THE UNIVERSITY of ADELAIDE

Stochastic Spatial Rainfall Modelling for Hydrological Design: Development of a Parsimonious Simulation Approach and Virtual Hydrological Evaluation Framework

Bree Sarah Bennett

B. Eng Civil & Structural Engineering (Hons)

B. Laws (Hons)

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Adelaide

Faculty of Engineering, Computer and Mathematical Sciences

School of Civil, Environmental and Mining Engineering

-June 2016-

Table of Contents

Abstract

The management of water, viewed either as a natural hazard or a vital resource, is critical for the safety and prosperity of communities. The risks associated with managing water availability, whether in scarcity or excess, are critical concerns for the design and operation of infrastructure as well as the implementation of public policy. The spatial variability of rainfall is a known driving force of catchment dynamics and water availability, but despite this, it is often poorly represented in hydrologic studies and designs.

This thesis focuses on improvements to the estimation, simulation and evaluation of spatial rainfall. Specifically these developments include: (i) the development of a generalised approach for spatial extreme rainfall estimation; (ii) the development of a flexible, continuous, and spatial stochastic model of rainfall and its corresponding evaluation; and (iii) an innovative framework for critically evaluating the performance of stochastic rainfall models via the assessment of simulated streamflow. Australian case study locations, with varying climates, are used to present and investigate these approaches.

A new approach for estimating extreme spatial rainfall intensities and a critical evaluation of current approaches for estimation are presented. Current techniques for estimating extreme spatial rainfall are reliant on areal reduction factors (ARF) to convert intensity estimates of extreme point rainfall to extreme spatial rainfall. It is common practice to ignore spatial variation in rainfall intensity and assume a constant ARF over a large region. Approaches using ARFs for estimating extreme spatial rainfall were demonstrated to be in error by 5% to 15%. A new approach that explicitly incorporates the variation of spatial rainfall over an area, referred to as Intensity Frequency Duration Area (IFDA) was developed to address this issue. IFDAs use spatially interpolated rainfall grids to directly estimate how extreme rainfall intensity varies with frequency, duration and area for a given location. The IFDA approach overcomes the shortcomings of existing approaches by avoiding the need to assume a fixed regional ARF value. IFDA provide direct and unbiased estimates of extreme spatial rainfall.

An alternative approach to spatially interpolated observations of extremes is to use data generated by a stochastic spatial rainfall model. A new model for continuously simulating fields of daily spatial rainfall in a parsimonious manner is developed in this thesis. A Gaussian latent variable approach is used because it is able to simultaneously generate rainfall occurrences as well as amounts. Parameter surfaces are produced via kriging which enables the model to produce stochastic replicates for any location of interest in the catchment. Additional benefits of the model are that it removes the need for interpolation to construct catchment average rainfall estimates, preserves the rainfall's volumetric properties and can be used with distributed hydrologic models. A comprehensive evaluation approach was developed to identify model strengths and weaknesses. This included a performance classification system that provided a systematic, succinct and transparent method to assess and summarize model performance over a range of statistics, sites and scales. The model showed many strengths in reproducing observed rainfall characteristics with the majority of statistics classified as either statistically indistinguishable from the observed or within 5% of the observed across the majority of sites and seasons.

A significant challenge when evaluating rainfall models is that the key variable of interest is resultant streamflow, not generated rainfall. Typical evaluation methods use a variety of rainfall statistics, but they provide limited understanding on (i) how rainfall influences streamflow generation; (ii) which rainfall characteristics are most important; and (iii) the trade-offs made when one or more features of rainfall are poorly reproduced. An innovative virtual hydrological evaluation framework is developed to evaluate whether deficiencies in simulated rainfall lead to deficiencies in resultant streamflow. The key feature of the framework is the use of a hydrological model to compare streamflow derived from observed and simulated rainfall at the same location. The framework allows the impact of an influencing month of simulated rainfall on streamflow in an evaluated month of interest to be isolated. Application of the virtual hydrological evaluation framework identified the importance of transition months May and June (late autumn/early winter) in the 'wetting-up' phase of the catchment cycle. Despite their low monthly flow volumes, the transition months contributed significantly to error in the annual total flow.

With improved representation and evaluation of spatial rainfall, this thesis ultimately demonstrates more realistic and accurate methods for hydrological estimation.

Statement of Originality

I, Bree Sarah Bennett, certify that this thesis contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

_____ _____

Bree Sarah Bennett Date

Acknowledgements

'Far and away the best prize that life has to offer is the chance to work hard at work worth doing.'

– Theodore Roosevelt 1903

I feel honoured and privileged to have been able to experience the pleasure of research. For this my most sincere gratitude goes to my supervisory team, Professor Martin Lambert, Associate Professor Mark Thyer, Professor Bryson Bates and Dr Michael Leonard, who have provided me with great support, inspiration and guidance throughout my candidature. Your combined skill, expertise and passion for research have been invaluable.

I would like to thank my family, friends and colleagues who have shown interest in my research. Particular thanks go to Nicole Arbon who has been especially steadfast in her support and encouragement.

Finally, thanks go to my husband, Phil, for his unwavering encouragement and understanding.

List of Figures

List of Tables

Chapter 1

Traditionally, urban development in Australia has occurred near rivers and on their flood plains, inextricably linking the management of water to community progress. Risks associated with water, such as floods or droughts, are therefore critical considerations for the development of public policy and infrastructure.

Rainfall is a fundamental component of the hydrological cycle. It is a driving force in catchment dynamics such that properties of incident rainfall have a significant influence on resulting runoff in terms of the timing, magnitude, range and behaviour of flows. Therefore the reliable estimation of rainfall is of great practical importance to hydrology and allied fields.

A key property of rainfall that affects the rainfall-runoff response is its spatial distribution. There are a number of challenges in obtaining and applying reliable estimates of spatial rainfall in hydrological modelling. These challenges include limitations of available rainfall data and rainfall modelling approaches as well as the evaluation of these models.

1.1 Limitations of Rainfall Data

Rainfall observations are commonly used as the basis for hydrological design rather than observed streamflow. Although rainfall requires transformation in terms of a rainfall-runoff model (i.e. a hydrological model) the benefits of rainfall-based runoff assessment are numerous. When compared to using streamflow records rainfall has several advantages: (i) there is a relative abundance of rainfall gauges and data sources (e.g. interpolated rainfall grids and stochastic rainfall models); (ii) they have relatively longer records; and (iii) they are largely unaffected by changes to the catchment over time. The reliance on rain gauges has acted to emphasise the temporal properties of rainfall over spatial features, which is largely because estimates of risk are primarily interested in rare events.

Spatial characteristics of rainfall are important for understanding water availability but have received less attention. For understanding variations in rainfall amount, the density of gauges at the daily timescale is suitable for reproducing the spatial

structure of rainfall. However, for understanding spatial features of individual events, studies must rely on pluviometer gauges and radar rainfall estimates which have relatively shorter record lengths and more complicated error structures. At the daily scale, the Bureau of Meteorology have developed the AWAP product (Raupach et al., 2009; Raupach et al., 2012), which is an interpolated dataset based on rain gauges. Fig. 1-1 shows the national coverage of the AWAP product and interpolated observed daily rainfall patterns for a selection of seasons (panels (b) – (e)). However as yet, there are few models capable of simulating stochastic replicates of spatial rainfall for design.

There is a long history of using interpolation methods to infill space between point rainfall measurements (Haberlandt, 2007; Hutchinson, 1995; Lebel and Laborde, 1988; Thiessen, 1911), including the use of geometric methods, thin plate splines and kriging frameworks. These methods are better suited to cases where there is little or no skewness in the distribution of values and where the variable is non-zero (as with aggregate totals at the annual or monthly scale). At daily and sub-daily scales these methods struggle due to the skewness of rainfall amounts, the patchiness of rainfall occurrences (i.e. wet-dry pattern) and complicated correlation structure. While, there have been some attempts to remedy this (e.g. Haberlandt, 2007) they have numerous limiting assumptions. One promising method which can accommodate the characteristics of skewness, patchiness and correlation is the latent variable modelling approach (Bardossy and Plate, 1992). It is developed further in this thesis.

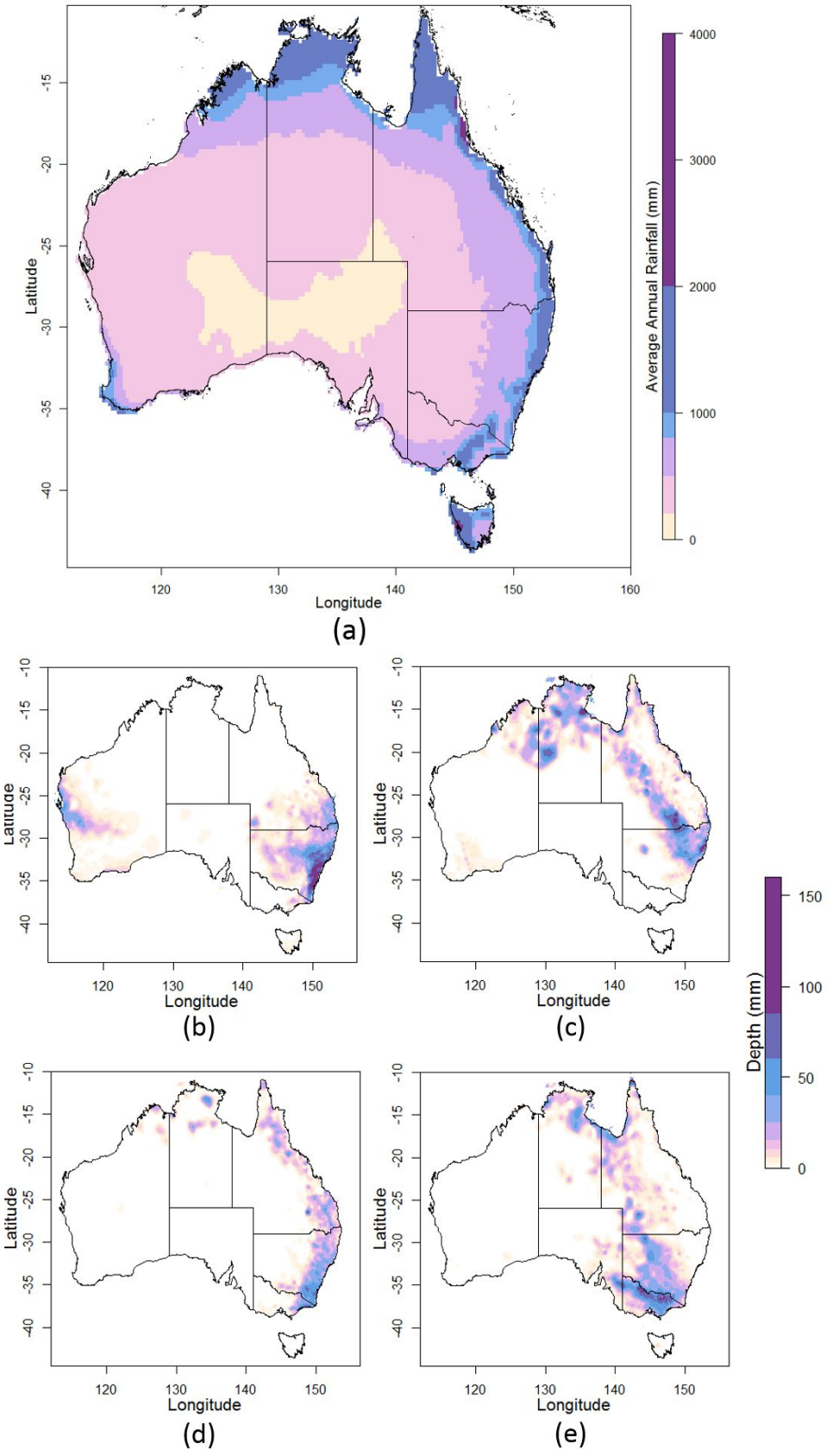


Fig. 1-1 AWAP interpolated daily rainfall (a) average annual rainfall, (b) 21 June 1975, (c) 4 March 1982, (d) 6 December 1995, and (e) 12 November 1998.

1.2 Limitations of Stochastic Rainfall Models and their Evaluation

Rainfall models have become increasingly sophisticated, in line with developments in computing ability, enabling the models to efficiently simulate and handle large amounts of data. To date, the majority of stochastic rainfall model development has been based at a single site (Heneker et al., 2001; Onof and Wheater, 1993; Rodriguez-Iturbe et al., 1988); some models have been developed for a multi-site configuration (Rasmussen, 2013; Sanso and Guenni, 2000; Srikanthan and Pegram, 2009; Wilks, 1998); and relatively few models infill space for long time periods (Leonard et al., 2008; Pegram and Clothier, 2001). The vast majority of truly spatiotemporal models tend to focus on individual storm events (Seed et al., 1999). There is a need to improve the spatial sophistication of rainfall models for simulating over long periods of time.

There are a number of challenges involved with spatiotemporal models. For example, one challenge is preserving the temporal correlation as well as cross-correlation of rainfall between locations in space. Another challenge is how well the process of rainfall amounts is associated with the process of rainfall occurrence (i.e. the wet-dry pattern). As a result, spatial rainfall models may produce unrealistic spatial patterns (Leonard et al., 2008); may have a scope restricted to individual storm events (Qin, 2010); or may have a scope restricted to just one aspect of the rainfall, such as the extremes (Davison et al., 2012). Improvements to rainfall models should seek to produce realistic patterns, with as few parameters as possible, for long time periods and for all relevant properties such as totals and extremes at multiple scales.

An issue related to the limited development of spatial rainfall models has been the compartmentalised and ad-hoc approach to model evaluation. Rainfall modelling is not an end in and of itself but a step towards resolving a hydrological problem. Despite this, there are almost no evaluations of simulated rainfall in terms of simulated streamflow. Hydrological modelling has a broad range of applications, each with its own set of streamflow characteristics of interest. For example, extreme streamflows are critical to flood frequency analysis (Brocca et al., 2011), low streamflows are critical to environmental flow evaluation (McMahon and Finlayson, 2003) and long-term wet- and dry-periods are critical for evaluation of drought risk (Frost et al., 2007). Streamflow characteristics may occur over a broad range of space and time scales and it is often not immediately clear which rainfall

characteristics are most required for reliably reproducing them. The best way to resolve this issue (and to demonstrate the importance of spatial rainfall) is to develop streamflow based methods of assessment for rainfall models.

1.3 Limitations of Spatial Rainfall in Hydrological Modelling

Hydrological modelling is an essential tool to assess the impact of natural hazards and their associated risks. Broadly, there are two approaches for hydrological design: event-based approaches and continuous simulation. Spatial limitations are relevant to both design approaches since floods (and droughts) are the result of a complex relationship between spatial rainfall and heterogeneous catchment conditions (Segond et al., 2007). Although spatial variability of rainfall is often acknowledged to be a key source of uncertainty in hydrological design (Faures et al., 1995), it is often neglected.

Consider the example of flood estimation using a design-event method. This approach has evolved from a legacy of methods reliant on point rainfall data and a series of assumptions aimed at reproducing the flood hydrograph peak (peak discharge). As a result the method entails many assumptions that overlook the complexity of spatial rainfall: (i) an areal reduction factor (ARF); (ii) a uniform distribution of rainfall over the catchment; (iii) rainfall estimated from a point rather than multiple gauges; and (iv) a flood frequency estimate that is forced to match the rainfall frequency that happens to be observed at the location of a rainfall gauge.

Consider the example of flood estimation using continuous simulation with stochastic spatial rainfall. To evaluate hydrological risks, continuous simulation requires long-term rainfall time-series. For example, to achieve a prediction error of less than 20% in the 1 in 100 year flood estimate, at least 10,000 years of rainfall are required (Li et al., 2014). While this type of simulation has been achieved with point rainfall models, there are few models capable of simulating spatial rainfall fields for long periods of time. There are a variety of rainfall models with spatiotemporal features (Groppelli et al., 2011; Northrop, 1998; Seed et al., 2013; Seed et al., 1999; Zhang and Switzer, 2007) but their complexities often make them unsuited for long-term continuous simulation. Models which rely on weather radar to represent the spatial structure of storm events and their evolution in space and time have typically been restricted to event-based simulation (Kim et al., 2009; Qin, 2010; Seed et al.,

2013). There is an ongoing need to develop spatiotemporal rainfall models which can efficiently generate long-term rainfall sequences for use in continuous simulation approaches.

1.4 Overall Research Objectives

The overall aim of this research is to improve the representation of spatial rainfall in hydrological models to provide better estimates of risk. Three specific research objectives have been identified, each of which has a number of sub-objectives:

Objective 1 – Direct spatial rainfall: To demonstrate a new method of estimating spatial extreme rainfall and improve the representation of the spatial variability in event-based hydrological approaches (*Paper 1*). This method is termed the Intensity-Frequency-Duration-Area (IFDA) approach.

> *Objective 1.1:* To demonstrate the utility of IFDAs and evaluate their spatial variation for eleven study regions across a range of climates within Australia.
>
> *Objective 1.2:* To evaluate the differences in spatial rainfall estimates obtained using the ARF-based approach and the 'true' spatial rainfall, the IFDA.
>
> *Objective 1.3:* To evaluate the characteristics of differences between the IFDA and the ARF-based approach with respect to properties of the extreme spatial rainfall such as area, frequency, duration and region.

Objective 2 – Stochastic spatial rainfall model: To develop and evaluate a new stochastic spatial rainfall model for use in continuous hydrological modelling (*Paper 2*).

> *Objective 2.1:* To develop a parsimonious latent variable rainfall model to generate spatial rainfall fields continuously; and
>
> *Objective 2.2:* To develop a comprehensive evaluation approach that provides a systematic, succinct and transparent method to assess and summarise rainfall model performance.
>
> *Objective 2.3*: To perform a comprehensive evaluation of the latent variable rainfall model's performance across a range of time and spatial scales.

Objective 3 – Hydrological evaluation: To develop a new framework for evaluating stochastic rainfall models based on their performance in simulating streamflow (*Paper 3*).

Objective 3.1: To introduce a framework for virtual hydrological evaluation of stochastic rainfall models.

Objective 3.2: To introduce two different tests which are part of the framework: an integrated test and a splice test. Combined use of these tests allows streamflow discrepancies to be attributed to their original source in the rainfall model according to site and month.

Objective 3.3: To use the virtual hydrological evaluation framework to evaluate a stochastic rainfall model and contrast the outcomes with standard evaluation methods.

1.5 Thesis Organisation

This contains five chapters with the main body of research presented in **Chapter 2** to **Chapter 4**. These chapters correspond to three journal papers (Bennett et al., 2016a; 2016b; 2016c).

Chapter 2 presents a new approach for the direct estimation of spatial extreme rainfall – IFDA (*Paper 1*). These new IFDA relationships are compared against current ARF-based spatial extreme rainfall estimates.

A new stochastic spatial rainfall model that uses a latent Gaussian variable approach is introduced in **Chapter 3**. In this Chapter a comprehensive evaluation of the rainfall model's performance is also presented (*Paper 2*). The supplementary materials referenced in *Paper 2* are reproduced in Appendix B.

In **Chapter 4** the virtual hydrologic evaluation framework for assessing stochastic rainfall models is introduced and contrasted against traditional approaches used to assess stochastic rainfall models (*Paper 3*).

The section and figure numbers have been modified in line with University guidelines but the manuscript material is otherwise unchanged. A copy of *Paper 1* is reproduced in Appendix A as published. *Paper 2* and *Paper 3* are under review at

present.

Conclusions are provided in **Chapter 5** and include a discussion of the contributions, limitations and future directions for the research.

Chapter 2

Estimating Extreme Spatial Rainfall Intensities (Paper 1)

Bree Bennett, Martin Lambert, Mark Thyer, Bryson C. Bates, Michael Leonard

Journal of Hydrologic Engineering, Volume 21 Issue 3

Statement of Authorship

Title of Paper	Estimating extreme spatial rainfall intensities.
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Bennett, B., Lambert, M., Thyer, M., Bates, B.C., Leonard, M., 2016. Estimating Extreme Spatial Rainfall Intensities. Journal of Hydrologic Engineering, 21(3): 04015074. DOI:10.1061/(ASCE)HE.1943-5584.0001316.

Principal Author

Name of Principal Author (Candidate)	Bree Sarah Bennett
Contribution to the Paper	Implementation and development of approach, visualisation and interpretation of results, preparation of manuscript, preparation of response to reviewers and acted as corresponding author.
Overall percentage (%)	85%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 2/6/16

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Martin Lambert
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.
Signature	Date 2/6/16

Name of Co-Author	Mark Thyer
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.
Signature	Date 2/6/16

Name of Co-Author	Bryson Bates		
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.		
Signature		Date	2/6/2016

Name of Co-Author	Michael Leonard		
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.		
Signature		Date	2/6/2016

Abstract

Determining the impact of catchment flooding requires an estimate of extreme spatial rainfall intensity. Current flood design practice typically converts a point estimate of rainfall intensity into a spatial rainfall intensity using an areal reduction factor, assumed constant across an entire region. Areal reduction factors do not explicitly consider regional variations in extreme rainfall. Here, a new approach for spatial estimates of extreme rainfall is introduced which directly incorporates the spatial area (A) into an intensity-frequency-duration relationship (IFD). This IFDA approach uses spatial rainfall fields to overcome shortcomings of the areal reduction factor by explicitly incorporating spatial variations in the extreme rainfall intensity. The IFDA approach is evaluated for eleven case study regions in Australia, across climates (tropical to Mediterranean), areas (25 to 7225 km^2), durations (1 to 4 days) and average recurrence intervals (ARI 2 to 100 years). The change in extreme spatial rainfall with respect to area varies markedly within each region suggesting that constant areal reduction factors for a region are inappropriate. Constant areal reduction factors are shown to underestimate extreme spatial rainfall intensities by 5% to 15%. The IFDA approach avoids these biases and is a promising new technique for use in design flood estimation.

2.1 Introduction

Of all natural disasters, floods have the highest global cost and affect the most people (Kousky and Walls, 2014; Miller et al., 2008; Strömberg, 2007). A key input for estimating flood risk is the spatial intensity of extreme rainfall events over a catchment. Current techniques for estimating extreme spatial rainfall rely on the use of an areal reduction factor (ARF) to convert intensity estimates of extreme point rainfall to extreme spatial rainfall. It is common practice to ignore the spatial variation in rainfall intensity and assume a fixed ARF applies over large regions. The aim of this paper is to introduce a new approach that explicitly incorporates the area and variation of spatial rainfall, referred to as Intensity Frequency Duration Area (IFDA).

The IFDA approach utilises spatially interpolated rainfall grids to directly provide an estimate of how spatial rainfall intensities vary with duration, frequency and area for a location. An IFDA adds the extra dimension of area (A) to an IFD curve to account for spatial variation in intensity over a catchment.

Existing design methods rely on interpolated maps of intensity-frequency-duration (IFD) rainfall calculated from point rainfall data. The ARF is then used to determine the areal average rainfall intensity from the point rainfall. The factor is defined as the ratio of extreme rainfall at a point to the extreme rainfall over an area for a given frequency (Asquith and Famiglietti, 2000). In brief, the ARF is a spatial correction factor used to fix limitations of a design methodology focused around pointwise rainfall estimates (IFDs).

The validity of assuming a fixed ARF for a region has been previously questioned (Catchlove and Ball, 2003; Durrans et al., 2002). However, these evaluations of ARF spatial variation have typically been limited to a single location (Catchlove and Ball, 2003) or climatic region (Durrans et al., 2002). In contrast, this study will evaluate these impacts across multiple climate regions.

This paper proposes that estimates of extreme spatial rainfall are better and more efficiently obtained by directly using gridded spatial rainfall in preference to scaling point-wise extreme rainfall. IFDAs, as direct spatial rainfall estimates, overcome the assumption that a single scaling relationship is sufficient for a region.

The key objectives of the paper are:

1. To demonstrate the utility of IFDAs and evaluate their spatial variation for eleven study regions across a range of climates within Australia.

2. To evaluate the differences in spatial rainfall estimates obtained using the ARF-based approach and the 'true' spatial rainfall, the IFDA.

3. To evaluate the characteristics of differences between the IFDA and the ARF-based approach with respect to properties of the extreme spatial rainfall such as area, frequency, duration and region.

2.2 The IFDA approach

The method of constructing IFD curves has been modified to directly incorporate spatial extent. In this approach, each IFDA relationship corresponds to a grid point of specific longitude and latitude. For each grid cell, a set of IFDAs are produced for a range of areas and durations. The approach is generic and is able to be applied to any gridded spatial rainfall data source such as spatially interpolated point rainfall (Bárdossy and Pegram, 2013), radar rainfall, and downscaled rainfall from climate model simulations.

An IFDA curve is constructed by first designating a grid cell within the subject region (Fig. 2-1a). This grid cell of specific latitude and longitude is the central point around which the spatial variation in rainfall is considered. A fixed area is designated around the centre grid cell (Fig. 2-1b). This designated area is illustrated as square in Fig. 2-1, but the procedure is general and can accommodate areas of any shape.

