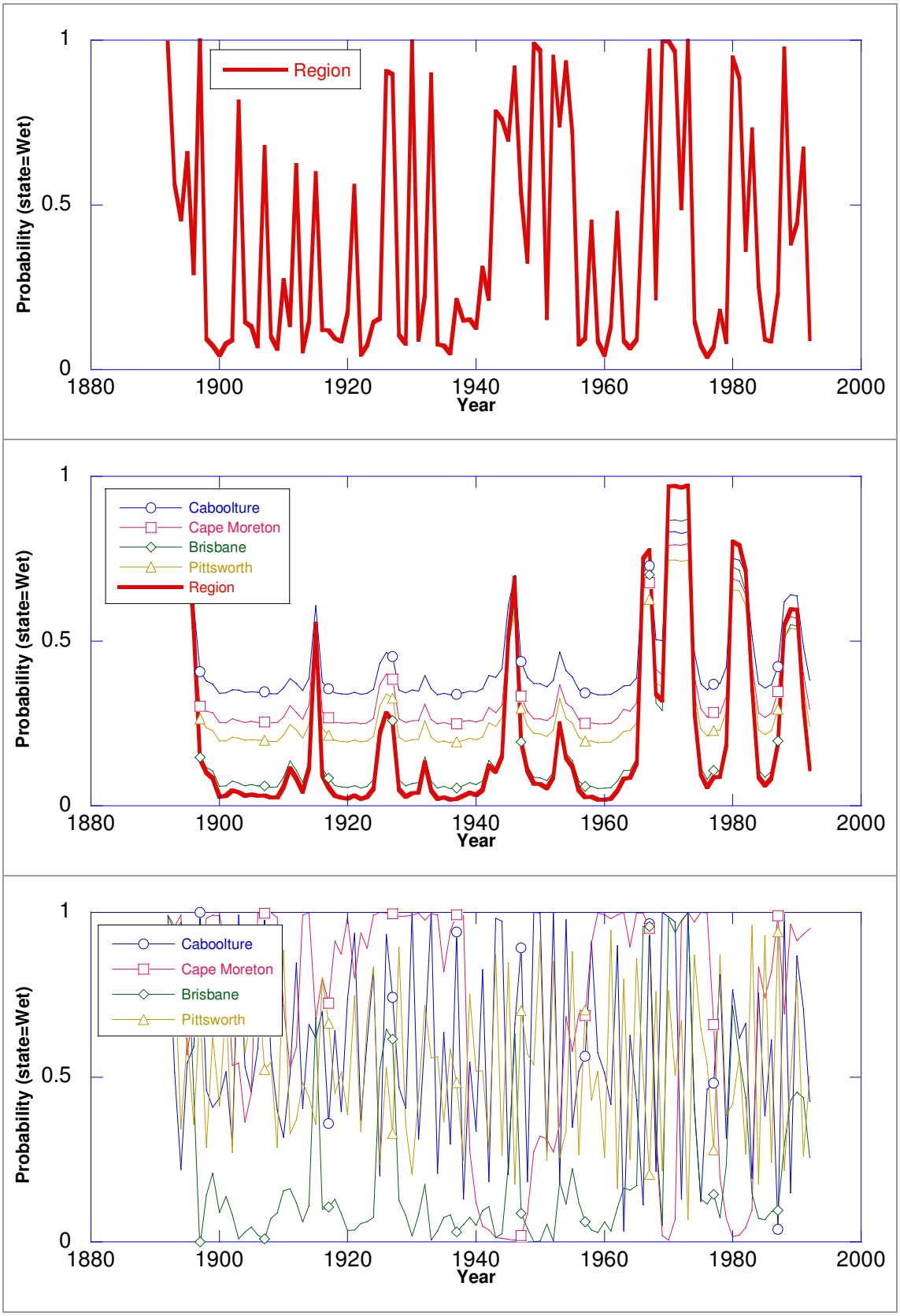


Figure 5.9 Brisbane Close Posterior State Series probabilities for (a) HMM, (b) Switch HMM and (c) Regional HMM.

Table 5.6 Brisbane Close posterior model probabilities

Model	Posterior Weight
HMM	0.0065
Switch HMM	0.0001
Regional HMM	0.9934

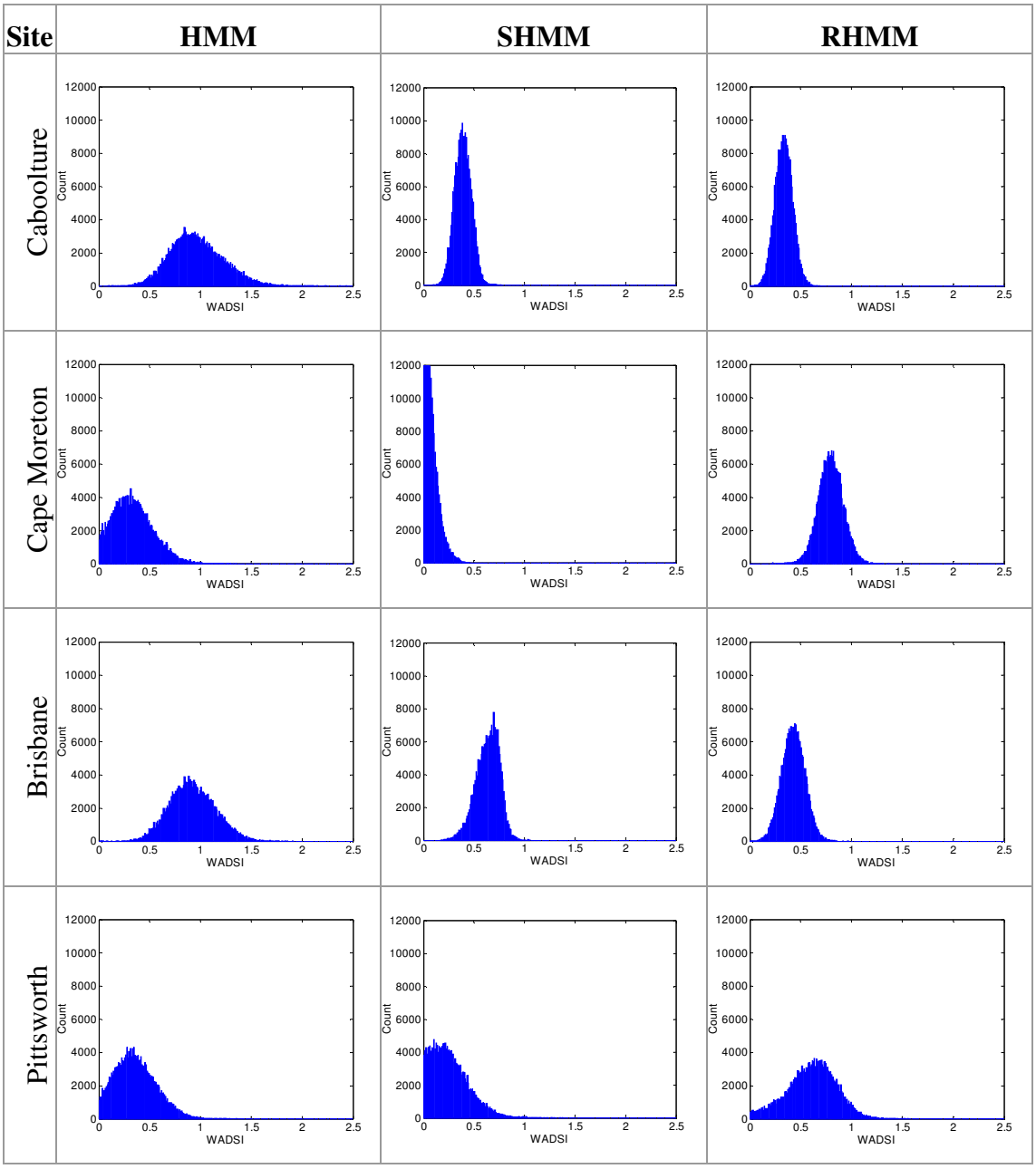


Figure 5.10 Brisbane Close Sampled Posterior WADSI distribution for (a) HMM, (b) Switch HMM and (c) Regional HMM.

5.3.4 Brisbane Far

The Brisbane Far sites consist of Cape Capricorn, Brisbane, Miles and Bingara. Bivariate histograms and plots of state series are shown in Figure 5.11 and Figure 5.12 respectively.

The HMM state series dithers considerably, with few years identified strongly. The associated transition probabilities are also poorly identified with a mass spread well over the parameter space. A significant proportion of the posterior cloud lies around the $p(r_t = D \mid r_{t-1} = W) = 1 - p(r_t = W \mid r_{t-1} = D)$ line indicating that the series could equally be modelled using a two state mixture model.

The regional state series of the SHMM identifies evidence of a Markovian structure, with a predominantly dry state series interspersed with a few wet periods of 10-20 years duration and with the transition probabilities showing less variability than that of the HMM. The Brisbane site specific state series follows the regional series closely. However, the remaining sites dither closely around the 0.5 mark. The Brisbane switch parameters are well identified near the origin, while the remaining sites, although being centred near the origin, show a much greater spread across the parameter space. The Brisbane state series of the RHMM is identified clearly, while the remaining sites dither around 0.5. The transition probabilities reflect this, with a strongly identified cloud for Brisbane, and the remaining sites showing diffuse posterior clouds. Thus the results suggest that Brisbane shows the strongest evidence for a two-state Markovian structure, while the other sites are providing little information on an overall controlling Markovian structure. The HMM state series suggest that if the sites are to be modelled under the one controlling structure, a two state mixture would be preferable; indeed it is possible that a single state structure may be preferable, at least for some sites.

Figure 5.13 illustrates the *WADSI* distributions for the three models. The Brisbane RHMM *WADSI* distribution is the only site with non-significant 0 posterior probability, showing the greatest degree of separation. The striking result here, in comparison to previous site grouping results, is the poor degree of separation, with a wide degree of variability displayed in the *WADSI*. Miles and Bingara show the greatest probability at zero, indicating these sites are very poorly identified. The

Brisbane HMM shows positive probability at zero, whereas this was not the case in comparison to the Brisbane Close HMM results. The more poorly identified sites are causing Brisbane to be less well identified.

To demonstrate further the possibility that these sites may be better modelled using a single state structure, a posterior bivariate histogram is plotted in Figure 5.14 for the mean parameters. There are four sets (each site) of mean (wet and dry) parameters for each model displayed. The label switching constraint $\mu_D^{site} < \mu_W^{site}$ is also plotted. The HMM posterior plots show significant density for all sites around the label switching constraint. Thus, the wet and dry means are not well separated from one another. The Brisbane and Cape Capricorn means show a greater degree of separation for SHMM and RHMM. The Miles and Bingara parameters remain in the same position for all models, close to the parameter constraint. *Srikanthan et al.* (2001) concluded for Brisbane, Cape Capricorn and Miles that the two state assumption of the single site HMM was 'possibly' justified, while Bingara was highly unlikely to exhibit persistence based on the *WADSI* . Visual inspection of the mean parameters in this study leads to the same conclusion, with the exception of Brisbane, which shows strong separation from the parameter constraint. The length of time series used in this group calibration (87 years) is significantly less than the length of individual site calibrations used in *Srikanthan et al.* (2001) (88,134,115 and114 years). Also in that study the water year showing the greatest *WADSI* for each site was months 7, 8, 2 and 3 respectively, while a water year starting in month 5 was used here. Given that such a short record was used, and also that we have not used the same water year, it is a testament to the RHMM's flexibility that the Brisbane state series was so well identified in the presence of other poorly identified sites.

Table 5.7 gives the BMS weights for the three models, with the probabilities again comprehensively favouring the RHMM. The probable reason for this choice is that the two state structure of the HMM is not justified at sites other than Brisbane (and possibly Cape Capricorn). Thus the identification of HMM and SHMM transition probabilities is biased by sites which should not be included in that single region. The RHMM on the other hand allows multiple regions, giving Brisbane the flexibility it requires to identify a strong state series, even with such a short amount of data.

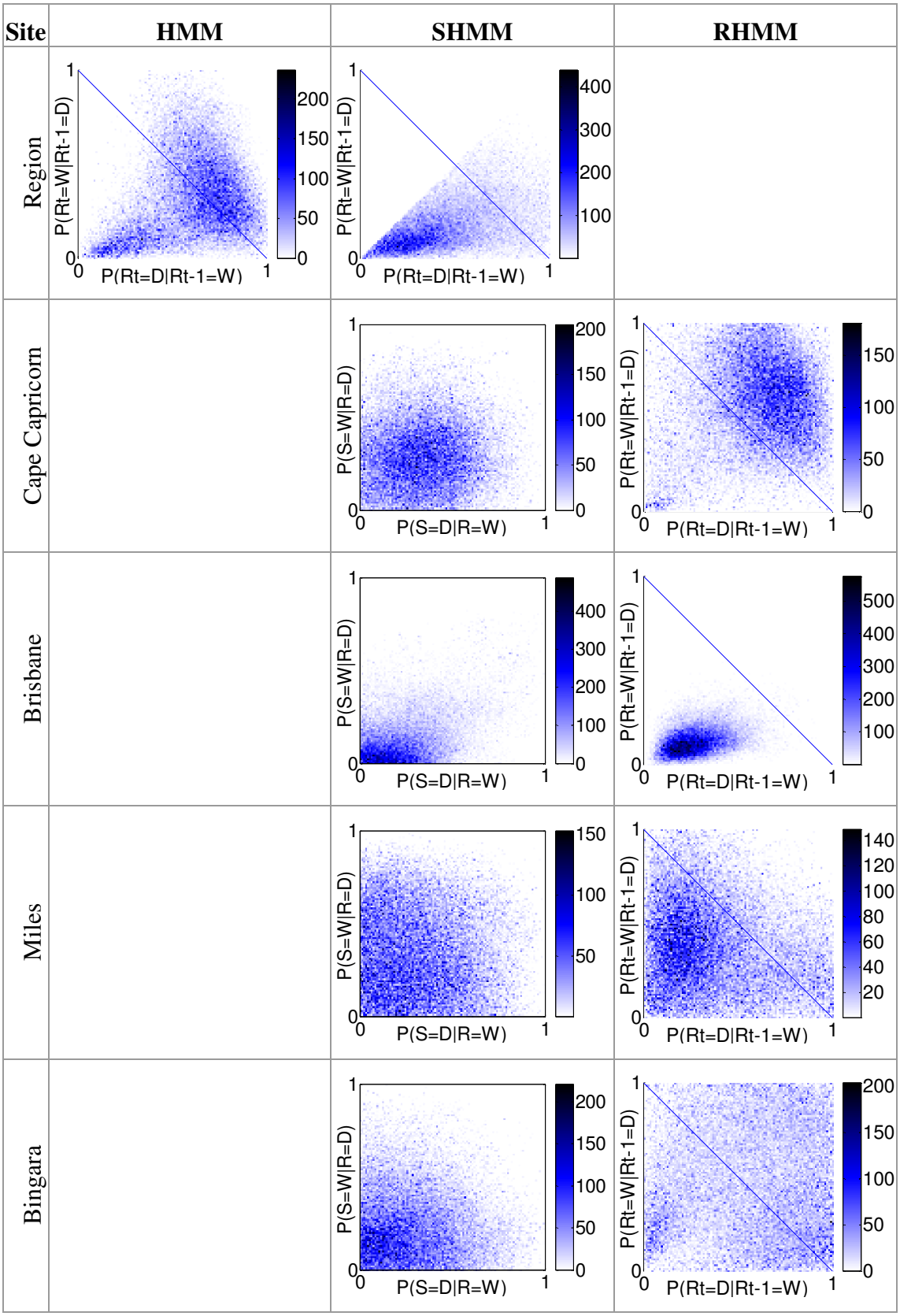


Figure 5.11 Brisbane Far Sampled Posterior distribution of Transition probabilities and Switch probabilities for (a) HMM, (b) Switch HMM and (c) Regional HMM.

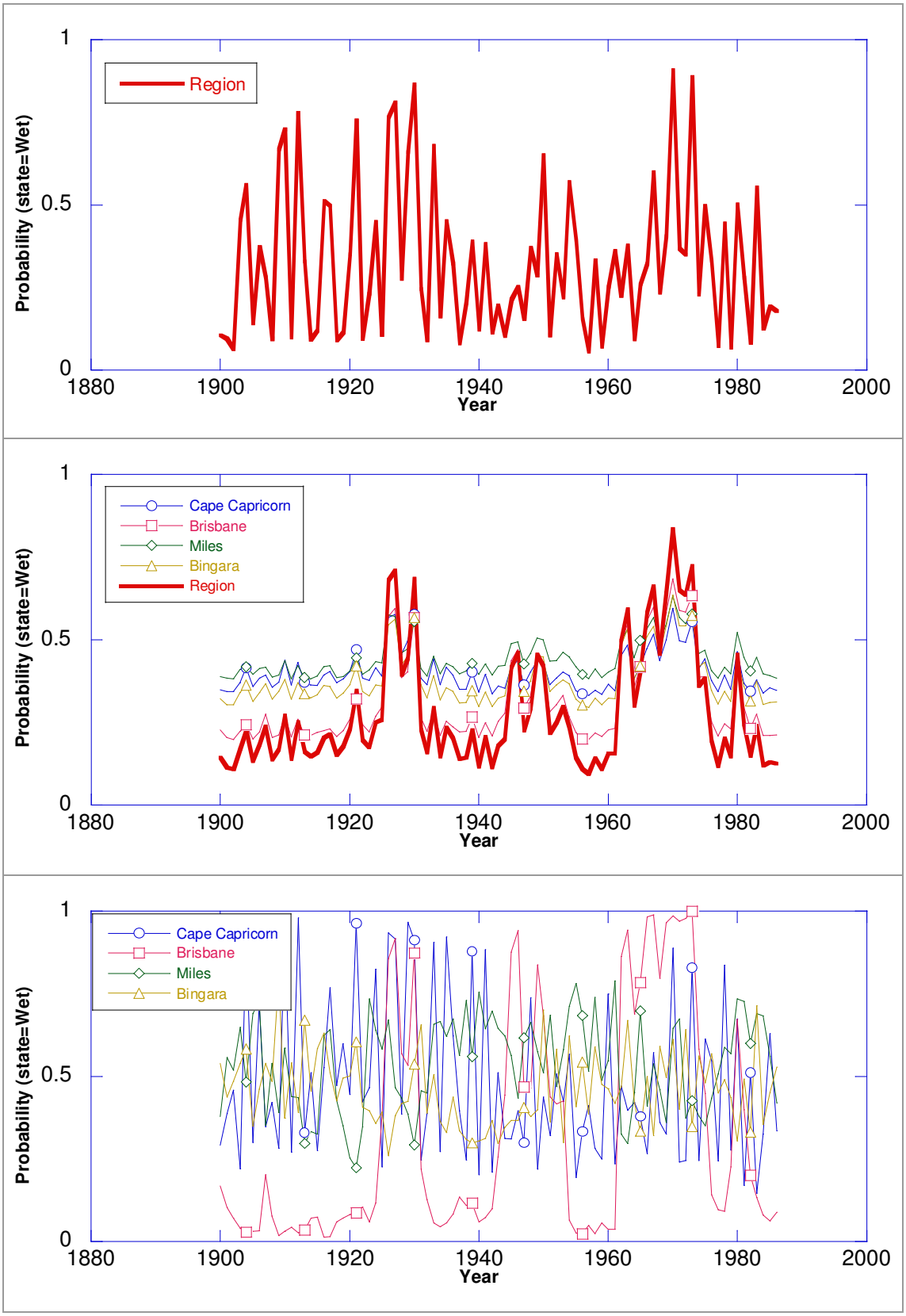


Figure 5.12 Brisbane Far Posterior State Series probabilities for (a) HMM, (b) Switch HMM and (c) Regional HMM.

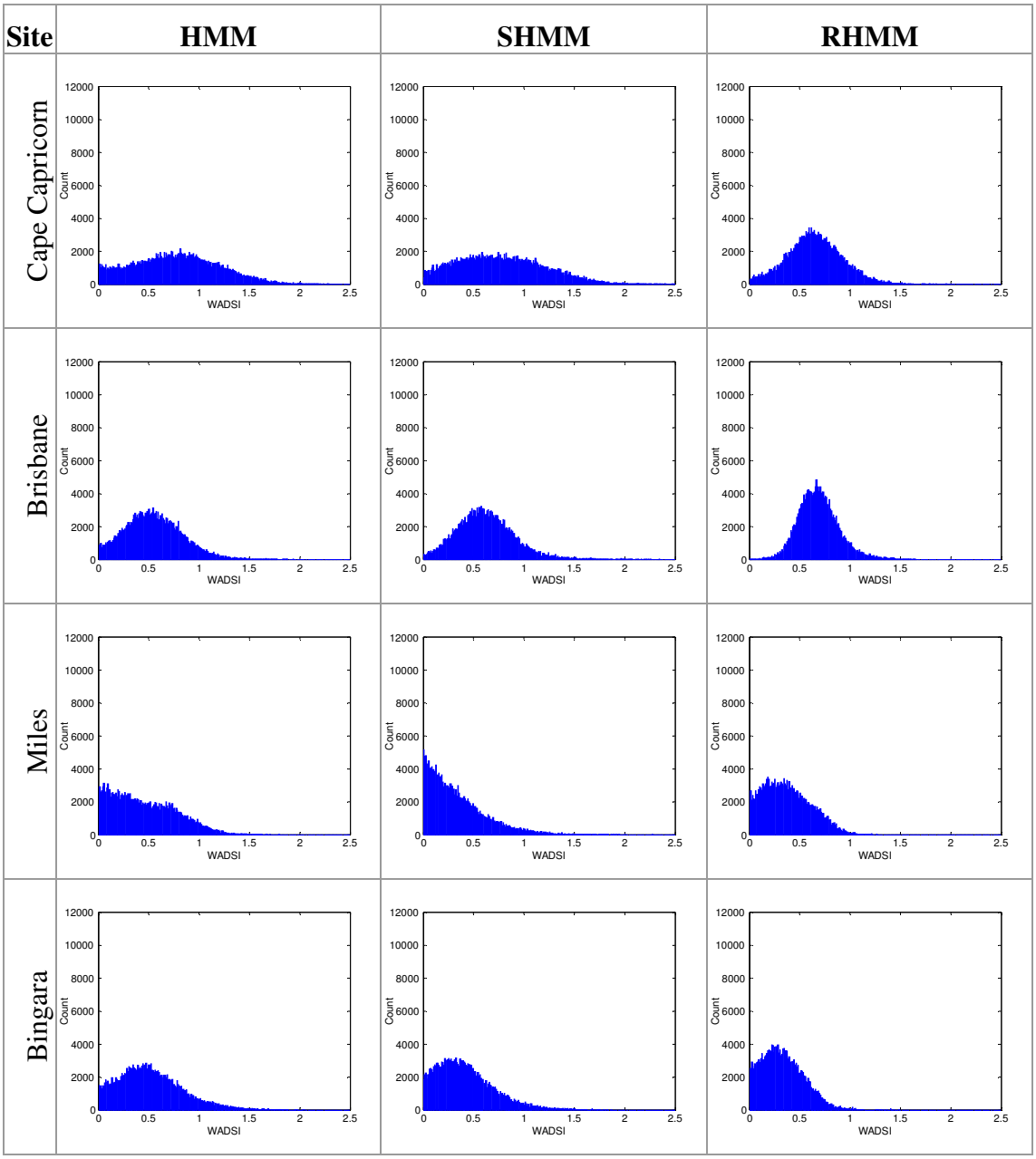


Figure 5.13 Brisbane Far Sampled Posterior WADSI distribution for (a) HMM, (b) Switch HMM and (c) Regional HMM.

Table 5.7 Brisbane Far posterior model probabilities

Model	Posterior Weight
HMM	0.0553
Switch HMM	0.0920
Regional HMM	0.8526

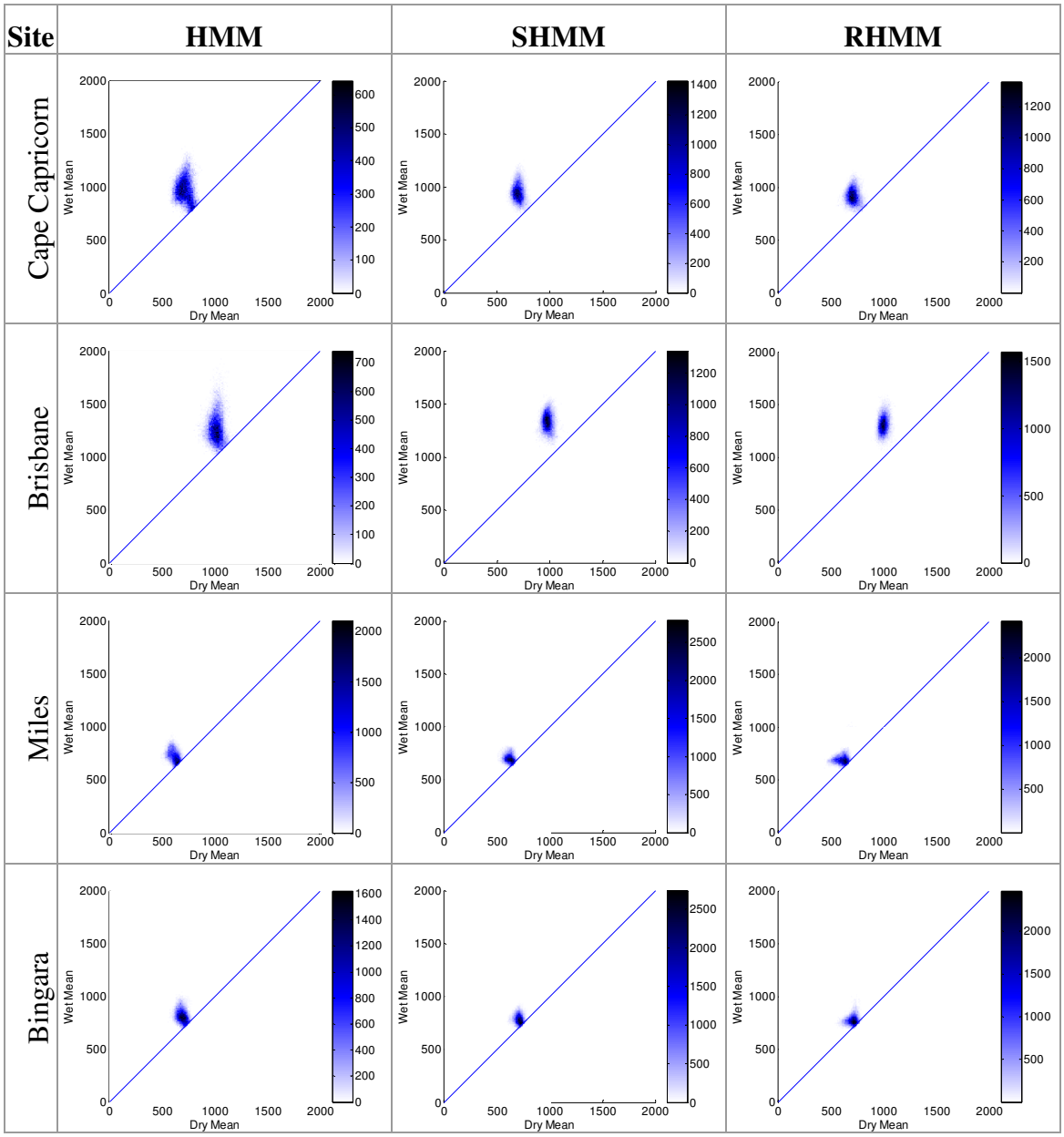


Figure 5.14 Brisbane Far Posterior mean distributions for (a) HMM, (b) Switch HMM and (c) Regional HMM.

5.3.5 Discussion of full model case studies

This section has presented application of the HMM, SHMM and RHMM to four case studies centred on Sydney and Brisbane. Different models performed better for different groupings. The HMM was superior for Sydney Close due to a strong state series being identifiable for the grouped sites. For Sydney Far the SHMM had the highest posterior probability due to it allowing more at-site variability than the HMM, yet still requiring an overall controlling structure. The RHMM performed best for the Brisbane Close sites presumably due to high correlations between sites being identified, and the state series

being used to add extra variability between sites. The RHMM was again found the better model for the Brisbane Far sites, but for a different reason, in that the sites other than Brisbane had poorly separated wet and dry means, with those sites probably better modelled with a single state structure.

Given that some sites/regions have identified a single state structure, a generalisation of this modelling framework to accommodate this would be to allow different numbers of states in different regions. This change would not be difficult to implement to the current model in terms of computer coding. However, the resulting number of models could become quite large, with longer computation required using the current model sampling technique. On the other hand, a model sampling technique such as Reversible Jump MCMC could circumvent the problem of calculating the marginal likelihood of each model. This suggests the strength of Reversible Jump in that many model combinations can be entertained; However, models are sampled at a rate proportional to their individual posterior probability. This contrasts with the sampling method used here where an equal number of parameters is sampled from each model (after initial optimisation) and marginal likelihoods are calculated using the posterior samples.

Overall, the two models, SHMM and RHMM, have served their purpose in identifying homogeneous climate regions. The SHMM shows clearly when a site differs from an identified regional state series (eg. Moss Vale). The RHMM confirms the results of the SHMM, often indicating that an anomalous site identified by the SHMM should be grouped into its own climate region.

5.4 Comparison of Switch and Regional HMM variants

Now that general relationships between the parameters, state series, *WADSI* and models have been recognised at each site, we will use this understanding to interpret comparison between the SHMM and RHMM variants. The term variant denotes all of the possible models ranging from the HMM to the full SHMM and RHMM. Each of these variants has a label with the HMM having the label $(\mathrm{F,F,F,F})$ or $(1,1,1,1)$ depending on whether it is being compared to SHMM or RHMM variants respectively. The F here denotes that switch probabilities are not used for a particular site, while T would indicate otherwise. The RHMM set of ones indicates that all sites are in region

one, while $(1,1,1,2)$ indicates site four is in region 2 with the other sites grouped in region 1.

The HMM, SHMM and RHMM variants are now applied to the four sets of data. Within each grouping only model averaged posterior state series are compared. Parameter plots are not compared due to the higher number of models.

5.4.1 Sydney Close

The Sydney Close model averaged posterior state series are shown in Figure 5.15, while the posterior model weights for the SHMM and RHMM variants are given in Table 5.8 and Table 5.9 respectively. Both the SHMM regional and the individual state series of the RHMM are similar (with the exception of Moss Vale) and clearly identified, with the 1900-1940 dry and 1940-1985 wet periods evident in both. The Moss Vale series differs greatly from the remaining sites in the RHMM in that there is more frequent switching between states. However, the Moss Vale series is clearly identified also. The individual SHMM series follow the regional series closely, with the exception of Moss Vale, while MtVic/Blackheath is almost identical to the regional series.

The model probabilities necessarily reflect the model averaged state series for the SHMM and RHMM. The SHMM variants indicate (over all results) that the models with switch probabilities on Moss Vale, and no switch on MtVic/Blackheath $(\mathrm{F},\cdot,\mathrm{T},\cdot)$ are preferred, with the highest posterior weight being $(\mathrm{F},\mathrm{F},\mathrm{T},\mathrm{F})$ and $(\mathrm{F},\mathrm{F},\mathrm{T},\mathrm{T})$. For the RHMM, model $(1,1,2,1)$ dominates, suggesting an individual state series is identified in Moss Vale, while the remaining sites are under the one control. The SHMM results also suggest that MtVic/Blackheath identifies the strongest state series, whereas Taralga and Sydney generally follow the overall control. Of note in these results is that the model averaged state series are much more clearly identified than the full model series, especially the RHMM.

Also given in Table 5.9 is a comparison between the best SHMM and RHMM models, with the ratio of posterior probabilities. In this case the RHMM variant $(1,1,2,1)$ has over twice the posterior weight of the SHMM maximum $(\mathrm{F},\mathrm{F},\mathrm{T},\mathrm{T})$, with

$p(M = \text{F,F,T,T}|\mathbf{Y})/p(M = 1,1,2,1|\mathbf{Y}) = 1/2.3$. This ratio is given here firstly to indicate which of the two modeling structures produced the more likely model, but secondly to provide a normalizing constant if the posterior weights for all the SHMM variants are to be compared against the RHMM variants.

