

Australian Rainfall & Runoff

Revision Projects

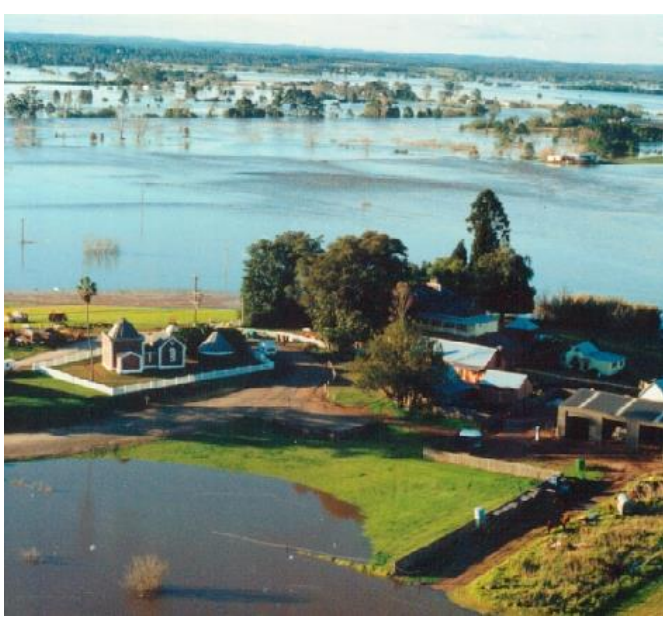
PROJECT 4

Continuous Rainfall Sequences
at a Point

STAGE 2 REPORT

P4/S2/014

June 2012



Australian Government




ENGINEERS
AUSTRALIA
Water Engineering

**AUSTRALIAN RAINFALL AND RUNOFF
PROJECT 4: CONTINUOUS RAINFALL SEQUENCES AT A POINT**

STAGE 2 REPORT

AUGUST, 2012

Project Project 4: Continuous Rainfall Sequences at a Point	AR&R Report Number P4/S2/014
Date 2 August 2012	ISBN 978-0-85825-877-8
Contractor UNSW Water Research Centre	Contractor Reference Number 2010/14
Authors Seth Westra Rajeshwar Mehrotra Ashish Sharma	Verified by 

COPYRIGHT NOTICE



This document, Project 4: Continuous Rainfall Sequences at a Point 2012, is licensed under the [Creative Commons Attribution 4.0 Licence](#), unless otherwise indicated.

Please give attribution to: © Commonwealth of Australia (Geoscience Australia) 2012

We also request that you observe and retain any notices that may accompany this material as part of the attribution.

Notice Identifying Other Material and/or Rights in this Publication:

The authors of this document have taken steps to both identify third-party material and secure permission for its reproduction and reuse. However, please note that where these third-party materials are not licensed under a Creative Commons licence, or similar terms of use, you should obtain permission from the rights holder to reuse their material beyond the ways you are permitted to use them under the 'fair dealing' provisions in the [Copyright Act 1968](#).

Further Information

For further information about the copyright in this document, please contact:

Intellectual Property and Copyright Manager

Corporate Branch

Geoscience Australia

GPO Box 378

CANBERRA ACT 2601

Phone: +61 2 6249 9367 or email: copyright@ga.gov.au

DISCLAIMER

The [Creative Commons Attribution 4.0 Licence](#) contains a Disclaimer of Warranties and Limitation of Liability.

ACKNOWLEDGEMENTS

This project was made possible by funding from the Federal Government through the Department of Climate Change. This report and the associated project are the result of a significant amount of in kind hours provided by Engineers Australia Members.



ENGINEERS
AUSTRALIA
Water Engineering

UNSW Water Research Centre

The University of New South Wales
Sydney, NSW, 2052

Tel: (02) 9385 5017

Fax: (02) 9313 8624

Web: <http://water.unsw.edu.au>



UNSW
THE UNIVERSITY OF NEW SOUTH WALES

water@
UNSW
water research centre

FOREWORD

AR&R Revision Process

Since its first publication in 1958, Australian Rainfall and Runoff (ARR) has remained one of the most influential and widely used guidelines published by Engineers Australia (EA). The current edition, published in 1987, retained the same level of national and international acclaim as its predecessors.

With nationwide applicability, balancing the varied climates of Australia, the information and the approaches presented in Australian Rainfall and Runoff are essential for policy decisions and projects involving:

- infrastructure such as roads, rail, airports, bridges, dams, stormwater and sewer systems;
- town planning;
- mining;
- developing flood management plans for urban and rural communities;
- flood warnings and flood emergency management;
- operation of regulated river systems; and
- prediction of extreme flood levels.

However, many of the practices recommended in the 1987 edition of AR&R now are becoming outdated, and no longer represent the accepted views of professionals, both in terms of technique and approach to water management. This fact, coupled with greater understanding of climate and climatic influences makes the securing of current and complete rainfall and streamflow data and expansion of focus from flood events to the full spectrum of flows and rainfall events, crucial to maintaining an adequate knowledge of the processes that govern Australian rainfall and streamflow in the broadest sense, allowing better management, policy and planning decisions to be made.

One of the major responsibilities of the National Committee on Water Engineering of Engineers Australia is the periodic revision of ARR. A recent and significant development has been that the revision of ARR has been identified as a priority in the Council of Australian Governments endorsed National Adaptation Framework for Climate Change.

The update will be completed in three stages. Twenty one revision projects have been identified and will be undertaken with the aim of filling knowledge gaps. Of these 21 projects, ten projects commenced in Stage 1 and an additional 9 projects commenced in Stage 2. The remaining two projects will commence in Stage 3. The outcomes of the projects will assist the ARR Editorial Team with the compiling and writing of chapters in the revised ARR.

Steering and Technical Committees have been established to assist the ARR Editorial Team in guiding the projects to achieve desired outcomes. Funding for Stages 1 and 2 of the ARR revision projects has been provided by the Federal Department of Climate Change and Energy

Efficiency. Funding for Stages 2 and 3 of Project 1 (Development of Intensity-Frequency-Duration information across Australia) has been provided by the Bureau of Meteorology.

Project 4: Continuous Rainfall Sequences at a Point

Continuous simulation of rainfall sequences are becoming increasingly important in design flood estimation as they represent, arguably, the most rigorous technique available to represent the joint behaviour of flood-producing extreme rainfall events, the preceding antecedent rainfall conditions, and the influence of non-stationary catchment conditions. This report describes the outcomes from the second stage of ARR research project 4. The objectives of this stage are to:

- (1) finalise the development of the regionalised state-based method of fragments approach as well as the development of a regionalised daily rainfall generation model; and
- (2) assess the performance of the method of fragments model using the same statistics and locations that were used in Frost et al [2004].

Arising from this project, methods were developed to allow for the generation of sequences of point-rainfall at the resolution of the pluviograph data (in this study taken to be in increments of 6-minutes) at any location in Australia. The testing conducted in this phase of work focused on statistics relevant for using continuous simulation in flood frequency estimation. Specifically, the method was tested in the context of the capacity to reproduce both extreme rainfall and the antecedent rainfall leading up to the annual maxima event, with the suite of methods generally performing well against these metrics.



Mark Babister
Chair Technical Committee for
ARR Research Projects



Assoc Prof James Ball
ARR Editor

AR&R REVISION PROJECTS

The 21 AR&R revision projects are listed below:

AR&R Project No.	Project Title
1	Development of intensity-frequency-duration information across Australia
2	Spatial patterns of rainfall
3	Temporal pattern of rainfall
4	Continuous rainfall sequences at a point
5	Regional flood methods
6	Loss models for catchment simulation
7	Baseflow for catchment simulation
8	Use of continuous simulation for design flow determination
9	Urban drainage system hydraulics
10	Appropriate safety criteria for people
11	Blockage of hydraulic structures
12	Selection of an approach
13	Rational Method developments
14	Large to extreme floods in urban areas
15	Two-dimensional (2D) modelling in urban areas.
16	Storm patterns for use in design events
17	Channel loss models
18	Interaction of coastal processes and severe weather events
19	Selection of climate change boundary conditions
20	Risk assessment and design life
21	IT Delivery and Communication Strategies

AR&R Technical Committee:

Chair: Mark Babister, WMAwater

Members: Associate Professor James Ball, Editor AR&R, UTS
 Professor George Kuczera, University of Newcastle
 Professor Martin Lambert, Chair NCWE, University of Adelaide
 Dr Rory Nathan, SKM
 Dr Bill Weeks, Department of Transport and Main Roads, Qld
 Associate Professor Ashish Sharma, UNSW
 Dr Bryson Bates, CSIRO
 Steve Finlay, Engineers Australia

Related Appointments:

ARR Project Engineer:

Monique Retallick, WMAwater

Assisting TC on Technical Matters:

Dr Michael Leonard, University of Adelaide

PROJECT TEAM

Project team:

- Dr Seth Westra, UNSW
- Dr Rajeshwar Mehrotra, UNSW
- A/Prof Ashish Sharma, UNSW
- Dr Sri Srikanthan, BOM

Independent review team:

- Emeritus Professor Geoff Pegram, University of KwaZulu-Natal, South Africa

EXECUTIVE SUMMARY

Continuous simulation of rainfall sequences is becoming an increasingly important tool in design flood estimation, as it represents arguably the most rigorous technique available to represent the joint behaviour of flood-producing extreme rainfall events and the preceding antecedent rainfall conditions. To inform the forthcoming revision of Australian Rainfall and Runoff (ARR), the aims of this project are to develop, test and validate the procedures for continuous rainfall simulation.

This report describes the outcomes from the second stage of ARR research project 4. The objectives of this stage are to: (1) finalise the development of the regionalised state-based method of fragments approach as well as the development of a regionalised daily rainfall generation model; and (2) assess the performance of the method of fragments model using the same statistics and locations that were used in Frost *et al* [2004], to enable direct comparison with the Disaggregated Rectangular Intensity Pulse (DRIP) model of Heneker *et al* (2001), and the Neyman-Scott Rectangular Pulse (NSRP) described by Cowpertwait *et al* [2002].

Regionalised state-based method of fragments disaggregation and daily Markov model

In this report we describe two regionalised methods which in combination enable the use of nearby rainfall records for cases where at-site records are unavailable or insufficiently long. The first of these methods is a regionalised version of the state-based method of fragments disaggregation model first described in Westra and Sharma [2010], where conditional to at-site daily rainfall, sub-daily rainfall fragments are drawn from nearby pluviograph gauges. The second method is a regionalised version of a Markov daily rainfall generation model, which enables generation of extended at-site daily rainfall sequences using information from a set of nearby daily rainfall gauges.

The regionalised state-based method of fragments logic uses the assumption of constant scaling between daily and sub-daily rainfall over some geographic region to combine at-site daily rainfall data with nearby sub-daily records. The scaling assumption was tested using a two-sample two-dimensional Kolmogorov-Smirnov test on a range of sub-daily rainfall attributes, with the results showing a high probability that the daily/sub-daily rainfall scaling at any two stations will be statistically similar provided that the distance between them is small. A logistic regression was then formulated to identify the main covariates that determine whether the daily to sub-daily scaling at two stations are similar, with the outcome that the scaling relationship is influenced by a combination of difference in latitude, longitude, elevation and distance to coast.

This approach allows for the generation of extended sequences of sub-daily rainfall at any location provided sufficient daily data is available, and thus makes use of the abundance of extended daily rainfall records across Australia relative to pluviograph records. Nevertheless

there are many regions in which extended daily data is not available. Furthermore, even if this data is available, the observational record only represents a single realisation of how daily rainfall might evolve in the future, and thus may result in undersimulation of the total variability in future daily rainfall.

As a result of these issues, a regionalised version of a Markov daily rainfall generation model also was developed, which allows extended realisations of daily rainfall to be generated at any location, regardless of whether or not sufficient daily data is available at that location. Similar to the regionalised state-based method of fragments model, this approach starts by identifying nearby daily-read gauges based on the scaling between annual and sub-annual rainfall. Unlike the regionalised method of fragments logic, however, this method does not directly draw daily rainfall fragments from nearby stations but rather uses the nearby stations to estimate the model parameters and then uses these parameters to generate daily rainfall at the location of interest. The method is based on an at-site version developed by Mehrotra and Sharma [2007a; b] and uses a Markov occurrence model conditional to both previous day's rainfall occurrence to account for short-memory persistence as well as aggregate rainfall over the previous 365 days to simulate longer timescale persistence. Amounts are simulated using a kernel density estimation procedure with conditional dependence on previous day's rainfall.

In combination, these two methods allow for the generation of sequences of point-rainfall at the resolution of the pluviograph data (in this study taken to be in increments of 6-minutes) at any location in Australia. An adapted and shortened version of chapters 2 to 4 have been accepted for publication as a two-paper series in *Water Resources Research*. The testing conducted in this phase of work focused on statistics relevant for using continuous simulation in flood frequency estimation. Specifically, the method was tested in the context of the capacity to reproduce both extreme rainfall and the antecedent rainfall leading up to the annual maxima event, with the suite of methods generally performing well against these metrics. The daily model was also tested for the number of wet days and intensity per wet day, with mean rainfall being well reproduced at most locations, although standard deviations were generally undersimulated. This was also reflected in the total annual rainfall distribution plots, where the mean of the simulated data was generally similar to the observations but the variance was too low. Although on balance the performance was satisfactory for the fully regionalised model, some deterioration could be observed particularly when the daily model was also regionalised, and this deterioration was most notable for the locations such as Alice Springs where there are limited nearby daily and/or sub-daily rain gauges.

Comparison with Frost et al [2004]

Having developed the suite of regionalised methods and tested them against a set of statistics relevant for flood estimation, the method was then evaluated in more detail using identical statistics to those used in Frost et al [2004] to test the DRIP and NSRP models. As the models used in Frost et al [2004] were based on at-site parameter estimates rather than

on the regionalised version of these models, the continuous simulation approach adopted here was based on the daily rainfall generation model described by Mehrotra and Sharma [2007a; b] and the at-site state-based method of fragments logic described by Sharma and Srikanthan [2006].

In evaluating the sub-daily rainfall statistics, the state-based method of fragments model performed similarly to the DRIP model, with both models on balance outperforming the NSRP model, with the exception of the autocorrelation statistics which were better reproduced in the NSRP model. Whereas the state-based method of fragments model outperformed DRIP in simulating annual maximum storm bursts for sub-daily durations, DRIP appeared superior in simulating the distribution of total annual rainfall. Nevertheless these differences were generally minor, and both models appear adequate for use in generating continuous rainfall data at the sub-daily timescale.

In evaluating the daily rainfall statistics, the state-based method of fragments performance is generally good across most statistics, with similar or better performance to both DRIP and NSRP. A weakness of the NSRP was a major overestimation of dry spell means and standard deviations at most locations, as well as an overestimation of wet spell means. The longer dry and wet spells simulated by the NSRP suggest a greater clustering of wet spells, which can have an impact on whether this method properly simulates antecedent rainfall conditions prior to the storm event. The main weakness of the Markov model is a slight underestimation of annual variability, with the method underestimating the probability of the driest and wettest years. The inclusion of the previous 365 day's rainfall as a predictor in the Markov model was designed to address this issue, however the results indicate that further work on this area is required. Although the issue appears to be systematic (occurring in most of the 10 locations studied), the magnitude of the underestimation is generally low, with the observations usually falling within the 90% confidence intervals.

Finally, although several alternative regional methods are available for continuous simulation, such as a regionalised version of DRIP by Jennings et al [2009] and regionalised versions of the Poisson cluster models by Gyasi-Agyei [1999], Gyasi-Agyei and Parvez Bin Mahbub [2007] and Cowpertwait and O'Connell [1997], there is still a significant research requirement in evaluating the performance of differing classes of regionalised models. Testing conducted on the regionalised state-based method of fragments and modified Markov models described in this report show fairly limited deterioration in model performance as the model becomes increasingly regionalised. The principal exceptions are for regions where there are limited nearby daily rainfall or pluviograph gauges, or for regions that are very climatologically different from the nearby gauged locations (for example in mountainous areas). Unfortunately other regionalised approaches, which generally involve estimating model parameters based on nearby gauges, would be likely to suffer from similar limitations. This highlights that despite the rapid advance in approaches in regionalising rainfall generation, there remains a continued need for collecting high quality point rainfall data at all timescales.

Table of Contents

1.	Overview	1
2.	Continuous Simulation Methodology Part I: A Non-parametric Approach to Rainfall Disaggregation	4
2.1.	Introduction.....	4
2.2.	Methodology	6
2.3.	Application	10
2.4.	Results	11
2.5.	Discussion	15
3.	Continuous Simulation Methodology Part II: A Regionalised Sub-daily Disaggregation Approach	16
3.1.	Introduction.....	16
3.2.	Data.....	17
3.3.	Methodology	19
3.3.1.	Regionalised state-based method of fragments algorithm.....	19
3.3.2.	Daily to sub-daily scaling	20
3.3.3.	Defining similarity.....	22
3.3.4.	Predictive model for statistical similarity.....	23
3.4.	Results	30
3.4.1.	Identifying ‘nearby’ stations - application to Sydney Airport.....	30
3.4.2.	Model validation.....	30
3.5.	Summary	37
4.	Continuous Simulation Methodology Part III: A Regionalised Approach to Daily Rainfall Generation	39
4.1.	Introduction.....	39
4.2.	Data.....	40

4.3.	Methodology	42
4.3.1.	Regionalised daily rainfall generation.....	42
4.3.2.	Identifying ‘nearby’ daily rainfall stations	44
4.4.	Results	53
4.4.1.	Annual and seasonal statistics.....	55
4.4.2.	Sub-daily statistics	58
4.5.	Summary	65
5.	Comparison with DRIP and NSRP	66
5.1.	Overview	66
5.2.	Sub-daily results	66
5.3.	Daily rainfall results.....	68
6.	Discussion and conclusions.....	70
6.1.	Regionalised state-based method of fragments and modified Markov models.....	70
6.2.	Comparison with DRIP and NSRP	71
6.3.	Recommendations and outstanding issues.....	71
	References	73
	Appendix A: State-Based Method of Fragments sub-daily evaluation statistics	77
	Appendix B: Daily Markov model - evaluation statistics.....	91

1. Overview

Continuous rainfall simulation represents an increasingly important tool in flood hydrology as it is capable of simulating the complete sequence of rainfall events which lead to flooding, including both the peak rainfall and the moisture conditions prior to the event. The benefits of such an approach were described in Kuczera et al. [2006], in which continuous simulation together with event joint probability methods based on Monte Carlo simulation were suggested as viable alternatives to the design storm approach, particularly for volume-sensitive systems where the role of antecedent rainfall conditions is important.

Although the field of continuous rainfall simulation has a long history with a wide range of modelling frameworks now available (see the Stage 1 report by Westra and Sharma [2010] for a detailed review of relevant literature), there are limited models which are able to use regionalised information to develop continuous rainfall sequences at the sub-daily timescale. The primary exceptions to this are a recently developed regionalised version of DRIP by Jennings et al [2009], regionalised versions of the Poisson cluster models by Gyasi-Agyei [1999], Gyasi-Agyei and Parvez Bin Mahbub [2007] and Cowpertwait and O'Connell [1997], and a regionalised k -nearest neighbour method first proposed in Westra and Sharma [2010] and described more fully herein.

This report provides the first detailed description of a suite of regionalised models which are able to integrate information from both nearby daily-read and sub-daily rainfall stations for cases where at-site data is unavailable or limited. An important advantage of this modelling framework is that it is able to make use of all the data available in the vicinity of the location of interest, rather than just a subset of long high-quality records as is the case with parametric alternatives. For example, by separating the daily and sub-daily rainfall generation algorithms, the methods are not impacted by the asymmetry of data availability between daily and sub-daily records (with an order of magnitude fewer pluviograph records being available than daily records). Furthermore, as the disaggregation method is based on resampling, it is able to use a combination of short and long pluviograph stations; thus, a pluviograph station with only, say, three years of record still can be included in the analysis alongside longer records. This substantially expands the value of the pluviograph record in Australia, as in many cases the records at individual stations are too short for meaningful analysis by themselves.

A second advantage is that a modelling framework, rather than a single model, is presented, providing significant flexibility in tailoring the model to intended applications, with examples of different model elements that have been developed at the University of NSW and the Australian Bureau of Meteorology provided in Table 1.1. Note that this list does not attempt to provide an exhaustive review of all possible combinations of models for stochastically generating precipitation, with a much more detailed review provided in [Westra and Sharma, 2010].

As can be seen in the table, the daily model allows for the inclusion of covariates into the Markov model such as previous 365-day rainfall occurrence or climate information such as that captured by climate indices, thereby allowing for the simulation of inter-annual and longer time-scale variability [Mehrotra and Sharma, 2007a; b]. A modified version of this approach also has been used to incorporate atmospheric covariates, and thus can be used for downscaling to enable simulation of daily rainfall under a future climate [Mehrotra and Sharma, 2005; 2006b; 2010]. Finally, a multi-site extension of the daily rainfall model is available [Mehrotra and Sharma, 2007a; b; Srikanthan and Pegram, 2009] in which dependencies across multiple sites are preserved. In terms of the sub-daily model, there is flexibility in choosing the number of sub-daily stations from which to draw data, with a greater number of stations meaning the capacity to simulate more variability and thus a larger diversity of extreme storm events. This may be useful when attempting to simulate very low exceedance probability events, but with the penalty of inducing more potential bias (with stations becoming increasingly far away from the location of interest and thus being less representative of at-site rainfall). Furthermore, the framework proposed here can potentially be adapted to a sub-daily downscaling approach which will complement the daily downscaling approach, and address the limitation that the daily timescale is too coarse to account for the type and intensity of individual storm systems which will occur in a future climate.

Table 1.1: Suite of continuous simulation models which have been developed at the University of New South Wales

	Daily to sub-daily disaggregation	Daily rainfall generation
At-site point rainfall	Sharma and Srikanthan [2006]; Chapter 2 of this report	Harrold et al [2003a; b]; Mehrotra and Sharma [2007a; b]
Regionalised point rainfall	[Westra and Sharma, 2010]; Chapter 3 of this report	Chapter 4 of this report
Low-frequency natural climate variability	Not applicable – assumes low-frequency variability is represented in daily rainfall sequences	Mehrotra and Sharma [2007a; b]
Multi-site rainfall	Not available	Mehrotra and Sharma [2007a; b]; Srikanthan and Pegram [2009]
Downscaling to account for future climate change	Currently under development	Mehrotra and Sharma [2005; 2006b; 2010];

In addition to this suite of continuous simulation models summarised above and described more fully in subsequent chapters, a range of other conceptual approaches to modelling sub-daily rainfall are available including:

1. Event-based models such as the Disaggregated Rectangular Intensity Pulse (DRIP) model of Heneker *et al* (2001);

2. Poisson cluster models, including the Bartlett-Lewis Rectangular Pulse and the Neyman-Scott Rectangular Pulse (NSRP) family of models; and
3. Multi-scaling models, such as the canonical and microcanonical family of models.

Each of these models was described at length in Westra and Sharma [2010], and a detailed summary of two studies which compared these models was also provided. One of the main conclusions was that the multi-scaling models performed poorly across a range of statistics, including incorrect simulation of hourly variance (with the microcanonical cascades model oversimulating variance and the canonical model undersimulating variance), problematic representation of wet spells highlighting issues with correctly simulating rainfall persistence, and significant biases in the representation of rainfall extremes. For this reason, this class of models was not considered further in the present study.

Of the remaining models, DRIP represents an example of the event based family of models which has been found to perform well, is available for free from the eWater CRC Stochastic Climate Library toolkit (<http://www.toolkit.net.au/Tools/SCL>), and is well known amongst Australian hydrology practitioners. The NSRP model described by Cowpertwait et al [2002] is a recent implementation in the class of Poisson cluster models, and also has been found to perform well. Both these models were reviewed in a comprehensive report by Frost et al [2004], with a total of 20 validation statistics covering timescales from sub-daily through to annual considered at ten locations across all major climate zones in Australia. To avoid unnecessary duplication, the present work repeats the Frost et al [2004] analysis for the state-based method of fragments disaggregation model and Markov daily model. Code was kindly provided by Andrew Frost to allow the present analysis to be conducted as similarly as possible to the original study.

The remainder of this report is structured as follows. The following three chapters describe the regionalised rainfall generation method. The next chapter provides a more detailed overview of the at-site state-based method of fragments model first developed in Sharma and Srikanthan [2006], with a particular emphasis on the resampling approach which considers both previous- and next-day wetness state. Chapter 3 then generalises this to a regionalised setting, by resampling sub-daily fragments from nearby pluviograph gauges conditional to at-site daily rainfall. Chapter 4 completes the regionalisation by also outlining a regionalised daily rainfall generation model. Part of the content of these chapters were also submitted as a two-paper series to Water Resources Research, which is presently under review. Having developed these methods, they are compared with both DRIP and NSRP in Chapter 5, with detailed results provided in Appendices A and B. Finally, discussion and conclusions are provided in Chapter 6.

2. Continuous Simulation Methodology Part I: A Non-parametric Approach to Rainfall Disaggregation

2.1. Introduction

The estimation of flood frequency statistics for ungauged catchments continues to be a problem of great practical interest. While the option used most is to simulate the flood hydrograph using a design rainfall storm based on a model whose parameters are regionalized as functions of catchment characteristics, interest is growing on estimating floods using continuous (uninterrupted) flow simulation, either through historical rainfall records, or via stochastic rainfall generation [Blazkova and Beven, 2002; Boughton and Droop, 2003; Cameron *et al.*, 2000; Lamb and Kay, 2004]. Some of the arguments for continuous simulation are:

- 1) its ease of use in catchments that have undergone or are planned to undergo anthropogenic changes while avoiding the need to make unrealistic assumptions related to antecedent conditions necessary in a design storm approach;
- 2) the associated increase in reliability given the length of available rainfall records (it is unusual to find streamflow time series that are longer than the rainfall series for a catchment);
- 3) the relative stationarity that can be associated with rainfall records compared with streamflow records; and
- 4) the wealth of research on stochastic generation that serves as the platform for generating sequences at a fine temporal resolution (daily and sub-daily).

However, as has been pointed out in many studies (see Beven [2002] and references therein), design flood estimates from the simulated continuous flow series exhibit high sensitivity to both rainfall-runoff model parameter uncertainty and the nature of the observed or generated rainfall sequences being used as inputs. The research presented here focuses on the latter of the above two issues – the stochastic generation of continuous rainfall sequences that offer a more realistic representation of observed rainfall records, especially the attributes in rainfall that lead to extreme flood events after the rainfall-runoff transformation.

Stochastic generation of rainfall has an extensive history of research, with a range of methods available for simulation at single or multiple locations, and at sub-daily, daily and longer time scales. Comprehensive reviews of the state of the literature in this area are provided by [Sharma and Mehrotra, 2010; Srikanthan and McMahon, 2001; Westra and Sharma, 2010]. Generation of sub-daily rainfall is again a well-researched topic, with most approaches being formulated as variations of the Bartlett-Lewis [Onof *et al.*, 2000] or Neyman-Scott [Cowpervait *et al.*, 2002] rectangular pulse models that parameterise rainfall generation through a representation of storm cells comprising a full storm and aggregates to form the complete rainfall time series. While these are useful alternatives especially in cases where model parameters can be regionalised for use at ungauged locations, they suffer from

the rigidity of the framework assumed, and the need for long, high-quality datasets for estimating the parameter values needed.

Another way to simulate continuous rainfall involves first stochastically generating the daily rainfall, followed by its disaggregation to a sub-daily time-step. This disaggregation-based logic makes use of the significantly longer and denser daily rainfall records that exist, consequently ensuring that the continuous sequences when aggregated to a daily time step exhibit greater consistency with observations. The downside of the approach lies in the use of fewer and shorter sub-daily records that are needed to build the disaggregation model, along with the issue of generating a sub-daily rainfall event that does not naturally continue from the previous or into the next day. The availability of long daily observations coupled with an increasing density of shorter duration records has seen the development of a range of such disaggregation alternatives over time.

Daily to sub-daily disaggregation models usually fall into one of two categories. The first set of models use a parameterised representation of the daily to sub-daily conversion process, a good example being the random cascades approach [*Sivakumar and Sharma, 2008*], or the Bartlett-Lewis model based recursive disaggregation procedure [*Koutsoyiannis, 2003*]. The second category involves use of a nonparametric resampling rationale where observed sub-daily rainfall patterns are conditionally prescribed to suitably selected daily rainfall amounts, the earliest such approach being presented by [*Snavidze, 1977*] and termed as “Method of Fragments”. While the use of parametric alternatives has distinct advantages in allowing the possibility of regionalising parameters for use in ungauged locations, or enabling the development of spatio-temporal models as per [*Bowler et al., 2006*], such approaches often suffer from an inability to simulate the broad distributional and persistence attributes that real rainfall data exhibits.

Simplistic nonparametric resampling strategies are often no better. For instance, Figure 2.1 shows the difference between the probability of sampling a sequence of three continuous wet days in 166 locations having at least 82 years of daily records, versus the probability of sampling the three continuous wet days when the middle day corresponds to the annual maximum rainfall. It is clear that the annual maximum rainfall has a greater likelihood of being encompassed by wet days on either side, something that cannot be simulated using either the parametric or the simple nonparametric resampling alternative discussed above, nor by the conventional simulation procedures that assume independence between rainfall occurrences and amounts. Problems such as the ones above become all the more critical as we get closer towards using synthetically generated continuous rainfall sequences to generate continuous flow series that serve as the basis for estimating the design flood. There is a need for generation of alternatives that lead to rainfall sequences that do more than just match intensity-frequency-duration relationships to those derived from the observed record. One such approach, aimed specifically at imparting longer-term persistence in rainfall, leading to a more realistic representation of antecedent conditions prior to a flood causing burst, is presented here.

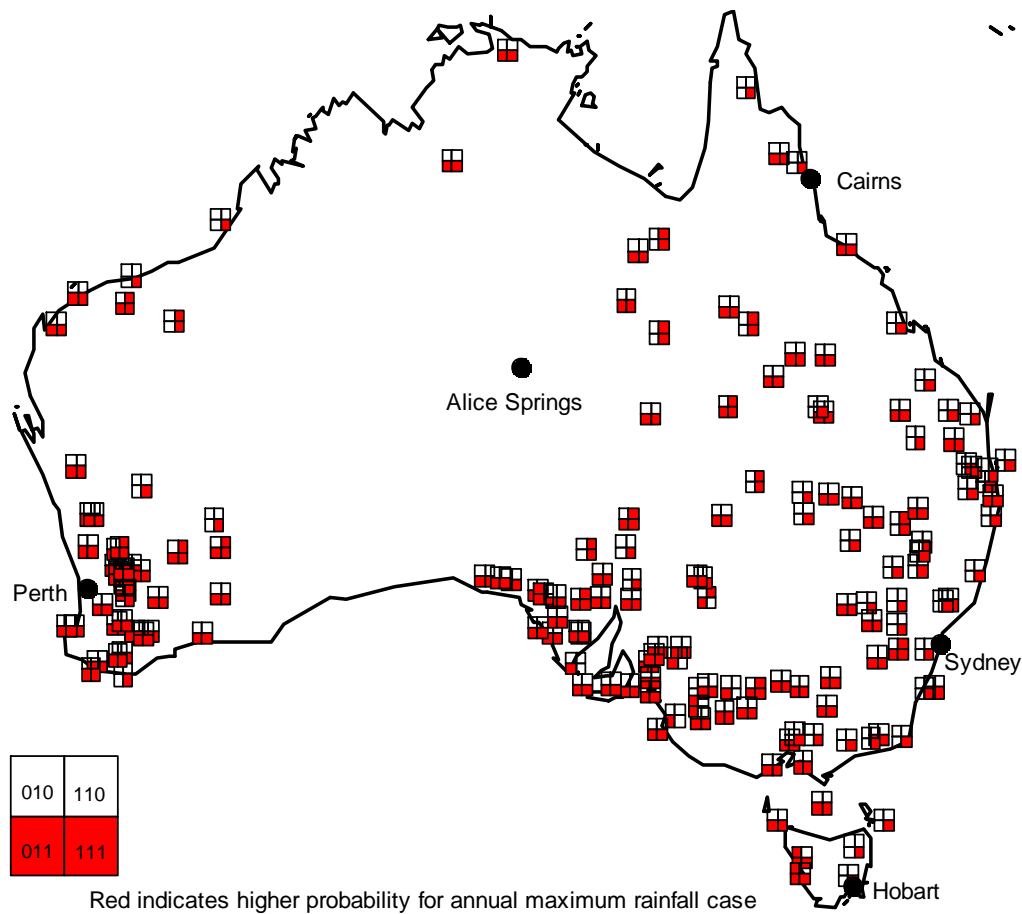


Figure 2.1. Differences in the probability of sampling a rainfall sequence ($i1j$), $i,j=[0,1]$, as compared to the case where the middle day represents the annual maximum rain. The percentage of stations where the probability conditional to the middle day being the annual maximum rainfall is greater than the marginal probability over the full rainfall record, equal 2.5% (010), 10.2% (110), 66.2% (011) and 95.2% (111) respectively. While the above numbers may be impacted by spatial dependence between locations, they generally point to a remarkably different dependence structure for sampling the annual maximum rain in comparison to rain on a usual day.

The rest of the chapter is organised as follows. The next sub-section presents the rationale behind the proposed disaggregation approach. This is followed by a description of the locations at which the approach was tested, along with the results obtained. Next, we present a discussion of the results, followed by conclusions and recommendations for future work.

2.2. Methodology

The rationale behind the continuous simulation approach proposed here is to ensure the representation of rainfall attributes that lead to an adequate representation of design flood

values. As has been demonstrated in Figure 2.1, this means ensuring an appropriate representation of the rainfall before and after the annual maximum events. In the context of sub-daily rainfall, this also implies an accurate representation of the antecedent conditions preceding the sub-daily annual maximum event, these antecedent conditions often being a result of the sub-daily rainfall on the day being disaggregated, along with the rain on preceding days. For continuous rainfall to lead to flows that result in a flood frequency relationship compatible with the one derived using observed flows, the following rainfall attributes must be represented well:

- (a) Within-day dependence and distributional representation - Generated sequences should be able to represent diurnal patterns such as increased probability of showers in morning or late afternoon and the dependence associated with differences in causative factors (convective versus frontal rainfall), while maintaining statistical correspondence with daily rainfall.
- (b) Representation of daily attributes - When sub-daily sequences are aggregated to a daily time scale, it is important that they maintain important attributes that impact on flow simulation. Some of these attributes are (i) seasonal representation, (ii) representation of spell characteristics (related to representation of dependence from one day to the next), and (iii) representation of low-frequency characteristics in rainfall that result in monthly or yearly rainfall being substantially different from one year to the next. While low frequency variability in rainfall is of less importance in the more extreme design rainfall events, our experience with arid and semi-arid catchments in Australia indicates an improper representation can lead to significant bias in the derived low-to-mid average recurrence interval (ARI) flood events [Weinmann *et al.*, 2002]. The usual approach to generating rainfall assumes a dependence structure that is incapable of simulating antecedent conditions prior to extreme events that are any different to those that occur the rest of the year. Figure 2.2 illustrates the probability distribution of the antecedent rainfall over a range of aggregation periods and design storm durations using long continuous rainfall data from Sydney. As can be inferred from the figure, the antecedent conditions prior to the annual maximum rain are markedly different to those that occur before more common rainfall events. Similar conclusions have been derived in earlier studies [Pui *et al.*, 2010a] and can also be inferred from Figure 2.1.

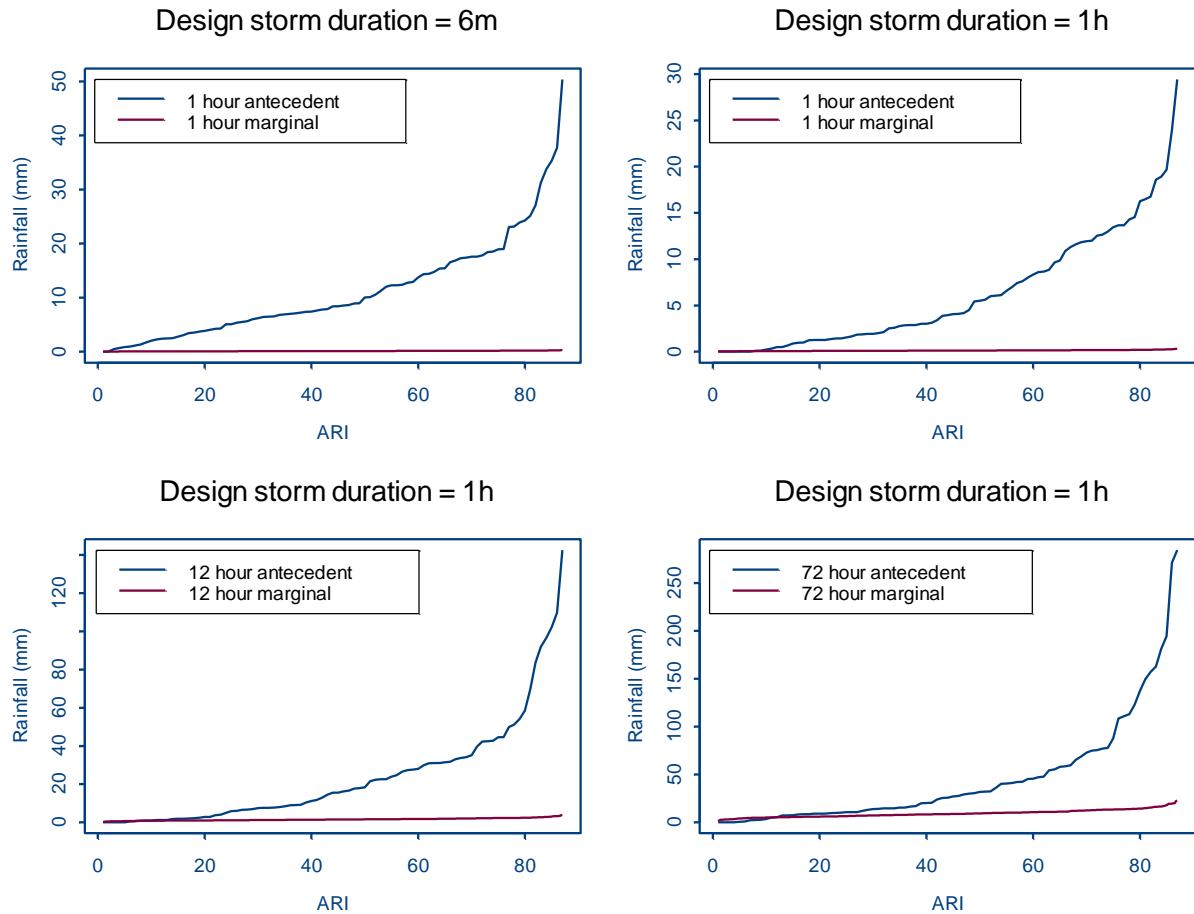


Figure 2.2: Antecedent rainfall distribution associated with design rainfall events using Observatory Hill (Sydney) continuous rainfall data. Note the difference in antecedent characteristics in comparison to the marginal distribution of rainfall, and the likely impact on the design flood if antecedent characteristics were assumed to represent “average” rainfall.

We present here a nonparametric alternative to generate continuous rainfall sequences that addresses the issues raised above. The proposed procedure resamples observed fractions of sub-daily rainfall with respect to the corresponding daily rainfall, through conditioning on a rainfall state that is defined based on the rainfall occurrence of the day before and after the day being disaggregated. The generation procedure can be expressed as follows:

Step 1: Form daily rainfall ($R_i = \sum_m X_{i,m}$) and sub-daily fragment ($f_{i,m} = X_{i,m} / R_i$) time series

where R_i represents the daily rainfall amount on day i , $X_{i,m}$ represents the sub-daily rainfall intensity on day i and sub-daily timestep m (with $m=1, \dots, 240$ for the 6-minute data used here), and $f_{i,m}$ represents the subdaily ‘fragment’.

Step 2: To disaggregate daily rainfall R_t for a particular day of the year t , form a moving window of length l days centred around day t . Segregate historical daily data into the following four rainfall classes:

$$\begin{aligned}
 \text{CLASS1:} & \quad R_j \geq 0 \mid (R_{j-1} = 0, R_{j+1} = 0) \\
 \text{CLASS2:} & \quad R_j \geq 0 \mid (R_{j-1} \geq 0, R_{j+1} = 0) \\
 \text{CLASS3:} & \quad R_j \geq 0 \mid (R_{j-1} = 0, R_{j+1} \geq 0) \\
 \text{CLASS4:} & \quad R_j \geq 0 \mid (R_{j-1} \geq 0, R_{j+1} \geq 0)
 \end{aligned} \tag{2.1}$$

where time j represents a day falling within the moving window centred on the current day t .

Step 3: Identify the class corresponding to the daily rainfall R_t that is to be disaggregated. Denote the class c_t where $c_t \in [1-4]$.

Step 4: Identify the k nearest neighbours of the conditioning vector $[R_t]$ as the days corresponding to the k lowest absolute departures $|R_j - R_t|$ where $c_j = c_t$. Specify $k = \sqrt{n}$ [Upmanu Lall and Sharma, 1996] where n represents the sample size of the class members falling within the moving window. The ranked daily rainfall from lowest absolute departure is then given as $R_{(j)}$, $j=1, \dots, k$, where the use of parentheses indicates use of ranked data. Sample neighbour j from the following conditional probability distribution [Upmanu Lall and Sharma, 1996; Mehrotra and Sharma, 2006a]:

$$p(j) = \frac{1/(j)}{\sum_{i=1}^k 1/i} \tag{2.2}$$

where $p(j)$ represents the probability of selecting neighbour (j) , with $(j)=1$ denoting the neighbour having the smallest absolute departure. Using a uniformly distributed random number $(0,1)$, select a neighbour $(R_{(j)}^o)$ using the probabilities in Equation 2. The fragments used to disaggregate can be specified as $f_{(j),m} = X_{(j),m}^o / R_{(j)}^o$.

Step 5: Increment t and repeat steps 2 to 4 until disaggregation is completed.

In the results reported in the next section, the above algorithm uses a moving window length of 15 days if the historical record is 40 years or longer, and 30 days if 20 years and shorter, with interpolated values in between. This data-dependent window length is chosen to ensure a sample size that is adequate for use in formulating the sub-daily time series, keeping in mind the typically few rainy days that span Australian rainfall records.