For each time step the total rainfall in the designated area is summed over the contributing grid cells (Fig. 2-1c). The increments of spatial rainfall are then summed over the duration and converted to intensity. This results in a time series of spatial rainfall intensities for the designated location, duration and spatial extent. From this time series the annual maximums are extracted and a generalised extreme value distribution is fitted (Green et al., 2012; Jordan et al., 2011; Siriwardena and Weinmann, 1996). The fitted distribution forms the IFDA corresponding to the centre grid cell for the designated area and duration.

The IFDA procedure for a single location can be summarised mathematically as follows: Let R_i denote the set of point rainfall intensities for a given duration for all points x in a spatial domain Ω and time increments t in Y_i, the i^{th} year (i = 1, ..., n), such that $R_i = \{R_i(x,t) | x \in \Omega, t \in Y_i\}$. Let $\text{annmax}[\ ]$ be a function that takes the maximum value across all the time increments t in Y_i. For the i^{th} year, the spatial annual maximum rainfall intensity for a catchment domain Ω with area A is defined as

$$C_i(A,t) = \frac{1}{A}\text{annmax}\left[\int_{\Omega} R_i(x,t)\,dx\right] \tag{1}$$

If the set of extracted events, $C_i = \{C_1, \ldots, C_n\}$ is ordered in terms of magnitude and assigned a frequency, then $C_F(A,t)$ denotes the spatial rainfall intensity for the catchment of area A and frequency F. The rainfall intensities of defined frequency, duration and area compose the IFDA relationship for that location. The repetition of this process at all locations throughout the domain creates the field of IFDAs and each cell within the region has its own set of IFDA relationships.

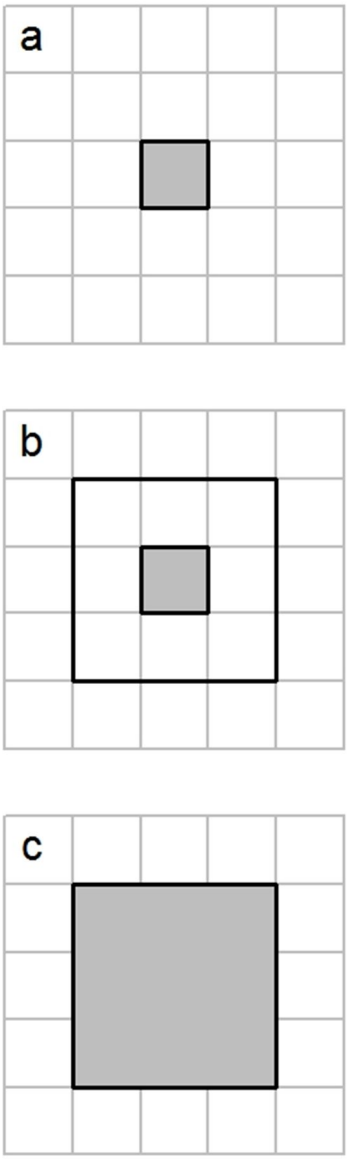


Fig. 2-1. Schematic of IFDA area designation (a) centroidal cell identified (b) area designated around centroidal cell (c) area consisting of aggregated contributing cells

2.3 Case study data

The Australian Water Availability Project (AWAP) gridded rainfall database provides daily rainfall depths on a 5 km square grid across Australia (Raupach et al., 2012). The grids are an interpolated product based on daily point rainfall records and are available from 1900 onwards. This study uses two subsets 1900-2011 and 1973-2011, where the shorter period provides a check against any changes in gauge density over the period of the longer record.

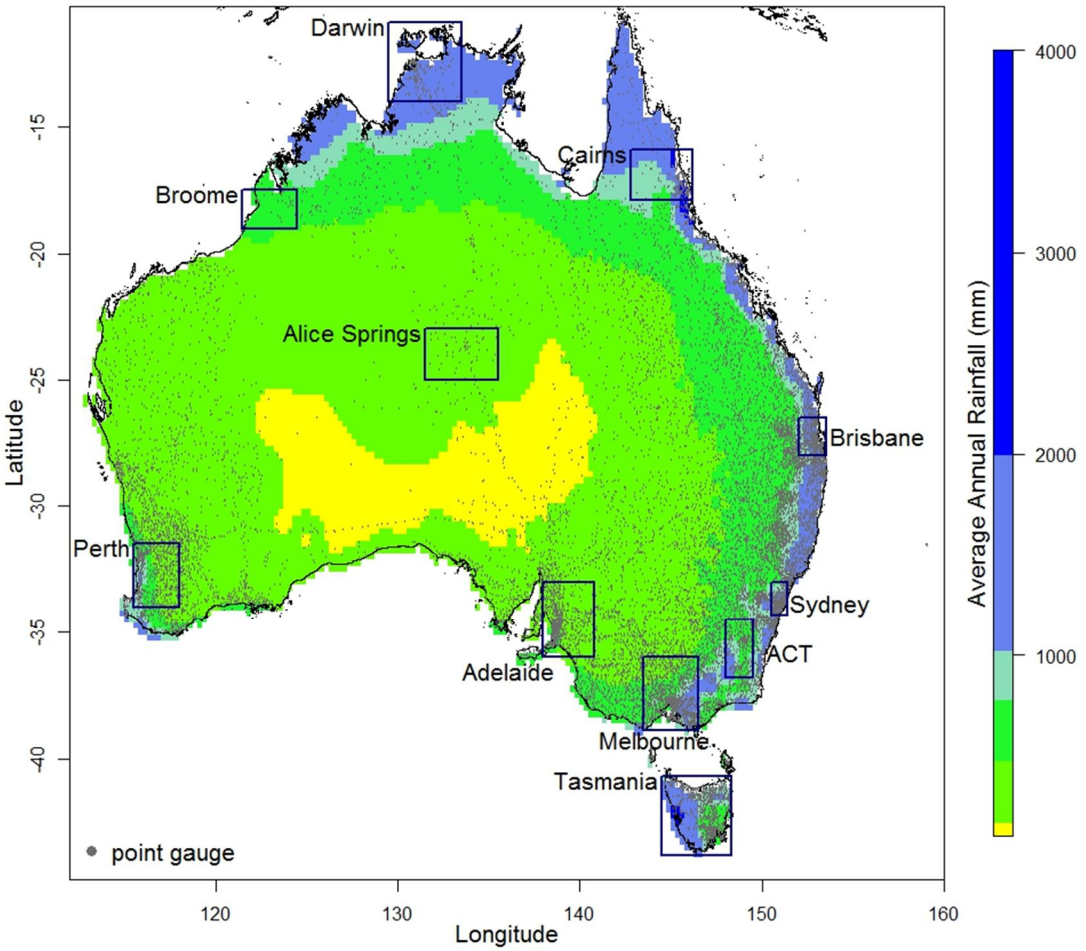


Fig. 2-2. Location of study regions with annual rainfall isohyets (mm)

A conversion factor of 1.15 is applied in this study to account for the restricted time period of the daily observations (e.g. 9am to 9am). Over restricted periods the maximum is lower than 24 hour totals that have been aggregated over an unrestricted time period (Boughton and Jakob, 2008; Jakob et al., 2005; van Montfort, 1990). For studies that focus on shorter durations (Catchlove and Ball, 2003) there is also an implicit need to apply the conversion factor to any daily observations that are used.

The IFDA approach is generic and can be applied to any spatial rainfall data set. For the purposes of this paper the gridded AWAP dataset is regarded as the best estimate of spatial 'truth'. While spatial interpolation can introduce artefacts (King et al., 2013), the development of better spatial interpolation procedures is a separate (and important) research area. An advantage of the IFDA is that it can easily be updated to take advantage of different or better data products, whereas the current ARF approach cannot (further discussion in Section 2.6.3).

Case study regions were selected for detailed analyses based on their gauge density and to ensure coverage of different climates (tropical: Broome, Cairns, Darwin; sub-tropical: Brisbane, Sydney; temperate: Melbourne, Tasmania, ACT; Mediterranean: Adelaide, Perth; and arid: Alice Springs) (Fig. 2-2). This paper focuses on the regions of Sydney and Melbourne to illustrate the IFDA relationships. Results from additional regions are included to illustrate that the same general behaviours of the IFDA relationship are exhibited at other locations. The analyses demonstrate the inappropriateness of fixed regional ARFs for estimating extreme spatial rainfall intensities. Melbourne and Sydney have especially dense observation networks, therefore the impact on the results of the AWAP interpolation procedure is minimised.

2.4 Methodology

To achieve objective 1, IFDA relationships are derived for all study regions and their behaviour is examined to evaluate variations in extreme spatial rainfall. To evaluate the differences between an ARF-based and IFDA approaches (objective 2), the distribution of at-site ARFs calculated using the gridded rainfall are compared with currently recommended fixed regional ARFs (ARF_R). Furthermore, the error induced by using ARF_R is evaluated as a percentage error of the direct estimate of extreme spatial rainfall. The characteristics of the resultant percentage errors are assessed against extreme spatial rainfall properties such as area, frequency, duration, and region (objective 3).

2.4.1 IFDA calculation

IFDAs were calculated following the method in Section 2.2 for all study regions and for areas of 25, 225, 625, 1225, 2025, 3025, 5625 and 7225 km^2. Durations of 1, 2, 3 and 4 days were investigated and results are presented for average recurrence intervals (ARIs) of 2 to 100 years. The IFDAs are presented as both location specific relationships and intensity fields. These intensity fields show the rainfall intensity for the specific IFDA area and duration plotted at the centre pixel. Therefore, as the area around the centre grid increases the boundaries of the region contract to ensure no points outside the original region are sampled in the calculation of intensities for larger areas.

2.4.2 Evaluation of fixed area ARFs against the IFDA approach

Objective 2 is to evaluate the differences in spatial rainfall estimates obtained from the ARF-based approach and the direct estimate of spatial rainfall obtained using the IFDA approach. Therefore, it assesses whether current practice is appropriate for providing robust estimates of extreme spatial rainfall.

Extreme spatial rainfall intensities are traditionally estimated using a representative extreme point rainfall and a regional ARF. The representative point rainfall of a certain frequency is denoted here as R_F, and there are two general approaches for defining it. The first is where the representative extreme point rainfall intensity is taken from a single location, typically the centre of the catchment area, denoted $\widehat{R}_F^c$. The second is where the spatial average of the extreme point rainfall values within the catchment domain is taken as the representative extreme point rainfall, denoted $\widehat{R}_F^{\overline{pt}}$. That is

$$\widehat{R}_F^{\overline{pt}}(t) = \frac{1}{A}\int_{\Omega} R_F(x,t)\,dx \tag{2}$$

This second approach is used in preference to the centre extreme point rainfall intensity when there is a strong spatial rainfall trend (e.g. due to elevation gradient) within a catchment. In this study, the impact of using both approaches is evaluated.

To evaluate whether the use of fixed regional ARF values, ARF_R, are appropriate for providing robust estimates of extreme spatial rainfall intensities, ARFs were calculated for all locations within each study region. The calculated ARFs are the values required at each sampled location to scale the representative extreme point-wise rainfall to obtain the true extreme spatial rainfall intensity of equivalent ARI. These at-site ARFs were then evaluated against currently recommended ARF_R values.

ARFs are a ratio of spatial rainfall to a representative point rainfall of equal recurrence interval and can be defined as

$$ARF_{(A,F,t)} = \frac{C_F(A,t)}{\widehat{R}_F(t)} \tag{3}$$

where $C_F(A,t)$ is the spatial rainfall intensity of specified area, A, frequency of F and duration t. This definition is not a storm-centred ARF which is based on the point and spatial rainfall of an individual storm (Svensson and Jones, 2010). Rather it is a statistical ARF which relates representative extreme point rainfall to extreme spatial rainfall with equal recurrence interval (Allen and DeGaetano, 2005; Durrans et al., 2002; Myers and Zehr, 1980; Sivapalan and Blöschl, 1998). ARFs are constructed by using Eq. (3) with a representative extreme point rainfall intensity for the two different cases outlined above. Where the centre extreme point rainfall intensity, $\hat{R}_F^c$ was used in Eq. (3), the resulting at-site ARF has been denoted as ARF_c, where the spatially averaged extreme point rainfall, $\hat{R}_F^{\overline{pt}}$, is used the at-site ARF is denoted as $ARF_{\overline{pt}}$. For this analysis the single grid cell (25 km^2 area) was nominated as the highest resolution approximation of point rainfall.

The at-site ARF_c and $ARF_{\overline{pt}}$ values are presented as boxplots to demonstrate the range and distribution of calculated ARF values throughout a region, with a comparison against the fixed regional ARF_R. These fixed regional ARF_R include values published in Pilgrim (1987) (following Myers and Zehr, 1980), updated values for New South Wales (Jordan et al., 2011) and updated values for Victoria (Siriwardena and Weinmann, 1996). The values used for reference in this study are for the areal rainfall intensity at 25 km^2.

To evaluate the bias present in using the ARF-based approach, the percentage error resulting from using this estimate compared against the spatial rainfall total is calculated. The percentage error is calculated as

$$\%Err = \frac{\hat{C}_F - C_F}{C_F} \times 100 \tag{4}$$

in which C_F is the directly obtained spatial rainfall intensity (i.e. the IFDA value at the required location with corresponding frequency F, Section 2.5.1) and $\hat{C}_F$ is the spatial rainfall intensity estimated using a representative extreme point rainfall via Eq. (5).

$$\hat{C}_F = ARF_F \hat{R}_F(t) \tag{5}$$

For the case $\hat{R}_F^c$, the estimated extreme spatial rainfall and resulting percentage error are denoted $\hat{C}_F^c$ and $\%Err_c$ respectively. For the case $\hat{R}_F^{\overline{pt}}$, the estimated extreme spatial rainfall and percentage error are denoted $\hat{C}_F^{\overline{pt}}$ and $\%Err_{\overline{pt}}$ respectively.

2.4.3 Identification of extreme spatial rainfall properties influencing errors in the ARF-based approach

To evaluate the differences between IFDA and ARF-based approaches four densely gauged regions are considered (Sydney, Brisbane, Melbourne and Perth). The comparison considers the two cases of $\hat{R}_F^c$ and $\hat{R}_F^{\overline{pt}}$. Specifically the coefficient of variation between the two cases is constructed for ARFs and percentage errors. Changes with respect to properties such as catchment area, frequency and location are assessed to evaluate which has the largest influence (Section 2.5.3).

2.5 Results

2.5.1 IFDA analysis

At each location within a region the IFDA relationship showed different trends with area. In all regions there were some locations with a large increase in intensity with increasing area, while other locations showed little or no increase in intensity with increasing area. For the Sydney region, an ARI of 50 years and duration of 1-day, Fig. 2-3a indicates a change of 41% between an area of 25 km^2 and 7225 km^2 whereas Fig. 2-3b illustrates little change in intensity with area. These locations are indicated on Fig. 2-3c and Fig. 2-3d as triangular and square points. The intensity patterns over the region change as the IFDA area increases. It is this spatial variation in IFDAs which produces the different IFDA curve behaviours exhibited in Fig. 2-3a and Fig. 2-3b. Fig. 2-4 similarly shows the change in IFDA with area for the Melbourne region. It similarly illustrates that extreme spatial rainfall intensity patterns change as IFDA areas increase. Again in Fig. 2-4 the triangle and square indicate locations at which intensity is highly variable or less variable with area, respectively. Spatial variation in IFDA curves was similarly observed for all other regions (not shown).

The changing spatial patterns of rainfall intensity for different areas can be clearly seen (Fig. 2-3c, Fig. 2-3d, Fig. 2-4c, Fig. 2-4d). The relationship between intensity and area is spatially heterogeneous and this leads to spatially heterogeneous fields of ARF values.

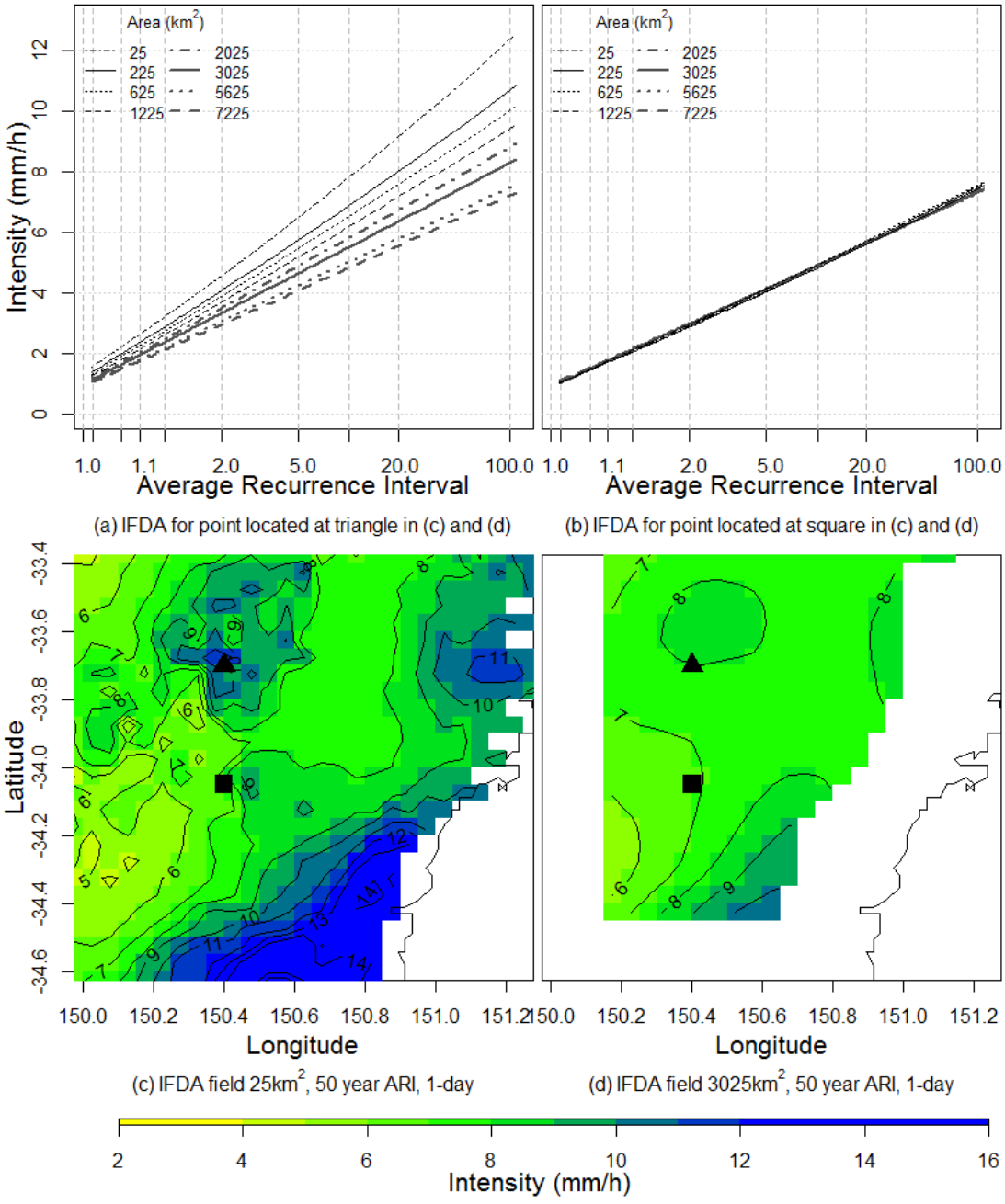


Fig. 2-3. Sydney region IFDA 1-day (a) 150.400°E 33.700°S (b) 150.400°E 34.050°S (c) IFDA field 25 km², 50 year ARI, 1-day (d) IFDA field 2025 km², 50 year ARI, 1-day

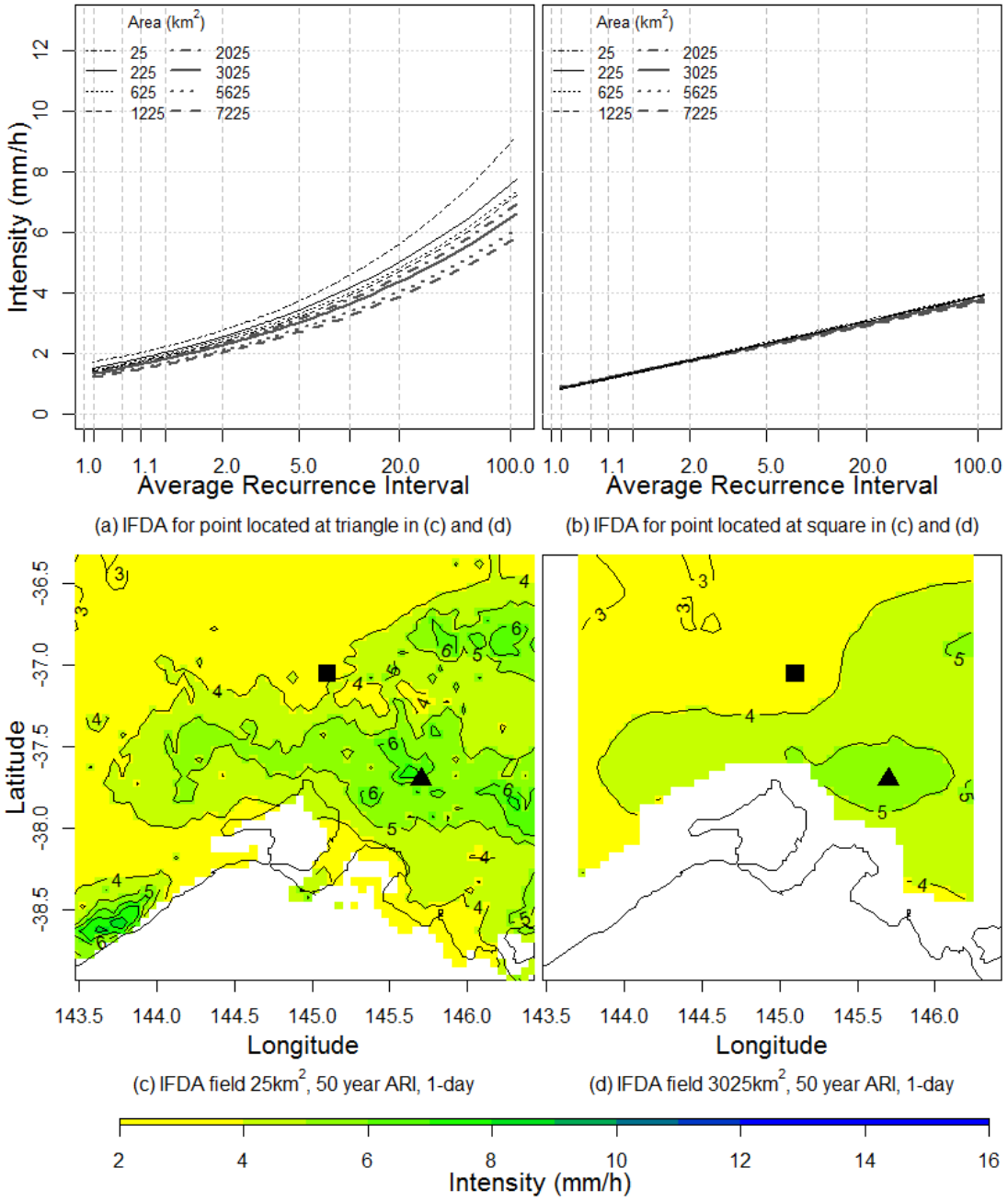


Fig. 2-4. Melbourne region IFDA 1-day (a) 145.700°E 37.700°S (b) 145.100°E 37.050°S (c) IFDA field 25 km^2, 50 year ARI, 1-day (d) IFDA field 3025 km^2, 50 year ARI, 1-day

2.5.2 Evaluation of the differences between estimates of extreme spatial rainfall obtained using the IFDA and ARF-based approaches

Whereas typical flood designs rely on a single ARF value for a given region, there is often considerable variation. To demonstrate this, Fig. 2-5 summarises the spatially heterogeneous ARF_c fields as boxplots, for all regions, an ARI of 50 years, a 1-day duration and areas of 625 km^2 and 3025 km^2. The boxplot whiskers extend to the

10% and 90% limits. The derived ARF_c distributions are compared with currently recommended ARF_R values for the Australian regions. The spread of these distributions varies from small to large and is approximately symmetrical. Values greater than one occur frequently. An ARF greater than one implies that the surrounding spatial rainfall intensity was greater than the individual centre point rainfall intensity for that frequency and duration.