Table 5.8 Sydney Close Switch HMM variants posterior model probabilities

Model Label	Posterior Weight	Model Label	Posterior Weight
F,F,F,F*	0.022	F,F,F,T	0.014
T,F,F,F	0.001	T,F,F,T	0.001
F,T,F,F	0.003	F,T,F,T	0.003
T,T,F,F	0.000	T,T,F,T	0.001
F,F,T,F	0.255	F,F,T,T	0.296
T,F,T,F	0.026	T,F,T,T	0.026
F,T,T,F	0.174	F,T,T,T	0.151
T,T,T,F	0.019	T,T,T,T	0.008

*Note: Transition probability constraint on all models including the non-switching HMM

Table 5.9 Sydney Close Regional HMM variants posterior model probabilities

Model Partition	Posterior Weight	Model Partition	Posterior Weight
1,1,1,1	0.016	1,2,2,2	0.000
1,1,1,2	0.000	1,2,2,3	0.003
1,1,2,1	0.728	1,2,3,1	0.039
1,1,2,2	0.007	1,2,3,2	0.001
1,1,2,3	0.165	1,2,3,3	0.000
1,2,1,1	0.000	1,2,3,4	0.005
1,2,2,1	0.034	$p(M = \text{F,F,T,T}|\mathbf{Y})/p(M = 1,1,2,1|\mathbf{Y}) = 1/2.3$	

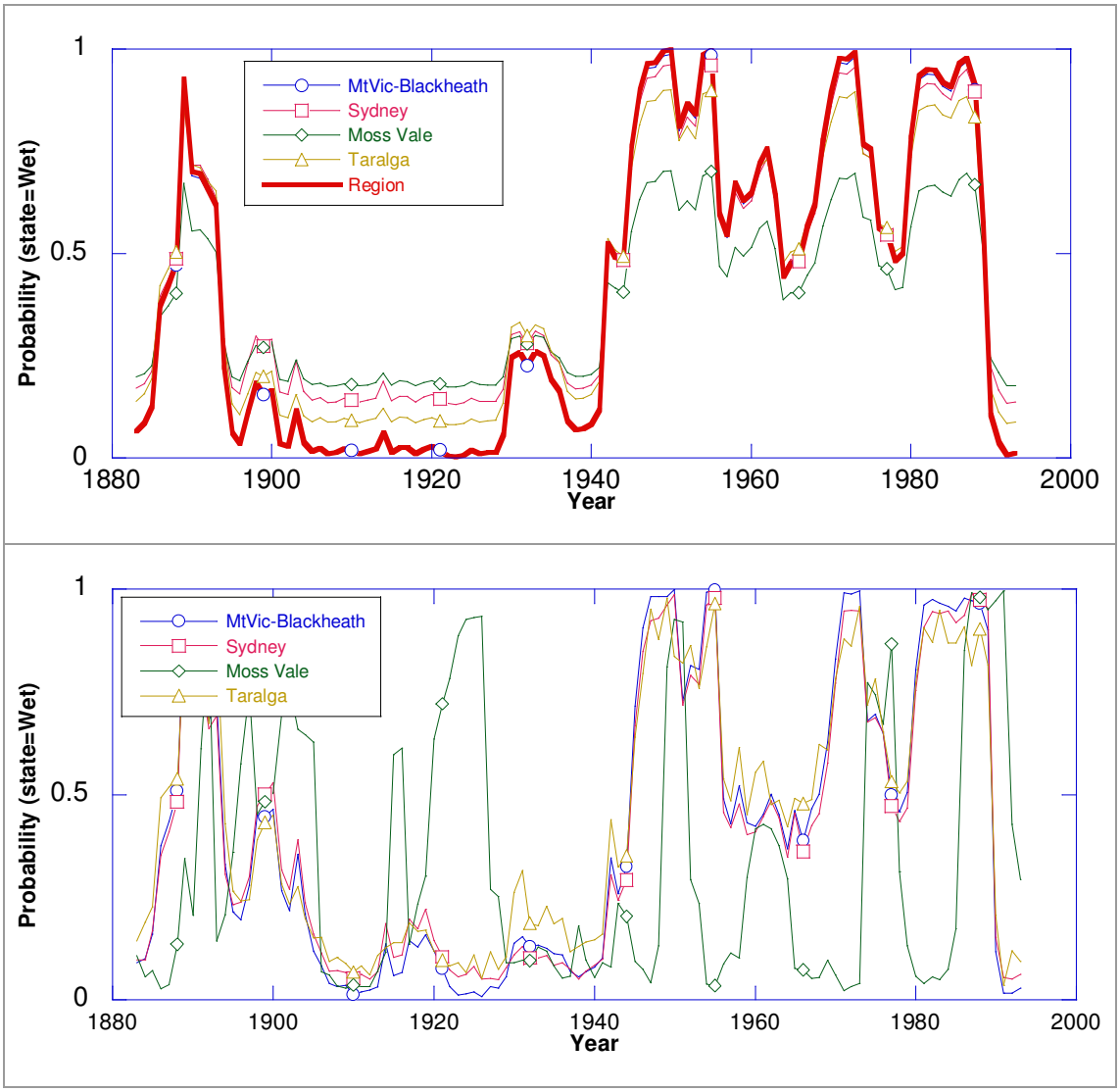


Figure 5.15 Sydney Close Model Averaged Posterior State Series probabilities for (a) Switch HMM variants and (b) Regional HMM variants

5.4.2 Sydney Far

The Sydney Far model averaged posterior state series are shown in Figure 5.16, while the posterior model weights for the SHMM and RHMM variants are given in Table 5.10 and Table 5.11 respectively. Again, the individual state series of the SHMM and the RHMM are similar (with the exception of Bingara) and clearly identified. There is a greater degree of transition between states than for Sydney Close. Nonetheless, the dry 1900-1940 and wet 1940-1985 periods remain discernible. Bingara shows the most deviation from the regional SHMM series and the other sites in the RHMM. The SHMM models with most significant posterior weight are $(\mathrm{T,F,F,F})$ and $(\mathrm{T,T,F,F})$, indicating that Bingara and Mudgee contribute least to the regional state series. For the

RHMM, $(1,2,2,2)$ and $(1,2,3,3)$ dominate the posterior weights, reaffirming the SHMM results by favouring models with Sydney and Moruya in the same region, and Bingara and Mudgee out. Like the Sydney Close results, the model averaged state series are more clearly identified than for the comparative full models, displaying much less dithering around 0.5. In this case the SHMM variant $(\mathrm{T,T,F,F})$ has significantly greater posterior weight than the RHMM maximum $(1,2,2,2)$, with $p(M = \mathrm{T,T,F,F}|\mathbf{Y}) / p(M = 1,2,2,2|\mathbf{Y}) = 7.4$.

The strong favouring of the SHMM variant is presumably due to the simplicity of the SHMM compared to the RHMM in identifying a regional state structure, whilst allowing some sites to differ from the overall control. In this case, Bingara does differ from the overall control somewhat. This causes the RHMM to separate Bingara into its own region. However, the SHMM recognizes that there is some information to be gained on a regional state within the Bingara record. The information gained on the SHMM regional series also explains some of the variability within the Bingara. The RHMM contrasts this information exchange between the regional series and Bingara, with Bingara in its own region, not having any influence on the regional state series at the other sites. Likewise, the regional series for the other sites do not have any influence on the variability of Bingara.

Table 5.10 Sydney Far Switch HMM variants posterior model probabilities

Model Label	Posterior Weight	Model Label	Posterior Weight
F,F,F,F*	0.008	F,F,F,T	0.001
T,F,F,F	0.364	T,F,F,T	0.023
F,T,F,F	0.007	F,T,F,T	0.004
T,T,F,F	0.461	T,T,F,T	0.019
F,F,T,F	0.001	F,F,T,T	0.004
T,F,T,F	0.040	T,F,T,T	0.013
F,T,T,F	0.002	F,T,T,T	0.012
T,T,T,F	0.027	T,T,T,T	0.015

*Note: Transition probability constraint on all models including the non-switching HMM

Table 5.11 Sydney Far Regional HMM variants posterior model probabilities

Model Partition	Posterior Weight	Model Partition	Posterior Weight
1,1,1,1	0.043	1,2,1,3	0.006
1,1,1,2	0.003	1,2,2,2	0.541
1,1,2,1	0.001	1,2,2,3	0.020
1,1,2,2	0.079	1,2,3,2	0.006
1,1,2,3	0.022	1,2,3,3	0.234
1,2,1,1	0.005	1,2,3,4	0.038
1,2,1,2	0.002	$p(M = \text{T,T,F,F}\mid\mathbf{Y})/p(M = 1,2,2,2\mid\mathbf{Y}) = 7.4$	

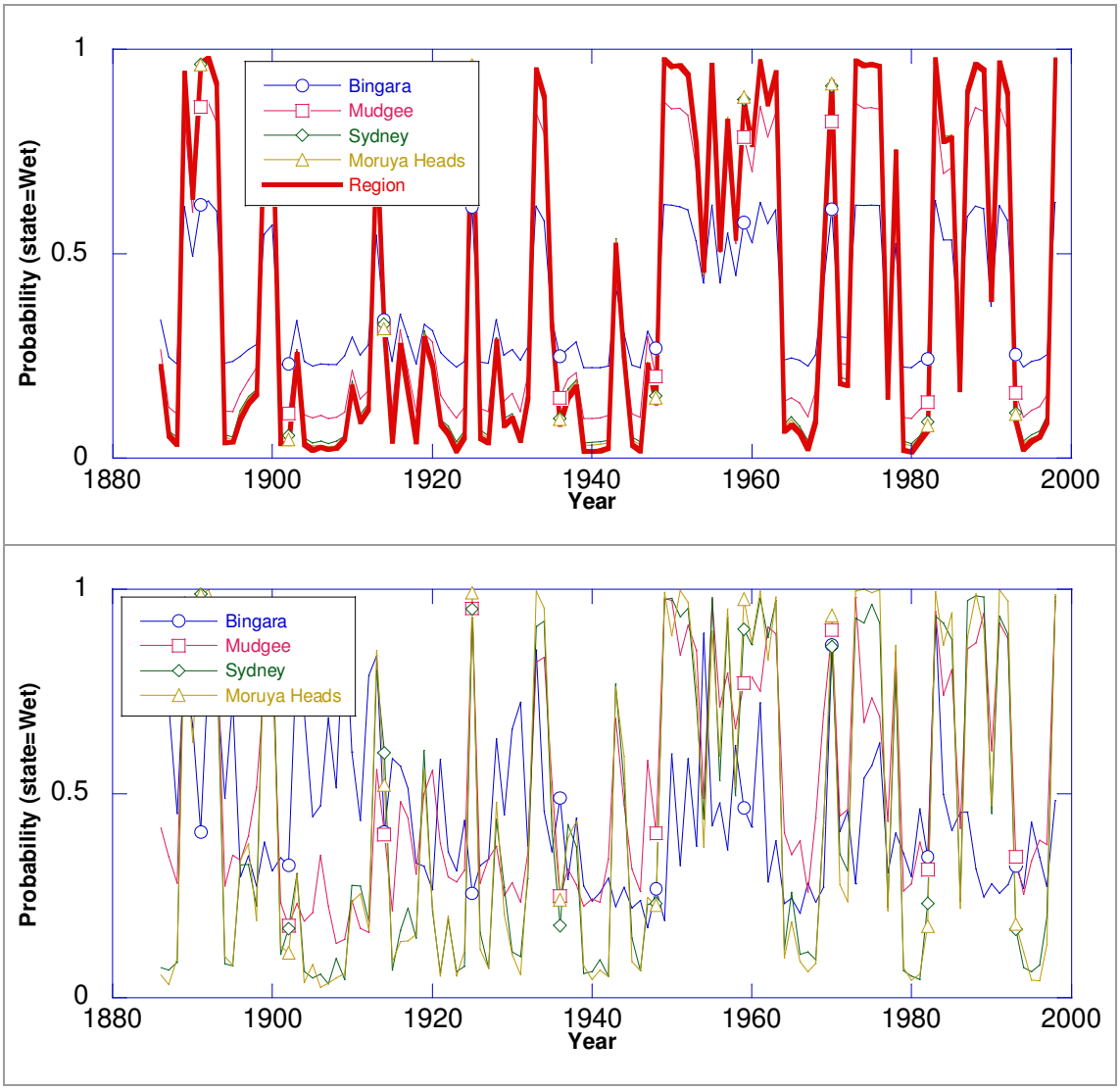


Figure 5.16 Sydney Far Model Averaged Posterior State Series probabilities for (a) Switch HMM variants and (b) Regional HMM variants

5.4.3 Brisbane Close

The Brisbane Close model averaged posterior state series are shown in Figure 5.17, while the posterior model weights for the SHMM and RHMM variants are given in Table 5.12 and Table 5.13 respectively. The SHMM state series strongly identifies Brisbane as following the regional series, whereas Caboolture, Cape Moreton and Pittsworth have state series closer to 0.5. The RHMM results differ markedly in that, Brisbane and Caboolture show a (similar) variable state series, with the 1900-1940 dry and 1940-1985 wet period being discernible in each. Caboolture's state series differs from the full RHMM in that it no longer dithers to the same degree, with it closely resembling the Brisbane state series. Cape Moreton shows a strongly identified series, being typically wet when Brisbane/Caboolture is dry. Pittsworth again dithers (with a greater degree of variation) around 0.5. The model selection results reflect these state series, with $(\mathrm{F,T,F,T})$ and $(1,2,1,3)$ showing significant posterior weights. The SHMM results suggest that Cape Moreton and Pittsworth are not influenced to the same degree by the climate control identified over Brisbane and Caboolture, while Cape Moreton and Pittsworth are in individual climate regions. The RHMM results agree in that Brisbane and Caboolture are grouped into the same climate region. In this case the RHMM variant $(1,2,1,3)$ has significantly greater posterior weight than the SHMM maximum $(\mathrm{F,T,F,T})$, with $p(M = \mathrm{F,T,F,T}|\mathbf{Y})/p(M = 1,2,1,3|\mathbf{Y}) = 1/108.3$.

As mentioned in the full model testing, the comprehensive posterior weighting in favour of the RHMM suggests that the RHMM can produce variability within the data using the Markov states better than the SHMM, splitting sites into different regions while using the Gaussian correlations to account for a greater degree of overall correlation. The result that the SHMM Cape Moreton and Brisbane state series are similar, while being quite different for the RHMM, is in conflict. This result is attributed to the models identifying high (Gaussian) correlations between sites, reducing the information available to identify common climate series across grouped sites. This issue is discussed further in Section 5.5.

Table 5.12 Brisbane Close Switch HMM variants posterior model probabilities

Model Label	Posterior Weight	Model Label	Posterior Weight
F,F,F,F*	0.048	F,F,F,T	0.167
T,F,F,F	0.048	T,F,F,T	0.177
F,T,F,F	0.086	F,T,F,T	0.250
T,T,F,F	0.073	T,T,F,T	0.139
F,F,T,F	0.000	F,F,T,T	0.002
T,F,T,F	0.002	T,F,T,T	0.004
F,T,T,F	0.001	F,T,T,T	0.002
T,T,T,F	0.001	T,T,T,T	0.000

*Note: Transition probability constraint on all models including the non-switching HMM

Table 5.13 Brisbane Close Regional HMM variants posterior model probabilities

Model Partition	Posterior Weight	Model Partition	Posterior Weight
1,1,1,1	0.001	1,2,2,1	0.000
1,1,1,2	0.002	1,2,2,2	0.000
1,1,2,1	0.000	1,2,2,3	0.001
1,1,2,2	0.000	1,2,3,1	0.007
1,1,2,3	0.000	1,2,3,3	0.008
1,2,1,1	0.055	1,2,3,4	0.121
1,2,1,3	0.805	$p(M = \text{F,T,F,T} \mid \mathbf{Y}) \big/ p(M = 1,2,1,3 \mid \mathbf{Y}) = 1 \big/ 108.3$	

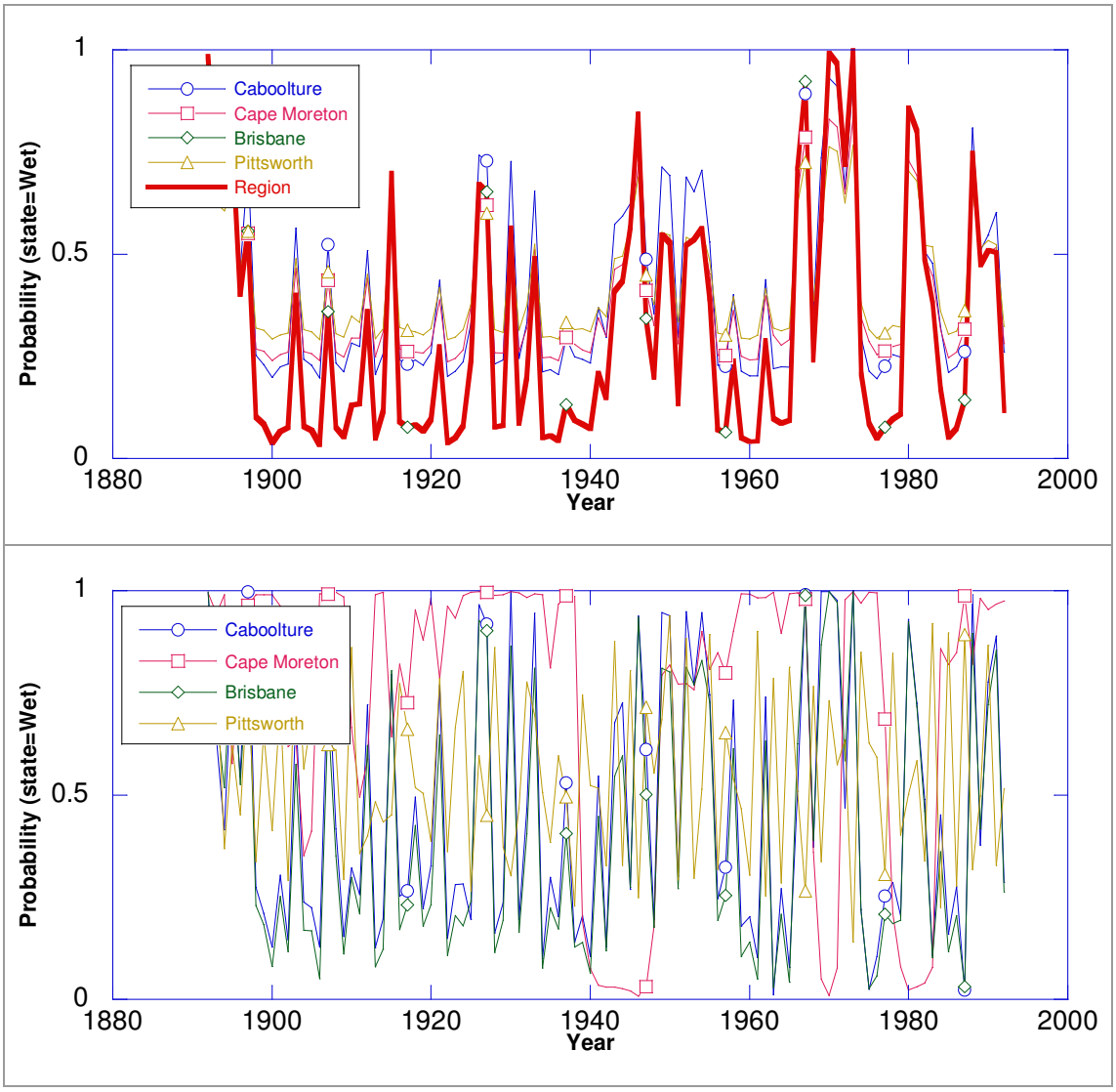


Figure 5.17 Brisbane Close Model Averaged Posterior State Series probabilities for (a) Switch HMM variants and (b) Regional HMM variants.

5.4.4 Brisbane Far

The Brisbane Far model averaged posterior state series are shown in Figure 5.18, while the posterior model weights for the SHMM and RHMM variants are given in Table 5.14 and Table 5.15 respectively. Both state series show that Brisbane is the most clearly identified site, with the other sites essentially dithering. When compared to the state series of the full models, there is little difference, except for some of the other sites showing more coherency in the state series for the model averaged RHMM. The SHMM models $(\mathrm{T,F,T,T})$ and $(\mathrm{T,F,T,F})$ have the greatest posterior weights. However, the weights are otherwise spread evenly between models. For the RHMM, models

$(1,2,3,3)$ and $(1,2,1,1)$ both show significant weight. These results indicate that Brisbane strongly identifies a Markov structure, while the other sites show poorly identified climate structure. Presumably, this result is due to the other sites (especially Miles and Bingara) not having an identifiable two state structure, as demonstrated in the *WADSI* posterior plots for the full models.

Table 5.14 Brisbane Far Switch HMM variants posterior model probabilities

Model Label	Posterior Weight	Model Label	Posterior Weight
F,F,F,F*	0.006	F,F,F,T	0.034
T,F,F,F	0.079	T,F,F,T	0.101
F,T,F,F	0.048	F,T,F,T	0.021
T,T,F,F	0.070	T,T,F,T	0.018
F,F,T,F	0.053	F,F,T,T	0.059
T,F,T,F	0.202	T,F,T,T	0.196
F,T,T,F	0.051	F,T,T,T	0.011
T,T,T,F	0.028	T,T,T,T	0.024

*Note: Transition probability constraint on all models including the non-switching HMM

Table 5.15 Brisbane Far Regional HMM variants posterior model probabilities

Model Partition	Posterior Weight	Model Partition	Posterior Weight
1,1,1,1	0.006	1,2,1,3	0.094
1,1,1,2	0.005	1,2,2,2	0.015
1,1,2,1	0.014	1,2,2,3	0.051
1,1,2,2	0.038	1,2,3,2	0.046
1,1,2,3	0.019	1,2,3,3	0.343
1,2,1,1	0.252	1,2,3,4	0.092
1,2,1,2	0.023	$p(M = \text{T,F,T,F}\mid\mathbf{Y})/p(M = 1,2,3,3\mid\mathbf{Y}) = 1/4.0$	

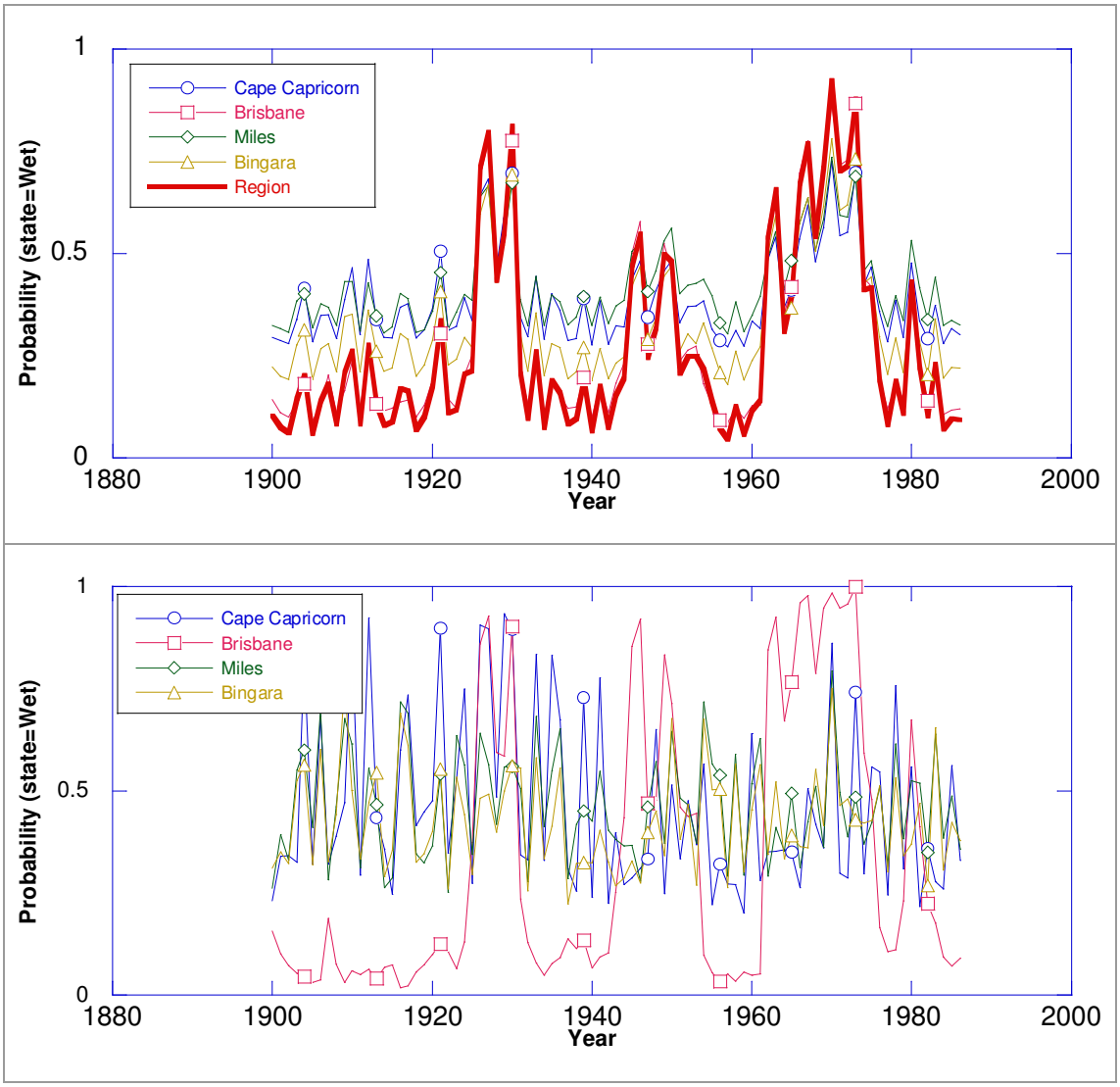


Figure 5.18 Brisbane Far Model Averaged Posterior State Series probabilities for (a) Switch HMM variants and (b) Regional HMM variants

5.4.5 Discussion of Switch and Regional variants case studies

This section has presented application of Bayesian model selection on SHMM and RHMM variants to four case studies centred on Sydney and Brisbane. When compared to the full model case studies of Section 5.3, the results are more interpretable in that the most complex models (the full SHMM and RHMM) are not identified as being the most probable models. Thus, the uncertainty of the extra parameters is reduced so that only parameters aiding identification are included, with a more clearly identified state series resulting.

The RHMM results have a much stronger meaning than those of HMM. Multiple state series are identified, yet sites are automatically grouped into homogeneous climate regions. The strongest groupings in terms of posterior weight affect the model averaged state series to the greatest degree.

When the SHMM variants are compared to the RHMM variants using BMS, the RHMM was favoured for all site groupings, with the exception of Sydney Far. This result indicates generally that the RHMM structure is more flexible in reproducing the data. The RHMM has the added benefit that it allows sites to be grouped into different regions, whilst the SHMM forces a common climate state on all sites. In some cases such as that observed for Sydney Far, this may be preferable to the RHMM. Here the common climate state with anomalous Bingara site allows stronger identification than the RHMM with Bingara in its own region. However, within the other site groupings, the common climate state assumption (with switch anomalies) could not be justified compared to the RHMM which allowed sites to be partitioned into their own climate region.

5.5 The Influence of Small-scale Spatial Correlation on Identification of State Series

5.5.1 Introduction

The case study involving the Brisbane Close group of sites produced some unusual results for the SHMM and RHMM calibrations. Even though Cape Moreton is situated closely to Caboolture and Brisbane, the model averaged state series of the RHMM indicated that Cape Moreton consistently identifies a wet state series, generally being in the opposite state of Brisbane and Caboolture. The SHMM Cape Moreton state series contradicted this, following the Brisbane state series reasonably closely. Given that these sites are close to one another, we would expect the probability of being under the same climate influence to be high.

Comparison of the RHMM variants strongly confirms the result that Cape Moreton is not grouped into the same region as Brisbane and Caboolture. The model selection process is choosing a model with a lesser amount of large-scale correlation than the single region RHMM. When the RHMM state series is examined, this result is stronger

in that there is little time where the state series of Brisbane and Cape Moreton are synchronised, producing very low large-scale correlation.

Given that there is a relationship between the number of sites grouped in each region (large-scale variation) and the correlation structure used (small-scale variation), several different ways of modelling the small-scale correlation are considered in this section. This section has a dual motivation: Firstly, it attempts to identify a reason for the anomalous behaviour of the Cape Moreton state series. Secondly it seeks to produce insight to guide future work for dealing with small-scale correlation structure - specifically, to determine whether the Exponential Decay correlation structure is flexible enough to compete with the Fitted Correlation structure.

This section details comparison of the small-scale correlation structures introduced in Section 4.4.3: the Zero Correlation, Empirical Correlation and Exponential Decay correlation structures along with the Fitted Correlation structure applied in Section 5.4. The study is undertaken in the context of comparing the effect of imposing these correlation structures upon the RHMM, and observing which RHMM variants are favoured depending on the correlation structure. RHMM variants were used here, as opposed to SHMM, as the RHMM provides opportunity for the most complex interaction of correlation structure and Markovian states and BMS clearly favours RHMM over SHMM. If an appropriate small-scale correlation structure is found for the RHMM, it is also likely that the correlation structure would be suitable for the less complicated SHMM.

The Exponential Correlation Decay model is then applied to the Sydney Close data to check if it is appropriate for sites that show non-anomalous grouping behaviour compared to the Brisbane Close state series; that is, for sites where the SHMM series were similar to the RHMM series.

Finally a nugget effect term (microscale variation) is added to the exponential decay model in an attempt to capture extra variability within the data.

5.5.2 Brisbane Close: Comparison of small-scale correlation structures on RHMM variants

The four approaches at modelling spatial correlation were applied to the Brisbane Close data, and the posterior model weights are given in Table 5.16. The model averaged state series are shown in Figure 5.19.

The Zero Correlation model strongly chooses the single region RHMM. Conversely, the Empirical Correlation approach produces posterior weights favouring the most complex RHMM, with the most significant model being every site separated into its own region $(1,2,3,4)$. The Exponential Decay Correlation produces a result similar to that observed in the Fitted Correlation studies, with the partition $(1,2,1,3)$ having maximum posterior weight. This partition puts Brisbane and Caboolture into the same climate region, while the other two sites are in individual regions.

The state series of the three tested models reflect the posterior model weights. The Zero Correlation state series, due to the grouping of all sites into one region, is a single line. This state series is quite variable with a high frequency of changing states. Nonetheless the states are strongly identified. However, the 1900-1940 dry and 1940-1985 wet periods are barely discernible. The strongly identified nature of this series and the grouping of all sites into one region reflects the lack of correlation at the small-scale. The state series is producing the overall spatial correlation.

The Empirical Correlation state series is somewhat similar to the full RHMM presented in Section 5.3.3, with strongly identified series for Brisbane and Cape Moreton, while Pittsworth and Caboolture dither. As in Section 5.3.3, the Brisbane and Caboolture state series are consistently in opposing states. This contradicts the model averaged RHMM series presented in 5.4.3 where Brisbane and Caboolture state series were quite similar due to the prominence of the model $(1,2,1,3)$ grouping these two sites.