While the above disaggregation approach is reasonably simple, its applicability rests on a number of assumptions. The first assumption is that the daily rainfall record to be disaggregated represents the observed daily rainfall with respect to both its distributional and its dependence attributes. For instance, the disaggregation assumes that the sub-daily rainfall fractions depend on the rainfall class along with the daily amount, which requires that the relationship of the daily rainfall and the various classes is similar to that in the historical record. A second assumption is that the relationship between the sub-daily rainfall fractions

and the corresponding daily rainfall can be characterised sufficiently well based on the sample within the moving window. A third assumption is that the relationship between the sub-daily rainfall on one day and the sub-daily rainfall on adjacent days can be expressed with reference to the relationships between the corresponding daily totals. While all of these assumptions can become questionable depending on the nature of the rainfall being modelled, these were found to be acceptable in the context of the Australian rainfall records we have analysed for deriving design flood estimates. It should, however, be noted that the approach may result in a biased representation of long spells that stretch across days, as would be the case for the sustained frontal or cyclonic events that form the basis for design in large catchments in the northern regions of Australia. Having said that, the proposed approach is likely to serve as a reasonable basis for deriving design estimates for the faster responding urban catchments for which the method has been developed.

2.3. Application

The continuous rainfall generation rationale described in the previous section was tested using sub-daily rainfall data from five climatologically different locations in Australia (listed in Table 2.1 with locations indicated on Figure 2.1). It is worth pointing that all locations except Alice Springs fall on the coast and are impacted by shorter duration convective rainfall events, with a more marked seasonal behaviour as one progresses to the north (stronger summer rainfall) or the south (stronger winter rain) of the country. The aim of our exercise was to evaluate whether the use of the above logic enabled an improved representation of antecedent rainfall characteristics of the type illustrated in Figure 2.2. On average there was <1% missing data across the five stations used for model evaluation, and these have been infilled using the data from the same day at nearby stations, after adjusting for differences in total annual rainfall.

Table 2.1: Locations of the continuous rainfall stations used. Actual locations of each of the stations are indicated in Figure 2.1.

Number	Station Name	Gauge number	Start year	Number of years of record	Köppen climate classification
1	Sydney airport	066037	1961	45	Temperate (warm summer)
2	Perth airport	009021	1960	46	Sub-tropical (dry summer)
3	Alice Springs airport	015590	1950	57	Desert/Grassland (hot, persistently dry)
4	Cairns airport	031011	1941	66	Tropical (monsoonal)
5	Hobart airport	094008	1959	47	Temperate (mild summer)

As the aim of the work described in this section was to assess the suitability of the class-based disaggregation logic, observed daily rainfall (formed by aggregating the sub-daily rainfall records) were used as the basis of performing the disaggregation. Hence, the daily rainfall R_t in the above algorithm was specified equal to the observed daily rainfall for day t . To ensure that the simulation procedure did not resample the sub-daily rainfall for the same day (t), the sub-daily data corresponding to the day in question was discarded from the moving window associated with day t . Hence, the procedure effectively forced selection of a sub-daily rainfall pattern that did not correspond to that which was observed, analogous to the leave-one-out cross-validation approach used as a substitute for split-sample validation in many applications. One hundred replicates of sub-daily rainfall each of length equal to that of the observed record were simulated using the observed daily record for the five locations.

2.4. Results

Figure 2.3 presents the intensity-frequency relationship ascertained using the disaggregated sequences for rainfall duration of 6 minutes for each of the locations considered. Figure 2.4 then presents the exceedance probability associated with the antecedent conditions prior to the annual maximum events reported in Figure 2.3. While the nonparametric nature of the disaggregation scheme can be expected to lead to greater concurrency between the simulated and observed design intensities as illustrated in Figure 2.3, the similarity between the antecedent conditions observed and simulated prior to the annual maximum events reflects the importance of the class-based approach to deriving the disaggregated rainfall. What is important to note is the marked difference that would have been observed in these results without using the four rainfall classes, as was visible in the two exceedance curves in Figure 2.2.

Table 2.2 presents design intensities and corresponding antecedent rainfall for longer durations than the 6 minute duration considered in Figures 2.3 and 2.4. As can be noted from the results, there appears to be a broad concurrence between the observed and simulated results, most importantly for the antecedent rainfalls prior to the annual maximum events.

It is emphasised that the above results use the actual daily rainfall as the basis for disaggregation, hence while adopting a leave-one-out cross-validation rationale, it is not possible to fully represent the uncertainty that is introduced when a stochastically generated daily rainfall sequence is used. Chapter 4 explores the change in the disaggregated rainfall properties when this additional uncertainty is introduced.

Equally importantly, the above rationale uses the observed sub-daily rainfall record to perform the disaggregation. It is to be expected that the disaggregated sequence represents the observed sequence well, given the availability of the historical record the procedure works off. The real challenge in the disaggregation is the situation where the sub-daily rainfall is not available, which is when the utility of the disaggregation will be best felt. This problem (or disaggregating daily rainfall to continuous in the absence of a sub-daily observed record), is addressed in Chapter 3 of this report.

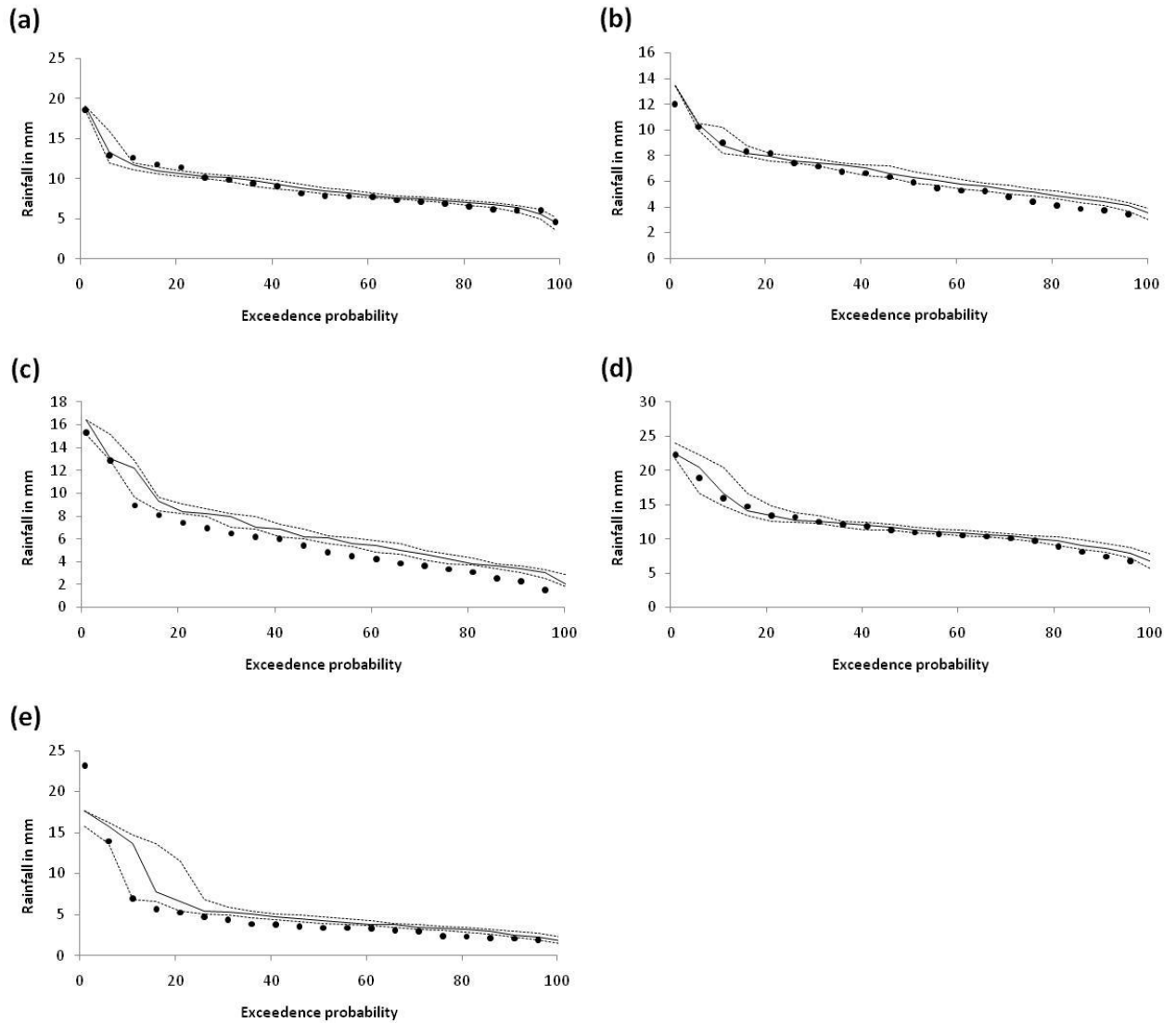


Figure 2.3: 6-minute annual maximum rainfall against exceedance probability for (a) Sydney, (b) Perth, (c) Alice Springs, (d) Cairns, and (e) Hobart. Black dots represents observed data, black solid line represents the median of 100 simulations, and black dotted lines represent the 5 and 95 percentile simulated values. Figure extracted from [Westra et al., 2012].

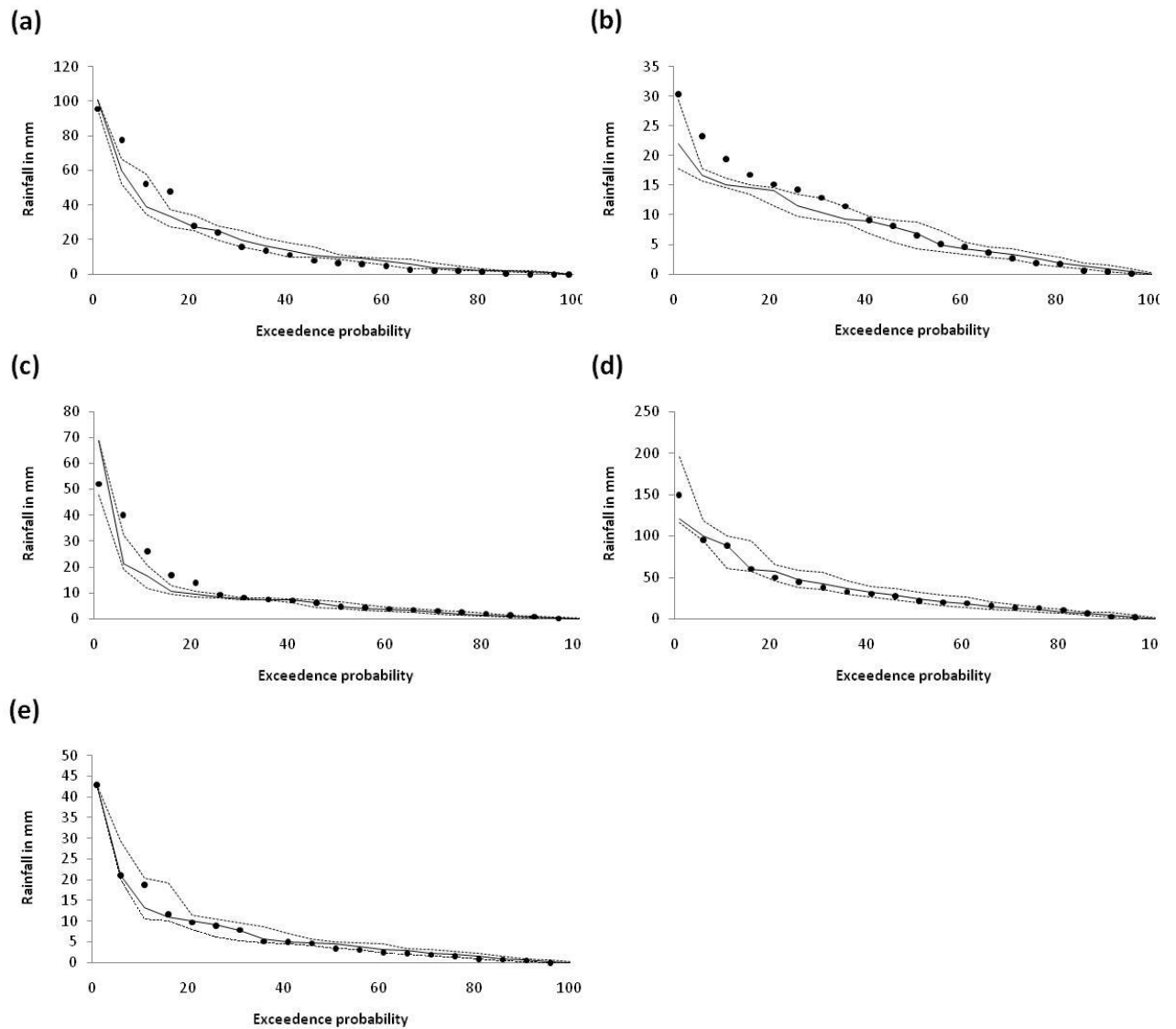


Figure 2.4: 6-hour antecedent rainfall prior to the 6-minute annual maximum storm burst plotted against exceedance probability for (a) Sydney, (b) Perth, (c) Alice Springs, (d) Cairns, and (e) Hobart. Black dots represents observed data, black solid line represents the median of 100 simulations, and black dotted lines represent the 5 and 95 percentile simulated values. Figure extracted from [Westra et al., 2012].

Table 2.2: Comparison of observed and simulated results for median annual maxima for different storm burst durations, and the antecedent rainfall prior to the 1 hour storm burst. The simulated median annual maxima represent the median of all 100 simulations.

	Sydney		Perth		Alice Springs		Cairns		Hobart	
	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)
Annual maxima										
6 min	8.87	9.06 (8.85-9.45)	6.18	6.57 (6.42-6.84)	5.50	6.68 (6.37-6.96)	11.6	12.1 (11.5-12.7)	4.51	5.64 (4.92-6.38)
30 min	25.7	23.3 (22.5-24.3)	14.7	15.3 (14.7-15.9)	16.7	18.3 (17.5-19.1)	34.9	35.3 (34.1-36.3)	11.3	12.5 (11.7-13.4)
1 hr	35.4	33.6 (33.3-34.9)	18.8	18.9 (18.1-19.6)	22.1	23.1 (22.2-24.0)	51.7	51.3 (49.7-53.4)	14.6	15.6 (14.9-16.3)
3 hr	55.4	52.1 (51.4-53.4)	29.0	28.2 (27.4-28.9)	32.6	32.4 (31.7-33.4)	83.5	86.8 (84.4-89.5)	22.9	23.2 (22.9-23.8)
6 hr	72.3	67.1 (65.5-69.8)	36.3	35.6 (34.8-36.4)	39.6	39.5 (38.8-40.0)	113	118 (115-123)	30.3	30.4 (29.9-30.9)
12 hr	91.8	84.9 (83.6-85.8)	45.4	44.3 (43.6-45.0)	48.2	47.3 (46.6-47.8)	147	155 (151-158)	39.6	39.1 (38.6-39.5)
Antecedent rainfall prior to 1-hr burst (mm)										
6 hr	15.4	13.5 (11.7-15.0)	6.76	5.73 (4.98-6.36)	6.10	5.20 (4.26-6.06)	25.4	27.2 (21.9-32.1)	6.31	5.31 (4.48-6.13)
12 hr	22.7	19.5 (16.8-23.4)	9.63	8.50 (7.60-9.22)	7.98	7.01 (6.06-8.46)	32.2	35.3 (28.7-40.4)	9.10	7.43 (6.29-8.47)
24 hr	31.4	28.6 (24.6-32.5)	11.9	12.4 (11.4-13.8)	10.6	11.0 (9.02-13.0)	40.3	51.4 (44.3-57.1)	9.09	11.1 (9.55-12.5)
48 hr	38.4	38.5 (35.2-43.8)	12.5	16.5 (14.7-18.5)	15.5	16.3 (13.2-19.1)	54.9	79.0 (72.3-89.8)	10.2	13.5 (11.6-15.3)

2.5. Discussion

This chapter presented a rationale for generating a near-continuous rainfall time series that aims to ensure that rainfall patterns prior to major storms events are represented in a realistic manner. The approach presented was nonparametric – it made no major assumptions about the nature of the relationship between continuous and aggregated rainfall – but was data-based and a sensible alternative to use as long as the daily and continuous rainfall data used in formulating the approach were representative of what can be expected for the location under study.

As mentioned before, the suitability of the proposed approach is linked strongly to the quality of daily and sub-daily rainfall data used. While daily data has a fairly broad coverage across Australia (and the world in general), allowing users to generate multiple realisations of daily rainfall for suitable record lengths, continuous rainfall records are available at fewer locations and for shorter lengths of time. In situations where these records are limited or not available, it is important that the approach described here be modified to use representative sub-daily records from other locations that are selected using a judicious and carefully designed procedure. The criteria that ought to be used in developing these representative sub-daily records, along with the issue of representing the increased uncertainty in the disaggregated rainfall through such a procedure, is discussed in the next chapter.

3. Continuous Simulation Methodology Part II: A Regionalised Sub-daily Disaggregation Approach

3.1. Introduction

In the previous chapter, a non-parametric k -nearest neighbour disaggregation model was presented in which sub-daily fragments of rainfall are re-sampled from the historical sub-daily pluviograph record conditional on the daily rainfall intensity and the previous and next day wetness state. This allows for the stochastic generation of continuous (uninterrupted) rainfall of any desired length, provided that daily data either is available or can be generated synthetically.

Although testing shows good performance across a range of statistics, the method suffers from two important limitations: firstly it is necessary to have long sub-daily records available at the location of interest, constraining the model to only a comparatively small number of locations where such long pluviograph records are available; and secondly that by re-sampling from the historical record, it is not possible to generate sequences that are more intense than the largest observed storm burst, which becomes important in flood frequency estimation when it is necessary to extrapolate beyond the largest observation.

Here we present a generalisation of the above approach by enabling the re-sampling of sub-daily fragments from pluviograph stations within some neighbourhood of the location of interest (henceforth referred to as the 'target' location), conditional on daily rainfall at that target location. This approach substantially expands the domain of applicability of the disaggregation logic to any location where sufficient daily data is available, with daily data generally being much more abundant than pluviograph data. To be able to perform this re-sampling, it is necessary to assume similar daily to sub-daily scaling at both the target location and the neighbouring locations from which the sub-daily fragments are to be sampled. The logic behind identifying the station 'neighbourhood' where such re-sampling is valid is the main contribution of this chapter.

The majority of work on regionalised disaggregation approaches described in the literature has thus far been based on the Poisson cluster family of models. For example, Cowpertwait et al [1996] and Cowpertwait and O'Connell [1997] developed a regionalised Neyman-Scott Rectangular Pulse (NSRP) model for generating sequences of hourly rainfall data across the UK, by regressing the NSRP parameters on site variables obtained from a relief map of the UK (namely: elevation, north-south distance, east-west effect and distance to coast). Cowpertwait et al [1996] also developed a disaggregation model that allows historical or generated hourly data to be disaggregated into totals for shorter time intervals. An alternative approach was proposed by Gyasi-Agyei [1999], who developed a regionalised version of the Gyasi-Agyei and Willgoose hybrid model based on the nonrandomised Bertlett-Lewis rectangular pulse and an autoregressive jitter [Gyasi-Agyei and Willgoose, 1997; 1999]. This approach uses observed daily statistics (namely dry probability, mean and variance) and two regionalised sub-daily parameter estimates, with promising results found in simulating sub-daily rainfall in central Queensland, Australia. This model was extended to Australia-wide data by Gyasi-Agyei and Parvez Bin Hahhub [2007], and found to be successful in simulating a range of statistics including extreme rainfall.

To our knowledge the approach presented here represents the first regionalised version of a resampling approach to continuous rainfall simulation, in which rather than identifying regionalised estimates of model parameters, we directly sample sub-daily fragments from nearby pluviograph stations. The benefits of such a re-sampling logic described in the previous chapter are also expected to be applicable here, particularly with respect to the manner in which the joint probability between extreme rainfall and antecedent rainfall conditions can be preserved.

The remainder of this chapter is structured as follows. In section 3.2 we provide an overview of Australia's continuous rainfall record. This is followed by a description of the proposed methodology, including the statistics used to determine the similarity between daily/sub-daily rainfall relationships at any two locations. Results are then presented in Section 3.4, including a preliminary analysis of the viability of the method at Sydney Airport, Australia, as well as more detailed results for five case study locations distributed throughout Australia. Finally, a brief summary is provided in Section 3.5.

3.2. Data

Continuous (sub-daily) rainfall data were obtained from the Australian Bureau of Meteorology at 1397 continuous pluviograph stations, in increments of 6 minutes. The location of each gauging station is shown in Figure 3.1, together with an indication of the length of record. Of the 1397 available gauging stations, 101 locations having length greater than 40 years, and a further 331 locations having length of between 20 and 40 years. In contrast, there are 17451 daily-read gauging stations in Australia, of which 2708 locations have records greater than 25 years and 1768 stations have more than 40 years of record, highlighting the potential benefits of developing a regionalised disaggregation approach which uses the conditional relationship between daily and sub-daily rainfall to generate sub-daily sequences at locations where only daily data is available or can be synthetically generated. As can be seen in Figure 3.1, the spatial distribution of the gauging stations is not homogeneous, with a high density of gauges in the populated regions particularly along the eastern coastal fringe of Australia, and lower density elsewhere.

The number of gauging stations with continuous rainfall records is plotted against the year of record in Figure 3.2. As can be seen, only a small number of gauging stations were available in the early 20th century (the longest available record being from Melbourne Regional Office, gauge number 086071, with data from 1873 to 2008), with significant increases in recording density apparent in the 1960s. To limit the effects of possible temporal variability in the daily/sub-daily characteristics, the remainder of the analysis only considers records between 1970 and 2005 with less than 20% of the record classified as 'missing', with a total of 232 stations meeting this criterion. 'Missing' data was defined as data which was flagged as either missing or presented as an accumulation over previous time steps, and in these cases the full day of record was removed from the analysis. As will be discussed further below, the proposed method is relatively insensitive to missing data.

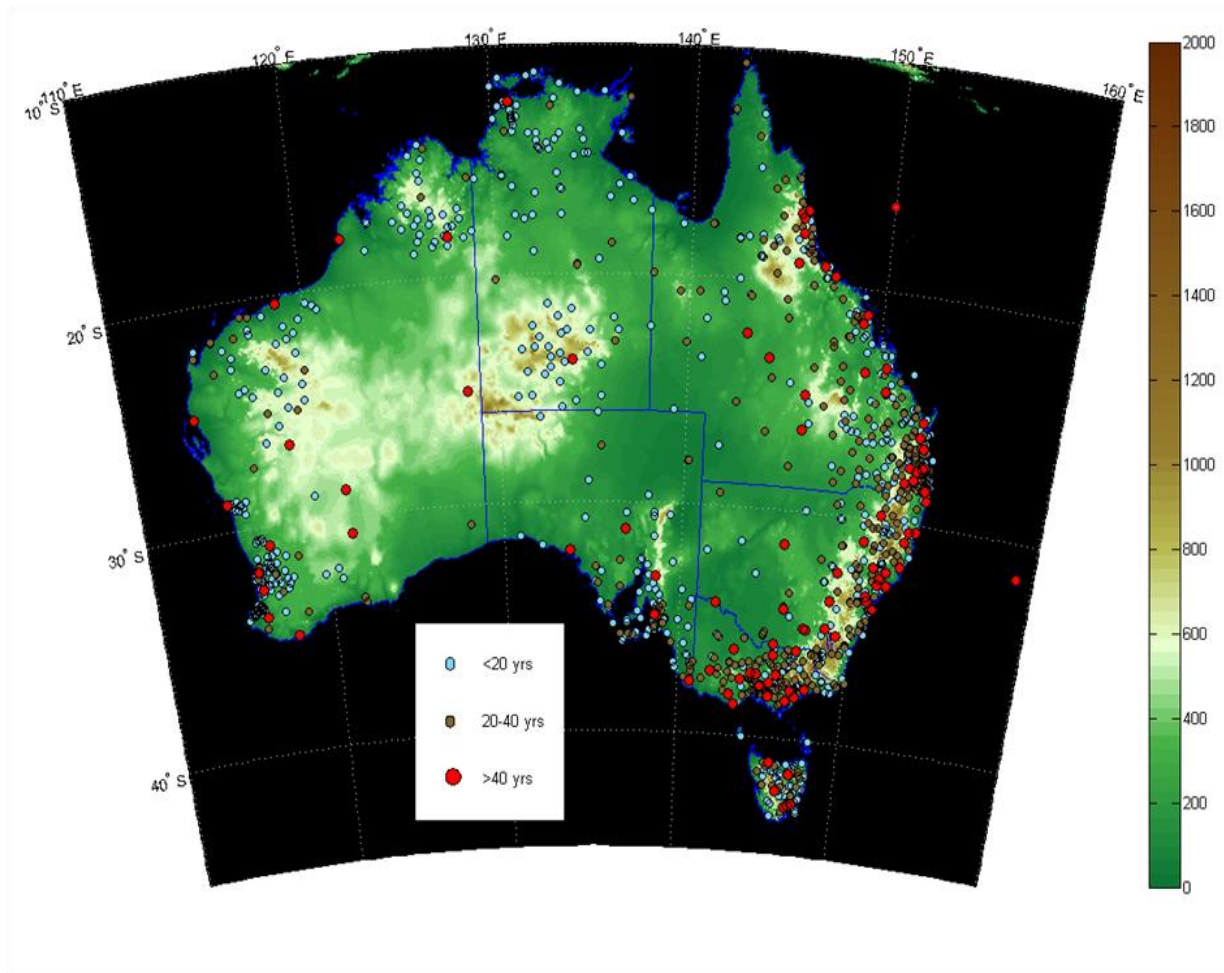


Figure 3.1: Spatial coverage and record length of the Australian sub-daily pluviograph record. Figure extracted from [Westra et al., 2012].

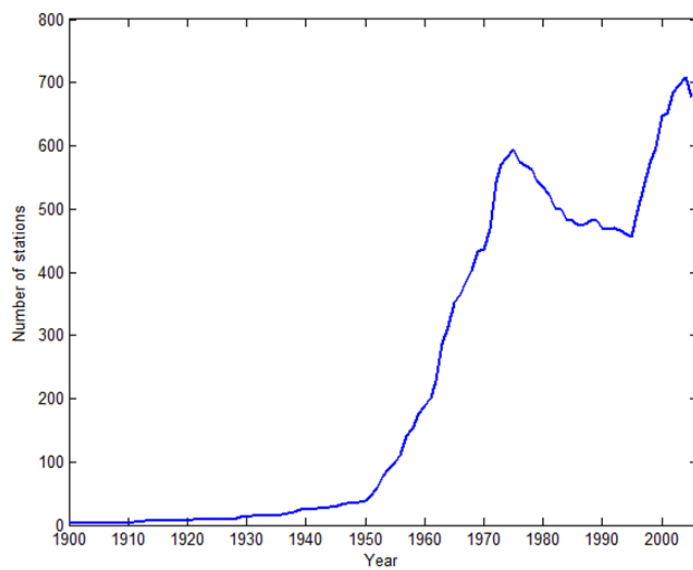


Figure 3.2: Number of Australia-wide pluviograph records against year of record, plotted from 1900. Figure extracted from [Westra et al., 2012].

3.3. Methodology

3.3.1. Regionalised state-based method of fragments algorithm

The daily to sub-daily rainfall class based disaggregation approach using at-site pluviograph data was described at length in Chapter 2. In this approach, we start by considering daily rainfall amount on day t , R_t^o ($t = 1, \dots, 365/6$ representing the calendar day of the year) together with previous and next day wetness state $I(R_{t-1}^o)$, $I(R_{t+1}^o)$, with I representing the indicator function ($I(R)=1$ for a wet day and 0 for a dry day). The disaggregation involves firstly identifying all the wet days within a defined moving window of t (we use a moving window of ± 15 days to ensure seasonal effects are correctly preserved), with the same previous and next day wetness state (i.e. $I(R_{i-1}^o) = I(R_{t-1}^o)$ and $I(R_{i+1}^o) = I(R_{t+1}^o)$), with subscript i representing an arbitrary day of the record whereas t represents the day for which the sub-daily fragment is sought. We now refer to R_j^o with subscript j as the rain days which meet these criteria. We then sort these stations by absolute deviation in rainfall amount ($|R_j^o - R_t^o|$) and select the k lowest ranked rain days $R_{(1)}^o, \dots, R_{(k)}^o$ with the use of parentheses in $R_{(j)}^o$, $j = 1, \dots, k$ indicating the data has been sorted. Sub-daily fragments then can be computed as $f_{(j),m}^o = X_{(j),m}^o / R_{(j)}^o$ for each of these k nearest neighbours. During re-sampling, any single fragment $f_{(j),m}^o$ is selected with probability $p(j)$, (with number of nearest neighbours given as $k = \sqrt{n}$ [Upmanu Lall and Sharma, 1996] where n represents the sample size of the class members falling within the moving window, and equation 2.2 of Chapter 2 to calculate $p(j)$ for a given k), and then the stochastically generated sub-daily rainfall series for day t can be calculated as $X_{t,m}^o = R_t^o \times f_{(j),m}^o$. Here we have added the superscript o to the above notation to emphasise that all fragments are derived from the same station.

We now describe a regionalised version of this disaggregation logic, in which we assume that we have daily data R_t^o available at the target location, but where sub-daily pluviograph data is either unavailable or insufficiently long for continuous simulation purposes. The extension is based on sampling sub-daily rainfall fragments $X_{(j),m}^s$ at neighbouring sites indexed by $s = 1, \dots, S$, where S represents the total number of pluviograph stations within the 'neighbourhood' of the target station. The methodology is identical to the disaggregation model described above using only at-site data, with the exception of substituting the nearby pluviograph records for at-site records. The regionalised methodology is summarised in the following algorithm:

Algorithm 3.1

- (1) For each rain day R_t^o at the target location, identify a moving window about t and find all rain days in each of the S 'nearest' locations to the target station with the same previous and next day wetness state ($I(R_{i-1}^s) = I(R_{t-1}^o)$ and $I(R_{i+1}^s) = I(R_{t+1}^o)$).
- (2) Rank each R_j^s by the absolute deviation from R_t^o , given by $|R_j^s - R_t^o|$. Selecting the k rain days with the lowest ranked deviations, we form the sequence $R_{(1)}^s, \dots, R_{(k)}^s$. For each $R_{(j)}^s$ compute the associate sub-daily rainfall fragment $f_{(j),m}^s = X_{(j),m}^s / R_{(j)}^s$.

- (3) Select a fragment, $f_{(j),m}^s$, from the fragments at neighbouring locations selected with probability $p(j)$ as described in Chapter 2 and compute the sub-daily rainfall via the relationship $X_{t,m}^o = R_t^o * f_{(j),m}^s$.

Although this approach is conceptually simple, the challenge is to identify the neighbourhood from which to sample the S pluviograph records. To achieve this, it is necessary to assume that the scaling between daily rainfall amount and the sub-daily rainfall fragments is consistent across the neighbourhood and thus the nearby sub-daily fragments can become substitutable for at-site sub-daily fragments. The basis for identifying whether the daily- to sub-daily scaling at two locations is similar and thus substitutable is described below.

3.3.2. Daily to sub-daily scaling

To enable substitution of sub-daily fragments, one needs to assume that for any day t , the conditional relationship between the daily rainfall amount R_t and the full sequence of sub-daily rainfall $X_{t,m}$ (with $m = 1, \dots, 240$ for six-minute duration rainfall) are statistically similar at both the target station and the nearby stations. This can be expressed as:

$$f(X_{t,m}^s | R_t^s) = f(X_{t,m}^o | R_t^o) \quad (3.1)$$

for all m and t , where $f(\cdot|\cdot)$ is used to express a conditional probability distribution. Given the difficulty of constructing separate conditional density functions for 240 separate increments of sub-daily rainfall, as well as the fact that for any wet day R_t there is a high probability that any sub-daily rainfall increment $X_{t,m}$ has no rainfall, we modify Equation 3.1 as follows:

$$f(Y_t^s | R_t^s) = f(Y_t^o | R_t^o) \quad (3.2)$$

where Y_t^s and Y_t^o represent scalar attributes of $X_{t,m}^s$ and $X_{t,m}^o$ for each day of record, respectively. The attributes to be considered include:

- Maximum intensity attributes: for each wet day, what is the maximum 6, 12, 30, 60, 120, 180 and 360-minute duration storm burst expressed as a fraction of the total rainfall amount for that day?
- Fraction zeros: for each day, what is the fraction of 6-minute time steps with no rainfall?
- Maximum intensity timing: for each wet day, what is the time of day when the maximum 6, 12, 30, 60, 120, 180 and 360-minute duration storm burst occurs?

In combination, these scalar attributes are expected to cover most of the information on the scaling and timing behaviour between daily rainfall and the fragments.

To illustrate these concepts, we present in Figure 3.3 the joint probability plot of daily rainfall and the maximum 12-minute storm burst at three locations in Australia: Hobart, Sydney and Darwin. These locations were selected as they have distinctly different climatology, with Hobart located in the south of Tasmania representing one of the most southerly pluviograph records, Darwin in the Northern Territory representing one of the most northerly pluviograph records, and Sydney

representing intermediate latitudes.

As can be seen in the daily rainfall histogram (lower panel), the marginal probabilities of daily rainfall at each station are distinctly different, with Darwin having a high probability of high daily rainfall amounts (the majority of rain days having well over 10mm rainfall) whereas Hobart has a large number of rain days with relatively little rainfall, with most days having significantly less than 10mm over the entire day. It should be emphasised, however, that our interest here is not on this marginal distribution; rather, we wish to know, conditional to some daily rainfall amount, whether the sub-daily rainfall properties are the same at any two locations. To determine whether this is the case, we started by plotting a loess smoother (support of 0.25 of the sample) [Hastie *et al.*, 2009] to represent the average value of the maximum 12-minute storm burst as a function of daily rainfall.

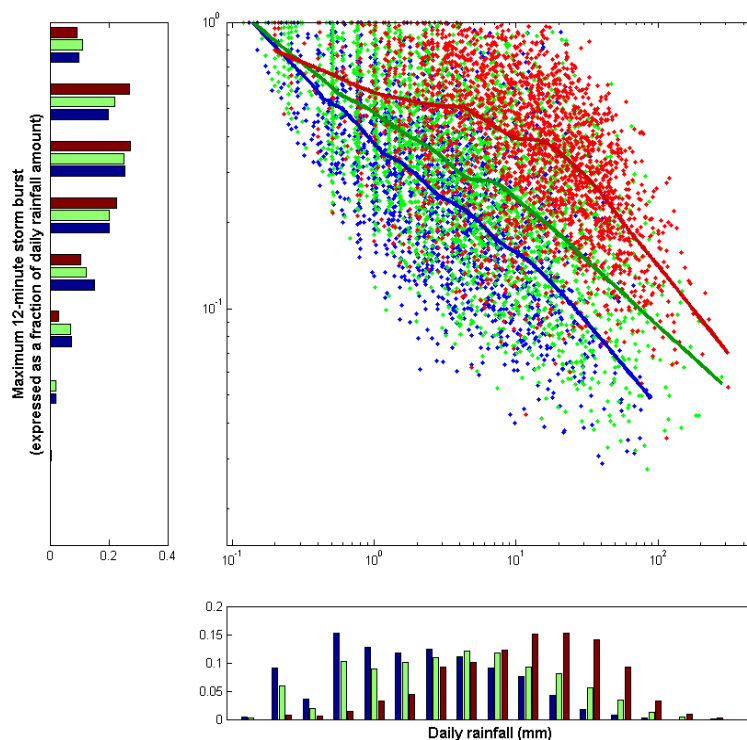


Figure 3.3: Scatter plot with daily rainfall and an attribute of sub-daily rainfall (the maximum 12-minute storm burst expressed as a fraction of the total daily rainfall) at three locations in Australia: Hobart (blue), Sydney (green) and Darwin (red). Histograms of daily rainfall and the maximum 12-minute storm burst are provided in the bottom and left figure panels, respectively, for each of the three locations. Figure extracted from [Westra *et al.*, 2012].

It is evident that the fraction of daily rainfall contained in the maximum 12-minute storm burst varies as a function of daily rainfall amount. This is unsurprising, as intuitively one would expect that for small daily rainfall amounts a smaller percentage of the day would be wet, and therefore there is a greater chance that the maximum 12-minute storm burst contains a large portion of

the daily rainfall. Interestingly, however, the loess smoother highlights that the relationship between daily rainfall and sub-daily rainfall is on average very different at the three locations, with Darwin typically having a greater fraction of the daily rainfall contained within the maximum 12-minute storm burst than Hobart. This suggests that even if both stations have the same daily rainfall amount, Darwin is more likely to have a larger number of short-duration, high-intensity rainfall events compared with Hobart, which appears sensible given the tropical nature of the Darwin climate. Although figures are not provided here, consistent conclusions can be drawn from considering other durations, as well as the fraction of each wet day that does not experience rainfall.

3.3.3. Defining similarity

We now wish to devise a metric to determine whether the conditional distributions in Equation 3.2 and illustrated in Figure 3.3 are in fact statistically equivalent. To simplify the analysis, rather than focus on the conditional distribution we consider whether the joint distribution of Y and R at any two stations is equivalent, given by:

$$f(Y^s, R^s) = f(Y^o, R^o) \quad (3.3)$$

This is a stricter criterion compared to the conditional distribution in Equation 3.2, since two locations having equivalent joint distributions imply that the conditional distribution must also be equivalent, although the opposite is not necessarily true (one can easily imagine two samples having an equivalent distribution of sub-daily rainfall conditional to daily rainfall amount, but different marginal distribution for the daily rainfall amount, and therefore different joint distributions).

To test the hypothesis that the joint distribution between daily rainfall and some attribute of sub-daily rainfall at any two locations are statistically similar, we use a two dimensional, two sample Kolmogorov-Smirnov (K-S) test. This represents a generalisation of the better known one-dimensional K-S test [Press *et al.*, 1992], and was developed by Fasano and Franceschini [1987]. The basis of the two-dimensional generalisation is that although a cumulative distribution function is not well defined over more than one dimension, the integrated probability in each of four quadrants around some point (x_i, y_i) in some arbitrary x and y dimensions provides a reasonable approximation. The two-dimensional K-S statistic D is the maximum difference (ranging over both data points and quadrants) of the integrated probabilities, and is given by [Press *et al.*, 1992]:

$$Probability(D > observed) = Q_{KS} \left(\frac{\sqrt{ND}}{1 + \sqrt{1 - r^2 \left(0.25 - \frac{0.75}{\sqrt{N}} \right)}} \right)$$

where

$$N = \frac{N_1 N_2}{N_1 + N_2}$$

with N_1 and N_2 representing the size of samples 1 and 2, respectively. In calculating the probability that the K-S statistic is above some defined level under the null hypothesis that the two samples are from the same population, it is necessary to evaluate the function:

$$Q_{KS}(\lambda) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 \lambda^2}$$

As an alternative to the K-S test, we also used a chi-square test which involves determining the probability that two different binned distributions are statistically similar. In many ways this approach is less attractive to the K-S test, as binning continuous data involves loss of information, and the selection of optimal bin sizes is generally a difficult problem for all but the simplest situations [e.g. *Scott, 1992*]. Nevertheless, given the importance of the evaluation of whether two samples are similar to the methodology developed here, the use of an alternative approach adds confidence to the conclusions. In summary, the chi-square statistic for two samples is given as:

$$\chi^2 = \sum_i \frac{(R_i - S_i)^2}{R_i + S_i}$$

where R_i and S_i are the number of occurrences in bin i for the first and second samples (i.e. rainfall data at any two stations).

An important parameter to be determined in the application of the chi-square method is the histogram bin width, h , with small h resulting in histograms with high variance (i.e. very rough or 'bumpy' histograms that follow the individual data points too closely), whereas high values of h result in high bias and miss important underlying features of the probability density function. For this reason, bin width is often referred to as a smoothing parameter, since it determines the amount of smoothing which is applied to the resulting histogram. To identify an optimal bin width, denoted as h^* , an approach presented in *Scott [1992]* for the two dimensional (bivariate) case in which there is potentially some correlation between the individual dimensions is given as:

$$h_j^* = 3.504 \sigma_j (1 - \rho^2)^{3/8} n^{-1/4}$$

where the subscript j represents the dimension (in our case one dimension being daily rainfall and the other being one of the sub-daily rainfall attributes), σ represents sample variance, n represents the sample size and ρ represents the correlation coefficient. As the chi-square test is used to compare two bivariate distributions, we calculate optimal bin width for the smallest sample and use this bin placement for both samples.

3.3.4. Predictive model for statistical similarity

In the previous sections we have developed a metric for determining whether the joint distribution between daily rainfall amounts and attributes of sub-daily rainfall are similar. As discussed earlier, to use this information to extend the continuous simulation approach to locations where pluviograph data is unavailable, it is necessary to draw sub-daily fragments from nearby stations conditional to daily rainfall at the target location. As such, we now wish to

determine: what factors influence whether the daily to sub-daily scaling at two stations will be similar?

To answer this, we consider each possible bivariate combination of the 232 pluviograph stations with at least 30 years of data, totalling 26796 station pairs, and calculate the two sample, two dimensional K-S statistic as well as the chi-squared statistic for each pair of stations and for each of the sub-daily rainfall attributes. We use a 5% significance level to evaluate whether two stations are similar, and then consider how the probability that any two stations are similar varies as a function of a range of possible covariates, including difference in latitude, longitude, distance to coast and elevation between each station pair. These predictors, summarised in Table 3.1, comprise a range of easily measurable physiographic characteristics which might be expected to determine the similarity between two stations. Seasonal variations in the daily to sub-daily rainfall relationship are accommodated by formulating the basis for identifying similar sub-daily stations with reference to the season of the year.

Table 3.1: Predictors used for the logistic regression model described in equation 1. The prefix ‘Diff_’ emphasises that it is the difference in each of the predictors between stations that is considered, rather than the absolute value.

Predictor	Units	Description / comments
Diff_lat	Degrees (expressed as a decimal)	Difference in latitude between each station pair, calculated as $\text{abs}(\text{Lat}_1 - \text{Lat}_2)$
Diff_lon	Degrees (expressed as a decimal)	Difference in longitude between each station pair, calculated as $\text{abs}(\text{Lon}_1 - \text{Lon}_2)$
Diff_lat Diff_lon *	Degrees (expressed as a decimal)	Interaction term, which would be greater than zero if it is the distance between stations, rather than the sum of the latitude and longitude, which is the dominant predictor.
Diff_dist_coast (normalised)	Dimensionless	Difference in distance to coast between each station pair, normalised by the average distance to coast for the station pair, calculated as $\text{abs}(\text{dist}_1 - \text{dist}_2) / \text{mean}(\text{dist}_1, \text{dist}_2)$.
Diff_elev	Metres	Difference in elevation between each station pair, calculated as $\text{abs}(\text{Elev}_1 - \text{Elev}_2)$

Thus we have a set of continuous predictors represented by \mathbf{V} (dimension 26796 x 5) which we wish to model against a binomial response represented by \mathbf{u} of length 26796 (where $u \in \{0, 1\}$)

represents the cases where the scaling between daily- and sub-daily rainfall at two stations are statistically different and similar, respectively, as calculated by the Kolmogorov-Smirnov test described in the previous section). This relationship can be modelled using a logistic regression,

in which:

$$\Pr(\mathbf{u} = \mathbf{1}) = \text{logit}(\mathbf{z}) = \frac{e^{\mathbf{z}}}{e^{\mathbf{z}} + 1} \quad (3.4)$$

transforms the continuous predictor variables to the range [0,1] as required when modelling a binomial response. In this equation, \mathbf{z} is defined as:

$$\mathbf{z} = \beta_0 + \beta_1 \mathbf{v}_1 + \dots + \beta_5 \mathbf{v}_5 \quad (3.5)$$

with β representing the regression coefficients.