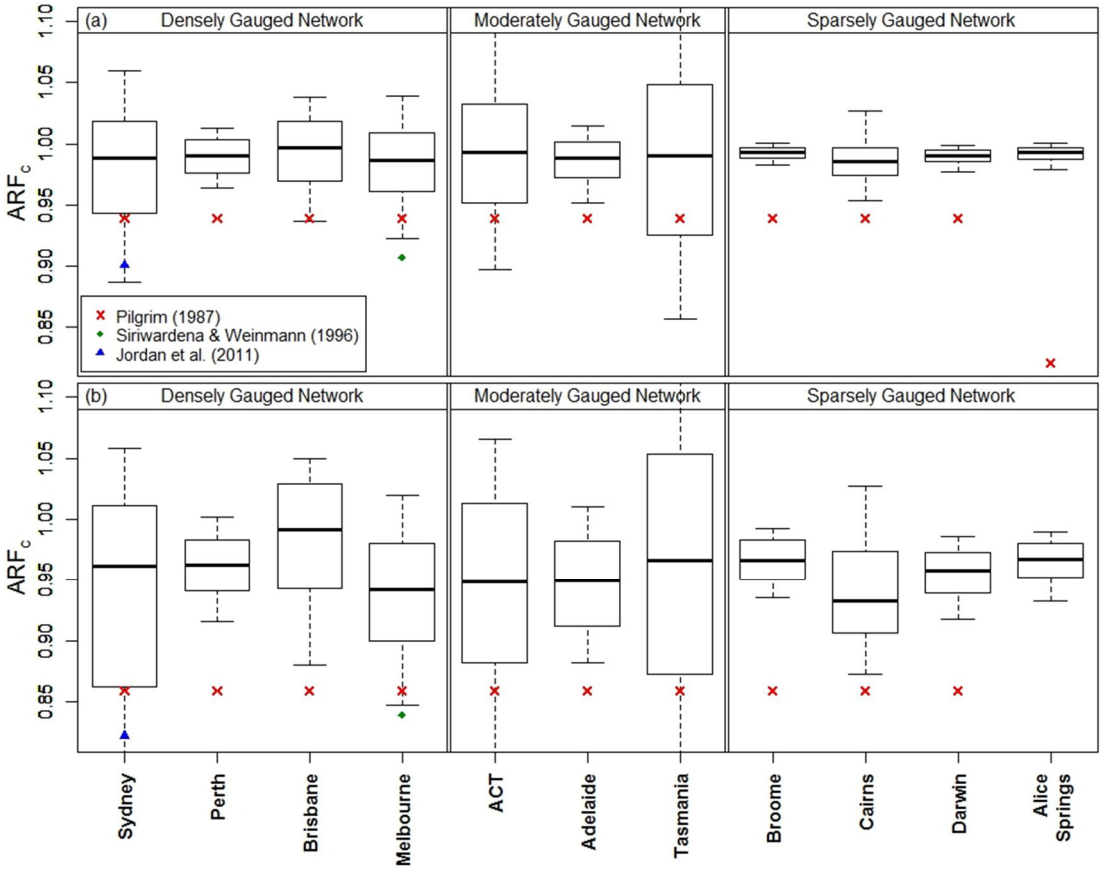


Fig. 2-5. ARF_c boxplots for 1-day, 50 year ARI rainfall (a) 625 km^2 (b) 3025 km^2 (10% and 90% limits shown)

Fig. 2-5 shows that although the regional ARF_R values lie within the range of the derived ARF_c distribution, the fixed values lie below the mean, often significantly so. For example, in Fig. 2-5b the mean ARF_c sits above the ARF_R value of Jordan et al. (2011) by approximately 14% for the Sydney region. The ARF_c distributions exhibit a number of distinctive features. They may be highly variable (Fig. 2-5a Sydney, Tasmania and Fig. 2-5b Sydney, Melbourne, ACT, Tasmania), indicating that ARF_R may differ significantly from ARF_c values for the majority of locations

throughout a region. Alternatively the ARF_c distribution may possess a narrow range of values centring on one (Fig. 2-5a Perth, Broome, Darwin) implying that the ARF_R are biased. Additionally, many ARF_c distributions exhibited values significantly greater than one (Fig. 2-5) indicating that it is common for extreme spatial rainfall surrounding a point to exceed the point rainfall. This behaviour was observed for all studied durations.

Fig. 2-6 presents similar results, using $ARF_{\overline{pt}}$ values. The intention of this alternative is to better account for the spatial variation in point rainfall over the area. However, this modification still leads to significant differences from the recommended ARF_R values. Notably, the $ARF_{\overline{pt}}$ are less variable and have less values above one.

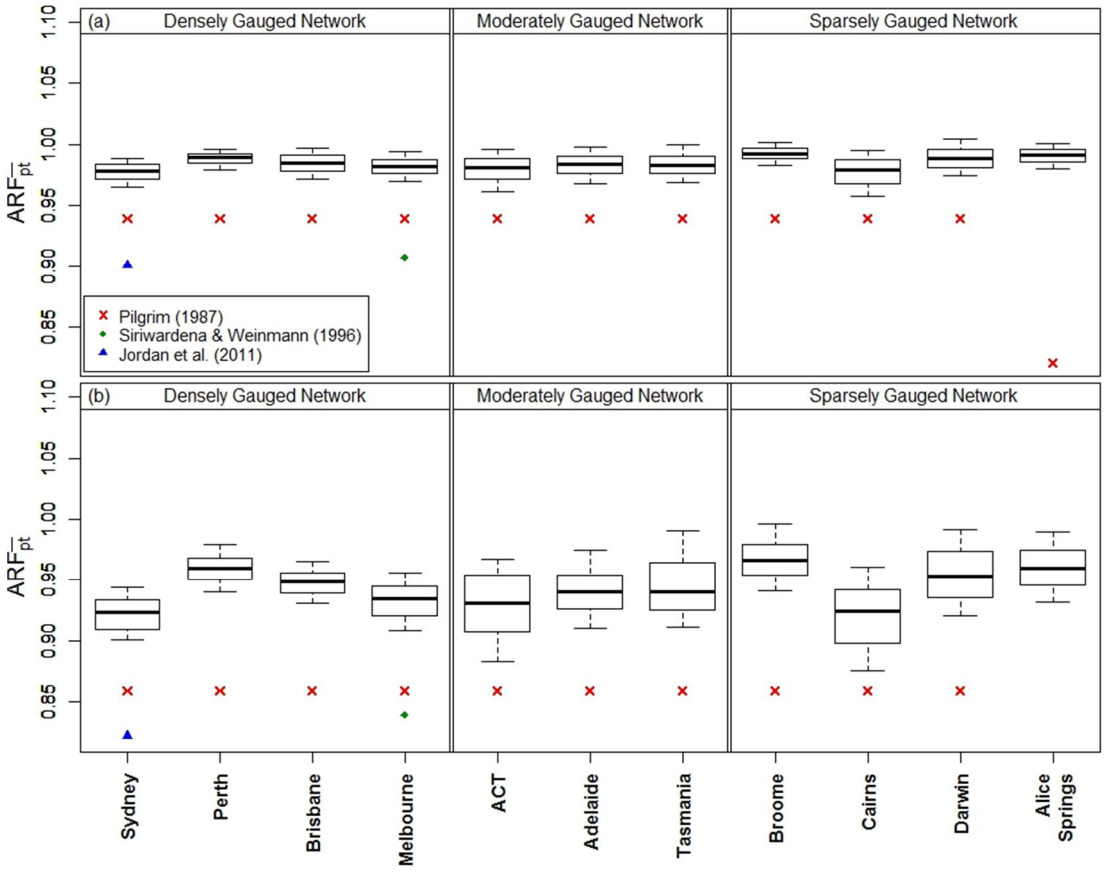


Fig. 2-6. $ARF_{\overline{pt}}$ boxplots for 1-day, 50 year ARI rainfall (a) 625 km^2 (b) 3025 km^2 (10% and 90% limits shown)

Fig. 2-7 shows percentage errors ($\%Err_c$) resulting from the central point rainfall combined with ARF_R values across areas of 625 km^2 and 3025 km^2. As a result, the extreme spatial rainfall is consistently underestimated for the majority of locations in all regions. Fig. 2-8 similarly shows the percentage error for the point-averaged rainfall method ($\%Err_{\overline{pt}}$) and this technique underestimates the extreme spatial rainfall for 95% to 100% of locations in the region. Therefore, the method of averaging point rainfall does not improve ARF-based techniques.

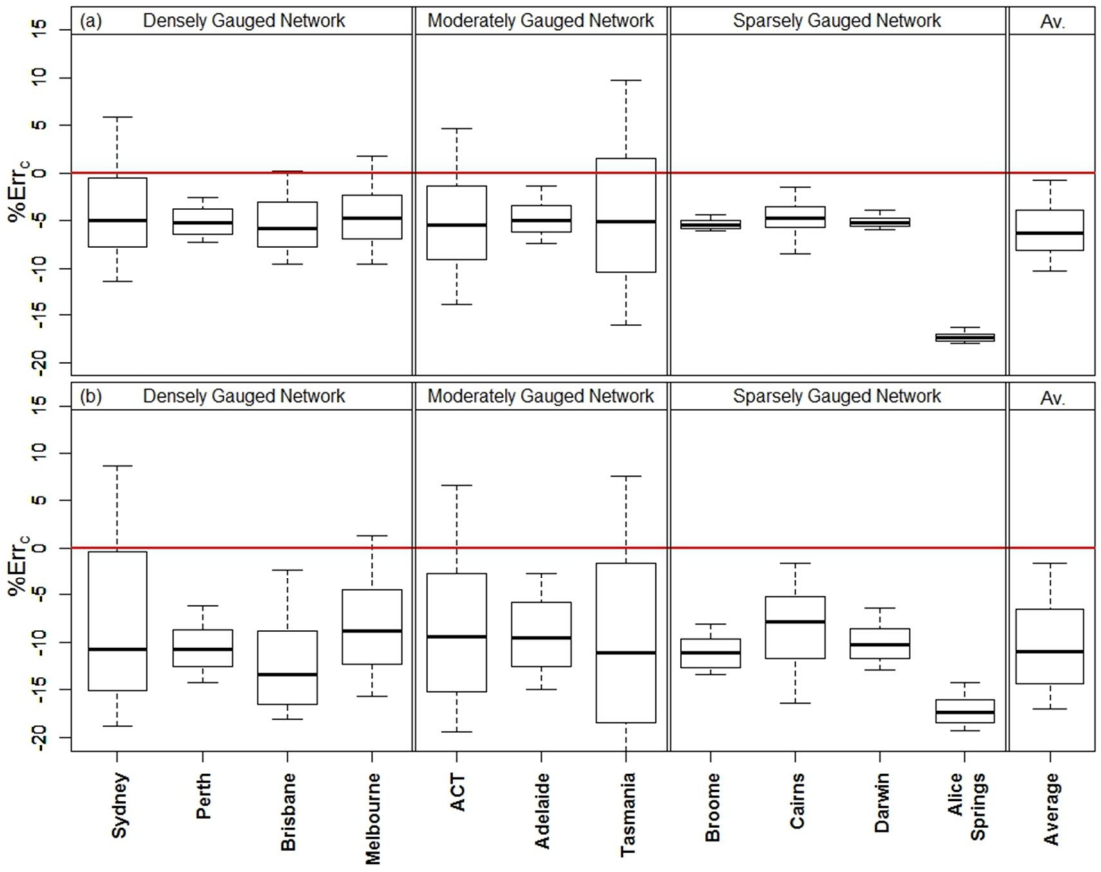


Fig. 2-7. $\%Err_c$ for 1-day, 50y ARI rainfall (a) 625 km^2 (b) 3025 km^2 (10% and 90% limits shown)

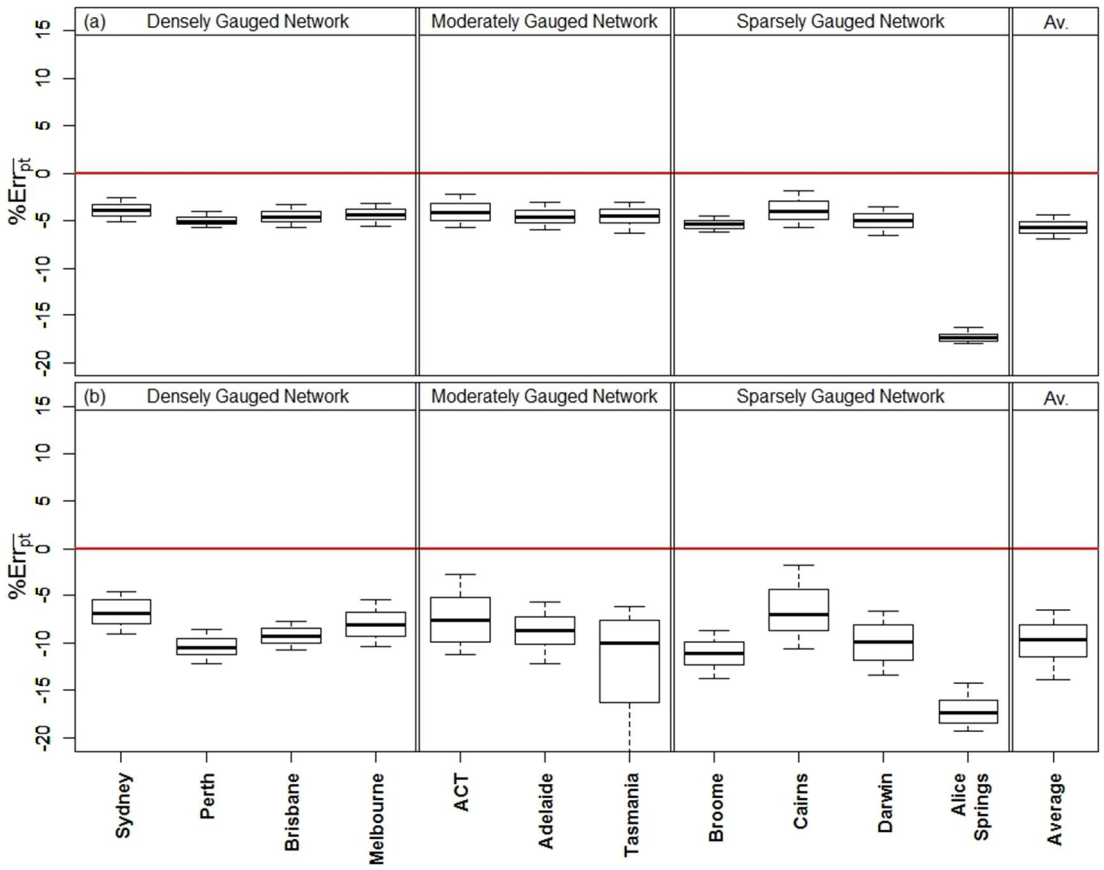


Fig. 2-8. $\%\overline{Err}_{pt}$ for 1-day, 50y ARI rainfall (a) 625 km^2 (b) 3025 km^2 (10% and 90% limits shown)

A further comparison was conducted as check for any influence of changing gauge density over the length of the record. Therefore, the analyses were repeated using a shorter 36 year series of daily grids corresponding to the latter third of the record which has the highest gauge density. Fig. 2-9 provides an example summary of calculated ARF and percentage error values for the Sydney region for an ARI of 25 years. The analyses did not find any significant effect of changing gauge density on the results.

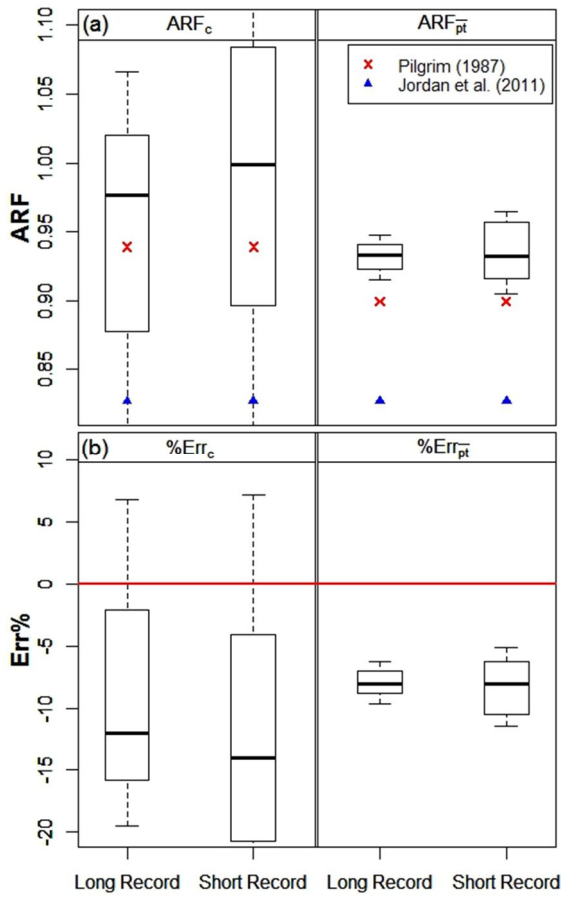


Fig. 2-9. Sydney ARF and percentage error boxplots for 25y ARI, 1-day rainfall comparing results of the long and short series analysis for an area of 3025 km^2 (10% and 90% limits shown) (a) ARF_c and $ARF_{\overline{pt}}$ (b) $\%Err_c$ and $\%Err_{\overline{pt}}$

2.5.3 Evaluation of extreme spatial rainfall properties influencing the differences between the ARF-based and IFDA approaches

Given the variability of the ARF values over the region, it is important to understand situations which are most affected. To assess this, the change in the coefficient of variation of the at-site ARFs against catchment area and ARI was evaluated. Fig. 2-10 shows typical results for both ARF_c and $ARF_{\overline{pt}}$ for four densely gauged regions (Perth, Melbourne, Brisbane and Sydney). In all cases the coefficient of variation increases with catchment area and $ARF_{\overline{pt}}$ is less variable than ARF_c. This observation was consistent for all ARIs and durations (not shown). This demonstrates that there is greater variability in estimates of extreme spatial rainfall produced using a fixed regional ARF, for larger catchment areas.

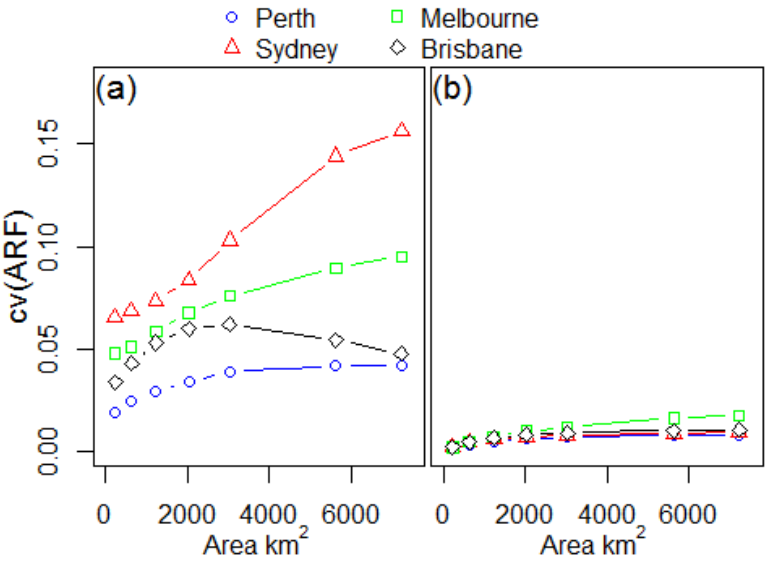


Fig. 2-10. Comparison of coefficient of variation of at-site ARFs against catchment area for 10y ARI, 1-day rainfall (a) ARF_c (b) $ARF_{\overline{pt}}$

Fig. 2-11 compares the mean percentage error $\%Err_c$ against ARI for areas of 625 km^2 and 3025 km^2. It can be seen that there is a bias towards underestimation for all locations and frequencies. For the 3025 km^2 area (Fig. 2-11b) the errors are larger and the differences between regions are more pronounced, especially for less frequent events. These observations are consistent over other catchment areas, regions and for $\%Err_{\overline{pt}}$ (not shown).

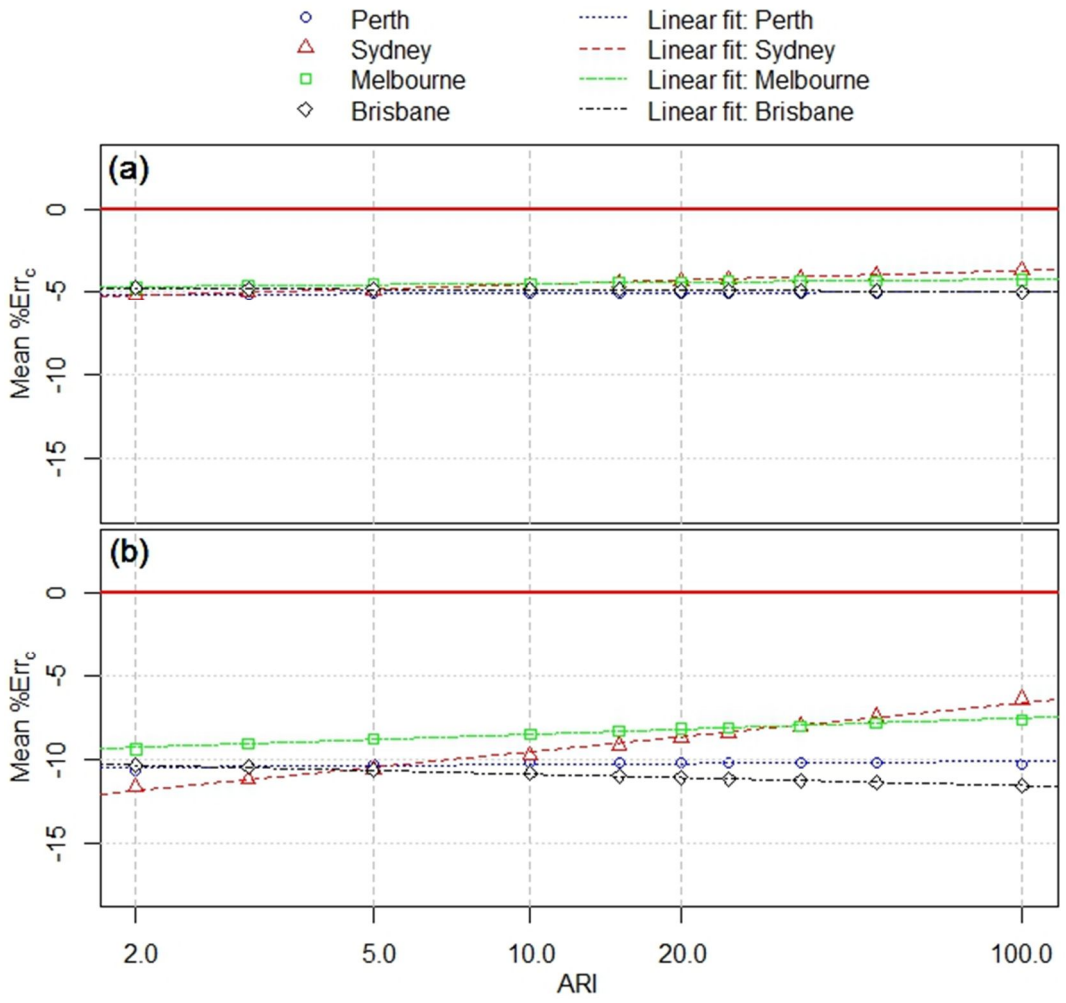


Fig. 2-11. Mean $\%Err_c$ against ARI for 1-day rainfall (a) 625 km^2 (b) 3025 km^2

2.6 Discussion and practical implications

This paper has demonstrated that current ARF-based approaches for estimating extreme spatial rainfall are biased, underestimating rainfall for the majority of locations. This motivates the need for more direct approaches that do not rely on fixed regional ARFs. Underestimating extreme spatial rainfall will typically lead to an underestimate of streamflow. There is a non-linear relationship between the flow and the cost of damage to infrastructure. Therefore, the resulting cost of damage from an underestimate may outweigh the alternative cost of adopting a larger design flood (See Botto et al., 2014). By improving design practice the economic impact and cost to society can be significantly reduced.

2.6.1 Sources of bias

Within a region, the relationship between point and areal rainfall is spatially heterogeneous. For two different locations within a region the scaling of intensity with area can be vastly different (Section 2.5.1). For both ARF approaches, ARF_c and $ARF_{\overline{pt}}$, the derivation of ARFs for each coordinate showed that ARFs varied spatially, leading to a range possible values for a given frequency, duration and spatial extent (Section 2.5.2). The use of a fixed ARF_R contradicts this observed spatial variation. Applying a fixed regional ARF_R introduces bias into estimates of extreme spatial rainfall for the majority of locations (e.g. Fig. 2-8).

The practice of using a fixed regional ARF_R is due to the historical legacy of poor spatial estimates of rainfall. Typically, ARFs have been derived using smaller catchments with higher gauge density and by pooling information from a range of catchments over larger regions (Jordan et al., 2011; Niemczynowicz, 1982; Pilgrim, 1987; Siriwardena and Weinmann, 1996; Yoo et al., 2007). More recently, remotely sensed estimates of rainfall have provided an alternative source to suggest that ARF values are not constant over a region (Durrans et al., 2002). However, the limited length of remotely sensed records prevents definitive conclusions.