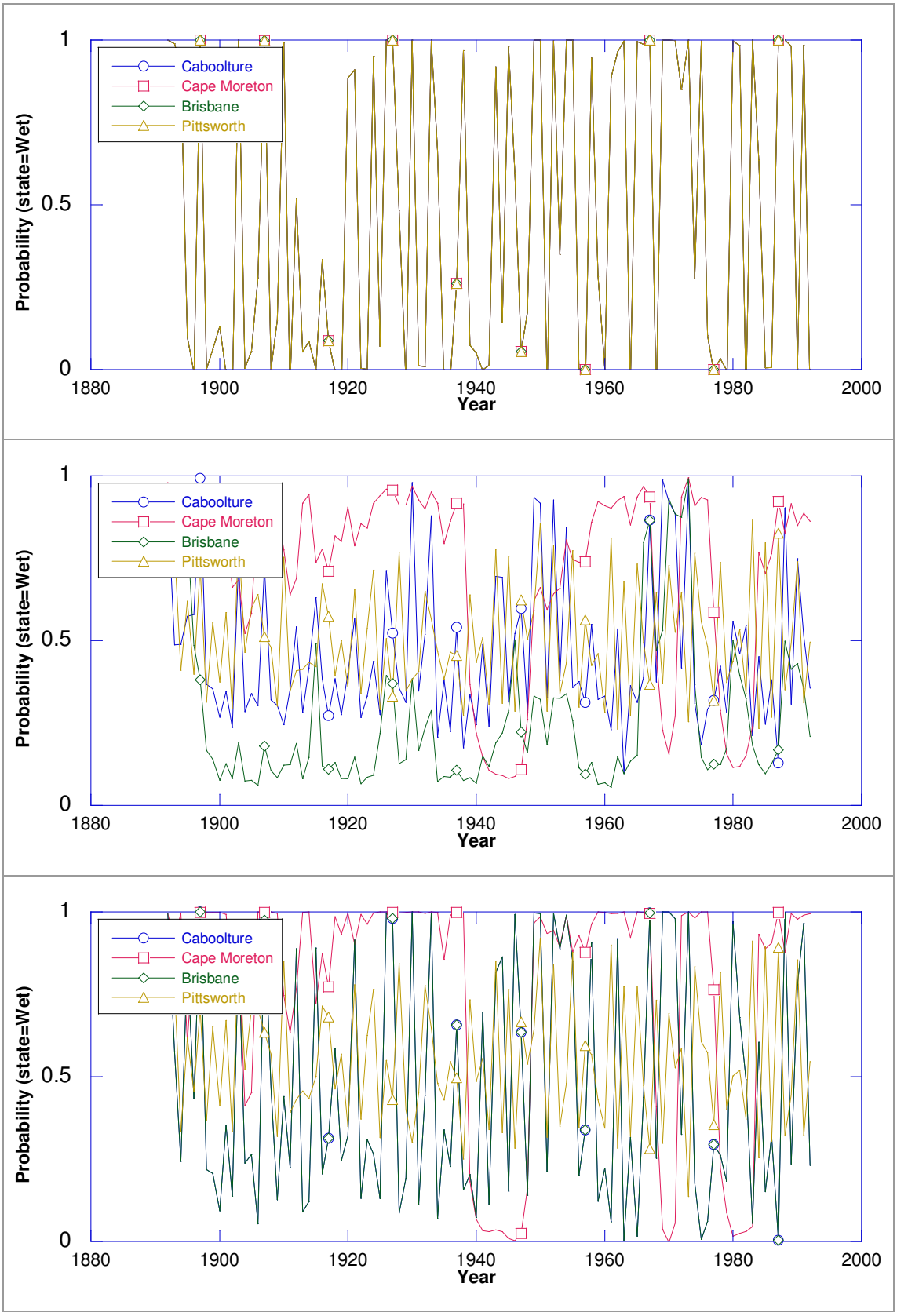


Figure 5.19 Brisbane Close Model Averaged Posterior State Series probabilities for (a) Zero, (b) Empirical and (c) Exponential Decay small scale variation models.

Table 5.16 Posterior model weights for RHMM variants modelling small scale variation with (a) Zero, (b) Empirical, (c) Exponential Decay and (d) Fitted Correlations.

Model	Zero Correlation	Empirical Correlations	Exponential Correlation	Fitted Correlations
1,1,1,1	1.00000	0.02583	0.00002	0.00080
1,1,1,2	0.00000	0.02390	0.00001	0.00240
1,1,2,1	0.00000	0.00061	0.00000	0.00001
1,1,2,2	0.00000	0.00064	0.00000	0.00000
1,1,2,3	0.00000	0.00330	0.00000	0.00008
1,2,1,1	0.00000	0.04527	0.03564	0.05498
1,2,1,3	0.00000	0.28043	0.96423	0.80508
1,2,2,1	0.00000	0.00208	0.00000	0.00008
1,2,2,2	0.00000	0.00143	0.00000	0.00006
1,2,2,3	0.00000	0.02480	0.00000	0.00090
1,2,3,1	0.00000	0.02832	0.00000	0.00677
1,2,3,3	0.00000	0.04340	0.00000	0.00753
1,2,3,4	0.00000	0.52000	0.00010	0.12131
Best Model	1,1,1,1	1,2,3,4	1,2,1,3	1,2,1,3
Relative Weight	6.6439E-40	2.9729E-05	9.9996E-01	1.1110E-05

The Exponential Decay state series is almost identical to the model averaged state series of the Fitted Correlation RHMM in Section 5.4.3, reflecting the grouping of Caboolture and Brisbane, whilst Cape Moreton is again predominantly wet, except for a few occasions near the years 1946,1970 and 1980. These dry Caboolture years (in terms of state probability) coincide with high Brisbane/Caboolture wet state probabilities. The RHMM has apparently identified conflicting rainfall patterns at these sites for these years.

Overall the state series of the Exponential Decay and Fitted Correlation state series are quite similar. The empirically estimated correlation structure differs in that Caboolture and Brisbane are not grouped, resulting in a poorly identified Caboolture series. The Zero Correlation series is much more variable than the other series reflecting the compensation for lack of small-scale correlation.

When the most likely models for each correlation structure are compared (at the bottom of Table 5.16), it is clear that the Exponential Decay structure is overwhelmingly superior, with the Empirical and Fitted correlation structures on approximately equal terms, and the Zero correlation structure lagging far behind.

The *WADSI* distributions for the most likely models in each class are presented in Figure 5.20. The Zero Correlation structure shows the greatest separation, with all sites having very few samples near $WADSI = 0$. The Exponential Decay model shows marginally more separation than the Empirical Correlation model (especially for Caboolture). The fact that the separation is greatest for the zero correlation model, yet the exponential decay model is overwhelmingly favoured in terms of model weight indicates that use of the *WADSI* to distinguish the better model (signified by greater separation) is not appropriate here. Due to parameter interactions associated with allowing small-scale correlation (discussed later in this section), such separations are not necessarily required to model the data to the same accuracy.

How can these results be explained? That is, why is the Exponential Decay favoured so strongly over the other parameterisations? Also, why do the Exponential Decay and Fitted Correlation structures group Caboolture and Brisbane into a single region while the empirically estimated correlations favour all sites being partitioned into individual regions? And finally, why is the state series of Cape Moreton constantly at odds (with the exception of the Zero Correlation state series) with Brisbane and Caboolture? The answer to these questions lies within the one altered variable across this comparison: the correlation between sites.

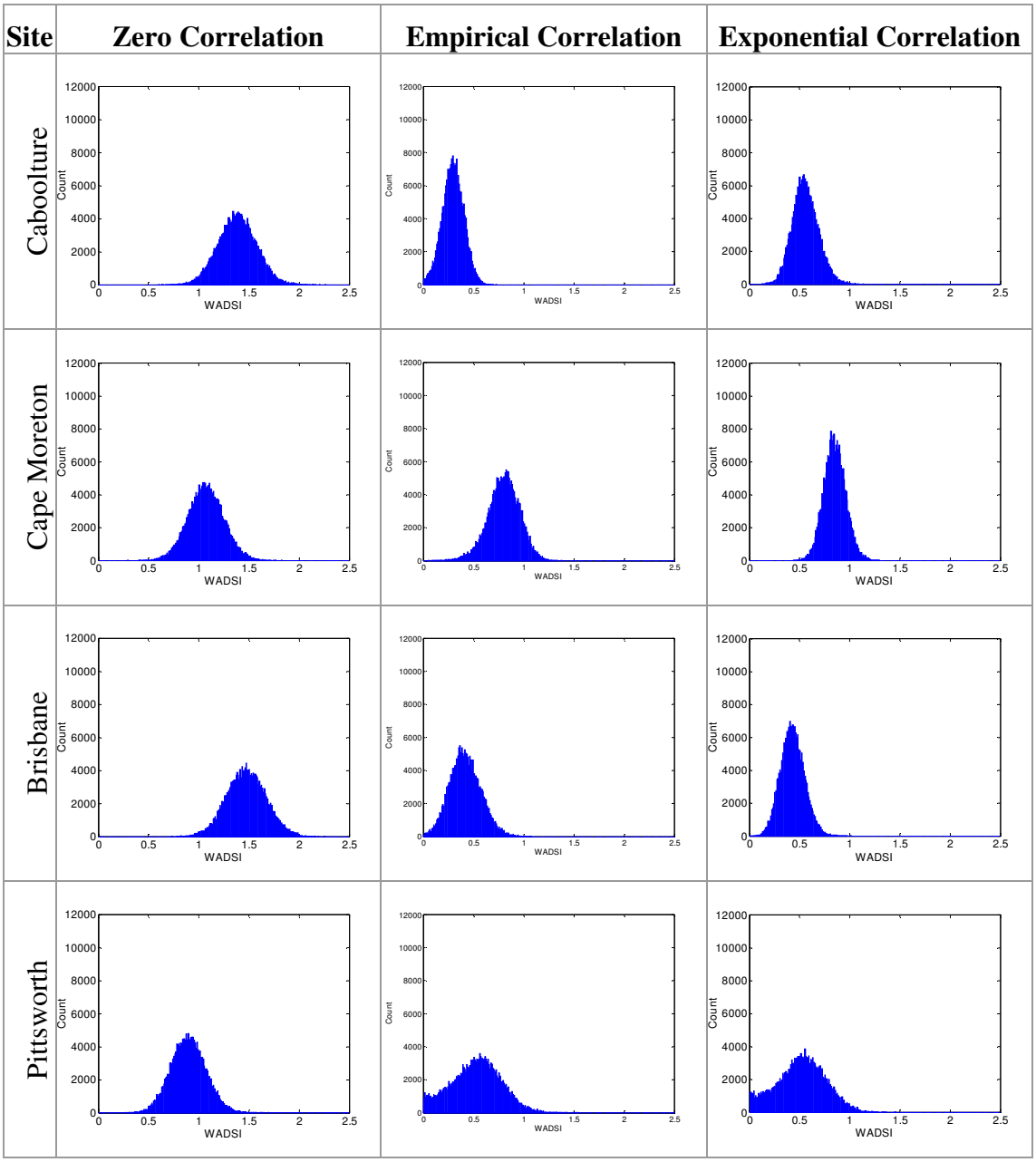


Figure 5.20 Brisbane Close RHMM Sampled Posterior WADSI distribution for (a) Zero, (b) Empirical and (c) Exponential small scale correlation structures.

The distribution of the correlation parameter between each site is plotted for the Exponential and the Fitted Correlation models in Figure 5.21. The correlation plots are presented for the maximum posterior weight model $(1,2,1,3)$ in both cases. The empirically estimated correlation is also plotted as a solid line.

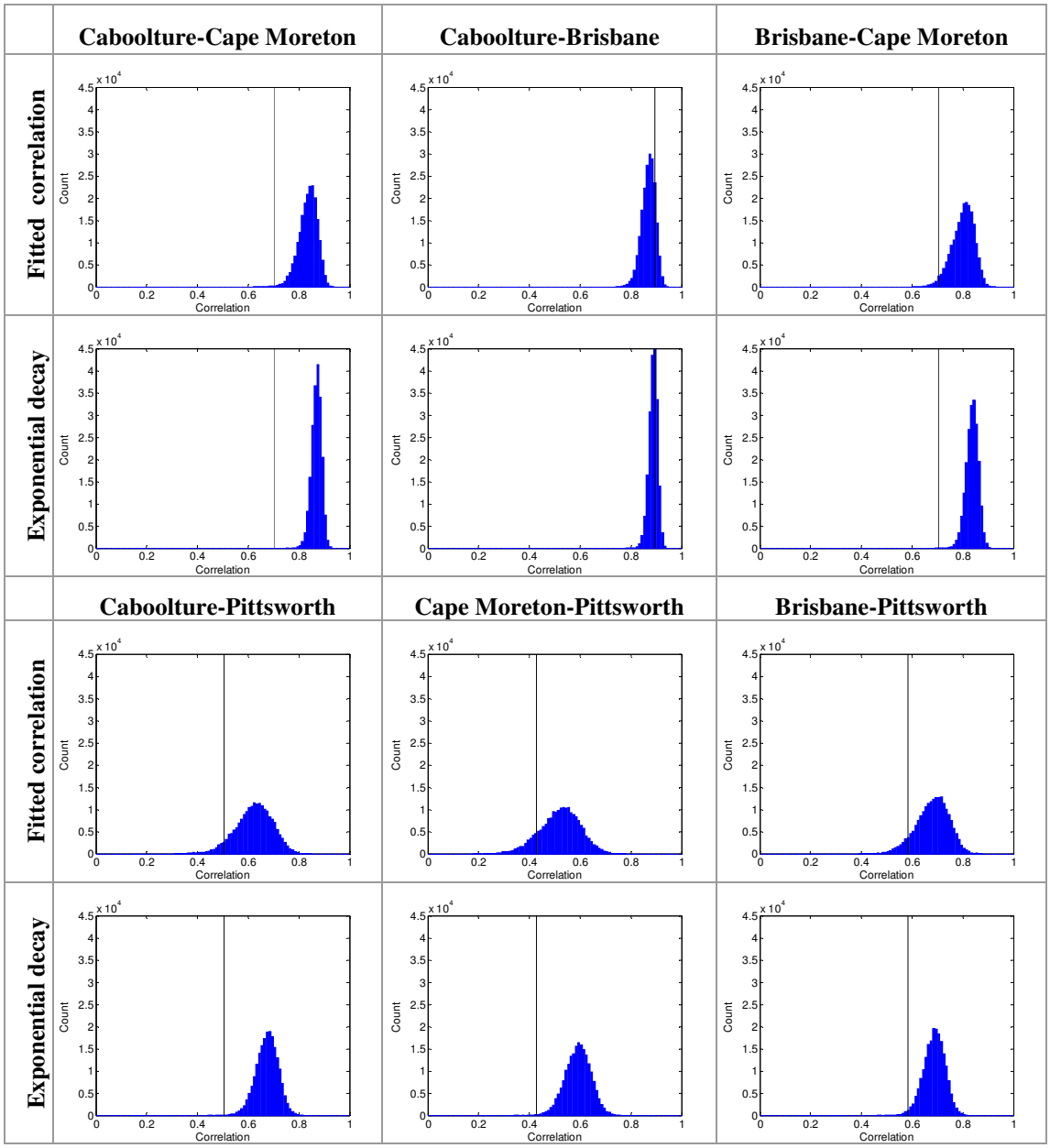


Figure 5.21 Brisbane Close correlation distributions for (a) 1,2,1,3 Fitted and (b) 1,2,1,3 Exponential Decay small scale correlation models. Empirical correlations are marked by solid line.

The Fitted Correlation RHMM offers the most flexibility in describing small-scale correlation structure. The other models have fewer parameters but are more constrained in their description of small-scale correlation. If a less parameterised correlation structure is applied and provides the same inference on the correlation parameters, BMS will choose the less parameterised model. Essentially, this is what is occurring with the Exponential Decay structure being favoured so heavily over the Fitted Correlation. As Figure 5.21 shows, the correlations of the Exponential Decay generally lie well within the correlation distributions for the Fitted Correlation. As there is only one parameter

for the Exponential Decay versus six for the Fitted Correlation model, the Exponential Decay model is favoured. The Fitted Correlation plots illustrate why the Empirical Correlation model is not chosen over the Exponential Decay. Although the Empirical correlations lie within the bulk of the Fitted Correlation distributions, the empirical correlations are near the lower tail for nearly all sites. Thus a parameterisation such as the exponential decay (only one extra parameter), with a distribution centred near the mode of the fitted correlation will be preferred. Similar comments apply to the Zero correlation model, with the Fitted Correlation distributions lying well away from zero.

We now turn to the second question; why do the Exponential Decay and Fitted Correlation structures group Caboolture and Brisbane into a single region while the empirically estimated correlations favour all sites being partitioned into individual regions? Of the correlation plots in Figure 5.21, the Caboolture-Brisbane empirical correlation is contained within the fitted (and exponential) correlation distribution to the greatest degree. Hence, as this empirical correlation does not differ from the fitted correlation significantly, the reason for grouping Caboolture and Brisbane into the same region must lie with the correlations between other groupings. For the remaining sites, the empirical correlation lies on the lower tail of the fitted correlation distribution, sometimes being separated well away from the fitted correlation distribution (eg. Caboolture-Cape Moreton). Of course, these correlations alone do not explain why one model is chosen over another, it is the interrelationship between all parameters of the model.

To demonstrate the relationship between these correlations and the selected model, a bivariate plot of annual rainfall between each site is given in Figure 5.22. Ellipses indicating the 95% probability region of the Gaussian distribution for each site state combination $\{DD, DW, WD, WW\}$ are also plotted. The covariance for each Gaussian distribution was calculated for each combination by taking the posterior expectation from the MCMC samples using the mean, standard deviation and correlation. The ellipses are plotted for the $(1,2,1,3)$ Fitted Correlation and $(1,2,3,4)$ Empirical Correlation models. The Fitted Correlation plot is shown here, rather than the Exponential Decay (which favoured the same grouping of sites), as Fitted Correlation

represents the most general correlation structure – in this way the grouping of sites cannot be attributed to the functional form of the Exponential Correlation structure.

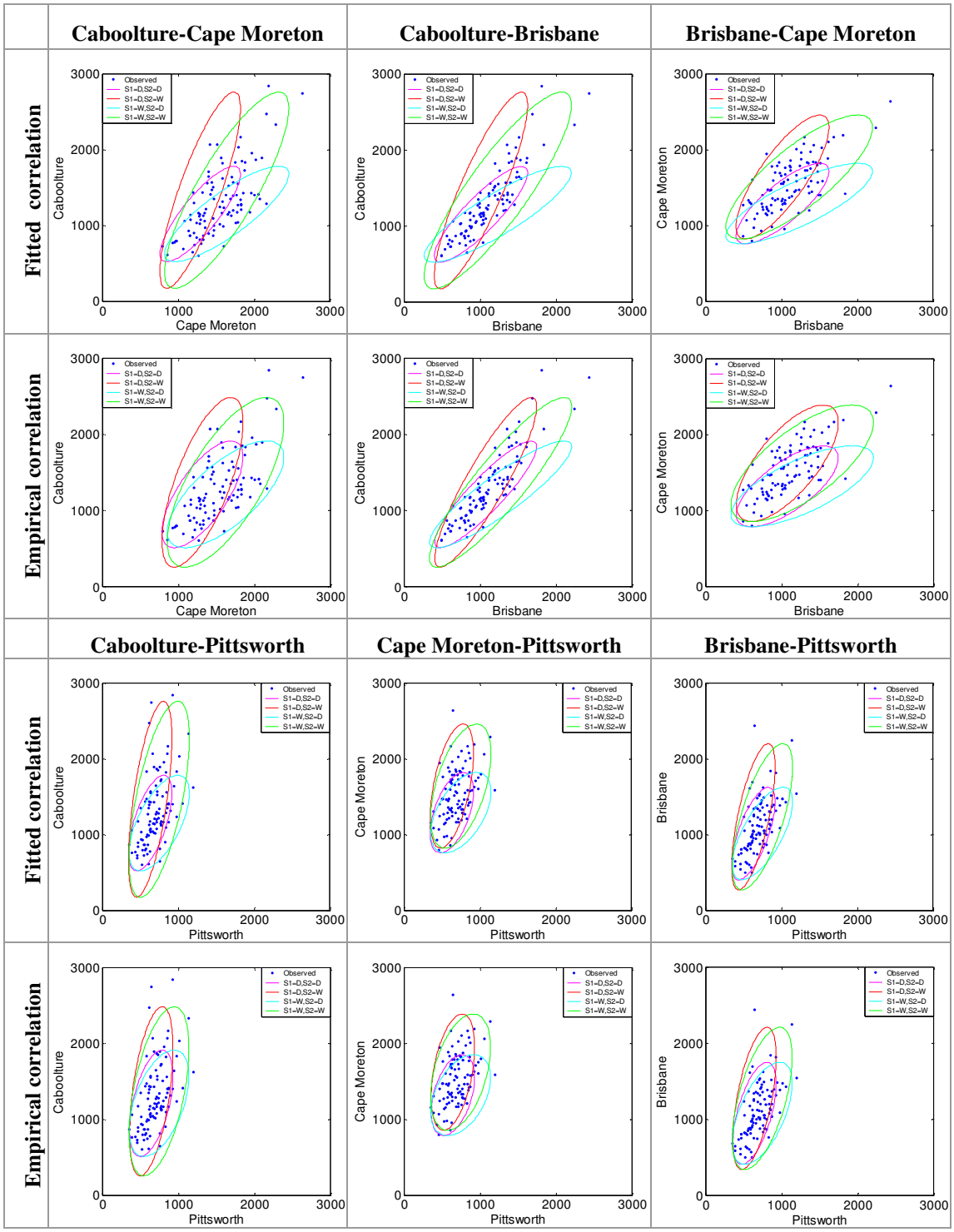


Figure 5.22 Brisbane Close observed annual rainfall and bivariate Gaussian distributions for (a) (1,2,1,3) Fitted and (b) (1,2,3,4) Empirical small-scale correlation models. Ellipses identify the 95% probability region for each of the site state combinations.

It is noted firstly that for all observed distributions, the majority show a fat upper tail. That is, there are several outliers compared to an approximately Gaussian distributed marginal. As the wet mean is constrained to be greater than the dry mean, it will be the wet distribution that accommodates these outliers. Given that the upper tail shows more variability, we would expect the wet distributions to also show more variance.

As the weight of each data point, in terms of log-likelihood, increases quadratically with distance from the mean of the Gaussian distributions according to:

$$\log\left(p\left(\mathbf{y}_t \mid \boldsymbol{\mu}_{R_t}, \boldsymbol{\Sigma}_{R_t}\right)\right) \propto -\frac{1}{2}\left(\mathbf{y}_t - \boldsymbol{\mu}_{R_t}\right)^T \boldsymbol{\Sigma}_{R_t}^{-1}\left(\mathbf{y}_t - \boldsymbol{\mu}_{R_t}\right) \qquad (5.1),$$

we would expect the outliers to have a disproportionate effect on the shape and location of the wet distribution. Indeed this is the case for all the Gaussian distributions shown here, with the *WW* state site combination (shown in green) consistently showing greater variance for each site than the *DD* (magenta). The *WW* distributions tend to accommodate the outliers along their major axis.

Returning to the issue of the Caboolture-Brisbane grouping, Caboolture and Brisbane are the only sites grouped in the Fitted Correlation model, while in the Empirical Correlation model they remain ungrouped. Consequently, the state combinations *DW* (x-axis site=dry, y-axis site=wet) shown in cyan and the *WD* (x-axis site=wet, y-axis site=dry) shown in red, for the Caboolture-Brisbane Fitted Correlation model are not used. That is, these Gaussian distributions are not used for the $(1,2,1,3)$ Fitted Correlation model. Given that the correlation for Caboolture-Brisbane is similar for both models, why are the *DW* and *WD* combinations not required for the fitted correlation model? As stated before, the answer lies with the other sites. It is of note here that the *WW* distribution shows greater variance for the fitted correlation model than the empirical, thus capturing the outliers to a greater degree. Overall, the fitted correlation model is allowing the individual Brisbane and Caboolture standard deviations to be larger.

Comparing the Empirical and Fitted Correlation ellipses across all sites, the most obvious difference between the two methods is that the wet variance of Caboolture is greater for the fitted correlation; this is especially evident within the Caboolture-

Pittsworth plot. Within that plot, the variance accommodates the outliers to a greater degree than the Empirical Correlation. The prior on the variance for both models is equal. Therefore the difference in variance is due to the constraint of the Empirical Correlation structure *a priori* fixing the correlation.

Figure 5.23 shows a regression plot between Caboolture-Pittsworth and Cape Moreton-Brisbane further identifying these outliers. The points $(642, 2742)$ and $(922, 2840)$ on the Caboolture-Pittsworth plot are respectively 3.9 and 3.1 standardized (Gaussian) residuals from the regression line. 95% of Gaussian distributed data should lie within two standard deviations. Given that one of the outliers mentioned here is nearly 4 standardized residuals from the regression line, the influence on the likelihood according to (5.1) will be substantial. There are similar outliers for the remaining site pairings, with the exception of Brisbane-Caboolture. The Pittsworth rainfall 642, is consistently an outlier across all site pairings.

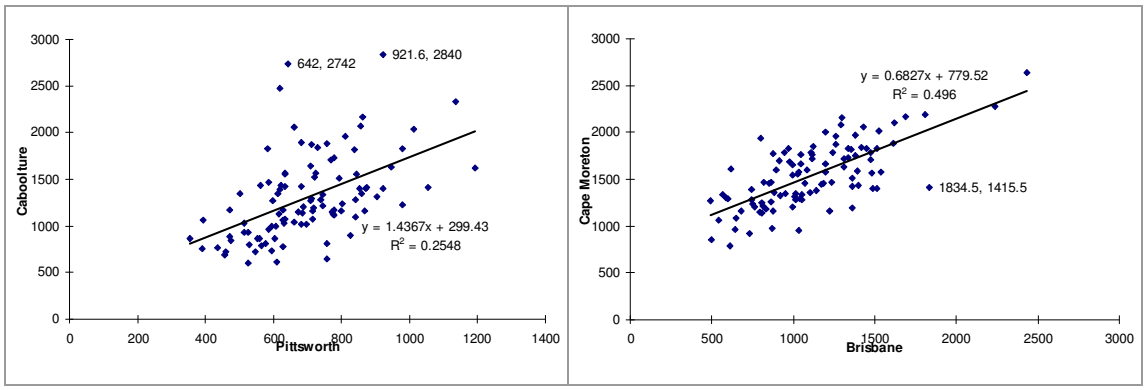


Figure 5.23 Annual rainfall regression plot for Caboolture-Pittsworth and Cape Moreton-Brisbane

The Caboolture-Pittsworth ellipse plots illustrate that the Fitted Correlation model has a greater slope for *WD*, attempting to accommodate the outlier to the right of the *WD* ellipse. To accommodate this outlier to the same extent (all other parameters being equal), the Empirical Correlation (having a lower coefficient of correlation) must increase either the Pittsworth wet or Caboolture dry variance. Indeed, it is observed that the Caboolture dry variance is increased for the Empirical Correlation fit. Likewise, in accommodating the outliers $(642, 2742)$ and $(922, 2840)$, the Empirical Correlation model must also increase either the Pittsworth dry or Caboolture wet variances to fit these outliers to the same degree. The model cannot increase variance on all distributions due to a few outliers, as the bulk of the remaining data points will be less

well modelled. Thus the Empirical Correlation standard deviations compromise, allowing the Pittsworth wet variance to increase, while decreasing the Pittsworth dry and Caboolture wet variances compared to their Fitted Correlation counterparts.

The ability of the Fitted Correlation model to identify a greater variance for Caboolture, due in some part to its relationship with Pittsworth, finally gives insight into the grouping of Caboolture and Brisbane. As a greater wet variance is permitted, and there are no outliers lying off the major axis of the WW Brisbane-Caboolture ellipse, it is suggested the DW and WD distributions are not required.

Considering the influence of outliers on identified parameters, it is not surprising to observe that for the remainder of sites that show outliers well off the major axis of the observed data, that grouping of these sites into the same region did not occur. This is particularly the case for Pittsworth, with the data point 642 consistently being a significant outlier for all site pairings, thus ensuring Pittsworth was not grouped with the other sites. For other site groupings, such as Cape Moreton-Caboolture and Brisbane-Cape Moreton, there are outliers off the major axis, with the DW and WD distributions being required for both the Fitted and Empirical correlation models. Thus, the data does not support grouping of Cape Moreton and its nearby neighbours Brisbane and Caboolture, answering the final question posed regarding anomalous results.

On examining the Cape Moreton data series, it is noted the Brisbane-Cape Moreton outlier $(1835,1416)$ occurs in 1970. According to the regression performed in Figure 5.23, this point is 3.2 standardised residuals away from the regression line, again being quite significant. Returning to the model averaged state series of Figure 5.17 and Figure 5.19, for all of the RHMM models with positive small-scall correlation (fitted, exponential and empirical correlations) 1970 identifies an uncharacteristically wet Brisbane and dry Cape Moreton state. Thus, the influence of the outlier is demonstrated even further, splitting the state series so as to not be in the same state. Of course, as much of the data is better modelled by the outlier ellipses (DW,WD) than the same state ellipses (DD,WW), the state series will prefer to have mixed states at each site for a time proportionate to the goodness of fit (defined in (5.1)) of the Gaussian density signified by the ellipse. This forces the state series between sites to be mirroring

themselves on some occasions. This behaviour is observed within the Brisbane and Cape Moreton state series, with the resulting transition probabilities accordingly showing symmetry about the $p(r_t = D \mid r_{t-1} = W) = p(r_t = W \mid r_{t-1} = D)$ line as demonstrated in Figure 5.8. Such effects on state series are a shortcoming of the current RHMM specification, as the state series appear to be affected largely by outliers, when in reality the sites may be under the same climate influence.

Relationships identified in this section between the correlation and variance parameters demonstrate three major points. Firstly, fitting correlations rather than applying an empirically estimated value provides more flexibility in reproduction of observed data, allowing accommodation of outliers to a greater degree. Secondly, if a functional correlation structure approximates the fitted correlation distribution well, it will be overwhelmingly favoured due to its parsimony. Finally, the non-Gaussian nature of the data can have a large effect on inference for a model which uses Gaussian distributions, with outliers potentially having a disproportionate effect on inferred parameters and state series.

5.5.3 Sydney Close: Exponential Decay correlation

The Exponential Decay model was applied to the Sydney Close group of sites. The resulting posterior model weights for the RHMM variants are given in Table 5.17, while the model averaged state series is presented in Figure 5.24.

The RHMM site grouping favoured by the Exponential Decay correlation structure is $(1,2,2,1)$, grouping MtVic/Blackheath-Taralga and Sydney-Moss Vale. This differs from the Fitted Correlation results where the $(1,1,2,1)$ model was preferred, leaving Moss Vale in its own region, with the other sites in a single group. The model averaged state series reflects this grouping with the MtVic/Blackheath and Taralga state series showing strongly identified dry periods from 1900-1930 and 1985-1994, with the remainder being strongly wet state. The Sydney and Moss Vale state series is predominantly dry with intermittent wet periods, showing a greater degree of variability, while not dithering. Also of note here is that the best Exponential Decay model when compared with the best Fitted Correlation models showed the posterior ratio of $p(M_{\text{exp}} = 1,2,2,1 \mid \mathbf{Y}) / p(M_{\text{fit}} = 1,1,2,1 \mid \mathbf{Y}) = 2.8$.

Table 5.17 Sydney Close Exponential Decay RHMM variants posterior model probabilities

Model Partition	Posterior Weight	Model Partition	Posterior Weight
1,1,1,1	0.001	1,2,2,2	0.000
1,1,1,2	0.037	1,2,2,3	0.196
1,1,2,1	0.081	1,2,3,1	0.172
1,1,2,2	0.000	1,2,3,2	0.001
1,1,2,3	0.015	1,2,3,3	0.000
1,2,1,1	0.000	1,2,3,4	0.035
1,2,2,1	0.460	$p(M_{\exp} = 1,2,2,1 \mid \mathbf{Y}) / p(M_{\text{fit}} = 1,1,2,1 \mid \mathbf{Y}) = 2.8$	

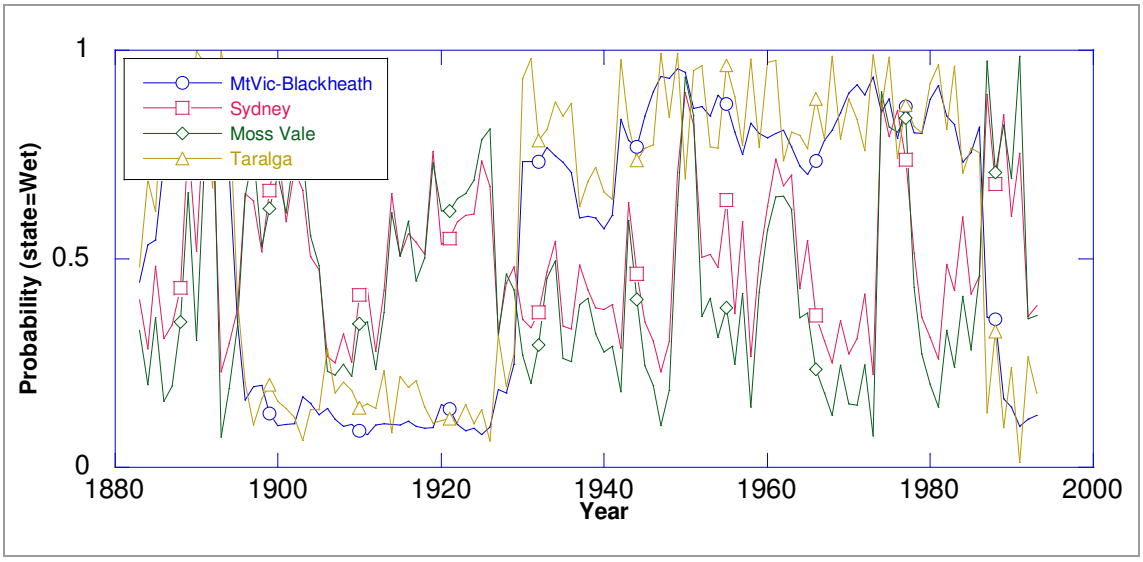


Figure 5.24 Sydney Close Model Averaged Posterior State Series probabilities for the Exponential Decay small scale variation model.