The results of the logistic regression model of Equation 3.4 are shown in Figure 3.4, plotted against the difference in latitude covariate. The results are presented for four attributes of sub-daily rainfall: 6 minute maximum storm burst, 1 hr maximum storm burst, fraction of day with no rainfall and time of day with the maximum 6-minute storm burst. A range of other attributes, listed in Section 3.3.2 were also examined, however the four attributes shown in Figure 3.4 were found to be representative of those other variables and are therefore the focus of subsequent analysis. The results are presented using both the K-S statistic (solid lines), and the chi-square statistic (dashed lines). Note that for the time attribute, we are only considering the marginal distribution of the time of day when the maximum 6-minute storm burst occurs, rather than a joint density.

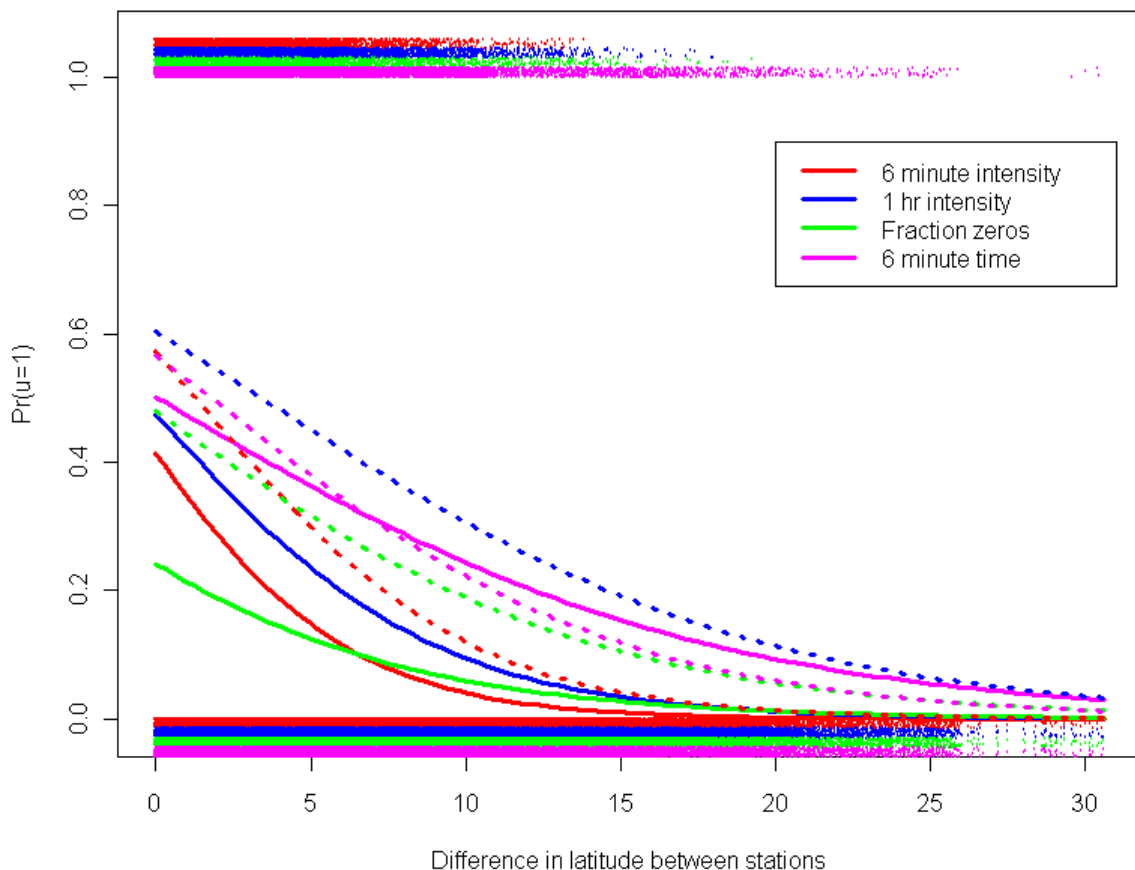


Figure 3.4: Logistic regression results against a single predictor – difference in latitude,

and four responses representing different sub-daily attributes. The responses have been calculated using both the Kolmogorov-Smirnov (solid lines) and Chi squared test statistics (dashed lines). The upper and lower part of the plot presents a scatter of the individual outcomes of the K-S statistics (with the upper part representing an outcome of '1' – i.e. the two stations are similar – while the lower part represents an outcome of '0'), and shows that a larger number of instances of '1' can be found for smaller difference in latitudes compared to instances of '0'. Figure extracted from [Westra et al., 2012].

Some interesting conclusions can be derived from Figure 3.4. Firstly, with the exception of the fraction zeros measured by the K-S statistic, there is a chance between 40% and 60% that the joint distribution of daily rainfall and each of the attributes are statistically similar provided the difference in latitude is small, with the probability decreasing rapidly as difference in latitude increases. This is interesting, as no account is made of any other physiographic information, such that longitude, elevation and distance to coast, such that stations may be located in opposite sides of the continent, or at very different elevations, and yet still have close to a 50% chance of having the same scaling between daily and sub-daily rainfall provided the latitude is the same. Secondly, the K-S statistic and chi-square statistic appear to be showing similar results, although the K-S statistic shows a slightly lower probability of two stations being equal. In the remainder of the section we will focus on the K-S statistic as this represents the more conservative metric. Finally, of all the metrics considered, the fraction of the day with no rainfall appears to vary most significantly between stations.

Consideration of just a single covariate – difference in latitude – as the only factor influencing similarity between station ignores other physiographic information which may be important. As such we develop a multivariate logistic regression model to consider the influence each of the plausible predictors mentioned above. The conceptual basis for this approach is illustrated in Figure 3.5. Given a target location of interest, we wish to define a zone for which the probability that daily to sub-daily scaling at two stations are statistically similar is greater than a pre-defined threshold. This zone is described by contours of equal probability, with the probability decreasing linearly (in the logistic transformed space) in each of the dimensions of the regression model. The shape of the contours is defined by the logistic regression coefficients. In the idealised example in Figure 3.5, we represent the case where the probability of two stations being statistically similar decreases at a faster rate in the latitude dimension compared to the longitude dimension. The ellipsoid shape of the contours are governed by the interaction term (latitude*longitude). Furthermore, the location of the target station is slightly offset from the centre of the contours, with this being governed by the influence of the relative difference in distance to coast.

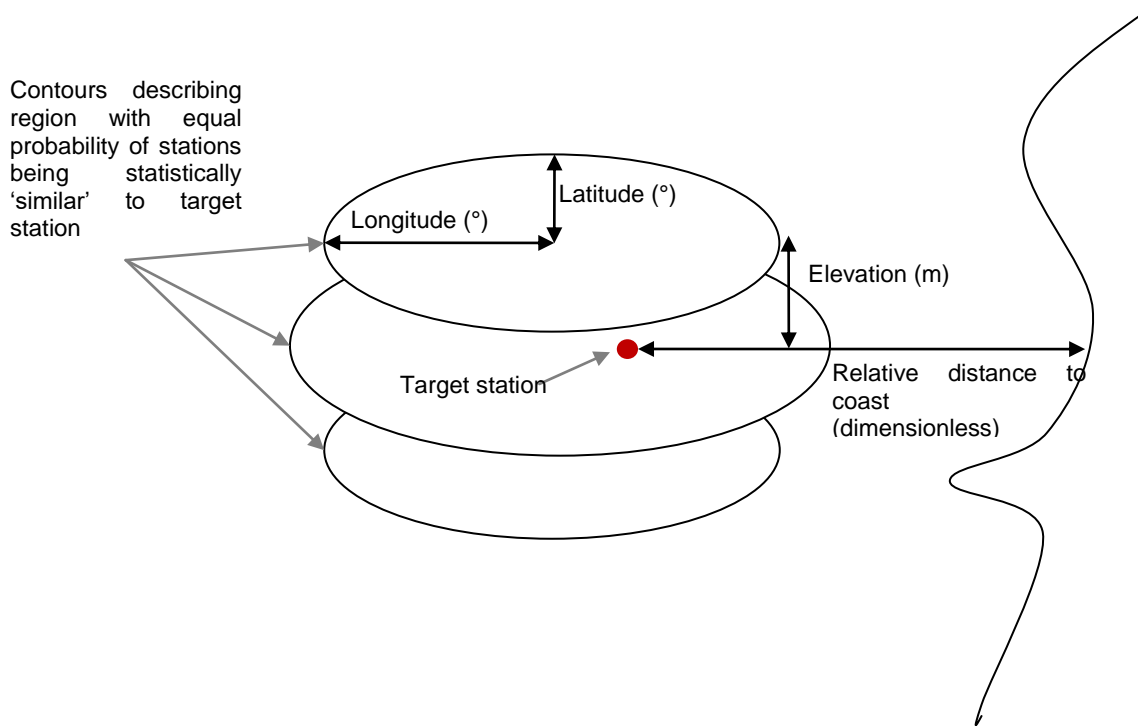


Figure 3.5: Diagrammatic representation of logistic regression results. The response is the probability that the joint distribution of daily rainfall amount and some attribute of sub-daily rainfall at a 'nearby' station is statistically similar to the target station. The predictors are the difference in latitude, longitude, latitude*longitude, elevation and a normalised distance to coast, with the logistic regression coefficients determining the relative decrease in the probability that two stations are similar in each of these dimensions. Figure extracted from [Westra et al., 2012].

The results of this multivariate regression are presented in Table 3.2, and once again plotted for the summer months against latitude in Figure 3.6, with the remaining predictors held at zero. As can be seen, the results in Figure 3.6 show notable improvements in the probability that two stations are equal compared to Figure 3.4, since now we are plotting the influence of latitude assuming that difference in longitude, elevation and relative distance to coast are all zero. In fact, with the exception of the fraction of zeros, the results show that for small values of each of the predictors there is between a 60 and 70% probability that the daily to sub-daily joint probability distributions are statistically similar.

Table 3.2: Logistic regression coefficients. All predictors were found to be statistically significant (usually with a p-value < 0.001 level), with the exception of several predictors labelled as 'NS' (not significant)

Season	Sub-daily rainfall attribute	Logistic regression coefficients					
		Intercept	Lat	Lon	Lat*lon	Dist_coast	Elev
DJF	6 minute intensity	0.426	-0.345	-0.0377	0.0064	-0.186	-0.00089
DJF	1 hour intensity	0.823	-0.333	-0.0425	0.0093	-0.231	-0.00075
DJF	Fraction zeros	-0.375	-0.253	-0.0318	0.0075	-0.242	-0.00065

DJF	6 minute time	0.979	-0.137	-0.0099	0.0022	-0.453	-0.00141
MAM	6 minute intensity	-0.067	-0.192	-0.0065	NS	-0.218	-0.00130
MAM	1 hour intensity	0.308	-0.178	-0.0074	NS	-0.107	-0.00098
MAM	Fraction zeros	-0.806	-0.157	-0.0105	0.0025	-0.165	-0.00060
MAM	6 minute time	1.256	-0.140	-0.0226	-0.0034	-0.227	-0.00092
JJA	6 minute intensity	-0.197	-0.097	-0.0110	0.0034	-0.096	-0.00198
JJA	1 hour intensity	0.471	-0.102	-0.0204	0.0033	NS	-0.00335
JJA	Fraction zeros	-0.365	-0.073	-0.0171	0.0031	-0.101	-0.00116
JJA	6 minute time	2.078	-0.098	-0.0321	0.0037	-0.156	-0.00069
SON	6 minute intensity	0.474	-0.387	-0.0722	0.0129	NS	-0.00146
SON	1 hour intensity	0.824	-0.325	-0.0835	0.0135	NS	-0.00132
SON	Fraction zeros	-0.382	-0.239	-0.0623	0.0104	-0.087	-0.00095
SON	6 minute time	1.028	-0.162	-0.0287	0.0042	-0.317	NS

It should be re-emphasised that this is in many ways a conservative estimate, as we have chosen to display the test statistic (the K-S statistic) which applies the harshest criterion to the data, and we consider the sub-daily attributes (e.g. fraction of zeros, 6 minute rainfall intensity) which are the most challenging to capture from daily data alone. Even more importantly, as can be seen in Figure 3.3, the number of samples in each bivariate distribution is large (30 years of data, 90 days per season, and about 30% of days being 'wet days' yields approximately 800 wet days) such that the 95% confidence intervals are very narrow (as the width of the confidence intervals is to a large degree governed by sample size).

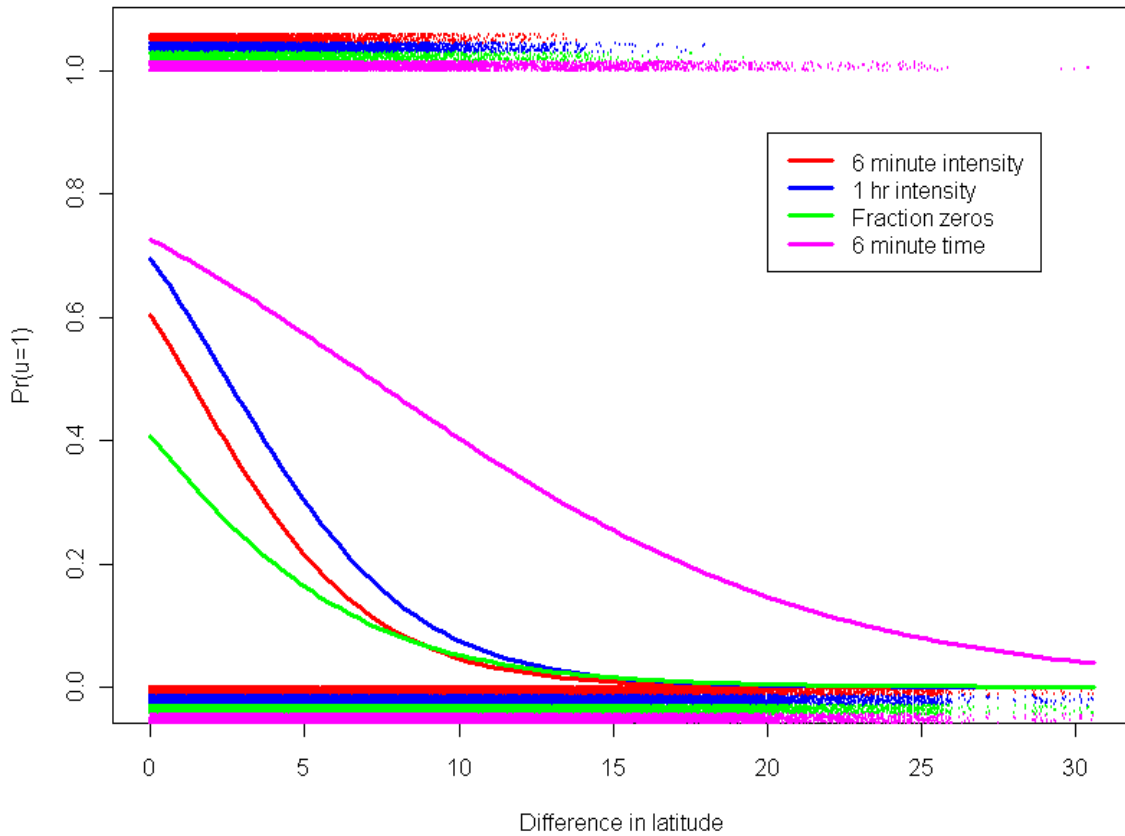


Figure 3.6: As per Figure 3.4, except the results represent the outcomes of the full multivariate regression. The probability that daily to sub-daily scaling is statistically similar is once again plotted against difference in latitude, however now all the remaining predictors are held at zero. Figure extracted from [Westra *et al.*, 2012].

3.4. Results

3.4.1. Identifying 'nearby' stations - application to Sydney Airport

We start by demonstrating a single application of the approach at one location: Sydney Airport (gauge number 066037). This location represents a relatively long-record pluviograph station, and therefore provides a useful record for verification of the method.

The approach to identifying 'nearby' stations is as follows:

- (1) For all the 1396 pluviograph stations in Australia (excluding the Sydney Airport gauge), calculate each of the regression predictors identified in Table 3.1; namely, difference in latitude, longitude, latitude*longitude, elevation and normalised distance to coast, relative to the Sydney Airport station;
- (2) Having developed the 1396 x 5 predictor matrix, apply the regression model presented in Equation 3.4 using the regression coefficients shown in Table 3.2 for each season and attribute to calculate the probability $\Pr(u=1)$;
- (3) Separately for each season and attribute, rank the probabilities from lowest to highest;
- (4) For each season calculate the average rank for each station across all attributes;
- (5) Select the S lowest-ranked stations for inclusion in the disaggregation model.

This algorithm yields different choices of stations for each season, as physiographic influences may vary depending on the dominant synoptic systems occurring and different times of the year. It is noted that the selection of the size of S represents a somewhat subjective decision, as larger values of S increase the probability of selecting stations which are statistically different to the target station, whereas smaller values of S will result in small sample sizes. For this case we selected $S = 13$, resulting in a total of 250 years of data distributed over the 13 stations.

These lowest-ranked 13 stations for the summer season are shown in Figure 3.7. As expected, the lowest ranked stations (i.e. those with the greatest chance of being 'similar' to Sydney Airport) are those which are most proximate to this station. Investigation of the locations of the selected stations suggests that they are generally within a small distance to coast, and all are at low coastal elevations. In this case, therefore, the stations appear to be selected over a wide range of latitudes, which is probably due to the strong increases in elevation and distance to coast with changing longitude.

3.4.2. Model validation

We now repeat the process of identifying nearby stations for five locations across Australia that have more than 50 years of pluviograph data, and which represent a diversity of climate zones. These stations are shown in Table 3.3. Having identified the pool of nearby stations from which to draw the fragments, we apply the approach described in Algorithm 3.1 to draw sub-daily rainfall fragments from nearby stations conditional to at-site daily rainfall, and compare these

sequences to the at-site pluviograph records.

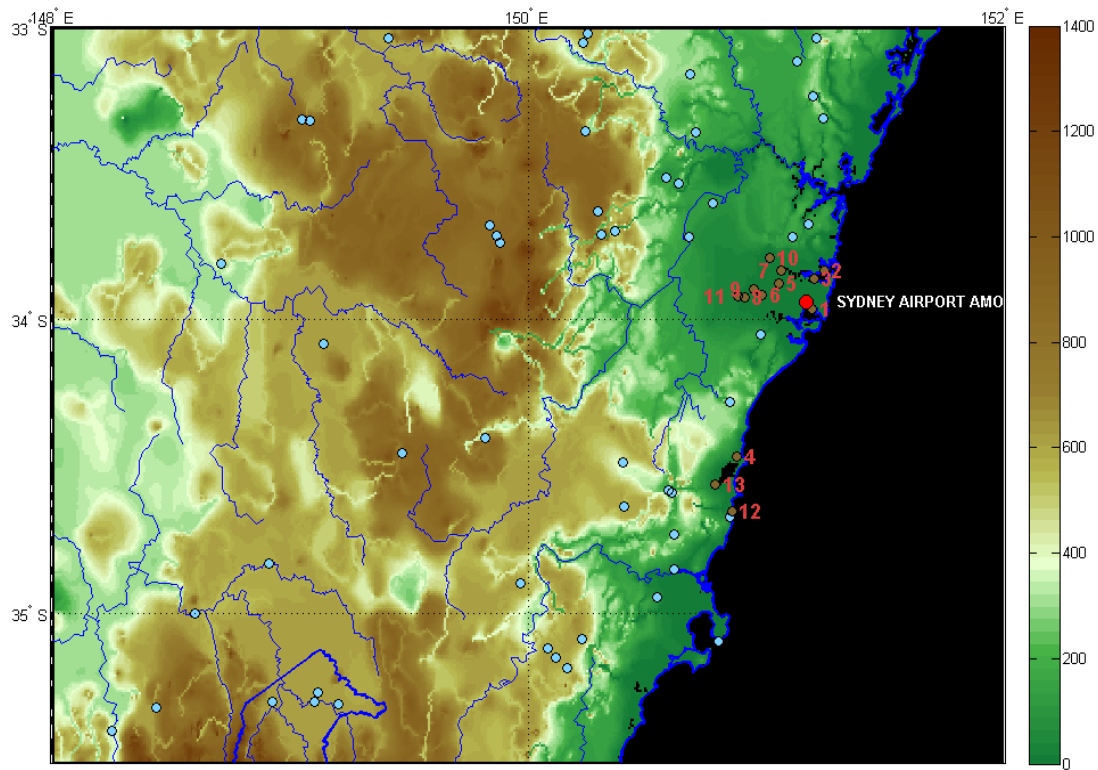


Figure 3.7: Sydney Airport (large red dot) and nearby pluviograph stations (blue and brown dots). The highest ranked 13 pluviograph stations (totalling approximately 250 years of pluviograph data) based the full logistic regression model are shown as brown dots, with the associated ranking. Figure extracted from [Westra et al., 2012].

Table 3.3: Data used to test continuous simulation model. All stations continue until 2007.

Station Name	Gauge number	Start year	Number of years of observed data	Latitude / longitude	Köppen climate classification
Sydney airport	066037	1961	45	-33.9411 / 151.1725	Temperate (warm summer)
Perth airport	009021	1960	46	-31.9275 / 115.9764	Sub-tropical (dry summer)
Alice Springs airport	015590	1950	57	-23.7951 / 133.8890	Desert/Grassland (hot, persistently dry)
Cairns airport	031011	1941	66	-16.8736 / 145.7458	Tropical (monsoonal)
Hobart airport	094008	1959	47	-42.8339 / 147.5033	Temperate (mild summer)

As with the results presented in Chapter 2, the use of a disaggregation model derived based on observed daily rainfall sequences implies that the daily and longer time scale statistics will be identical to the observational dataset. As such, it is necessary to test the capacity of the model using a range of sub-daily rainfall characteristics. Reflecting the likely application of this model for flood estimation, the statistics considered here are based on: (a) whether the model is capable of reproducing the extreme rainfall intensity correctly; and (b) whether the model captures the antecedent rainfall prior to the flood-producing rainfall event.

Considering first the annual maxima statistics, we present in Figure 3.8 a plot of the annual maximum 6-minute rainfall against exceedance probability for both the observed data at the target location, as well as the results of 100 simulation runs with the same length of series as the original target pluviograph time series to make for easier comparison. As can be seen, the observed data is generally within the 90 percent confidence interval for most of the stations, with the exception of Alice Springs, for which the generated sequences tend to overestimate rainfall for all exceedance probabilities, and for Hobart in which the simulated sequences underestimate the low exceedance probability rainfall events. Possible reasons for this behaviour are provided in section 3.5 below. These results are also presented in the upper half of Table 3.4 for a range of storm burst durations from 6-minute through to 12-hour. Once again the observed and simulated sequences are generally similar, with no systematic under- or over-estimation biases.

We next consider the antecedent rainfall prior to the design storm burst event, plotted in Figure 3.9. The justification for focusing on the antecedent rainfall exceedance probability plot was described at length in Chapter 2. As can be seen, the simulated data appear to follow the observed data reasonably well, although there are several points outside the 90% confidence interval. Importantly, no systematic biases could be identified, with performance varying depending on the location. This is also shown in the lower half of Table 3.4 with the antecedent rainfall of different durations prior to the 1-hour storm burst. Once again the observed antecedent rainfall is within the 90% confidence interval, with the exception of Cairns in which antecedent rainfall is underestimated for 6-hour depth prior to the 1- hour storm burst, and overestimation for longer durations.

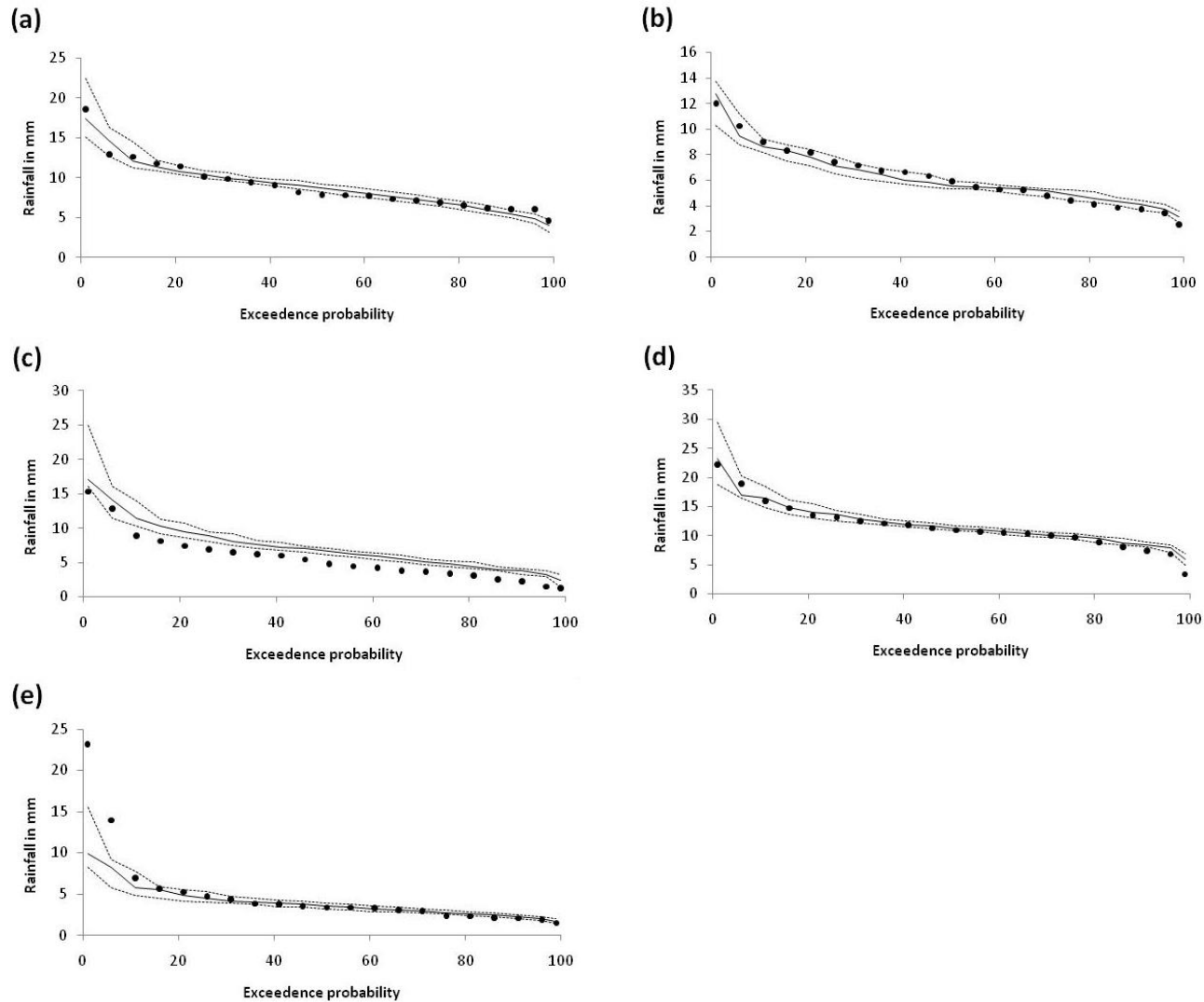


Figure 3.8: 6-minute annual maximum rainfall against exceedance probability for (a) Sydney, (b) Perth, (c) Alice Springs, (d) Cairns, and (e) Hobart. Black dots represents observed data, black solid line represents the median of 100 simulations, and black dotted lines represent the 5

and 95 percentile simulated values. Figure extracted from [Westra et al., 2012].

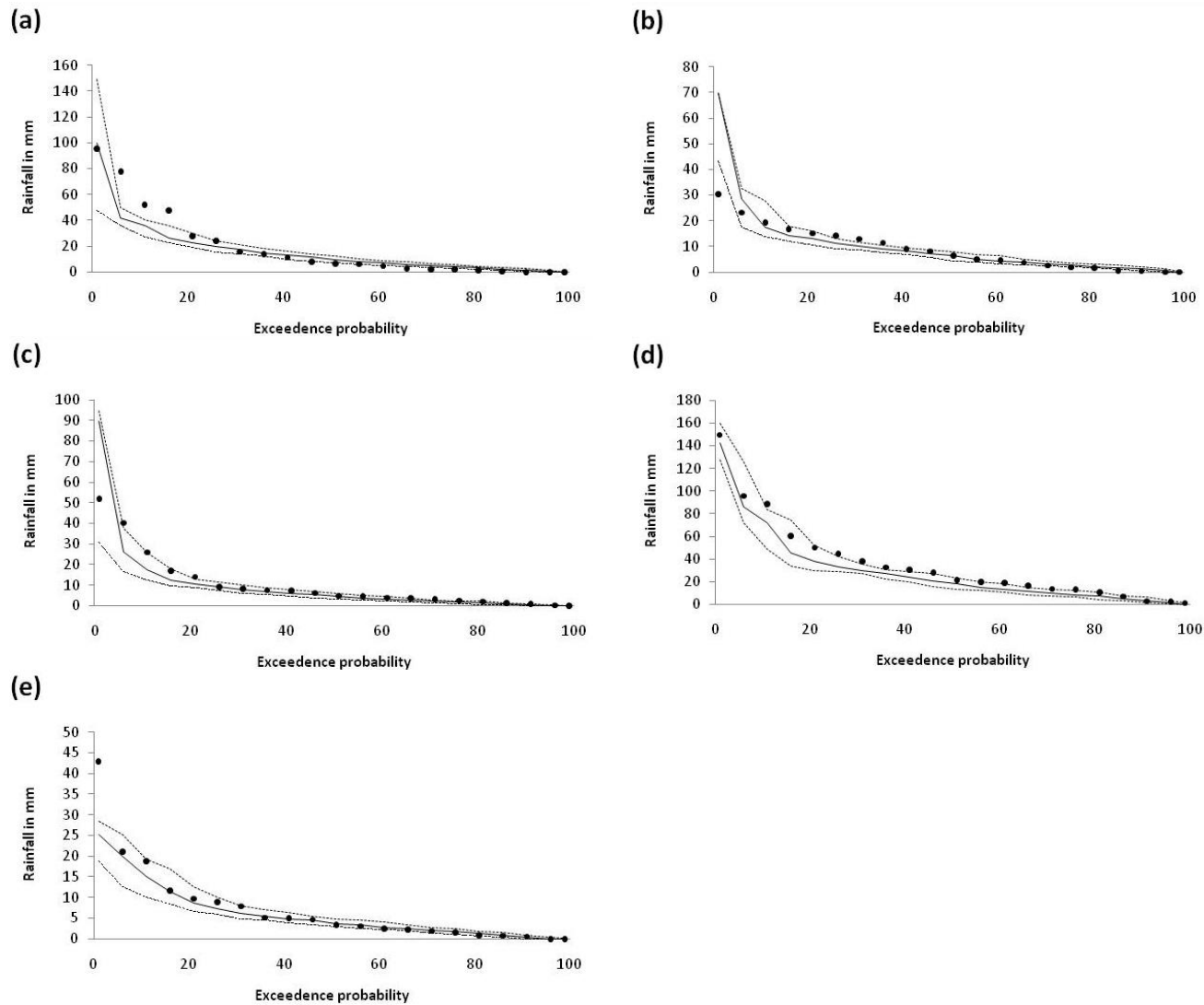


Figure 3.9: 6-hour antecedent rainfall prior to the 6-minute annual maximum storm burst plotted against exceedance probability for (a) Sydney,

(b) Perth, (c) Alice Springs, (d) Cairns, and (e) Hobart. Black dots represents observed data, black solid line represents the median of 100 simulations, and black dotted lines represent the 5 and 95 percentile simulated values. Figure extracted from [Westra *et al.*, 2012].

Table 3.4: Comparison of observed and simulated results for median annual maxima for different storm burst durations, and the antecedent rainfall prior to the 1 hour storm burst. The simulated median annual maxima represent the median of all 100 simulations.

	Sydney		Perth		Alice Springs		Cairns		Hobart	
	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)
Annual maxima										
6 min	8.87	8.95 (8.63-9.38)	6.18	6.16 (5.80-6.51)	5.50	7.21 (6.79-7.66)	11.6	11.9 (11.4-12.4)	4.51	3.97 (3.54-4.32)
30 min	25.7	23.6 (22.5-24.8)	14.7	13.7 (13.2-14.4)	16.7	18.9 (18.1-19.9)	34.9	36.6 (35.2-38.0)	11.3	9.15 (8.57-9.70)
1 hr	35.4	31.7 (30.3-33.7)	18.8	18.4 (17.6-19.2)	22.1	24.4 (23.0-25.8)	51.7	54.8 (52.2-57.5)	14.6	12.3 (11.6-13.0)
3 hr	55.4	49.0 (47.0-51.6)	29.0	27.8 (26.7-28.7)	32.6	32.6 (31.2-34.2)	83.5	88.3 (85.4-91.5)	22.9	20.4 (19.6-21.3)
6 hr	72.3	63.2 (61.6-65.8)	36.3	34.7 (33.9-35.8)	39.6	39.0 (37.5-40.7)	113	118 (114-121)	30.3	28.1 (27.1-29.3)
12 hr	91.8	82.0 (80.3-84.4)	45.4	44.1 (43.2-45.1)	48.2	47.6 (47.0-48.2)	147	151 (147-153)	39.6	36.8 (36.0-37.7)
Antecedent rainfall prior to 1-hr burst (mm)										
6 hr	15.4	13.1 (10.4-16.6)	6.76	7.53 (6.57-8.74)	6.10	6.09 (5.05-7.21)	25.4	20.4 (16.5-24.3)	6.31	5.91 (4.84-7.05)
12 hr	22.7	18.2 (14.8-22.5)	9.63	10.5 (9.26-11.8)	7.98	8.97 (7.58-10.1)	32.2	29.1 (23.9-34.4)	9.10	7.61 (6.01-9.15)
24 hr	31.4	26.8 (22.0-32.5)	11.9	14.0 (12.1-15.8)	10.6	13.1 (10.9-15.4)	40.3	49.8 (42.5-58.4)	9.09	10.1 (8.09-12.1)
48 hr	38.4	35.5 (29.0-43.3)	12.5	18.0 (15.6-20.7)	15.5	18.7 (16.0-22.4)	54.9	79.9 (69.9-90.6)	10.2	13.4 (10.4-16.9)

3.5. Summary

In this chapter, a framework was described where continuous (6-minute increment) rainfall can be generated at any location of interest provided that daily data is either available or can be synthetically generated. The basis of this approach is to randomly draw sub-daily fragments from 'nearby' pluviograph stations conditional to the daily rainfall amount and previous- and next-day wetness state at the target station. The identification of 'nearby' stations is based on a distance metric which considers latitude and longitude as well as elevation and distance to coast, with the relative importance of each variable determined by looking at the similarity in the daily to sub-daily scaling at 232 long pluviograph stations across Australia.

This approach seeks to address several important limitations associated with the Australian pluviograph record. Firstly, compared to daily rainfall data, there is approximately one order of magnitude less pluviograph stations, and the records at each station are generally shorter than their daily-read counterparts. Thus, by combining longer, more abundant and more reliable daily data at the target location with the information contained in a number of pluviograph records in the neighbourhood of the target location, it is possible to make the best use of the both types of data. Secondly, by drawing records from multiple nearby pluviograph records rather than relying on a single record, it is possible to also consider information from records only several years long, which would usually be discarded as being too short for meaningful analysis. This provides a significant advantage over regionalised parametric models described in the introduction to this chapter, as the estimation of parameters generally requires many years of pluviograph data. Finally, pluviograph data flagged as missing or unreliable simply can be discarded from the analysis, even if there is a systematic bias in the missing data (e.g. pluviograph recording tends to fail during major storm events). This is because, provided the daily rainfall data is reliable, and there is sufficient data at other pluviograph stations to capture a diversity of rainfall events across a range of magnitudes, such possible systematic pluviograph recording biases are unlikely to be translated into the final synthetically generated sequences.

The evaluation of the method on a range of statistics which are relevant for flood estimation, notably the annual maximum statistics and the antecedent rainfall prior to the flood-producing storm burst, suggests that the method compares reasonably well with at-site data for the five test locations considered. Nevertheless, the method does require the presence of representative pluviograph gauges in the vicinity of the target location, which may not be possible at every location. For example, poorer performance at Alice Springs compared to other stations is likely to be due to the low density of pluviograph records in the centre of Australia compared to other regions (see Figure 3.1), highlighting the importance of having sufficient pluviograph records within a close vicinity of the location of interest.

A different situation was experienced for the Hobart Airport station. As shown in Figures 3.8 and 3.9, the largest observed value for Hobart both for the annual maximum plots and the antecedent conditions was much greater than any of the values of the synthetically generated data. In particular, the maximum recorded 6-minute storm burst was 23.14mm occurring on the 24th April 1972, comprising a very intense storm burst for that latitude. Aggregating the pluviograph record for that full day showed 192.2mm falling on that day, which contrasted with

the daily-read station at the same location recording only 42.2mm for that day. We next examined the nearest pluviograph and daily-read station pairing, namely gauge number 94029 located 15.6km from the Hobart Airport gauge, and found the aggregated daily rainfall from the pluviograph to be 27.94mm, compared with 27.9mm from the daily rain gauge at that same location. Furthermore, the maximum 6-minute increment rainfall intensity was found to be 1.74mm, substantially smaller than that recorded at Hobart Airport. This therefore indicates that a recording error probably occurred at the pluviograph gauge at Hobart Airport, and that the synthetically generated sequence is more likely to capture the correct behaviour of sub-daily rainfall.

Finally, we note that although daily data is much more abundant than pluviograph data across Australia, in many regions the length or reliability of daily rainfall may not be sufficient for stochastic generation of rainfall sequences. This is the subject of the next chapter, in which the approach presented herein is generalised to any location in Australia, regardless of the availability of daily or pluviograph data.

4. Continuous Simulation Methodology Part III: A Regionalised Approach to Daily Rainfall Generation

4.1. Introduction

Daily rainfall constitutes a basic meteorological input to hydrological, agricultural, ecological and other environmental systems. Stochastic generation of daily rainfall is necessary in these systems to augment or use in place of recorded rainfall data, particularly when observed daily records are short, contain missing records or are unavailable, or where multiple plausible realizations of rainfall are required. The generation of such rainfall sequences is typically achieved using a class of statistical models referred to as ‘weather generators’, which seek to generate a time series of daily (or other time-step) rainfall and other weather variables in a manner that represents statistical properties such as the mean, variance, day-to-day and longer-term persistence and extreme behaviour as present in the instrumental rainfall record [Wilks and Wilby, 1999]. Although weather generators also can be used to characterize other weather variables, the approach presented here has been developed for generation of daily rainfall only.

Generation of daily rainfall proceeds in two distinct stages: firstly the generation of rainfall occurrence to specify the sequencing of wet days, followed by generation of rainfall amounts on the generated “wet” days. One of the earliest – and still most widely used – rainfall occurrence models is the first-order Markov model developed by Gabriel and Neumann [1962], in which the probability of a wet or dry day is defined conditional only on the previous day’s rainfall state. Deficiencies of such ‘short memory’ process models (in which precipitation is only dependent on the past through the most recent day’s rainfall occurrence) include under-simulation of long dry spells and interannual variability, with these issues being addressed in more recent work using higher-order Markov models and Markov models that consider exogenous climate variables as additional predictors [Wilks and Wilby, 1999]. To generate precipitation amounts, *Todorovic and Woolhiser* [1975] used an exponential model to simulate the rainfall amount for each wet day, with two-parameter gamma distributions and mixed exponential distributions also commonly used. An alternative that does not need to assume the probability distribution associated with the rainfall, is presented in the nonparametric weather generation literature [Brandsma and Buishand, 1998; Buishand and Brandsma, 2001; Harrold et al., 2003a; b; U. Lall et al., 1996; Mehrotra and Sharma, 2007a; b; Rajagopalan and Lall, 1999; Rajagopalan et al., 1996; Sharma and O’Neill, 2002; Sharma et al., 1997]. In addition to the above referenced papers, a more detailed review of stochastic generation of rainfall for current as well as climate change conditions is presented in [Sharma and Mehrotra, 2010].

The aim of the methodology presented here is to extend the generation of daily rainfall to locations where daily rainfall records are not available. Traditionally such regionalised extensions have been achieved via the use of spatial interpolation or extrapolation of model parameters [Guennia and Hutchinson, 1998; Johnson et al., 2000; Kyriakidis et al., 2004; Wilks, 2008]. This chapter describes an alternative approach in which sequences are developed using daily rainfall records at ‘similar’ locations which are meteorologically consistent with the rainfall record at the location of interest.

The regionalised procedure presented here uses the Modified Markov Model (MMM) for stochastic generation of daily rainfall as presented in Mehrotra and Sharma [2007b], in which the occurrence model comprises a Markov chain conditional on previous day's rainfall occurrence as well as aggregate rainfall over a number of previous days (e.g. 365 day aggregate number of wet days) to account for low-frequency persistence, and the amounts model uses a nonparametric kernel density estimation procedure with conditional dependence on previous day's rainfall. The daily rainfall sequences generated using the proposed regionalised model, are then disaggregated based on the approaches presented in chapters 2 and 3. The resulting sub-daily rainfall is then compared to the sequences that are derived using observed daily and sub-daily rainfall as per chapter 2, and then using the observed daily rainfall without access to the historical sub-daily rain as per chapter 3. The regionalised procedure presented here is developed using 2708 daily rain gauge locations across Australia as discussed in Section 4.2. In Section 4.3 we summarise the proposed algorithm, and describe the basis for identifying meteorologically 'similar' stations. The results from detailed testing at both a daily and sub-daily time scale are presented in Section 4.4, followed by conclusions in Section 4.5.

4.2. Data

Daily rainfall data are obtained from the Australian Bureau of Meteorology for 17,451 gauging stations, with a maximum of about 8000 daily rain gauges recording rainfall in any given year. The distribution of the daily rainfall network is illustrated in Figure 4.1, in which the number of recording rain gauges is plotted as a time series from 1850 until 2007, with low numbers of stations recording in the mid 1800s, and a build-up of rainfall gauges in the decades surrounding 1900 to approximately present levels. This can be contrasted with the series of sub-daily rainfall presented as Figure 3.2, in which there are a maximum of only around 600-700 sub-daily rainfall stations recording at any time, and with very few recording prior to the 1960s.