This study has relied on spatially interpolated rainfall and was able to show significant spatial variation in ARF values. The challenge with using interpolated data is the inherent spatial smoothing and artefacts introduced by the algorithm used (King et al., 2013). One of the potential impacts of spatial smoothing is the true at-site ARF values may exhibit more distinct spatial variations than observed in the current study. As a result, the bias in at-site estimates of extreme spatial rainfall are likely to be greater than shown in Fig. 2-7 and Fig. 2-8.

It is not often acknowledged that gauge-based estimates of ARFs also rely on an algorithm for constructing the spatial rainfall estimate. Therefore, no matter what method is used there is the potential for influence by an interpolation algorithm. To mitigate this influence we focused on densely gauged regions in this study. One of the advantages of the IFDA approach is that it is generic and independent of the interpolation algorithm. Hence it can easily take advantage of advances in spatial rainfall interpolation techniques.

Both derivation approaches, ARF_c and ARF_{pt}, exhibited values exceeding one, indicating the sampled point rainfall was less than the surrounding rainfall (Sections 2.5.2). This contrasts with the widely used recommendation that ARF are upper bounded at one. Only a few previous studies (e.g. Catchlove and Ball, 2003) have reported ARF values greater than one for three main reasons. Firstly, calculation of an ARF starts with the assumption that extreme point rainfall is more intense than extreme spatial rainfall. This is not always the case, for example, due to the presence of a dominant storm path. Secondly, the focus on obtaining a regional ARF value has meant that spatial variations in extreme spatial rainfall are smoothed out. Thirdly, point gauge data is underrepresented in hard to access areas with strong rainfall gradients (Prudhomme and Reed, 1999; Svensson and Jones, 2010).

2.6.2 Conditions where bias has the greatest impact

The paper demonstrated that a significant change in mean percentage error in estimates of extreme spatial rainfall occurs with increasing area. The largest mean percentage errors were observed for the largest catchments suggesting that current practice is least effective for large catchments. The poor performance for large catchments is likely due to the typically long distances over which rainfall events are correlated at a daily time step. Due to the spatial correlation structure of large rainfall events, the areal reduction of a sampled point rainfall using an ARF_R is likely to underestimate extreme spatial rainfall intensity.

Significant differences in the percentage errors of the extreme spatial rainfall estimates were observed between each region. The difference between the mean percentage errors at different case study locations is greater than the change in mean percentage error across different frequencies (Fig. 2-11). Thus the frequency of the event seems to be a less decisive factor than location in producing errors in extreme spatial rainfall estimates. The poor performance of the ARF-based approach across all studied regions implies that further investigation into the governing extreme spatial rainfall properties specific to each region is required. These extreme spatial rainfall properties may include seasonality, topography and rainfall mechanisms contributing to extreme events.

2.6.3 Advantages and limitations of IFDAs

To summarise advantages and limitations Table 2-1 presents a comparison of the IFDA and ARF-based approaches across a range of criteria.

Table 2-1. Comparison of ARF-based and IFDA Approaches

Item	Criteria	ARF-based Approach	IFDA Approach
1	Spatial scaling relationship	Assumed to be the same for all locations in the region. This causes complications for large catchments or catchments with steep rainfall gradients.	Data driven and allowed to vary throughout the region.
2	Spatial rainfall estimate	Spatial rainfall estimate has conventionally been coupled to the ARF derivation method.	Made directly from spatial gridded estimates. Therefore, the spatial interpolation approach is separate from methodology.
3	Conceptual approach	ARF and IFD isolate separate components of an event that are later combined but can potentially be inconsistent.	Single metric directly incorporates intensity, frequency, duration and area of design event.
4	Reliance on single representative point rainfall	The ARF scales the extreme rainfall at a key location in the region to yield the spatial estimate.	The spatial interpolation is estimated from all gauges in the region without the need to rely on a single location as being representative.
5	Type of method for assigning frequency to design event	Indirectly approximates design intensity based on quantile matching of point and areal rainfall not necessarily derived from same event.	Directly assigns frequency to design event of certain intensity, area and duration. This is more consistent with actual events.
6	Assumptions of approach	Assumes extreme point rainfall intensity is always greater than the spatial rainfall intensity (i.e. ARF<1).	Assumes spatial interpolated estimates are appropriate.

ARF values calculated using point rainfall data are assumed to be spatially homogeneous within a general climatic zone (Jordan et al., 2011; Pilgrim, 1987; Siriwardena and Weinmann, 1996). As IFDAs are location specific estimates of extreme spatial rainfall, any impact of location or climatic region is directly accounted for in the IFDA.

Current practice relies on an ARF and IFD that have been developed separately. This concern was also raised by Panthou et al. (2014) who, in a different approach, sought statistical consistency between ARF and IFD models rather than adopting direct extreme spatial rainfall estimates. The derivation of IFDAs provides unbiased design rainfall estimates, by incorporating the spatial variation in spatial rainfall intensity directly. Furthermore, by estimating extreme spatial rainfall directly the IFDA approach ensures statistical consistency in the extreme spatial rainfall estimate. However, they are not without limitation, for example, the application of the approach to irregular catchment areas and near-miss storms requires further investigation.

An ARF is a ratio of spatial rainfall to a representative point rainfall of equal recurrence interval. Its derivation requires an estimate of spatial extreme rainfall, but the limitations of this interpolation are rarely separated from the limitations of the ARF. It is important to note that while the quality of the spatial interpolation method will influence the derivation of IFDAs as they are direct estimates of rainfall intensity, it is nonetheless preferable to separate out the method of interpolation. Ultimately, the use of IFDAs depends on the availability of high-quality spatial rainfall data sources. The application of IFDAs to real-world catchments can be improved with further development of continuous spatial rainfall models (Kleiber et al., 2012; Leonard et al., 2008), the continued collection and processing of radar data and continued development of interpolated spatial rainfall algorithms (See Bárdossy and Pegram, 2013; Beesley et al., 2009).

Challenges exist in determining the best method to apply IFDAs. This is because as each coordinate within a catchment possesses its own IFDA relationship, for a given ARI/area/duration there will be range of spatial rainfall intensities applicable to the catchment. This characteristic is perhaps a benefit by allowing spatial variability to be incorporated into design estimates.

2.7 Conclusions

The key objective of this paper was to introduce and evaluate the IFDA approach for extreme spatial rainfall estimation. An IFDA, which is calculated directly from spatial rainfall fields, adds the extra dimension of area to an IFD curve to account for spatial variation over a catchment. The paper demonstrated that the standard practice of using a fixed regional ARF introduces bias into estimates of extreme spatial rainfall and discussed subsequent implications. The existing approaches for estimating extreme spatial rainfall were typically in error by 5% to 15% (Fig. 2-7 and Fig. 2-8). Analysis of the characteristics of these errors showed that they varied with region, catchment area and frequency.

Gridded daily rainfall from the Australian Water Availability Project (AWAP) for eleven case study regions in Australia (which span a wide range of climatic zones) was used. The analysis demonstrated that the IFDA relationship varied within regions. This is in contrast with current practice which assumes that a fixed regional ARF is appropriate. The analysis also demonstrated that ARF values were frequently greater than one, due to cases where the extreme spatial rainfall exceeded the intensity of the sampled point rainfall.

The IFDA approach overcomes the shortcomings of existing approaches primarily by avoiding the need to assume a fixed regional ARF value. The IFDA methodology is proposed as a promising technique for obtaining direct and unbiased estimates of extreme spatial rainfall. The application of the method relies on robust methods for interpolating rainfall, and is therefore benefited by improvements to interpolation algorithms.

2.8 Acknowledgments

This work was supported by an Australian Research Council Discovery grant: A new flood design methodology for a variable and changing climate DP1094796. Additional financial support was provided by the CSIRO Climate Adaptation Flagship. The authors gratefully thank Mark Babister for valuable discussions and constructive comments.

2.9 References

Allen, R.J., DeGaetano, A.T., 2005. Areal reduction factors for two eastern united states regions with high rain-gauge density. Journal of Hydrologic Engineering, 10(4): 327-335. DOI:10.1061/(asce)1084-0699(2005)10:4(327)

Asquith, W.H., Famiglietti, J.S., 2000. Precipitation areal-reduction factor estimation using an annual-maxima centered approach. J. Hydrol., 230(1-2): 55-69.

Bárdossy, A., Pegram, G., 2013. Interpolation of precipitation under topographic influence at different time scales. Water Resources Research, 49(8): 4545-4565. DOI:10.1002/wrcr.20307

Beesley, C.A., Frost, A., Zajaczkowski, J., 2009. A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets, 18th World IMACS / MODSIM Congress, Cairns, Australia.

Botto, A., Ganora, D., Laio, F., Claps, P., 2014. Uncertainty compliant design flood estimation. Water Resources Research, 50(5): 4242-4253. DOI:10.1002/2013wr014981

Boughton, W., Jakob, D., 2008. Adjustment Factors for Restricted Rainfall. Australian Journal of Water Resources, 12(1): 37-48.

Catchlove, R.H., Ball, J.E., 2003. A Hydroinformatic Approach to the Development of Areal Reduction Factors, 28th International Hydrology and Water Resources Symposium. Institution of Engineers, Australia, Wollongong, Australia, pp. 1.9-1.15.

Durrans, S.R., Julian, L.T., Yekta, M., 2002. Estimation of depth-area relationships using radar-rainfall data. Journal of Hydrologic Engineering, 7(5): 356-367. DOI:10.1061/(asce)1084-0699(2002)7:5(356)

Green, J., Xuereb, K., Johnson, F., Moore, G., The, C., 2012. The Revised Intensity-Frequency-Duration (IFD) Design Rainfall Estimates for Australia, The 34th Hydrology and Water Resources Symposium. Engineers Australia, Sydney, pp. 808.

Jakob, D., Taylor, B.F., Xuereb, K.C., 2005. A Pilot Study to Explore Methods for Deriving Design Rainfalls for Australia - Part 1., Bureau of Meteorology.

Jordan, P.W. et al., 2011. Areal Reduction Factors for Estimation of Design Rainfall

Intensities for New South Wales and the Australia Capital Territory, 33rd Hydrology & Water Resources Symposium. Engineers Australia, Brisbane.

King, A.D., Alexander, L.V., Donat, M.G., 2013. The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia. International Journal of Climatology, 33(10): 2376-2387. DOI:10.1002/joc.3588

Kleiber, W., Katz, R.W., Rajagopalan, B., 2012. Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. Water Resources Research, 48(1): W01523. DOI:10.1029/2011wr011105

Kousky, C., Walls, M., 2014. Floodplain conservation as a flood mitigation strategy: Examining costs and benefits. Ecological Economics, 104(0): 119-128. DOI:http://dx.doi.org/10.1016/j.ecolecon.2014.05.001

Leonard, M., Lambert, M.F., Metcalfe, A.V., Cowpertwait, P.S.P., 2008. A space-time Neyman-Scott rainfall model with defined storm extent. Water Resources Research, 44(9): W09402.

Miller, S., Muir-Wood, R., Boissonnade, A., 2008. An exploration of trends in normalized weather-related catastrophe losses. In: Diaz, H.F., Murnane, R.J. (Eds.), Climate Extremes and Society. Cambridge University Press.

Myers, V.A., Zehr, R.M., 1980. A methodology for point to area ratios. 24, National Water Service, Silver Spring, MD.

Niemczynowicz, J., 1982. Areal intensity-duration-frequency curves for short term rainfall events in Lund. Nordic Hydrology, 13(4): 193-204.

Panthou, G., Vischel, T., Lebel, T., Quantin, G., Molinié, G., 2014. Characterizing the space–time structure of rainfall in the Sahel with a view to estimating IDAF curves. Hydrol. Earth Syst. Sci. Discuss., 11(7): 8409-8441. DOI:10.5194/hessd-11-8409-2014

Pilgrim, D. (Ed.), 1987. Australian Rainfall & Runoff - A Guide to Flood Estimation. Institute of Engineers, Australia, Barton, ACT.

Prudhomme, C., Reed, D.W., 1999. Mapping extreme rainfall in a mountainous region using geostatistical techniques: A case study in Scotland. International Journal of Climatology, 19(12): 1337-1356.

Raupach, M. et al., 2012. Australian Water Availability Project, CSIRO Marine and Atmospheric Research, Canberra, Australia.

Siriwardena, L., Weinmann, P., 1996. Derivation of areal reduction factors for design rainfalls in Victoria. Rep. No. 96, 4.

Sivapalan, M., Blöschl, G., 1998. Transformation of point rainfall to areal rainfall: Intensity-duration-frequency curves. J. Hydrol., 204: 150-167.

Strömberg, D., 2007. Natural Disasters, Economic Development, and Humanitarian Aid. Journal of Economic Perspectives, 21(3): 199-222. DOI:doi: 10.1257/jep.21.3.199

Svensson, C., Jones, D.A., 2010. Review of methods for deriving areal reduction factors. Journal of Flood Risk Management, 3(3): 232-245. DOI:10.1111/j.1753-318X.2010.01075.x

van Montfort, M.A.J., 1990. Sliding maxima. Journal of Hydrology, 118(1–4): 77-85. DOI:http://dx.doi.org/10.1016/0022-1694(90)90251-R

Yoo, C., Kim, K., Kim, H.S., Park, M.J., 2007. Estimation of areal reduction factors using a mixed gamma distribution. J. Hydrol., 335(3-4): 271-284. DOI:10.1016/j.jhydrol.2006.11.026

Chapter 3

Comprehensive Evaluation of a Latent Variable Approach to Continuously Simulate Daily Rainfall Fields (Paper 2)

Bree Bennett, Mark Thyer, Michael Leonard, Martin Lambert and Bryson Bates

Journal of Hydrology, submitted January 2016

Statement of Authorship

Title of Paper	Comprehensive evaluation of a latent variable approach to continuously simulate daily rainfall fields.
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Bennett, B., Thyer, M., Leonard, M., Lambert, M., Bates, B.C. 2016. Comprehensive evaluation of a latent variable approach to continuously simulate daily rainfall fields. Journal of Hydrology, (submitted).

Principal Author

Name of Principal Author (Candidate)	Bree Sarah Bennett
Contribution to the Paper	Development and implementation of approach, visualisation and interpretation of results, preparation of manuscript and acted as corresponding author.
Overall percentage (%)	85%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 2/6/16

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Mark Thyer
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.
Signature	Date 2/6/16

Name of Co-Author	Michael Leonard
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.
Signature	Date 2/6/16

Name of Co-Author	Martin Lambert
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.
Signature	Date 2/6/16

Name of Co-Author	Bryson Bates
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.
Signature	Date 2/6/2016

Abstract

The spatial distribution of rainfall has a significant influence on catchment dynamics and the generation of streamflow time series. However, there are few stochastic models that can simulate long sequences of stochastic rainfall fields continuously in time and space. To address this issue, this paper presents a stochastic model that produces daily rainfall fields across the catchment. A latent variable approach is used because it is able to parsimoniously simulate rainfall occurrences as well as amounts. A comprehensive evaluation approach was developed to identify model strengths and weaknesses. This included a performance classification system that provided a systematic, succinct and transparent method to assess and summarize model performance over a range of statistics, sites, scales and seasons. The utility of the model is demonstrated using a case study from the Onkaparinga catchment in South Australia. The model showed many strengths in reproducing the observed rainfall characteristics with the majority of statistics classified as either statistically indistinguishable from the observed or within 5% of the observed across the majority of sites and seasons. These included rainfall occurrences/amounts, wet/dry spell distributions, annual volumes/ extremes and spatial patterns, which are important from a hydrological perspective. One of the few weaknesses of the model was that the total annual rainfall in dry years (lower 5%) was over-estimated by 15% on average over all sites. An advantage of the comprehensive evaluation was that it was able identify the source of this over-estimation was poor representation of the annual variability of rainfall occurrences. Given the strengths of this continuous daily rainfall field model it has a range of potential hydrological applications because it provides the ability to estimate streamflow over an entire catchment.

3.1 Introduction

Robust assessments of the hydrological impacts of floods and droughts, climate and land-use change across catchments requires the use of spatially-distributed hydrological models. As these models rely on spatially-distributed rainfall fields it is essential to have realistic simulations of rainfall fields that can reproduce the practically relevant temporal and spatial characteristics over a broad range of scales. Despite the significance of this need, there are as yet few models for long-term continuous simulation of spatial rainfall fields over a region at daily or sub-daily scales.

Although rainfall models have become increasingly sophisticated over recent decades, the majority of models have been based on a single site (Heneker et al., 2001; Onof and Wheater, 1993; Rodriguez-Iturbe et al., 1988) or the extension of these methods to represent multiple sites in a catchment (Rasmussen, 2013; Srikanthan and Pegram, 2009; Wilks, 1998). Broadly, there are three main approaches for developing rainfall models based on rainfall gauges that have been extended to simulating spatial rainfall fields: (i) a conceptual generating process that combines amounts and occurrences together (Leonard et al., 2008) (ii) a two-step approach that simulates the wet-dry occurrences and then the conditional amounts (Mehrotra et al., 2006) and (iii) a transformed latent (i.e. hidden) variable that maps the wet and dry occurrences to a single distribution so that dry values stem from a lower truncated portion and the amounts stem from the upper portion (Baxevani and Lennartsson 2015). In contrast to the first two approaches, the latter approach allows the process of wet-dry occurrences to be parsimoniously combined with the process of generating rainfall amounts, as well as reproducing realistic patterns of spatial rainfall. The structure of latent variable models is flexible, as demonstrated by their wide range of applications including the analysis of satellite data (Bell, 1987), downscaling (Allcroft and Glasbey, 2003) and continuous simulation (Bardossy and Plate, 1992; Sanso and Guenni, 2000), and thus it has been adopted in this study.

This paper describes a parsimonious model for spatial rainfall fields that is simple to calibrate and an evaluation of its performance and usefulness over a range of space and time scales. The model uses a Gaussian latent-variable approach (Rasmussen, 2013) that simulates rainfall occurrence and amounts using a simple power

transformation, taking full advantage of the parsimonious nature of the transformed latent-variable approach. Kriging is used to produce parameter surfaces since the Gaussian latent variable representation has been noted to be particularly suited to kriging (Cressie, 1993; Kleiber et al., 2012). Additional features of this approach are that it: 1) removes the need for interpolation methods to construct annual totals (it is surprising that sophisticated multisite models are popularly combined with the Thiessen interpolation method (Candela et al., 2012; Kwon et al., 2011) despite the known limitations of this geometric approach); 2) provides stochastic replicates for any location of interest in the catchment; 3) preserves the volumetric properties of rainfall and avoids the need for areal reduction factors (Bennett et al., 2015); and 4) can be used conveniently with distributed models.

While there are a number of studies in the literature that have presented significant advances in the continuous simulation of rainfall fields using a latent variable approach (Baxevani and Lennartsson, 2015; Kleiber et al., 2012), there is, in general, a need for more rigorous assessment of model performance. These previous studies have typically presented results using only selected statistics, sites and months. In this paper, a more comprehensive approach to model evaluation is presented which succinctly summarises model performance over all sites, scales and seasons. This enables clear identification of model strengths and weakness. Furthermore, in previous studies, cross-validation was typically undertaken for only a few select sites. Another benefit of a succinct approach is that it enables evaluation on the basis of full cross-validation across all sites within the domain, rather than simply a select few, as undertaken in previous work.

This paper has two objectives: (1) to present a parsimonious latent variable rainfall model to generate spatial rainfall fields continuously; and (2) to perform a comprehensive evaluation of model performance across a range of time and spatial scales. To assist in assessing performance and rationalising the trade-offs across statistics, scales, sites and seasons, the paper implements a transparent performance classification scheme. This scheme succinctly enables a comparison of performance across a range of model properties, which is beneficial to a systematic assessment of areas where the model is performing well and where model improvements are required. The paper is divided into six sections. Section 3.2 describes the model, calibration procedure and introduces the method of model evaluation. Section 3.3

presents the case study while Section 3.4 demonstrates use of the model for simulation and presents the results of the model evaluation. Discussion and conclusions are summarised in Sections 3.5 to 3.6.

3.2 Methodology

3.2.1 Spatial Latent Variable Model

Latent variable approaches to rainfall modelling have received attention in a range of applications, such as downscaling, continuous simulation and modelling extremes (Allcroft and Glasbey, 2003; Bardossy and Plate, 1992; Baxevani and Lennartsson, 2015; Davison et al., 2012; Durban and Glasbey, 2001; Kleiber et al., 2012; Qin, 2010; Rasmussen, 2013; Sanso and Guenni, 2000). An implementation is presented here that is parsimonious and represents rainfall continuously in space. Let r_{it} be the rainfall at site $i = 1, \dots, N$ and at time $t = 1, \dots, T$, which is related to a normally distributed latent variable l_{it} via left-truncation at zero and via a power transformation,

$$r_{it} = \begin{cases} l_{it}^{\beta} & l_{it} > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where β is the power transformation parameter. The vector of latent variables at all sites is $\boldsymbol{l}_t = \{l_{1t}, \dots, l_{Nt}\}$ and is related to the subsequent timestep $t+1$ according to the multivariate normal distribution

$$\begin{bmatrix} \boldsymbol{l}_{t+1} \\ \boldsymbol{l}_t \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \boldsymbol{\mu}_{t+1} \\ \boldsymbol{\mu}_t \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{t+1,t+1} & \boldsymbol{\Sigma}_{t+1,t} \\ \boldsymbol{\Sigma}_{t,t+1} & \boldsymbol{\Sigma}_{t,t} \end{bmatrix} \right), \tag{2}$$

where $\boldsymbol{\mu}_t$ and $\boldsymbol{\mu}_{t+1}$are the means at time steps t and $t+1$ respectively, $\boldsymbol{\Sigma}_{t+1,t+1}$ and $\boldsymbol{\Sigma}_{t,t}$ are the lag-0 covariance matrices at respective timesteps, and $\boldsymbol{\Sigma}_{t+1,t}$ and $\boldsymbol{\Sigma}_{t,t+1}$ are the lag-1 cross covariance matrices. The latent values at each subsequent time step are obtained as the multivariate regression:

$$\boldsymbol{l}_{t+1} | \boldsymbol{l}_t \sim \text{MVN}\left(\boldsymbol{\mu}_{t+1} + \boldsymbol{\Sigma}_{t+1,t}\boldsymbol{\Sigma}_{t,t}^{-1}(\boldsymbol{l}_t - \boldsymbol{\mu}_t),\ \boldsymbol{\Sigma}_{t+1,t+1} - \boldsymbol{\Sigma}_{t+1,t}\boldsymbol{\Sigma}_{t,t}^{-1}\boldsymbol{\Sigma}_{t,t+1}\right) \tag{3}$$

This specification is akin to a lag-1 multivariate autoregressive model (Rasmussen, 2013) and a number of further simplifying assumptions have been made to facilitate ease of calibration and extension of the model to be continuous in space. These

simplifcations are:

1. The mean and covariance are assumed here to be constant within each month such that $\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_{t+1,t+1} = \boldsymbol{\Sigma}_{t,t}$.
2. The use of a separable cross-covariance $\boldsymbol{\Sigma}_{t+1,t} = \bar{\varphi}_t \boldsymbol{\Sigma}_{t,t}$ where the temporal component, denoted by a scalar autoregressive parameter $\bar{\varphi}_t$, is separate from the spatial component $\boldsymbol{\Sigma}_{t,t}$ (Genton, 2007).

The reasoning for a separable covariance function warrants further discussion. When a vector of autoregressive parameters is used, this is referred to as a contemporaneous model (Rasmussen, 2013) and it represents a significant reduction in the number of estimated parameters (N^2 lag-0 and N lag-1 parameters) compared to explicitly modelling all cross-covariances (N^2 lag-0 and N^2 lag-1 parameters).

Here, a single temporal parameter $\bar{\varphi}_t$ (along with parameters of a spatial covariance function) is used that is common to all sites, which represents a further reduction in the number of parameters. The main reason for this simplification is that multisite models do not have the requirement of being continuous in space and therefore have more flexibility in the specification of their covariance, leading to a better fit of calibration and validation statistics (e.g. those associated with the temporal autocorrelation such as the number of consecutive wet days). A further reason is that the scalar constant ensures that the covariance matrix $\boldsymbol{\Sigma}_{t+1,t}$ maintains a block Toeplitz structure, which is important for utilising efficient simulation methods (Dietrich and Newsam, 1993).

Explicitly modelling a covariance structure has the benefit that asymmetry in the covariance matrix (due to advection) can be modelled (Rasmussen, 2013), but for smaller catchments at the daily timescale, this asymmetry should not be evident. Asymmetry is likely to be more prominent an issue for subdaily timescales and larger regions, and in those cases a Lagrangian framework can be used to avoid the issue (Seed et al., 1999). Overall, the impact of using a spatially constant covariance structure will depend on how homogeneous a study region is in terms of observed daily auto-correlations.