To explain the differing groupings occurring for the Exponential Decay correlation, the correlation distributions of both the $(1,1,2,1)$ Fitted and $(1,2,2,1)$ Exponential models are plotted versus distance in Figure 5.25. The correlation histograms for these two site groupings are given in Figure 5.26, whilst the associated Gaussian 95% ellipses for each site pairing are produced in Figure 5.27.

It is noted firstly that the Exponential Decay correlation distributions do not correspond to the Fitted correlation distributions as well as the Exponential Decay correlations of the Brisbane Close groupings shown in Figure 5.21. This is because the fitted correlations do not display a consistent relationship with distance between sites. This is

exemplified by the Moss Vale-Taralga site pair with the Exponential Decay correlation being significantly higher than the fitted correlation due to Moss Vale-Taralga having the shortest distance between sites. Over all site groupings, MtVic/Blackheath-Moss Vale and Moss Vale-Sydney are the only site pairings to be lower than the fitted correlation, while MtVic/Blackheath-Taralga is within the fitted distribution. Sydney-Taralga and Moss Vale-Taralga have correlation greater than that of the fitted distributions. A compromise is made by the Exponential decay model, with sites having higher or lower correlations than would be produced from fitting individually depending on distance between sites. Given that a correlation trend with distance has been applied to the RHMM, yet the fitted correlation does not display this trend (for all sites), it is not surprising to observe that the best model for the Exponential Decay RHMM was not as overwhelmingly favoured compared to its Fitted Correlation counterpart as in the Brisbane Close testing. That is, the Bayes factor gap between the Fitted Correlation and Exponential Decay model is decreased (for Sydney Close) as the Fitted Correlation is identifying a more complex spatial structure than the Exponential Decay can provide. However, the Exponential Correlation still has a greater marginal likelihood due to its parsimony.

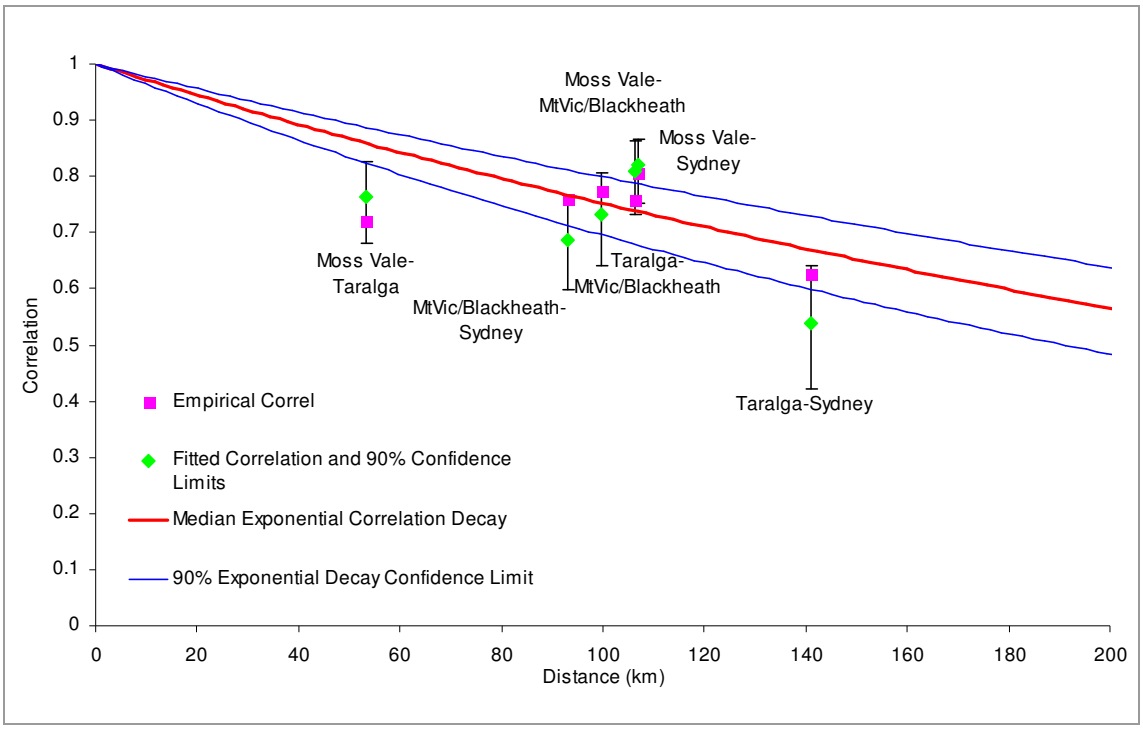


Figure 5.25 Empirical, Fitted and Exponential Decay Correlation versus distance for Sydney Close RHMM.

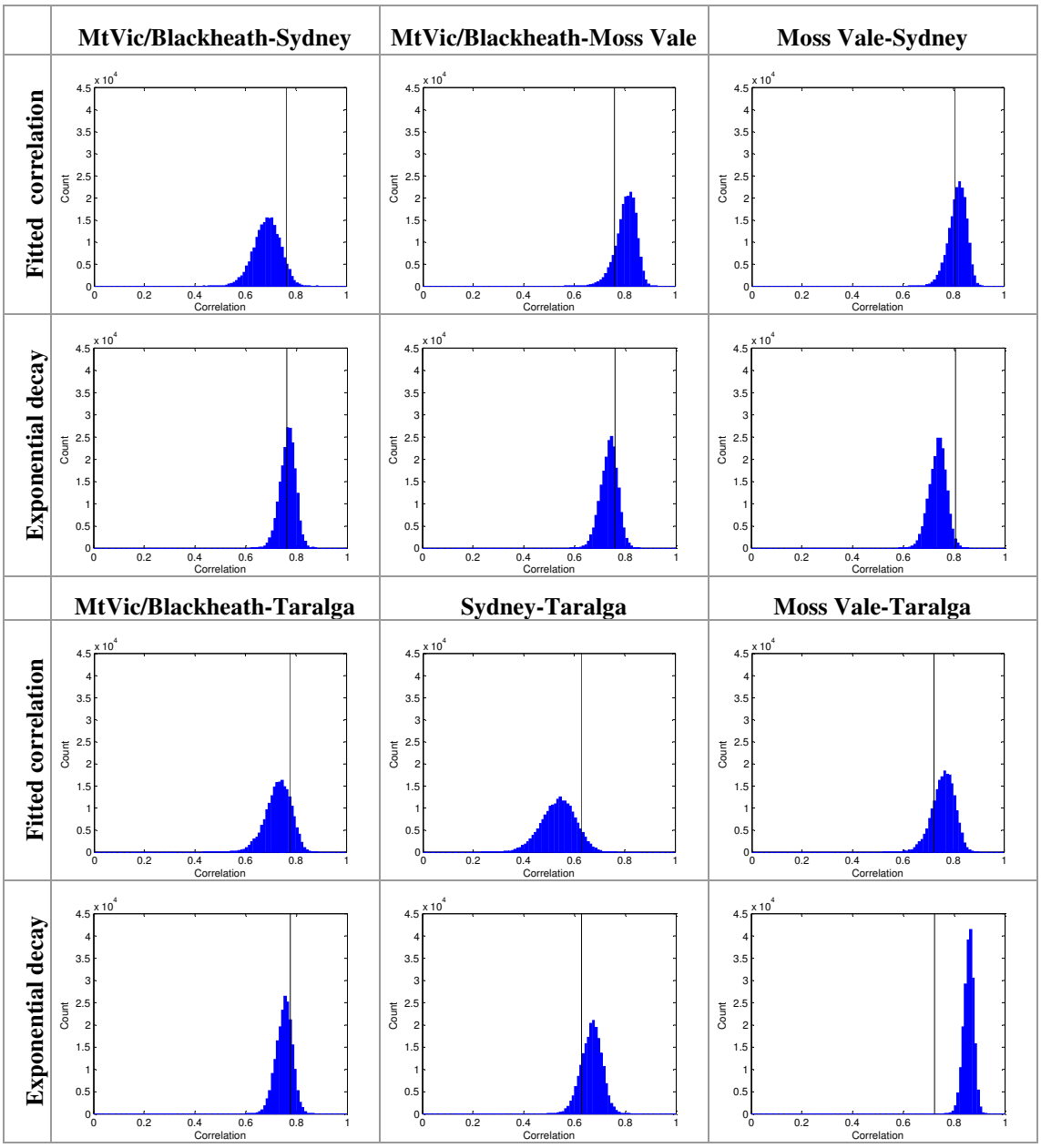


Figure 5.26 Sydney Close correlation distributions for (a) (1,1,2,1) Fitted and (b) (1,2,2,1) Exponential Decay small scale correlation models. Empirical correlations are marked by solid line.

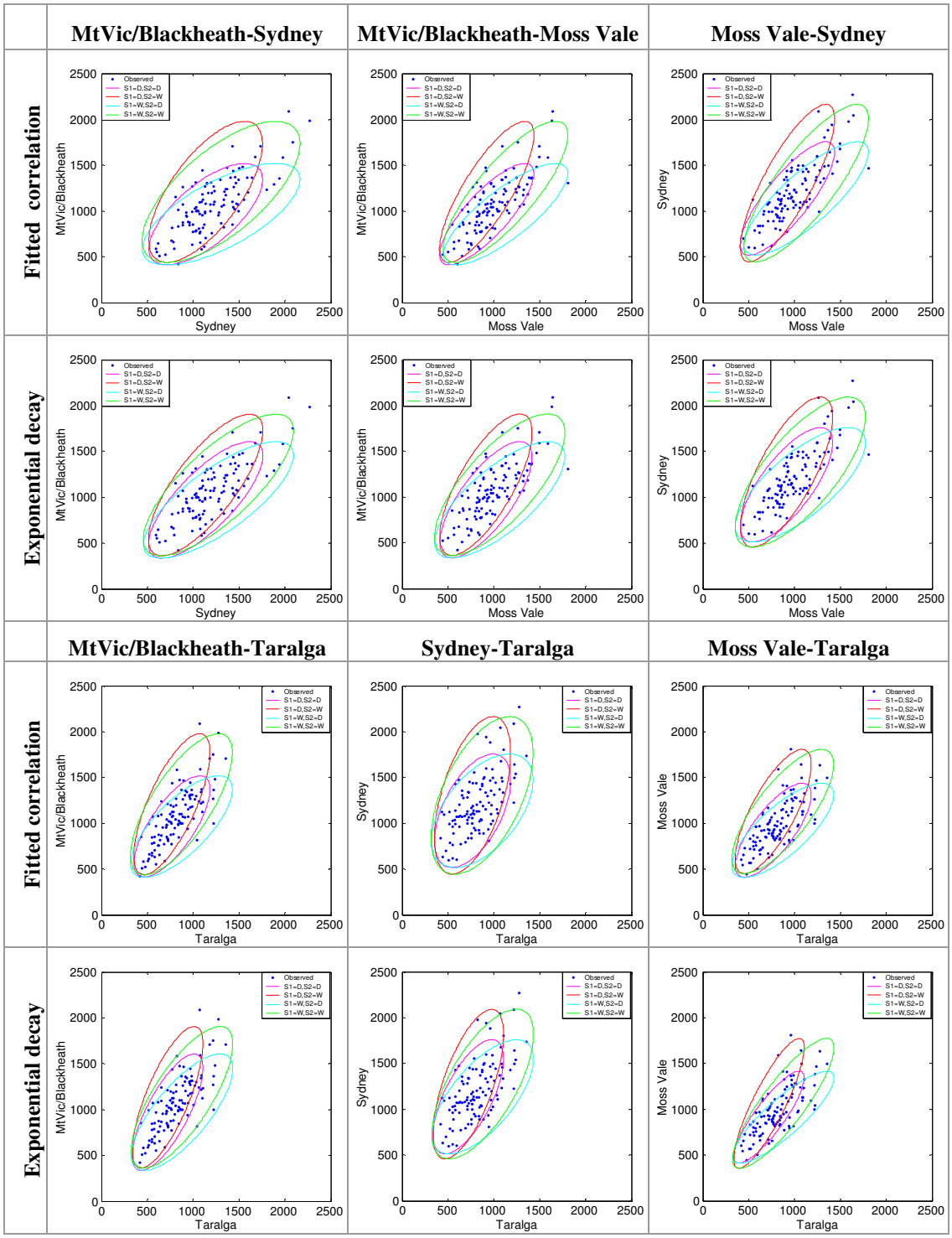


Figure 5.27 Sydney Close observed annual rainfall and bivariate Gaussian distributions for (a) (1,1,2,1) Fitted and (b) (1,2,2,1) Exponential small-scale correlation models. Ellipses identify the 95% probability region for each of the site state combinations.

The 95% Gaussian ellipses shown in Figure 5.27 give a clue to the changes in grouping. As some changes in correlation structure have been caused by the simplicity of the Exponential Correlation Decay model, the ellipses will be forced to be thinner with a

higher correlation, or fatter with lower correlation. On some occasions the variance of the affected ellipse decreases to compensate for forced lower correlations (e.g. MtVic/Blackheath-Moss Vale). For other site pairs (e.g. Moss Vale-Sydney), site grouping occurs due to the fattening of the ellipse, with the grouped model no longer requiring the DW and WD ellipses as the outliers can be accommodated by the enlarged WW and DD distributions. This grouping did not occur for MtVic/Blackheath-Moss Vale due to the presence of a significant outlier lying just outside the WD ellipse. On some occasions the variance of the affected thinned ellipse increases to compensate for forced higher correlations (e.g. Moss Vale-Taralga). For other site pairs (Sydney-Taralga, MtVic/Blackheath-Sydney), site splitting occurs due to the thinning of the ellipse, with the split model requiring the DW and WD ellipses to account for outliers that cannot be accommodated by the thinned WW and DD distributions. Site grouping did not occur for the fitted correlation Moss Vale-Taralga, even though fatter ellipses were present compared to the exponential decay, again due to the presence of outliers. The splayed shape of the distribution requires the DW and WD ellipses.

Some insight has been gained on interpreting the differing results between the fitted and exponential correlation RHMM models. Of note is that the Exponential Decay will not necessarily produce correlation distributions, and hence site groupings, similar to that of the Fitted Correlation model. Lower correlation coefficient distributions tend to produce higher amount of grouping, and *vice versa* for higher correlations. Whether or not the exponential correlation decay is more justifiable than the Fitted Correlation model, is a problem of model selection. Sydney Close RHMM variants slightly favoured the $(1,2,2,1)$ exponential correlation over the $(1,1,2,1)$ in terms of ability to reproduce the data (marginal likelihood). However, the model averaged state series differed significantly. Thus, the marginal likelihood alone is not an indicator how well the state series represents the data, but rather how well the data is explained by the overall model.

It is noted here that the Exponential Decay generally produced slightly greater marginal likelihoods (and therefore more probable models) than the fitted correlations, for all site groupings (see Appendix B), for both the SHMM and RHMM. However, at the stage of

writing this thesis, in depth study of variation in hyperparameters on the Exponential Decay has not been undertaken. In line with the testing of Table 5.16, it is expected that forcing a lower degree of small-scale correlation with the Exponential Decay model, would cause a higher degree of site grouping. The weakly informative prior chosen here was based on observed correlations between all sites used in this study. However, it is clear that the overall variation is contributed to by both the small-scale Gaussian correlations, and the correlations induced by coherent state series between sites. Future testing may find that forcing a lower degree of small-scale correlation through the prior may produce more probable models than those used here.

5.5.4 The nugget effect

Although not fully presented here (see Appendix B) some initial calibrations to the Sydney Close and Brisbane Close groupings with a microscale variation (nugget effect) term added were undertaken. The extra variation term was applied by the imposition of a non-correlated Gaussian component with constant variation over all sites according to Section 4.4.3. For the testing with the fitted correlations, this made little difference to the marginal likelihoods. However, for the Exponential Decay correlation models, significantly higher marginal likelihoods resulted for Sydney Close RHMM. As the fitted correlation distributions were not reproduced by the Exponential Decay model for Sydney Close, the nugget effect term is allowing an extra source of variability to account for these differences. The extra variability is not required in cases where the exponential decay matches the fitted correlations well (Brisbane Close). It is expected as more sites are included in the analysis, a nugget effect term such as this will become more identifiable, and more justifiable. Although a nugget term was not used in the Fitted Correlation results, where enough flexibility was presumably provided to compensate for its absence, fitting correlations is less tenable for a larger number of sites. Hence, a functional relationship (such as the Exponential Decay) for the small-scale correlation, coupled with a microscale variation term providing an extra source of variation becomes more attractive.

5.5.5 Discussion of small-scale correlation structure

The examination of the small-scale correlation structures has shed light on the effect of outliers on state series. The small-scale correlation structures used here (both fitted and

exponential decay correlation) are vulnerable to disproportionate influence by outliers, sometimes having marked effect on the state series produced. The Brisbane Close study demonstrated this with Cape Moreton not being grouped with its nearby neighbours Caboolture and Brisbane. This anomalous splitting/grouping of sites is caused by the non-Gaussian splayed nature of the data. An approach to remedy this, while still maintaining the current model structure would be to transform the rainfall data to more closely resemble a Gaussian distribution. Annual hydrological data is often transformed using the Box-Cox transformation (as defined in the case studies of Section 3.10.2). Alternatively, the power transform:

$$y_t = \begin{cases} \omega^{\gamma} & \omega > 0 \\ 0 & \omega \leq 0 \end{cases} \qquad (5.2),$$

where $\gamma > 0$, was used by *Sanso and Guenni* (2000) for daily rainfall. It is expected that use of such a transform would reduce the effect of outliers, especially for the *WW* state site combination. The *WW* distribution was the most affected in all case studies here as the outliers tended to be at the upper tail of the rainfall distribution. By reducing the effect of outliers, the model has a higher chance of identifying the state series it was designed for.

The preliminary testing has indicated that an exponential correlation decay structure provides an appropriate method of incorporating small-scale variation. This single parameter structure, is less parameterised than fitting correlations between every site, and is therefore attractive in terms of parsimony. Although not reproducing the fitted correlation distribution for the Sydney Close sites, the Exponential Decay was still favoured over the Fitted Correlation model in terms of marginal likelihood.

The introduction of an extra source of variability using micro-scale variability (or the nugget effect) may increase the marginal likelihood of the Exponential Decay model. When applied to a larger number of sites, it is expected that the nugget effect will be of more importance, compensating for the simplicity of the Exponential Decay.

5.6 Discussion

Consideration of the full model calibrations, variant model comparisons and testing of correlation structures suggests two issues need to be addressed in a discussion, firstly

whether the new model structures are justified, and secondly, the future directions of research.

5.6.1 Justification of SHMM and RHMM assumptions

The SHMM and RHMM have both identified state series significantly different than the HMM, with each providing different mechanisms for individual sites to deviate from the HMM regional controlling state. Although, more flexibility is allowed in state series, the SHMM and RHMM assumptions must be compared to the original HMM specification if they are to be justified. The tool used to test model hypotheses in this thesis, Bayesian model selection, is used to determine which modelling structure is more appropriate given a particular set of sites. In one case study (Sydney Far), an SHMM variant provided the most likely model (as opposed to the HMM or RHMM), indicating that the extra flexibility of allowing anomalous sites states are justified. In the remaining case studies, the RHMM variants show greatest posterior weight, indicating that the multiple region RHMM is more justifiable than an single region HMM. In no cases (except the zero small-scale correlation Brisbane Close test) does the HMM produce a maximum marginal likelihood comparable to the SHMM and RHMM.

Given that the HMM generalisations have produced variants with significantly greater posterior weight than the HMM, it is believed that the generalisations are justified. Of course, further evaluation of the models is suggested on more sites, in different areas. What has been demonstrated here is that the SHMM and RHMM provide a means of determining which sites should be included in a HMM analysis. Moreover, the BMS model averaging technique removes the need to choose which sites are included, weighting the models according to their posterior weight.

Although not undertaken for the case studies presented previously, weighting of the site grouping edges according to the posterior model weights could give some idea of the overall (model averaged) state correlation introduced by the RHMM variants. An example of such a map is given in Figure 5.28 for two preliminary 7 site RHMM calibrations (see Appendix B) using the Exponential Decay correlation structure coupled with a nugget effect term. For a larger number of sites, this provides an easily viewable map of which sites are likely to be grouped into the same region, and therefore

into the same state. Within Figure 5.28, the thicker the edge connecting site nodes, the greater the probability the two sites are within the same region.

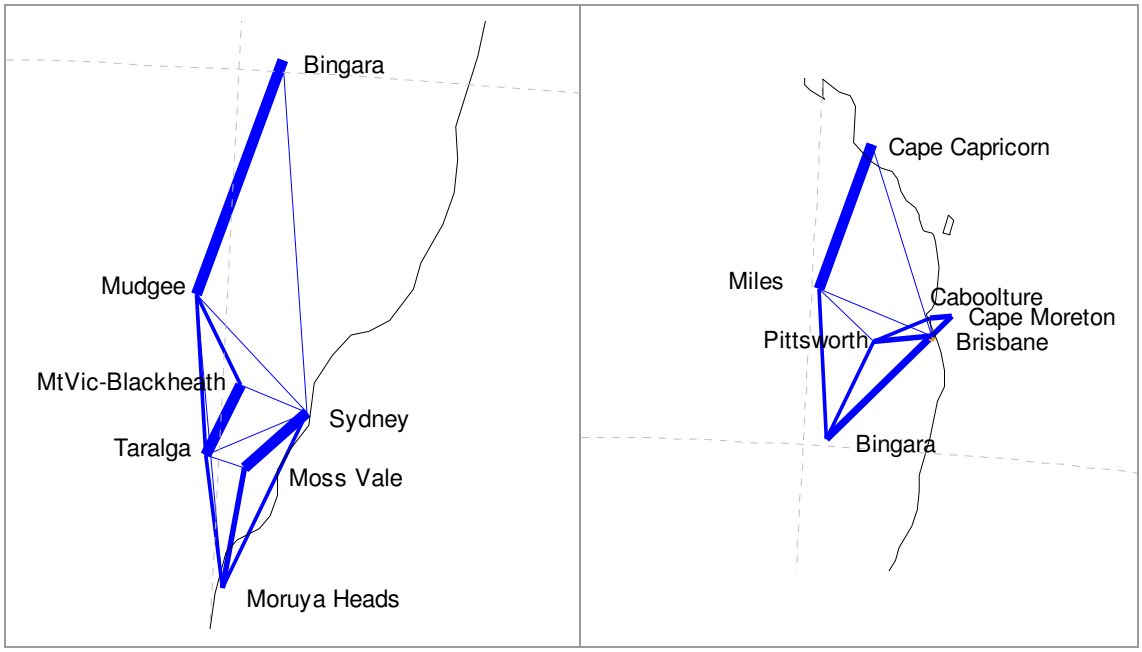


Figure 5.28 RHMM 7 Site exponential decay with nugget effect regional probability map.

5.6.2 Future directions of research

This study has made some inroads to modelling the spatio-temporal stochastic structure of annual rainfall. However, there are significant opportunities for extending this work.

In terms of overall model structure, other generalisations of the HMM are possible. A spatio-temporal HMM with each site's state dependent not only on the previous timestep state, but also the state of surrounding sites seems a worthwhile direction. This approach is essentially a hybrid of HMM for time series and a Markov random field on a lattice. Such a model may reduce the reliance on BMS in choosing which model variant is most appropriate.

Alternatively, related state space models such as the dynamic linear models (*Sanso and Guenni*, 2000) or the shifting mean models (*Fortin et al.*, 2002, *Sveinsson et al.*, 2003) provide a flexible structure that could be used to capture the conceptual climate state variability. To the author's knowledge, altering the climate state across different sites has not been undertaken in these models, and as such the approach taken here for the RHMM, could be applied to these models, breaking the sites into different climate

regions with a climate state modelled in each region. Also, the setup of the RHMM regional partitioning defining the model could be altered such that partitioning is undertaken through a hierarchical indicator within the model itself. This however would require reformulation of the indicator setup for identification purposes.

Hierarchical models show great promise in modelling complex spatio-temporal processes (*Wikle et al.*, 2001). There is opportunity within HMM to use a greater hierarchical structure. For example, means and variances could be modelled as coming from a regional pool/distribution. Also, only annual rainfall has been used in the calibration of the models used in this study. Other related atmospheric indices and variables such as the Southern Oscillation Index or the Interdecadal Pacific Oscillation could be incorporated into the HMM structure. This would at least make the simulations/predictions consistent with such indices automatically, such as in the GCM downscaling approach exemplified in *Hughes et al.* (1999).

The Markovian state structure and site groupings identified have shown some sensitivity to the small-scale correlation structure. An exponential correlation decay with a Gaussian nugget effect for microscale variation should be trialled in further studies. It is expected that calibrations to a greater number of sites will provide insight into whether such a structure is justifiable, with many other well known correlations relationships possible. Of significant importance is the issue of outliers, with the Gaussian distributions being affected strongly. The Box-Cox transformation or that provided (5.2) may normalise the data sufficiently to reduce the splitting effect observed for the Brisbane Close case studies.

Regarding model selection, the possibility of using Reversible Jump MCMC should be investigated further. It is believed this model jumping technique provides a much more efficient means at estimating posterior model weight, as it removes the need to evaluate the marginal likelihood for every model. It is warned however, that the RJMCMC is quite complex, and it is probable that implementation would require a significant amount of work.

Methods of incorporating missing data have not been included in this study. Given that the models are currently limited by the length of concurrent data, the model is currently throwing away much data at the ends of the rainfall series based on the shortest record.

Much of this information on the state series could be used to aid overall identification. Such a method is vital in areas with short $(< 80\ years)$ annual rainfall records.

5.7 Conclusion

This chapter has presented application of the HMM, Switch HMM and Regional HMM to four sets of data located in Eastern Australia. The first section provided a detailed analysis of the three models, comparing parameter uncertainty, and its relationship with the individual site state series. From this analysis, interpretation of the testing of the SHMM and RHMM variants could be undertaken. Within the variant testing, the two case studies located around the key site of Sydney (Close and Far), identified a predominantly dry period from 1900-1940, and predominantly wet period from 1940-1985. The case studies centred on Brisbane showed a less identifiable persistence structure, with some of the sites indicating evidence in favour of a single state model.

The HMM generalisations found greater posterior weight according to Bayesian model selection than the original HMM, indicating that the relaxed regional climate state assumption was justifiable. Generally, the RHMM variants produced models with greater marginal likelihood than the SHMM variants, indicating the RHMM allowed more flexibility to reproduce the variability within the data. As such, the RHMM model averaged state series will be used in Chapter 6 to condition the event based rainfall model DRIP. Further development of the small-scale correlation structure was discussed, with functional relationships such as the exponential decay model being preferable to fitting correlations at every site as the number of sites increases.

Chapter 6 Application of HMM to DRIP

6.1 Introduction

The Australian climatic regime is influenced by several global climate circulations. These circulations produce the high variability and intra- and inter-annual persistence that is a common feature of Australian hydro-climatic data. Current short-timescale point rainfall models do not explicitly account for this persistence. As a result, the variability of annual rainfall may be significantly underestimated. One way to deal with this problem is to condition the short-timescale rainfall model on the climate state simulated by a model of long-term persistence such as the HMM (*Katz and Zheng*, 1999, *Thyer and Kuczera*, 2000).

The point rainfall model we have chosen, detailed in *Heneker et al.* (2001), is an extension of the work of *Eagleson* (1978) who presented a simple event based rectangular intensity pulse model. The stochastic rainfall model, DRIP which stands for Disaggregated Rectangular Intensity Pulse, has previously shown good results for reproducing aggregated rainfall statistics, IFD curves and mean annual rainfall. However, DRIP was unable to satisfactorily simulate annual rainfall variability. This is because DRIP, typical of its genre, only allows for intra-annual seasonality of the rainfall process. Although seasonality is accommodated, storm arrivals nonetheless are independent events. As a result, DRIP has no mechanism to simulate intra- and inter-annual persistence. Where this is significant, DRIP is unable to reproduce the variability of rainfall at aggregation scales of a year.

This chapter explores the inclusion of the HMM in the simulation of event-based rainfall. While particular attention is given to the reproduction of the distribution of annual rainfall, important short timescale statistics are checked to ensure that DRIP's overall performance has not been compromised. Satisfactory reproduction of both long and short timescale statistics is necessary if the rainfall model is to be used to simulate flood and drought climate sequences.

A case study (Case Study I) based on the work detailed in *Frost et al.* (2002) is presented first. This study demonstrates the calibration of the DRIP model at Brisbane and Sydney using conditioning on the original HMM of *Thyer* (2001). The conditioned

two-state DRIP model is compared to the original DRIP (single state) specification, and an improved reproduction of annual variability is observed.

The annual rainfall sites used for calibration of the HMM in *Frost et al*. (2002) were Caboolture, Cape Moreton and Brisbane (for Brisbane), and Blackheath, Sydney and Moss Vale (for Sydney). Although this choice of sites provided an improved reproduction of annual rainfall variability, it was not clear whether introduction of other sites into the HMM analysis would improve on those DRIP results. That is, the HMM sites were not chosen by any formal means. Reiterating one of the major motivations for this study, a method for choosing which sites to include in a HMM analysis had not yet been devised.

Bayesian model selection in conjunction with the RHMM provides a formal method of choosing (averaging over) which sites should be included in a HMM region. This suggests that the RHMM posterior state series be used to condition the DRIP model. In Case Study II two model averaged RHMM state series from the Close and Far case studies of the previous chapter were used to condition DRIP at Brisbane and Sydney. The resulting reproduction of annual variability is then compared with that observed and with that of the previous HMM application.

6.2 Model Description

6.2.1 General Overview of DRIP

DRIP simulates the inter-event time, storm duration, average event intensity and temporal distribution characteristics of point rainfall. It can be used to simulate long sequences of rainfall events at time-scales down to 6 minutes. Figure 6.1 illustrates the key conceptualisation, a schematic of a time series of rectangular rainfall pulses or storm events characterised by three random variables: t_a, the inter-event time; t_d, the storm duration; and i, the average rainfall intensity. The storm depth is defined as the product of i and t_d. Monthly parameters are used to ensure seasonal variations are taken into account.

As intra-storm rainfall is of interest in flood estimation, individual rectangular pulses are disaggregated using a dimensionless mass curve scheme similar to that employed by

Woolhiser and Osborn (1985) and *Koutsoyiannis and Foufoula-Georgiou* (1993). This scheme uses the gross characteristics of each rainfall event (t_d and i) to perform a conditional random walk through dimensionless mass curve space. This enables the temporal distribution within a storm to be simulated at short timescales of the order of 6 minutes or less.

Inter-event time and storm duration distributions are described by a generalised exponential distribution with a combination of the Generalized Pareto (GP) and the power law kernel (*Lambert and Kuczera*, 1998). The distribution of rainfall intensity is conditioned on the storm duration, and is described by the GP distribution whose mean and variance are described by a broken line function – see *Heneker et al.* (2001) for further details.

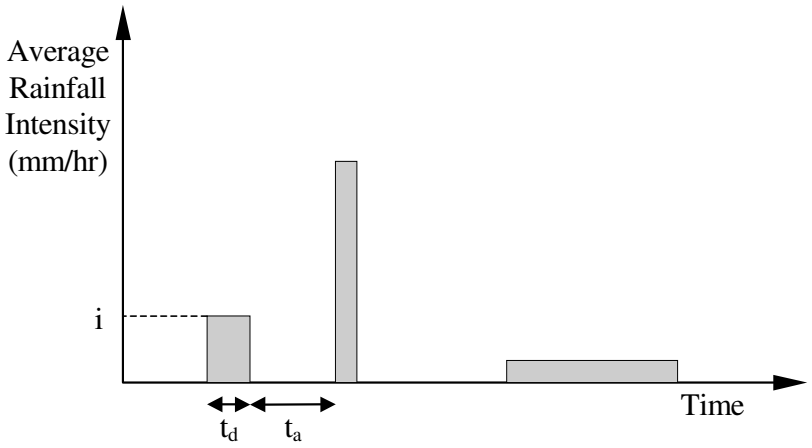


Figure 6.1 Drip model of precipitation events

6.2.2 Incorporating Inter-annual Persistence into DRIP

The RHMM averaged state time series described in the Chapter 5 gives the posterior probability of any rainfall year used in the calibration being in a dry (or wet) state. This probability time series is used to condition the DRIP rainfall model during calibration. The question is, how do we condition the DRIP model on this state series?