Of these daily gauging stations, we selected a subset of 2708 locations (Figure 4.2) that have longer than 25 years of continuous record and less than one percent missing values for developing the similarity metric. Of these stations, 940 have less than 40 years of record, 1437 have between 40 and 100 years and a further 331 stations have records of more than 100 years. In spite of large network of rain gauges, the spatial distribution of the gauging stations is not homogeneous, with a higher density of gauges in the populated regions particularly along the eastern part of Australia. For the remaining analysis we only focus on this set of 2708 stations and fill in the small percentage of missing data using the records of same date from nearby stations.

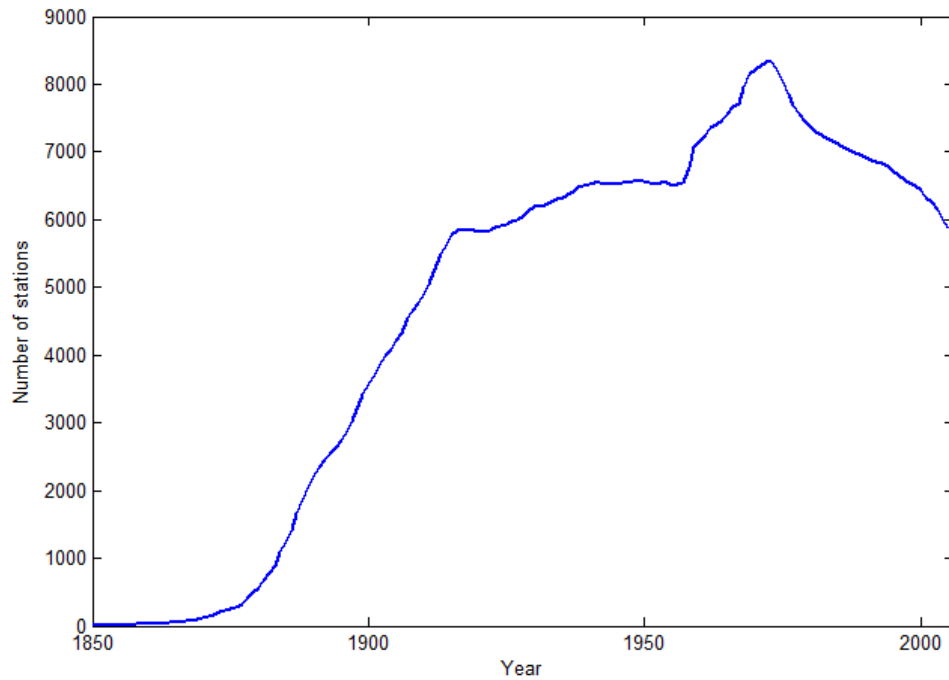


Figure 4.1: Number of Australia-wide daily rainfall records against year of record, plotted from 1850, considering only stations with <1% data missing. Figure extracted from [Mehrotra et al., 2012].

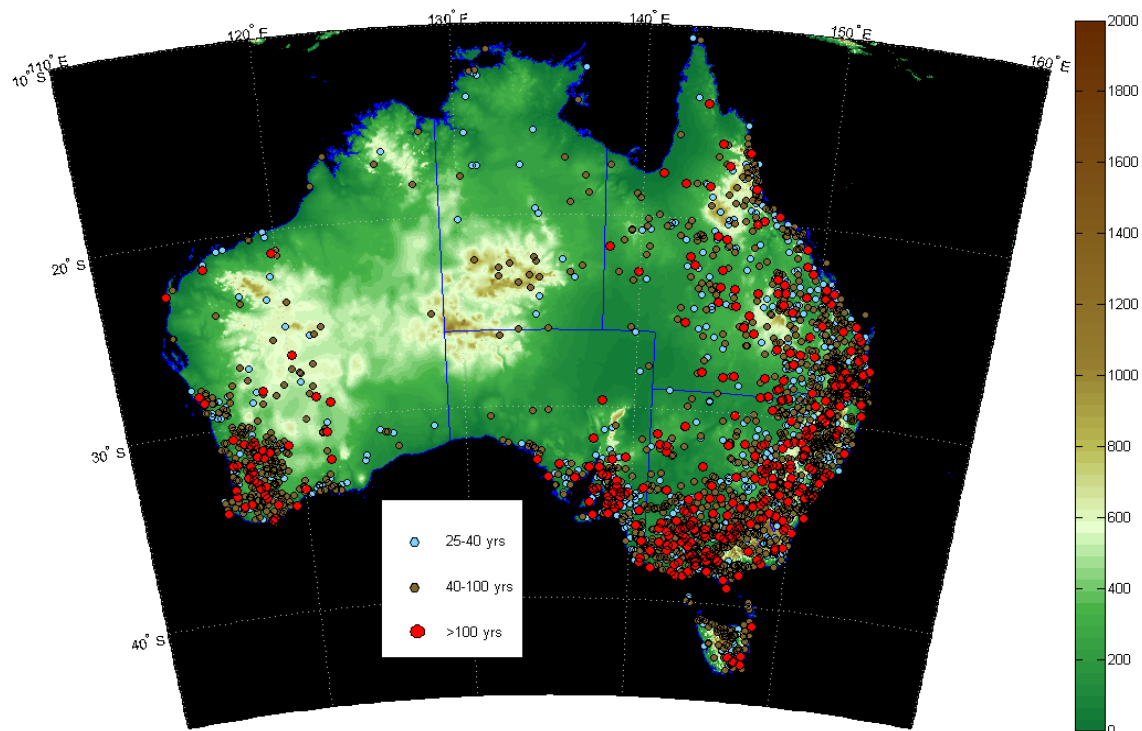


Figure 4.2: Spatial coverage and record length of the Australian daily rainfall record. Only locations with < 1% data missing and length > 25 years are presented, totalling 2708

stations. Figure extracted from [Mehrotra et al., 2012].

4.3. Methodology

Here we present a regionalised approach to generating daily rainfall data at any location of interest regardless of the presence of at-site gauged rainfall data, by sampling the daily rainfall from a number of nearby rain gauges which are considered to be meteorologically ‘similar’. The methodology uses a scaling logic similar to that described in Chapter 3 to identify and define similarity, except that in this case the scaling relationship being investigated relates annual and daily rainfall. Prior to describing the regionalisation approach, we will briefly summarise the daily rainfall generator which is based on the Modified Markov Model (MMM) of Mehrotra and Sharma [2007b], and which was developed to preserve variability across multiple timescales.

4.3.1. Regionalised daily rainfall generation

4.3.1.1. Modified Markov Model for generation of daily rainfall sequences

As in Chapters 2 and 3, we denote R_t as the rainfall amount at a given station on day t (where $t = 1, \dots, 365$ represents the calendar day), and a rainfall occurrence as $I(R_t) = 1$ if $R_t \geq 0.3$ mm and $I(R_t) = 0$ otherwise with $I()$ representing the indicator function. In a traditional Markov order one model, we can express the transition probabilities via $P(I(R_t) | I(R_{t-1}))$, with transition probabilities for each day t estimated separately over a sliding moving window of 15 days either side of t .

The Markov order one model is limited in that it is only dependent on rainfall occurrence on the previous day, and thus cannot represent low-frequency variability which is known to exist in precipitation data [Buishand, 1978]. To ensure such low-frequency variability is correctly maintained, Mehrotra and Sharma [2007b] showed that it is possible to include the vector \mathbf{Z}_t representing long-term predictors, with the predictor matrix including either aggregated rainfall statistics over some number of previous time steps, exogenous predictors such as climate indices, or both. For the present study we focus on a single predictor, namely the aggregate number of rainfall occurrences over the previous 365 days, defined as:

$$Z_t = \sum_{l=1}^{365} I(R_{t-l}) \quad (4.1)$$

The transition probabilities of interest can thus be given by:

$$P(I(R_t) | I(R_{t-1}), \mathbf{Z}_t) \quad (4.2)$$

We use Equation 4.1 of Mehrotra and Sharma [2007b] to calculate the transition probabilities in Equation 4.2 above based on a parametric multivariate normal model. This model requires estimation of 3 parameters, namely wet day transition probabilities and mean and variance of \mathbf{Z}_t , for each day t using the observed data within the moving window.

Having developed the sequence of wet and dry days, it is now necessary to generate rainfall amounts R_t for each wet day. In *Mehrotra and Sharma* [2007b] and *Mehrotra and Sharma* [2010], the rainfall amounts were generated by formulating the conditional probability $f(R_t|\mathbf{C}_t)$ where \mathbf{C}_t is a vector of conditioning variables containing rainfall amounts on previous days as well as possibly exogenous climate indices. In this section we simplify the approach to only consider previous day's rainfall depth, R_{t-1} , as the predictor of current-day rainfall depth, such that we only need to specify the conditional density $f(R_t|R_{t-1})$. This simplification requires us to assume that low-frequency variability in rainfall can be fully accounted for by simulating low-frequency variability in rainfall occurrences, with further evidence to support this assumption coming from a related study which finds that low frequency climate modes such as the El Niño-Southern Oscillation tend to influence rainfall occurrences much more strongly than rainfall amounts on wet days [*Pui et al.*, 2010b].

A Gaussian kernel density estimate [*Sharma and O'Neill*, 2002; *Sharma et al.*, 1997] is used to define $f(R_t|R_{t-1})$. Once again, the density is estimated for each day t using data in a moving window of 15 days on either side of t . This density estimation procedure is described in detail in earlier papers by [*Harrold et al.*, 2003a; *Mehrotra and Sharma*, 2007b; *Sharma*, 2000; *Sharma and O'Neill*, 2002; *Sharma et al.*, 1997].

The full Modified Markov Model for generation of both rainfall occurrences and rainfall amounts as used in the current study is reproduced as Algorithm 4.2.

4.3.1.2. Regionalised extension of daily rainfall generation model

The regionalised extension of this model is somewhat different to the regionalised daily rainfall disaggregation model described in Chapter 3. In particular, rather than re-sampling sub-daily fragments from nearby stations, we use the information from these nearby daily stations to estimate the parameters for the Markov rainfall occurrence model and form the nonparametric kernel density estimate for the rainfall amounts model. The algorithm is described below:

Algorithm 4.1

- 1) Identify a total of S 'nearby' daily rainfall gauging stations with the greatest probability of exhibiting statistically 'similar' rainfall characteristics to the target location [see **Section 4.3.2** for the basis of defining these stations].
- 2) Estimate the parameters of both the occurrence and amounts models at nearby stations using data at that location. At a given day, choose any one of these stations at random with probability $p(j)$, and using the selected station parameters, generate the daily rainfall.
- 3) Normalise the rainfall data by total annual rainfall and move on to the next day, using interpolated contour maps of total annual rainfall supplied by the Bureau of Meteorology.

More details on the regionalised rainfall generation procedure are provided in **Section 4.3.2.3** and in Algorithm 4.2.

4.3.2. Identifying ‘nearby’ daily rainfall stations

Similar to the methodology described in Chapter 3, the regionalised approach relies on using data from nearby rainfall stations (in this case daily-read stations) as a substitute for at-site data for cases where at-site data is either unavailable or too short. As such it is necessary to: (1) identify metrics by which we determine whether two stations are ‘similar’; and (2) predict the probability that stations within a ‘neighbourhood’ of the target location are similar by regressing against physiographic indicators such as latitude, longitude, elevation and distance to coast.

4.3.2.1. Annual and within year daily rainfall characteristics

To enable substitution of daily rainfall series from stations within a neighbourhood of the target location, one needs to consider the equivalence not only of the marginal distributions of annual and within-year rainfall but also the joint relationship between the annual and within-year rainfall at the target station as indicated by superscript ‘o’, and at nearby station indicated by superscript ‘s’. This can be expressed as:

$$f(R_{yt}^s, A_y^s) = f(R_{yt}^o, A_y^o) \quad (4.3)$$

with R_{yt} representing daily rainfall amount for year y , and A_y representing the total annual rainfall for that same year, and $f()$ representing the joint probability density function relating the two variables. For this section we only consider data from the 2708 locations for which long daily rainfall records are available, and assume that such relationships will hold in other locations in Australia for which data is missing. For convenience the subscript y will be omitted from subsequent notation, however when referring to conditional or joint probabilities between annual and daily rainfall, it is implicit that the daily rainfall is sampled from the same year as the aggregate annual rainfall.

A difficulty with this formulation is that R_t^s and R_t^o represent a time series for each year of record ($t = 1, \dots, 365/6$) whereas A^s and A^o represents the total rainfall amount for that year and is therefore a scalar. We therefore modify Equation 4.3 as follows:

$$f(Y^s, A^s) = f(Y^o, A^o) \quad (4.4)$$

where Y^s and Y^o represent within-year scalar attributes of R_t^s and R_t^o for each year of record, respectively. The within-year rainfall behaviour is characterised by various daily, seasonal and spell related rainfall attributes. The attributes to be considered include:

- Maximum daily intensity attributes: for each year, the maximum daily rainfall in each season.
- Maximum wet spells: for each year, the maximum length of sequence of wet days in each season.
- Maximum dry spells: for each year, the maximum length of sequence of dry days in each season.

- Rainfall in maximum wet spells: for each year, the total rainfall in the maximum length of sequence of wet days in each season.
- Amount per wet day: for each year, the average rainfall amount per wet day for each season.
- 7 days rainfall totals: for each year, the maximum 7 days rainfall amount for each season.
- Seasonal rainfall: for each year, the total rainfall amount for each season.
- Seasonal wet days: for each year, the total number of wet days for each season.
- Annual wet days: for each year, the total number of wet days.

In combination, these scalar attributes are expected to cover most of the information on the scaling and timing between annual and within-year rainfall distribution behaviour.

To illustrate these concepts, we present in Figure 4.3 a bivariate scatter plot of annual and summer rainfall at five locations in Australia: Sydney, Perth, Alice Springs, Cairns and Hobart. These locations, which are the same as was used in Chapters 2 and 3, were selected as they have distinctly different climatology, with Hobart located in the south of Tasmania representing one of the southernmost records with temperate climate, Cairns in the north of Queensland representing a location having a moist tropical climate, Alice Springs in the centre of Australia with semi-arid climate, Perth in western Australia representing one of the westernmost records with a mixture of Californian and Mediterranean climates, and Sydney in eastern Australia representing intermediate latitudes.

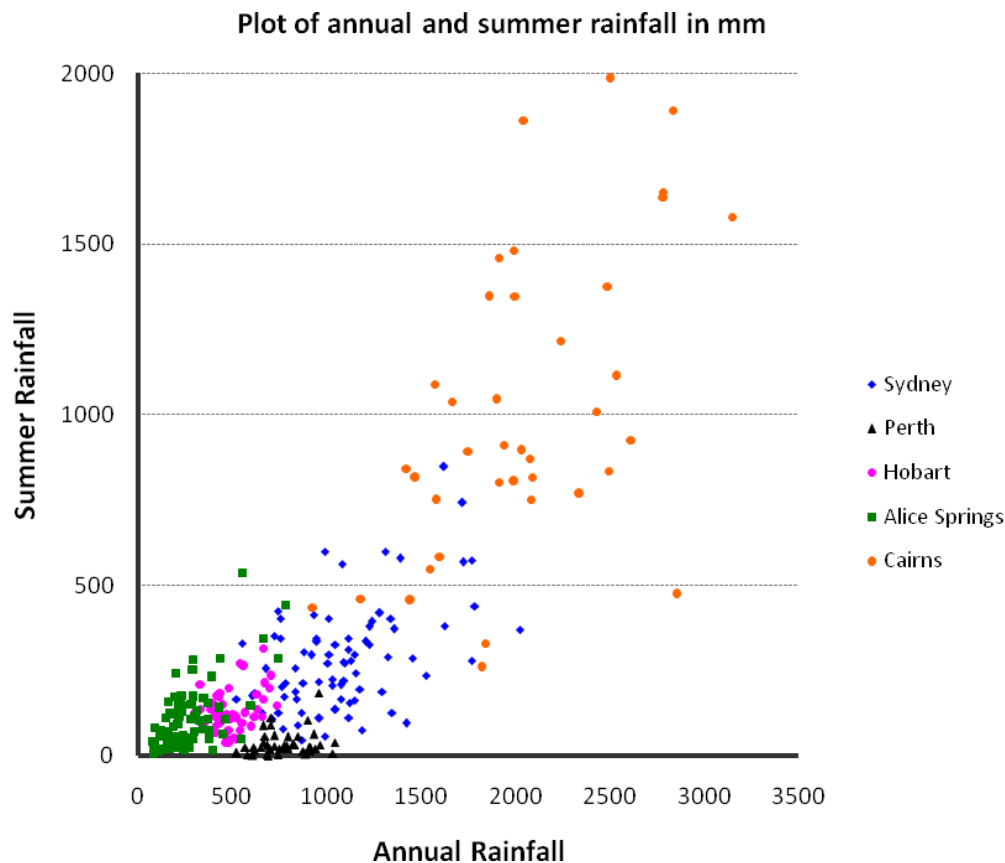


Figure 4.3: Plot of annual rainfall amount and an attribute of with-in-year rainfall (the summer rainfall amount) at five locations in Australia. Figure extracted from [Mehrotra et al., 2012].

As can be seen from this figure, the relationship between seasonal and annual rainfall at each station are distinctly different with Cairns having high annual and summer rainfall amounts whereas Hobart and Alice Springs have relatively little annual and summer rainfall, with summer rainfall being 25% of annual for Hobart and about 40% of annual for Alice Springs. Sydney and Perth have intermediate values of annual rainfall, although a much lower fraction of annual rainfall occurs in summer in Perth compared to Sydney. It is this relationship between annual average rainfall and various sub-annual attributes which is of interest for this study, as it enables a range of climate regimes to be clearly distinguished. Although figures are not provided here, similar conclusions can be drawn from consideration of other within year rainfall attributes.

Another important consideration while dealing with rainfall regionalisation relates to the high spatial variability in rainfall. To highlight this aspect consider rainfall observations at Sydney Observatory Hill. The observed average annual rainfall at the station on the basis of a 150 year long record is 1216 mm, while the observed average annual rainfall at locations within a 20 km radius of Sydney Observatory Hill varies significantly (e.g. Sydney airport, 1087 mm (79 years); Concord golf club, 1135 mm (69 years) and Potts Hill reservoir 917 mm (113 years)). The best

estimate of average annual rainfall from nine nearby stations is 1096 mm, which is 10% below the estimate of the Sydney Observatory Hill annual average rainfall. It is therefore quite likely that identified nearby stations, in spite of having similarity in other rainfall attributes such as seasonality and wet spell characteristics, might contain a bias in annual rainfall relative to the target location. In the following discussions we assume that a good estimate of long term average annual rainfall at the target location is known from some other reliable sources, for example, from the long-term relationships that have been developed by the Australian Bureau of Meteorology for annual rainfall across Australia (http://www.bom.gov.au/jsp/ncc/climate_averages/rainfall/index.jsp). This estimate is then used to scale the generated daily rainfall at nearby stations following a scaling procedure described in Algorithm 4.2.

4.3.2.2. Defining the neighbourhood

Having identified metrics by which to measure the annual and sub-annual rainfall characteristics at any station, we now need to define a neighbourhood over which the annual to sub-annual (seasonal/daily) rainfall scaling is equivalent. Given our assumption that an estimate of total annual rainfall is available and has sufficient accuracy at any target location in Australia, once we have identified the region with consistent annual to sub-annual scaling, we can use the sub-annual (daily) data at nearby locations and finally correct for differences in the total annual rainfall.

Consistent with the methodology described in Chapter 3, for all pairs of daily rainfall stations across Australia we first examine the bivariate distribution $f(Y^s, A^s) = f(Y^o, A^o)$ for annual rainfall and each of the sub-annual rainfall attributes described in the previous section, and test whether they are statistically similar using the two-dimensional, two-sample Kolmogorov-Smirnov (K-S) test as described more fully in that chapter. This test was developed by [Fasano and Franceschini, 1987] and is summarised in Chapter 3. The chi-square test described in Chapter 3 was not used here due to the much smaller sample sizes (between 25 and 150 samples at any location, with one sample for each year of record); the two dimensional two sample K-S test was specifically developed for such small sample sizes [Fasano and Franceschini, 1987] and therefore remains appropriate.

In total, 2708 separate rain gauge stations with at least 25 years of data were used to formulate this relationship, totalling 3,665,278 station pairs. We classify a station pair to be statistically similar based on the K-S test using a 95% confidence threshold, and thus have a vector of length 3,665,278 with all the classifications of whether the stations are statistically similar or different.

Figure 4.4 presents changes in the percentage of station pairs which are statistically 'similar', with increases in absolute difference in latitude and longitude between station pairs based on a frequency binning approach. The percentage of significant stations is calculated by counting the number of statistically similar station pairs in each bin (using a total of 50 bins), and are presented for seven attributes of within year rainfall: maximum wet spell summer, maximum dry spell winter, daily maximum rainfall summer, 7 days cumulative rainfall summer, rainfall in

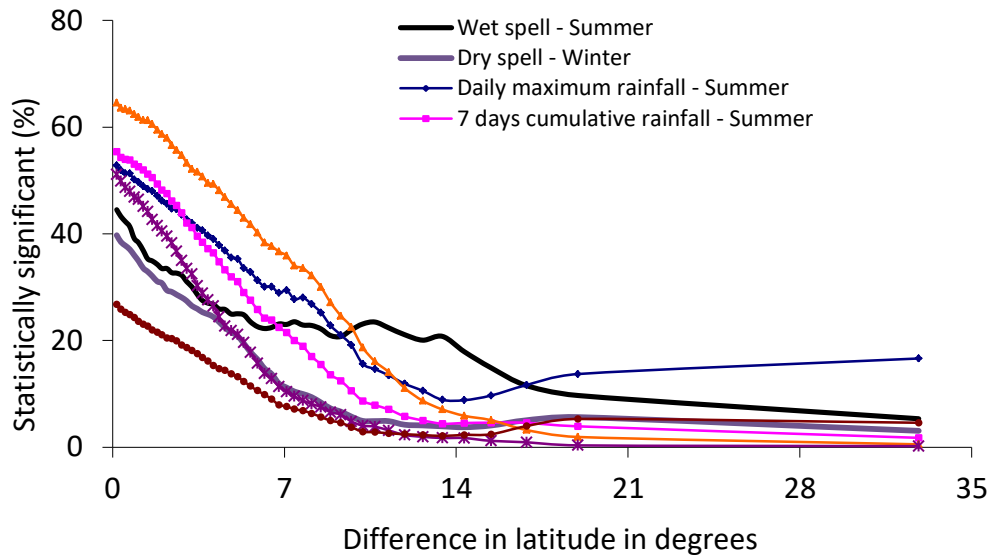
maximum wet spell summer seasonal rainfall summer and number of wet days winter.

Some interesting conclusions can be derived from this figure. Firstly, with the exception of the number of wet days in winter, there is between a 35% and 65% chance that the joint distribution of annual rainfall and each of the within year rainfall attributes are statistically similar provided the difference in latitude or longitude is small, with the probability decreasing rapidly as difference in latitude or longitude increases. This is interesting, as in Figure 4.4a, no account is made of any other physiographic information, such as longitude, elevation and distance to coast, such that stations may be located in opposite sides of the continent, or at very different elevations, and yet still have close to a 50% chance of having the same scaling between annual and with-in-year rainfall provided they are at similar latitudes. Secondly, while the probability that two stations are similar decreases with increasing difference in longitude for small differences, the probability increases again once the difference in longitude reaches about 20 to 25 degrees. This result is due to the clustering of stations as shown in Figure 4.2, with groups of stations in the south west and southern parts of the continent showing similar climatology. For subsequent analysis we only consider predictors for station pairs with a difference in latitude less than 15 degrees, difference in longitude less than 10 degrees, and difference in elevation less than 350 metres, with a total of 1,646,664 station pairs meeting these criteria. This ensures that the probability that two stations are similar can be represented as a smoothly varying function which decreases monotonically as the magnitude of each of the predictors increases.

We now use a logistic regression model to find the probability that any two stations are similar conditional to a range of physiographic metrics, such as the difference in latitude, longitude, elevation and distance to coast between each station pair. This formulation is equivalent to the formulation specified in Equation 3.4 and will not be repeated here. This model is developed in a multivariate setting using all of the above physiographic metrics as predictors, with the conceptual basis shown in Figure 3.5.

The results of this multivariate regression for all key rainfall attributes are presented in Table 4.1, and once again plotted for the selected rainfall attributes in Figure 4.5, against an amalgamated variable comprising the mean of all the predictors when expressed as a percentage of their maximum range. As can be seen, the results show notable improvements in the probability that two stations are similar compared to Figure 4.4, since now we are considering the influence of other predictors as well. In fact, with the exception of the number of wet days and maximum dry spells in winter season, the results show that for small values of each of the predictors there is more than 80% probability that the annual to within year joint probability distributions are statistically similar. This forms the basis for our assertion that, provided an adequate estimate of annual rainfall is available at the location of interest, it is possible to draw data from daily-read gauges within a neighbourhood of that location.

(a) Difference in latitude



(b) Difference in longitude

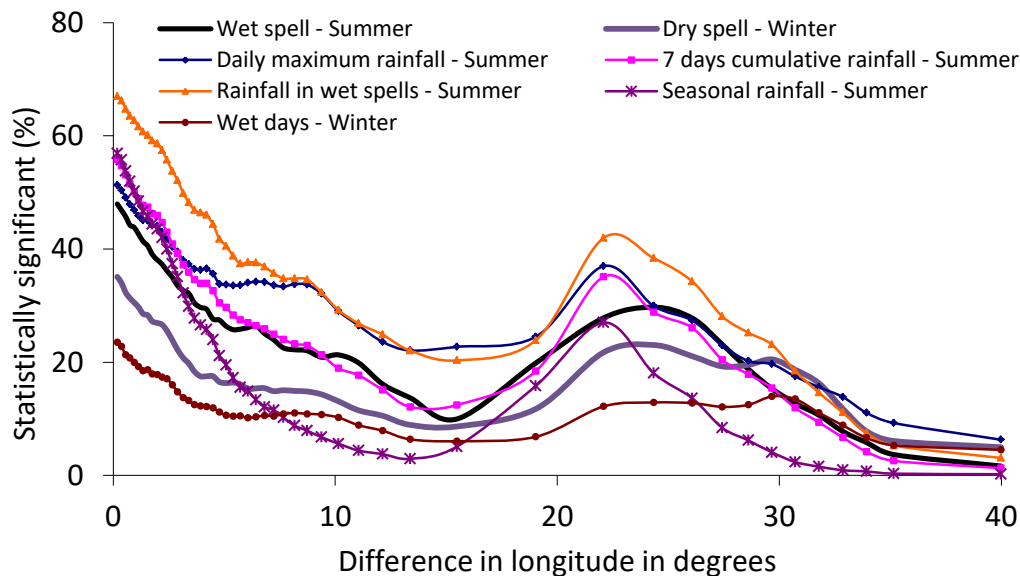


Figure 4.4: Percent of statistically similar stations against a single predictor – (a) difference in latitude and (b) difference longitude, and seven responses representing different with-in-year rainfall attributes. The responses have been calculated using the Kolmogorov-Smirnov test statistic. Vertical axis shows number of station pairs being statistically similar in terms of annual and with-in-year rainfall attributes at 95% confidence level out of total number of station pairs in a bin. Figure extracted from [Mehrotra et al., 2012].

Table 4.1: Logistic regression coefficients. All predictors were found to be statistically significant

(usually with a p-value < 0.001 level).

Season	With-in-year rainfall attribute	Logistic regression coefficients					
		Intercept	Lat	Lon	Elev	Dist_coast	Lat*Lon
DJF	Maximum daily rainfall	1.94	-0.311	-0.217	-0.00006	-0.996	0.0292
DJF	Maximum wet spells	1.57	-0.124	-0.238	-0.00097	-1.84	0.0194
DJF	Maximum dry spells	1.27	-0.0815	-0.299	-0.00067	-1.99	0.0260
DJF	Maximum 7 days cumulative rainfall	2.10	-0.359	-0.233	0.00022	-0.528	0.0166
DJF	Rainfall in maximum wet spell	2.69	-0.331	-0.228	0.00061	-0.664	0.0149
DJF	Amount per wet day	0.715	-0.159	-0.142	-0.00092	-2.03	0.0240
DJF	Total rainfall	2.26	-0.421	-0.351	0.00035	-0.280	0.0146
DJF	Number of wet days	0.687	-0.102	-0.283	-0.00129	-1.78	0.0220
MAM	Maximum daily rainfall	1.81	-0.134	-0.175	0.00103	-1.22	0.0169
MAM	Maximum wet spells	1.41	-0.0949	-0.0953	-0.00044	-2.88	0.0124
MAM	Maximum dry spells	1.60	-0.0989	-0.136	-0.00087	-3.17	0.0183
MAM	Maximum 7 days cumulative rainfall	2.45	-0.173	-0.159	0.0006	-1.29	0.0168
MAM	Rainfall in maximum wet spell	3.13	-0.212	-0.194	0.00052	-1.29	0.0217
MAM	Amount per wet day	0.748	-0.142	-0.162	-0.0007	-1.93	0.0250
MAM	Total rainfall	3.43	-0.145	-0.109	-0.00028	-2.58	0.0077
MAM	Number of wet days	0.704	-0.168	-0.123	-0.00067	-2.47	0.0228
JJA	Maximum daily rainfall	1.82	-0.135	-0.227	0.00107	-1.71	0.0178
JJA	Maximum wet spells	0.655	-0.265	-0.120	-0.0001	-0.817	0.0183
JJA	Maximum dry spells	0.740	-0.326	-0.167	0.00036	-0.527	0.0111
JJA	Maximum 7 days cumulative rainfall	1.94	-0.226	-0.223	0.00047	-0.897	0.0115
JJA	Rainfall in maximum wet spell	2.15	-0.189	-0.181	0.00044	-0.993	0.0115
JJA	Amount per wet day	0.477	-0.165	-0.190	-0.00026	-1.44	0.0221
JJA	Total rainfall	1.53	-0.337	-0.277	0.00021	-0.231	0.0059
JJA	Number of wet days	0.0353	-0.318	-0.139	0.00015	-0.504	0.0136
SON	Maximum daily rainfall	2.15	-0.135	-0.175	-0.00038	-2.23	0.0197
SON	Maximum wet spells	1.14	-0.174	-0.154	0.00015	-1.75	0.022
SON	Maximum dry spells	1.17	-0.389	-0.170	0.00058	-1.00	0.0337
SON	Maximum 7 days cumulative rainfall	2.69	-0.157	-0.152	0.00052	-2.15	0.0124
SON	Rainfall in maximum wet spell	3.38	-0.197	-0.0942	0.00065	-1.63	0.0141
SON	Amount per wet day	0.602	-0.147	-0.116	-0.00085	-2.15	0.0216
SON	Total rainfall	2.559	-0.222	-0.138	0.00025	-1.62	0.0161
SON	Number of wet days	0.324	-0.212	-0.167	0.00014	-1.32	0.0266
Annual	Annual wet days	-0.500	-0.147	-0.0981	-0.00039	-1.62	0.0199

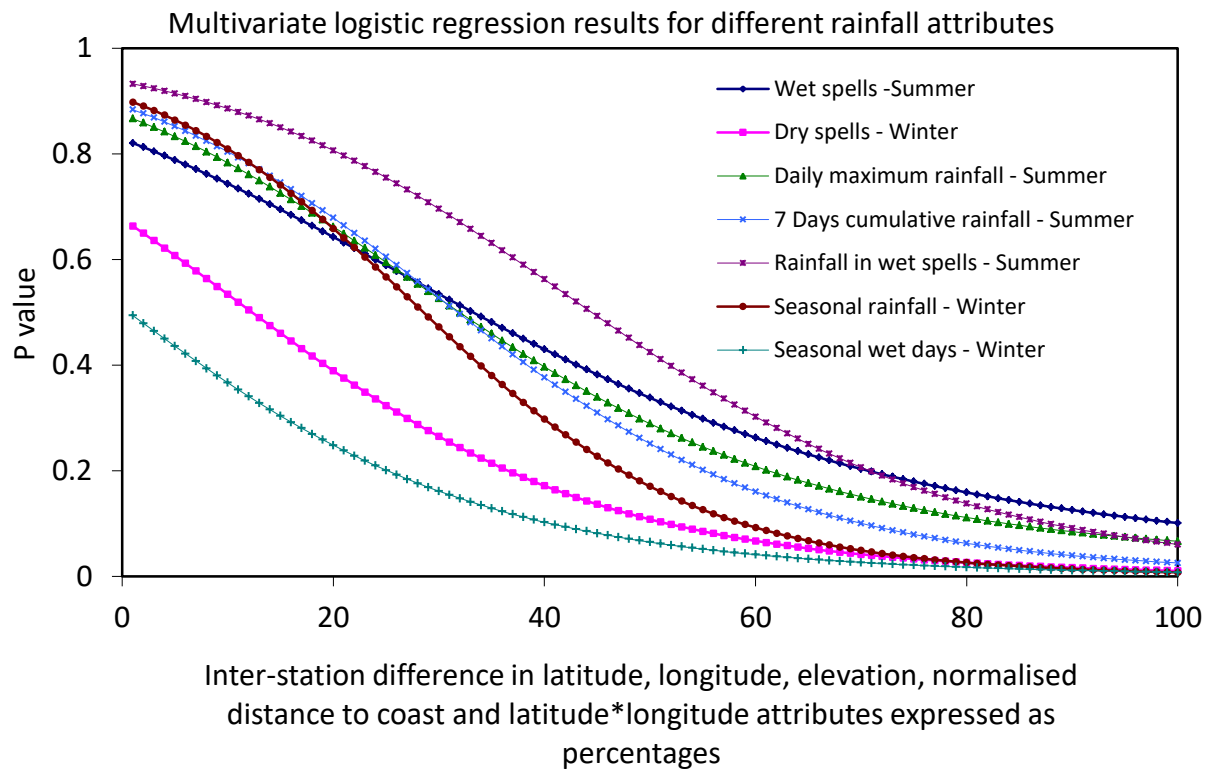


Figure 4.5: As per Figure 4.4, except the results represent the outcomes of the full multivariate logistic regression. The probability that annual and with-in-year rainfall attributes are statistically similar is plotted against percent differences in latitude, longitude, elevation, normalised distance to coast and latitude*longitude with 100 percent representing 15 degree difference in latitude, 10 degree difference in longitude, 350 metre difference in elevation, 1 unit of scaled difference in distance to coast and 75 squared degree latitude*longitude. Figure extracted from [Mehrotra et al., 2012].

4.3.2.3. Model Implementation

On the basis of the methodology described in the previous section, multivariate logistic relations are developed for all key rainfall attributes, with regression coefficients as shown in Table 4.1. Owing to a large pool of rainfall attributes, the developed relationships are examined closely and a few important rainfall attributes are selected encompassing the full distribution of relationships and also capturing the overall seasonal variations. The finally selected rainfall attributes include: (a) rainfall in maximum wet spells – winter; (b) rainfall in maximum wet spells – summer; (c) number of wet days – winter; (d) number of wet days – summer; (e) total rainfall amount – winter; (f) total rainfall amount – summer and; (g) maximum wet spells – summer, totalling seven rainfall attributes.

The approach to identifying ‘nearby’ stations is as follows:

1. For any location of interest (the ‘target’ location), identify the probability (u) that each of the 2708 daily rain gauge stations in Australia is statistically similar using the logistic regression coefficients provided in Table 4.1.

2. Rank the probabilities from lowest to highest for each rainfall attribute, and calculate the average rank, r_i , for each of the 2708 stations across all rainfall attributes.
3. The S lowest-ranked stations represent 'statistically similar nearby stations' for inclusion in the daily rainfall generation model.
4. Calculate the weight associated with each nearby stations using the following:

$$w_i = \frac{1/r_i}{\sum_{k=1}^S 1/r_k} \quad (4.5)$$

where the w_i represents the weight associated with the i th station. Lowest ranked stations with statistically similar rainfall attributes will have higher weight and therefore have a high probability of being selected in the rainfall generation algorithm.

The selection of the size of S is somewhat subjective, as larger values of S increase the probability of selecting stations which are statistically different to the target station, whereas smaller values of S will result in small sample sizes. For this study we selected $S = 5$, resulting in an average of approximately 125 - 200 years of data distributed over the 5 stations. The stepwise procedure of rainfall generation at the target station using the daily records of S nearby stations is given in the algorithm below:

Algorithm 4.2a – Identification of nearby stations and model parameter estimation

1. Identify the S nearby stations following the procedure outlined in section 4.3.2.3. Calculate the weight w_s associated with each nearby station s using equation 4.5. Low ranked stations with statistically similar rainfall attributes will have higher weights. Transform these weights to probabilities (P_s) and cumulative probabilities (P_{ws}) using the following:

$$P_s = \frac{w_s}{\sum_{i=1}^S w_i} \text{ and } P_{ws} = P_{ws-1} + P_{ws} \text{ for } s>1 ; P_{w1} = P_1 \quad (4.6)$$

2. Calculate the average annual rainfall, \bar{A}^s , at these stations, and the average annual rainfall at the target station, \bar{A}^o , using a spatially interpolated map of total annual rainfall across Australia (can be obtained from the Bureau of Meteorology web site : http://www.bom.gov.au/jsp/ncc/climate_averages/rainfall/index.js).
3. At each identified nearby location s , for all calendar days of the year, calculate the transition probabilities of the standard first order Markov model and, conditional means and variances of the higher time scale predictor variable Z (previous 365 days wetness state) using the observations falling within the sliding window of 15 days on either side of the current day. Denote these transition probabilities as $P_{1,1}(s)$ for current day being wet, previous day was also wet and $P_{1,0}(s)$ for current day being wet, previous day was dry and, conditional means and variances of predictor variable Z as μ and \mathbf{V} (of dimension 4), respectively for all four cases of current and previous days being wet or day. Also for each calendar day (t), look for wet days ($I(R_t)=1$) within the same sliding window and form

series of current day rainfall amount ($R_j(s)$) and associated previous days rainfall value ($R_{j-1}(s)$). For a given day t , let j index varies from 1 to N . Calculate variances and covariances (Σ) of $R_j(s)$ and $R_{j-1}(s)$ series.

4. Before the start of simulation, select at random a nearby station. Pick a short segment (one year) of the historical sequence at this station to use for the initial specification of Z_t . The first day in the generated sequence is the day immediately after the end of this start up sequence.

Algorithm 4.2b – Generation of rainfall occurrences

1. At a given day t , generate a uniformly distributed random number u and identify the position s^* such that $P_{ws^*-1} < u \leq P_{ws^*}$, thereby selecting a nearby station s . Assign appropriate transition probability to the day t based on previous day's rainfall state of the generated series at the target station. If previous day is wet, assign probability P as $P_{11}(s)$ otherwise assign $P_{10}(s)$.
2. Calculate the value of the previous 365 days wetness state (number of wet days) prior to the day t using equation (1) and the available generated sequence $I(R^*)$ at target station, where R^* defines the generated rainfall series. Modify the transition probability P of earlier step using equation (A2) and, conditional means and variances of higher time scale predictor for the generated day t at the nearby station s :

$$\hat{P} = P_{1,i} \frac{\frac{1}{(V_{1,i})^{1/2}} \exp\left\{-\frac{1}{2}(Z_t - \mu_{1,i})V_{1,i}^{-1}(Z_t - \mu_{1,i})'\right\}}{\left[\frac{1}{(V_{1,i})^{1/2}} \exp\left\{-\frac{1}{2}(Z_t - \mu_{1,i})V_{1,i}^{-1}(Z_t - \mu_{1,i})'\right\} P_{1,i}\right] + \left[\frac{1}{(V_{0,i})^{1/2}} \exp\left\{-\frac{1}{2}(Z_t - \mu_{0,i})V_{0,i}^{-1}(Z_t - \mu_{0,i})'\right\} (1 - P_{1,i})\right]} \quad (4.7)$$

where the $\mu_{1,i}$ parameter represents the mean $E(Z_t | I(R_t) = 1, I(R_{t-1}) = i)$ and $V_{1,i}$ is the corresponding variance. Similarly, $\mu_{0,i}$ and $V_{0,i}$ represent, respectively, the mean and the variance of Z when $(I(R_{t-1}) = i)$ and $(I(R_t) = 0)$. The $P_{1,i}$ represents the baseline transition probability of the first order Markov model defined by $P(I(R_t) = 1 | I(R_{t-1}) = i)$ with i being either 0 or 1.

3. Denote the modified transition probability as \hat{P} . Generate a uniformly distributed random number u and compare it with \hat{P} . If u is $\leq \hat{P}$, assign rainfall occurrence, $I(R_t^*)$ for the day t as wet (1) otherwise dry (0). If day is simulated as dry, move on to the next day ignoring rainfall amount generation steps.

Algorithm 4.2b – Generation of rainfall amount on wet days

1. The conditional simulation of rainfall amount on a day t using a bivariate Gaussian kernel density estimate is given by the following [Sharma et al. 1997]:

$$\hat{f}(R_t | R_{t-1}) = \sum_{j=1}^N \frac{1}{(2\pi \Sigma')^{0.5} \lambda} \omega_j \exp\left(-\frac{(R_t - b_j)^2}{2\lambda^2 \Sigma'}\right) \quad (4.8)$$

where,

$$\omega_j = \frac{\exp\left(-\frac{(R_{t-1} - R_{j-1})^2}{2\lambda^2 \Sigma_{22}}\right)}{\sum_{j=1}^N \exp\left(-\frac{(R_{t-1} - R_{j-1})^2}{2\lambda^2 \Sigma_{22}}\right)}; \Sigma' = \Sigma_{11} - \frac{\Sigma_{12}^2}{\Sigma_{22}}; b_j = R_j + (R_{t-1} - R_{j-1}) \frac{\Sigma_{12}}{\Sigma_{22}} \quad (4.9)$$

here, $\hat{f}(R_t | R_{t-1})$ is the conditional probability density estimate, λ is the bandwidth, b_j is the conditional mean associated with each kernel slice, ω_j is the weight associated with each kernel slice that constitutes the conditional probability density and N is total number of data points falling within the sliding window and satisfying the condition $(I(R_j)=1, j=1, N)$. Σ_{11} , Σ_{12} etc. are the terms in the covariance matrix Σ , expressed as:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad (4.10)$$

The bandwidth λ is adopted here is the Gaussian reference bandwidth λ_{ref} following

[Scott, 1992] and is expressed as: $\lambda_{ref} = \left(\frac{4}{m+2}\right)^{1/(m+4)} N^{(-1/(m+4))}$ where, m equals

number of conditioning variables which is one in our case here.