Table 3-1 shows the reduction in the required number of estimated parameters following the use of a separable covariance function in comparison to the multisite approach of explicitly modelling all cross-covariances for different numbers of sites. Table 3-1 shows that the reduction in parameters, for increases from 30% for two sites to 94% for 25 sites which represents a significant gain in parsimony. The importance of model parsimony is further discussed in Section 3.5.1.

Table 3-1 Comparison of the number of parameters required to simulate at ***N*** sites per season modelled

No. Sites modelled	Multisite model	Spatial latent variable model	Reduction in parameters
2	14	10	30%
5	65	19	71%
10	230	34	85%
25	1325	79	94%

The lag-0 covariance matrix $\mathbf{\Sigma}_{t,t}$ has elements $\Sigma_{ij} = \sigma_{it}\sigma_{jt}\rho\left(d_{ij}\middle|\nu_t, \alpha_t, \lambda_t\right)$ for all pairs of sites i and $j = 1, \dots, N$ where $\sigma_i$and σ_j are the standard deviations at each site, $d_{ij}$is the distance between the sites, and ν_t, α_t and λ_t are the parameters of an isotropic powered-exponential correlation function defined by

$$\rho\left(d_{ij}\middle|\nu_t, \alpha_t, \lambda_t\right) = \begin{cases} 1 & d_{ij} = 0 \\ (1 - \nu_t)\exp\left(-\left(\frac{\mathrm{d_{ij}}}{\alpha_t}\right)^{\lambda_t}\right) & d_{ij} > 0 \end{cases} \quad (4)$$

where for time step t, α_t is the range parameter, λ_t is the power term and ν_t is the nugget. The full parameter specification for the multisite model is therefore $\boldsymbol{\mu}_t = \{\mu_{1t}, \dots, \mu_{Nt}\}, \boldsymbol{\sigma}_t = \{\sigma_{1t}, \dots, \sigma_{Nt}\}$, $\boldsymbol{\beta}_t = \{\beta_{1t}, \dots, \beta_{Nt}\}$, $\bar{\varphi}_t$, ν_t, α_t and λ_t, where the parameter values remain constant for all time steps t in a common month. Recent methods have alternatively used parameter sets that vary for each day of the year by defining a cyclic relationship between the parameter surface and day of the year (Baxevani and Lennartsson, 2015; Kleiber et al., 2012). In contrast this paper assessed whether defining fixed parameter sets for each month was able to reproduce aggregate totals and intra-annual variability.

To simulate the model continuously across a region $\boldsymbol{\mu}_t$, $\boldsymbol{\sigma}_t$ and $\boldsymbol{\beta}_t$ need to be specified for all locations over the entire field, thus additional parameters are required to interpolate these parameter surfaces from the observed sites. Table 3-1 showed the model is relatively parsimonious compared to multisite rainfall models and this can be further simplified by making parameters common across months or sites as relevant to the study region. Further potential enhancements are discussed in Section 3.5.3.

3.2.2 Calibration

Calibration of the model proceeds in a step-wise manner. The first step is the estimation of the marginal distribution parameters at each site. The second step is to estimate the at-site lag-1 temporal correlation. The third step is to calibrate the spatial correlation function, and the fourth step is the regionalisation of the parameters to simulate the spatial rainfall field.

In the first step for estimating the marginal distribution parameters, both the method of moments and the maximum likelihood method are valid, but the method of moments has been reported as giving better quality fit to the upper tail of rainfall amounts (Rasmussen, 2013) and thus is used in this study. Consider that the observed time-series of daily rainfall at a site has been partitioned according to a number, n_d, of 'dry' zero values and the truncated set of n_w 'wet' values, i.e. $\boldsymbol{r}_w = \{r_t|, t = 1, \dots, n_w\}$. The proportion of dry values is determined as $\hat{p}_d = n_d/(n_d + n_w)$. Let $\boldsymbol{l}_w$ denote the latent values corresponding to $\boldsymbol{r}_w$ after transformation. The observed first and second order non-central moments of the truncated latent distribution are determined as

$$\mathrm{E}[\boldsymbol{l}_w] = \frac{1}{n_w}\sum_{t=1}^{n_w} r_t^{1/\widehat{\beta}} \tag{5}$$

$$\mathrm{E}[\boldsymbol{l}_w^2] = \frac{1}{n_w}\sum_{t=1}^{n_w} \left(r_t^{1/\widehat{\beta}}\right)^2. \tag{6}$$

Consider the left-truncated normal distribution with known truncation point (Johnson et al., 1994). The parameter $\hat{\beta}$ can be estimated by solving the following two equations

$$\frac{\hat{\sigma}}{\mathrm{E}[\boldsymbol{l}_w]} = \left(\frac{\phi(\hat{\delta})}{1-\Phi(\hat{\delta})} - \hat{\delta}\right)^{-1} \tag{7}$$

$$\left(\frac{\hat{\sigma}}{\mathrm{E}[\boldsymbol{l}_w]}\right)\left(\frac{\hat{\sigma}}{\mathrm{E}[\boldsymbol{l}_w]} - \hat{\delta}\right) = \frac{\mathrm{E}[\boldsymbol{l}_w^2]}{\mathrm{E}[\boldsymbol{l}_w]^2} \tag{8}$$

where $\mathrm{E}[\boldsymbol{l}_w]$ and $\mathrm{E}[\boldsymbol{l}_w^2]$ are defined in Eq. (5) and Eq. (6), $\phi(\,)$ is the standard normal density, $\Phi(\,)$ is the cumulative normal density, and $\hat{\delta} = \hat{\mu}/\hat{\sigma}$, which is the truncation point as a standardised deviate. To obtain the deviate of the truncation point, the procedure first equates $\hat{\delta} = \Phi^{-1}(\hat{p}_d)$, then $\frac{\hat{\sigma}}{E[L_w]}$ is determined using Eq. (7). Following this, the left hand side of Eq. (8) is reduced to a constant, whilst the right hand side is dependent on the $\hat{\beta}$ parameter through Eq. (5) and Eq. (6). Subsequently, Eq. (8) can be used to estimate $\hat{\beta}$ to give the best fit using root finding techniques (Rasmussen, pers. comm., Jan. 2014).

Having estimated $\hat{\beta}$, the parameters $\hat{\mu}$ and $\hat{\sigma}$ can be estimated by minimising the objective function

$$min\ ((\mathrm{E}[\boldsymbol{r}_w] - \hat{m}_w)^2 + (\mathrm{VAR}[\boldsymbol{r}_w] - \hat{s}_w^2)^2) \tag{9}$$

where $\hat{m}_w$ and $\hat{s}_w^2$ are the mean and variance of the truncated wet values

$$\hat{m}_w = \frac{1}{n_w}\sum_{t=1}^{n_w} r_t \tag{10}$$

$$\hat{s}_w^2 = \frac{1}{n_w-1}\sum_{t=1}^{n_w}(r_t - \hat{m}_w)^2 \tag{11}$$

and the corresponding moments, $\mathrm{E}[\boldsymbol{r}_w]$ and $\mathrm{VAR}[\boldsymbol{r}_w]$ in terms of the marginal parameters $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\beta}$ are obtained by integration over the wet values,

$$\mathrm{E}[\boldsymbol{r}_w] = (1 - \hat{p}_d)^{-1}\int_0^\infty x f_X(x)dx \tag{12}$$

$$\mathrm{VAR}[\boldsymbol{r}_w] = (1 - \hat{p}_d)^{-1}\int_0^\infty (x - E[\boldsymbol{r}_w])^2 f_X(x)dx. \tag{13}$$

where the $(1 - \hat{p}_d)$ renormalises the density due to the truncation at zero and $f_X(x)$ is given by

$$f_X(x) = \left(2\pi\hat{\sigma}^2\hat{\beta}^2\right)^{-1/2} x^{(-1+1/\hat{\beta})} exp\left(-\hat{\sigma}^{-2}/2\left(x^{1/\hat{\beta}} - \hat{\mu}\right)^2\right), x > 0 \tag{14}$$

In the second step, having estimated the at-site marginal distribution parameters for each site and month, the lag-1 autocorrelation (l_{t+1}, l_t) can be estimated for all pairs of points above the zero threshold. This estimate corresponds to an estimate of correlation in the left-truncated bivariate normal distribution and can be related analytically (Weiler, 1959) to the underlying autocorrelation parameter $\hat{\varphi}_i$ of the non-truncated bivariate distribution at site i. The analytic relationship can be numerically solved for $\hat{\varphi}_i$ since the other marginal parameters $\hat{\mu}$ and $\hat{\sigma}$ have been determined, resulting in $\hat{\varphi}_i$ estimates for all months at all sites.

In the third step, due to the separable covariance function, only the lag-0 cross-covariances $\hat{\Sigma}_{ij}$ remain to be estimated and from them the parameters of the spatial correlation function. As with the autocorrelation, lag-0 cross-covariances can be estimated from the non-zero latent values corresponding to correlation in a left-truncated bivariate normal distribution. The $\hat{\Sigma}_{ij}$ are found by solving the Weiler (1959) relationship with known $\hat{\mu}$ and $\hat{\sigma}$. A sum of squared errors approach is then used to fit this spatial correlation function, Eq. (4), to these estimates and obtain the parameters $\hat{v}, \hat{\alpha}$ and $\hat{\lambda}$.

In the fourth step, the spatial field of marginal distribution parameters ($\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\beta}$) are estimated by interpolating the at-site parameter estimates ($\hat{\mu}$, $\hat{\sigma}$ and $\hat{\beta}$), obtained from step 1, with dependent variables of distance between grid points and elevation. Each parameter is kriged independently. The use of elevation as a dependent variable helps the method to reproduce the spatial rainfall gradient of the region. It is possible that independently kriging rather than jointly kriging the parameter surfaces could lead to spurious parameter combinations that affect the marginal distribution of rainfall. This is tested by comparing the results from full calibration versus via cross-validation. For the lag-one correlation, a single representative $\hat{\bar{\varphi}}$ for each month is specified as the geometrical average (e.g. Thiessen weighted value) of the $\hat{\varphi}_i$ at-site estimates. The model performance evaluation of the temporal statistics (wet/dry spell duration) will test whether the assumption of a spatially constant, monthly lag-one temporal correlation is valid.

3.2.3 Model Performance Evaluation

The model evaluation is split over the model's performance at the observed sites ('at-site' performance), the spatial performance and cross-validation of the model.

At-site performance

The evaluation of at-site performance focuses primarily on temporal statistics. First, the model's ability to preserve the mean and standard deviations of wet day amounts as well as the probability of a wet day is considered. These statistics are used in calibration and are therefore a foundational assessment. Secondly, the wet spell and dry spell length distributions for all seasons and sites are presented. A 'spell' is a block of consecutive time steps having the same state, either 'wet' or 'dry', and relates to the intermittence of the daily amounts, based on the rainfall auto-correlation.

Whereas the daily timescale is characterised by wet-dry sequences, distributions of monthly and annual rainfall are important for preserving seasonal characteristics and inter-annual variability. For this reason, distributions of monthly total rainfall, annual total rainfall and the number of wet days annually are evaluated. The ability of the model to reproduce these aggregate totals is presented as quantile-quantile plots of representative statistics of the examined distributions (mean, standard deviation, lower tail indicator - 5^{th} percentile, upper tail indicator - 95^{th} percentile). Correlations between annual totals are also evaluated at each site to assess whether the model reproduces inter-annual variability. The correlations between consecutive monthly total rainfall amounts (i.e. January-to-February) are evaluated to further understand the structure of variability in the annual rainfall.

The simulation of extremes is a demanding test of model performance. Whereas the calibration focuses on means and variances of amounts, the extremes are an emergent property of the model. Underestimation of extreme daily rainfall values is a common problem for stochastic rainfall simulators. Many point rainfall models and spatial rainfall models struggle to simulate extremes due to issues with cascade generators and resampling approaches, limitations in adopted amount generation distribution, and a lack of correlation between weather states and extreme precipitation amounts (see Hundecha et al., 2009; Li et al., 2012 and references therein).

Spatial field performance

There are three parts to the evaluation of the spatial rainfall fields. First, the distributions of the number of jointly wet sites for each month are compared. While this is an assessment of the spatial correlation properties, it is strictly a multisite assessment rather than an assessment of the interpolated spatial features of the model. Second, the catchment domain aggregated behaviour of the observation sites is evaluated (See also Baxevani and Lennartsson, 2015; Kleiber et al., 2012). The domain averaged rainfall is the catchment average rainfall time-series estimated by Thiessen weighting of the rainfall at each site on each day. The domain averaged series for the observed rainfall and the simulated rainfall are compared using the aforementioned metrics to assess at-site rainfall statistics (daily mean, standard deviation, proportion of wet days, wet spell/dry spell distributions, monthly totals, annual totals, monthly and annual correlations, intensity-frequency-duration relationships). Thirdly, the spatial rainfall fields produced by the model are evaluated by comparing the field of average annual total rainfalls produced by the model against the observed average annual total rainfalls that have been kriged (again using elevation as a covariate) over the region to infill space and reproduce the rainfall gradient. It is challenging to truly assess the spatial features of the rainfall model since the rainfall is observed at points, thus any spatial comparison to observations must also rely on interpolation of the observations. While comparison to radar data is possible, this can be problematic, since radar records are short and subject to measurement errors that require correction against the same underlying rainfall gauges.

Cross-validation performance

A cross-validation of both the parameters predicted by the kriging and at-site model performance is conducted to assess the error associated with spatial interpolation. Using the calibrated at-site parameters (μ, σ, β) the region is kriged using the distance between grid points and elevation as covariates, each time omitting one site and comparing the estimated surfaces at the omitted location (μ^*, σ^*, β^*) with the calibrated parameter value. The cross-validation then evaluates all at-site statistics previously discussed using the interpolated parameters at the omitted site, repeated for all sites.

Performance classification

In addition to visual comparison of the observed versus simulated statistics, the models performance is assessed using a classification system. The use of a classification system provides a systematic, transparent and succinct approach to compare the relative performance of the model over a large range of statistics, sites, spatial scales and time scales (daily, monthly and annual)

For the purposes of this paper, a two stage classification system was developed. In the first stage, a three level categorisation system was developed, where model performance of a single statistic for given site/month was classified into one of three categories; 'good', 'fair' and 'poor' performance. Table 3-2 summarises the quantitative tests for each performance class and provides an example of each category. 'Good' performance indicates that less than 10% of the observations lie outside the simulation's 90 % probability limits and therefore the simulated rainfall is statistically indistinguishable from the observed for that evaluated statistic (Fig. 3-1, case (1)). 'Fair' performance indicates that the statistic derived from the observed rainfall sits within three standard deviations of the simulated mean - assuming the uncertainty in the statistics is normally distributed, this represents the 99.7% limits (Fig. 3-1, case (2)), or the absolute relative difference between the observation and the simulated mean is less than 5% (Fig. 3-1, case (3)). The absolute relative difference is calculated as

$$RD = |100\,(x_{obs} - E[x_{sim}])/x_{obs}| \tag{15}$$

where RD is the absolute relative distance, x_{obs} is the evaluated statistic's observed value, $E[x_{sim}]$ is the expected value of the statistic, $x_{sim,i}$ for over all realisations i. Otherwise, performance is classified as 'poor' (Fig. 3-1, case (4)).

Table 3-2 Performance classification criteria

Performance Classification	Key	Test
'good'		Less than 10% of observations outside 90% limits (case 1)
		More than 10% of observations are outside 90% limits but within the 99.7% limits (case 2)
'fair'		OR
		Absolute relative difference between the observation and simulated mean is 5% or less (case 3)
'poor'		Otherwise (case 4)

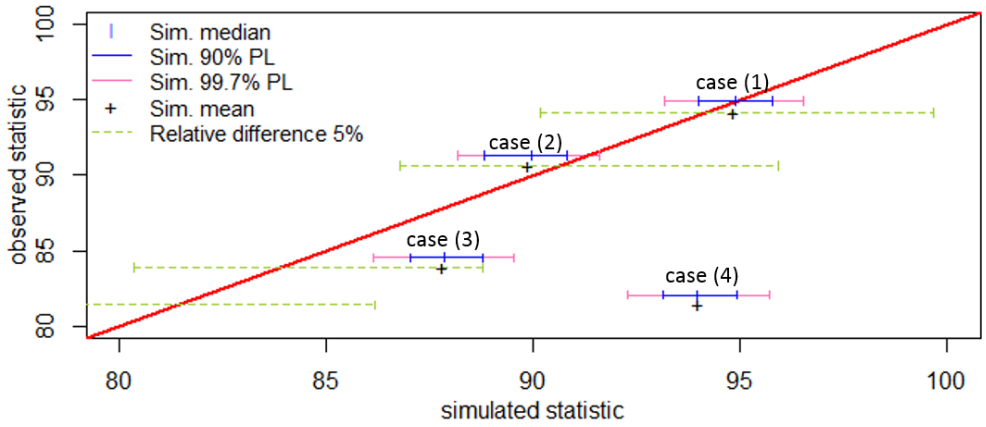
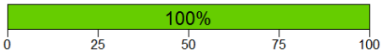
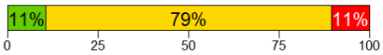
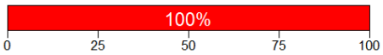
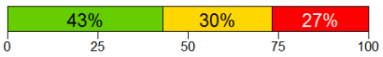
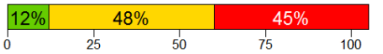
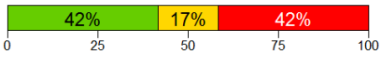


Fig. 3-1 Illustration of performance classification, case (1) shows 'good' performance, cases (2) and (3) show 'fair' performance and case (4) shows 'poor' performance.

The second stage of the performance classification summaries the 'cumulative' model performance for multiple sites or months for a given statistic. Table 3-3 provides these 'cumulative' performance classifications, which range from 'Overall Good' to 'Overall Poor', and the quantitative tests used to determine them, which are based on the percentage of cases (sites/months) which are classified as 'good', 'fair' or 'poor' in the first stage. For example, 'Overall Highly Variable' occurs when the percentage of cases classified as 'good' and 'poor' are greater than the percentage of cases deemed 'fair'. This second stage enables an assessment of the model's ability to capture a particular statistic over multiple sites and months.

Table 3-3 Cumulative performance classification criteria

Cumulative Performance Classification	Test	Example
'Overall Good'	'good' > 50 %	
'Overall Fair'	'fair' > 50 %	
'Overall Poor'	'poor'> 50%	
'Overall Fair – Good'	'fair' & 'good' > 'poor'	
'Overall Fair – Poor'	'fair' & 'poor'> 'good'	
'Overall Highly Variable'	'good' & 'poor' > 'fair'	

The advantage of classifying the model performance for each statistic for each site and month is that it provides a systematic and transparent method to describe and discuss the multitude of results. By using a quantitative approach to classify model performance it reduces the often used, ad-hoc nature of descriptive assessments, (words such as 'adequate', 'suitable', or 'reasonable'). However, the choice of categories still involves an element of subjectivity (e.g. Evin et al., 2014). For example, the use of only three categories, 'good', 'fair' and 'poor' or the choice of 3 standard deviations to delineate 'fair' from 'poor' is subjective and whether the model performance is good or poor, ultimately depends on the practical application of the model. Nonetheless, this classification system has clear advantages over an ad-hoc descriptive assessment, and is particularly convenient where there a large number of statistics.

3.3 Case Study

The Onkaparinga catchment in South Australia is used as a case study (Fig. 3-2). The catchment contains the Mt Bold Reservoir, which is the largest reservoir supplying metropolitan Adelaide, and is supplemented by water from the Murray River via a pipeline to the east. Modelling rainfall over the catchment is important for understanding the natural flow regime, which informs understanding of dependence on the Murray River for water security.

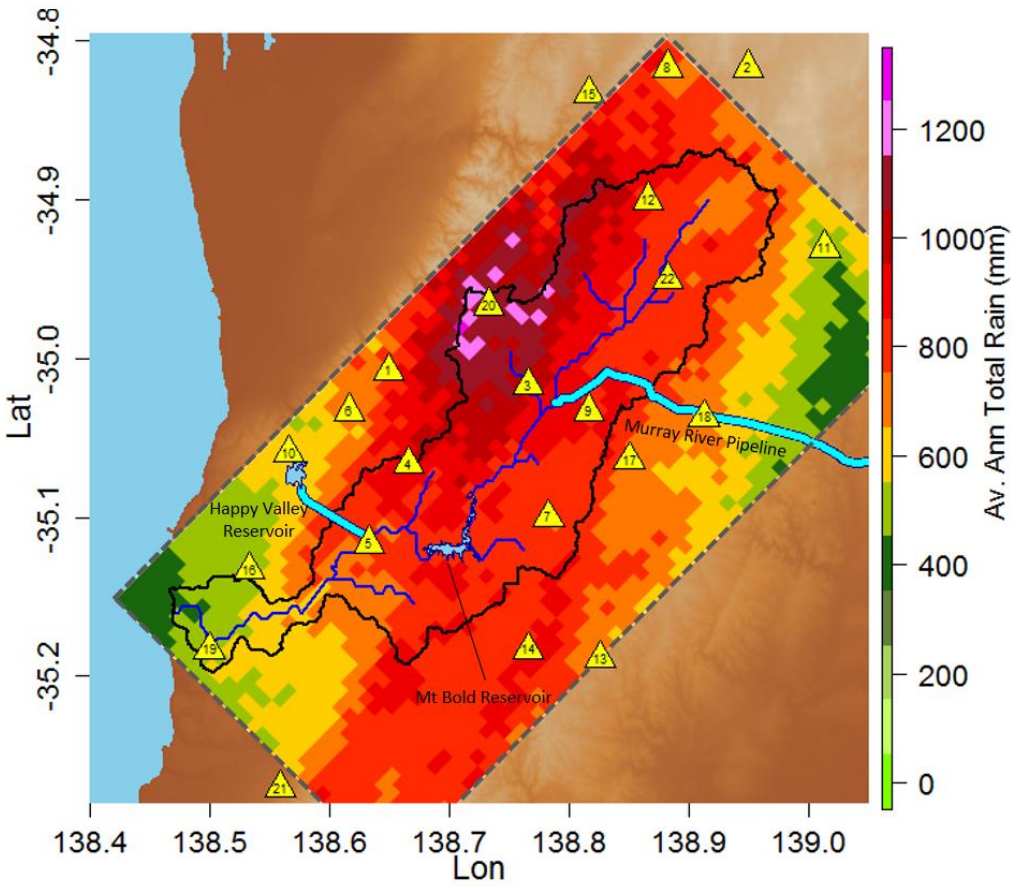


Fig. 3-2 Locations of rainfall observation sites, Onkaparinga catchment and study region.

There are 22 daily rainfall gauges within and surrounding the Onkaparinga catchment (Fig. 3-2 and Table 3-4) obtained from the SILO database (Jeffrey et al., 2001). Their records span the period from 1900 to present, but to minimise any potential impact of missing values in the records, the period 1914 to 1986 was selected, since this period had minimal missing data. The data were quality checked for trends and inhomogeneities (Westra et al., 2014) by comparing against the Happy Valley site (23721) which is part of the high quality network of gauges. All sites were evaluated as suitable. From Fig. 3-2, it is clear there is a strong rainfall gradient with average annual rainfall ranging from 522 mm at the mouth of the Onkaparinga (Site No. 19) at elevation 7 m up to 1088 mm at Uraidala (Site No. 20) at an elevation of 499 m. The catchment rainfall is highly seasonal with the majority of rainfall occurring in the seasons of winter (June, July and August) and spring (September, October and November) and with negligible rainfall occurring throughout summer (December, January and February).

Table 3-4 Site names and locations

Site No.	Site Name	Elev (m)	Ann. Ave. Rain (mm)	Site No.	Site Name	Elev (m)	Ann. Ave. Rain (mm)
1	Belair	386	786	12	Lobethal	470	882
2	Birdwood	385	723	13	Macclesfield	302	730
4	Bridgewater	376	1046	14	Meadows	384	869
4	Cherry gardens	345	924	15	Cudlee Creek	311	831
5	Clarendon	223	818	16	Morphett Vale	90	562
6	Coromandel Valley	234	714	17	Mount Barker	349	766
7	Echunga	375	805	18	Nairne	403	678
8	Gumeracha	346	793	19	Old Noarlunga	7	522
9	Hahndorf	347	845	20	Uraidla	499	1088
10	Happy Valley	148	638	21	Willunga	158	642
11	Harrogate	335	552	22	Woodside	387	801

Nineteen rainfall gauges lie inside the boundary of the Onkaparinga catchment and are used to evaluate at-site model performance. The three gauges that lie outside the catchment are used for the spatial interpolation of parameters to reduce edge effects, but are not used in model evaluation. The simulation experiment consisted of 100 replicates using 0.88 km square grids over the case study region.