It is beneficial when modelling rainfall to keep the model parsimonious or as simple as possible. Fewer parameters reduce calibration time and increase the chance of successful regionalisation of the model because misleading correlations between parameters are reduced. The two-state HMM provides a simple framework for

conditioning an event based rainfall model on inter- and intra-annual persistence. Intra-annual persistence is enforced because the annual state, wet or dry, applies to all months in the water year. Inter-annual persistence arises when the HMM identifies a temporal Markovian structure within the annual data, producing consecutive years in a wet or dry state. Of course the HMM may degenerate to a mixture model, which would produce extra variability through the intra-annual persistence alone rather than through intra- and inter-annual persistence. Conversely, the model could be generalised further to allow more intra-annual variability by relaxing the assumption of a constant climate state throughout each year.

In keeping with the goal of developing a parsimonious model it is hypothesised that only the number of storm arrivals per time period is modulated by the HMM. Thus, the two-state HMM is imposed only on those parameters affecting the number of storm arrivals i.e. inter-event time and storm duration. Storm temporal distribution and average rainfall intensity parameters are not adjusted. Accordingly, two separate sets of parameters are calibrated for the storm arrival parameters, one for years classified as dry and the other for years classified as wet.

As we cannot be sure which state a particular historic year is in, the hidden state probability series produced by the HMM is used to estimate the probability of the historic year being in either a wet or dry state. One possible way of using this series in the calibration of the DRIP model would be to classify each year as either wet or dry using some threshold value. However, this approach is subjective and does not allow the full amount of information from the data to be utilised by both states.

A more objective method is to weight event calibration data using the hidden state probability. This weighting is undertaken on a generalised exponential distribution that takes the form:

$$P\left(X \le x_i \mid \boldsymbol{\theta}_t, s_t\right) = 1 - \exp\left(-g\left(x_i, \boldsymbol{\theta}_t, s_t\right)\right) \qquad x_i > 0 \;\; i = 1,...,n \tag{6.1}$$

where n is the number of data points, X is the independently distributed random variable describing either storm duration or inter-event time, t is the time at the start of event X and is expressed in years, $\boldsymbol{\theta}_t$ is a parameter vector dependent on t, s_t is the

Markov state at time t, and $g(x, \boldsymbol{\theta}_t, s_t)$ is the exponential kernel function. $\boldsymbol{\theta}_t$ contains the parameter subvectors $\{\boldsymbol{\theta}_t^W, \boldsymbol{\theta}_t^D\}$ where $\boldsymbol{\theta}_t^W$ and $\boldsymbol{\theta}_t^D$ are those parameters relating to the wet and dry state respectively. Total probability is used to remove the dependence on the state in equation (6.1) yielding the probability density:

$$p(x \mid \boldsymbol{\theta}_t) = p(x \mid \boldsymbol{\theta}_t^W) P(s_t = W) + p(x \mid \boldsymbol{\theta}_t^D) P(s_t = D) \qquad (6.2),$$

where $P(s_t = W)$ denotes the hidden state probability given the year the event X occurred. If there is a high probability of being in a wet state, the wet state parameters exert a greater influence on the random variable X than do the dry state parameters, and vice versa for a low probability.

The overall likelihood is, assuming independence of events, the product of the densities for all the data points given by:

$$p(x_1, ..., x_n) = \prod_{i=1}^{n} p(x_i \mid \boldsymbol{\theta}_t) \qquad (6.3).$$

The parameter sets $\boldsymbol{\theta}_t$ were estimated using the maximum likelihood optimisation techniques described in *Lambert and Kuczera* (1998).

The maximum likelihood parameter estimation technique used here differs from the Bayesian parameter uncertainty model calibration techniques used in previous chapters. Although such a technique does not incorporate parameter uncertainty, maximum likelihood will give an indication of the best possible fit for the particular model used. If one model's simulations do not correspond to the observed data well, it is an indication that the model hypothesis is not as justified compared to another that does. Incorporating parameter uncertainty into DRIP is beyond the scope of this study. As parameter uncertainty has not been incorporated here, use of BMS and model averaging between different DRIP specifications is not possible. Again, this is an intended direction of future research.

6.3 Data

Pluviograph data in six-minute increments from three Australian capital cities (Sydney, Brisbane and Melbourne) were used by *Heneker et al.* (2001) to calibrate the model parameters. They found that simulated annual rainfall significantly underestimated the observed variability for Brisbane and Sydney. However, the Melbourne variability was reproduced satisfactorily. This result concurs with the work of *Thyer and Kuczera* (2000) which found that locations significantly influenced by tropical weather systems (Brisbane, Sydney) revealed an identifiable persistence structure. In the case of Melbourne, which is not influenced significantly by tropical weather systems, the data did not support the persistence assumptions used in the HMM. Therefore, this work will focus on validation against pluviograph data from the Brisbane and Sydney gauges. The length of the pluviograph data is from January 1913-November 1991 for Sydney, and January 1908-December 1991 for Brisbane as detailed in Table 5.1 in the previous chapter.

The original HMM applied in Case Study I (from *Frost et al.*, 2002) requires the input of annual rainfall totals from the region surrounding the pluviograph site. The use of multiple sites enables space-for-time substitution which should strengthen the identification of persistence (provided that the sites belong to the same persistence region). However, at the time this case study was undertaken, a rigorous rationale for choosing sites to be used in the HMM was yet to be devised. Accordingly, in that study, a heuristic approach was adopted in which a site is only added to the HMM if its posterior distribution of the transition probabilities was consistent with that of the already accepted sites. Of course, it is recognised that the requirement of similar transition probabilities does not guarantee individual site state series are coherent. Using this heuristic Sydney, Blackheath (63009) and Moss Vale gauges were selected for the Sydney multi-site analysis, whereas Brisbane, Caboolture and Cape Moreton were selected for the Brisbane analysis. The Mount Victoria/Blackheath composite of the previous chapter replaced the Blackheath record used here as it only had a length of 1898-1993, reducing the overall length of annual rainfall series. The Caboolture record used in the *Frost et al.* (2002) study was shortened by four years (1888-1891) in the previous chapter as these years contained two completely missing rainfall months (April 1890, November 1891). Such missing months could have considerable effect on the

RHMM, and hence were removed. However, their influence on the HMM is less pronounced, with the outlier effects outlined in the previous chapter not being possible. Therefore, the concatenated length of HMM annual rainfall series is 1898-1993 (May-April water year) for Sydney, and 1888-1993 for Brisbane (April-March water year). The start of the rainfall year was chosen according to the *SSI* index of *Thyer* (2001). The resulting state series for the HMM is shown in Figure 6.2a.

The model averaged RHMM state series produced in the previous chapter and used in Case Study II are shown in Figure 6.2b and c, and correspond to the Close and Far site groupings respectively.

Comparison of the HMM and RHMM state series reveals that the Brisbane Close RHMM state series is quite different to that of the original HMM, displaying a much greater degree of variability. It is interesting to note in Figure 6.2b and c that for some years the Sydney and Brisbane state HMM series are somewhat similar, suggesting that the same climate controls may be occurring over both regions. The Sydney Close RHMM series does show similarities to that of the HMM, however the 1940-1985 wet period is more pronounced in the RHMM series. The Brisbane and Sydney Far RHMM series show a greater degree of variability of state, with a proportionately higher amount of time in the wet state, and a greater degree of dithering around 0.5.

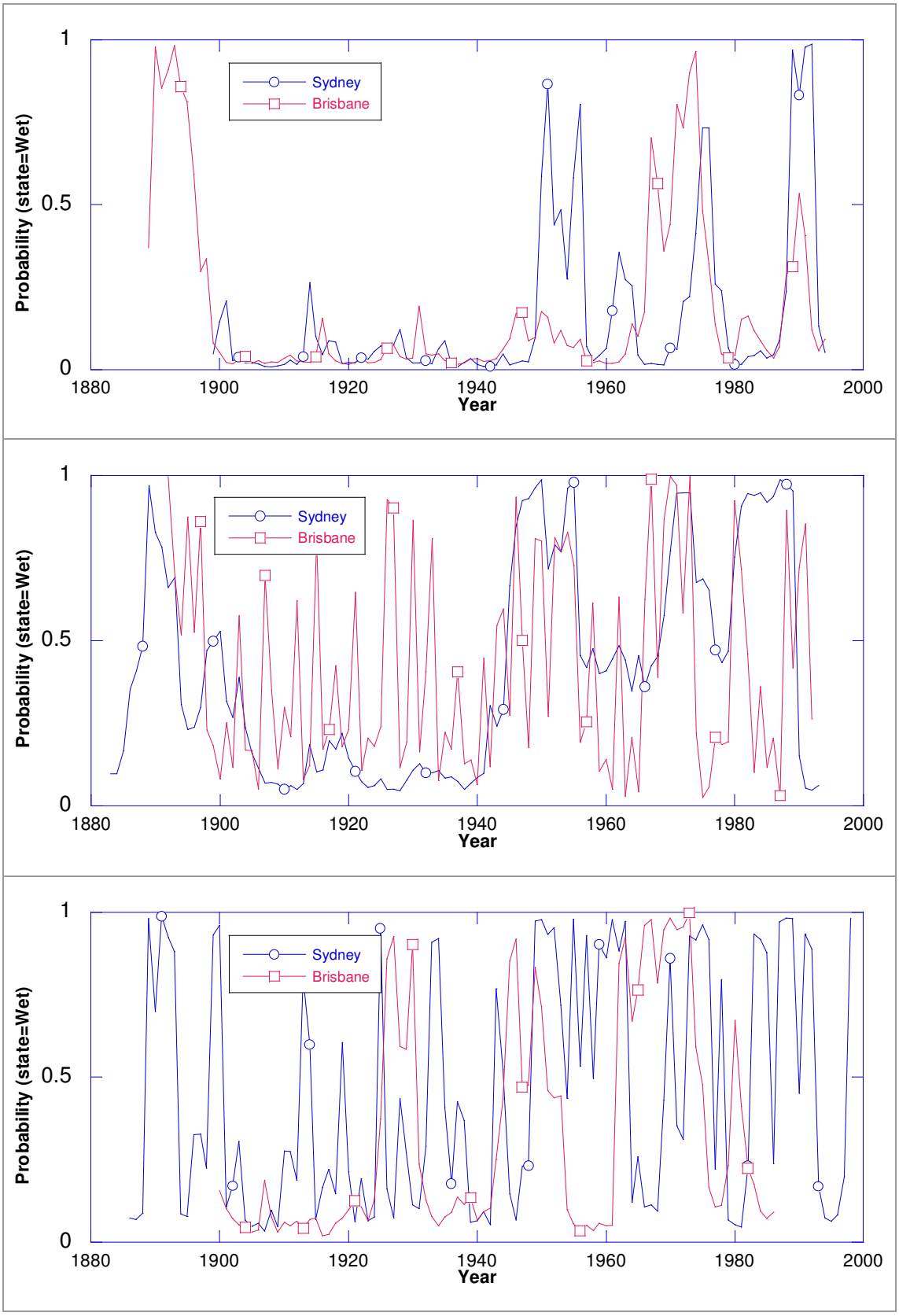


Figure 6.2 Posterior State Series probabilities for the (a) HMM used in *Frost et al.* (2002), and model averaged RHMM with fitted correlation for (b) Close and (c) Far site groupings.

6.4 Case Study I : DRIP conditioned on HMM state series

This section details the conditioning of the DRIP model on the posterior state series of the original HMM as discussed in *Frost et al*. (2002). The two HMM series shown in Figure 6.2a were used in the calibration of the DRIP model, weighting wet and dry storm duration and inter-storm duration parameters according to this series. The two state DRIP model is validated by comparison with observed annual rainfall and the simulations produced by the single state DRIP specification. Other short timescale statistics important in design (IFD, hourly/daily probability of no rain, hourly/daily mean/variance) are also checked against that observed.

6.4.1 Results

Annual Rainfall

A comparison between simulated annual rainfall for Brisbane and Sydney using a two-state versus a single-state model is shown in Table 6.1. When comparing observed and simulated statistics it is important to quantify the sampling uncertainty in the statistic. Accordingly, 90% confidence limits are reported as well as the simulated median value of the statistic, which were calculated by ranking 1000 replicated simulations each having the same length as the observed record. If the observed values lie between the upper and lower confidence limits, then the model is not inconsistent with the observed data. Note however, because the model has not been calibrated to the validation statistics, the median values do not necessarily have to follow the data. Table 6.1 shows that the observed mean is reproduced well by both models. However, the two-state model provides a marked improvement in reproducing the annual standard deviation. Of the statistics presented within Table 6.1, the auto correlation is least well produced by the two state models, with the observed value lying just within the upper 90% confidence limit for both Sydney and Brisbane. This could be attributed to sampling variability, that is, this has just occurred by chance. Alternatively, the two-state DRIP model may be inducing too much auto-correlation from one year to the next.

Table 6.1 Simulated and Observed Annual Mean, Standard Deviation and Auto correlation with 90% confidence limit for Sydney 1859-1998 and Brisbane 1887-1993.

	Sydney						Brisbane					
	Mean (mm)		Standard Deviation (mm)		Auto correlation		Mean (mm)		Standard Deviation (mm)		Auto correlation	
Observed	1221.7		332.0		0.10		1125.4		350.7		0.05	
Single State Simulated	1228.2	1265.3	259.5	285.7	0.00	0.13	1091.2	1134.0	251.3	288.3	-0.01	0.16
		1191.6		233.0		-0.15		1051.1		222.3		-0.16
Two State Simulated	1232.0	1311.3	335.0	393.5	0.30	0.45	1125.0	1204.0	313.4	378.6	0.22	0.40
		1163.0		282.2		0.10		1052.0		258.0		0.03

Observed annual rainfall distributions for Sydney and Brisbane are compared to those produced by the DRIP single and two-state simulation in Figure 6.3 and Figure 6.4. For both sites, incorporating two-state simulation provides a marked improvement to the single state distributions. The Sydney (two-state) expected simulation closely matches the observed values for most of the distribution with some suggestion of overprediction of low rainfalls. Likewise, the Brisbane (two-state) simulation reproduces the distribution for the majority of the curve except for the lower extremity where the observed rainfall lies just below the 90% confidence limit. The problems with the low rainfall may be due to misclassification of an actual wet year as a dry year. Alternatively the hypothesis that only the distributions of inter-event and storm duration are affected by climate state may need to be reconsidered.

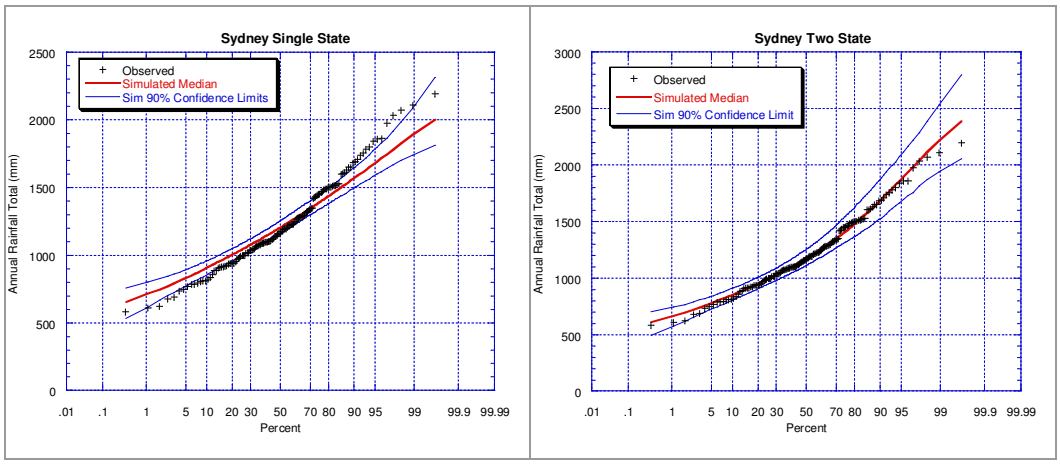


Figure 6.3 Observed Annual Rainfall versus 1000 Replicated Simulations; Sydney Observed 1859-1998 vs. 140yr Repeated Simulation for (a) Single State DRIP and (b) Two State DRIP

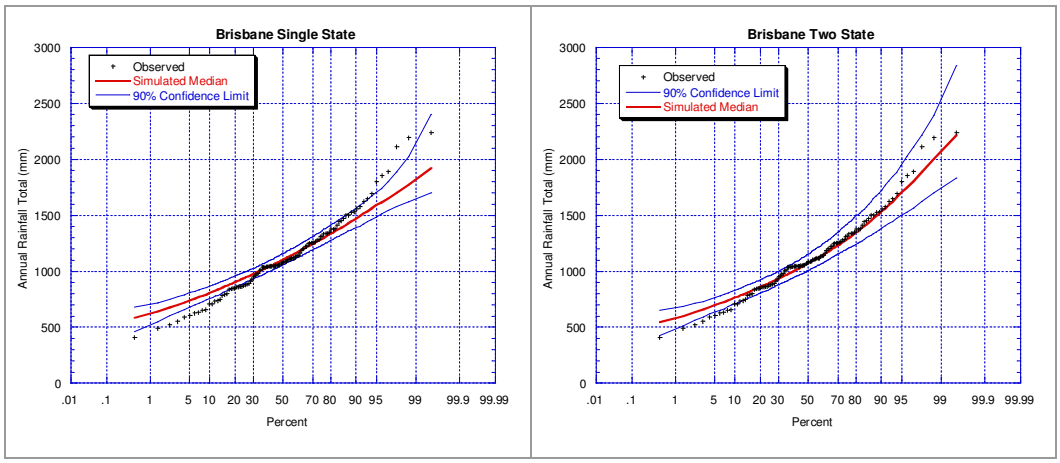


Figure 6.4 Observed Annual Rainfall versus 1000 Replicated Simulations; Brisbane Observed 1887-1993 vs. 107yr Repeated Simulation for (a) Single State DRIP and (b) Two State DRIP

IFD Curves

Comparison of observed and simulated Intensity-Frequency-Duration curves provides a thorough test of the model's ability to simulate the temporal nature of rainfall. As yearly extremes of short-duration rainfall are not calibrated within DRIP, accurate representation of extreme distributions adds credence to the model's usefulness in design. Observed and simulated (two-state) curves are shown in Figure 6.5. As can be seen, with the exception of the Sydney 72 hour IFD curve, the simulated results correspond very well with the observed values.

Aggregated Rainfall Statistics and Probability of No Rain

Other general statistics such as daily and hourly rainfall mean and standard deviation need to be well reproduced if the model is to be considered robust. Also, the probability of no rainfall over different timescales gives an indication of the validity of the model conceptualisation. The ability of the model to account for seasonality can be tested by comparing these statistics on a monthly basis. Hourly and daily aggregated rainfall statistics are shown in Figure 6.6 and Figure 6.7 respectively. The simulated versus observed probability of no rain is shown in Figure 6.8. The simulated values satisfactorily reproduce all of the statistics. The simulated probability of no rain is slightly lower than that observed for all the months shown. It is hypothesised here that the observed dry probability statistic is biased upwards due to missing values being predominantly wet. The seasonal variation for each observed statistic is matched closely by the majority of simulated values.

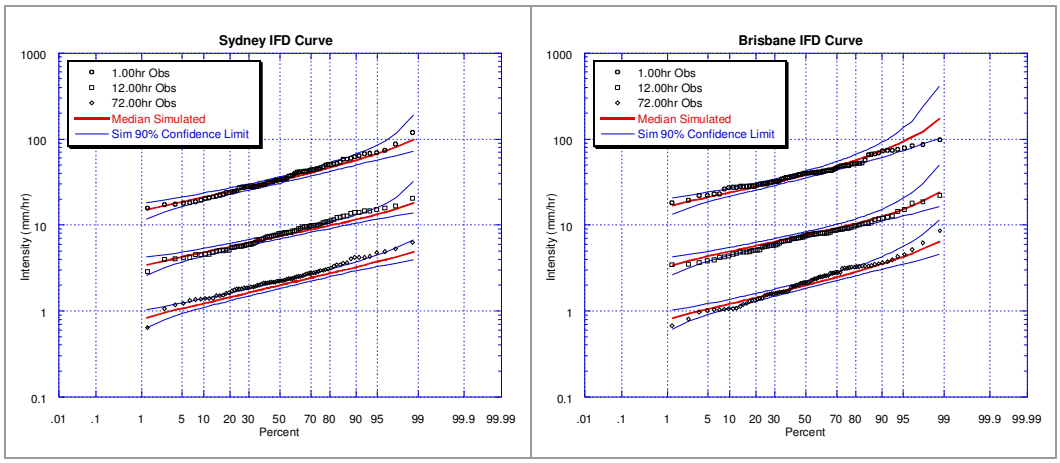


Figure 6.5 Observed and Simulated IFD Curves for 1hr, 12hr and 72hr duration at sites (a) Sydney and (b) Brisbane

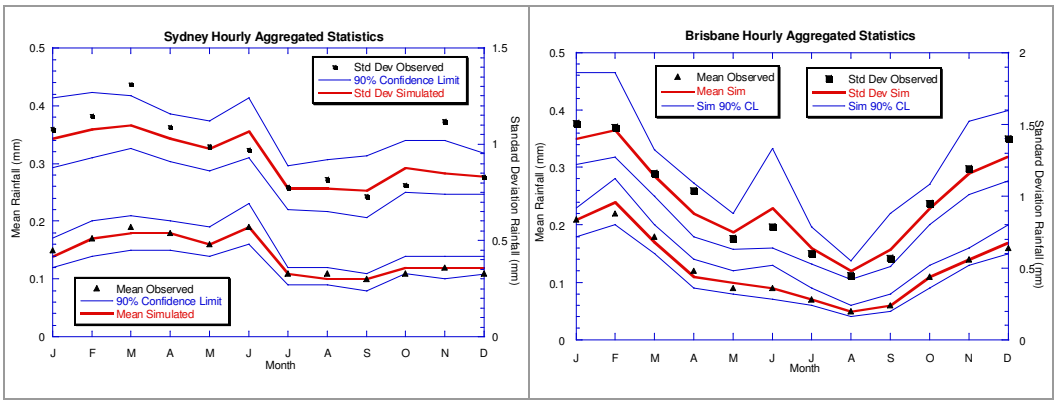


Figure 6.6 Observed and Simulated Hourly Rainfall Mean and Standard Deviation for (a) Sydney and (b) Brisbane

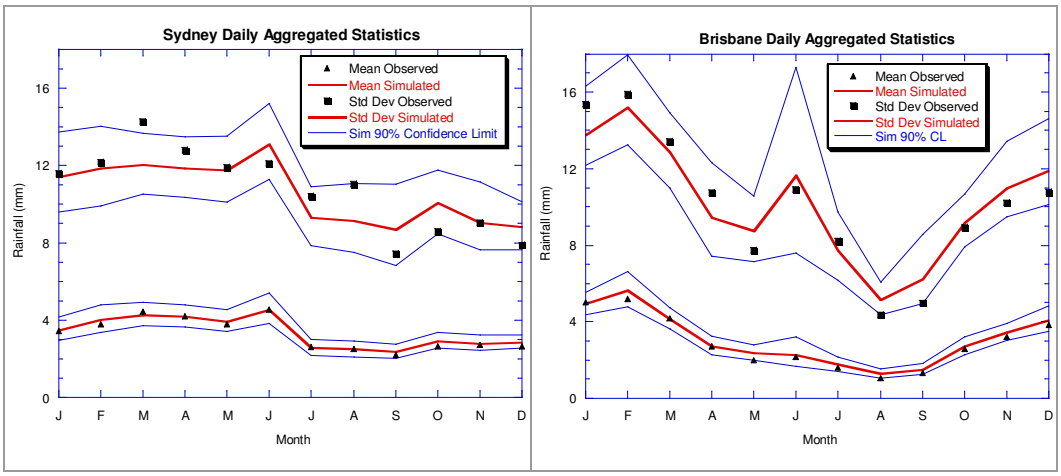


Figure 6.7 Observed and Simulated Daily Rainfall Mean and Standard Deviation for (a) Sydney and (b) Brisbane

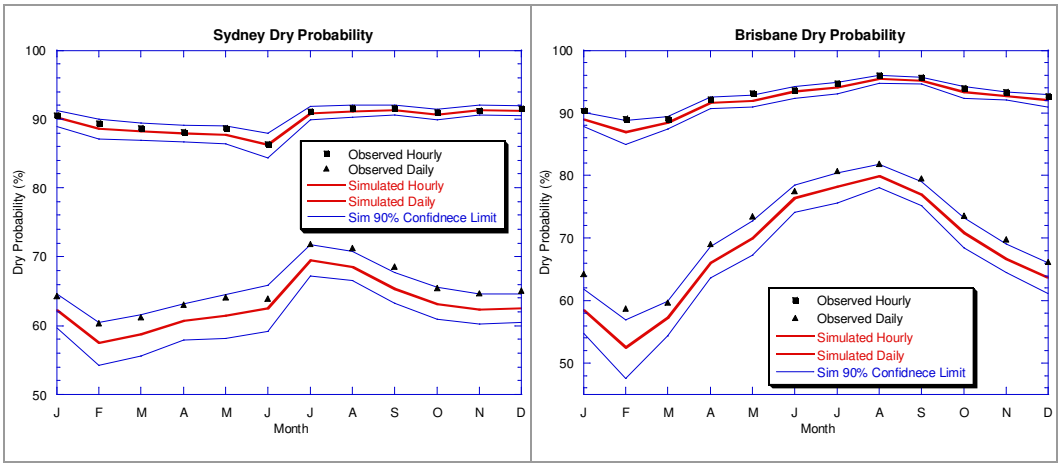


Figure 6.8 Observed and Simulated Hourly and Daily Dry Probabilities for (a) Sydney and (b) Brisbane

6.4.2 Discussion of HMM case study

The incorporation of a two-state HMM in the event-based rainfall model DRIP has enabled an improved reproduction of the annual rainfall distribution. This improvement is attributed to the ability of the HMM to conceptually incorporate the intra- and inter-annual persistence apparent in Australian rainfall. The HMM was applied only to those processes within DRIP that influence the number of storm arrivals, namely storm and inter-storm duration. The credibility of the model was tested by comparison of simulated and observed values of IFD curves, monthly aggregated rainfall statistics and the probability of no rain. Simulated values are shown to correspond well with observed values and also show general improvement when compared to single state DRIP.

Of the annual statistics tested, the two-state DRIP model auto-correlation appears to be over inflated. An approach to reduce this auto-correlation, whilst also modelling inter-annual persistence, would be to allow the state to vary within the water year also. That is, there is a Hidden Markov monthly state given the overlying annual Hidden Markov state, relaxing the rigidity of the intra-annual persistence structure of the current two-state DRIP-HMM. Such a model would allow a greater degree of flexibility in reproducing variability, yet would not force the degree of dependence from one year to the next to produce this variability.

6.5 Case Study II : DRIP conditioned on model averaged RHMM state series

This section examines the conditioning of the DRIP model on the model averaged posterior state series of the fitted correlation RHMM presented in the previous chapter.

6.5.1 Results: Fitted Correlation RHMM-DRIP

The two sets of RHMM (Close and Far) series, shown in Figure 6.2b and c, were used in the calibration of the DRIP model, weighting wet and dry storm duration and inter-storm duration parameters. The two state DRIP model is validated by comparison with observed annual rainfall and the simulations produced by the single state DRIP specification. Short-scale statistics produced were similar to those observed for Case Study I. However, they are not presented here because reproducing the variability of annual rainfall is the main focus of this section.

Observed annual rainfall distributions for Sydney and Brisbane are compared to those produced by the DRIP single and two-state simulation in Figure 6.9a, b and c. There are two sets of two-state results (Figure 6.9b and c) pertaining to the Close and Far grouping of sites respectively. The annual summary statistics of the DRIP simulations are also presented within Table 6.2.

For the Sydney annual distributions, an improved reproduction of observed data is produced using the Sydney Far state series, with the observed values predominantly lying on the simulated median line. However, the Sydney Close results do not significantly improve upon those of the single state calibration, underestimating the variability of the annual rainfall. The RHMM Brisbane Close simulation shows a significantly higher degree of variance (greater slope) than the single state simulation, yet the overall variance remains underestimated. The Brisbane Far simulation on the other hand does not significantly improve upon the annual distribution of the single state model, again underestimating the overall variance. The annual summary statistics shown in Table 6.2 reflect these results, with the median standard deviation consistently underestimating that observed for all simulations. The auto correlation statistic, for the simulations which showed the greatest variance (Sydney Far and Brisbane Close) were

within acceptable bounds. This contrasts the HMM results where the autocorrelation was high compared to that observed.

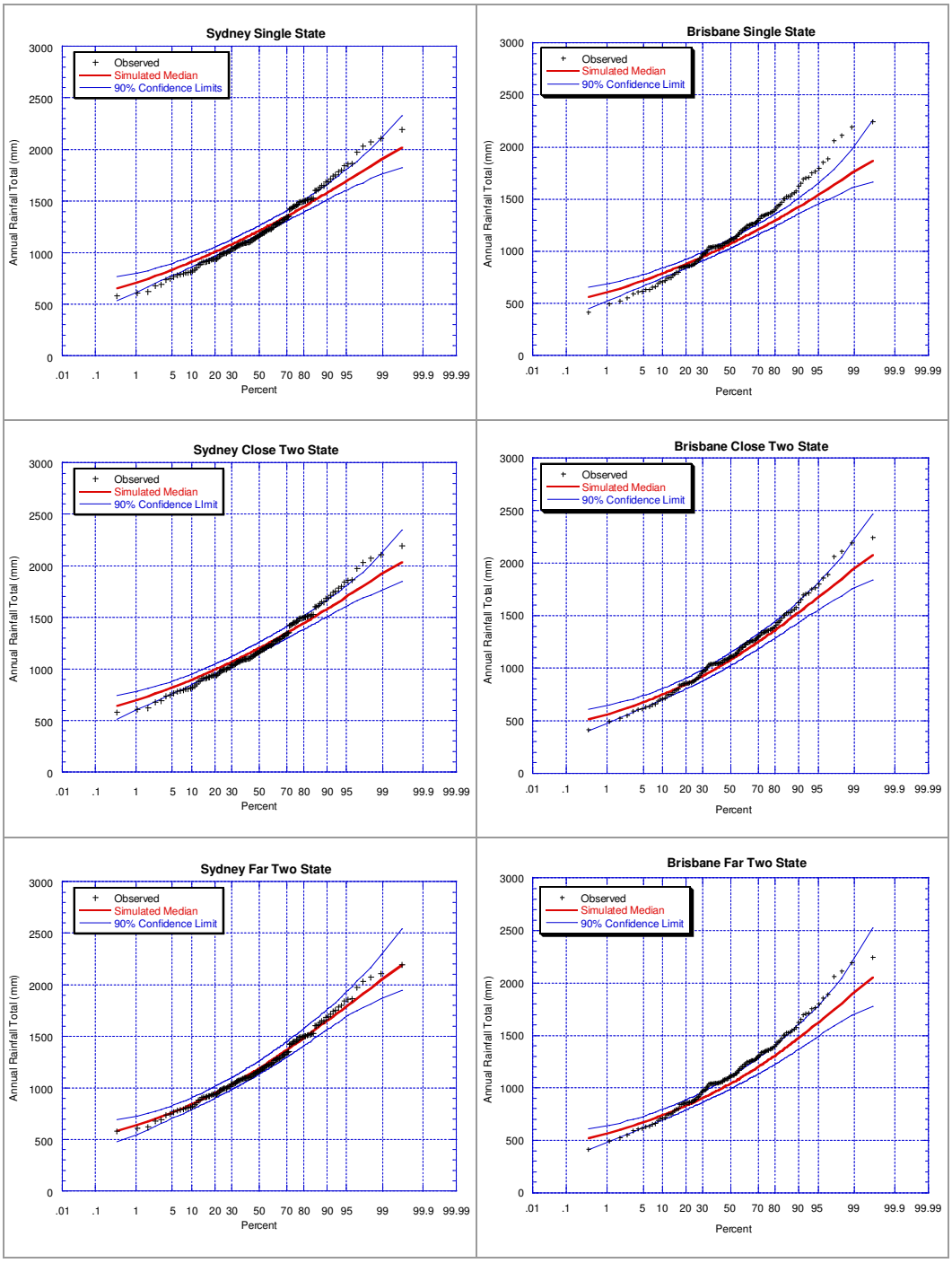


Figure 6.9 Observed Annual Rainfall versus 1000 Replicated Simulations; Sydney Observed 1859-1998 vs. 140yr Repeated Simulation, Brisbane Observed 1860-1993 vs. 134yr Repeated Simulation for (a) Single State DRIP and (b) RHMM-DRIP Close sites and (c) RHMM-DRIP Far sites. RHMM uses fitted correlation.