2. Conditional rainfall simulation proceeds by estimating weights ω_j for the kernel slices for all N data points that are associated with each data pair (R_j, R_{j-1}) and R_{t-1}^* using equation (A4). These weights represent the contribution that each kernel has in forming the conditional probability density. These weights are transferred to cumulative probability P_j using the following:

$$p_j = \frac{\omega_j}{\sum_{i=1}^N \omega_i} \text{ and } P_j = P_{j-1} + p_j \text{ for } j>1; P_1 = p_1 \quad (4.11)$$

3. Generate a uniformly distributed random number u and identify the position j^* such that $P_{j^*-1} < u \leq P_{j^*}$, thereby selecting an $R_j(s)$ value from the $R(s)$ series. Calculate b_j as the conditional mean associated with the kernel $R_j(s)$ using equation (4.9).

4. Sample R_t^* as a random variate from the kernel centred on b_j (the conditioned kernel slice being a Gaussian PDF with a mean b_j and a variance equal to $(\lambda^2 \Sigma')$,

$$R_t^* = b_j + \lambda(\sqrt{\Sigma'})W_t \quad (4.12)$$

where W_t is a random variate from a normal distribution with mean of 0 and variance of 1, Σ' is a measure of spread of the conditional density given by equation 4.9 and R_t^* is the generated rainfall at a day t . If generated rainfall is less than rainfall threshold of 0.3 mm, go back to step 3 else move on to the next step.

5. Rescale the generated daily rainfall by multiplying it by the ratio \bar{A}^o/\bar{A}^s .
6. Move to the next day in the generated sequence and repeat above steps (starting from rainfall occurrences) until the desired length of generated sequence is obtained.

4.4. Results

We tested the applicability of the logic outlined in Section 4.3 in this section. Specifically, we assessed the capability of the regionalised daily simulation model (formulated to not use the observed record for the location being modelled) in representing attributes derived from the observed daily record, followed by an assessment of the continuous rainfall sequences derived through disaggregation from the generated daily sequence. Our assessment is based on daily and sub-daily rainfall data at five climatologically different locations in Australia (Sydney, Perth, Alice Springs, Cairns and Hobart). It should be noted that the regionalised daily generation model reported here uses all available daily rain gauges in Australia (totalling to 2708 gauges with a minimum of 25 years of data), and not just the 1282 gauges that were used in the development of the logistic regression relationships (where only station pairs with a difference in latitude less than 15 degrees, difference in longitude less than 10 degrees, and difference in elevation less than 350 metres were used). The assessment results in the following sub-sections are based on 100 realisations, each equalling the record length of the historical data available at each location.

4.4.1. Annual and seasonal statistics

The seasonal and annual means and standard deviations of wet days and rainfall amounts from the simulated and observed daily rainfall time series are presented in Table 4.2. The means of both number of wet days and rainfall amounts are reproduced reasonably well, with the simulated results generally within 10% of the observed data. The primary exception to this is for Alice Springs, in which the simulated mean number of wet days is between 17.1% and 50.0% below the observed number of wet days, with the rainfall amount also being underestimated by 18.4% for the winter season. The reason for this discrepancy is likely to be the sparse sampling of rainfall in the vicinity of Alice Springs leading to the selection of 'nearby' gauges which are not reflective of at-site daily rainfall; furthermore the arid nature of the Alice Springs climate may also contribute to results, with much of the rainfall being contained in a small number of wet years potentially leading to less consistent results. In all cases the average annual observed and

simulated rainfall amounts correspond exactly, as each simulated series is scaled to the observed rainfall amounts. In a setting where observed data is not available such scaling will be achieved using a spatially interpolated total annual rainfall product, therefore inducing an additional source of uncertainty. Unlike the mean rainfall, the annual standard deviations are generally under-simulated, by an average of about 12% for number of wet days and by an average of 19% for rainfall amounts.

Table 4.2: Observed and simulated rainfall statistics for five selected locations.

Season/ Station	Number of wet days				Rainfall amounts (mm)			
	Mean		Standard deviation		Mean		Standard deviation	
	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)
Sydney (066037)								
Autumn	33.1	33.9 (32.7-35.6)	8.4	7.1 (6.3-8.1)	320	336 (319-357)	159	123 (105-141)
Winter	28.4	27.4 (26.3-28.7)	8.3	6.7 (5.8-7.7)	267	255 (238-272)	147	107 (93-123)
Spring	30.0	30.1 (28.9-31.6)	7.9	6.4 (5.6-7.3)	214	217 (204-232)	110	83 (71-96)
Summer	31.9	32.8 (31.6-34.1)	8.4	6.5 (5.6-7.3)	285	280 (263-297)	158	102 (87-118)
Annual	123.3	124.3 (121.6- 127.3)	21.0	16.2 (13.7-18.8)	1086	1087 (1087-1087)	317	222 (187-252)
Perth (009021)								
Autumn	23.5	22.9 (21.8-24.2)	5.4	6.2 (5.3-7.0)	161	167 (155-177)	63	63 (54-73)
Winter	49.5	49.1 (47.7-51.5)	7.4	7.8 (6.6-8.9)	438	424 (412-438)	89	91 (78-110)
Spring	28.2	29.9 (28.6-31.0)	6.9	6.2 (5.4-7.2)	144	153 (145-160)	47	45 (39-51)
Summer	8.3	8.8 (8.0-9.6)	4.1	6.2 (5.4-7.2)	35	38 (32-43)	35	26 (20-33)
Annual	109.5	110.7 (108.3- 114.0)	15.8	12.9 (11.0-14.9)	776	781 (781-782)	143	125 (109-145)
Alice Springs (015590)								
Autumn	7.8	5.4 (4.6-6.5)	5.1	3.8 (3.2-4.7)	67	66 (56-77)	76	60 (48-73)
Winter	6.9	3.6 (2.9-4.2)	5.4	2.6 (2.1-3.1)	38	29 (24-35)	45	28 (22-34)
Spring	11.6	8.0 (7.1-9.6)	5.5	4.3 (3.4-5.3)	58	59 (51-67)	42	42 (33-55)
Summer	14.2	10.7 (9.6-12.9)	6.0	5.5 (4.6-6.8)	117	125 (115-139)	102	92 (73-115)
Annual	40.6	27.8 (24.9-32.0)	12.9	10.9 (8.2-14.1)	280	279 (279-279)	152	144 (114-176)
Cairns (031011)								
Autumn	49.2	48.4 (46.7-50.5)	8.2	7.7 (6.5-9.2)	722	730 (701-765)	327	222 (181-267)
Winter	25.0	26.2 (24.8-27.5)	8.3	6.9 (5.8-8.0)	105	140 (129-151)	51	56 (46-69)
Spring	24.3	24.7 (23.4-25.9)	8.9	6.6 (5.8-7.5)	165	187 (172-205)	110	83 (71-100)
Summer	49.4	47.7 (46.3-49.1)	9.0	7.3 (6.2-8.6)	1008	933 (900-968)	414	270 (235-313)
Annual	147.9	147.0 (144.7- 149.9)	17.8	14.2 (12.1-16.3)	2000	1991 (1991-1991)	555	366 (306-421)
Hobart (094008)								

Autumn	29.4	28.1 (26.0-30.7)	6.7	6.5 (5.1-8.4)	115	109 (101-118)	53	38 (31-46)
Winter	34.7	33.0 (31.1-35.3)	8.2	6.4 (5.2-7.6)	119	121 (112-129)	42	35 (29-41)
Spring	35.9	35.6 (33.2-39.5)	7.2	7.4 (6.2-8.6)	131	139 (130-148)	46	43 (35-51)
Summer	26.2	26.0 (23.2-29.7)	5.8	6.8 (5.0-8.6)	131	128 (116-139)	60	48 (40-58)
Annual	126.2	122.7 (114.8-133.5)	19.8	19.5 (14.1-24.2)	496	496 (496-496)	110	92 (76-105)

Figure 4.6 presents the year-to-year distribution of the annual rainfall amounts and annual number of wet days across a range of exceedance probabilities. As can be seen, for total annual rainfall amounts although the median is well simulated, the variability is low for most locations, with the upper and lower bounds of the extremes being underestimated. In contrast, the number of wet days is generally well reflected. The exception to this is once again Alice Springs, where the distribution of annual rainfall is accurately represented whereas the number of wet days is underestimated. This can be explained by the transition probability parameters provided in Table 4.3, which are generally within 10% of the at-site parameters for all locations except for Alice Springs.

The results show overall good agreement between the observed and simulated statistics at all stations. The underestimation of variability at annual time scale is attributable more to the structure and assumptions of the daily rainfall generation model adopted here than to the regionalisation procedure. The simplified structure of daily rainfall generation model (a single predictor as aggregate number of rainfall occurrences over the previous 365 days and use of global bandwidth in kernel density estimation procedure) and the assumption of normal distribution in equation 4.7 may result in these discrepancies in the results. To check whether the under estimation of variability is due to the regionalisation procedure adopted here, we used the same model for rainfall generation at these sites using the observed at site rainfall record and obtained the similar results (not included). Experimenting with a larger number of predictors (Mehrotra and Sharma, 2007a), using the local bandwidth in rainfall simulation procedure (Sharma et al, 1997) and using aggregated wet day predictor(s) in the rainfall amount simulation stage (Harrold et al, 2003b) might help further improve the representation of observed year to year variability in the simulations.

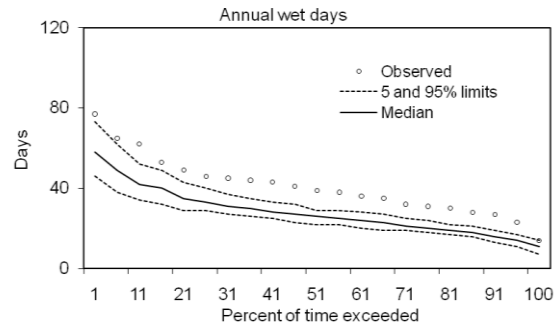
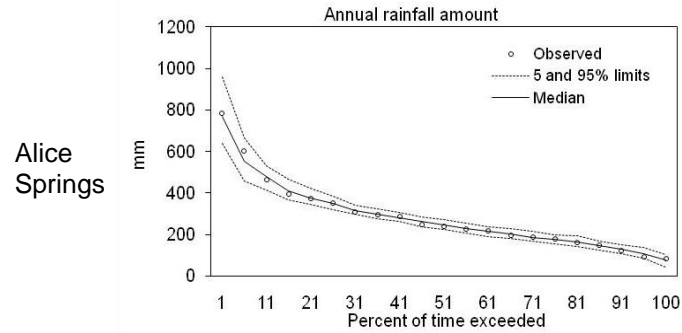
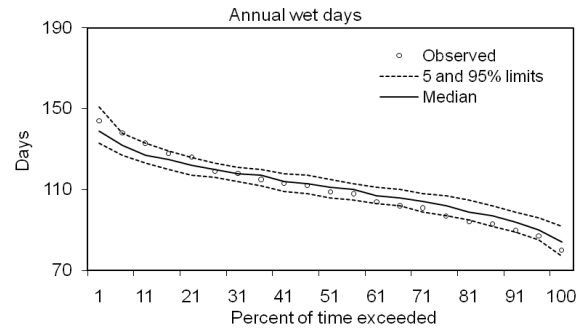
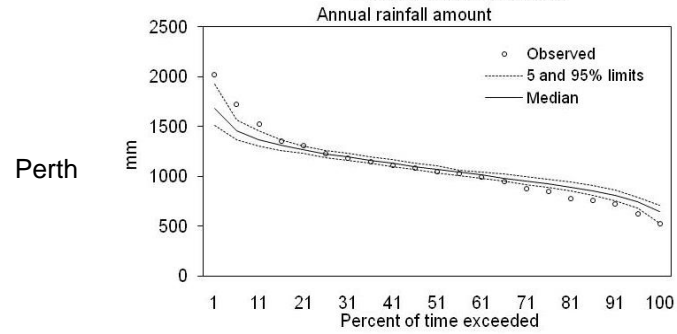
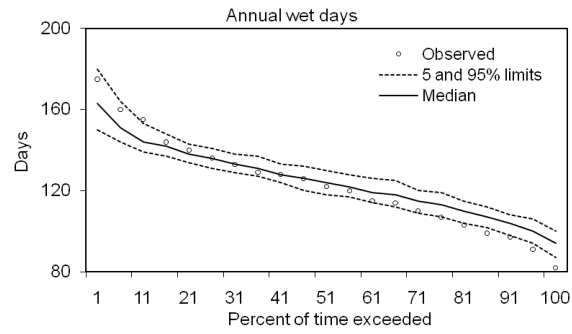
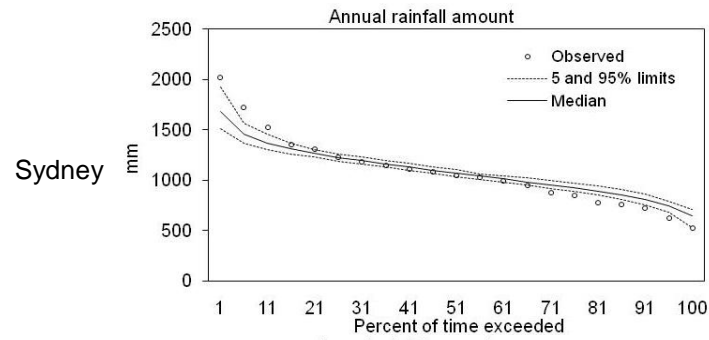
Table 4.3: Observed and simulated rainfall transition probabilities for five selected locations. Also shown are the percent differences in the brackets.

Station/ Probability	Sydney		Perth		Alice Springs		Cairns		Hobart	
	Observed	Simulated	Observed	Simulated	Observed	Simulated	Observed	Simulated	Observed	Simulated
p ₁₀	0.153	0.155 (1.3%)	0.117	0.117 (-0.1%)	0.060	0.045 (-25.0%)	0.119	0.128 (7.4%)	0.178	0.178 (-0.5%)
p ₁₁	0.184	0.185 (0.7%)	0.184	0.186 (1.3%)	0.051	0.031 (-39.5%)	0.285	0.274 (-3.7%)	0.168	0.158 (-6.0%)
p ₁₁₁	0.103	0.102 (-1.0%)	0.116	0.120 (3.8%)	0.022	0.013 (-41.8%)	0.207	0.189 (-8.7%)	0.082	0.075 (-7.9%)
p ₁₁₀	0.081	0.084 (2.9%)	0.068	0.066 (-2.8%)	0.029	0.018 (-37.8%)	0.077	0.085 (9.8%)	0.086	0.082 (-4.3%)
p ₀₁₀	0.072	0.071 (-0.4%)	0.049	0.051 (3.5%)	0.032	0.027 (-13.6%)	0.042	0.044 (2.9%)	0.092	0.095 (3.1%)
P ₀₁₁	0.081	0.084 (2.9%)	0.068	0.066 (-2.8%)	0.029	0.018 (-37.8%)	0.077	0.085 (9.8%)	0.086	0.082 (-4.3%)

4.4.2. Sub-daily statistics

Results based on the disaggregation of the generated daily rainfall to a sub-daily time step are presented in Table 4.4 and Figures 4.7 and 4.8. These results are analogous to Table 3.4 and Figures 3.8 and 3.9 in which at-site daily rainfall was used but sub-daily fragments were sourced from nearby pluviograph stations. Thus, the comparison of these results can be used to determine the impact on precipitation extremes and antecedent rainfall for the case when daily rainfall is also simulated using nearby station records.

As can be seen, the results are very similar to those presented in Chapter 3 for all cases, although the confidence intervals are slightly wider suggesting that sourcing daily rainfall information from a greater range of stations increases variance in both extremes and the antecedent conditions leading up to the mean. Nevertheless, these changes are minor and suggest that the regionalisation of the daily rainfall model does not result in significant deterioration of simulated sub-daily rainfall statistics.



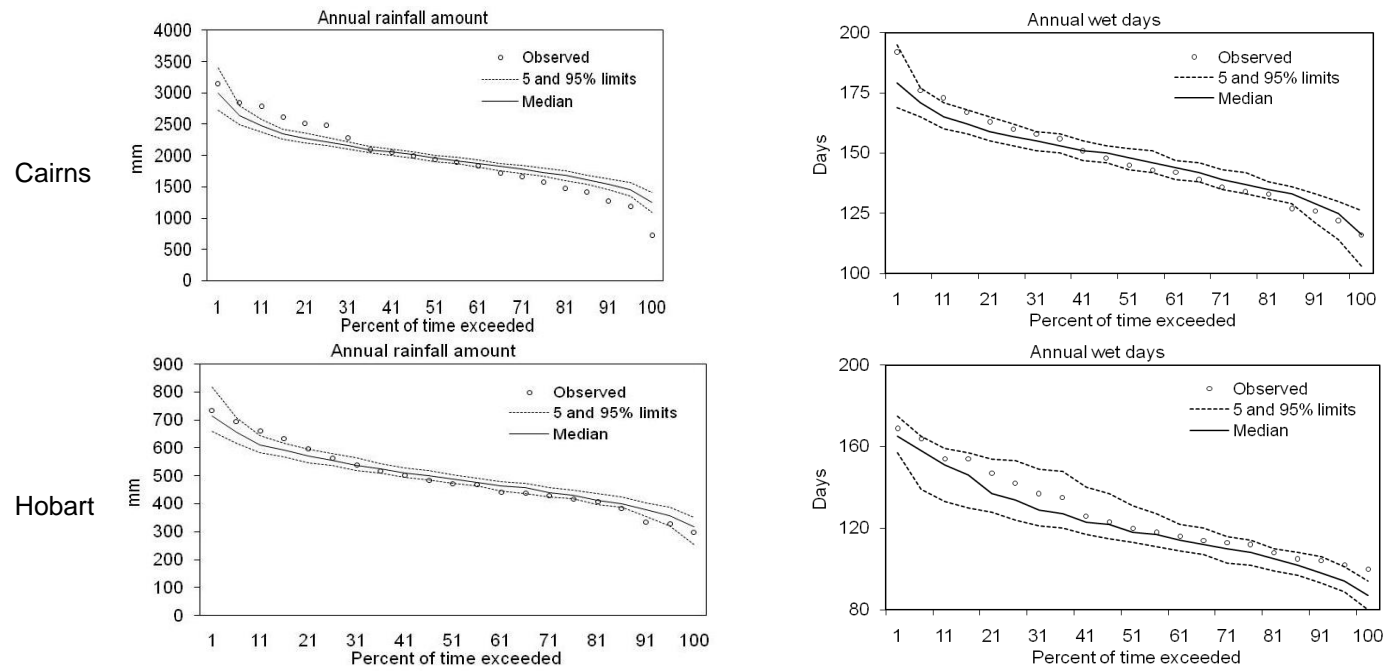


Figure 4.6: Distribution plots of observed and model simulated annual number of wet days and rainfall amount for five selected locations. Figure extracted from [Mehrotra et al., 2012].

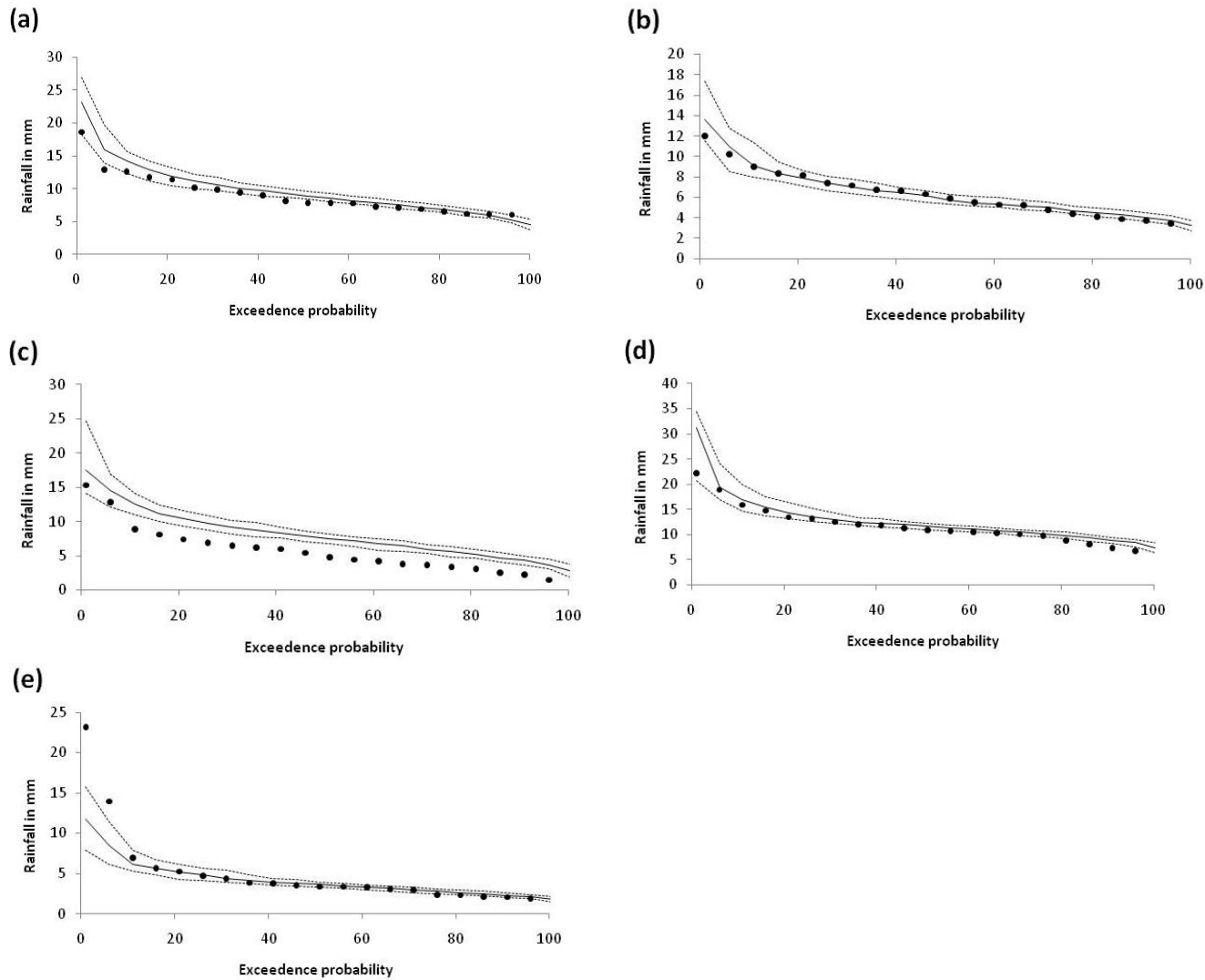


Figure 4.7: 6-minute annual maximum rainfall against exceedance probability for (a) Sydney, (b) Perth, (c) Alice Springs, (d) Cairns, and (e) Hobart. Black dots represents observed data, black solid line represents the median of 100 simulations, and black dotted lines represent the 5

and 95 percentile simulated values. Figure extracted from [Mehrotra et al., 2012].

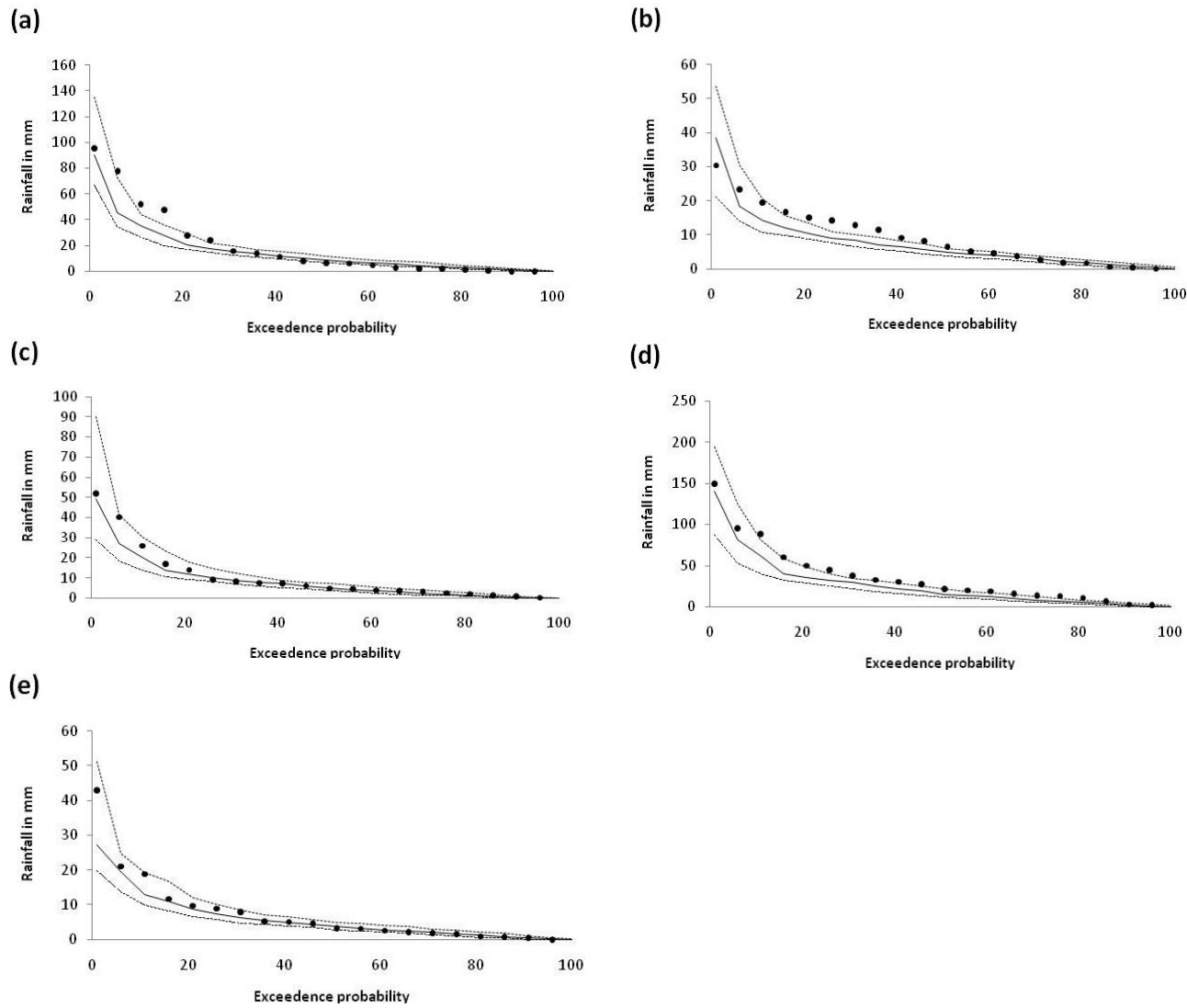


Figure 4.8: 6-hour antecedent rainfall prior to the 6-minute annual maximum storm burst plotted against exceedance probability for (a) Sydney,

(b) Perth, (c) Alice Springs, (d) Cairns, and (e) Hobart. Black dots represents observed data, black solid line represents the median of 100 simulations, and black dotted lines represent the 5 and 95 percentile simulated values. Figure extracted from [Mehrotra et al., 2012].

Table 4.4: Comparison of observed and simulated results for median annual maxima for different storm burst durations, and the antecedent rainfall prior to the 1 hour storm burst. The simulated median annual maxima represent the median of all 100 simulations.

	Sydney		Perth		Alice Springs		Cairns		Hobart	
	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)	Observed	Simulated (5 and 95% confidence bounds)
Annual maxima										
6 min	8.87	9.70 (8.97 – 10.3)	6.18	6.36 (5.87 – 6.83)	5.50	8.02 (7.24 – 8.60)	11.6	12.5 (11.9 – 13.6)	4.51	4.10 (3.70 – 4.73)
30 min	25.7	25.0 (23.6-27.2)	14.7	14.3 (13.3-15.5)	16.7	21.4 (19.4 – 23.6)	34.9	37.9 (36.2 – 39.6)	11.3	9.77 (8.90 – 10.8)
1 hr	35.4	33.4 (31.3-36.6)	18.8	18.1 (17.0 – 19.6)	22.1	27.3 (24.1 – 29.9)	51.7	55.1 (52.5 – 58.0)	14.6	12.9 (12.0 – 14.0)
3 hr	55.4	50.5 (47.6-55.8)	29.0	26.8 (25.0 – 28.8)	32.6	34.9 (31.1 – 38.9)	83.5	88.5 (84.0 – 93.4)	22.9	20.3 (18.8 – 21.9)
6 hr	72.3	64.7 (61.1-70.0)	36.3	33.8 (31.8-36.0)	39.6	41.0 (36.9 – 44.8)	113	116 (109 – 123)	30.3	26.7 (24.3 – 28.7)
12 hr	91.8	82.4 (77.2-88.7)	45.4	41.5 (39.1 – 44.4)	48.2	47.1 (42.3 – 51.5)	147	148 (138 – 158)	39.6	32.8 (30.0 – 35.1)
Antecedent rainfall prior to 1-hr burst (mm)										
6 hr	15.4	13.1 (10.5-17.6)	6.76	5.36 (4.20-6.96)	6.10	4.31 (2.59 – 6.32)	25.4	20.8 (16.0 – 27.6)	6.31	5.04 (3.76 – 6.71)
12 hr	22.7	17.4 (13.9-22.9)	9.63	7.02 (5.51-9.01)	7.98	5.71 (3.47 – 8.10)	32.2	27.2 (20.4 – 37.2)	9.10	6.14 (4.54 – 7.93)
24 hr	31.4	21.9 (17.9-28.1)	11.9	9.20 (7.28 – 11.6)	10.6	8.63 (6.03 – 11.3)	40.3	35.7 (26.5 – 46.9)	9.09	7.24 (5.67 – 9.79)
48 hr	38.4	25.7 (21.3-33.0)	12.5	12.9 (10.5 – 16.3)	15.5	11.9 (8.97 – 15.8)	54.9	47.6 (38.4 – 59.4)	10.2	8.85 (7.35 – 11.5)

4.5. Summary

The objectives of this chapter were to present a framework for the substitution of 'nearby' daily rainfall records in cases where daily rainfall at the target location is either unavailable or too short, and to evaluate the performance of the approach at a range of locations.

The stations which are likely to be statistically similar to the target location were identified using a range of predictors including location parameters and difference in elevation and difference in proximity to the coast. The model parameters were then estimated using the data at these locations, and the generated data using these parameters were transferred to the target location after an adjustment was made for annual average rainfall.

The procedure was tested in a fully cross-validation setting, so that simulations for target locations did not involve parameters fitted to rainfall observations at nearby locations. The results show that the method performs well in reproducing the number of wet days and rainfall amounts when there are a large number of daily stations in the vicinity of the target location, although performance did deteriorate for Alice Springs which is located in a data-sparse region of Australia. In contrast, the standard deviation of both wet days and amounts is typically undersimulated at all locations.

Interestingly, the sub-daily statistics, namely the annual maxima and the antecedent conditions, are well preserved, and the use of the regionalised daily model results in little deterioration in performance compared to using recorded daily data. This suggests the model is well suited for flood simulation which requires correct representation of peak rainfall and the moisture conditions in the hours and days leading up to the event.

Finally we conclude that although regionalised methods of rainfall generation enable the generation of rainfall time series at locations where no data is recorded, the models should not be expected to perform as well as models which are trained using high-quality at-site rainfall data. This is particularly the case where a location is climatologically anomalous compared to surrounding gauges, or where the density of nearby gauging stations is sparse, and highlights the value of maintaining a high-quality recording network. Nevertheless performance is generally reasonable across most statistics, particularly those necessary for flood estimation.

5. Comparison with DRIP and NSRP

5.1. Overview

In this chapter we compare the output of the continuous simulation model described in the previous three chapters with the Disaggregated Rectangular Intensity Pulse (DRIP) model of Heneker et al [2001] and a single site version of the Neyman-Scott Rectangular Pulse (NSRP) model of Cowpertwait et al [2002]. Both the DRIP and NSRP models were evaluated in detail by Frost et al [2004] at ten locations across Australia, and thus for this report we reproduce the analysis only for the at-site state-based method of fragments / daily modified Markov model presented in earlier chapters.

The pluviograph and daily rain gauge locations were summarised in Tables 1 and 2 of Frost et al [2004]. As with that study, one hundred replicates with the length of the original series were generated, with these sequences used to estimate confidence intervals for each of the statistics. Missing years were discarded from the analysis, with the average fraction missing data over the period of record (excluding those years which were missing in their entirety) being around 5%. As the generated sequences do not contain missing data, the generated sequences are slightly longer than the observed sequences, and the missing data may slightly affect the observed statistics. Nevertheless, each of the continuous simulation methods – namely the state-based method of fragments, DRIP and NSRP – are treated similarly, so that this is not likely to impact on the conclusions.

Both the DRIP and NSRP models were based on at-site data so that, to ensure the models were comparable, we considered only the results of the state-based method of fragments disaggregation model using at-site pluviograph data described in Chapter 2, together with a daily rainfall model generated using the modified Markov model described in Chapter 4. Although some deterioration might be expected when moving to a regionalised framework, this issue was covered in detail using the test statistics described in Chapters 2-4 and is not repeated here.

5.2. Sub-daily results

The results from our state-based method of fragments / daily modified Markov model are summarised in Table 5.1, with figures for all statistics for all months at all ten locations provided in Appendix A. The figures are directly comparable with Figures A1-A9 and A12 (for the DRIP model) and Figures B1-B9 and B12 (for the NSRP model), which were provided in Frost et al [2004]. The only statistics not considered were the dryspell-wetspell correlation and the wetspell-dryspell correlation, as we were not able to reproduce the statistics for the observed data. The confidence bands for these statistics in Frost et al [2004] were generally very wide, suggesting that limited information is provided on model performance from these statistics. Therefore these statistics were not considered here.

The results in Table 5.1 show generally good performance for the state-based method of fragments disaggregation, with most statistics being very similar to the DRIP model. The main exception was the treatment of extreme rainfall in terms of the distribution of annual maxima,

with the state-based method of fragments showing closer correspondence with the observations compared to DRIP.

Table 5.1: Evaluation statistics for state-based method of fragments model against observed pluviograph data at 10 locations (detailed results provided in Appendix A). A comparison with the DRIP and NSRP results described in Frost et al [2004] is also provided.

Statistic	Performance	Comparison with DRIP/NSRP	Figure
Dry probability (%)	Good – slight downward bias	Similar to DRIP; less variable than NSRP	A1
Mean rainfall (mm)	Good	Similar to DRIP/NSRP	A2
Standard deviation of rainfall (mm)	Good	Similar to DRIP/NSRP	A3
Coefficient of skew of rainfall (-)	Good	Similar to DRIP/NSRP	A4
Lag one autocorrelation	Slight underestimation all durations	DRIP underestimates 1hr and overestimates 24hr; NSRP performs well across all durations	A5
Dry spell duration mean (hr)	Good	Similar to DRIP; significantly improves on NSRP	A6
Dry spell duration standard deviation (hr)	Good	Similar to DRIP; significantly improves on NSRP	A7
Wet spell duration mean (hr)	Overestimates 24hr duration	Similar to DRIP; slight improvement to NSRP	A8
Wet spell duration standard deviation (hr)	Overestimates 24hr duration	Similar to DRIP; slight improvement to NSRP	A9
Annual maximum intensity distribution plot (Intensity-Frequency-Duration)	Good – possible upward bias for low non-exceedance probability events however this may be due to quality of raw pluviograph data	Superior to DRIP; similar to NSRP.	A10

An interesting result was that, for cases where biases were present, they tended to be in the same direction, of the same magnitude, during the same months and at the same locations, as was the case for DRIP. The state-based method of fragments and DRIP are two fundamentally different approaches, and thus it is possible that some of the observed biases may be due to the issue of missing data described previously, as the stochastically generated data by definition does not contain any missing records. Nevertheless, given the generally satisfactory

performance of both DRIP and the state-based method of fragments across almost all the statistics, this issue was not explored further.

Due to the similarity between the performance of the state-based method of fragments and DRIP, the conclusions of Frost et al [2004] with regards to the difference between DRIP and NSRP also hold for the difference between the state-based method of fragments and NSRP. In particular, both the state-based method of fragments and DRIP are superior in representing dry and wet spell duration means and standard deviations, whereas NSRP is superior in terms of lag-one autocorrelation. NSRP performs better than DRIP in terms of the annual maximum intensity distribution, with results generally comparable to the state-based method of fragments. In general, however, all methods perform satisfactorily in reproducing sub-daily statistics.

5.3. Daily rainfall results

The results based on the daily statistics are summarised in Table 5.2, with figures for all statistics at all ten locations provided in Appendix B. The figures have been developed to be directly comparable with Figures A13-A22 (DRIP) and Figures B13-B22 (NSRP) in Frost et al [2004].

Once again the results show generally good performance across most statistics, with similar or better performance to both DRIP and NSRP. A weakness of the NSRP identified in Frost et al [2004] was a major overestimation of dry spell means and standard deviations at most locations, as well as an overestimation of wet spell means. The longer dry and wet spells simulated by the NSRP model suggest that a greater clustering of wet spells, which can have an impact on whether this method properly simulates antecedent rainfall conditions prior to the storm event.

The main weakness of the Markov model is a slight underestimation of annual variability, with the method underestimating the probability of the driest and wettest years. The inclusion of the previous 365 day's rainfall as a predictor in the Markov model was designed to address this issue; however additional work is required to completely resolve this issue. Although the issue appears to be systematic (occurring in most of the 10 locations studied), the magnitude of the underestimation is generally low, with the observations usually falling within the 90% confidence intervals.

Table 5.2: Evaluation statistics for Daily Markov model against observed daily rainfall data at 10 locations (detailed results provided in Appendix B). A comparison with the DRIP and NSRP results described in Frost et al [2004] is also provided.

Statistic	Performance	Comparison with DRIP/NSRP	Figure
Annual rainfall distribution plot	Reasonable; slight underestimation of variability	DRIP and NSRP are slightly superior although both significantly underestimate Perth rainfall across all most exceedance probabilities. NSRP generally performs best.	B1
Dry probability	Good – slight overestimation for Hobart	Slight improvement on DRIP; significant improvement on NSRP	B2
Mean rainfall (mm)	Good	Similar to DRIP and NSRP	B3
Standard deviation of rainfall (mm)	Good	Similar to DRIP and NSRP	B4
Coefficient of skew of rainfall (-)	Good	Similar to DRIP and NSRP	B5
Lag one autocorrelation	Underestimates all locations	DRIP also underestimates autocorrelation by similar magnitude; NSRP performs best	B6
Dry spell duration mean (days)	Slight overestimation for some location	Similar performance to DRIP. Both outperform NSRP.	B7
Dry spell duration standard deviation (days)	Good	Similar performance to DRIP. Both outperform NSRP.	B8
Wet spell duration mean (days)	Slightly underestimates	Slight improvement over DRIP, which also underestimates but by a greater amount. Significant improvement over NSRP.	B9
Wet spell duration standard deviation (days)	Good	Performs better than both DRIP and NSRP.	B10

6. Discussion and conclusions

This report describes the outcomes from the second stage of Project 4 – Continuous Simulation at a Point, which is conducted as part of the revision of Australian Rainfall and Runoff. The emphasis of this stage was to: (1) finalise the development of the regionalised state-based method of fragments approach as well as the development of a regionalised daily rainfall generation model; and (2) assess the performance of the state-based method of fragments using the same statistics and locations that were used in Frost et al [2004], to enable direct comparison with the DRIP and NSRP models.

6.1. Regionalised state-based method of fragments and modified Markov models

The complete at-site state-based method of fragments methodology was described in Chapter 2, which was based on generating near-continuous rainfall time series that aim to ensure the rainfall patterns prior to the major storm event are represented in a realistic manner. Testing using both extreme statistics (such as the Intensity-Frequency-Duration, or IFD, relationships) and antecedent rainfall-based statistics show this model performs well, although it is heavily reliant on having access to long records of sub-daily rainfall at the location of interest. Furthermore, by only sampling from historical sub-daily rainfall at the location of interest, the diversity of possible rainfall events which could fall on the catchment is likely to be under-simulated.

To address these issues, the regionalised version of this model was then presented in Chapter 3, and involves sampling from a set of nearby sub-daily rain gauges, conditional to at-site daily rainfall records. Model evaluation based on the two statistics most relevant for flood simulation: namely the Intensity-Frequency-Duration (IFD) statistics and the associated antecedent rainfall, showed the model performed generally well in most locations. The primary exception was for Alice Springs, for which the regionalised state-based method of fragments model resulted in an overestimation of the IFD statistics. This is attributed to the lack of sufficient nearby pluviograph records, such that the algorithm was forced to draw from geographically distant pluviograph data to develop the continuous sequences. Such biases could also be expected from other regionalised approaches, however, as the estimation of model parameters would also be based on the availability of sufficient pluviograph records. By contrast, the regionalised state-based method of fragments approach identified a likely recording error in Hobart Airport, such that the synthetic series is likely to be more representative of the IFD statistics at Hobart Airport than the historical data itself.

A regionalised Markov model was then described in Chapter 4, which was designed to enable the generation of continuous sequences of daily rainfall at any location in Australia regardless of the presence of gauged data. Testing of this method showed generally reasonable performance, although a slight underestimation of the standard deviation of both wet days and amounts was found at all locations. Nevertheless, combining this modified Markov model with the regionalised state-based method of fragments approach, it was found that the IFD and antecedent rainfall statistics were well represented at all locations, once again with the exception of Alice Springs.

6.2. Comparison with DRIP and NSRP

The at-site version of the state-based method of fragments disaggregation and daily rainfall modified Markov models were then compared with the DRIP and NSRP models. In general, the model performed similarly to the DRIP model, and with both models performing better than the NSRP model. Whereas the state-based method of fragments model outperformed DRIP in simulating annual maximum storm bursts for sub-daily durations, DRIP appeared superior in simulating the distribution of total annual rainfall. Nevertheless these differences were in general minor, and both models appear adequate for use in generating continuous rainfall data at the sub-daily timescale.

Finally, although several alternative regional methods are available for continuous simulation, such as a regionalised version of DRIP by Jennings et al [2009] and regionalised versions of the Poisson cluster models by Gyasi-Agyei [1999], Gyasi-Agyei and Parvez Bin Mahbub [2007] and Cowpertwait and O'Connell [1997], there is still a significant research requirement in performance of differing classes of regionalised models. Testing conducted on the regionalised state-based method of fragments and Markov models described in this report show fairly limited deterioration in model performance as the model becomes increasingly regionalised. The principal exceptions are for regions where there are limited nearby daily rainfall or pluviograph gauges, or for regions what are very climatologically different from the nearby gauged locations (for example in mountainous areas). Unfortunately other regionalised approaches, which involve estimating model parameters based on nearby gauges, would be likely to suffer from similar limitations. This highlights that despite the rapid advance in approaches in regionalising rainfall generation, there remains a continued need for maintaining a high-quality observational network of point rainfall data.