3.4 Results

The results present a wide range of statistics as described in Section 3.2.3, and uses the performance classification system from Section 3.2.3. Section 3.4.1 assess at-site performance of the model, Sections 3.4.2 assess spatial field performance of the model and Section 3.4.3 presents the performance in cross-validation. Due to the multitude of results, only selected key statistics are presented in the main paper, with further detailed results for each site and statistic can be located in Supplementary Material A—D (see Appendix B).

3.4.1 At-site performance

Daily rainfall occurrence and amounts

The model shows ‘Overall Good’ performance in reproducing the following statistics for each month; the mean wet day rainfall amounts, standard deviation of wet day rainfall and the mean number of wet days each month (and therefore the probability of a wet day) (Fig. 3-3a to c). This demonstrates that the model structure is suitable for reproducing the observed daily marginal rainfall statistics. However, Fig. 3-3d

indicates that the model under-predicts variability in the number of wet days for some months (February, May, June, August, October and November) and over-predicts it for one month (January).

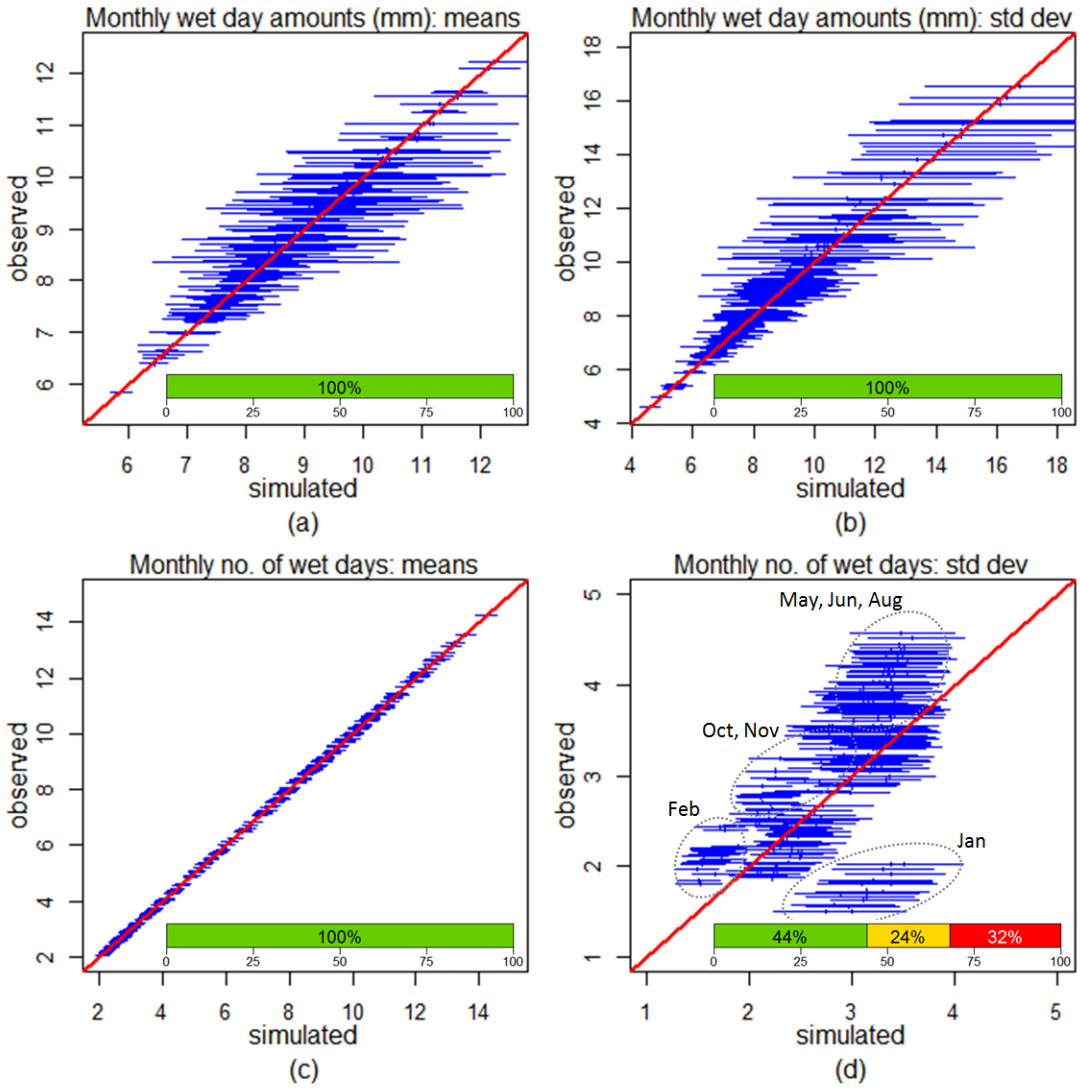


Fig. 3-3 At site daily statistics for all sites and months, 90% probability limits shown, cumulative performance shown as a percentage of all sites and months.

Fig. 3-4 shows an example of the model performance in simulating wet spell and dry spell length distributions for the month of March at the Lobethal site. The model shows 'Overall Good' performance in simulating the wet spell length distribution for the autumn and winter months and 'Overall Fair – Good' performance for the spring and summer months when examined over all sites (Fig. 3-4a). The majority of instances in which performance was classified as 'poor' occurred in February when the catchment has very little rainfall. The model shows 'Overall Good' performance

in simulating the distribution of dry spell lengths in all months and sites (Fig. 3-4b). This demonstrates that the use of a constant regional average autoregressive parameter for each month yields 'Overall Fair– Good' performance in producing realistic temporal patterns and rainfall persistence.

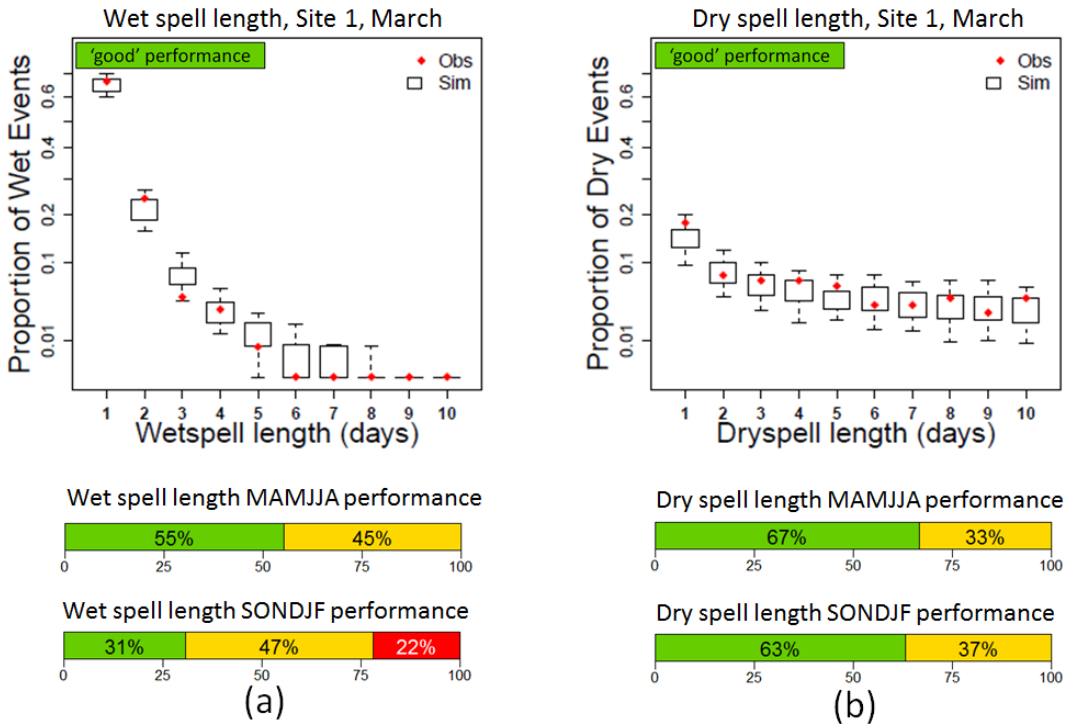


Fig. 3-4 Distribution of event lengths (a) wet spell length distribution and (b) dry spell length distribution, 90% probability limits shown.

Monthly and annual statistics

The model exhibits 'Overall Good' performance in simulating total monthly rainfall amounts (Fig. 3-5). Fig. 3-5a shows that the performance in simulating the mean monthly total rainfall is 'Overall Good' for all months and sites. The simulated standard deviation of monthly rainfall totals is 'Overall Good' for the majority of months and sites (Fig. 3-5b) except for some sites with higher monthly standard deviation in May, June and August. The simulation of both the lower and upper tails of the monthly total rainfall distribution is also 'Overall Good' for the majority of sites (Fig. 3-5c to d).

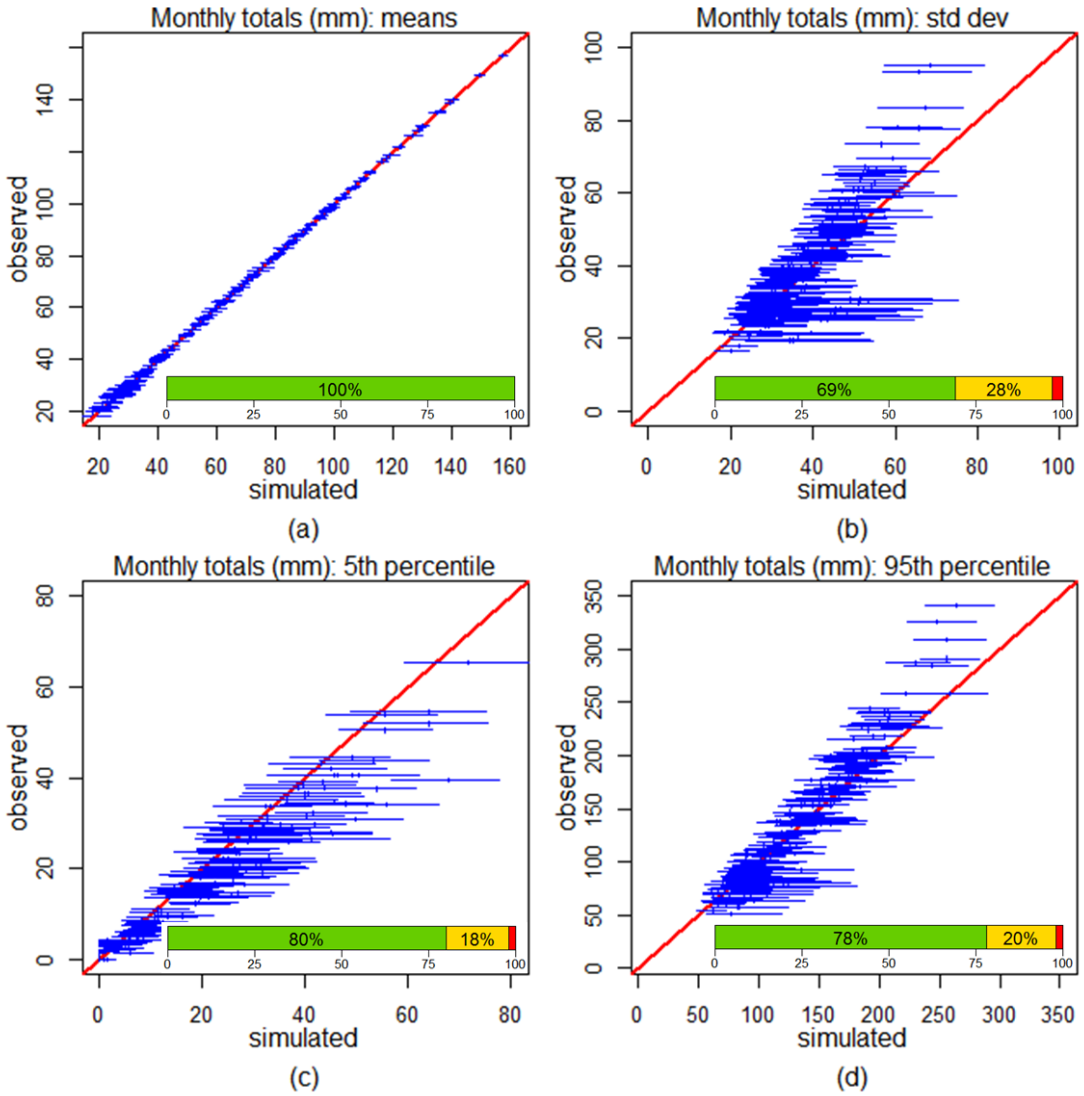


Fig. 3-5 At site monthly totals for all sites and months (a) means, (b) standard deviations, (c) lower 5th percentile and (d) upper 95th percentile, 90% probability limits shown, cumulative performance as a percentage of all sites and months.

Fig. 3-6a shows that the model exhibits 'Overall Good' performance in simulating the mean total annual rainfall for all sites. However, the model underestimates the variability of the total annual rainfall, exhibiting 'Overall Fair – Poor' performance in the simulation of the total annual rainfall standard deviation (Fig. 3-6b). This results in the model not reproducing the drier rainfall years. This is seen in the 'Overall Fair – Poor' performance in simulating the lower tail (lower 5%, 5th percentile) of the total annual rainfall, with the simulated rainfall being larger than the observed by on average 15% (Fig. 3-6c). The performance of the simulation of the upper tail (upper 5%, 95th percentile) of the total annual rainfall distribution is 'Overall Good'. Thus the underestimation of annual variability is less of a problem,

in this case, for the simulation of wetter years.

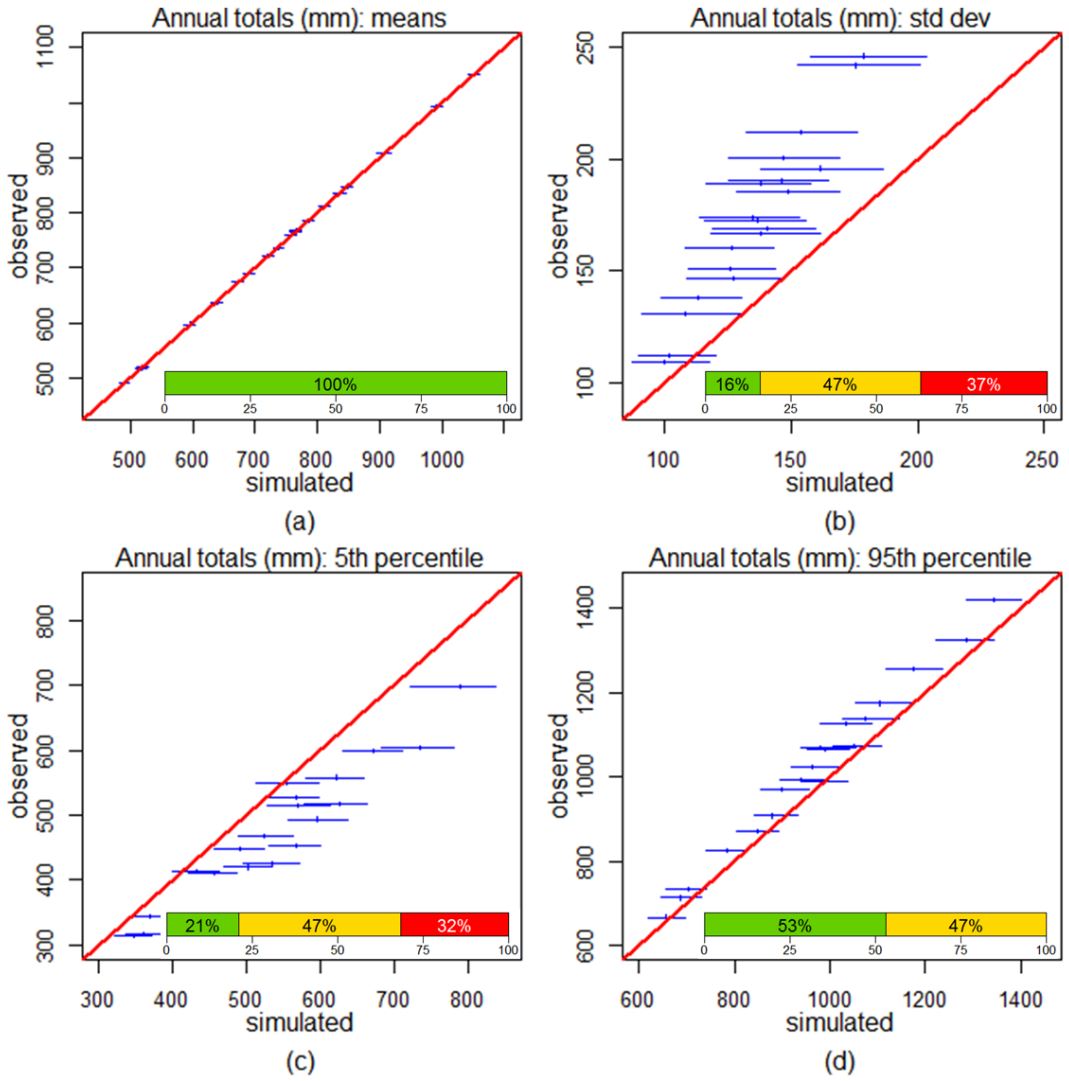


Fig. 3-6 At site annual totals for all sites and months (a) means, (b) standard deviations, (c) 5th percentile and (d) 95th percentile, 90% probability limits shown, cumulative performance as a percentage of sites.

Fig. 3-7 shows that the simulation demonstrates 'Overall Good' performance annually in simulating the mean wet-day amount, the standard deviation of wet-day amounts and the mean annual number of wet days at all sites (Fig. 3-7a to c). However, the model's performance is 'Overall Poor' for all sites in simulating variability in the number of wet days annually as the annual variance is underestimated (Fig. 3-7d). By evaluating the annual occurrences and amounts it is shown that the simulation of higher annual rainfall totals in the lower tail of the annual total rainfall distribution (Fig. 3-6c) is due to a lack of variability in the

simulation of the number of wet days annually (Fig. 3-7d). As the simulation of rainfall amounts and variability of those amount is 'Overall Good' for all sites (Fig. 3-7a and b) the 'Overall Fair – Poor' performance in simulating the annual total rainfall in drier years does not result from any issue in rainfall amount generation. This deficiency is due to the lack of variability in the simulation of occurrences. The practical implication of this deficiency is that the model is poorer at simulating the overall drier rainfall years as discussed with respect to Fig. 3-6c.

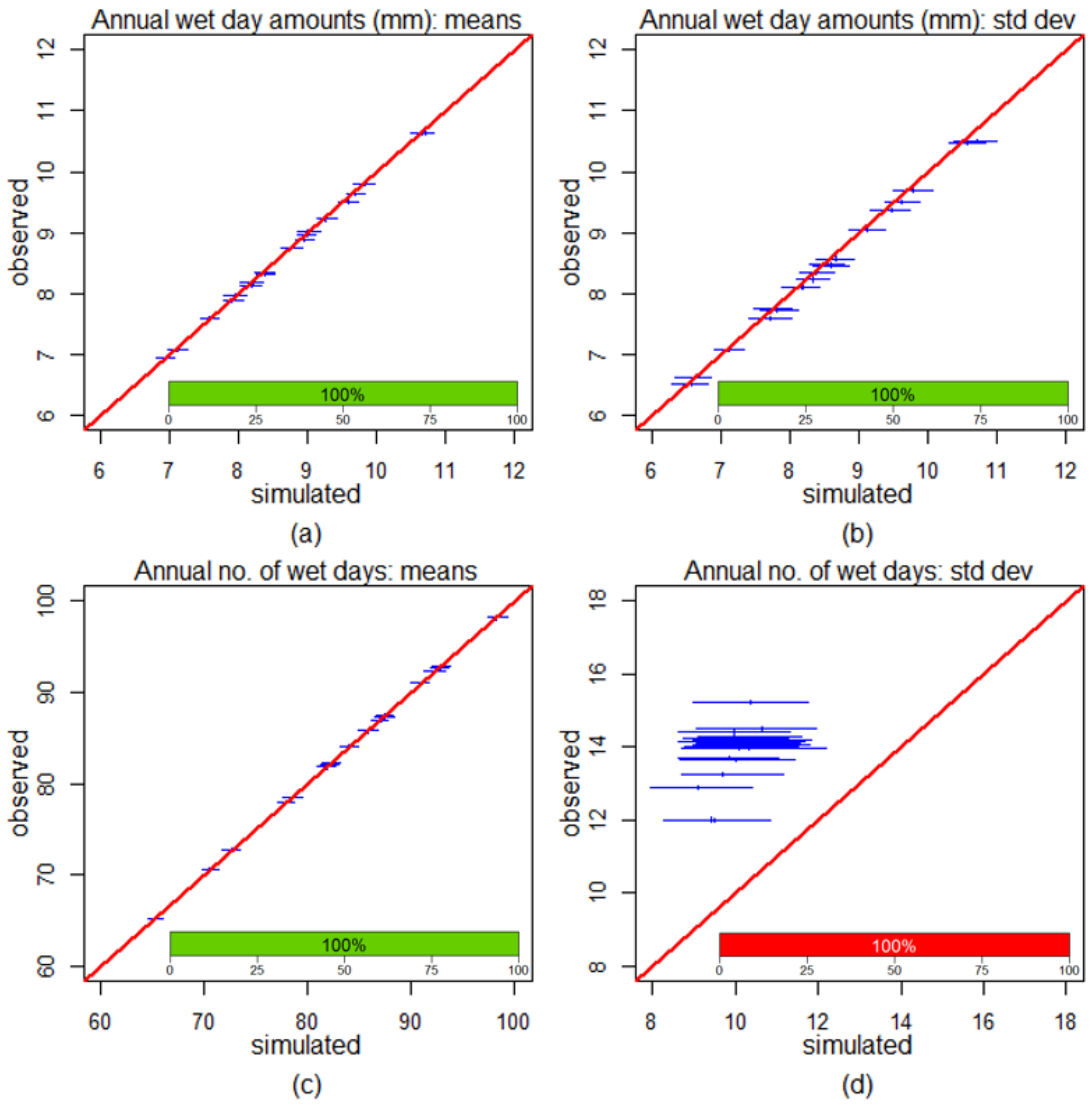


Fig. 3-7 At site annual mean wet day amounts (a) means and (b) standard deviations; number of wet days for all sites (c) means, (d) standard deviations, 90% probability limits shown, cumulative performance as a percentage of sites.

Temporal correlation of annual and monthly totals

Fig. 3-8 shows the correlations between consecutive monthly (Fig. 3-8a) and annual (Fig. 3-8b) rainfall totals for the simulated and observed rainfall. The model does not include month-to-month correlation, thus the simulations are centred on zero. As there is low monthly persistence at this location, the correlations between monthly consecutive rainfall totals show 'Overall Good' performance for all sites and months (Fig. 3-8a). However, correlations in the consecutive totals for June and July show a number of sites that are deemed 'fair' (Fig. 3-8a). This may be of concern as these months are part of the wet (winter) season for the catchment in which a large proportion of the annual rainfall occurs. At the annual scale, the observed correlation between consecutive annual rainfall totals has 'Overall Good' performance (Fig. 3-8b) as there is little inter-annual persistence for rainfall in this catchment.

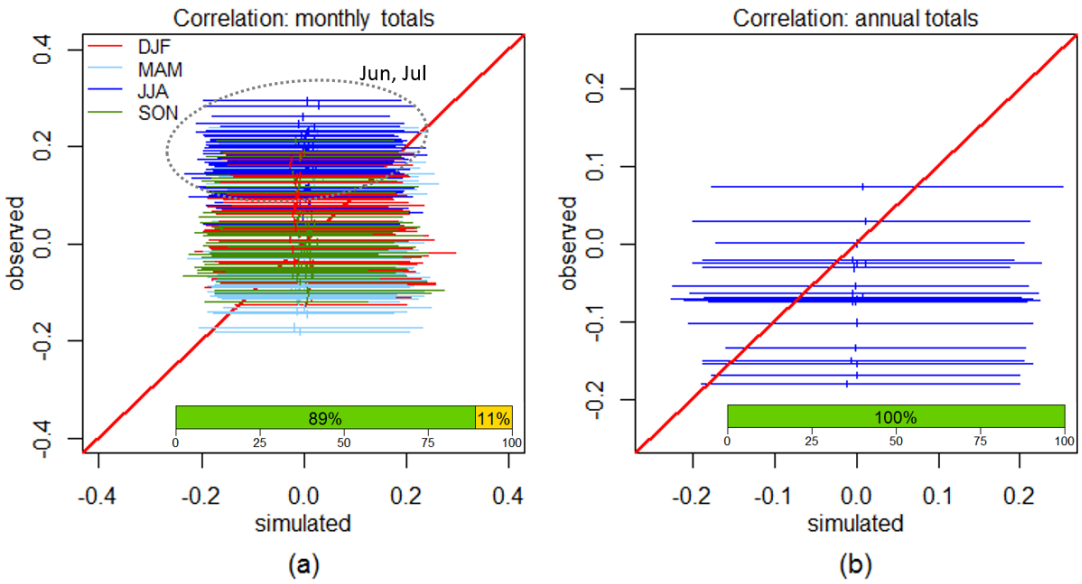


Fig. 3-8 At-site correlations (a) monthly rainfall totals, cumulative performance as a percentage of all sites and months; and (b) annual rainfall totals, 90% probability limits shown, cumulative performance as a percentage of all sites.