Table 6.2 Simulated and Observed Annual Mean, Standard Deviation and Auto correlation with 90% confidence limit for Sydney 1859-1998 and Brisbane 1860-1993.

	Sydney						Brisbane					
	Mean (mm)		Standard Deviation (mm)		Auto correlation		Mean (mm)		Standard Deviation (mm)		Auto correlation	
Observed	1221.7		332.0		0.10		1154.2		371.0		0.06	
Two State Close	1229.1	1273.5	268.4	297.1	0.03	0.17	1119.2	1172.1	306.0	346.9	0.11	0.25
		1186.1		241.8		-0.11		1069.0		273.9		-0.03
Two State Far	1228.6	1286.2	315.0	350.2	0.17	0.31	1084.0	1141.7	292.6	337.5	0.19	0.34
		1177.2		285.9		0.03		1031.8		256.9		0.03

6.5.2 Discussion: Fitted Correlation RHMM-DRIP

How can these contrasting results be explained? That is, why does the Sydney Far state series enable a better fit of DRIP simulations to the annual data than Sydney Close data (and *vice versa* for Brisbane)? Also, why does the single region HMM of Case Study I reproduce the variance of Sydney to a greater degree of that of the Sydney Far calibration, yet have markedly different state series?

The better reproduction of the autocorrelation by the Sydney Far and Brisbane Close simulations compared to the single region HMM is attributed to the more variable state series for the RHMM. The HMM state series shows a high degree of persistence, whereas the RHMM generally shows more frequent switching between states. This results in a lower degree of autocorrelation.

The general underestimation of the variance by all simulations presented can in part be attributed to the non-stationary nature of annual rainfall. That is, the simulations presented were based on calibration of the RHMM to records of shorter length than those presented in Table 6.2. Likewise, the length of the pluviograph records used for DRIP calibration were significantly shorter than the annual records used for comparison. The variance of annual rainfall over this longer period is greater than that over the length of the DRIP record. However, it was considered that using the entire annual records for comparison was a stronger test for the DRIP-RHMM model, cross-validating using data not used in calibration.

In view of the Sydney single region HMM and Sydney Far RHMM results, there obviously is more than one way to produce the variance observed in the annual rainfall. Providing that lower mean rainfall within the pluviograph record coincides with dry RHMM years, and higher mean rainfall coincides with wet RHMM years, a greater degree of variability will be produced by the two state DRIP model. However, it is not clear which of the two methods, using model averaged RHMM or HMM state series is superior; this is a question of hypothesis testing (or model selection).

Central to the choice of model for conditioning DRIP is the misclassification of wet and dry years. If the state series misclassifies events, especially large events on the tails of the inter-storm and storm distributions, this will bias the means of the DRIP dry and wet distributions towards one another, thus producing less variability. As the means move closer to each other, the distribution becomes more like that of the single state model.

As the RHMM uses a 'hidden' state series to produce the variability present in annual rainfall, it is not clear upon examination whether the state series produced from calibration is appropriate to condition the calibration of DRIP. Other empirical relationships observed must be used to check the classification. The El Niño Southern Oscillation (ENSO) has been identified as having influence on Australian rainfall and runoff (*Ropelewski and Halpert*, 1996) – with El Niño years generally being correlated with low rainfall years, and high rainfall in La Niña years for this area (*Kiem and Franks*, 2001b,, 2001a). Misclassification is now examined in terms of ENSO, comparing the state series of the RHMM versus the ENSO classification. A state series, classifying years into either El Niño, Neutral and La Niña years is shown in Figure 6.10. The state series spanning 1950-2001 was derived from the six month (October-March) average values of the MEI as described in *Kiem and Franks* (2001b). The 1900-1949 classifications were based on the six month (October-March) average values of the Nino3 indicator also described in *Kiem and Franks* (2001b). These classifications were used to stratify the pluviograph data over the May-April water year (as used for the HMM and RHMM calibration), and calculate average dry spell duration, storm duration and intensity for the El Niño, Neutral and La Niña periods. Likewise, the RHMM state series used in this section was used to calculate average wet and dry event durations and intensities. The resulting average event properties are presented in Table 6.3.

Comparing the Sydney Close state series to the HMM and Sydney Far RHMM, Sydney Close shows a lot less variability of state (lower transition probabilities) than that of the HMM or Sydney Far. Conversely, the Brisbane Close series shows greater variability than both the single region HMM and the Brisbane Close RHMM series. Of the state series, Brisbane Close resembles the ENSO series the closest, with Sydney Far also showing a similar degree of variability. Given that significant El Niño's occurred in 1977, 1979 and 1982, we would expect there to be considerably lower rainfalls during these years. In turn, a state series indicating a high probability of being in the dry state would be expected. This is the case for the HMM and Sydney Far state series, however the Sydney Close state series is predominantly wet during this period, not venturing much below the 0.5 mark.

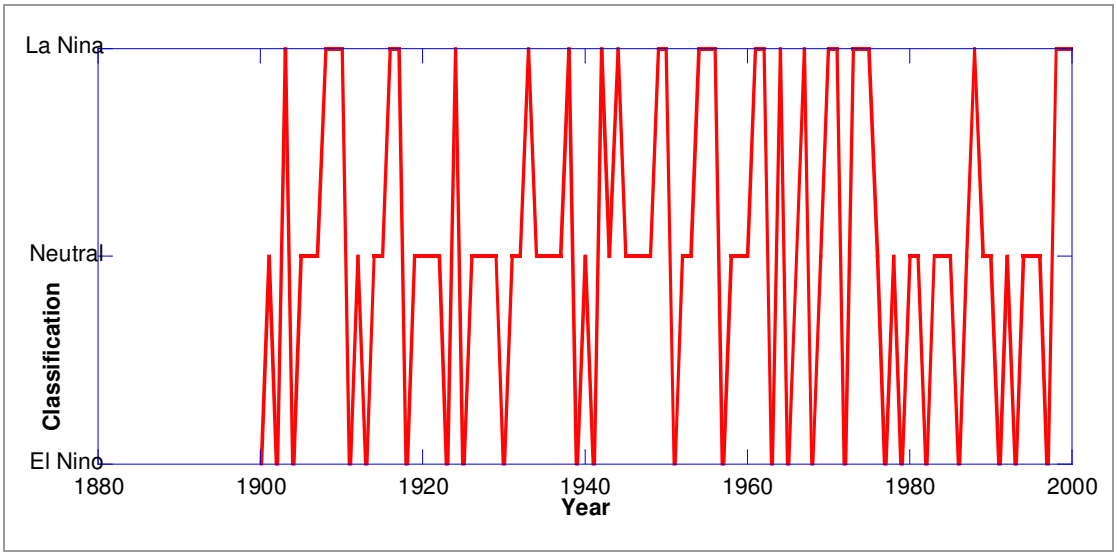


Figure 6.10 ENSO classified years according to Nino3 (1900-1949) and MEI (1950-2000) in *Kiem and Franks* (2001b)

Table 6.3 Sydney and Brisbane average storm event data stratified into El Niño/La Niña and RHMM wet and dry years.

Site			Dryspell Duration	Storm Duration	Storm Intensity
Sydney	Entire record		45.00	4.10	1.71
	El Niño years		49.40	4.03	1.74
	Neutral years		45.28	4.06	1.71
	La Niña years		41.61	4.22	1.69
	Close RHMM	Dry	46.04	4.11	1.61
		Wet	43.85	4.10	1.84
	Far RHMM	Dry	47.73	3.99	1.64
		Wet	42.45	4.20	1.78
Brisbane	Entire record		53.58	3.44	2.41
	El Niño years		60.92	3.11	2.40
	Neutral years		53.71	3.49	2.42
	La Niña years		48.52	3.58	2.38
	Close RHMM	Dry	55.08	3.32	2.44
		Wet	51.65	3.59	2.37
	Far RHMM	Dry	55.70	3.42	2.42
		Wet	50.60	3.45	2.40

It appears the Sydney Close state series is misclassifying data in strong El Niño years, which is in turn reducing the variability produced from the DRIP simulation. The other two series (HMM Sydney and RHMM Sydney Far), having classified these major years correctly are more likely to have more variable annual distributions. Although the state series for HMM Sydney and RHMM Sydney Far are quite different, these two series are presumably coinciding in appropriate years (large wet or dry events). The remaining years which do not coincide must not have significant influence on the DRIP mean.

The average event properties of the ENSO and RHMM classified data are examined in Table 6.3. Comparing the average dryspell duration for El Niño and La Niña years, the dryspell is longer for El Niño years, for both Sydney and Brisbane. The storm duration is shorter in El Niño years, whilst storm intensity is marginally greater. The RHMM weighted dry/wet storm durations show similar trend to the El Niño/La Niña, with longer dryspells for the dry state. The dry/wet storm durations also show similar trends

to the ENSO classified data, with Sydney Far and Brisbane Close producing wet storm durations on average of greater length than the dry storm durations. The exceptions lie with Sydney Close and Brisbane Far, the sites showing poorest reproduction of annual variability, with similar average storm durations for wet and dry years. The wet/dry average storm intensities show little variability between wet and dry years and ENSO states across all site calibrations, with Sydney intensities being slightly greater for wet years, whilst being slightly less for the Brisbane RHMM. The significance of the average dryspell and storm duration changes can be considered by noting that the average number of storms per year is defined as:

$$Average\,no\,storms/\,year = \frac{24*365}{Mean(dryspell) + Mean(storm\,duration)} \qquad (6.4).$$

Due to the relative size of the dryspell compared to the storm duration, changes in the dryspell have a greater influence over the number of storms in a year. If the dryspell rose with El Niño while the storm duration dropped by the same magnitude, the same number of storms in the year would result. This is not the case here, with the El Niño/ La Niña average number of storms being 164/191 for Sydney and 137/168 for Brisbane. Thus La Niña years have 17% and 23% more storms on average per year than in El Niño years.

The lack of separation of Sydney Close and Brisbane Far average wet and dry storm duration data is *prima facie* evidence that these series have a higher degree of misclassification compared to the other site groupings. Of course, there is no such thing as the correct state series, as the RHMM is just a stochastic conceptualisation of reality. Likewise, the ENSO state series is just an empirically observed phenomenon, not necessarily reflecting the correct state series. However, empirical relationships such as the ENSO phenomenon are useful in identifying shortcomings within the RHMM.

The apparent misclassification of El Niño/La Niña rainfall event data demonstrates two points. Firstly, if the pluviograph data is misclassified into a wet year, when there are extreme dry events occurring, the DRIP simulation will be constrained in reproducing variability as was possible if otherwise classified. Secondly, climate indices such as the MEI used in *Kiem and Franks* (2001a) are useful in checking for misclassification within the state series of the HMM and its variants. This checking allowed clear

identification of periods in the state series which may not be consistent with the known influence of ENSO.

This comparison against data stratified into El Niño/Neutral/La Niña years raises the possibility of calibrating DRIP using a two state El Niño/La Niña or three state El Niño/Neutral/La Niña model. This removes the need to calibrate the RHMM model. However, determination of which sites should be conditioned using this (ENSO state series) returns us to the motivating reason for using the RHMM – how do we choose sites to be grouped into homogenous climate regions? Rather than using the ENSO series as a replacement for RHMM, it is recommended that the ENSO series is incorporated into RHMM calibration so that regions where ENSO effects are identifiable can be identified formally according to BMS.

6.5.3 Results: Exponential Correlation Decay RHMM-DRIP

To demonstrate further the influence of differing state series on the calibration of DRIP, two-state DRIP is now calibrated using the model averaged state series of the exponential correlation decay RHMM. The state series for the Close and Far site groupings are shown in Figure 6.11a and b respectively.

Although in preliminary testing of this model the exponential decay correlation structure produced greater marginal likelihoods than that of the fitted correlation model, the state series produced were affected by anomalous behaviour related to the complex correlation structure. Some site groupings (especially Brisbane Close) produced state series that were not in keeping with the intended hierarchical structure of the RHMM. The model averaged state series, with the exception of Brisbane Far, show a much greater degree of variability than the fitted correlation RHMM series, with Brisbane Close and Sydney Far resembling oscillating noise about the 0.5 probability line.

Observed annual rainfall distributions for Sydney and Brisbane are compared to those produced by the DRIP two-state simulation for Close and Far site groupings in Figure 6.12a and b. The annual summary statistics of the DRIP simulations are also presented within Table 6.4.

The Brisbane simulations do not produce annual rainfall distributions significantly different from that of single state DRIP – with the variance underestimated. Conversely,

both Sydney simulations produce annual rainfall distribution with a variance (and mean) that is much greater than that observed, with the right hand tail showing significantly larger rainfall years.

6.5.4 Discussion: Exponential correlation decay RHMM-DRIP

Apparently, the Sydney DRIP calibrations have overcompensated for the variability within the annual rainfall record, with Sydney Close showing the greatest degree of over dispersion. The state series used here allows a greater degree of separation of DRIP parameters (and hence wet and dry means), perhaps indicating that the pluviograph data is more appropriately classified. Although the data may be more appropriately classified, it is not clear whether this state series should be used in calibration when other state series produce simulated annual rainfall distributions much closer to that observed. Again this is a question of hypothesis testing/model selection, which was not undertaken here as the DRIP model had not yet incorporated the Bayesian parameter uncertainty framework to apply BMS. These comments regarding hypothesis testing are equally applicable for Brisbane, where the use of the two-state series in calibration does not show any significant increase in quality of reproduction of annual variance.

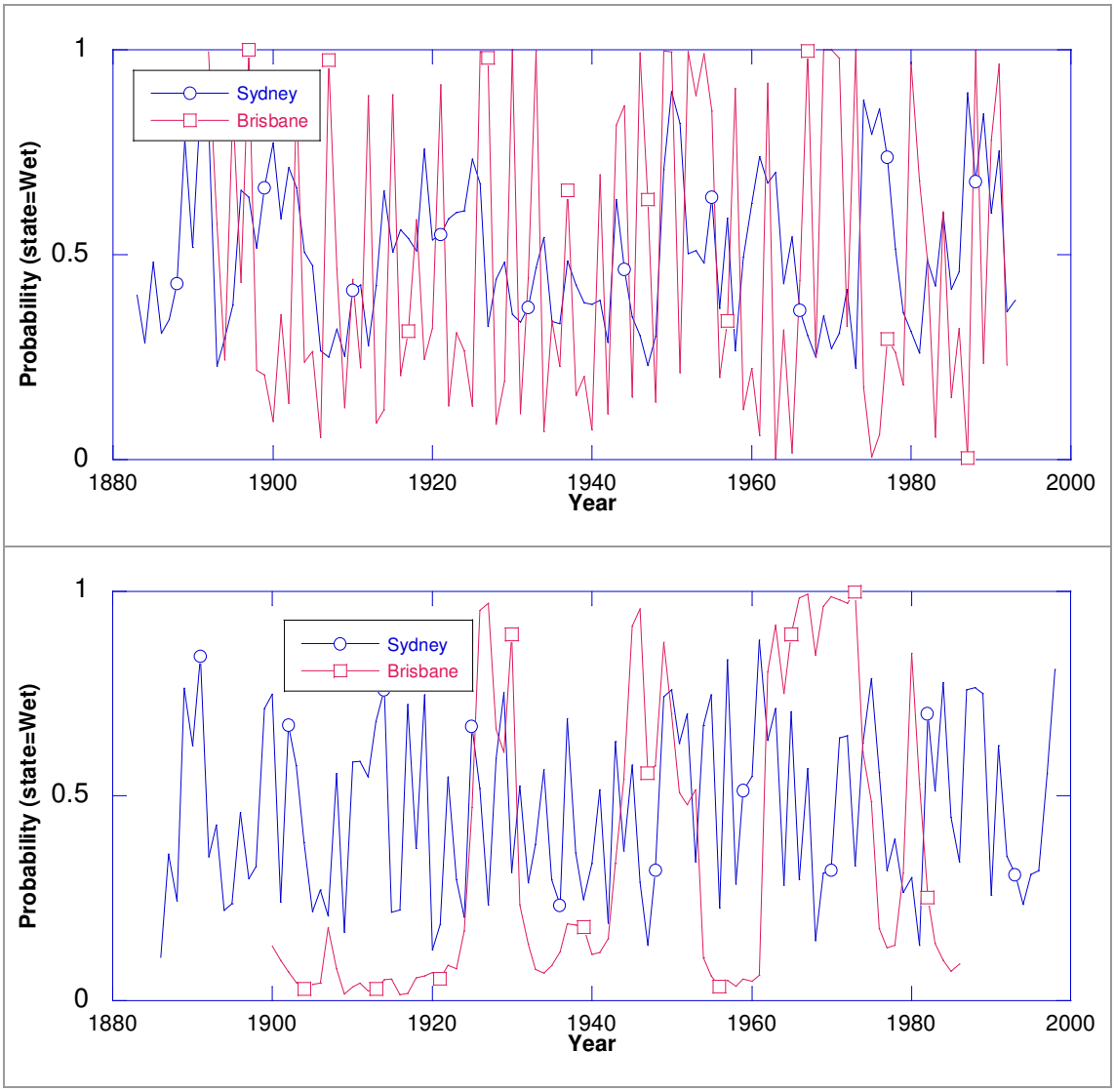


Figure 6.11 Posterior State Series probabilities for the model averaged RHMM with exponential correlation decay for (a) Close and (b) Far site groupings.

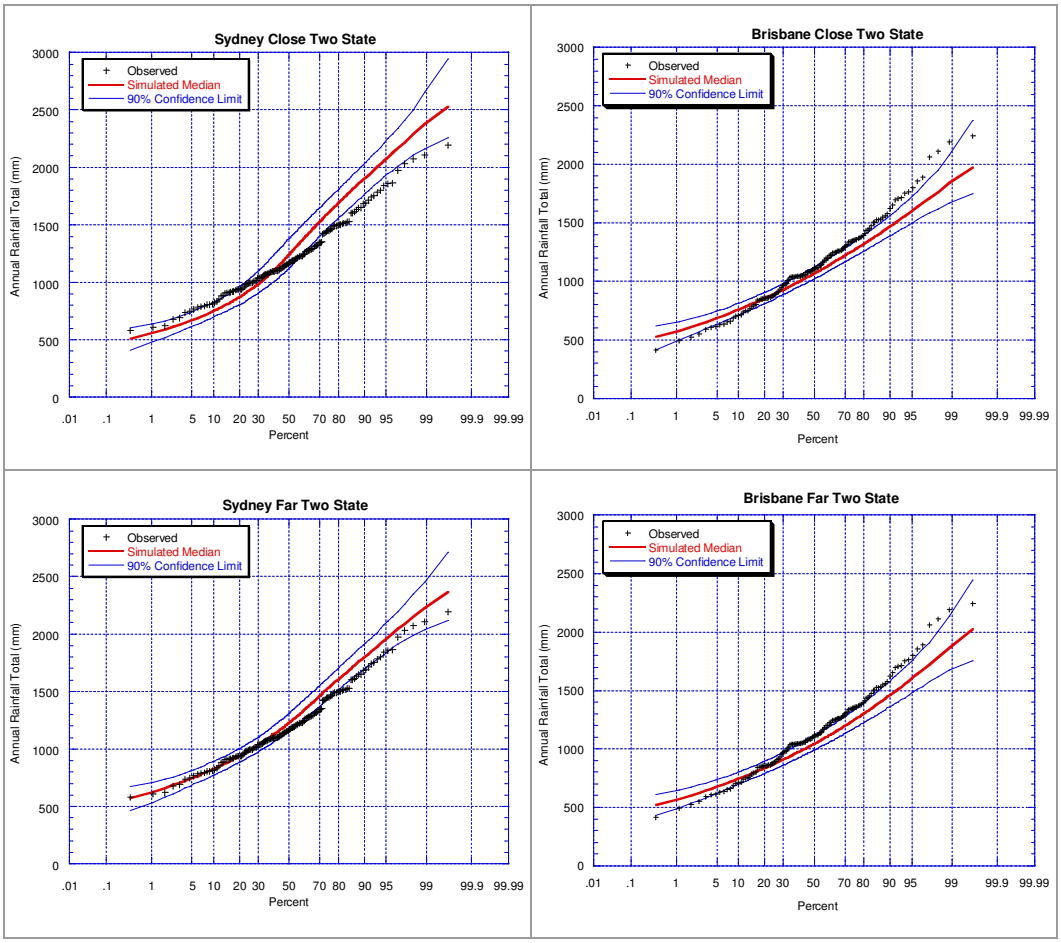


Figure 6.12 Observed Annual Rainfall versus 1000 Replicated Simulations; Sydney Observed 1859-1998 vs. 140yr Repeated Simulation, Brisbane Observed 1860-1993 vs. 134yr Repeated Simulation for exponential correlation decay RHMM-DRIP (a) Close sites and (b) RHMM-DRIP Far sites.

Table 6.4 Exponential Decay Simulated and Observed Annual Mean, Standard Deviation and Auto correlation with 90% confidence limit for Sydney 1859-1998 and Brisbane 1860-1993.

	Sydney						Brisbane					
	Mean (mm)		Standard Deviation (mm)		Auto correlation		Mean (mm)		Standard Deviation (mm)		Auto correlation	
Observed	1221.7		332.0		0.10		1154.2		371.0		0.06	
Two State Close	1294.4	1386.0	442.4	481.5	0.37	0.47	1101.3	1142.7	280.3	315.6	0.03	0.17
		1205.3		400.6		0.26		1058.0		248.4		-0.11
Two State Far	1292.0	1355.3	388.0	427.2	0.19	0.32	1082.8	1141.9	286.8	325.8	0.16	0.31
		1223.0		348.4		0.05		1028.4		250.1		0.02

A possible reason for the exponential decay RHMM for Sydney producing overdispersed annual rainfall simulations, is that the two-state assumption for both the

storm duration and inter-storm time is not justified. That is, the two-state assumption may be justified for only the storm duration or only the inter-storm duration. A simulation presenting the Sydney Close DRIP-RHMM calibration, with single state storm duration and two-state inter-storm time is given in Figure 6.13.

The single state storm duration and two-state inter-storm time simulation reproduces the annual rainfall distribution accurately. This may indicate that the two-state assumption is justified only for the inter-storm duration, and not the storm duration. Once again, whether or not this is the case is a question of model selection. This requires parameter uncertainty being incorporated into the DRIP model so as to allow determination of the model weights according to BMS.

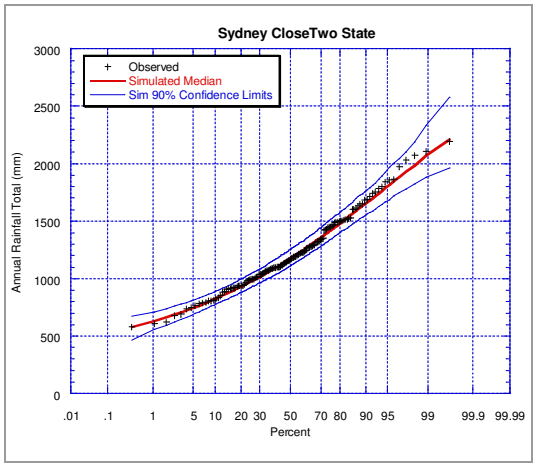


Figure 6.13 Observed Annual Rainfall versus 1000 Replicated Simulations; Sydney Observed 1859-1998 vs. 140yr Repeated Simulation for RHMM-DRIP Close sites with two state inter-storm duration only.

6.6 Discussion of case studies

The case studies demonstrate two major points:

1. Misclassification of major pluviograph events (on the tail of the associated distributions) into wet or dry years can reduce the amount of variability produced by the two-state DRIP simulation. As the DRIP calibration is currently independent of the calibration of the hidden state series, state series inconsistent with the pluviograph event data can result.

2. Whether or not DRIP should include two-state/single state, inter-storm durations/storm durations/intensity distributions is a question of hypothesis testing. Currently the DRIP calibration framework does not allow formal model comparison (through BMS).

An approach which would reduce the amount of pluviograph data misclassified according to state series would be to calibrate the HMM (or RHMM) concurrently with the DRIP model. This would ensure that the state series is consistent with both the annual rainfall used previously in the HMM calibration, and also the pluviograph data used for DRIP.

Although beyond the scope of this thesis the joint calibration of DRIP and HMM is described below. The likelihood (demonstrated here for the HMM) calculation is generalised from Section 3.4.1 to include the rainfall event data $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_t)$ where $\mathbf{x}_t$ is the collection of event data for a particular year t. This possibly empty subvector $\mathbf{x}_t = (x_{a(t)}, ..., x_{b(t)})$ contains data points $a(t)$ to $b(t)$, where $a(\cdot)$ and $b(\cdot)$ are indexes for the first and last data point in each year. A typical likelihood term, analogous to (3.12) for the original HMM (with $\mathbf{y}_t$, the annual rainfall, and $\mathbf{x}_t$ both being considered data) is:

$$\begin{aligned}
&p\left(\mathbf{y}_t, \mathbf{x}_t \mid \mathbf{Y}_1^{t-1}, \mathbf{X}_1^{t-1}, \boldsymbol{\theta}_{HMM}, \boldsymbol{\theta}_{DRIP}\right) \\
&\quad = \sum_{r_t} p\left(\mathbf{y}_t, \mathbf{x}_t \mid r_t, \mathbf{Y}_1^{t-1}, \mathbf{X}_1^{t-1}, \boldsymbol{\theta}_{HMM}, \boldsymbol{\theta}_{DRIP}\right) \; p\left(r_t \mid \mathbf{Y}_1^{t-1}, \mathbf{X}_1^{t-1}, \boldsymbol{\theta}_{HMM}, \boldsymbol{\theta}_{DRIP}\right) \\
&\quad = \sum_{r_t} p\left(\mathbf{y}_t, \mathbf{x}_t \mid r_t, \boldsymbol{\theta}_{DRIP}\right) \qquad\qquad p\left(r_t \mid \mathbf{Y}_1^{t-1}, \mathbf{X}_1^{t-1}, \boldsymbol{\theta}_{HMM}, \boldsymbol{\theta}_{DRIP}\right) \\
&\quad = \sum_{r_t} p\left(\mathbf{y}_t \mid \mathbf{x}_t, r_t, \boldsymbol{\theta}_{DRIP}\right) p\left(\mathbf{x}_t \mid r_t, \boldsymbol{\theta}_{DRIP}\right) \quad p\left(r_t \mid \mathbf{Y}_1^{t-1}, \mathbf{X}_1^{t-1}, \boldsymbol{\theta}_{HMM}, \boldsymbol{\theta}_{DRIP}\right) \\
&\quad \approx \sum_{r_t} p\left(\mathbf{y}_t \mid r_t, \boldsymbol{\theta}_{DRIP}\right) \left(\prod_{i=a(t)}^{b(t)} p\left(x_i \mid r_t, \boldsymbol{\theta}_{DRIP}\right) \right) p\left(r_t \mid \mathbf{Y}_1^{t-1}, \mathbf{X}_1^{t-1}, \boldsymbol{\theta}_{HMM}, \boldsymbol{\theta}_{DRIP}\right)
\end{aligned} \tag{6.5}$$

where $\boldsymbol{\theta}_{HMM}$ and $\boldsymbol{\theta}_{DRIP}$ are the HMM and DRIP parameters respectively. The final line assumes independence of the annual rainfall data $\mathbf{y}_t$ and the rainfall event data $\mathbf{x}_t$ given the regional state. This independence assumption may require modification as annual rainfall data has dependence on pluviograph data measured nearby (given that annual data can be considered accumulated pluviograph data). It is a question of whether the regional state explains this dependency sufficiently. The final line is equivalent to the

weighting undertaken in (6.2) with the addition of the annual data. The remainder of the likelihood calculation (revolving around the transition probabilities) remains the same, with analogous calculations to (3.11) and (3.13) used. The annual rainfall distribution $p(\mathbf{y}_t \mid r_t, \boldsymbol{\theta}_{DRIP})$ could be modelled as a Gaussian distribution, as was done for the HMM, however differing in that the annual mean and variance parameters are inferred from DRIP simulation i.e. $p(\mathbf{y}_t \mid r_t, \boldsymbol{\theta}_{DRIP}) \sim N(y_t; \mu(r_t, \boldsymbol{\theta}_{DRIP}), \sigma(r_t, \boldsymbol{\theta}_{DRIP}))$. Here the mean μ and standard deviation σ can be estimated from a suitably long DRIP simulation given the parameters $\boldsymbol{\theta}_{DRIP}$ and state r_t. Alternatively, nonparametric estimation of $p(\mathbf{y}_t \mid r_t, \boldsymbol{\theta}_{DRIP})$ from DRIP simulation would obviate the need for an annual rainfall Gaussian distribution assumption.

Returning to the subject of hypothesis testing, incorporating a parameter uncertainty framework (applying priors to all model parameters such that the posterior is proper and identifiable), and using the techniques of BMS and model averaging, would allow formal testing of the various DRIP model hypotheses. This method, in tandem with the joint HMM-DRIP calibration technique mentioned above would thus allow the most appropriate state series to be grouped with the most appropriate DRIP model, considering both the pluviograph and annual data.

Has the conditioning of DRIP on the hidden state of the HMM (or its variants) produced a model which improves reproduction of the observed inter-annual variability? The increased variability observed in the case studies compared to the single state DRIP would tend to indicate so. Whether or not this is adequate, or would better be modelled using some other hypothesis, is question to be addressed by future research when formal model and parameter uncertainty is incorporated into the DRIP calibration framework.

Finally, it is noted that the conditional weighting of an event based model using the output series of a HMM (or its variants) can be used for any event based model which uses likelihood based inference.

6.7 Conclusion

This chapter has demonstrated the conditioning of the event based rainfall model DRIP on the hidden state series of the original two-state HMM of *Thyer* (2001) and model

averaged hidden state series of the RHMM presented in Chapter 5. Both of these case studies were based on pluviograph data collected at Sydney and Brisbane.

The first case study (using the hidden state series of the HMM) demonstrated an improved reproduction of the annual rainfall distribution. This improvement is attributed to the ability of the HMM to conceptually incorporate the intra- and inter-annual persistence apparent in Australian rainfall. The HMM was applied only to those processes within DRIP that influence the number of storm arrivals, namely storm and inter-storm duration. The credibility of the model was tested by comparison of simulated and observed values of IFD curves, monthly aggregated rainfall statistics and the probability of no rain. Simulated values were shown to correspond well with observed values and also showed general improvement when compared to single state DRIP. Nonetheless, the lack of a formal method for selecting sites in the HMM resulted in a heuristic approach which required all sites to have similar transition probabilities but could not guarantee individual state series were coherent.

The second case used a model averaged RHMM state series derived independently of the DRIP model. It was found that the annual distribution produced by the two-state DRIP model showed some sensitivity to the choice of state series with under or overestimation of annual variability resulting. These differences were attributed to possible misclassification of pluviograph data into inappropriate wet or dry years. In some cases the variability produced by the two-state DRIP model was greater than that observed on the annual rainfall records, suggesting that the hypothesis of two-state storm durations and inter-storm durations can be questioned.

It is recommended that to avoid misclassification of DRIP events a joint RHMM-DRIP calibration be undertaken, thus allowing the DRIP data to have an influence on the hidden state series produced. Also, as parameter uncertainty has not been incorporated in DRIP calibration so far, model selection using BMS has not been possible. Applying BMS to the joint RHMM-DRIP model would thus identify state series that are more consistent with the pluviograph data and enable formal model comparison of DRIP variants conditional on the resultant state series.

Chapter 7 Conclusions

7.1 Introduction

This thesis has presented new approaches to stochastically modelling long-term persistence in rainfall at multiple sites. A feature of these approaches is that they allow for spatially non-homogenous conceptual climate effects. In addition this thesis has considered the problem of downscaling long-term persistence to short timescale stochastic rainfall models. The analyses were conducted using a Bayesian framework to explicitly account for parameter uncertainty and select between competing hypotheses.

This chapter summarises the principal findings of this thesis and examines future research directions.

7.2 Summary

7.2.1 Objectives

The main objective of this thesis was to address the spatially varying long-term effects of climate on rainfall over a range of timescales ranging from sub-hourly to multi-year.