6.3. Recommendations and outstanding issues

Based on the outcomes described in this report, it is concluded that there are a number of approaches now available which allow for the generation of extended sequences of continuous rainfall at point locations. Comparison of these approaches at locations where significant at-site data is available shows similar performance, particularly for the state-based method of fragments and DRIP models. Nevertheless, a range of outstanding issues remain, which are summarised as follows:

- 1) The regionalised state-based method of fragments and modified Markov models have not been compared with other regionalised approaches such as the regionalised version of DRIP by Jennings et al [2009] and regionalised versions of the Poisson cluster models by Gyasi-Agyei [1999], Gyasi-Agyei and Parvez Bin Mahbub [2007] and Cowpertwait and O'Connell [1997]. Such a comparison may be beneficial to facilitate wider uptake of these methods.
- 2) Given the complexity of continuous simulation models, wide update of such methods by engineering practice is unlikely in the absence of software products. Although software is freely available for the at-site version of DRIP, software is lacking for most of the remaining methods.

- 3) There is a linkage between the outcomes of this project and ARR Revision Project 8: Use of Continuous Simulation for Design Flow Estimation. Testing of different continuous rainfall simulation methods in their capacity to simulate design flows may be beneficial to evaluate the implications of moving to a continuous simulation approach for design flood estimation.
- 4) A multi-site extension to the modified Markov model was described by [*Mehrotra and Sharma, 2007a*], in which stochastic sequences of daily rainfall can be generated at multiple point locations in a manner that preserves spatial dependence. This could be extended to a regionalised setting by evaluating how the spatial dependence varies spatially, and may also serve as an alternative basis for estimating areal reduction factors.
- 5) All the methodologies here have been developed to generate continuous rainfall sequences based on historical climate conditions. One of the principal advantages of continuous simulation for future climate is that all changes to the character of precipitation (mean and extreme rainfall, seasonality, intermittency, etc) can be accommodated. The capacity of extending the approaches described in this report for simulating precipitation sequences that are representative of future climate is likely to be of significant benefit for future design flood estimation.

References

- Beven, K. J. (2002), *Rainfall-Runoff Modelling: The Primer*, 360 pp., John Wiley & Sons.
- Blazkova, S., and K. Beven (2002), Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty), *Water Resources Research*, 38(8).
- Boughton, W., and O. Droop (2003), Continuous simulation for design flood estimation - a review, *Environmental Modelling & Software*, 18(4), 309-318.
- Bowler, N. E., C. E. Pierce, and A. W. Seed (2006), STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP, *Quarterly Journal of the Royal Meteorological Society*, 132, 2127-2155.
- Brandsma, B., and A. T. Buishand (1998), Simulation of extreme precipitation in the Rhine basin by nearest neighbour resampling
Hydrological Earth Systems Science 2, 195-209.
- Buishand, A. T. (1978), Some remarks on the use of daily rainfall models, *Journal of Hydrology*, 36, 295-308.
- Buishand, A. T., and B. Brandsma (2001), Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest neighbour resampling, *Water Resources Research*, 37, 2761-2776
- Cameron, D., K. Beven, J. Tawn, and P. Naden (2000), Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation), *Hydrology and Earth System Sciences*, 4(1), 23-34.
- 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 System Sciences*, 1, 71-80.
- Cowpertwait, P. S. P., C. G. Kilsby, and P. E. O'Connell (2002), A space-time Neyman-Scott model of rainfall: Empirical analysis of extremes, *Water Resources Research*, 38(8).
- Cowpertwait, P. S. P., P. E. O'Connell, A. V. Metcalfe, and J. A. Mawdsley (1996), Stochastic point process modelling of rainfall. II. Regionalisation and disaggregation, *Journal of Hydrology*, 175, 47-65.
- Fasano, G., and A. Franceschini (1987), Monthly Notices of the Royal Astronomical Society, edited, pp. 155-170.
- Frost, A. J., R. Srikanthan, and P. S. P. Cowpertwait (2004), Stochastic generation of rainfall data at subdaily timescales: a comparison of DRIP and NSRP. *Rep. 04/9*.
- Gabriel, K. R., and J. Newmann (1962), A Markov chain model for daily rainfall occurrence at Tel Aviv, *Quarterly Journal of the Royal Meteorological Society*, 88, 90-95.
- Guennia, L., and M. F. Hutchinson (1998), Spatial interpolation of the parameters of a rainfall model from ground-based data, *Journal of Hydrology*, 212-213, 335-347.
- Gyasi-Agyei, Y. (1999), Identification of regional parameters of a stochastic model for rainfall disaggregation, *Journal of Hydrology*, 223, 148-163.
- Gyasi-Agyei, Y., and G. R. Willgoose (1997), A hybrid model for point rainfall modelling, *Water Resources Research*, 33(7), 1699-1706.
- Gyasi-Agyei, Y., and G. R. Willgoose (1999), Generalisation of a hybrid model for point rainfall, *Journal of Hydrology*, 219(3-4), 218-224.
- Gyasi-Agyei, Y., and S. M. Parvez Bin Mahbub (2007), A stochastic model for daily rainfall disaggregation into fine time scale for a large region, *Journal of Hydrology*, 347, 358-370.

- Harrold, T. I., A. Sharma, and S. J. Sheather (2003a), A nonparametric model for stochastic generation of daily rainfall amounts, *Water Resources Research*, 39(12).
- Harrold, T. I., A. Sharma, and S. J. Sheather (2003b), A nonparametric model for stochastic generation of daily rainfall occurrence, *Water Resources Research*, 39(12), 1343.
- Hastie, T., R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*.
- Heneker, T. M., M. Lambert, and G. Kuczera (2001), A point rainfall model for risk-based design, *Journal of Hydrology*, 251, 65-87.
- Jennings, S., M. Lambert, and G. Kuczera (2009), A high resolution point rainfall model calibrated to short pluviograph or daily rainfall data, *Journal of Hydrology* - submitted.
- Johnson, G. L., C. Daly, G. H. Taylor, and C. L. Hanson (2000), Spatial variability and interpolation of stochastic weather simulation model parameters, *Journal of Applied Meteorology*, 39, 778-796.
- Koutsoyiannis, D. (2003), Rainfall Disaggregation Methods: Theory and Applications, in *Workshop on Statistical and Mathematical Methods for Hydrological Analysis* edited, Rome.
- Kuczera, G., M. Lambert, T. M. Heneker, S. Jennings, A. J. Frost, and P. J. Coombes (2006), Joint probability and design storms at the crossroads, *Australian Journal of Water Resources* 10(1).
- Kyriakidis, P. C., N. L. Miller, and J. Kim (2004), A spatial time series framework for simulating daily precipitation at regional scales, *Journal of Hydrology*, 297, 236-255.
- Lall, U., and A. Sharma (1996), A nearest neighbor bootstrap for time series resampling, *Water Resources Research*, 32(3), 679-693.
- Lall, U., B. Rajagopalan, and D. G. Tarboton (1996), A nonparametric wet/dry spell model for resampling daily precipitation, *Water Resources Research*, 32, 2803-2823.
- Lamb, R., and A. L. Kay (2004), Confidence intervals for a spatially generalized, continuous simulation flood frequency model for Great Britain, *Water Resources Research*, 40(7).
- Mehrotra, R., and A. Sharma (2005), A nonparametric nonhomogenous hidden Markov model for downscaling of multi-site rainfall occurrences, *Journal of Geophysical Research*, 110(D16108).
- Mehrotra, R., and A. Sharma (2006a), Conditional resampling of hydrologic time series using multiple predictor variables: A k-nearest neighbour approach, *Advances in Water Resources*, 29, 987-999.
- Mehrotra, R., and A. Sharma (2006b), A nonparametric stochastic downscaling framework for daily rainfall at multiple locations, *Journal of Geophysical Research*, 111(D15101).
- Mehrotra, R., and A. Sharma (2007a), A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability, *Journal of Hydrology*, 335, 180-193.
- Mehrotra, R., and A. Sharma (2007b), Preserving low-frequency variability in generated daily rainfall sequences, *Journal of Hydrology*, 345, 102-120.
- Mehrotra, R., and A. Sharma (2010), Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment, *Water Resources Research*, 46(W07526).
- Mehrotra, R., S. Westra, A. Sharma, and R. Srikanthan (2012), Continuous Rainfall Simulation: 2 - A regionalised daily rainfall generation approach, *Water Resources Research In press*.
- Onof, C., R. E. Chandler, A. Kakou, P. Northrop, H. S. Wheeler, and V. Isham (2000),

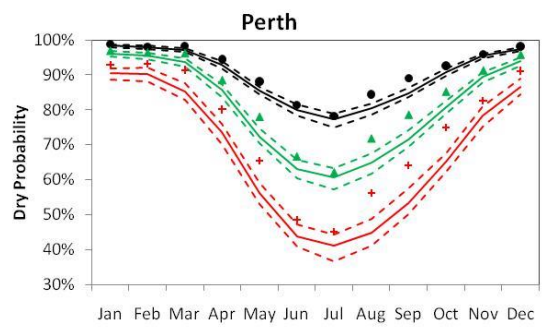
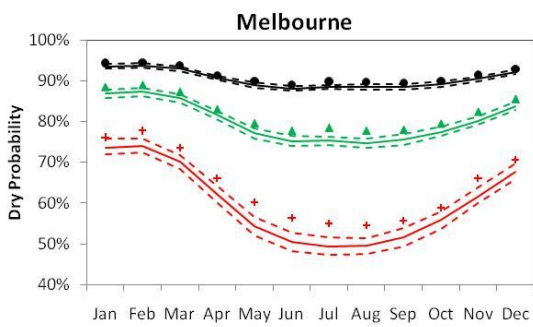
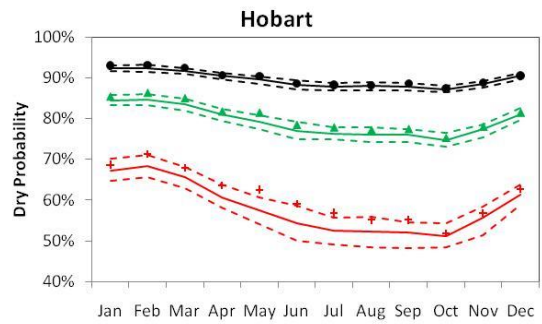
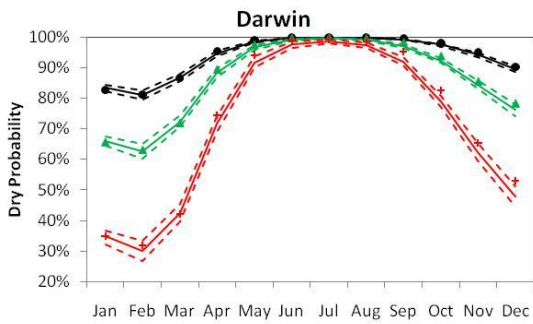
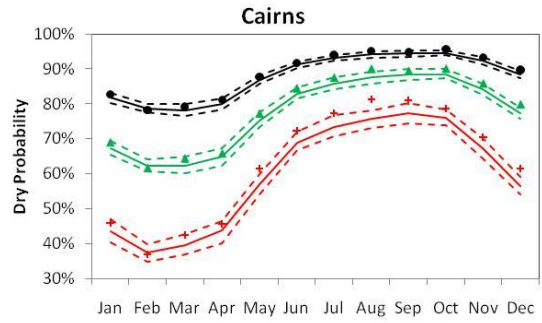
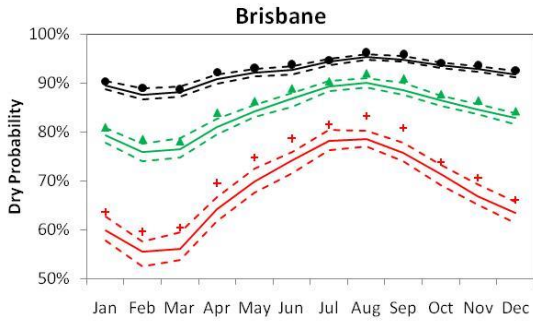
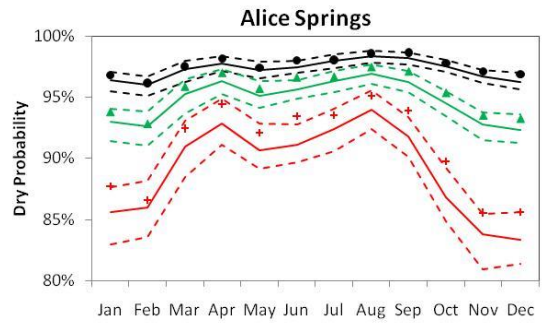
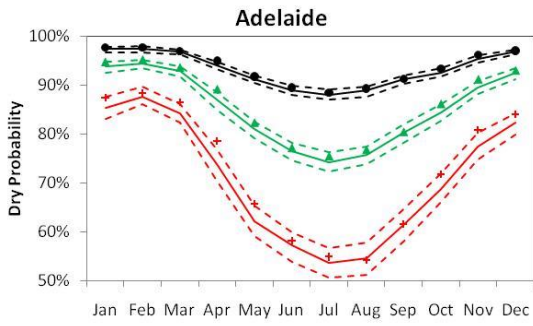
- Rainfall modelling using Poisson-cluster processes: a review of developments, *Stochastic Environmental Research and Risk Assessment*, 14(6), 384-411.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992), *Numerical Recipes in Fortran - The Art of Scientific Computing*, Second Edition ed., 963 pp., Cambridge University Press, Cambridge.
- Pui, A., A. Lall, and A. Sharma (2010a), How does the Interdecadal Pacific Oscillation affect Design Floods in Eastern Australia? , *Water Resources Research*, under review (submitted Feb 2010).
- Pui, A., S. Westra, and A. Sharma (2010b), How does the El Nino-Southern Oscillation (ENSO) and other climate modes affect multi-scale temporal rainfall variability in Australia?, *in press Monthly Weather Review*.
- Rajagopalan, B., and U. Lall (1999), A nearest neighbour bootstrap resampling scheme for resampling daily precipitation and other weather variables, *Water Resources Research*, 35(10), 3089-3101.
- Rajagopalan, B., U. Lall, and D. G. Tarboton (1996), A nonhomogeneous Markov model for daily precipitation simulation, *Journal of Hydrologic Engineering*, 1(1), 33-40.
- Scott, D. W. (1992), *Multivariate Density Estimation - Theory, Practise and Visualization*, 317 pp., John Wiley and Sons, Inc, New York.
- Sharma, A. (2000), Seasonal to interannual rainfall probabilistic forecasts for improved water supply management. Part 3. A nonparametric probabilistic forecast model, *Journal of Hydrology*, 239, 249-258.
- Sharma, A., and R. O'Neill (2002), A nonparametric approach for representing interannual dependence in monthly streamflow sequences, *Water Resources Research*, 38(7).
- Sharma, A., and R. Srikanthan (2006), Continuous rainfall simulation: a nonparametric alternative, in *30th Hydrology and Water Resources Symposium*, edited, Launceston, Tasmania.
- Sharma, A., and R. Mehrotra (2010), Rainfall Generation, in *Rainfall: State of the Science*, edited by F. Testik and M. Gebremichael, p. 32, American Geophysical Union.
- Sharma, A., D. G. Tarboton, and U. Lall (1997), Streamflow simulation: a nonparametric approach, *Water Resources Research*, 33(2), 291-308.
- Sivakumar, B., and A. Sharma (2008), A cascade approach to continuous rainfall generation at point locations, *Stochastic Environmental Research and Risk Assessment (SERRA)*(DOI 10.1007/s00477-007-0145-y), 1-9.
- Snavidze, G. G. (1977), *Mathematical Modeling of Hydrologic Series*, Water Resources Publications, Littleton, Colorado.
- 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., and G. G. S. Pegram (2009), A nested multisite daily rainfall stochastic generation model, *Journal of Hydrology*, 371, 142-153.
- Todorovic, P., and D. A. Woolhiser (1975), A stochastic model of n-day precipitation, *Journal of Applied Meteorology*, 14, 17-24.
- Weinmann, P. E., A. Rahman, T. M. T. Hoang, E. M. Laurenson, and R. J. Nathan (2002), Monte Carlo Simulation of Flood Frequency Curves from Rainfall - the Way Ahead, *Australian Journal of Water Resources*, 6(1).
- Westra, S., and A. Sharma (2010), *Australian Rainfall and Runoff Revision Project 4, Stage 1: Continuous Rainfall Sequences at a Point*, Engineers Australia.
- Westra, S., R. Mehrotra, A. Sharma, and R. Srikanthan (2012), Continuous Rainfall Simulation: 1 - A regionalised sub-daily disaggregation approach, *Water Resources Research*, *In press*.

Wilks, D. S. (2008), High-resolution spatial interpolation of weather generator parameters using local weighted regressions, *Agricultural and Forest Meteorology*, 148, 111-120.

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.

Appendix A: State-Based Method of Fragments sub-daily evaluation statistics

For all statistics plotted, the observed values are plotted as a point value. The 1, 6 & 24 hr results use the \circ (black), Δ (green) and $+$ (red) symbols, respectively. Simulated median (thick line) and 5 and 95% confidence levels (dotted line) are also plotted.



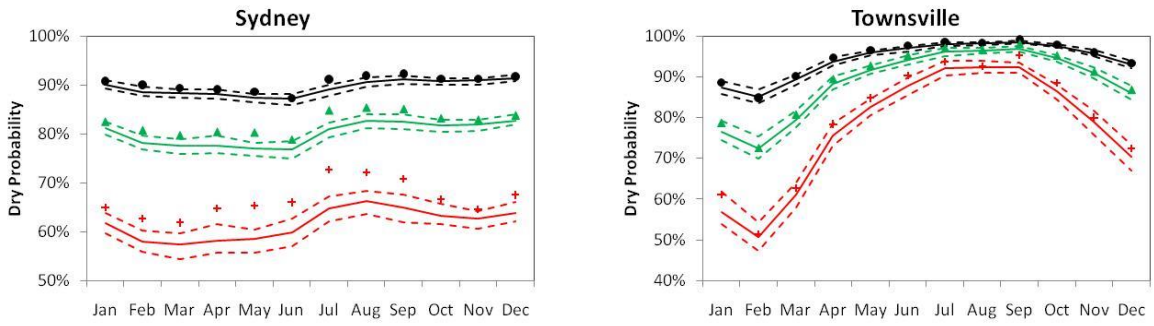
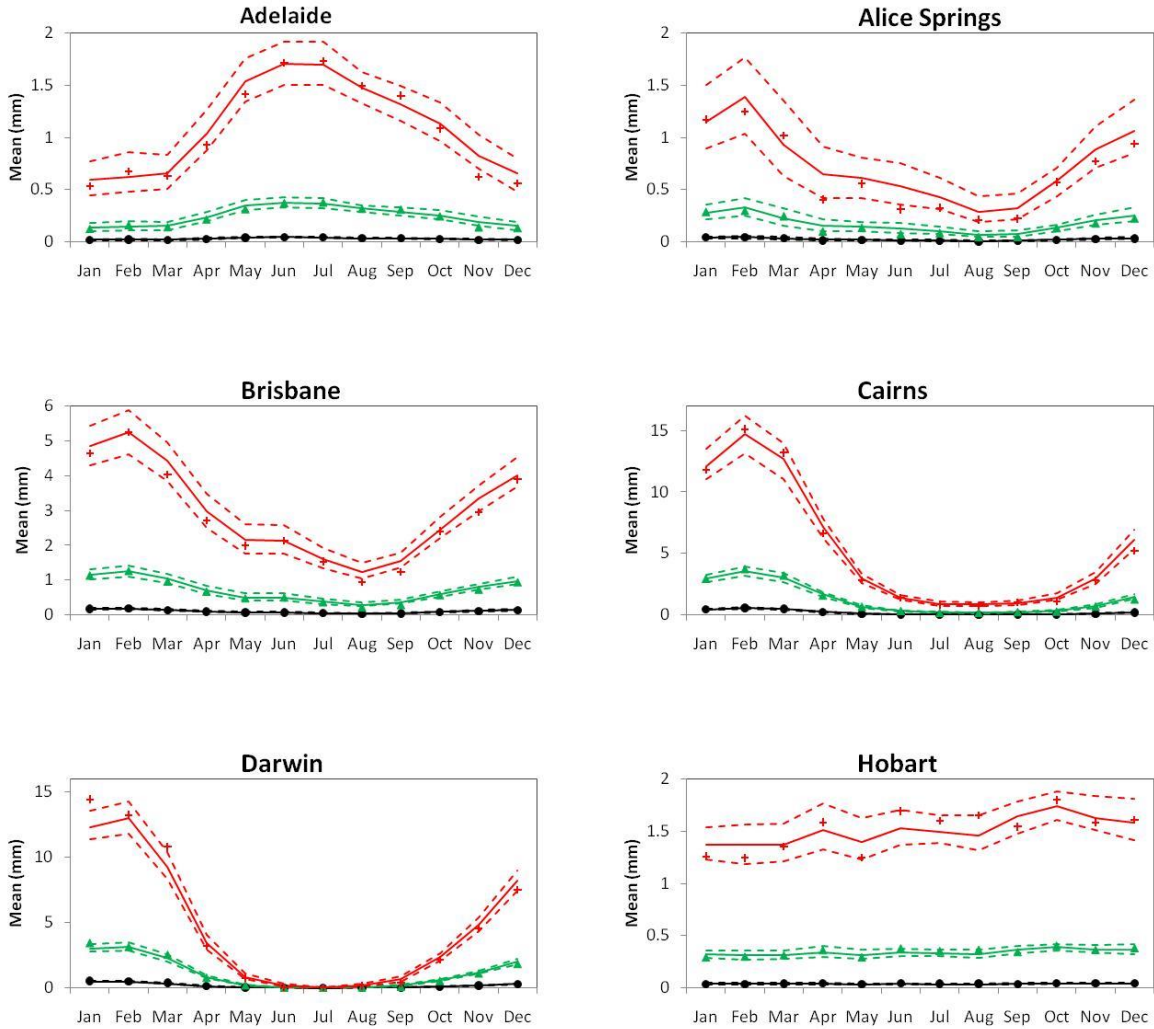


Figure A1: Method of fragments dry probability: 1, 6 & 24 hr statistics



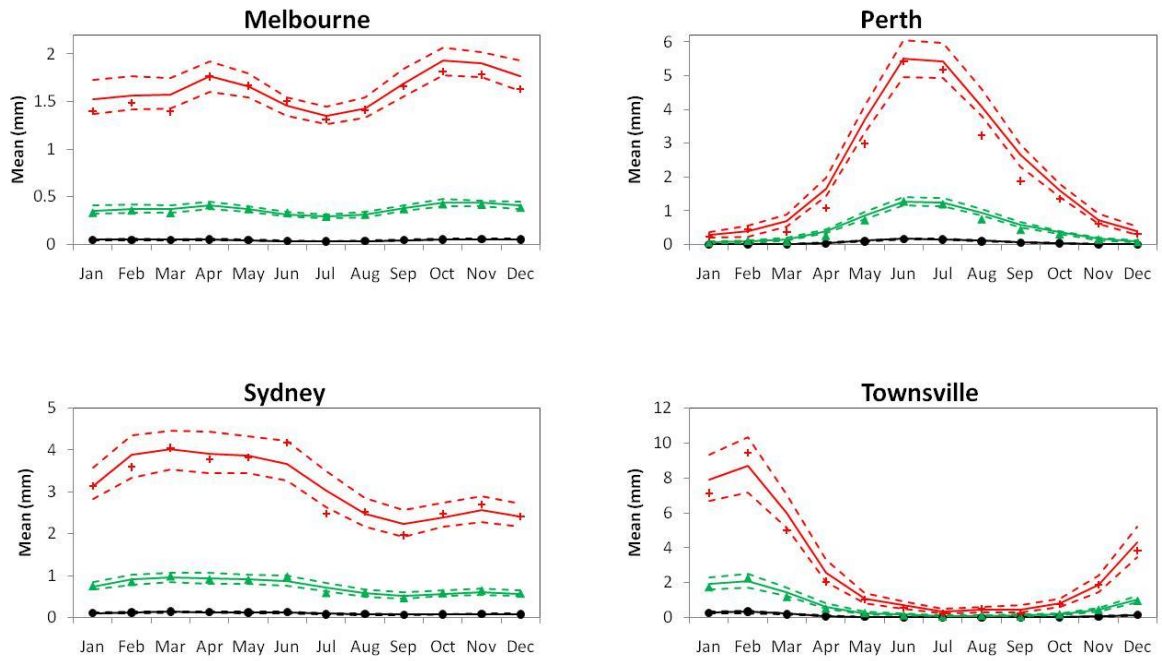
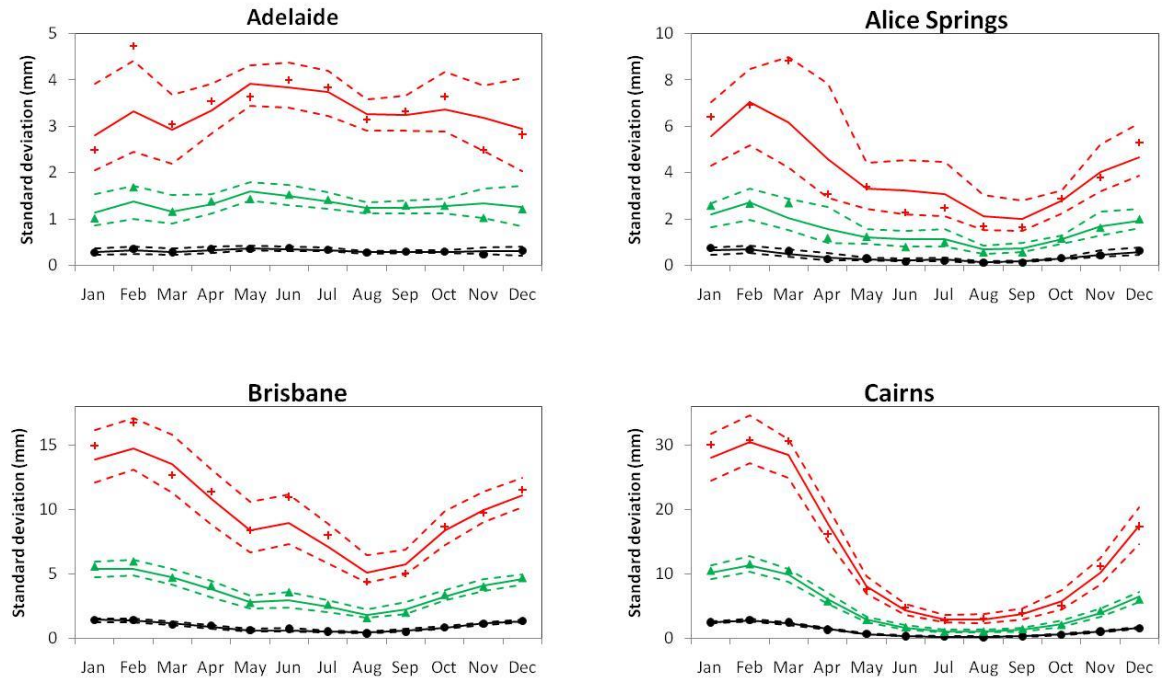


Figure A2: Method of fragments monthly mean: 1, 6 & 24 hr statistics



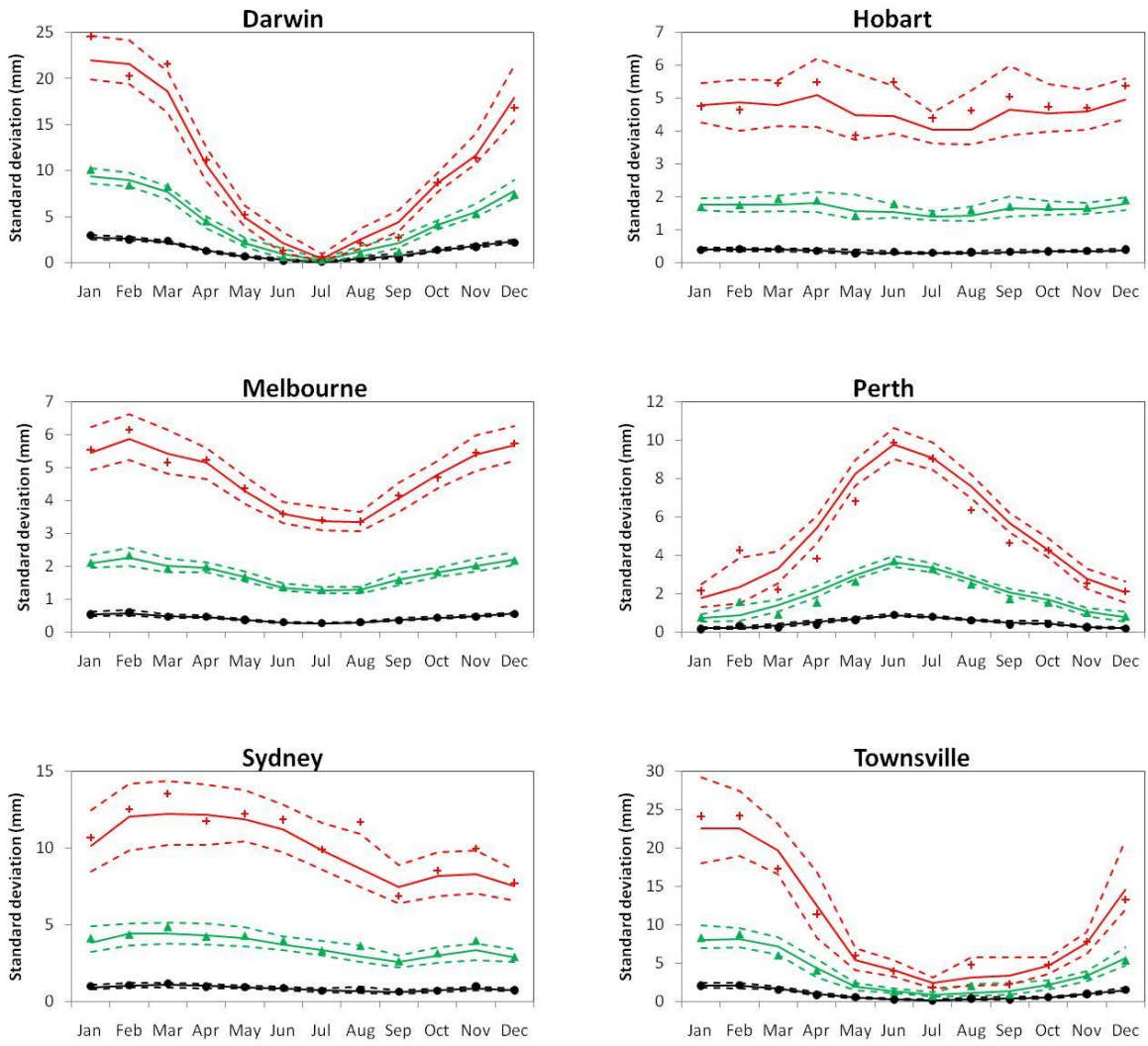
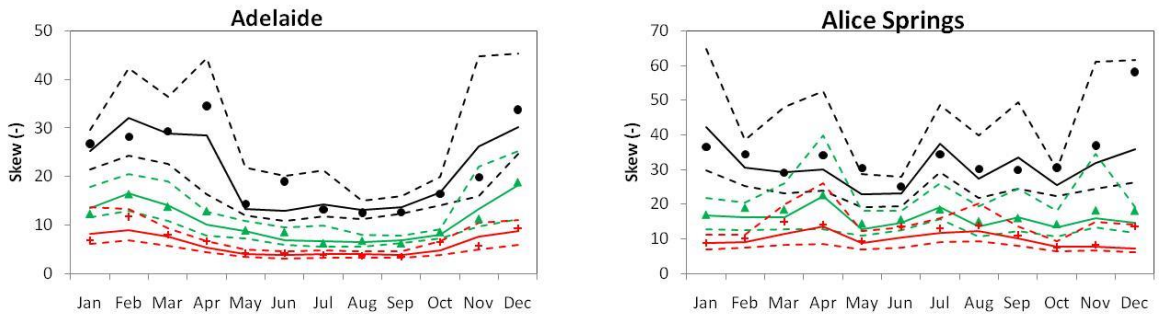


Figure A3: Method of fragments monthly standard deviation: 1, 6 & 24 hr statistics



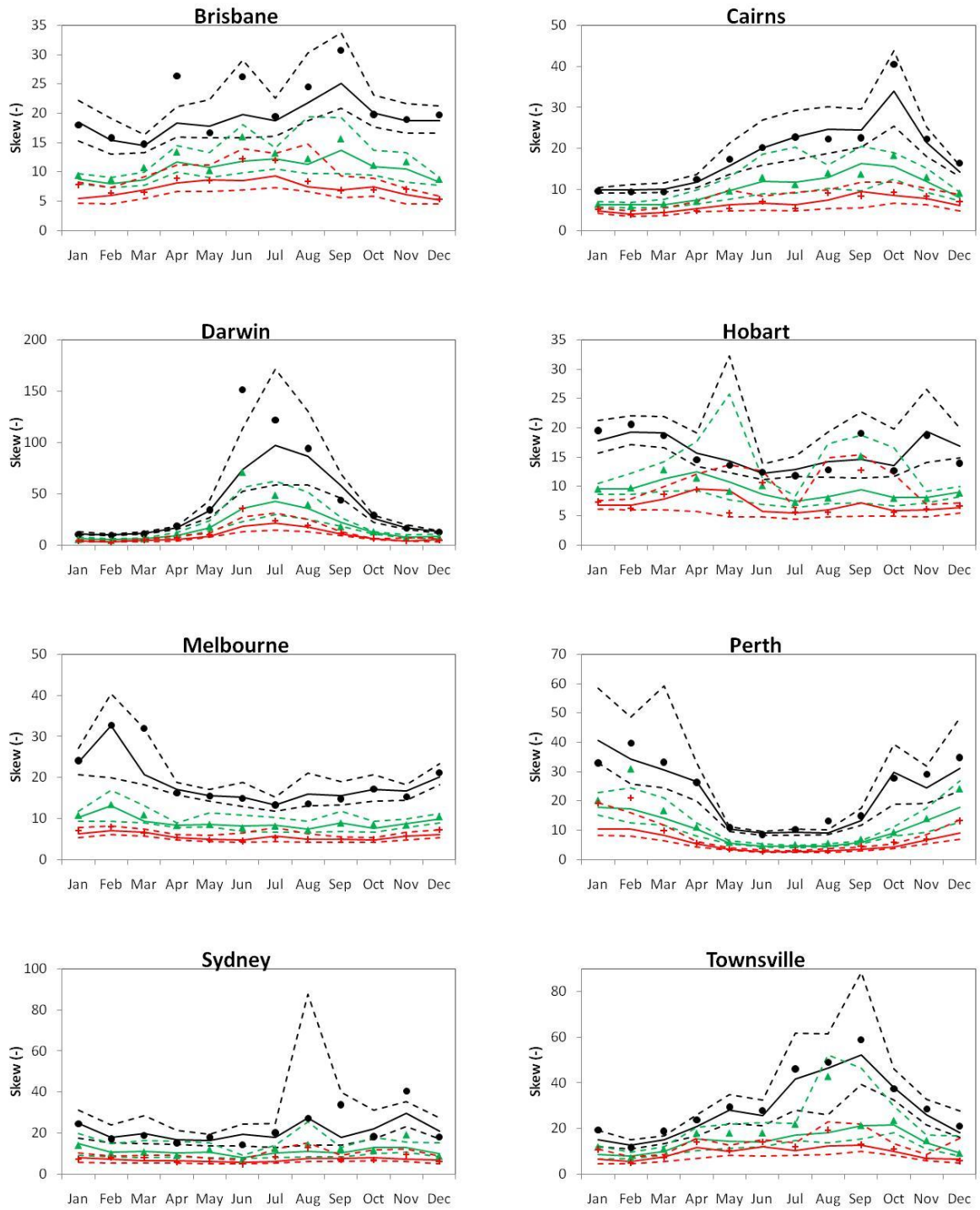
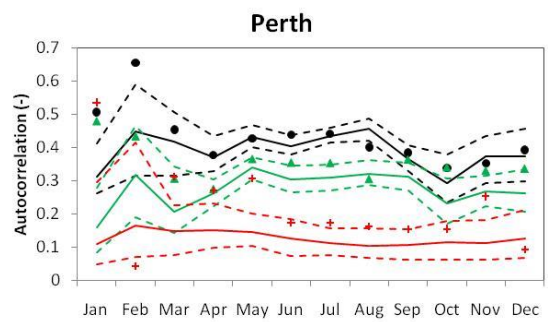
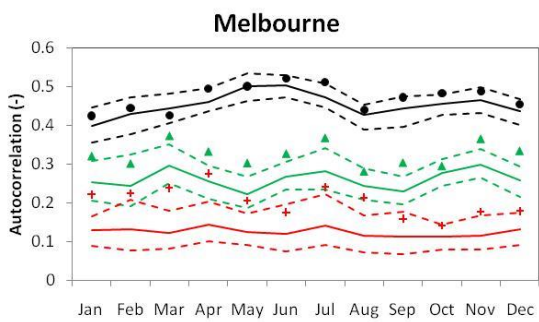
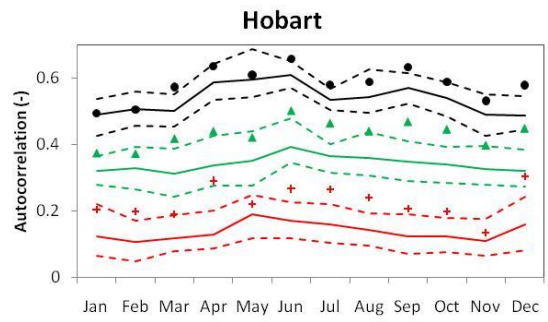
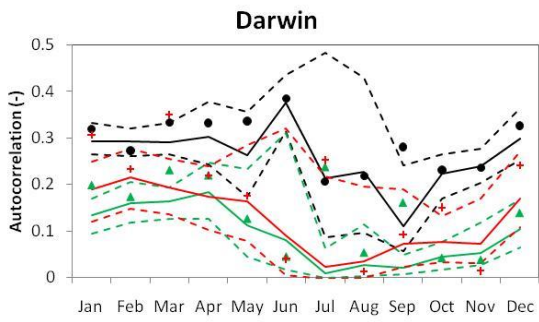
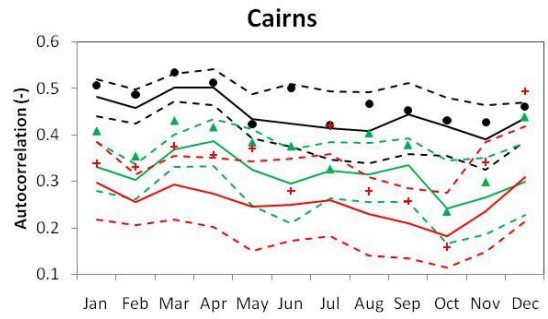
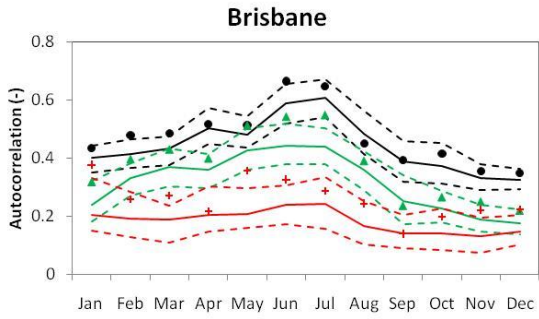
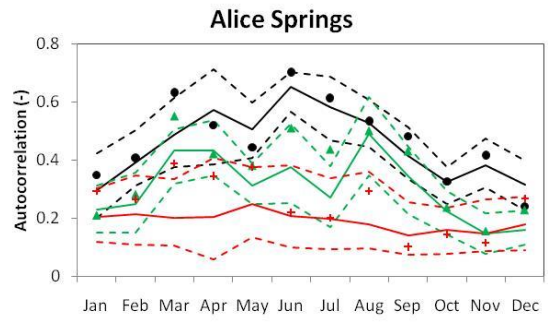
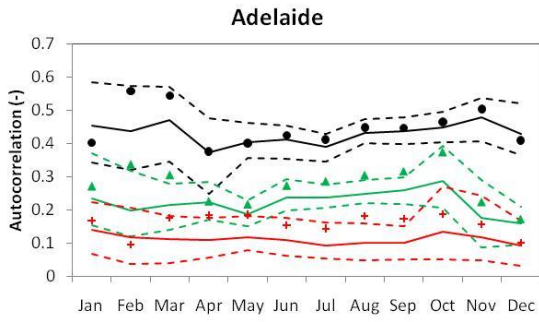