Daily rainfall extremes

Fig. 3-9 shows a comparison of the observed and simulated daily annual maximum rainfall for an example site, Coromandel Valley, and the cumulative performance over all sites. Ten sites (53%) across the catchment the model showed 'good' performance in reproducing the Intensity-Frequency-Duration (IFD) relationship, four sites (21%) exhibited 'fair' performance in reproducing the IFD relationship and

five sites (26%) exhibited ‘poor’ performance (Fig. 3-9). The model exhibited ‘Overall Good’ performance in reproducing these rainfall extremes.

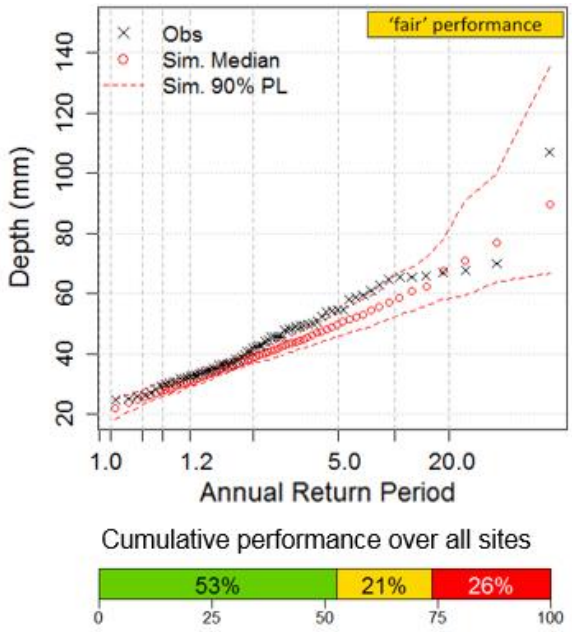


Fig. 3-9 Simulated and observed daily annual maxima example from site 6 (Coromandel Valley) and cumulative performance as a percentage of sites.

3.4.2 Spatial field performance

Multi-site occurrences

In contrast to multisite models which parameterise the covariance for each pair of sites, the continuous model has a single cross-correlation structure over the region for each month. The top panel of Fig. 3-10 illustrates three categories of spatial rainfall coverage: ‘sparse’ rain (Fig. 3-10a), ‘patchy’ rain (Fig. 3-10b) and ‘dense’ rain (Fig. 3-10c). The middle panel shows an example of the distribution of jointly wet sites (March) and the bottom panel summarises model performance over all months for each of the three illustrative categories. The model shows ‘Overall Good’ performance for each category from ‘sparse’ and ‘patchy’ rain coverage (Fig. 3-10a and b) but is deemed ‘Overall Highly Variable’ for the ‘dense’ rain category (Fig. 3-10c) due to the model over-predicting the number of instances in which all 19 sites were wet. Baxevani and Lennartsson (2015) similarly noted the higher probability of observing rainfall at all sites (right hand side of Fig. 3-10c) compared to partial coverage of the region, but their model under-predicted instances where the sites were either all dry or all wet.

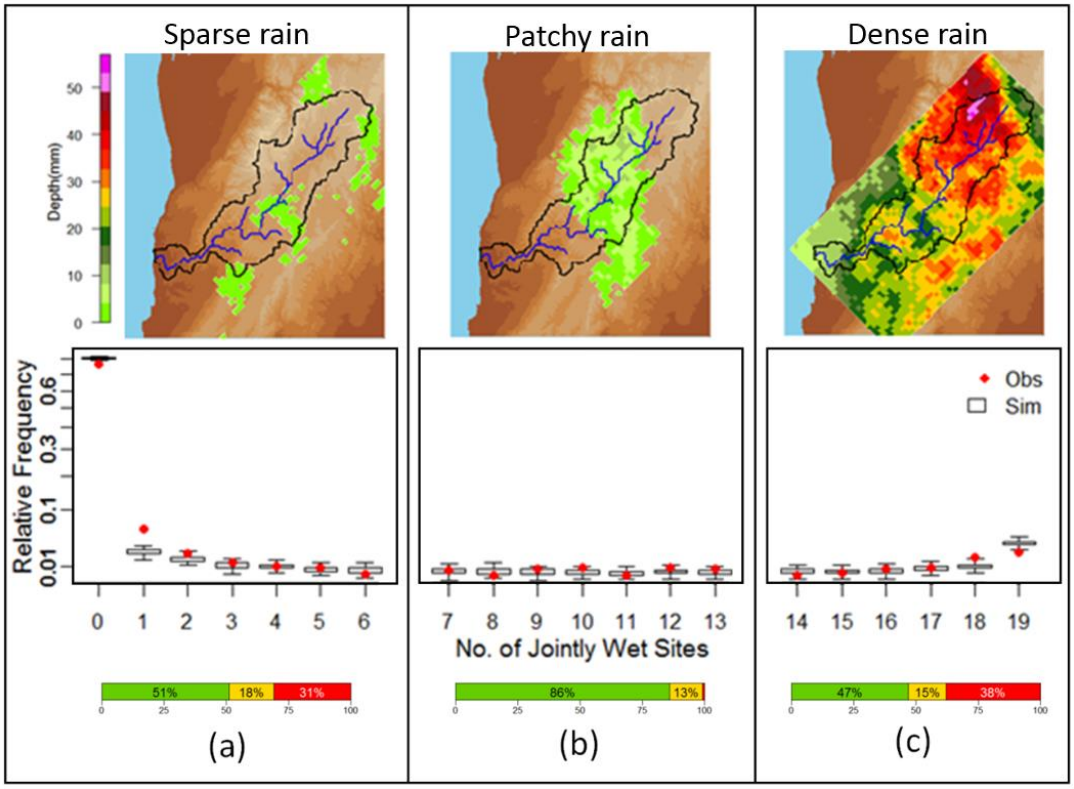


Fig. 3-10 Distribution of number of jointly wet sites for (a) 'sparse' rain, (b) 'patchy rain' and (c) 'dense' rain. Example shown for March. Cumulative performance shown as a percentage of months and options within a category.

Catchment rainfall

The domain average behaviour of the sites can be assessed using the same statistics as for individual sites. This approach showed that the domain average behaviour had 'Overall Fair – Good' performance when reproducing the daily statistics, wet spell/dry spell distributions, annual and monthly totals, correlations and extremes. As with the individual sites, the model showed poorer performance in reproducing the lower-tail of the annual total rainfall distributions (See Supplementary Material A), but this was to be expected.

The model's ability to reproduce the spatial rainfall gradient over the domain shown in the region's average annual total rainfall was evaluated. This evaluation compared the simulated average annual total rainfall distributions and the corresponding interpolated surface of the observed average annual total rainfall (shown in Fig. 3-2) at each grid point in the region. The interpolated observed average annual total rainfall fell within the 90% limits of the simulated mean annual total rainfall for 78% of the region indicating 'good' performance. Another 22% of the region showed

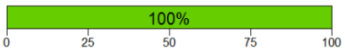
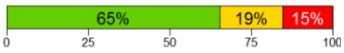
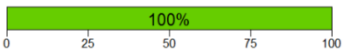
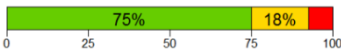
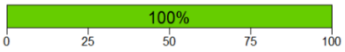
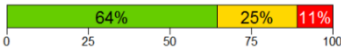
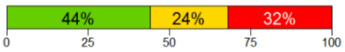
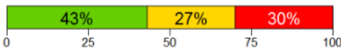
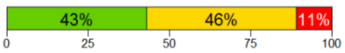
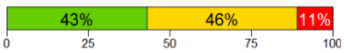
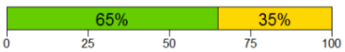
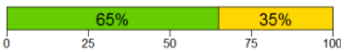
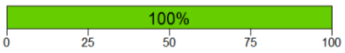
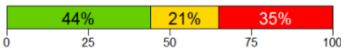
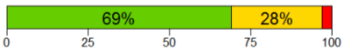
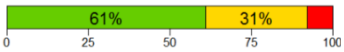
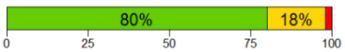
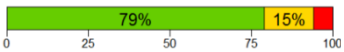
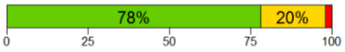
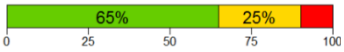
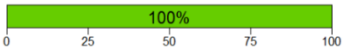
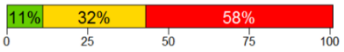
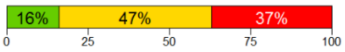
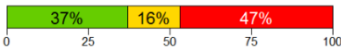
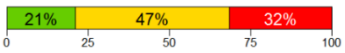
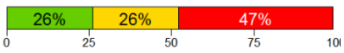
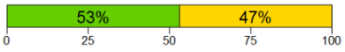
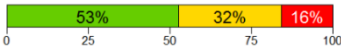
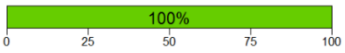
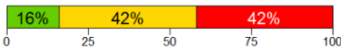
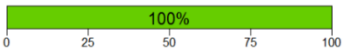
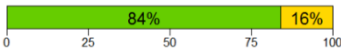
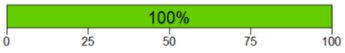
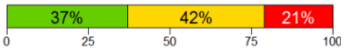
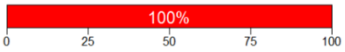
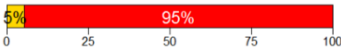
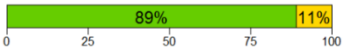
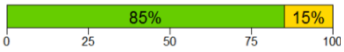
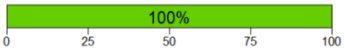
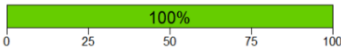
'fair' performance. The instances having the greatest difference occurred at the very high elevations and near the boundaries, which reflects a limitation of the interpolation approach. Nevertheless, the large majority (78%) of the grid points having 'good' performance indicates that the gradient in the annual rainfall total is 'Overall Good'.

3.4.3 Cross-validation performance

The model was evaluated using a leave-one-out cross-validation approach (See Section 3.2.3). There was minimal difference between the observed and predicted parameters over the region (See Supplementary Material B), suggesting that the regression against elevation and the variogram parameters are appropriate. This is further assessed by comparing the at-site performance of the model calibrated using all data against the cross-validated at-site performance.

Table 3-5 summarises the performance of the model when using all sites in calibration and the performance at each site when that site is removed from calibration. The cross-validation shows some decrease in performance, but this decrease typically occurs when sites nearer the boundary (e.g. Site No. 11) are left out and there is little other nearby information to assist the interpolation. This issue is a property of the spatial interpolation scheme more than the model framework. Nevertheless, the performance of the model for monthly and annual rainfall distributions, correlations in rainfall totals and extreme rainfall is predominantly 'Overall Fair – Good'.

Table 3-5 Comparison of 'all data' and 'cross-validation' performance. Overall performance measure summarised to the right of each bar using Table 3-3 classification scheme.

	All data		Cross-validation	
Metric	**Percent of cases**	**Overall classification**	**Percent of cases**	**Overall classification**
Monthly				
Wet day amounts – means		Good		Good
Wet day amounts – std dev		Good		Good
No. wet days – means		Good		Good
No. wet days – std dev		Highly Variable		Highly Variable
Wet spell distribution		Fair – Good		Fair – Good
Dry spell distribution		Good		Good
Total rainfall - means		Good		Highly Variable
Total rainfall – std dev		Good		Good
Total rainfall – lower tail		Good		Good
Total rainfall – upper tail		Good		Good
Annual				
Total rainfall - means		Good		Poor
Total rainfall – std dev		Fair – Poor		Highly Variable
Total rainfall – lower tail		Fair – Poor		Fair – Poor
Total rainfall – upper tail		Good		Good
Wet day amounts - mean		Good		Fair – Poor
Wet day amounts – std dev		Good		Good
No. wet days – means		Good		Fair – Good
No. wet days – std dev		Poor		Poor
Correlations				
Monthly total rainfall		Good		Good
Annual total rainfall		Good		Good
Extremes				
Intensity-Frequency-Duration curve		Good		Highly Variable

Several statistics are worth noting from Table 3-5. As previously discussed (Section 3.4.1), the variability in the number of wet days annually remains under predicted and classified as 'Overall Poor' performance. The performance of the simulation of the mean annual total rainfall drops from 'Overall Good' to 'Overall Poor' between the calibration and validation scenarios. However, the relative difference between the simulated mean and the observed mean annual total rainfall is within 10% for 14 sites (75%). This issue is due to the lack of variability in the simulated mean annual total rainfall, such that changes to the interpolated mean can easily lie outside the 90% limits (See Supplementary Material C). Likewise, the same issue occurred for the simulated mean annual number of wet days, which dropped from 'Overall Good' to 'Overall Fair – Good' between the calibration and validation scenarios. This drop in the number of sites with 'good' performance also occurred in the case of the mean monthly total rainfall for the same reason, but, the cumulative classification became 'Overall Highly Variable' (See Supplementary Material D). Although classed as 'Overall Highly Variable' the relative difference between the simulated mean and the observed mean annual total rainfall is within 10% for 81% of the sites and months.

The simulation of the variability in annual total rainfall changes from 'Overall Fair – Poor' to 'Overall Highly Variable' also due to small changes in the simulation parameters as many of the sites deemed 'fair' in the calibration scenario were near the boundary of being classified as 'good' or 'poor. This was also determined to cause the drop in performance for the IFD curves. The wet spell and dry spell distributions, due to the use of an averaged temporal correlation parameter and the model's low sensitivity to changes in the temporal correlation, do not show a drop in performance between the all data and validation scenarios.

3.5 Discussion

3.5.1 The importance of model parsimony in continuous simulation

Continuous hydrological simulation for applications such as flood and drought risk typically requires long-term sequences of rainfall time series. For example, Li et al. (2014) calculate that to achieve a prediction error of less than 20% in the 1 in 100 year flood estimate 10,000 years of rainfall is required. The greater the level of parsimony in the model, the easier it will be to generate long-term sequences for practical applications to assess hydrological risks. In the literature there is a wide

variety of models having spatiotemporal features (Groppelli et al., 2011; Northrop, 1998; Seed et al., 2013; Seed et al., 1999; Zhang and Switzer, 2007). However, their complexity means they typically are not suitable for long-term continuous simulation of a catchment. For example, spatiotemporal models that are developed for forecast applications using weather radar (Kim et al., 2009; Seed et al., 2013), implement high levels of complexity to represent the spatial structure of storm events and their spatiotemporal evolution, however they typically restrict their focus to single events. While these complex spatial-temporal rainfall models provide insight on spatiotemporal structure of individual rainfall events, it is not clear how they could be used to generate long-term rainfall sequences suitable for continuous hydrological simulation of a catchment.

The latent variable approach used in this study has a number of features that make it more parsimonious than existing approaches. Firstly, it implicitly accounts for temporal correlations in the wet-dry pattern (Section 3.4.2) as well as the rainfall amounts (Section 3.4.2) and is thus more parsimonious compared to models which simulate rainfall amounts conditional on wet-dry patterns (Kleiber et al., 2012; Wilks, 2009). Secondly, the use of a spatially continuous covariance function has meant that significantly less parameters are used than in multisite models to represent the spatial correlation structure. To specify the spatial correlation structure the number of parameters in a multisite model scales N^2 with the number of sites N, whereas the continuous model has a constant number of parameters ($\bar{\varphi}, \nu, \alpha$ and λ) over the region. Thirdly, while more complex composite transformations have been used (Baxevani and Lennartsson, 2015), in this study a power transformation was adopted. For this case study, the comprehensive evaluation undertaken did not suggest that a more complicated composite transformation was needed because at-site performance evaluation showed the wet/dry occurrences and rainfall amounts, and extremes evaluations were classified as ‘Overall Good’ (Table 3-5).

3.5.2 The importance of a comprehensive evaluation framework in continuous simulation

This comprehensive evaluation led to the identification of a model deficiency (lack of variability at annual scales) and, importantly, provided a means to assess the origin of this deficiency. The under prediction of variability in aggregate totals,

termed overdispersion, is a well-known issue with many classes of stochastic precipitation generation models (Katz and Parlange, 1998; Mehrotra and Sharma, 2007; Paschalis et al., 2013; Wilks, 1999). Often, this is attributed to lack of model persistence at the inter-annual timescale, or the lesser acknowledged issue of intra-annual month-to-month variability. However, the comprehensive analysis showed that inter-annual correlations were 'Overall Good' (Section 3.4.1) and the intra-annual correlations were 'Overall Good'. In this instance, the lack of variability in the number of wet days simulated annually was determined to be the cause of a lack of variability in annual total rainfall amounts (See Section 3.4.1). Specifically, the model showed 'poor' performance in simulating drier years. This diagnosis demonstrates the value of a comprehensive evaluation, because identifying the root cause of the issue can lead to a differing remedy. In this instance, the poor model performance in drier years suggests model improvement might potentially consider drier years in more detail rather than focus on the issue of inter-annual persistence.

Another key advantage of the comprehensive evaluation has been a direct comparison between the performance using all data and against the performance using leave-one-out cross-validation ('LCV') for all sites. This is rare in studies that present continuous simulation approaches. This analysis showed that overall there was not a large decrease in performance for LCV, which provides greater confidence in model performance ability. The largest differences in LCV was for locations with less adjacent surrounding gauges or higher elevations, most likely due to larger uncertainty arising from the spatial interpolation, which concurs with Kleiber et al. (2012). The overall low change for LCV may be because of the relatively dense rain network and relatively small elevations in the case study – this may not be the case in other regions. Approaches to remedy this are discussed in the following section.

Spatial properties were also included as part of the comprehensive evaluation, while the results shown provided no evidence to suggest any major deficiencies, this component of model evaluation could be further enhanced by the use of spatial rainfall information (e.g. radar, satellite).

3.5.3 Future development and applications of framework

The comprehensive evaluation identified the variance in annual totals and occurrences as being a limiting feature of the model for the given case study. Future

versions of the model may address this issue, for example, by conditioning the model on weather states, conditioning on covariates and model nesting over multiple time scales (Sharma and Mehrotra, 2013).

The LCV evaluation identified some sites that with larger decrease in performance, which was postulated to be due to the spatial interpolation. This could be addressed by incorporating the uncertainty in the interpolation approach, as undertaken by Kleiber et al. (2012), or developing more sophisticated interpolation techniques.

Future research will also include evaluation of the approach at wide range of locations and/or additional data sources (e.g radar) and applications in different contexts, such as conditional simulation (e.g. Renard et al., 2011) or as a weather generator simulating fields of variables such as temperature or evapotranspiration (Srikanthan and McMahon, 2001). These extensions may highlight the need for further model enhancements.

3.6 Conclusions

The goal of this study was to develop a model capable of generating long-term continuous rainfall fields suitable for hydrological simulation for assessing flood and drought risk. Hence model parsimony and ease of calibration were important. For this reason, a latent variable approach was adopted because it provides a parsimonious method to jointly generate rainfall occurrence and amount. Furthermore, a parsimonious approach was adopted for the formulation of temporal and spatial correlation structure. A comprehensive evaluation approach was developed using a performance classification system that provided a systematic, succinct and transparent method to assess and summarize model performance over a range of statistics, sites, scales and seasons. Importantly it was able to identify model strengths and weaknesses.

The evaluation of the results used a wide range of statistics which were important from a hydrological perspective. This included rainfall occurrence/amounts, wet/dry spell distributions, seasonality, annual maximum extremes, spatial gradients, temporal and spatial correlations across range of time scales from daily to annual. The model showed many strengths in reproducing the observed rainfall characteristics with the majority of statistics classified as either statistically

indistinguishable from the observed or within 5% of the observed across the majority of sites and seasons. One of the few weaknesses of the model was that the total annual rainfall in dry years (lower 5%) was over-estimated by 15% on average over all sites. An advantage of the comprehensive evaluation was that it was able identify the source of this over-estimation was poor representation of the annual variability of rainfall occurrences.

Further research will address these weaknesses, and then apply the model in different regions using the comprehensive evaluation framework to identify if further enhancements are required. Given the strengths of this continuous daily rainfall field model it has a range of potential hydrological applications because it provides the ability to estimate streamflow over an entire catchment.

3.7 Acknowledgements

This work was supported by an Australian Research Council Discovery grant: A new flood design methodology for a variable and changing climate DP1094796. Additional support was provided by the CSIRO Climate Adaptation Flagship. The authors gratefully thank Peter Rasmussen for his help in calibrating his 2013 multisite precipitation model.

3.8 References

Allcroft, D.J., Glasbey, C.A., 2003. A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. Journal of the Royal Statistical Society Series C-Applied Statistics, 52: 487-498. DOI:10.1111/1467-9876.00419

Bardossy, A., Plate, E.J., 1992. Space-time model for daily rainfall using atmospheric circulation patterns. Water Resources Research, 28(5): 1247-1259. DOI:10.1029/91wr02589

Baxevani, A., Lennartsson, J., 2015. A spatiotemporal precipitation generator based on a censored latent Gaussian field. Water Resources Research.

Bell, T.L., 1987. A space-time stochastic model of rainfall for satellite remote-sensing studies. Journal of Geophysical Research: Atmospheres, 92(D8): 9631-9643. DOI:10.1029/JD092iD08p09631

Bennett, B., Lambert, M., Thyer, M., Bates, B., Leonard, M., 2015. Estimating Extreme Spatial Rainfall Intensities. Journal of Hydrologic Engineering, 0(0): 04015074. DOI:doi:10.1061/(ASCE)HE.1943-5584.0001316

Candela, L., Tamoh, K., Olivares, G., Gomez, M., 2012. Modelling impacts of climate change on water resources in ungauged and data-scarce watersheds. Application to the Siurana catchment (NE Spain). Science of The Total Environment, 440: 253-260. DOI: http://dx.doi.org/10.1016/j.scitotenv.2012.06.062

Cressie, N., 1993. Statistics for Spatial Data: Wiley Series in Probability and Statistics. Wiley: New York, NY, USA.

Davison, A.C., Padoan, S.A., Ribatet, M., 2012. Statistical Modeling of Spatial Extremes. 161-186. DOI:10.1214/11-STS376

Dietrich, C.R., Newsam, G.N., 1993. A fast and exact method for multidimensional gaussian stochastic simulations. Water Resources Research, 29(8): 2861-2869. DOI:10.1029/93WR01070

Durban, M., Glasbey, C.A., 2001. Weather modelling using a multivariate latent Gaussian model. Agricultural and Forest Meteorology, 109(3): 187-201. DOI:10.1016/s0168-1923(01)00268-4

Evin, G., Thyer, M., Kavetski, D., McInerney, D., Kuczera, G., 2014. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. Water Resources Research, 50(3): 2350-2375.

Genton, M.G., 2007. Separable approximations of space-time covariance matrices. Environmetrics, 18(7): 681-696.

Groppelli, B., Bocchiola, D., Rosso, R., 2011. Spatial downscaling of precipitation from GCMs for climate change projections using random cascades: A case study in Italy. Water Resources Research, 47(3): W03519. DOI:10.1029/2010WR009437

Heneker, T.M., Lambert, M.F., Kuczera, G., 2001. A point rainfall model for risk-based design. J. Hydrol., 247(1-2): 54-71. DOI:Doi: 10.1016/s0022-1694(01)00361-4

Hundecha, Y., Pahlow, M., Schumann, A., 2009. Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes. Water resources research, 45(12).

Jeffrey, S.J., Carter, J.O., Moodie, K.B., Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environ. Modell. Softw., 16(4): 309-330. DOI:http://dx.doi.org/10.1016/S1364-8152(01)00008-1

Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous univariate distributions, vol. 1-2. New York: John Wiley & Sons.

Katz, R.W., Parlange, M.B., 1998. Overdispersion phenomenon in stochastic modeling of precipitation. Journal of Climate, 11(4): 591-601.

Kim, S., Tachikawa, Y., Sayama, T., Takara, K., 2009. Ensemble flood forecasting with stochastic radar image extrapolation and a distributed hydrologic model. Hydrological Processes, 23(4): 597-611. DOI:10.1002/hyp.7188

Kleiber, W., Katz, R.W., Rajagopalan, B., 2012. Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. Water Resources Research, 48: 17. DOI:10.1029/2011wr011105

Kwon, H.-H., Sivakumar, B., Moon, Y.-I., Kim, B.-S., 2011. Assessment of change in design flood frequency under climate change using a multivariate downscaling model and a precipitation-runoff model. Stochastic Environmental Research and Risk Assessment, 25(4): 567-581. DOI:10.1007/s00477-010-0422-z

Leonard, M., Lambert, M.F., Metcalfe, A.V., Cowpertwait, P.S.P., 2008. A space-time Neyman-Scott rainfall model with defined storm extent. Water Resources Research, 44(9): W09402.