Chapter 2 reviewed previously published methods of incorporating long-term persistence within stochastic rainfall models. The reviewed models differ from one another in the degree to which data is conditioned on previously occurring data. Some models directly relate current rainfall to rainfall in preceding timesteps (e.g. AR1). Other models use latent states (as a conceptual climate state), with the rainfall being considered a degraded observation of this state. The degree of persistence identified by these models depends on the lag of the dependence permitted by the modelling structure. Of these methods, few specifically address long-term persistence, with the majority focussing on day-to-day and seasonal non-stationarity. Of the models that do consider long-term persistence, non-homogeneity of the conceptual climate influence is rarely accounted for.

The hidden Markov model (HMM) introduced by *Thyer* (2001) for modelling annual rainfall was used to identify long-term persistence within Australian hydroclimatological series. This HMM generally requires multiple site rainfall input if the state series is to be identified given the relatively short length of annual rainfall

records (<100 years). However, a formal method of choosing sites to be included in a HMM analysis had not been devised prior to this study. Hence, developing such a method by generalising the HMM was a major objective of this study.

7.2.2 HMM generalisations: Switch HMM and Regional HMM

This thesis extends the previous work of *Thyer* (2001) who applied a multi-site HMM to model annual rainfall within Australia. This work differs from that of *Thyer* (2001) in that the generalisations introduced do not assume the same climate state at each site. The Switch HMM allows at-site anomalous states, whilst still maintaining a regional control. The Regional HMM, on the other hand, allows sites to be partitioned into different Markovian state regions. Chapter 4 provided a detailed discussion of these HMM generalisations, including derivation of likelihood functions for use in Bayesian analysis. The extent to which these new models capture the non-homogenous nature of inter-annual persistence is of primary interest in this thesis.

7.2.3 Bayesian modelling framework and model selection

Chapter 3 outlined the Bayesian framework that was used throughout this thesis. This framework was employed because it rigorously deals with parameter uncertainty when making inferences and evaluating competing hypotheses. Moreover, Bayesian model averaging provides a rational means for allowing for model and parameter uncertainty when making inferences. This model uncertainty framework was discussed in detail as it has rarely been used in hydrological studies.

An MCMC sampling technique, the random walk Metropolis-Hastings (MH) sampler was introduced and compared to the Gibbs sampler, the model calibration technique used in the study of *Thyer* (2001). The MH sampler was chosen in this study due to its comparative simplicity. Use of the MH sampler obviated the need to simulate the hidden state series as part of the sampling process, thus reducing the chance of occurrence of 'trapping states' encountered with the Gibbs sampler for HMM problems. The simplification arose because the MH sampler could employ the Baum-Welch likelihood formulation for the HMM which integrated out the hidden state.

A test study demonstrated some of the perils involved in Bayesian model selection (BMS), along with comparisons to other widely used model selection methods.

Significantly, in cases where there are little data and/or highly dissimilar models, the Schwarz criterion (or Bayesian information criterion) can yield a poor approximation to the Bayes Factor. Methods for estimating Bayes factors using MCMC posterior samples were investigated. It was concluded that the Gelfand-Dey estimator provided the most reliable estimate of marginal likelihood.

Bayesian model averaging, the extension of Bayesian principles to model space, was discussed. This method was preferred over model selection as it removed the need to select particular models according to some (usually *ad hoc*) criteria. This model averaging technique was used for comparison of the HMM and its generalisations.

7.2.4 Spatial dependency and the HMM variants

Spatial dependency was incorporated into the HMM variants through two mechanisms: large-scale distance-independent correlation induced by a regional climate state and small-scale distance-dependent correlation preserved by the multivariate normal rainfall distributions. Several parameterisations of the small-scale Gaussian correlation were proposed for testing, namely Fitted Correlation, Exponential Decay Correlation, Empirical and Zero Correlation. The Fitted Correlation parameterisation was the most flexible because it treated the correlation coefficients as completely unknown and therefore required fitting individual site-to-site correlations. This parameterisation was chosen for use in the majority of the case studies.

7.2.5 Switch HMM and Regional HMM case studies

Chapter 5 reported on the application of the HMM, SHMM and RHMM to four groupings of four sites located on the Eastern coast of Australia. These groupings were denoted Close and Far (with the Close sites being within 200km of the key site), with either Sydney or Brisbane as the key site. The SHMM and RHMM both identified state series significantly different from the HMM series, with each providing different mechanisms for individual sites to vary from the HMM regional controlling state. In one case study (Sydney Far), a SHMM variant was selected as the most likely model (as opposed to the HMM or RHMM). In the remaining case studies (Brisbane Close and Far, Sydney Close) the RHMM variants showed greatest posterior weight, indicating that the data favoured the multiple region RHMM over the single region HMM or the SHMM variants. In no cases (except the zero small-scale correlation Brisbane Close

test) did the HMM produce the maximum marginal likelihood when compared to the SHMM and RHMM.

Given that the HMM generalisations produced variants with significantly greater posterior weight than the HMM, it is believed that the generalisations are justified. Of course, further evaluation of the models is suggested on more sites, in different areas. What has been demonstrated here is that the SHMM and RHMM provide a means of determining which sites should be included in a HMM analysis. It should be noted that the BMS model averaging technique removes the need to choose which sites are included, weighting the models according to their posterior weight.

Generally, the RHMM variants produced models with greater marginal likelihood than the SHMM variants, indicating the RHMM allowed more flexibility to reproduce the variability within the data. As such, the RHMM model averaged state series was used to condition the event based rainfall model DRIP.

7.2.6 Small-scale correlation structures: the influence of outliers

The Fitted Correlation case study involving the Brisbane Close group of sites produced some unusual results for the SHMM and RHMM calibrations. Even though Cape Moreton is situated closely to Caboolture and Brisbane, the model averaged state series of the RHMM indicated that Cape Moreton consistently identified a wet state series, generally being in the opposite state of Brisbane and Caboolture. The SHMM Cape Moreton state series contradicted this, following the Brisbane state series reasonably closely. Given that these sites are close to one another, we would expect the probability of being under the same climate influence to be high.

Given that there is a relationship between the number of sites grouped in each RHMM region (large-scale variation) and the correlation structure used (small-scale variation), several different ways of modelling the small-scale correlation were tested. This testing had a dual motivation: Firstly to identify a reason for the anomalous behaviour of the Cape Moreton state series; and secondly to produce insight to guide future work for dealing with small-scale correlation structure - specifically, to determine whether the Exponential Decay correlation structure is flexible enough to compete with the Fitted Correlation structure.

Relationships identified in Section 5.5 between the correlation and variance parameters demonstrated three major points. Firstly, fitting correlations rather than applying an empirically estimated value provides more flexibility in reproduction of observed data, allowing accommodation of outliers to a greater degree. Secondly, if a functional correlation structure (such as Exponential Decay) approximates the fitted correlation distribution well, it will be overwhelmingly favoured due to its parsimony. Finally, the non-Gaussian nature of the data can have a large effect on inference for a model which uses Gaussian distributions, with outliers potentially having a disproportionate effect on inferred parameters and state series. This final point is quite significant, with future work being required to address this disproportionate influence.

7.2.7 Conditioning DRIP on HMM state series

Chapter 6 explored conditioning a short timescale event-based stochastic rainfall model called DRIP on the HMM state series to preserve annual rainfall statistics – this represented the original motivation for this thesis. *Frost et al.* (2002) conditioned the DRIP model on the hidden state series of the original two-state HMM of *Thyer* (2001). This conditioning was motivated by the observation that the single-state DRIP model, typical of its genre, was underestimating annual variability because there was no explicit structure to incorporate long-term persistence. Although an improved reproduction in annual variability was observed using this conditioning, it was not clear which or how many sites to use in the HMM analysis. This motivated the introduction of the HMM generalisations coupled with Bayesian model averaging.

Chapter 6 reported on this original case study and the conditioning on the model averaged hidden state series of the RHMM. Both of these case studies were based on pluviograph data collected at Sydney and Brisbane.

The first case study (from *Frost et al.,* 2002) demonstrated an improved reproduction of the annual rainfall distribution. This improvement was attributed to the ability of the HMM to conceptually incorporate the intra- and inter-annual persistence apparent in Australian rainfall. The HMM was applied only to those processes within DRIP that influence the number of storm arrivals, namely storm and inter-storm duration. Nonetheless, the lack of a formal method for selecting sites in the HMM resulted in a

heuristic approach which required all sites to have similar transition probabilities but could not guarantee state series were coherent.

The second case study used a model averaged RHMM state series derived independently of the DRIP model. It was found that the annual distribution produced by the two-state DRIP model showed some sensitivity to the choice of state series with under or overestimation of annual variability resulting. These differences were attributed to possible misclassification of pluviograph data into inappropriate wet or dry years. In some cases the variability produced by the two-state DRIP model was greater than that observed in the annual rainfall records, suggesting that the hypothesis of two-state storm durations and inter-storm durations can be questioned.

It is noted that the conditional weighting of an event based model using the output series of a HMM (or its variants) can be used for any event based model which uses likelihood based inference.

7.3 Future Work

This study has made original and significant inroads to modelling the spatio-temporal stochastic structure of annual rainfall. However, there are significant opportunities for extending this work.

7.3.1 Model structure

In terms of overall model structure, other generalisations of the HMM are possible. A spatio-temporal HMM with each site's state dependent not only on the previous timestep state, but also the state of surrounding sites seems a worthwhile extension to pursue. This approach is essentially a hybrid of HMM for time series and a Markov random field on a lattice. Such a model may reduce the reliance on BMS in choosing which model variant is most appropriate.

Alternatively, related state space models such as the dynamic linear models (*Sanso and Guenni*, 2000) or the shifting mean models (*Fortin et al.*, 2002, *Sveinsson et al.*, 2003) could provide a flexible structure to capture the conceptual climate state variability. To the author's knowledge, altering the climate state across different sites has not been undertaken in these models, and as such the approach taken here for the RHMM, could be applied to these models, breaking the sites into different climate regions with a

climate state modelled in each region. Also, the setup of the RHMM regional partitioning defining the model could be altered such that partitioning is undertaken through a hierarchical indicator within the model itself. This however would require reformulation of the indicator setup for identification purposes.

Hierarchical models show great promise in modelling complex spatio-temporal processes (*Wikle et al.*, 2001). There is opportunity within HMM to further exploit hierarchical structure. For example, means and variances could be modelled as coming from a regional pool/distribution. Also, only annual rainfall has been used in the calibration of the models used in this study. Other related atmospheric indices and variables such as the Southern Oscillation Index or the Interdecadal Pacific Oscillation could be incorporated into the HMM structure. This would at least make the simulations/predictions consistent with such indices automatically, such as in the GCM downscaling approach exemplified in *Hughes et al.* (1999).

7.3.2 HMM variant structure

The Markovian state structure and site groupings identified in the case studies have shown some sensitivity to the small-scale correlation structure. Exponential correlation decay with a Gaussian nugget effect for microscale variation should be trialled in further studies. Of significant importance is the issue of outliers, with Gaussian distributions being affected strongly. Use of a transformation (most likely mild) that makes the transformed data more closely approximated by a Gaussian distribution may reduce the splitting effect observed for the Brisbane Close RHMM case studies.

7.3.3 MCMC sampling method

Regarding model selection, the possibility of using Reversible Jump MCMC should be investigated further. It is believed this model jumping technique provides a much more efficient means at estimating posterior model weight, as it removes the need to evaluate the marginal likelihood for every model. It is warned however, that the RJMCMC is quite complex, and it is likely that implementation would require a significant amount of work.

7.3.4 Missing data within the HMM variants

Methods of incorporating missing data have not been included in this study. Given that the models are currently limited by the length of concurrent data, the model is currently

throwing away much data at the ends of the rainfall series based on the shortest record. Much of this information on the state series could be used to aid overall identification. Such a method is vital in areas with short $(< 80\ years)$ annual rainfall records.

7.3.5 Joint DRIP-HMM calibration in Bayesian framework

It is recommended that to avoid misclassification of DRIP events a joint RHMM-DRIP calibration be undertaken, thus allowing the DRIP data to have an influence on the identification of the hidden state series. Because parameter uncertainty has not been incorporated in DRIP calibration so far, model selection using BMS has not been possible. Applying BMS to the joint RHMM-DRIP model would thus identify state series that are more consistent with the pluviograph data and enable formal model comparison of DRIP variants conditional on the resultant state series.

7.4 Final Remarks

Two new spatio-temporal HMM's have been introduced in this thesis, with the purpose of capturing the persistent, spatially non-homogeneous nature of climate influence on rainfall series observed in Australia. Both of these methods are quite different to other attempts at modelling these series, in that non-homogenous persistent conceptual climate effects are explicitly incorporated in the model design.

The conditioning of the event based rainfall model DRIP on the hidden state series of the HMM (or its variants) demonstrates the practical value of the HMM models. Modelling annual rainfall itself is not the major purpose. They have been used here to capture spatio-temporal variability and downscale hierarchically to a small timescale stochastic rainfall model.

References

2000. The Oxford English dictionary online. Oxford University Press, Oxford.

Aitkin, M., 1991. Posterior Bayes Factors. *Journal of the Royal Statistical Society Series B-Methodological 53*(1):111-142.

Aitkin, M., 1997. The calibration of P-values, posterior Bayes factors and the AIC from the posterior distribution of the likelihood. *Statistics and Computing 7*:253-261.

Akaike, H., 1973. Information Theory and an Extension of the Maximum Likelihood Principle. Pages 267 *in* B. N. Petrox and F. Caski, editors. Second International Symposium on Information Theory, Budapest: Akedemai Kiado.

Andrieu, C., and C. P. Robert, 2001. Controlled MCMC for Optimal Sampling. *Preprint*.

Bardossy, A., and E. J. Plate, 1992. Space-Time model for Daily Rainfall Using Atmospheric Circulation Patterns. *Water Resources Research 28*(5):1247-1259.

Bartlett, M. S., 1957. Comment on "A Statistical Paradox" by D. V. Lindley. *Biometrika 44*:533-534.

Bellone, E., 2000. Nonhomogeneous hidden Markov models for downscaling atmospheric patterns to precipitation amounts. PhD Thesis. University of Washington.

Bellone, E., J. P. Hughes, and P. Guttorp, 2000. A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Climate Research 15*(1):1-12.

Bengio, Y., 1999. Markovian Models for Sequential Data. *Neural Computing Surveys 2*:129-162.

Beran, J., 1994. Statistics for long-memory processes. Chapman & Hall, New York.

Berger, J. O., 1985. Statistical decision theory and Bayesian analysis, 2nd edition. Springer-Verlag, New York.

Berger, J. O., 2000. Bayesian Analysis: A Look at Today and Thoughts of Tomorrow. *Journal of the American Statistical Association 95*(452):1269-1276.

Berger, J. O., and J. M. Bernardo, 1992. On the development of reference priors *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, Bayesian Statistics 4. Oxford University Press, Oxford.

Berger, J. O., and L. R. Pericchi, 1996. The Intrinsic Bayes Factor for Model Selection and Prediction. *Journal of the American Statistical Association 91*(433):109-122.

Berger, J. O., and T. Sellke, 1987. Testing a point null hypothesis: the irreconcilability of *P* values and evidence. *Journal of the American Statistical Association 82*:112-122.

Berliner, L. M., 1996. Hierarchical Bayesian time series models *in* K. Hanson and R. Silver, editors, Maximum Entropy and Bayesian Methods:15-22. Kluwer Academic Press.

Bernardo, J. M., 1979. Reference posterior distributions for Bayesian inference (with discussion). *Journal of the Royal Statistical Society Series B-Methodological 36*:192-236.

Bernardo, J. M., and R. Rueda, 2002. Bayesian hypothesis testing: a reference approach. *International Statistical Review 70*(3):351-350.

Bernardo, J. M., and A. F. M. Smith, 2000. Bayesian Theory, 2nd edition. John Wiley & Sons, Ltd., Chichester.

Besag, J., 1997. On Bayesian Analysis of Mixtures with an Unknown Number of Components - Discussion. *Journal of the Royal Statistical Society Series B-Methodological 59*(4):758-792.

Box, G. E. P., and G. M. Jenkins, 1976. Time series analysis : forecasting and control, Rev. edition. Holden-Day, San Francisco.

Bras, R. L., and I. Rodriguez-Iturbe, 1985. Random Functions and Hydrology. Addison-Wesley Publishing Company.

Campbell, E. P., D. R. Fox, and B. C. Bates, 1999. A Bayesian approach to parameter estimation and pooling in nonlinear flood event models. *Water Resources Research 35*(1):211-220.

Carlin, B. P., and S. Chib, 1995. Bayesian Model Choice Via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society Series B-Methodological 57*(3):473-484.

Carlin, B. P., and T. A. Louis, 2000. Bayes and empirical Bayes methods for data analysis, 2nd edition. Chapman & Hall.

Celeux, G., M. Hurn, and C. P. Robert, 2000. Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association 95*(451):957-970.

Charles, S. P., B. C. Bates, and J. P. Hughes, 1999. A spatiotemporal model for downscaling precipitation occurrence and amounts. *Journal of Geophysical Research-Atmospheres 104*(D24):31657-31669.

Chib, S., 1995. Marginal Likelihood From the Gibbs Output. *Journal of the American Statistical Association 90*(432):1313-1321.

Chib, S., 1996. Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models. *Journal of Econometrics 75*(1):79-97.

Chib, S., and E. Greenberg, 1995. Understanding the Metropolis-Hastings Algorithm. *American Statistician 49*(4):327-335.

Chib, S., and I. Jeliazkov, 2001. Marginal Likelihood From the Metropolis-Hastings Output. *Journal of the American Statistical Association 96*(453):270-281.

Chiew, F. H. A., and T. A. McMahon, 2003. El Nino/ Southern Oscillation and Australian rainfall and streamflow. *Australian Journal of Water Resources 6*(2):115-129.

Chiew, F. H. A., T. C. Piechota, J. A. Dracup, and T. A. McMahon, 1998. El Nino Southern Oscillation and Australian Rainfall, Streamflow and Drought - Links and Potential for Forecasting. *Journal of Hydrology 204*(1-4):138-149.

Cornish, P. M., 1977. Changes in seasonal and annual rainfall in New South Wales. *Search 8*(1-2):38-40.

Cowpertwait, P. S. P., and P. E. O'Connell, 1997. A regionalised Neyman-Scott model of rainfall with convective and stratiform cells. *Hydrology and Earth Sciences 1*:71-80.

Cressie, N. A. C., 1991. Statistics for spatial data. Wiley, New York.

Dellaportas, P., J. J. Forster, and I. Ntzoufras, 2002. On Bayesian Model and Variable selection Using MCMC. *Statistics and Computing 12*:27-36.

DiCiccio, T. J., R. E. Kass, A. E. Raftery, and L. Wasserman, 1997. Computing Bayes Factors By Combining Simulation and Asymptotic Approximations. *Journal of the American Statistical Association 92*(439):903-915.

Eagleson, P. S., 1978. Climate, Soil, and Vegetation 2. The Distribution of Annual Precipitation Derived From Observed Storm Sequences. *Water Resources Research 14*(5):713-721.

Folland, C. K., J. A. Renwick, M. J. Salinger, and A. B. Mullan, 2002. Relative influences of the Interdecadal Pacific Oscillation and ENSO on the South Pacific Convergence Zone - art. no. 1643. *Geophysical Research Letters 29*(13):1643.

Fortin, V., L. Perreault, J.-C. Ondo, and R.-C. Evra, 2002. Bayesian Long-term forecasting of Annual Flows with a Shifting-level Model. *in* 2002 Conference on Water Resources Planning and Management. Environmental and Water Research Institute, Roanoke, VA.

Fortin, V., L. Perreault, and J. D. Salas, 2003. Retrospective Analysis and Forecasting of Streamflows Using a Shifting Level Model. *Journal of Geophysical Research Submitted*.

Franks, S. W., and G. Kuczera, 2002. Flood frequency analysis: Evidence and implications of secular climate variability, New South Wales. *Water Resources Research 38*(5):1062.

Frost, A. J., S. Jennings, M. Thyer, M. Lambert, and G. Kuczera, 2002. Incorporating long-term climate variability into a short-timescale rainfall model using a hidden state Markov model. *Australian Journal of Water Resources 6*(1):63-70.

Fruhwirth-Schnatter, S., 2001. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association 96*(453):194-209.

Fruhwirth-Schnatter, S., 2002. Model Likelihoods and Bayes Factors for Switching and Mixture Models. Technical Report Vienna University of Economics and Business Administration, Vienna.

Gelfand, A. E., and D. K. Dey, 1994. Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society Series B-Methodological 56*(3):501-514.

Gelman, A., and J. B. Carlin, 1992. Inference from iterative simulation using multiple sequences. *Statistical Science 7*:457-472.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 1995. Bayesian Data Analysis. Chapman and Hall, London.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2004. Bayesian Data Analysis, 2nd edition. Chapman & Hall/CRC, New York.

Geman, S., and D. Geman, 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*:721-741.

Godsill, S. J., 1998. On the relationship between MCMC model uncertainty methods. Technical CUED/F-INFENG/TR. 305, Cambridge University Engineering Department, Cambridge.

Grayson, R. B., R. M. Argent, R. J. Nathan, T. A. McMahon, and R. G. Mein, 1996. Hydrological Recipes: Estimation Techniques in Australian Hydrology. Cooperative Research Centre for Catchment Hydrology.

Green, P. J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*(4):711-732.

Haario, H., E. Saksman, and J. Tamminen, 1999. Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics 14*:375-395.

Haario, H., E. Saksman, and J. Tamminen, 2001. An Adaptive Metropolis Algorithm. *Bernoulli 7*(2):223-242.

Han, C., and B. P. Carlin, 2000. MCMC Methods for Computing Bayes Factors: A Comparative Review. Research report 2000-001, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis.

Handcock, M. S., and J. R. Wallis, 1994. An approach to statistical spatial-temporal modeling of Meteorological fields. *Journal of the American Statistical Association 89*:368-390.

Hastings, W. K., 1970. Monte-Carlo sampling methods using Markov Chains and their applications. *Biometrika 57*:97-109.

Hay, L. E., G. J. McCabe, D. M. Wolock, and M. A. Ayers, 1991. Simulation of Precipitation by Weather Type Analysis. *Water Resources Research 27*(4):493-501.

Heneker, T. M., M. F. Lambert, and G. Kuczera, 2001. A point rainfall model for risk-based design. *Journal of Hydrology 247*(1-2):54-71.

Hughes, J. P., and P. Guttorp, 1994. A Class of Stochastic Models for Relating Synoptic Atmospheric Patterns to Regional Hydrologic Phenomena. *Water Resources Research 30*(5):1535-1546.

Hughes, J. P., P. Guttorp, and S. P. Charles, 1999. A non-homogeneous hidden Markov model for precipitation occurrence. *Journal of the Royal Statistical Society Series C-Applied 48*(1):15-30.

Hurst, H. E., 1951. Long Term Storage Capacities of Reservoirs. *Transactions American Society of Civil Engineers 116*:770-799.

Isdale, P. J., B. J. Stewart, K. S. Tickle, and J. M. Lough, 1998. Paleohydrological Variation in a Tropical River Catchment - a Reconstruction Using Fluorescent Bands in Corals of the Great Barrier Reef, Australia. *Holocene 8*(1):1-8.

Jeffreys, H., 1961. Theory of probability, 3rd edition. Clarendon Press, Oxford.

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. *Journal for Basic Engineering (AMSE) 82D*:35-45.

Kane, R. P., 1997. On the Relationship of ENSO with Rainfall over Different Parts of Australia. *Australian Meteorological Magazine 46*(1):39-49.

Kass, R. E., and A. E. Raftery, 1995. Bayes Factors. *Journal of the American Statistical Association 90*(430):773-795.

Kass, R. E., and L. Wasserman, 1995. A Reference Bayesian Test for Nested Hypothesis and Its Relationship to the Schwarz Criterion. *Journal of the American Statistical Association 90*(431):928-934.

Katz, R. W., 1981. On Some Criteria for Estimating the Order of a Markov Chain. *Technometrics 23*(3):243-249.

Katz, R. W., and M. B. Parlange, 1993. Effects of an Index of Atmospheric Circulation on Stochastic Properties of Precipitation. *Water Resources Research 29*(7):2335-2344.

Katz, R. W., and X. G. Zheng, 1999. Mixture model for overdispersion of precipitation. *Journal of Climate 12*:2528-2537.

Kiem, A. S., and S. W. Franks, 2001a. Determining the importance of rapidly changing ENSO indices when forecasting ENSO events. Research Report 219.12.2001, University of Newcastle, Newcastle.

Kiem, A. S., and S. W. Franks, 2001b. On the identification of ENSO-induced rainfall and runoff variability: a comparison of methods and indices. *Hydrological Sciences Journal-Journal des Sciences Hydrologiques 46*(5):715-727.

Kiem, A. S., S. W. Franks, and G. Kuczera, 2003. Multi-decadal variability of flood risk. *Geophysical Research Letters 30*(2).

Koutsoyiannis, D., and E. Foufoula-Georgiou, 1993. A Scaling Model of a Storm Hyetograph. *Water Resources Research 29*(7):2345-2361.

Kuczera, G., and E. Parent, 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology 211*(1-4):69-85.

Lambert, M., and G. Kuczera, 1998. Seasonal Generalized Exponential Probability Models with Application to Interstorm and Storm Durations. *Water Resources Research 34*(1):143-148.

Lavery, B., G. Joung, and N. Nicholls, 1997. An Extended High-Quality Historical Rainfall Dataset for Australia. *Australian Meteorological Magazine 46*(1):27-38.

Lee, P. M., 1989. Bayesian statistics : an introduction. Oxford University Press, New York.

Lindley, D. V., 1957. A Statistical Paradox. *Biometrika 44*:187-192.

Lough, J. M., 1993. Variations of Some Seasonal Rainfall Characteristics in Queensland, Australia - 1921-1987. *International Journal of Climatology 13*(4):391-409.

Lough, J. M., 1997. Regional Indices of Climate Variation - Temperature and Rainfall in Queensland, Australia. *International Journal of Climatology 17*(1):55-66.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997. A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production. *Bulletin of the American Meteorological Society 78*(6):1069-1079.

Marden, J. I., 2000. Hypothesis Testing: From *p* Values to Bayes Factors. *Journal of the American Statistical Association 95*(452):1316-1320.

Matheron, G., 1962. Traite de Geostatistique Appliquee, Tome I. Memoires du Bureau de Recherches Geologiques et Minieres, No. 14 Editions Technip, Paris.

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

Mengersen, K., C. P. Robert, and C. Guihenneuc-Jouyaux, 1998. MCMC Convergence Diagnostics: A "Reviewww" *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, Bayesian Statistics 5, Oxford.

Mengersen, K. L., and R. L. Tweedie, 1996. Rates of Convergence of the Hastings and Metropolis Algorithms. *Annals of Statistics 24*(1):101-121.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics 21*:1087-1091.

Morris, C. N., 1983. Parametric empirical Bayes inference: Theory and implications. *Journal of the American Statistical Association 78*:47-65.

Newton, M. A., and A. E. Raftery, 1994. Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society Series B-Methodological 56*(1):3-48.

Nicholls, N., B. Lavery, C. Frederiksen, W. Drosdowsky, and S. Torok, 1996. Recent Apparent Changes in Relationships between the El Nino - Southern Oscillation and Australian Rainfall and Temperature. *Geophysical Research Letters 23*(23):3357-3360.

Ntzoufras, I., 1999. Aspects of Bayesian Model and Variable Selection Using MCMC. PhD Thesis. Athens University of Economics and Business, Athens.

O'Hagan, A., 1995. Fractional Bayes Factors for Model Comparison. *Journal of the Royal Statistical Society Series B-Methodological 57*(1):99-138.

Perreault, L., J. Bernier, B. Bobee, and E. Parent, 2000a. Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. *Journal of Hydrology 235*:242-263.

Perreault, L., E. Parent, J. Bernier, B. Bobee, and M. Slivitzky, 2000b. Retrospective multivariate Bayesian change-point analysis: A simultaneous single change in the mean of several hydrological sequences. *Stochastic Environmental Research and risk Assessment 14*:243-261.

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

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*(5):319-324.

Richardson, S., and P. J. Green, 1997. On Bayesian Analysis of Mixtures with an Unknown Number of Components. *Journal of the Royal Statistical Society Series B-Methodological 59*(4):731-792.

Robert, C. P., 1996. Mixtures of distributions: inference and estimation *in* W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, Markov Chain monte Carlo in Practice:441-464. Chapman & Hall, London.

Roberts, G. O., and R. L. Tweedie, 1996. Geometric Convergence and Central Limit Theorems for Multidimensional Hastings and Metropolis Algorithms. *Biometrika 83*(1):95-110.

Ropelewski, C. F., and M. S. Halpert, 1996. Quantifying Southern Oscillation - Precipitation Relationships. *Journal of Climate 9*(5):1043-1059.

Salas, J. D., 1993. Analysis and Modeling of Hydrologic Time Series *in* D. Maidment, editor. Handbook of Hydrology. McGraw-Hill, New York.

Sanso, B., and L. Guenni, 2000. A Nonstationary Multisite Model for Rainfall. *Journal of the American Statistical Association 95*(452):1089-1100.

Schervish, M. J., 1995. Theory of statistics. Springer-Verlag, New York.

Schwarz, G., 1978. Estimating the Dimension of a Model. *The Annals of Statistics 6*:461-464.

Spiegelhalter, D. J., N. G. Best, B. R. Carlin, and A. van der Linde, 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Methodological 64*(4):583-616.

Srikanthan, R., G. Kuczera, M. Thyer, and T. A. McMahon, 2002. Generation of annual rainfall data for Australian stations. *in* 27th Hydrology and Water Resources Symposium. IEAust, Melbourne.

Srikanthan, R., and T. A. McMahon, 1985. Stochastic generation of rainfall and evaporation data. Technical Report 84, Australian Water Resources Council, Canberra.

Srikanthan, R., and T. A. McMahon, 2001. Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences 5*(4):653-670.

Srikanthan, R., T. A. McMahon, G. G. S. Pegram, M. Thyer, and G. Kuczera, 2001. Generation of Annual Rainfall for Australian Stations. Technical Report Cooperative Research Centre for Catchment Hydrology, Melbourne.

Stanton, D., and D. White, 1986. Constructive combinatorics. Springer-Verlag, New York.

Stephens, M., 1997. Bayesian Methods for Mixtures of Normal Distributions. PhD. University of Oxford, Oxford.

Stephens, M., 2000. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society Series B-Methodological 62*(4):795-809.

Sveinsson, O. G. B., J. D. Salas, D. C. Boes, and R. A. Pielke Sr., 2003. Modeling the Dynamics of Long term Variability of Hydroclimatic Processes. *Journal of Hydrometeorology 4*(2):489-505.

Taylor, K., 1999. Rapid climate change. *American Scientist 87*(4):320-327.

Thyer, M., 2001. Modelling Long-Term Persistence in Hydrological Time Series. PhD Thesis. University of Newcastle, Newcastle.

Thyer, M., and G. Kuczera, 2000. Modeling long-term persistence in hydroclimatic time series using a hidden state Markov model. *Water Resources Research 36*(11):3301-3310.

Thyer, M., and G. Kuczera, 2003a. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach. *Journal of Hydrology 275*:12-26.

Thyer, M., and G. Kuczera, 2003b. A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 2. Real data analysis. *Journal of Hydrology 275*:27-48.

Trenberth, K. E., and T. J. Hoar, 1996. El Niño-Southern Oscillation Event: Longest on record. *Geophysical Research Letters 23*:57-60.

Wasserman, L., 2000. Bayesian Model Selection and Model Averaging. *Journal of Mathematical Psychology 44*:92-107.

Waymire, E., V. K. Gupta, and I. Rodriguez-Iturbe, 1984. A Spectral theory of Rainfall Intensity at the Meso-β Scale. *Water Resources Research 20*(10):1453-1465.

Wikle, C. K., 2003. Hierarchical Models in Environmental Science. *International Statistical Review 71*(2):181-200.

Wikle, C. K., and N. Cressie, 1999. A Dimension-Reduction Approach to Space-Time Kalman Filtering. *Biometrika 86*:815-829.

Wikle, C. K., R. F. Milliff, D. Nychka, and L. M. Berliner, 2001. Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association 96*(454):382-397.

Wilks, D. S., and R. L. Wilby, 1999. The weather generation game: a review of stochastic weather models. *Progress in Physical Geography 23*(3):329-357.

Woolhiser, D. A., and H. B. Osborn, 1985. A Stochastic Model of Dimensionless Thunderstorm Rainfall. *Water Resources Research 21*(4):511-522.

Yevjevich, V., 1963. Fluctuations of Wet and Dry Years, Part 1, Research Data Assembly and Mathematical Models. Hydrology Paper 1, Colorado State University, Fort Collins.