Figure A4: Method of fragments monthly skew: 1, 6 & 24 hr statistics



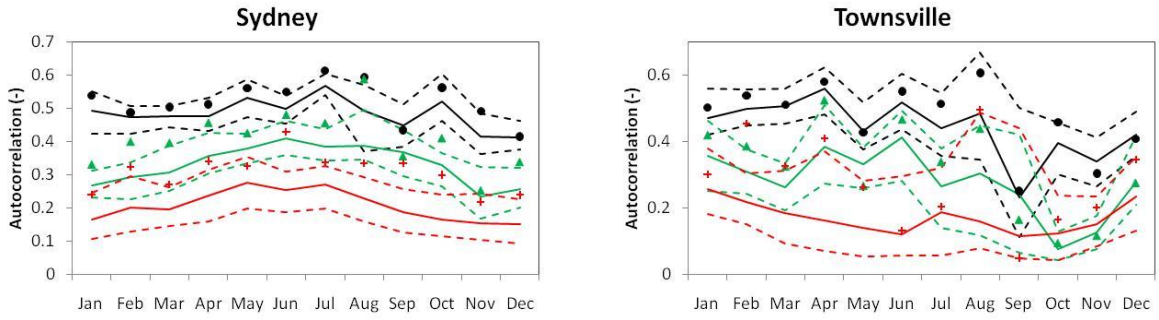
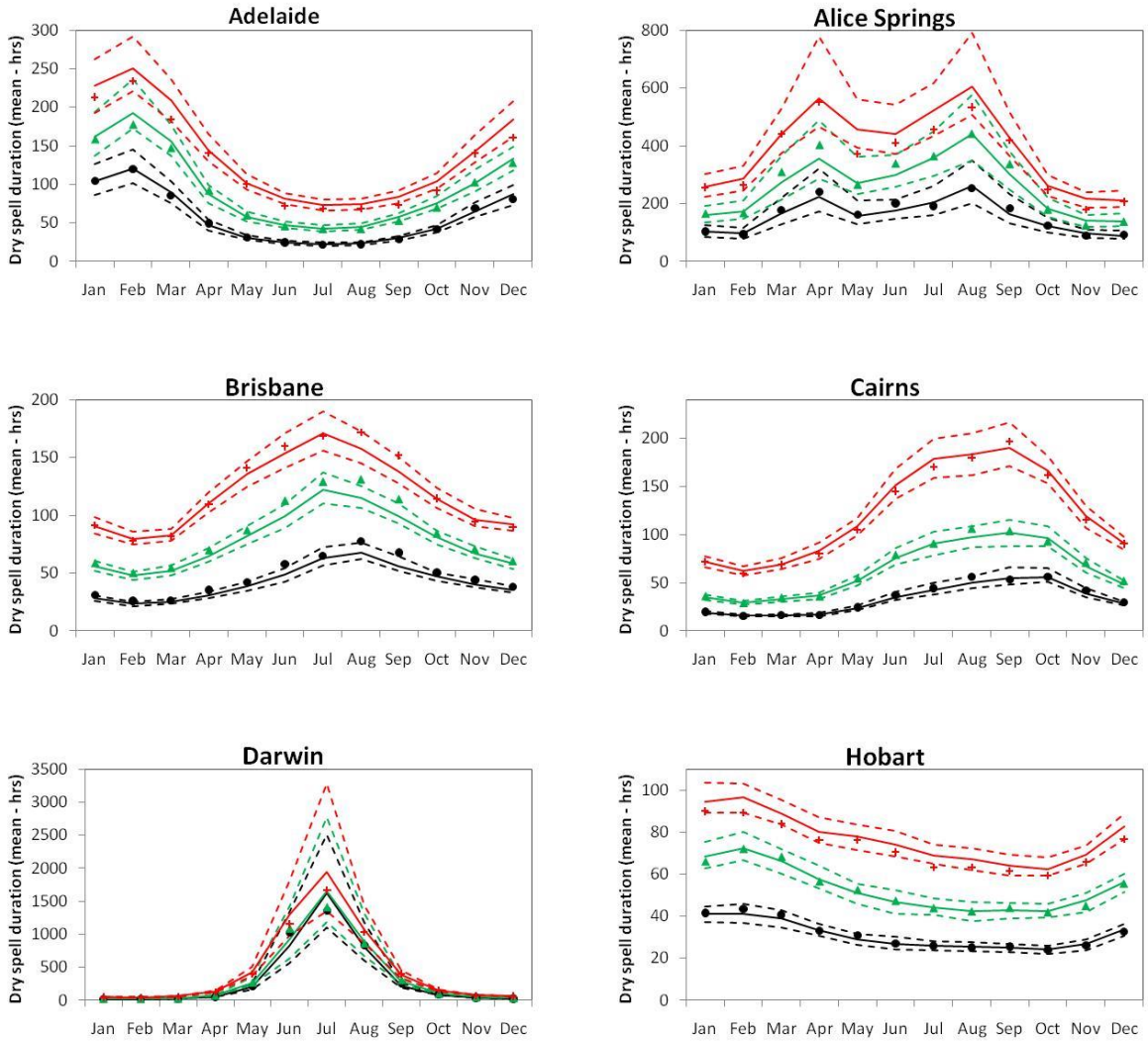


Figure A5: Method of fragments monthly autocorrelation: 1, 6 & 24 hr statistics



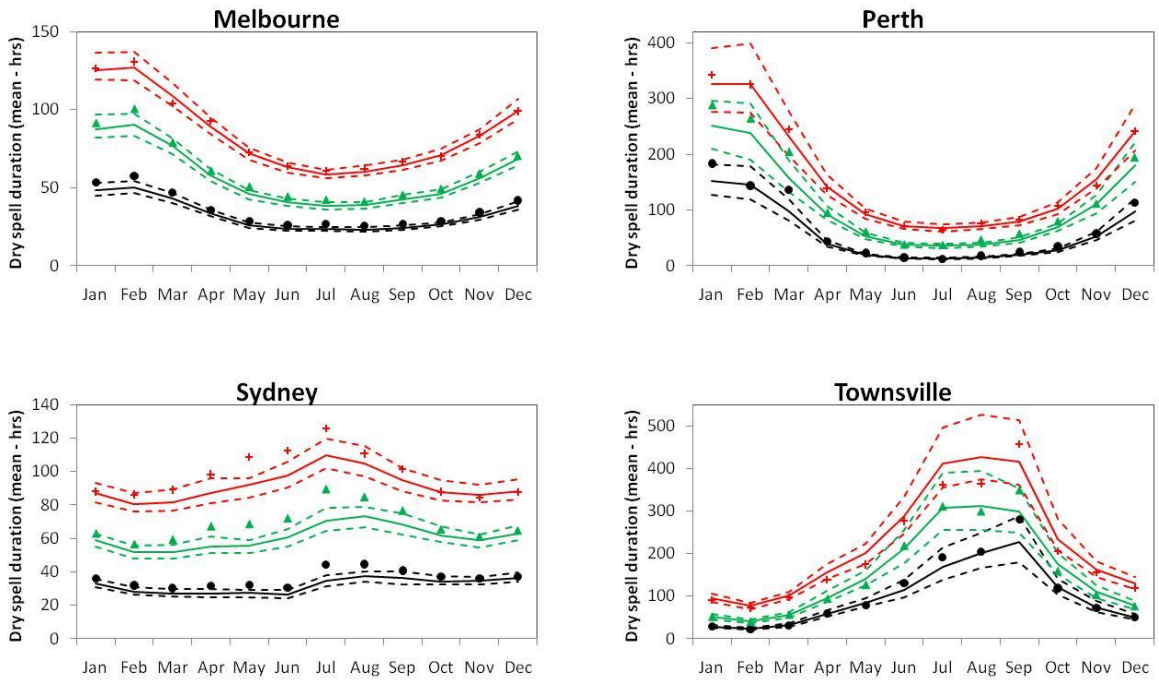
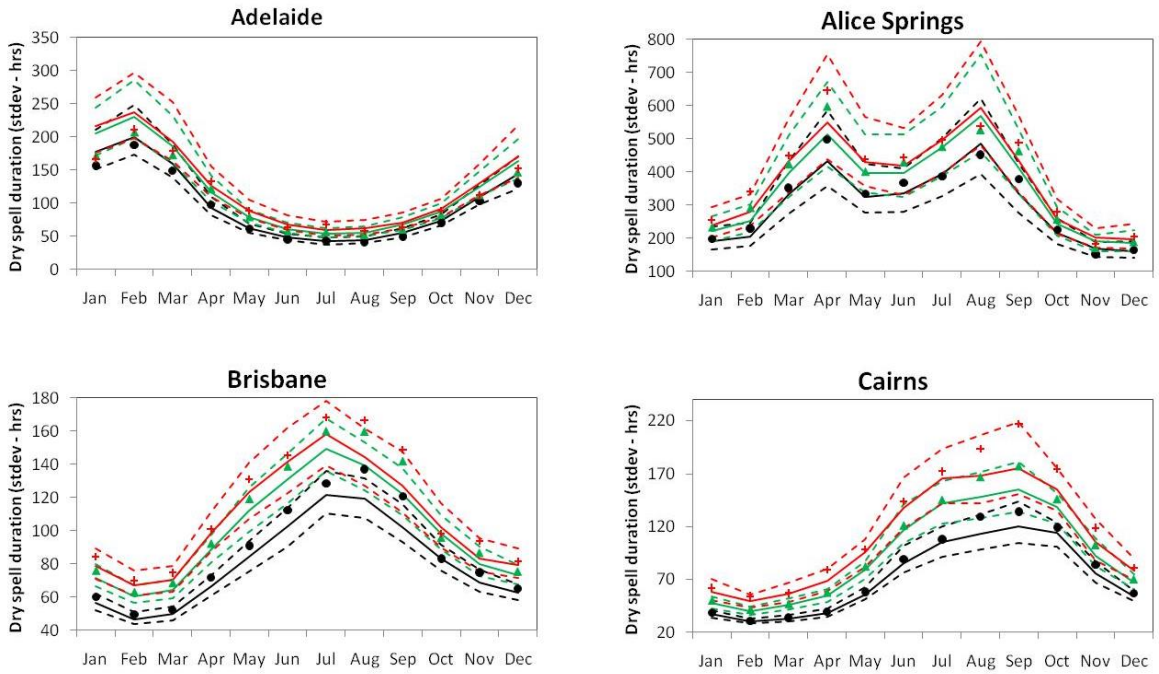


Figure A6: Method of fragments monthly dryspell mean: 1, 6 & 24 hr statistics



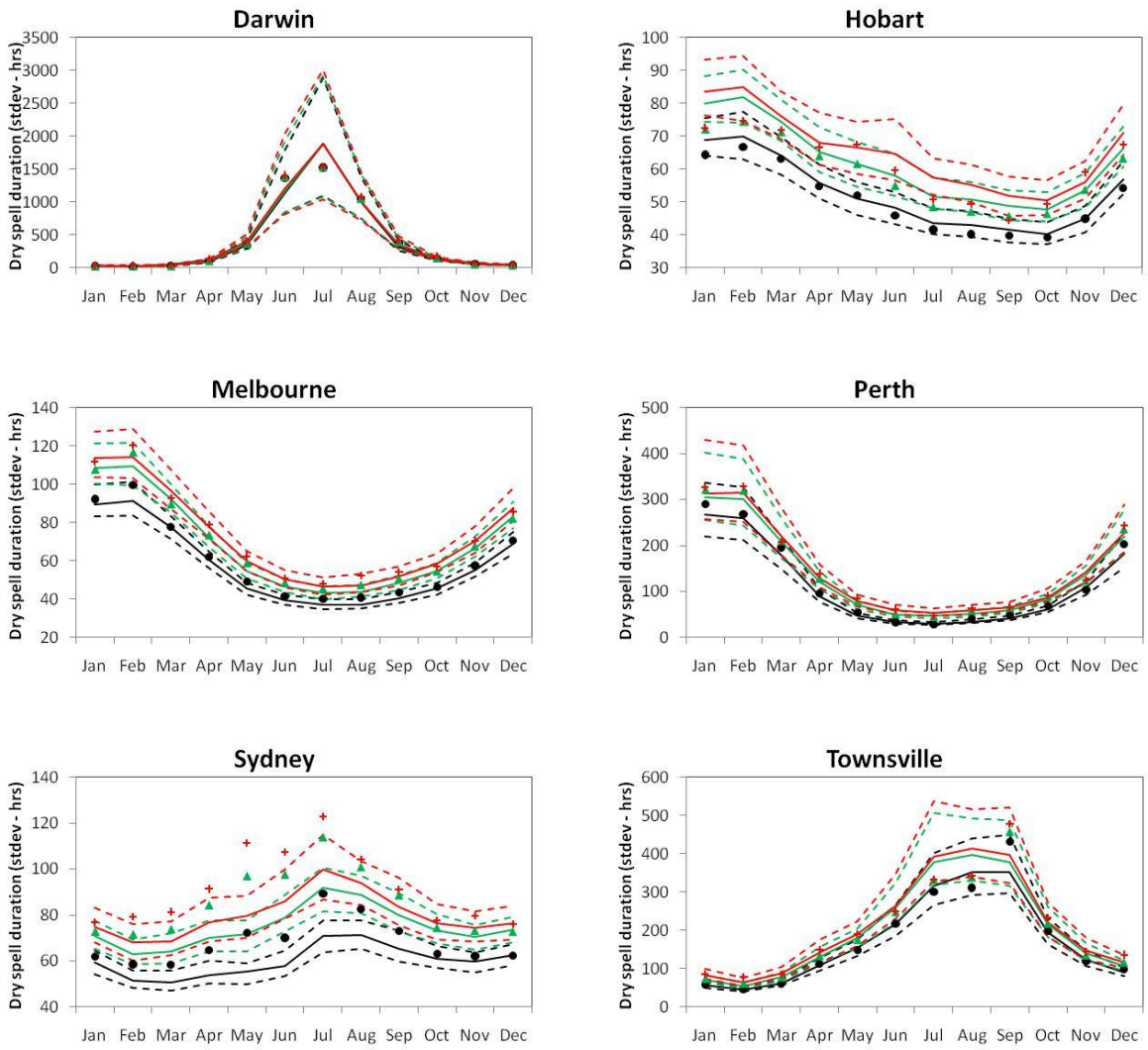
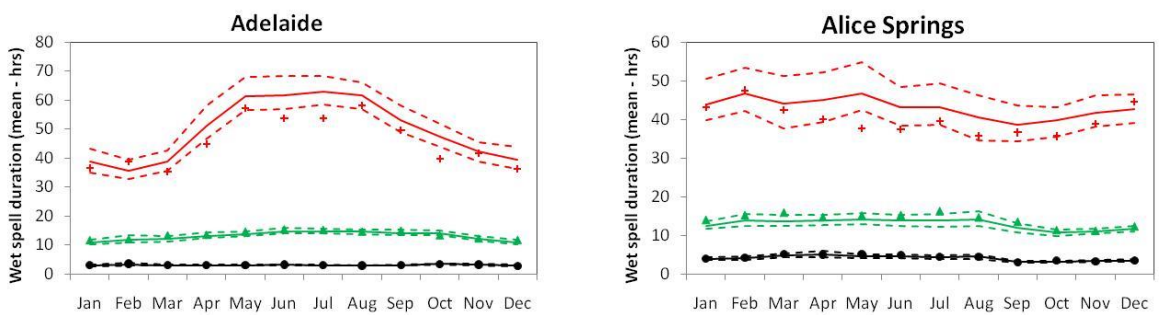


Figure A7: Method of fragments monthly dryspell standard deviation: 1, 6 & 24 hr statistics



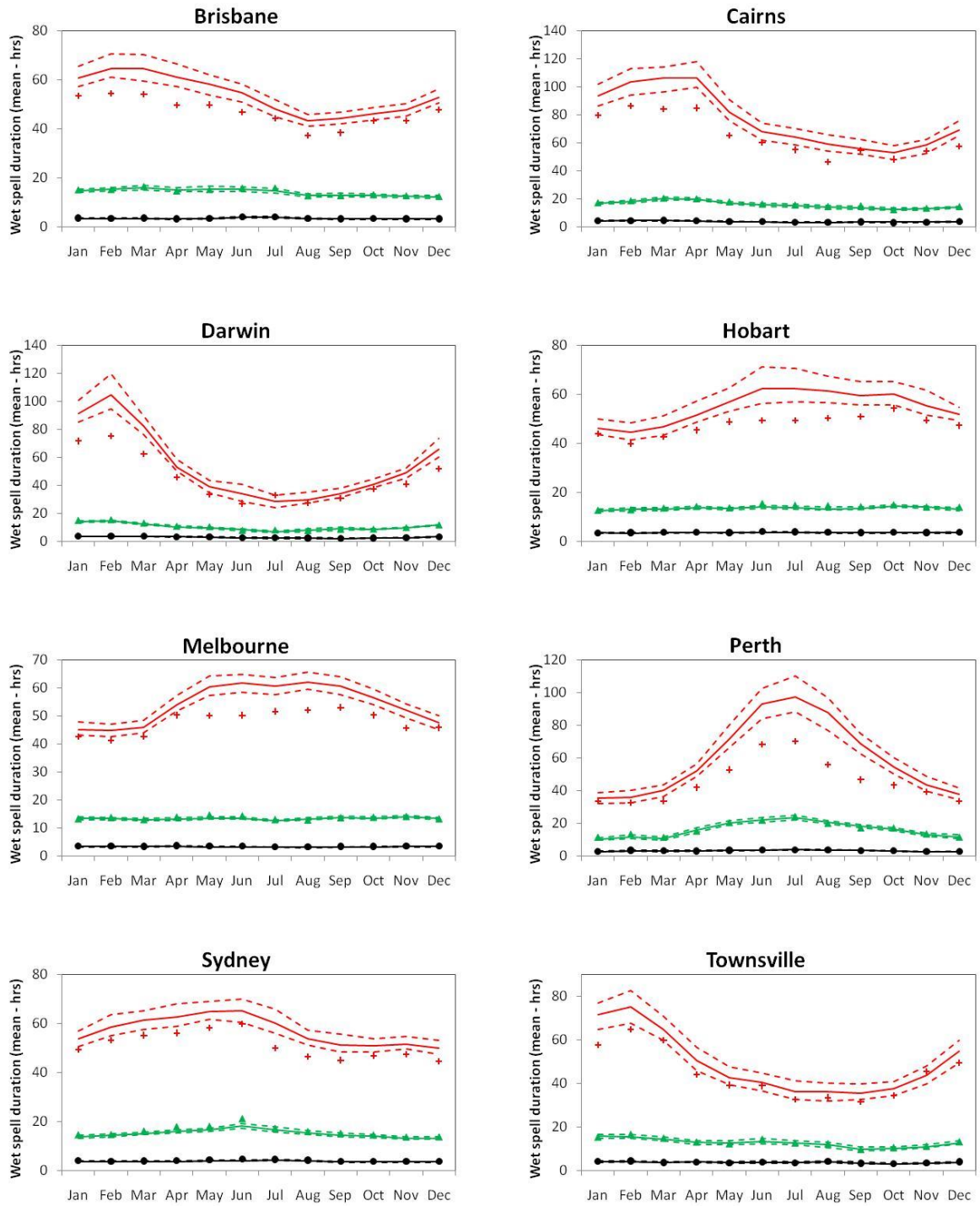


Figure A8: Method of fragments monthly wetspell mean: 1, 6 & 24 hr statistics

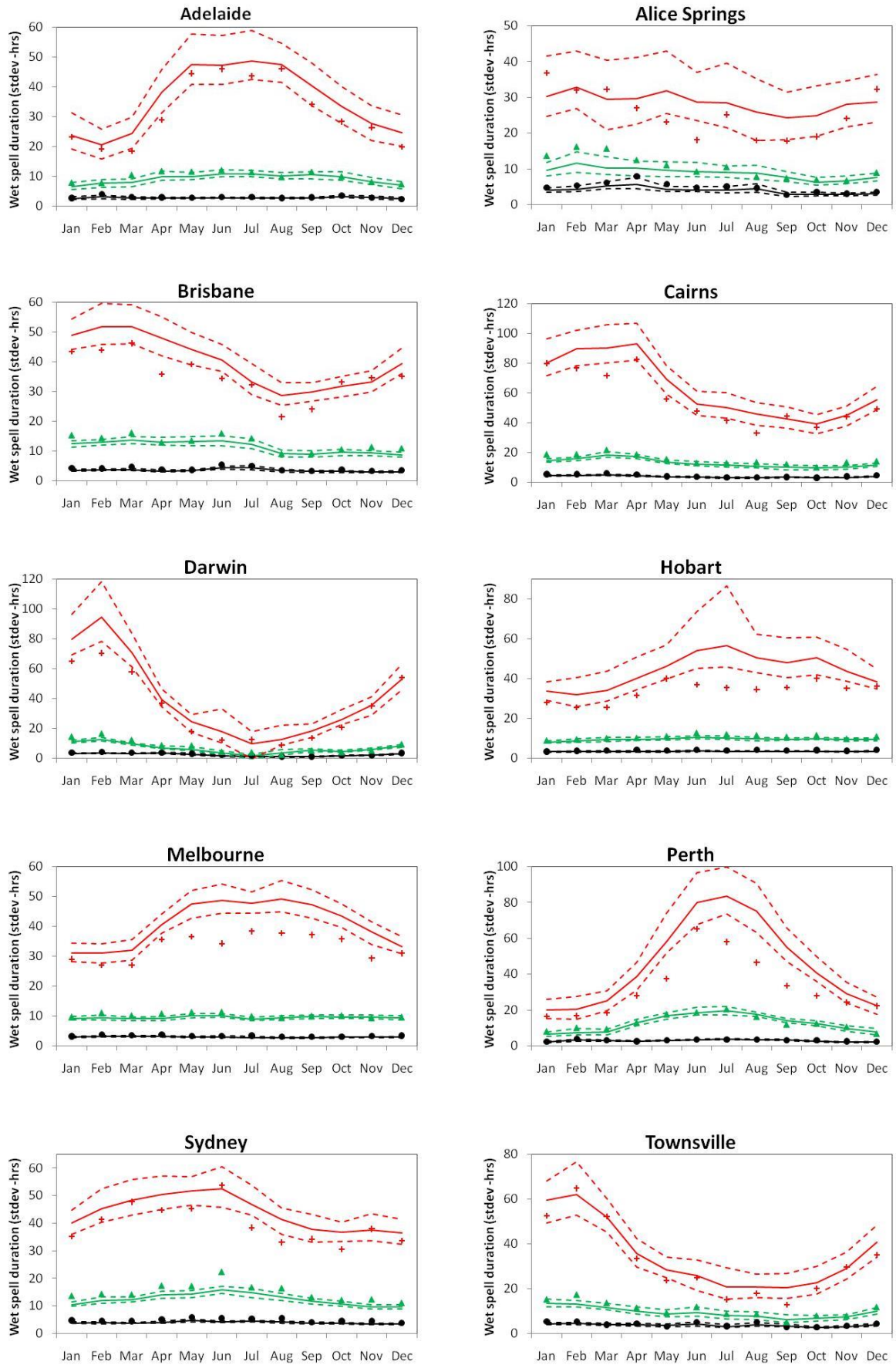
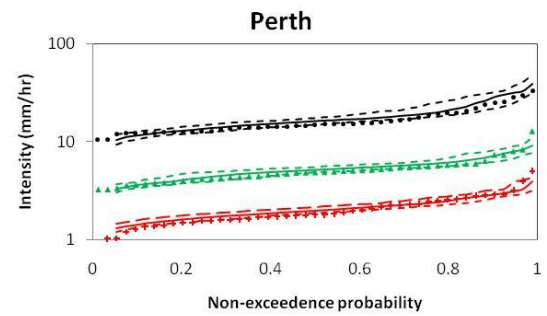
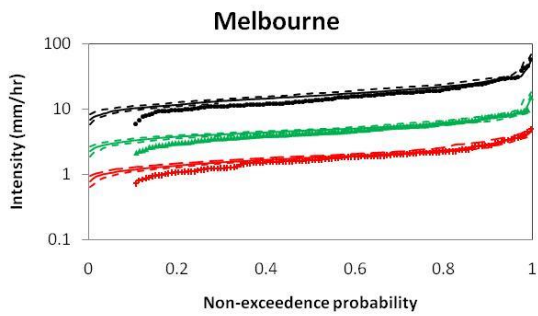
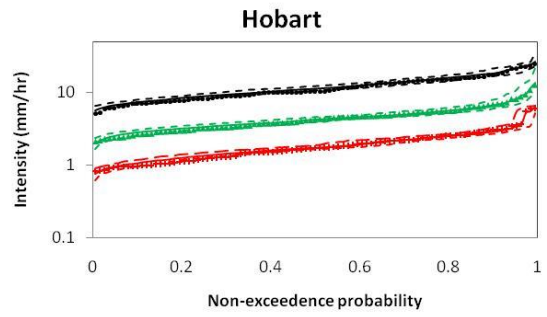
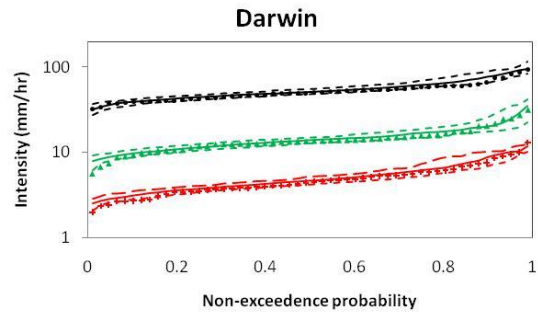
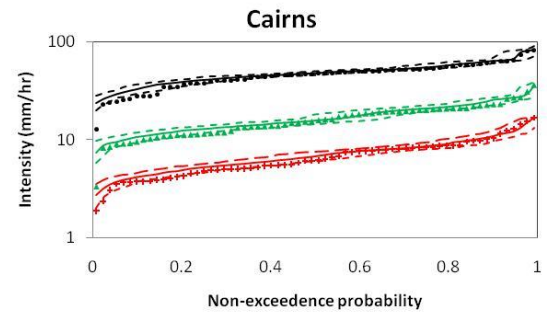
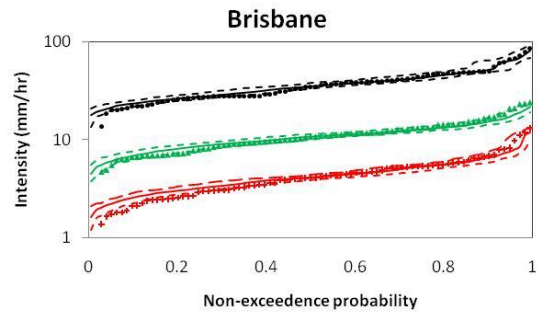
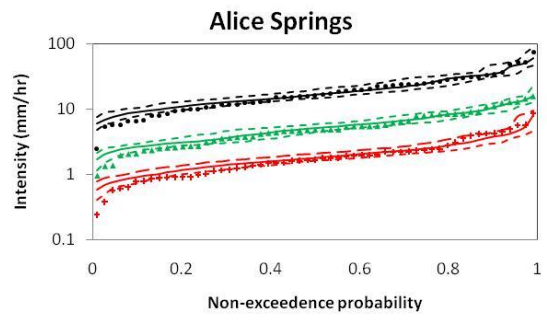
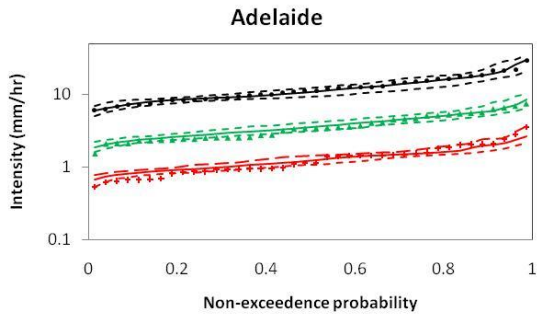


Figure A9: Method of fragments monthly wetspell standard deviation: 1, 6 & 24 hr

statistics



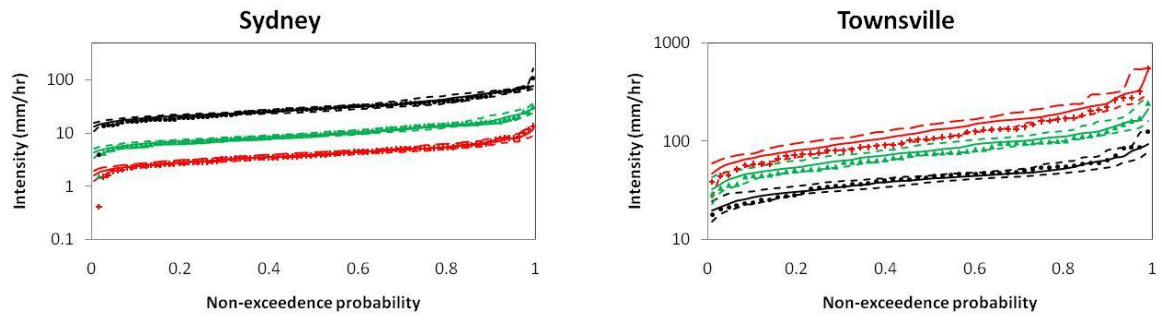
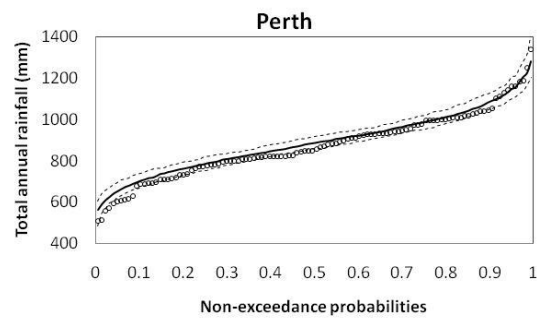
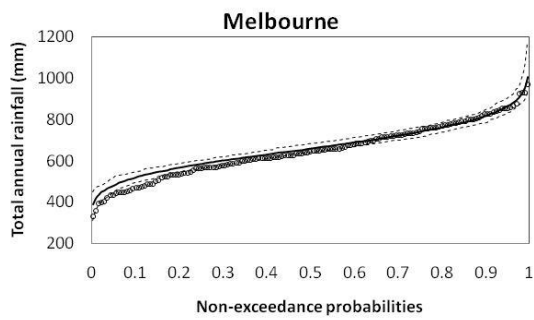
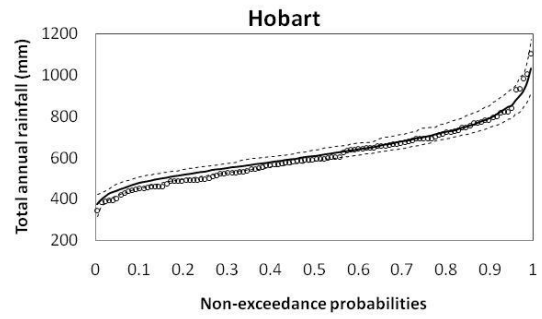
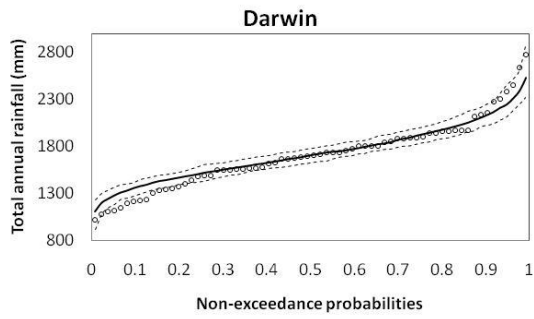
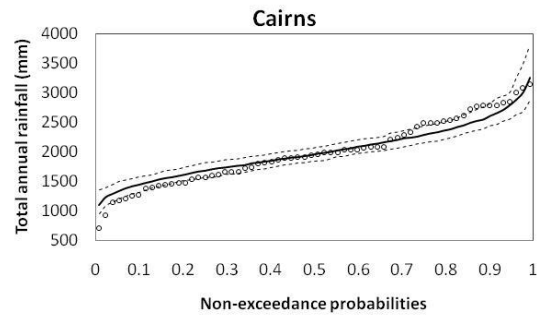
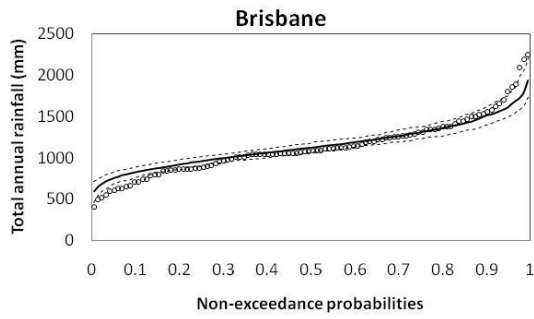
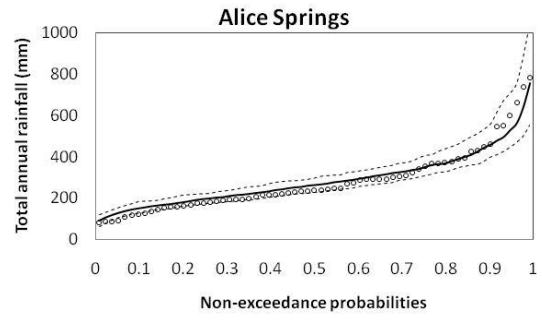
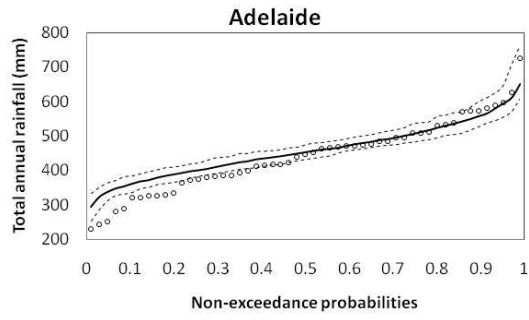


Figure A10: Method of fragments 1, 6 & 24 hr Intensity-Frequency-Duration curves

Appendix B: Daily Markov model - evaluation statistics

For all statistics plotted, the observed values are plotted as a point value. Simulated median (thick line) and 5 and 95% confidence levels (dotted lines) are also plotted.



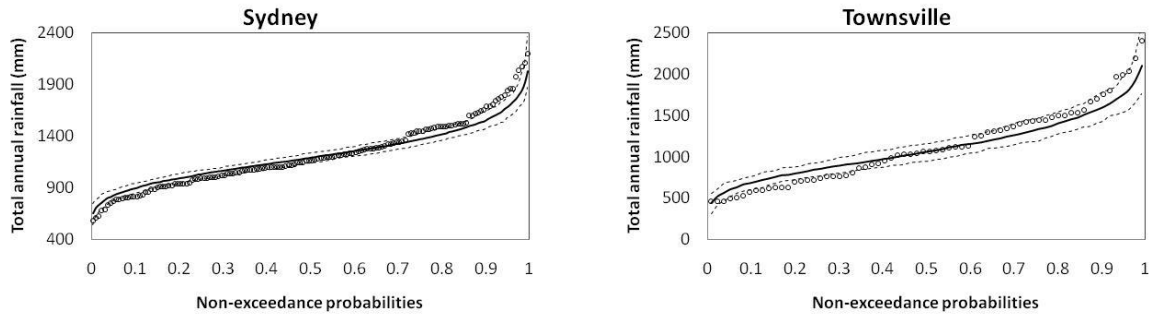
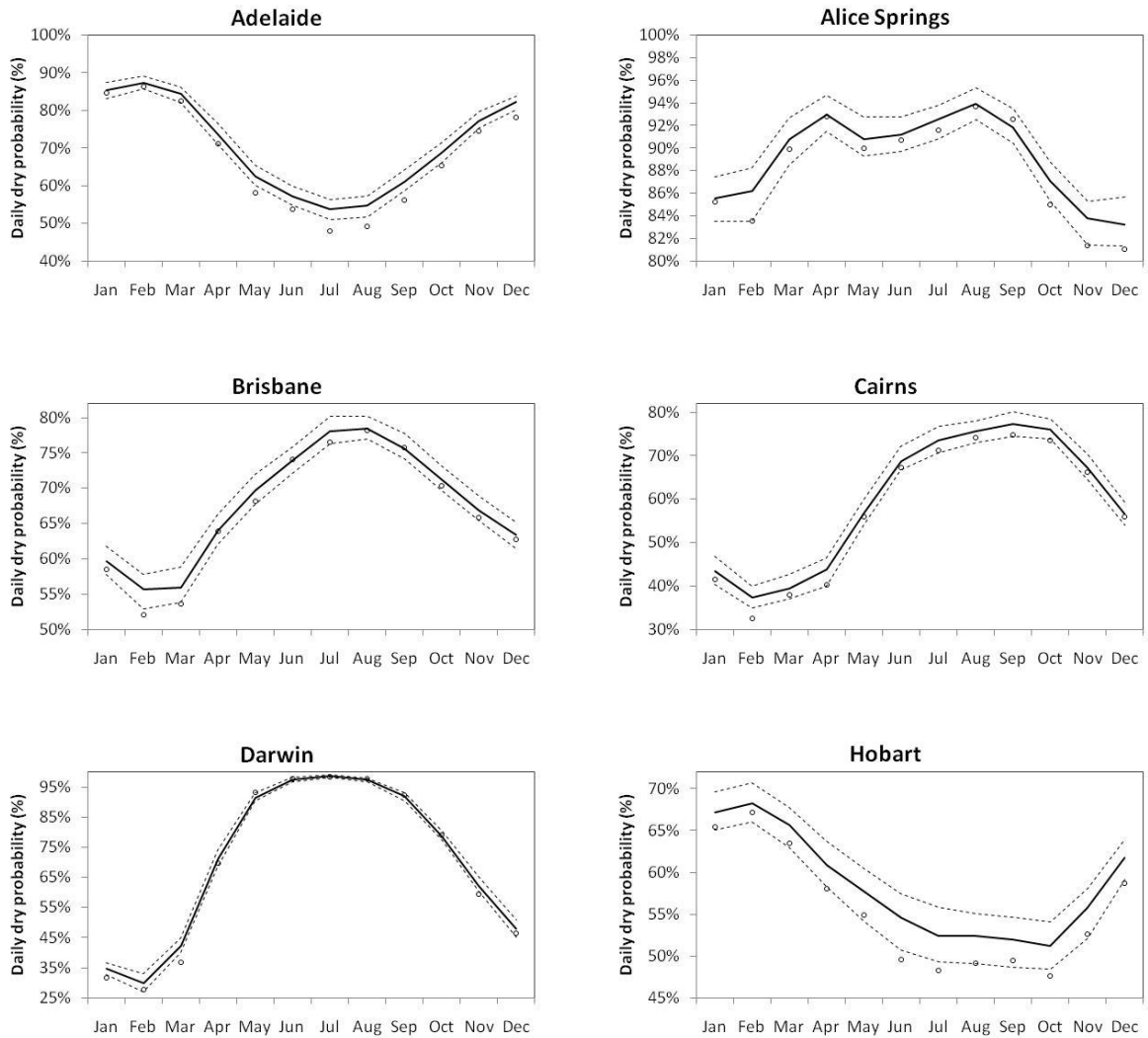


Figure B1: Daily Markov model - annual rainfall distribution



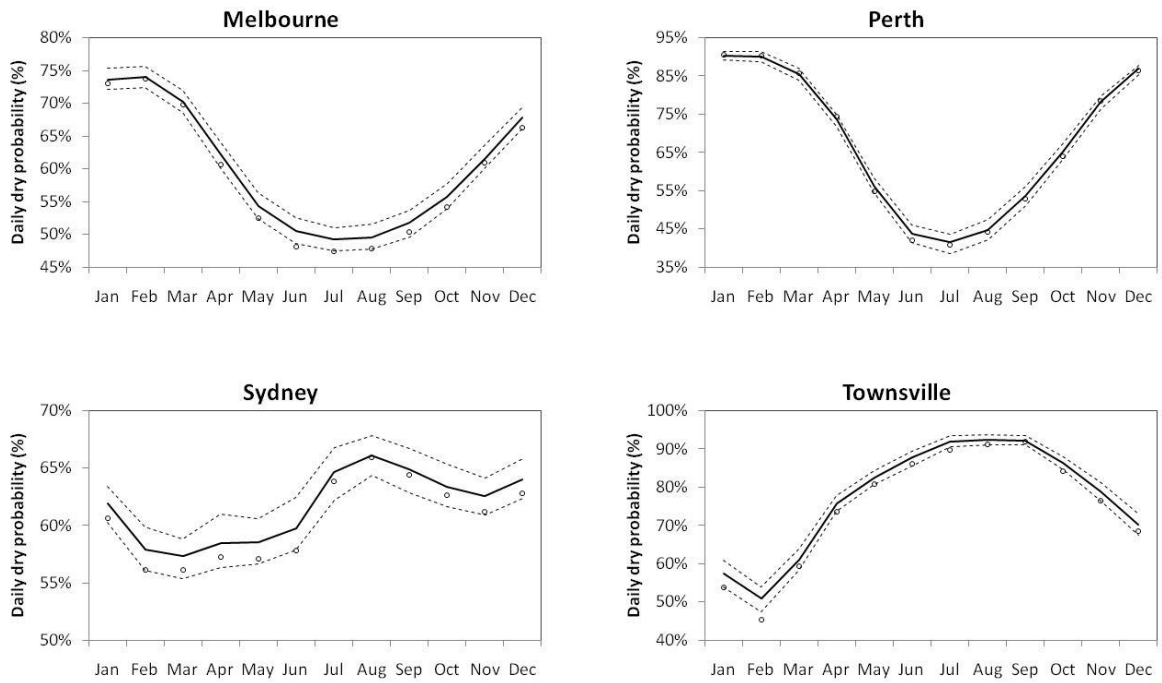
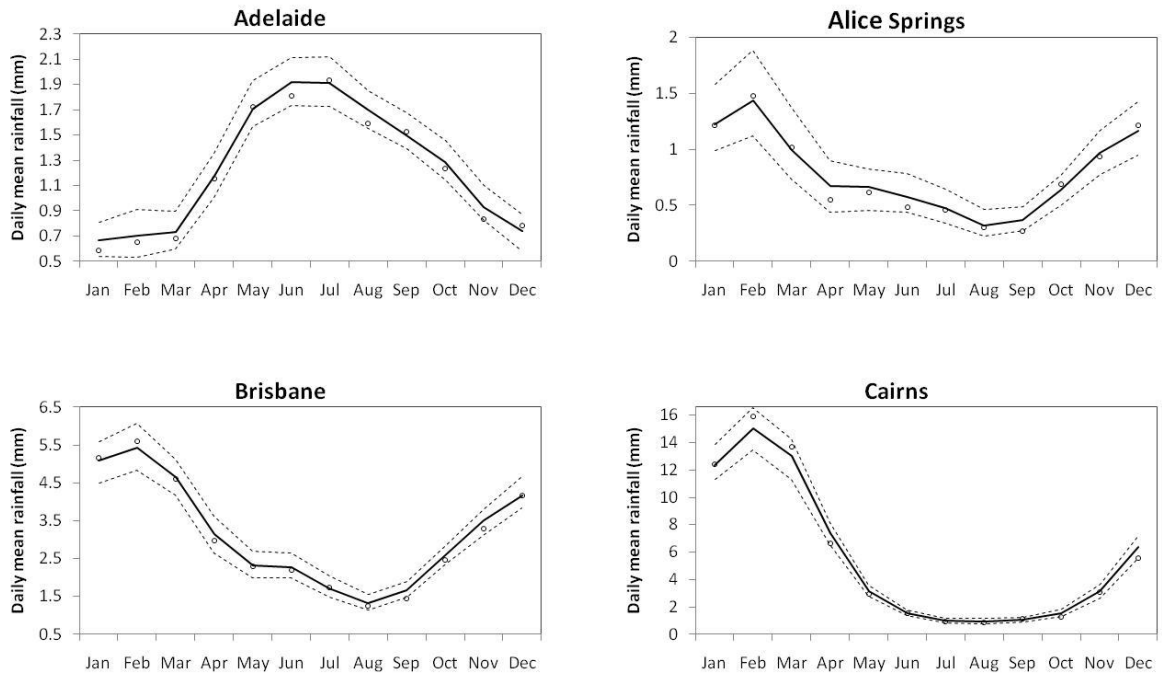