Zhang, P., 1993. On the convergence rate of model selection criteria. *Communications in Statistics 22*:2765-2775.

Zheng, X. G., 1996. Unbiased Estimation of Autocorrelations of Daily Meteorological Variables. *Journal of Climate 9*(9):2197-2203.

Appendices

Appendix A– Marginal Likelihood Normalising Constants

A.1 Introduction

Calculation of the marginal likelihood under parameter constraints was discussed in Section 4.5.2. If marginal likelihoods are to be compared, the prior distribution of the parameter that the constraint applies to must be normalised according to (4.13).

Of the models considered in this study, calculation of these normalising constants were required for several parameters. Bounding on the mean parameters $\mu_i^{site} : site = 1,...,d, i = D,W$ between upper ($ubnd$) and lower ($lbnd$) bounds was performed. Also, in calculating the Fitted Correlation marginal likelihoods, the uniform priors on the correlation parameters require normalisation for combinations which do not result in positive definite covariance matrices.

A.2 Bounding constraint on wet and dry means

Reiterating Section 4.5.3, the priors on the individual mean and variance parameters are specified by:

$$\begin{aligned} \mu_i^{site} \mid \sigma_i^{site^2} &\sim N\left(\mu_0, {\sigma_y}^2 / \kappa\right) \\ \sigma_i^{site^2} &\sim Inv-\chi^2\left(\nu_0, \sigma_0^2\right) \end{aligned} \qquad \text{(A.1)},$$

with the joint probability of the state means and variance is defined as:

$$\begin{aligned} p\left(\mu_i^{site}, \sigma_i^{site^2}\right) &= p\left(\mu_i^{site} \mid \sigma_i^{site}, lbnd < \mu_i^{site} < ubnd\right) p\left(\sigma_i^{site^2}\right) \\ &= \frac{f_N\left(\mu_i^{site}; \mu_0, \left(\sigma_i^{site}\right)^2 / \kappa\right) I\left[lbnd < \mu_i^{site} < ubnd\right]}{P\left(lbnd < \mu_i^{site} < ubnd\right)} f_{Inv-\chi^2}\left(\sigma_i^{site^2}; \nu_0, \sigma_0^2\right) \\ &= \frac{f_N\left(\mu_i^{site}; \mu_0, \left(\sigma_i^{site}\right)^2 / \kappa\right) I\left[lbnd < \mu_i^{site} < ubnd\right] f_{Inv-\chi^2}\left(\sigma_i^{site^2}; \nu_0, \sigma_0^2\right)}{\int_{\mu_i^{site}=lbnd}^{ubnd} \int_{\sigma_i^{site^2}=0}^{\infty} f_N\left(\mu_i^{site}; \mu_0, \left(\sigma_i^{site}\right)^2 / \kappa\right) f_{Inv-\chi^2}\left(\sigma_i^{site^2}; \nu_0, \sigma_0^2\right) d\sigma_i^{site^2} d\mu_i^{site}} \end{aligned} \qquad \text{(A.2)},$$

where f_N and $f_{Inv-\chi^2}$ are the distribution functions for the Normal and Inverse-χ^2 distributions respectively. These functions are:

$$p(\mu|\mu_0,\sigma^2/\kappa) = f_N(\mu;\mu_0,\sigma^2/\kappa) = \frac{\sqrt{\kappa}}{\sqrt{2\pi}\sigma} e^{\left(-\frac{\kappa}{2\sigma^2}(\mu-\mu_0)^2\right)} \quad \text{(A.3)},$$

for the Normal, and:

$$p(\sigma^2 \mid \nu_0,\sigma_0) = f_{Inv-\chi^2}(\sigma^2;\nu_0,\sigma_0) = \frac{(\nu_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)}(\sigma_0)^{\nu_0}\,\sigma^{-(\nu_0/2+1)}e^{-\nu_0\sigma_0^2/(2\sigma^2)} \quad \text{(A.4)}.$$

For $f_{Inv-\chi^2}$. For further details of these distributions see *Gelman et al.* (1995, Appendix A).

Thankfully the denominator term in (A.2) can be calculated analytically for the hyperparameters used in this study (see Table 4.2). Substituting $\mu_0 = \overline{y}_{site}$, $\kappa = 1$, $\sigma_0^2 = \overline{s}_{site}^{\,2}$, $\nu_0 = 2$ and $lbnd = 0$, the denominator becomes:

$$\begin{aligned}
&\int_{\mu_i^{site}=lbnd}^{ubnd}\int_{\sigma_i^{site^2}=0}^{\infty} f_N\left(\mu_i^{site};\sigma_i^{site},\mu_0=\overline{y}_{site},\kappa=1\right) f_{Inv-\chi^2}\left(\sigma_i^{site^2};\nu_0=2,\sigma_0^2=\overline{s}_{site}^2\right) d\sigma_i^{site^2} d\mu_i^{site} \\
&= \int_{\mu_i^{site}=0}^{ubnd} \frac{\overline{s}_{site}^2}{\left(\overline{y}_{site}^2 - 2\overline{y}_{site}\mu_i^{site} + \left(\mu_i^{site}\right)^2 + 2\overline{s}_{site}^2\right)^{3/2}} d\mu_i^{site} \\
&= \frac{\overline{y}_{site}\sqrt{\left(\overline{y}_{site}^2 - 2\overline{y}_{site}ubnd + 2\overline{s}_{site}^2 + (ubnd)^2\right)} + \sqrt{\left(\overline{y}_{site}^{\,2} + 2\overline{s}_{site}^2\right)(ubnd - \overline{y}_{site})}}{2\sqrt{\left(\overline{y}_{site}^2 + 2\overline{s}_{site}^2\right)}\sqrt{\left(\overline{y}_{site}^2 - 2\overline{y}_{site}ubnd + (ubnd)^2 + 2\overline{s}_{site}^2\right)}}
\end{aligned} \quad \text{(A.5)}.$$

A.3 Positive definite constraint on correlation matrix

The priors on the individual correlation coefficient parameters were uniform and bounded between $[0,0.95]$ for all site pairs:

$$\rho_{ij} \sim Uniform[0,0.95] : i, j = 1,...,d \quad \text{(A.6)}.$$

The joint probability of all correlation coefficients (over the unconstrained parameter space Θ) is therefore:

$$\begin{aligned} p(\boldsymbol{\rho} \mid \boldsymbol{\rho} \in \Theta) &= \prod_{i=1,j=1}^{d} p(\rho_{ij}) \ : i \neq j \\ &= \prod_{i=1,j=1}^{d} \left(\frac{1}{0.95-0} \right) : i \neq j \\ &= \left(\frac{1}{0.95-0} \right)^{d(d-1)/2} \end{aligned} \quad \text{(A.7).}$$

However, this prior must be normalised for combinations of correlation coefficients which do not result in positive definite matrices. Rather than analytically calculating the normalising constant, numerical integration was used. The uniform prior for all sites was sampled from, producing multiple sets of correlation coefficient matrix. The normalising constant is determined by the proportion of these correlation matrix which are positive definite. That let Θ' be the constrained (positive definite) parameter space, the normalising constant for the prior (A.7) is:

$$\begin{aligned} P(\boldsymbol{\rho} \in \Theta') &= \int_{\Theta'} p(\boldsymbol{\rho} \mid \boldsymbol{\rho} \in \Theta) d\boldsymbol{\rho} \\ &= \frac{1}{ns} \sum_{i=1}^{ns} I\left[\boldsymbol{\rho}^{(i)} \in \Theta' \right] : \ \boldsymbol{\rho}^{(i)} \leftarrow p(\boldsymbol{\rho} \mid \boldsymbol{\rho} \in \Theta) \end{aligned} \quad \text{(A.8).}$$

This normalising constant was calculated for correlation coefficient matrix of dimension four as this is the number of sites used in case studies. 300,000,000 samples were used to ensure that the resulting acceptance ratio was not in error. The acceptance ratio, along with its log-space equivalent is presented in Table A.1.

Table A.1 Full model marginal likelihoods and model weights

Acceptance ratio	0.55478
Log-space	-0.58918

Appendix B– Case Study Marginal Likelihoods and Posterior Weights

B.1 Introduction

Within Chapter 5, several case studies were presented comparing the newly introduced HMM variants on four site groupings; Sydney Close/Far and Brisbane Close/Far. This analysis was based around the marginal likelihood, an integral factor in calculating model weight. The model marginal likelihoods are presented here, along with the resulting model weights (assuming uniform model priors).

The results are divided into four sections. Section B.2 presents the full HMM, SHMM and RHMM marginal likelihoods (see Section 5.3 for details). Section B.3 presents the SHMM and RHMM variant marginal likelihoods, for both the Fitted Correlation (Section 5.4) and Exponential Decay Correlation parameterisations (Section 5.5). The marginal likelihood for the RHMM with a nugget effect term for two site groupings (Sydney and Brisbane Close – see Section 5.5.4) is given in Section B.4. The 7 site Fitted Correlation with nugget effect RHMM calibrations (mentioned in Section 5.6.1) are listed within B.5.

B.2 Fitted Correlation Model Marginal Likelihoods

Table B.2 Full model marginal likelihoods and model weights

Site Grouping		HMM	SHMM	RHMM
Sydney Close	$\log\left(p(\mathbf{y}\mid M)\right)$	-2971.65	-2972.27	-2972.75
	Posterior Weight	0.5333	0.2884	0.1783
Sydney Far	$\log\left(p(\mathbf{y}\mid M)\right)$	-3072.51	-3071.44	-3072.65
	Posterior Weight	0.2085	0.6090	0.1825
Brisbane Close	$\log\left(p(\mathbf{y}\mid M)\right)$	-2887.86	-2885.28	-2877.45
	Posterior Weight	0.0000	0.0004	0.9996
Brisbane Far	$\log\left(p(\mathbf{y}\mid M)\right)$	-2370.04	-2369.53	-2367.30
	Posterior Weight	0.0553	0.0920	0.8526

B.3 HMM Variant Marginal Likelihoods

Table B.3 Sydney Close RHMM variants posterior model probabilities

RHMM Model Label	Fitted Correlation		Exponential Decay	
	Posterior Weight	Marginal Likelihood	Posterior Weight	Marginal Likelihood
1,1,1,1	0.016	-2971.65	0.001	-2972.75
1,1,1,2	0.000	-2975.85	0.037	-2969.32
1,1,2,1	0.728	-2967.84	0.081	-2968.54
1,1,2,2	0.007	-2972.43	0.000	-2977.74
1,1,2,3	0.165	-2969.32	0.015	-2970.26
1,2,1,1	0.000	-2978.63	0.000	-2973.69
1,2,2,1	0.034	-2970.89	0.460	-2966.81
1,2,2,2	0.000	-2976.57	0.000	-2979.49
1,2,2,3	0.003	-2973.49	0.196	-2967.66
1,2,3,1	0.039	-2970.75	0.172	-2967.79
1,2,3,2	0.001	-2974.35	0.001	-2973.33
1,2,3,3	0.000	-2976.43	0.000	-2977.85
1,2,3,4	0.005	-2972.75	0.035	-2969.38

Table B.4 Sydney Close SHMM variants posterior model probabilities

SHMM Model Label	Fitted Correlation		Exponential Decay	
	Posterior Weight	Marginal Likelihood	Posterior Weight	Marginal Likelihood
F,F,F,F	0.022	-2971.25	0.001	-2972.21
T,F,F,F	0.001	-2974.56	0.000	-2976.01
F,T,F,F	0.003	-2973.24	0.000	-2973.33
T,T,F,F	0.000	-2975.61	0.000	-2975.16
F,F,T,F	0.255	-2968.82	0.014	-2969.03
T,F,T,F	0.026	-2971.09	0.002	-2971.04
F,T,T,F	0.174	-2969.20	0.038	-2968.06
T,T,T,F	0.019	-2971.43	0.007	-2969.81
F,F,F,T	0.014	-2971.75	0.015	-2968.95
T,F,F,T	0.001	-2974.56	0.016	-2968.91
F,T,F,T	0.003	-2973.20	0.005	-2970.02
T,T,F,T	0.001	-2974.29	0.105	-2967.04
F,F,T,T	0.296	-2968.66	0.152	-2966.67
T,F,T,T	0.026	-2971.10	0.011	-2969.28
F,T,T,T	0.151	-2969.34	0.600	-2965.30
T,T,T,T	0.008	-2972.27	0.034	-2968.17

Table B.5 Sydney Far RHMM variants posterior model probabilities

RHMM Model Label	Fitted Correlation		Exponential Decay	
	Posterior Weight	Marginal Likelihood	Posterior Weight	Marginal Likelihood
1,1,1,1	0.043	-3072.51	0.000	-3077.32
1,1,1,2	0.003	-3075.16	0.003	-3074.61
1,1,2,1	0.001	-3076.36	0.005	-3074.10
1,1,2,2	0.079	-3071.90	0.247	-3070.20
1,1,2,3	0.022	-3073.18	0.712	-3069.15
1,2,1,1	0.005	-3074.65	0.000	-3079.14
1,2,1,2	0.002	-3075.58	0.000	-3077.52
1,2,1,3	0.006	-3074.56	0.002	-3075.06
1,2,2,2	0.541	-3069.98	0.000	-3079.15
1,2,2,3	0.020	-3073.30	0.000	-3076.80
1,2,3,2	0.006	-3074.44	0.002	-3074.95
1,2,3,3	0.234	-3070.82	0.001	-3076.24
1,2,3,4	0.038	-3072.65	0.027	-3072.42

Table B.6 Sydney Far SHMM variants posterior model probabilities

SHMM Model Label	Fitted Correlation		Exponential Decay	
	Posterior Weight	Marginal Likelihood	Posterior Weight	Marginal Likelihood
F,F,F,F	0.008	-3072.04	0.000	-3077.37
T,F,F,F	0.364	-3068.22	0.000	-3077.18
F,T,F,F	0.007	-3072.20	0.000	-3077.30
T,T,F,F	0.461	-3067.98	0.000	-3076.47
F,F,T,F	0.001	-3074.14	0.005	-3073.12
T,F,T,F	0.040	-3070.42	0.081	-3070.26
F,T,T,F	0.002	-3073.48	0.017	-3071.83
T,T,T,F	0.027	-3070.80	0.253	-3069.11
F,F,F,T	0.001	-3074.01	0.002	-3073.92
T,F,F,T	0.023	-3070.98	0.001	-3075.19
F,T,F,T	0.004	-3072.68	0.001	-3075.02
T,T,F,T	0.019	-3071.17	0.017	-3071.79
F,F,T,T	0.004	-3072.82	0.150	-3069.63
T,F,T,T	0.013	-3071.55	0.124	-3069.83
F,T,T,T	0.012	-3071.64	0.164	-3069.54
T,T,T,T	0.015	-3071.44	0.185	-3069.42

Table B.7 Brisbane Close RHMM variants posterior model probabilities

RHMM Model Label	Fitted Correlation		Exponential Decay	
	Posterior Weight	Marginal Likelihood	Posterior Weight	Marginal Likelihood
1,1,1,1	0.000	-2887.86	0.000	-2882.02
1,1,1,2	0.000	-2886.45	0.000	-2882.06
1,1,2,1	0.000	-2891.03	0.000	-2895.67
1,1,2,2	0.000	-2888.30	0.000	-2893.80
1,1,2,3	0.000	-2885.89	0.000	-2892.26
1,2,1,1	0.001	-2884.50	0.043	-2872.62
1,2,1,3	0.012	-2881.71	0.932	-2869.54
1,2,2,1	0.001	-2884.49	0.000	-2884.15
1,2,2,2	0.003	-2883.17	0.000	-2883.67
1,2,2,3	0.038	-2880.56	0.000	-2880.09
1,2,3,1	0.024	-2881.02	0.000	-2877.94
1,2,3,3	0.057	-2880.17	0.001	-2877.06
1,2,3,4	0.864	-2877.45	0.025	-2873.16

Table B.8 Brisbane Close SHMM variants posterior model probabilities

SHMM Model Label	Fitted Correlation		Exponential Decay	
	Posterior Weight	Marginal Likelihood	Posterior Weight	Marginal Likelihood
F,F,F,F	0.001	-2887.25	0.003	-2881.45
T,F,F,F	0.095	-2882.27	0.001	-2883.20
F,T,F,F	0.000	-2888.25	0.076	-2878.28
T,T,F,F	0.030	-2883.43	0.049	-2878.72
F,F,T,F	0.000	-2891.61	0.000	-2885.16
T,F,T,F	0.005	-2885.13	0.000	-2887.75
F,T,T,F	0.000	-2890.21	0.010	-2880.36
T,T,T,F	0.001	-2886.87	0.001	-2882.41
F,F,F,T	0.002	-2885.99	0.007	-2880.65
T,F,F,T	0.699	-2880.27	0.002	-2882.19
F,T,F,T	0.001	-2887.07	0.337	-2876.80
T,T,F,T	0.129	-2881.96	0.383	-2876.67
F,F,T,T	0.000	-2888.80	0.000	-2883.90
T,F,T,T	0.032	-2883.36	0.030	-2879.22
F,T,T,T	0.000	-2888.79	0.076	-2878.28
T,T,T,T	0.005	-2885.28	0.026	-2879.37

Table B.9 Brisbane Far RHMM variants posterior model probabilities

RHMM Model Label	Fitted Correlation		Exponential Decay	
	Posterior Weight	Marginal Likelihood	Posterior Weight	Marginal Likelihood
1,1,1,1	0.043	-2370.04	0.002	-2371.27
1,1,1,2	0.003	-2370.16	0.000	-2373.40
1,1,2,1	0.001	-2369.16	0.001	-2372.09
1,1,2,2	0.079	-2368.18	0.002	-2371.29
1,1,2,3	0.022	-2368.87	0.001	-2372.19
1,2,1,1	0.005	-2366.30	0.580	-2365.46
1,2,1,2	0.002	-2368.67	0.002	-2371.03
1,2,1,3	0.006	-2367.28	0.023	-2368.71
1,2,2,2	0.541	-2369.10	0.005	-2370.13
1,2,2,3	0.020	-2367.90	0.031	-2368.38
1,2,3,2	0.006	-2368.00	0.013	-2369.27
1,2,3,3	0.234	-2365.99	0.201	-2366.52
1,2,3,4	0.038	-2367.30	0.140	-2366.88

Table B.11 Brisbane Far SHMM variants posterior model probabilities

SHMM Model Label	Fitted Correlation		Exponential Decay	
	Posterior Weight	Marginal Likelihood	Posterior Weight	Marginal Likelihood
F,F,F,F	0.006	-2370.82	0.009	-2370.93
T,F,F,F	0.079	-2368.32	0.035	-2369.61
F,T,F,F	0.048	-2368.82	0.067	-2368.94
T,T,F,F	0.070	-2368.45	0.112	-2368.43
F,F,T,F	0.053	-2368.73	0.009	-2370.90
T,F,T,F	0.202	-2367.38	0.045	-2369.35
F,T,T,F	0.051	-2368.76	0.146	-2368.17
T,T,T,F	0.028	-2369.38	0.137	-2368.23
F,F,F,T	0.034	-2369.18	0.002	-2372.54
T,F,F,T	0.101	-2368.08	0.102	-2368.52
F,T,F,T	0.021	-2369.67	0.007	-2371.20
T,T,F,T	0.018	-2369.79	0.038	-2369.52
F,F,T,T	0.059	-2368.61	0.008	-2371.03
T,F,T,T	0.196	-2367.41	0.200	-2367.85
F,T,T,T	0.011	-2370.31	0.036	-2369.57
T,T,T,T	0.024	-2369.53	0.046	-2369.32

B.4 RHMM Variant Marginal Likelihoods – Nugget effect

Table B.12 Sydney Close HMM variants with nugget effect posterior model probabilities

RHMM Model Label	Fitted Correlation with nugget effect		Exponential decay with nugget effect	
	Posterior Weight	Marginal Likelihood	Posterior Weight	Marginal Likelihood
1,1,1,1	0.014	-2972.72	0.007	-2969.09
1,1,1,2	0.000	-2977.58	0.004	-2969.72
1,1,2,1	0.735	-2968.76	0.102	-2966.44
1,1,2,2	0.018	-2972.48	0.000	-2974.43
1,1,2,3	0.170	-2970.23	0.002	-2970.22
1,2,1,1	0.000	-2979.54	0.003	-2969.98
1,2,2,1	0.016	-2972.62	0.674	-2964.56
1,2,2,2	0.000	-2977.20	0.000	-2972.94
1,2,2,3	0.002	-2974.49	0.018	-2968.18
1,2,3,1	0.039	-2971.71	0.185	-2965.85
1,2,3,2	0.002	-2974.46	0.000	-2973.10
1,2,3,3	0.000	-2976.90	0.000	-2974.14
1,2,3,4	0.004	-2974.10	0.004	-2969.66

Table B.13 Brisbane Close HMM variants with nugget effect posterior model probabilities

RHMM Model Label	Fitted Correlation with nugget effect		Exponential decay with nugget effect	
	Posterior Weight	Marginal Likelihood	Posterior Weight	Marginal Likelihood
1,1,1,1	0.001	-2766.75	0.000	-2759.70
1,1,1,2	0.005	-2765.26	0.000	-2759.99
1,1,2,1	0.000	-2771.33	0.000	-2777.08
1,1,2,2	0.000	-2770.55	0.000	-2769.37
1,1,2,3	0.000	-2768.37	0.000	-2776.37
1,2,1,1	0.068	-2762.66	0.050	-2752.00
1,2,1,3	0.867	-2760.12	0.950	-2749.05
1,2,2,1	0.000	-2768.88	0.000	-2765.36
1,2,2,2	0.000	-2769.90	0.000	-2767.40
1,2,2,3	0.001	-2767.21	0.000	-2764.26
1,2,3,1	0.002	-2766.23	0.000	-2761.83
1,2,3,3	0.006	-2765.06	0.000	-2761.24
1,2,3,4	0.049	-2762.99	0.000	-2757.68

B.5 RHMM Variant 7 Site initial tests

Note: Culling of models was performed to reduce the computation time The number of regions had a maximum of three. The minimum number of sites in a region was 2.

Table B.14 7 Site Sydney 1886-1993 RHMM variants with Exponential Decay correlation and nugget effect posterior model probabilities

Model Label	Posterior Weight	Marginal Likelihood	Model Label	Posterior Weight	Marginal Likelihood
1,1,1,1,1,1,1	0.000011	-5018.96	1,1,2,2,3,3,2	0.021922	-5011.40
1,1,1,1,1,2,2	0.000002	-5020.52	1,1,2,2,3,3,3	0.000008	-5019.28
1,1,1,1,2,1,2	0.000009	-5019.14	1,1,2,3,2,3,2	0.000009	-5019.24
1,1,1,1,2,2,1	0.004241	-5013.04	1,1,2,3,3,1,2	0.192125	-5009.23
1,1,1,1,2,2,2	0.000001	-5021.24	1,1,2,3,3,2,2	0.235293	-5009.03
1,1,1,2,1,1,2	0.000000	-5024.14	1,1,2,3,3,3,2	0.135846	-5009.58
1,1,1,2,1,2,1	0.011464	-5012.05	1,2,1,1,1,1,2	0.000008	-5019.33
1,1,1,2,2,1,1	0.075175	-5010.17	1,2,1,1,1,2,1	0.000002	-5020.87
1,1,1,2,2,1,2	0.000015	-5018.68	1,2,1,1,1,2,2	0.000026	-5018.15
1,1,1,2,2,2,1	0.285629	-5008.83	1,2,1,1,2,1,2	0.000000	-5024.68
1,1,1,2,2,2,2	0.000016	-5018.62	1,2,1,1,2,2,1	0.000000	-5026.98
1,1,1,2,2,3,3	0.003268	-5013.30	1,2,1,1,2,2,2	0.000000	-5024.24
1,1,1,2,3,2,3	0.000065	-5017.22	1,2,1,1,3,2,3	0.000001	-5022.05
1,1,1,2,3,3,2	0.000019	-5018.45	1,2,1,1,3,3,2	0.000006	-5019.54
1,1,2,1,1,1,2	0.003278	-5013.30	1,2,2,1,1,1,1	0.000000	-5023.91
1,1,2,1,1,2,2	0.000853	-5014.65	1,2,2,1,1,1,2	0.001106	-5014.39
1,1,2,1,2,1,2	0.000005	-5019.83	1,2,2,1,1,2,1	0.000000	-5023.50
1,1,2,1,2,2,2	0.000142	-5016.44	1,2,2,1,1,2,2	0.001088	-5014.40
1,1,2,1,3,3,2	0.009452	-5012.24	1,2,2,1,1,3,3	0.000043	-5017.64
1,1,2,2,1,1,1	0.000000	-5023.23	1,2,2,1,2,1,2	0.000027	-5018.09
1,1,2,2,1,1,2	0.000988	-5014.50	1,2,2,1,2,2,1	0.000000	-5025.57
1,1,2,2,1,2,1	0.000011	-5018.97	1,2,2,1,2,2,2	0.000036	-5017.80
1,1,2,2,2,1,1	0.006142	-5012.67	1,2,2,1,3,1,3	0.000001	-5021.27
1,1,2,2,2,1,2	0.000066	-5017.21	1,2,2,1,3,2,3	0.000000	-5024.35
1,1,2,2,2,2,1	0.003900	-5013.13	1,2,2,1,3,3,1	0.000000	-5022.19
1,1,2,2,2,2,2	0.000089	-5016.90	1,2,2,1,3,3,2	0.001502	-5014.08
1,1,2,2,2,3,3	0.001484	-5014.09	1,2,2,1,3,3,3	0.000000	-5024.64
1,1,2,2,3,1,3	0.000012	-5018.95	1,2,3,1,1,2,3	0.003414	-5013.26
1,1,2,2,3,2,3	0.000007	-5019.52	1,2,3,1,2,2,3	0.000844	-5014.66
1,1,2,2,3,3,1	0.000348	-5015.54	1,2,3,1,3,2,3	0.000000	-5023.35

Sites : Bingara, Mudgee, Mt Vic/Blackheath, Sydney, Moss Vale, Moruya Heads, Taralga

Table B.15 7 Site Brisbane 1900-1986 RHMM variants with Exponential Decay correlation and nugget effect posterior model probabilities

Model Label	Posterior Weight	Marginal Likelihood	Model Label	Posterior Weight	Marginal Likelihood
1,1,1,1,1,1,1	0.019869	-4100.15	1,2,2,1,2,1,2	0.000000	-4117.91
1,1,1,1,1,2,2	0.000276	-4104.42	1,2,2,1,2,2,1	0.000000	-4119.43
1,1,1,1,2,1,2	0.002636	-4102.17	1,2,2,1,2,2,2	0.000090	-4105.55
1,1,1,1,2,2,1	0.001070	-4103.07	1,2,2,1,2,3,3	0.001933	-4102.48
1,1,1,1,2,2,2	0.016528	-4100.33	1,2,2,1,3,1,3	0.001204	-4102.95
1,1,1,2,1,1,2	0.000000	-4119.44	1,2,2,1,3,3,1	0.033805	-4099.62
1,1,1,2,2,1,1	0.000257	-4104.49	1,2,2,1,3,3,3	0.099795	-4098.53
1,1,1,2,2,1,2	0.000000	-4116.86	1,2,2,2,1,1,1	0.081125	-4098.74
1,1,2,2,1,1,1	0.202388	-4097.83	1,2,2,2,1,1,2	0.000200	-4104.75
1,1,2,2,1,1,2	0.015539	-4100.39	1,2,2,2,2,1,1	0.005633	-4101.41
1,2,1,1,2,1,1	0.001263	-4102.90	1,2,2,2,2,1,2	0.000025	-4106.83
1,2,1,1,2,1,2	0.000000	-4119.87	1,2,2,2,3,1,3	0.000725	-4103.46
1,2,1,1,2,2,1	0.000274	-4104.43	1,2,2,3,1,1,3	0.000882	-4103.26
1,2,1,1,2,2,2	0.017869	-4100.25	1,2,2,3,2,1,3	0.000409	-4104.03
1,2,1,1,2,3,3	0.035436	-4099.57	1,2,2,3,3,1,1	0.075837	-4098.81
1,2,2,1,1,1,1	0.018581	-4100.22	1,2,2,3,3,1,3	0.004366	-4101.66
1,2,2,1,1,2,2	0.000000	-4118.38	1,2,3,3,2,1,1	0.333421	-4097.33
1,2,2,1,1,3,3	0.001854	-4102.52	1,2,3,3,2,1,2	0.017665	-4100.27
1,2,2,1,2,1,1	0.001032	-4103.11	1,2,3,3,2,1,3	0.008015	-4101.06

Sites : Cape Capricorn, Caboolture, Cape Moreton, Brisbane, Pittsworth, Miles, Bingara.

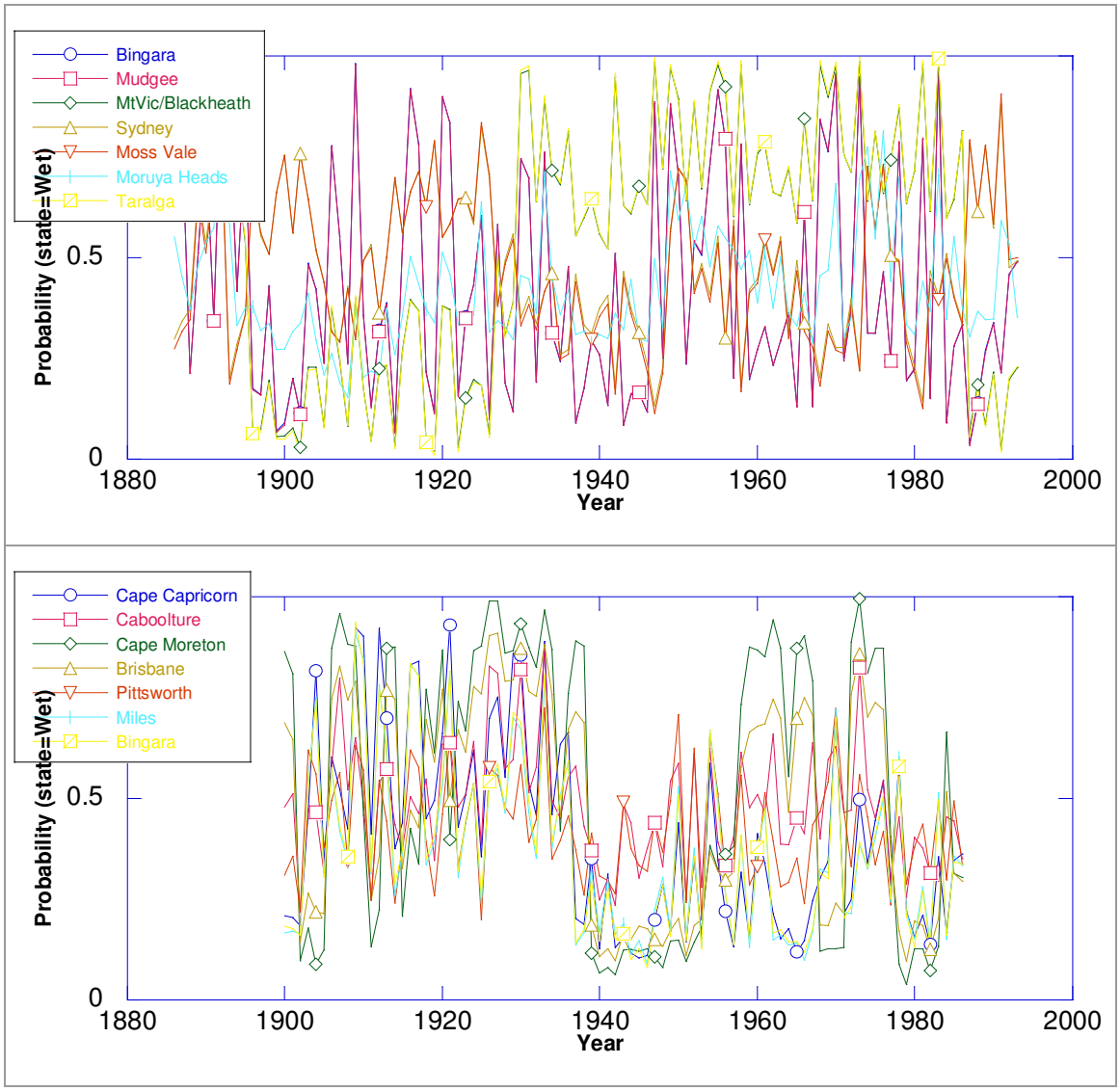


Figure B.1 7-Site State Series probabilities for (a) Sydney and (b) Brisbane