Figure B2: Daily Markov model – dry probability



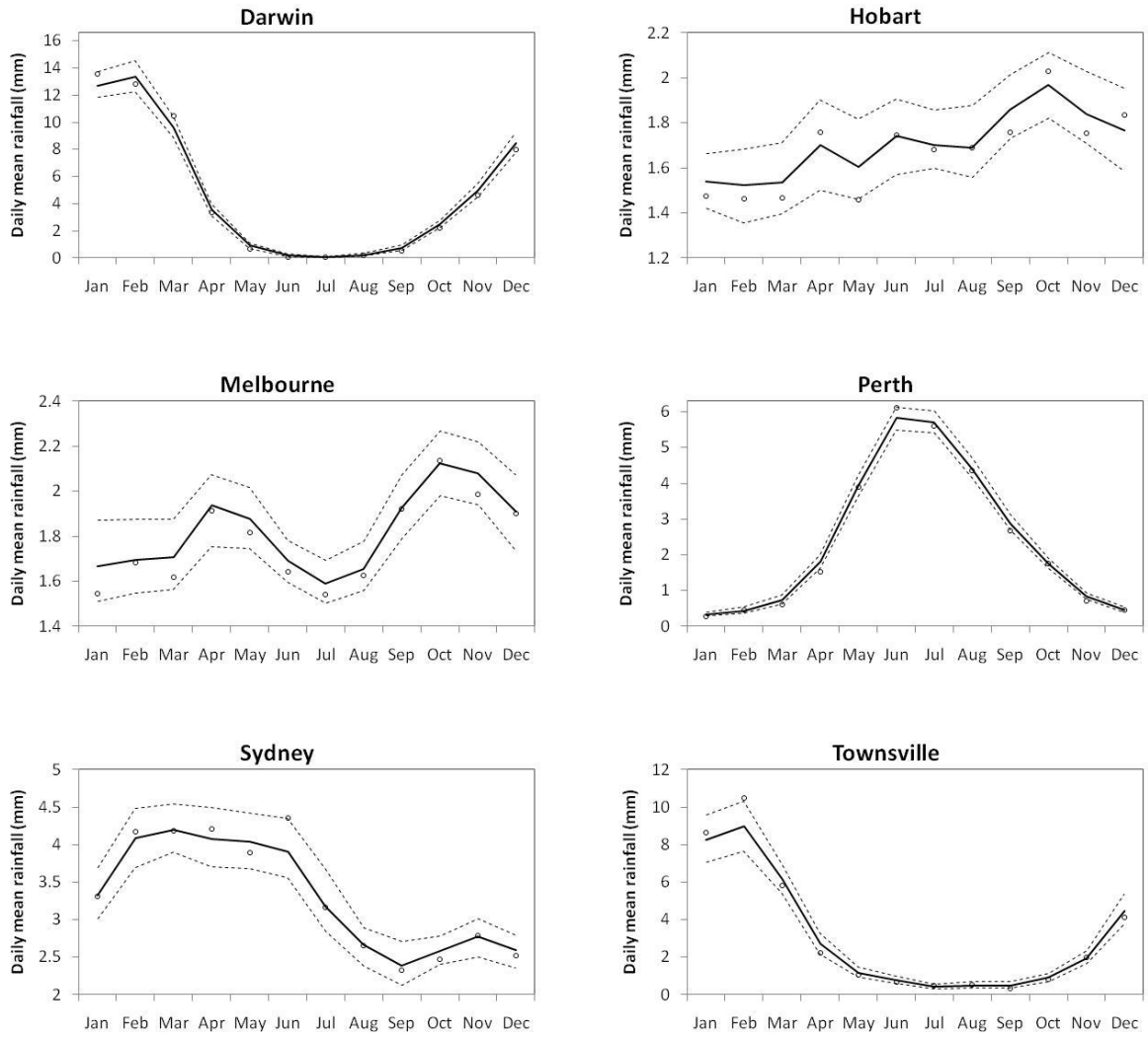
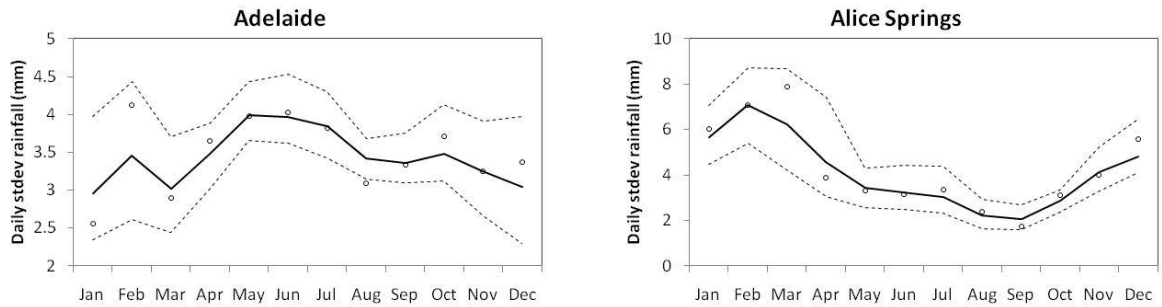


Figure B3: Daily Markov model – daily mean rainfall



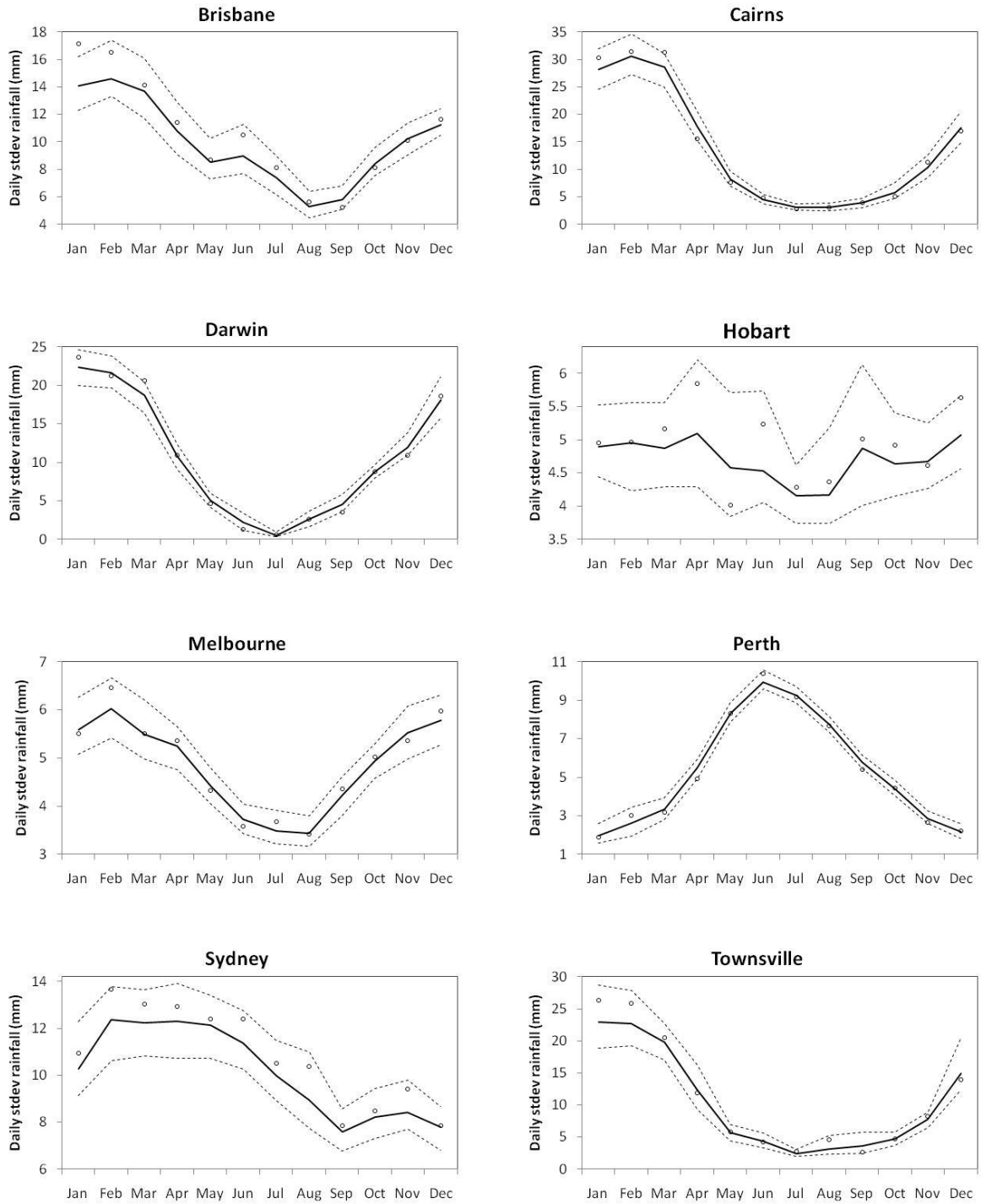
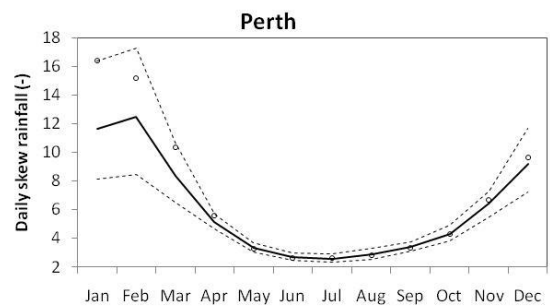
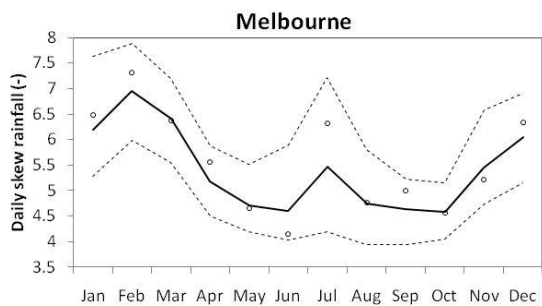
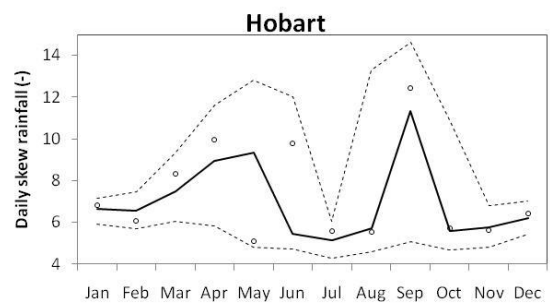
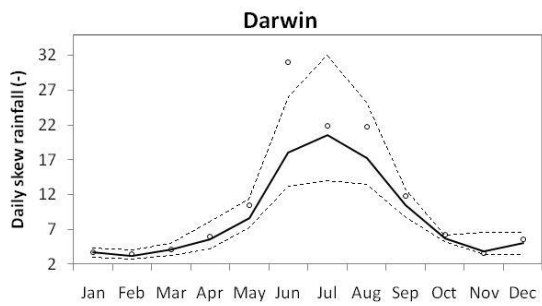
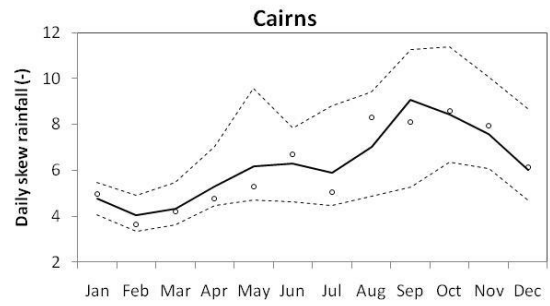
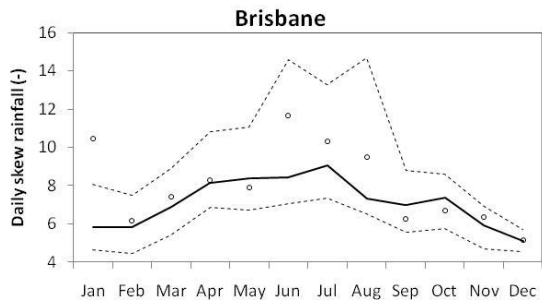
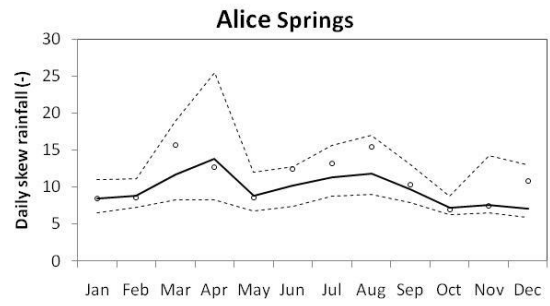
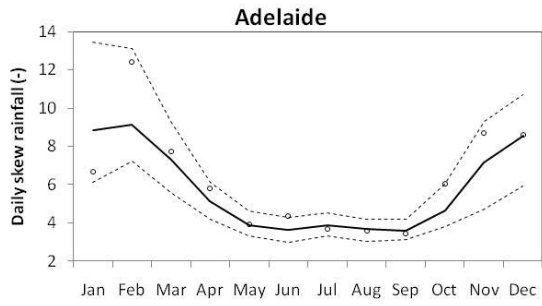


Figure B4: Daily Markov model – daily standard deviation



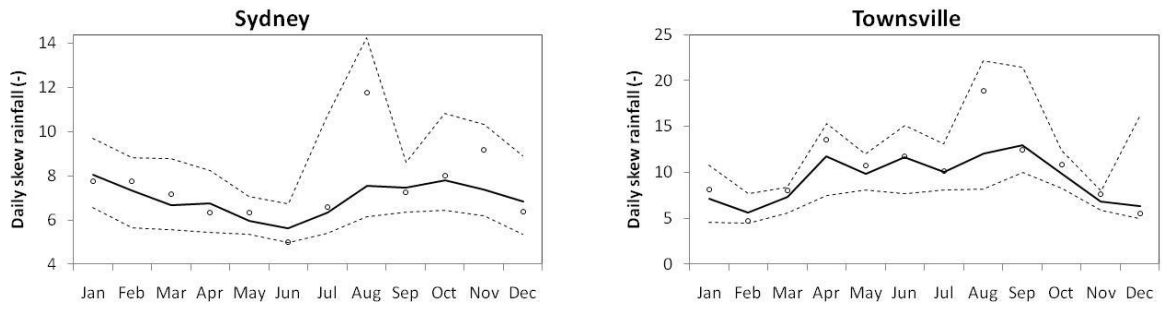
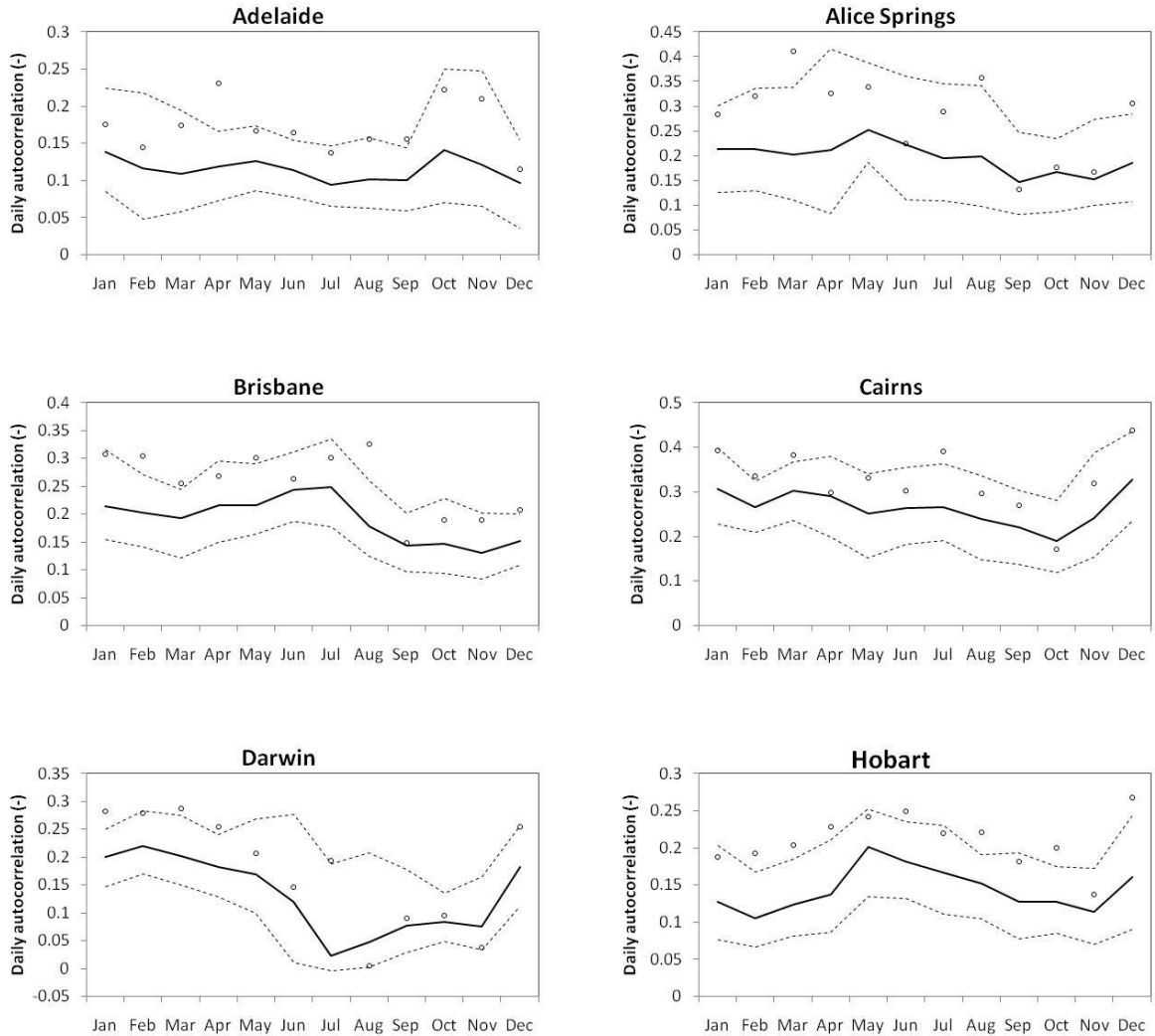


Figure B5: Daily Markov model – daily skew



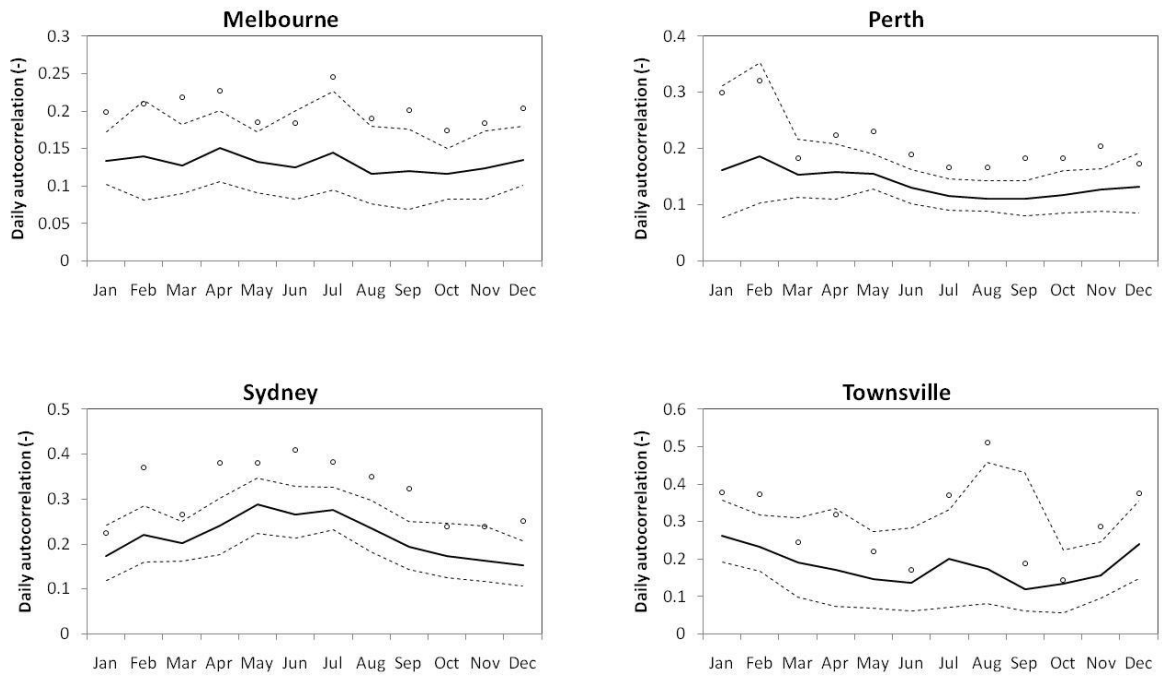
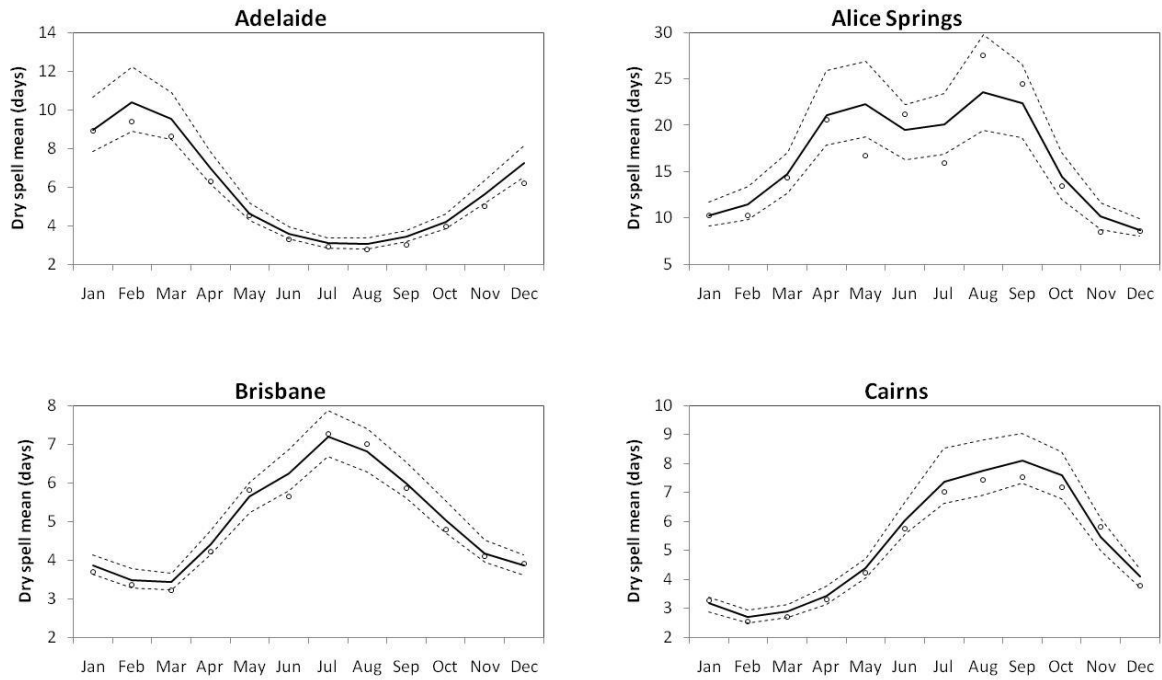


Figure B6: Daily Markov model – autocorrelation



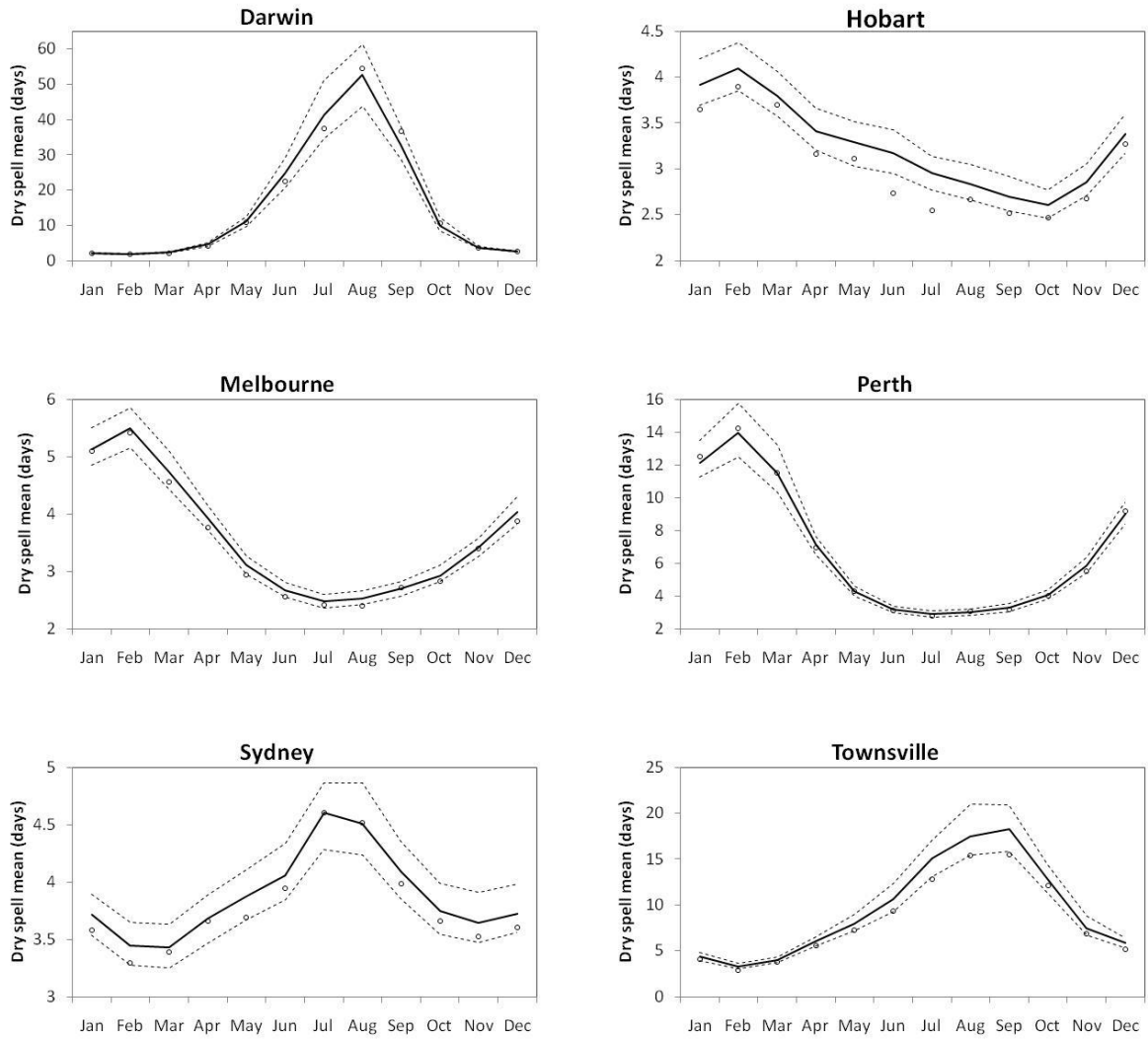
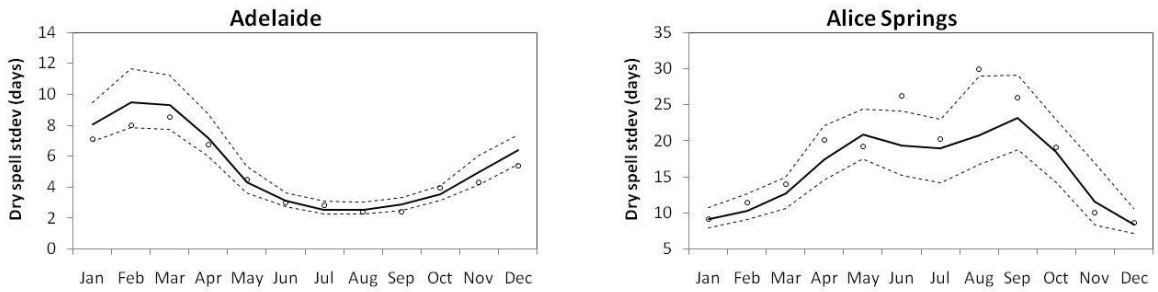


Figure B7: Daily Markov model – dry spell mean



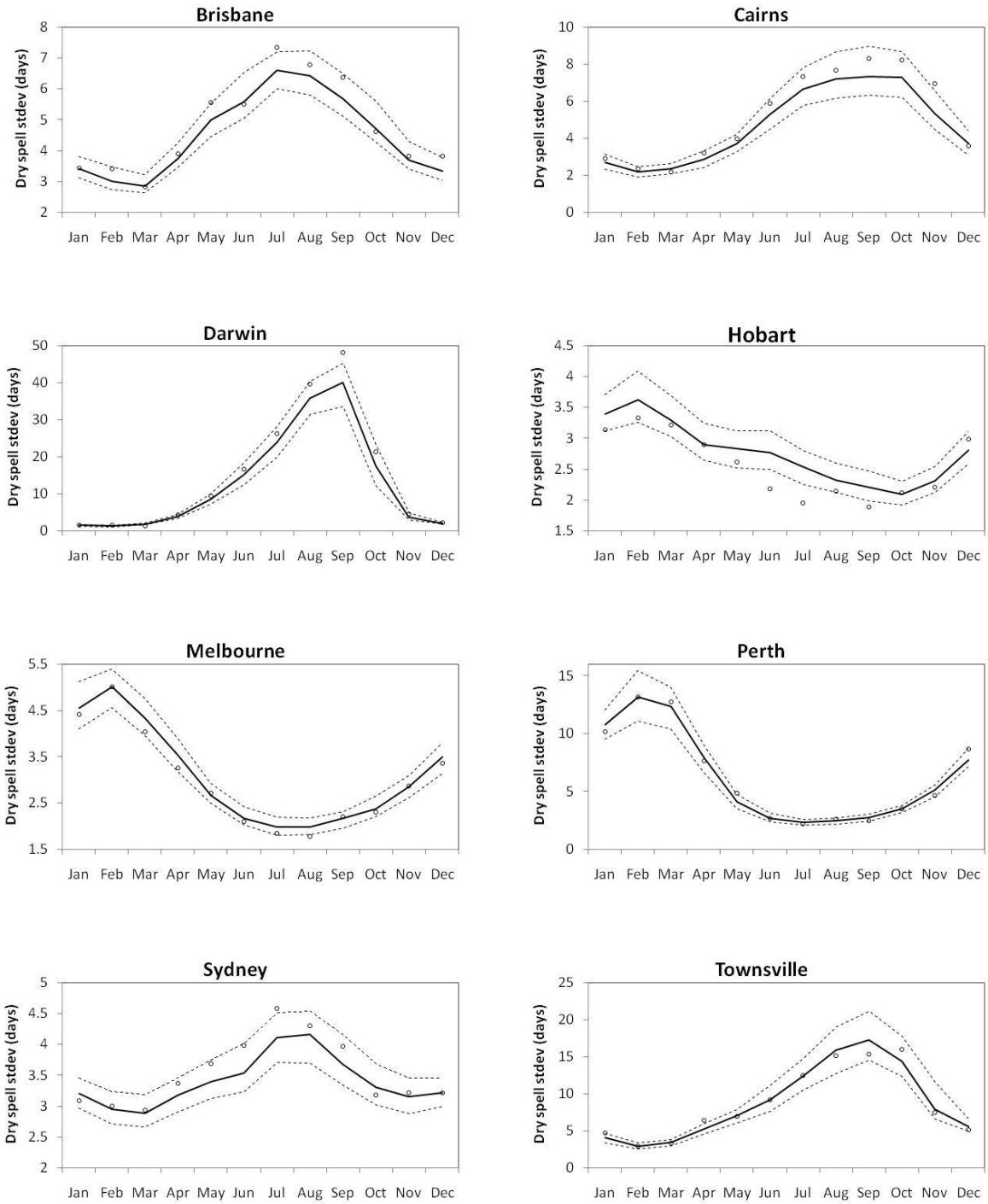
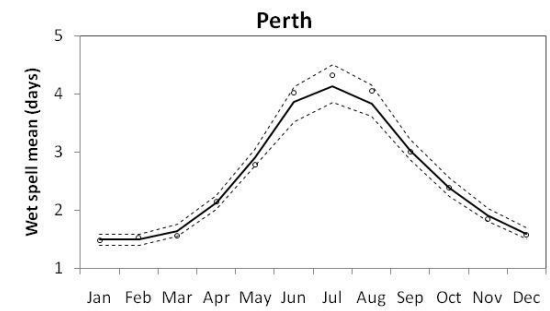
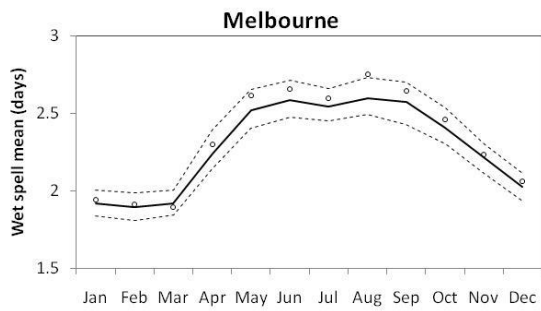
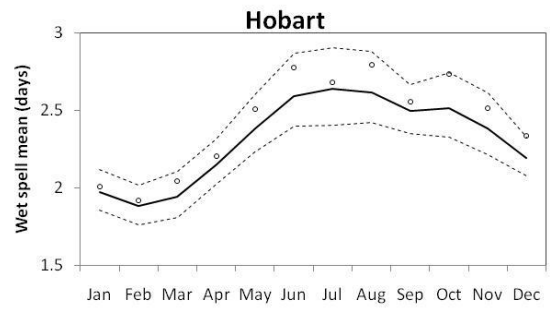
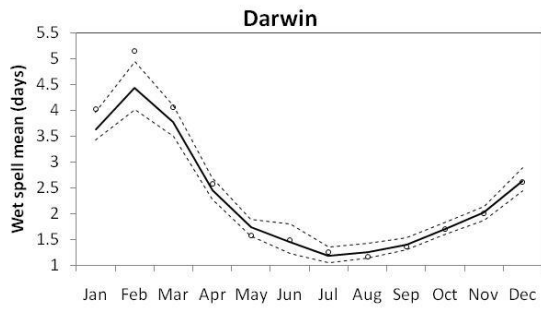
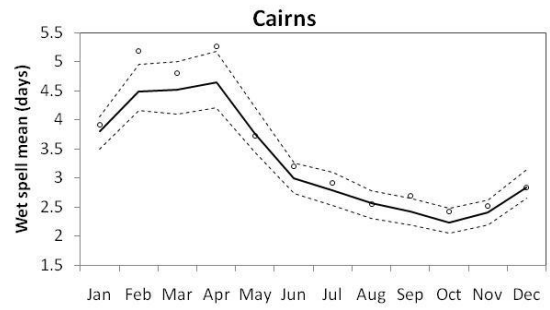
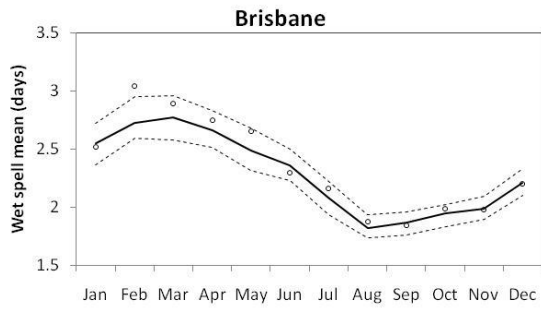
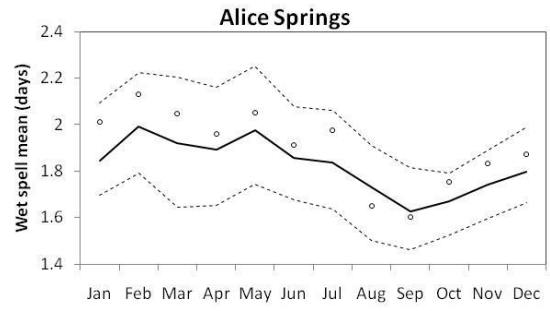
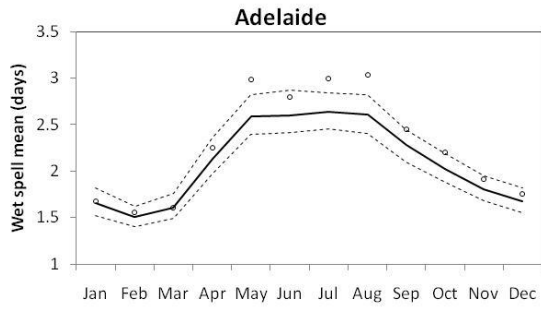


Figure B8: Daily Markov model – dry spell standard deviation



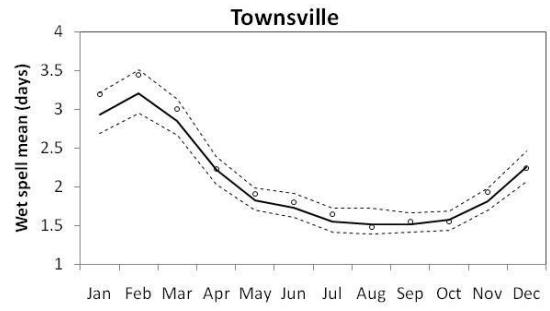
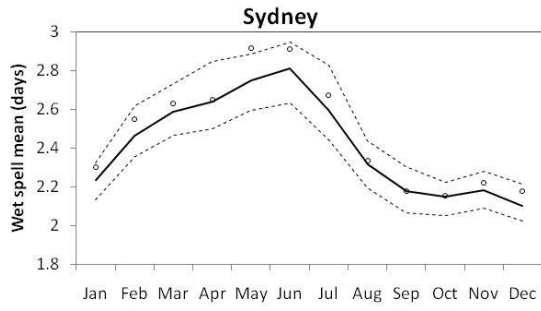
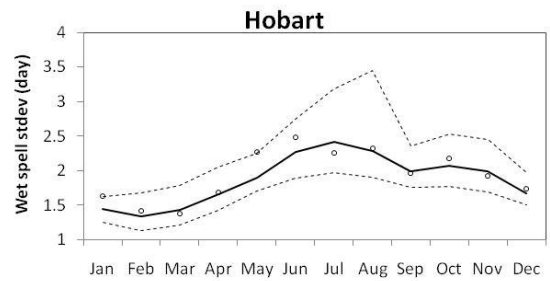
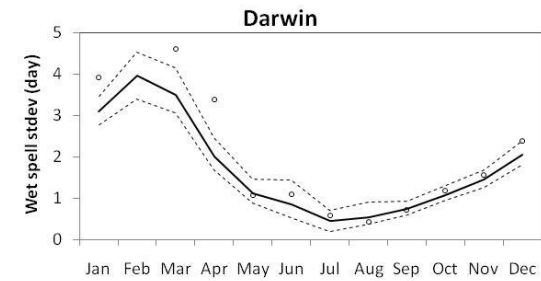
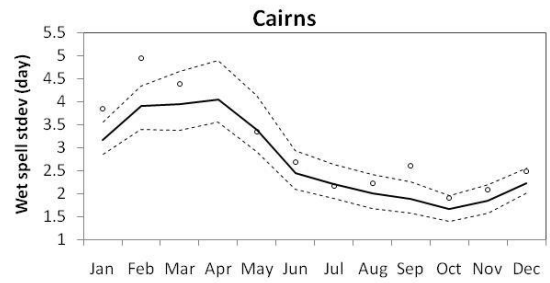
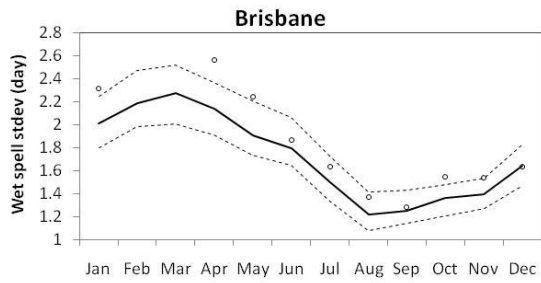
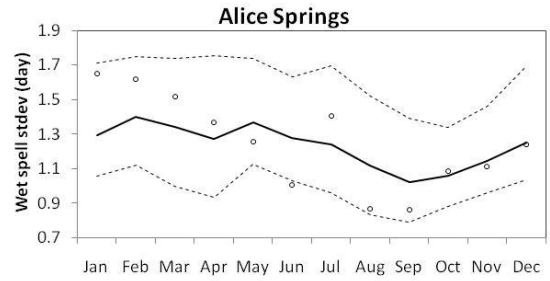
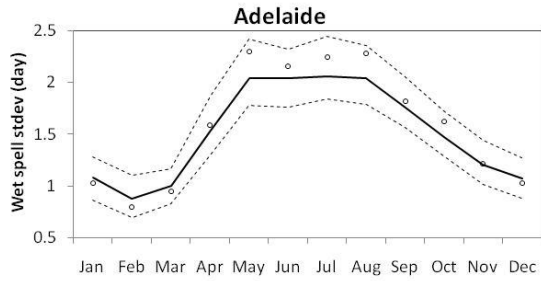


Figure B9: Daily Markov model – wet spell mean



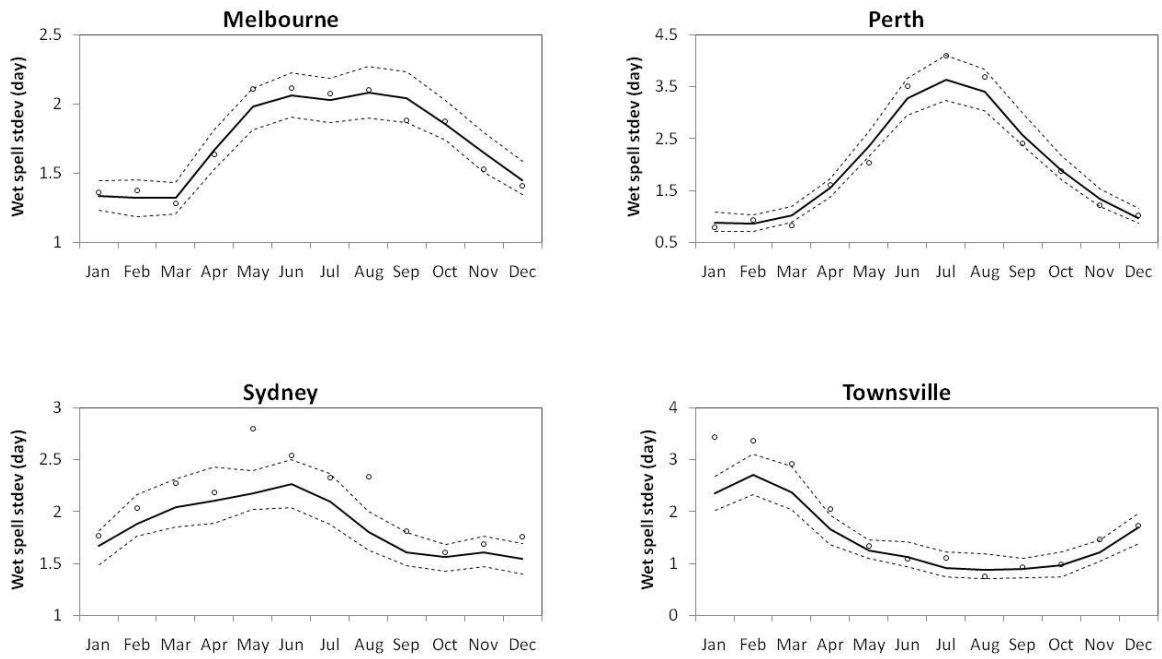


Figure B10: Daily Markov model – wet spell standard deviation