Li, C., Singh, V.P., Mishra, A.K., 2012. Simulation of the entire range of daily precipitation using a hybrid probability distribution. Water Resources Research, 48(3).

Li, J., Thyer, M., Lambert, M., Kuczera, G., Metcalfe, A., 2014. An efficient causative event-based approach for deriving the annual flood frequency distribution. J. Hydrol., 510: 412-423.

Mehrotra, R., Sharma, A., 2007. A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability. J. Hydrol., 335(1): 180-193.

Mehrotra, R., Srikanthan, R., Sharma, A., 2006. A comparison of three stochastic multi-site precipitation occurrence generators. J. Hydrol., 331(1–2): 280-292. DOI:http://dx.doi.org/10.1016/j.jhydrol.2006.05.016

Northrop, P., 1998. A clustered spatial-temporal model of rainfall. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 454(1975): 1875-1888. DOI:10.1098/rspa.1998.0238

Onof, C., Wheater, H.S., 1993. Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulse Model. J. Hydrol., 149(1): 67-95. DOI:http://dx.doi.org/10.1016/0022-1694(93)90100-N

Paschalis, A., Molnar, P., Fatichi, S., Burlando, P., 2013. A stochastic model for high-resolution space-time precipitation simulation. Water Resources Research, 49(12): 8400-8417. DOI:10.1002/2013WR014437

Qin, J., 2010. A High-Resolution Hierarchical Model for Space-time Rainfall. PhD Thesis, University of Newcastle, Newcastle.

Rasmussen, P., 2013. Multisite precipitation generation using a latent autoregressive model. Water Resources Research, 49(4): 1845-1857.

Renard, B. et al., 2011. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. Water Resources Research, 47(11).

Rodriguez-Iturbe, I., Cox, D.R., Isham, V., 1988. A Point Process Model for Rainfall: Further Developments. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 417(1853): 283-298. DOI:10.1098/rspa.1988.0061

Sanso, B., Guenni, L., 2000. A nonstationary multisite model for rainfall. Journal of the American Statistical Association, 95(452): 1089-1100. DOI:10.2307/2669745

Seed, A.W., Pierce, C.E., Norman, K., 2013. Formulation and evaluation of a scale decomposition-based stochastic precipitation nowcast scheme. Water Resources Research, 49(10): 6624-6641. DOI:10.1002/wrcr.20536

Seed, A.W., Srikanthan, R., Menabde, M., 1999. A space and time model for design storm rainfall. Journal of Geophysical Research: Atmospheres, 104(D24): 31623-31630. DOI:10.1029/1999JD900767

Sharma, A., Mehrotra, R., 2013. Rainfall Generation, Rainfall: State of the Science. American Geophysical Union, pp. 215-246. DOI:10.1029/2010GM000973

Srikanthan, R., McMahon, T.A., 2001. Stochastic generation of annual, monthly and daily climate data: A review. Hydrology and Earth System Sciences, 5(4): 653-670.

Srikanthan, R., Pegram, G.G.S., 2009. A nested multisite daily rainfall stochastic generation model. J. Hydrol., 371(1–4): 142-153. DOI:http://dx.doi.org/10.1016/j.jhydrol.2009.03.025

Weiler, H., 1959. Means and standard deviations of a truncated normal bivariate distribution. Australian Journal of Statistics, 1(3): 73-81.

Westra, S., Thyer, M., Leonard, M., Kavetski, D., Lambert, M., 2014. Impacts of Climate Change on Surface Water in the Onkaparinga Catchment. Final Report Volume 1: Hydrological Model Development and Sources of Uncertainty, Goyder Institute for Water Research, Adelaide, South Australia.

Wilks, D.S., 1998. Multisite generalization of a daily stochastic precipitation generation model. J. Hydrol., 210(1–4): 178-191. DOI:http://dx.doi.org/10.1016/S0022-1694(98)00186-3

Wilks, D.S., 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agricultural and Forest Meteorology, 93(3): 153-169. DOI:http://dx.doi.org/10.1016/S0168-1923(98)00125-7

Wilks, D.S., 2009. A gridded multisite weather generator and synchronization to observed weather data. Water resources research, 45(10).

Zhang, Z., Switzer, P., 2007. Stochastic space-time regional rainfall modeling adapted to historical rain gauge data. Water Resources Research, 43(3): W03441. DOI:10.1029/2005WR004654

Chapter 4

A Virtual Hydrological Framework to Evaluate Stochastic Rainfall Models (Paper 3)

Bree Bennett, Mark Thyer, Michael Leonard, Martin Lambert and Bryson C. Bates

Water Resources Research, submitted May 2016

Statement of Authorship

Title of Paper	A virtual hydrological framework to evaluate stochastic rainfall models.
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Bennett, B., Thyer, M., Leonard, M., Lambert, M., Bates, B.C. 2016. A virtual hydrological framework to evaluate stochastic rainfall models. Water Resources Research, (submitted).

Principal Author

Name of Principal Author (Candidate)	Bree Sarah Bennett
Contribution to the Paper	Development and implementation of approach, visualisation and interpretation of results, preparation of manuscript and acted as corresponding author.
Overall percentage (%)	85%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 2/6/16

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Mark Thyer
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.
Signature	Date 2/6/16

Name of Co-Author	Michael Leonard
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.
Signature	Date 2/6/16

Name of Co-Author	Martin Lambert		
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.		
Signature		Date	2/6/16

Name of Co-Author	Bryson Bates		
Contribution to the Paper	Supervised research, helped to evaluate and edit the manuscript.		
Signature		Date	2/6/2016

Abstract

Stochastic rainfall modelling is a commonly used technique for evaluating the impact of flooding, drought or climate change in a catchment. While considerable attention is given to the development of stochastic rainfall models, significantly less attention is given to performance evaluation methods. Typical evaluation methods employ a variety of rainfall statistics, but they give limited understanding about which rainfall characteristics are most important for reliable streamflow prediction when one or more aspects of the observed rainfall are poorly reproduced by the rainfall model. To address this issue a new evaluation method for rainfall models is introduced, with three key features: (i) streamflow-based — to give a direct evaluation of modelled streamflow performance, (ii) virtual — to avoid the issue of confounding errors in hydrological model or data, and (iii) targeted — to isolate the source of errors according to specific sites and months. The virtual hydrologic evaluation framework is applied to a case study of 22 sites in South Australia. The framework demonstrated that apparently 'good' modelled rainfall can produce 'poor' streamflow predictions, whilst 'poor' modelled rainfall may lead to 'good' streamflow predictions. The framework identified the importance of rainfall model performance in the 'wetting-up' months of the catchment cycle (May and June in this case study) for providing reliable predictions of streamflow over the entire year despite their low monthly flow volume. This insight would not have been found using existing methods, and highlights the importance of the virtual hydrologic evaluation framework for stochastic rainfall model evaluation.

.

4.1 Introduction

Stochastic rainfall model simulations are primarily used as inputs to a hydrological model, for simulating realisations of streamflow. Streamflow simulations are then used to assess hydrological risks, such as flood risks (e.g. Camici et al., 2011; Li et al., 2016) or drought risks (e.g. Henley et al., 2013; Mortazavi-Naeini et al., 2015; Paton et al., 2013). When evaluating the efficacy of stochastic rainfall models, current approaches, which make comparisons to observed rainfall and/or streamflow, are limited. This is because they are unable to make a targeted evaluation of the stochastic rainfall model's ability to reproduce streamflow characteristics of practical interest. This paper introduces a new virtual framework that enables targeted, hydrological evaluation of stochastic rainfall models.

Stochastic rainfall models are typically evaluated against a wide range of observed rainfall statistics (Baxevani and Lennartsson, 2015; Bennett et al., 2016; Rasmussen, 2013; Srikanthan and Pegram, 2009; Wilks, 2008). Observed-rainfall evaluation involves a multitude of comparisons between observed and simulated rainfall statistics. For the example of a daily stochastic rainfall model, these statistics might include the mean and standard deviation of: rainfall on wet days, wet-spell and dry-spell durations, wet-day proportions, extreme values, and daily total rainfall. This basic list would involve 12 daily-scale statistics per site per month. Aggregated statistics might include the mean, standard deviation and autocorrelation at aggregated monthly and annual time scales, producing another 39 statistics per site. For a multi-site rainfall model having 10 gauges, the evaluation would involve 1830 temporal statistics before considering the statistics of spatial aggregates. Ever-increasing numbers of evaluation statistics makes it unclear which are the most important and how trade-offs in respective performance between them might be evaluated.

Typically, observed rainfall evaluation shows rainfall model performance as 'mixed': the majority of statistics are reproduced well but some are poorly reproduced in a manner that varies depending on the statistic, timescale and site. For example, Bennett et al. (2016) undertook a comprehensive evaluation of a spatial daily rainfall model, and found that at the daily scale the majority of statistics were well reproduced, except for the standard deviation of the number of wet days, with the results varying from month to month and site to site. In January almost 100% of sites

were classified as ‘poor’, whilst in July 100% sites were classified as ‘good’, and other months varied between these cases. This type of ‘mixed’ outcome presents a number of challenges for evaluating predictive performance. Firstly, it is difficult to ascertain if the rainfall model’s performance is sufficient in terms of predictions of practical interest, which are typically streamflow-based. Secondly, it is unclear if it is necessary to invest time and effort to address instances of poor performance, when the majority of statistics are well reproduced.

To overcome the limitation of observed-rainfall evaluation, the alternative is to evaluate the rainfall model’s performance in terms of streamflow, hereafter referred to as ’observed-streamflow evaluation’. Rainfall modelling is not an end in itself, therefore hydrological statistics of interest (such as properties of the flow duration curve as well as annual and monthly total flows), can give better perspective on suitability of the rainfall model for its intended application (see Cowpertwait, 2006). For example, observed-streamflow evaluation is applied to continuous simulation approaches which are used to predict the annual flood frequency distribution (e.g. Blazkova and Beven, 2002; 2009; Camici et al., 2011; McMillan and Brasington, 2008).

Observed-streamflow evaluation typically involves (1) a stochastic model that produces simulations of rainfall, that are (2) input to a hydrological model to produce simulated streamflow, which is (3) converted to the predictions of interest (e.g. for continuous simulation approaches, this is the flood frequency distribution), and (4) compared against the observed streamflow predictions of interest, (e.g. flood frequency distribution based on the observed streamflow, such as in Camici et al. (2011)). The challenge with this approach is that when there is poor predictive performance (i.e. a significant discrepancy between the observed and predicted streamflow) it is difficult to ascertain if the poor performance was caused by the hydrological model or the stochastic rainfall model. Hydrological model performance can vary substantially due to data errors (Andreassian et al., 2001; Kuczera and Williams, 1992; McMillan et al., 2010), and structural errors (Clark et al., 2008; Renard et al., 2010; Renard et al., 2011; Smith et al., 2008). If large data error or large structural error is present, the calibrated hydrological model will poorly represent catchment processes. Even hydrological models with ‘good’ performance can have predictive errors with a standard deviation at 25% of the predicted

streamflow (Evin et al., 2014). With predictive errors of this magnitude or worse it makes it difficult to evaluate the performance of the stochastic rainfall model and identify opportunities for improvement.

The focus of this paper is the development and application of a virtual framework for streamflow-based evaluation of stochastic rainfall models. Virtual experiments have been used previously in a variety of contexts, including the evaluation of hydrological model sensitivity (Ball, 1994; Nicótina et al., 2008; Paschalis et al., 2013; Shah et al., 1996; Wilson et al., 1979), evaluating how well simpler conceptual hydrological models characterise complex processes (Li et al., 2015b), and for developing new techniques for flood frequency analysis (Li et al., 2014; 2016). However, virtual frameworks have not previously been used for evaluating stochastic rainfall models. Virtual-observed streamflow (from a hydrological model having observed rainfall as an input) is a better baseline than observed streamflow for evaluating the stochastic rainfall. The framework overcomes the limitations of observed-rainfall evaluation and observed-streamflow approaches because it: (1) evaluates the performance of the stochastic rainfall model in terms of the key streamflow characteristics of interest; and (2) enables targeted evaluation of the stochastic rainfall model by avoiding confounding issues implied by using observed streamflow (structural errors in the hydrological model and the impact of data error on hydrological model calibration).

The key objectives of this paper are:

1. To introduce a framework for virtual hydrological evaluation of stochastic rainfall models.

2. To introduce two different tests which are part of the framework: an integrated test and a splice test. Combined use of these tests allows streamflow discrepancies to be attributed to their original source in the rainfall model according to site and month.

3. To use the virtual hydrological evaluation framework to evaluate a stochastic rainfall model and contrast the outcomes with standard evaluation methods.

The basic framework of virtual hydrologic evaluation is explained in Section 4.2.1 with the procedure for the integrated test and splice test outlined in Section 4.2.2. The Onkaparinga catchment in South Australia (Section 4.3) is used to illustrate the

procedure (Section 4.4). Discussion and conclusions revisit the features of the framework and the alternative implications it gives for improving rainfall model performance (Sections 4.5 and 4.6).

4.2 Methodology

4.2.1 Evaluation framework

To date, the method of observed-rainfall evaluation has been the most common method for evaluating stochastic rainfall models (Bennett et al. 2016). It involves the comparison of simulated and observed rainfall statistics at a given site for a concurrent period of interest. Virtual hydrological evaluation (Fig. 4-1(a)) is a complementary framework to observed-rainfall evaluation. A hydrological model is used so that simulated streamflow statistics can be compared to virtual-observed streamflow statistics. It is important to clearly define these terms:

- Simulated streamflow — is streamflow produced by the hydrological model by inputting simulated rainfall at a given site.
- Virtual-observed streamflow — is streamflow produced by the hydrological model by inputting observed rainfall at the same given site.

A defining feature of the evaluation framework is that no comparison is made to observed streamflow (though it may have been indirectly used to calibrate the hydrological model). Instead, the framework identifies discrepancies in distributions of flow between the virtual-observed streamflow and the simulated streamflow. Because the hydrological evaluation is a relative comparison of the observed and simulated rainfall, it is important that all other parameters and extraneous variables (e.g. potential evapotranspiration) relating to the hydrological model are kept the same in all instances. Because the hydrological evaluation is virtual, it is not bound by a comparison to observed streamflow sites and avoids the requirement to construct estimates of catchment rainfall. Without a virtual approach, the framework would be unable to evaluate each rainfall site separately, as necessary for isolating rainfall model deficiencies.

The illustration in Fig. 4-1(a) is referred to as an ‘integrated test’ because it demonstrates the performance of the rainfall model as a whole. The hydrological model evaluates the effects of rainfall over longer timescales due to the integrating

property of conceptual stores in the catchment. While the integrated test can isolate rainfall performance at individual sites, it cannot isolate the influence of individual rainfall periods.

Building on the integrated test, an additional test is illustrated in Fig. 4-1(b), referred to as a 'splice test'. The splice test investigates the impact of simulated rainfall in a given influencing month on the production of streamflow in an evaluated month of interest. This is achieved by splicing observed and simulated rainfall into a single time series which is used to produce simulated streamflow.

Consider the time series of observed, R^{obs}, and simulated, R^{sim}, daily rainfall for each year (and replicate) at a given site. Fig. 4-1(b) illustrates the embedding of simulated rainfall R_k^{sim} in an influencing month, k, within observed rainfall R_m^{obs} for all other months $m \in \{1, \dots, 12 | \, m \neq k\}$. The resulting spliced rainfall time series $R_{(k)}^{spl}$ is denoted with respect to the influencing month, and has the same length as the corresponding observed R^{obs} and simulated R^{sim} time series.

$$R_{(k)}^{spl} = \bigcup^{12} \begin{cases} R_m^{sim} ; m = k \\ R_m^{obs} ; m \neq k \end{cases} \tag{1}$$

For example, if June ($k = 6$) is selected as the influencing month, each year of the spliced time series, $R_{(6)}^{spl}$, would be composed as follows:

$$R_{(6)}^{spl} = \left\{ R_1^{obs}, \dots, R_5^{obs}, R_6^{sim}, R_7^{obs}, \dots, R_{12}^{obs} \right\} \tag{2}$$

The ensemble of $k = 1, \dots, 12$ spliced rainfall time series $R_{(k)}^{spl}$ for all influencing months is transformed according to a hydrological model $g[\,]$ to produce an ensemble of simulated streamflows. However, the hydrological evaluation centres on a specific subset of flows $Q_{(t,k)}^{sim}$ relating to the evaluated month, t,

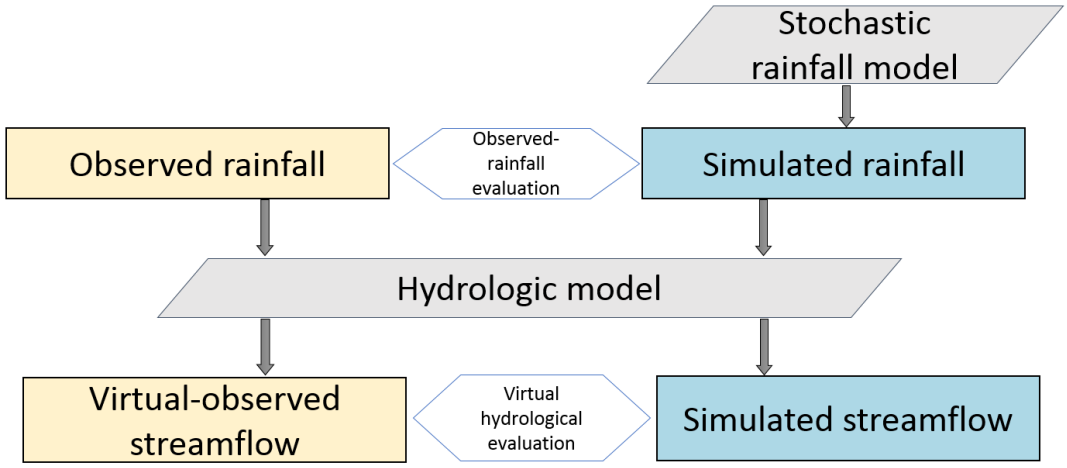
$$Q_{(t,k)}^{sim} \subset_{\mathrm{t}} g\left[R_{(k)}^{spl}\right] \tag{3}$$

The ensemble of simulated streamflows corresponding to influencing month k and evaluated month t are used to determine errors with respect to the virtual-observed streamflow in the evaluated month. Using the function $h[\,]$ to denote a calculated statistic of interest, the relative error in an evaluated month is given by

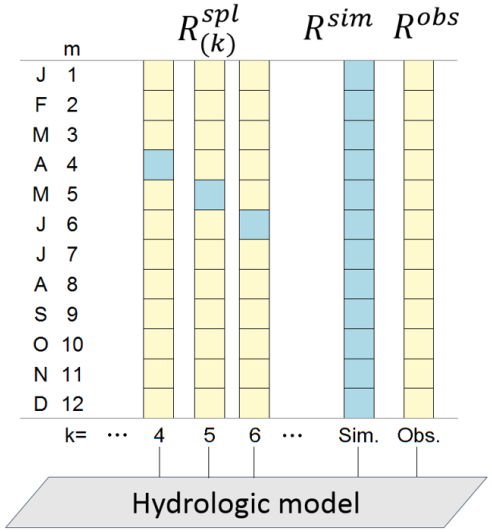
$$\%Err_{(t)} = \frac{h\left[Q_{(t)}^{sim}\right] - h\left[Q_{(t)}^{vo}\right]}{h\left[Q_{(t)}^{vo}\right]} \times 100 \quad (4)$$

where $Q_{(t)}^{vo}$ is the virtual-observed streamflow and $Q_{(t)}^{sim}$ is the simulated streamflow in the evaluated month (whether from the integrated test or any of the 12 splice tests). Thus, for each site, statistic and evaluated month there are 13 errors to compare.

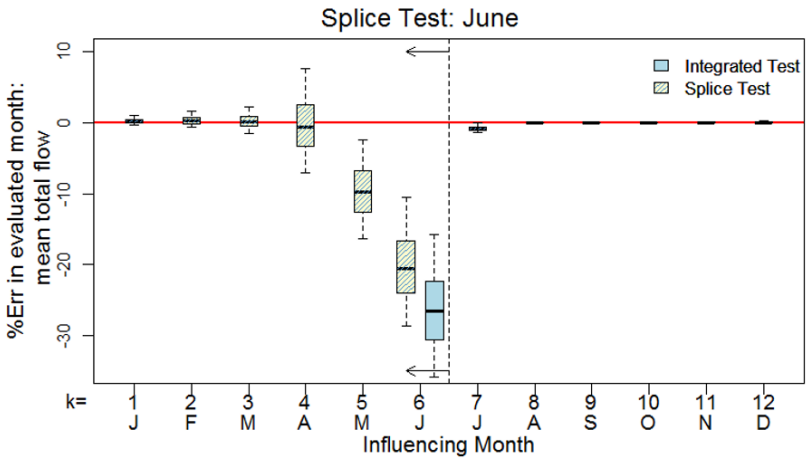
By construction, spliced rainfall is identical to the observed rainfall for all months other than the influencing month, so any errors in streamflow statistics can be attributed to the influencing month free from other factors. Fig. 4-1(c) illustrates for the evaluated month of June ($t = 6$), a typical error profile from a splice test of the mean monthly flow. Fig. 4-1(c) shows the integrated test produced a median error of 27% (blue shaded boxplot), when the influencing month is June ($k = 6$) the median error is 20%, when the influencing month is May ($k = 5$) the median error is 10%, and when the influencing month is April ($k = 4$) the median error is negligible. Therefore, the bias in mean June streamflow is primarily due to rainfall model deficiencies in June and May respectively.



(a)



(b)



(c)

Fig. 4-1 Schematic of (a) the virtual hydrologic evaluation framework where simulated streamflow is compared against virtual-observed streamflow, (b) the method of constructing a splice test by embedding simulated months in an observation time series, and (c) the error profile produced when using the integrated and splice tests for the evaluated month of June (90% limits shown).

4.2.2 Evaluation procedure

Virtual hydrologic evaluation is an additional procedure that follows from, but does not replace, an observed-rainfall evaluation. The procedure for virtual hydrologic evaluation is implemented in a stepwise manner.

The first step is to select a streamflow characteristic of interest, herein termed the 'primary streamflow characteristic'. The reason for specifying a primary streamflow characteristic is to enable a method (carried out in step four) for filtering sites and concentrating the investigation of the rainfall model on sites that perform poorly in terms of its intended application. For example, a flood frequency distribution would be a suitable characteristic for a flood risk study, or the distribution of annual total flow when investigating yield.

The second step is to select an appropriate hydrological model to simulate the streamflow. The hydrological model should be selected on the basis that it is capable of simulating streamflow for the timescales, magnitudes and physical processes of interest to the intended application. For example, a capability for simulating peak flows is important for flood risk studies.

The third step is to conduct an integrated test for each rainfall site. The integrated test serves as an overall test of the rainfall model's performance. The test uses the observed rainfall time series and all replicates of the simulated rainfall time series as inputs to the hydrological model (Fig. 4-1 (a)).

The fourth step assumes a multi-site evaluation and is optional if only a single site is evaluated. The set of investigated sites is narrowed to a smaller subset that poorly simulated the primary streamflow characteristic during the integrated test.

The fifth step is to conduct a detailed analysis of the integrated test results at the monthly scale for the remaining subset of sites. Evaluating monthly total flows is a valuable test of rainfall model performance as the production of monthly total flow volumes relies on the integration of many daily rainfall characteristics (amount, duration, persistence). Errors in reproducing virtual-observed streamflow are contrasted against the observed-rainfall evaluation so that specific sites and months can be identified for further investigation in steps six and seven. The comparison between the observed-rainfall evaluation and the integrated test can be summarised graphically (see Section 4.4.1) for the subset of sites showing:

- simulated daily rainfall statistics (mean (*m*) daily amounts, standard deviation (*sd*) of daily amounts, mean number of wet days (nwet) and the standard deviation of the number of wet days);
- aggregate rainfall statistics (mean and standard deviation of total rainfall); and
- aggregate streamflow statistics (mean and standard deviation of total flow).

For convenience, model performance is categorised following Bennett et al. (2016) as 'good', 'fair' or 'poor'. Performance was categorised as 'good' if the observed/virtual statistic fell within the 90% limits of the simulated statistic, as 'fair' if the observed/virtual statistics fell outside the 90% limits of the simulated statistic but within the 99.7% limits and otherwise as 'poor'.

The sixth step is to implement the splice test for each site of interest. The full set of spliced rainfall (e.g. spliced rainfall for each month designated as the influencing month $R_{(k)}^{spl}; k = 1, \dots, 12$) is input to the hydrological model. This step is repeated for all available replicates of the spliced time series.

The seventh step is to investigate and compare the results of the splice test (Step 6) and the integrated test (Step 5) selecting each month as the evaluated month in turn. Side-by-side comparison of the results of the integrated test and the splice test is given in terms of the errors for selected monthly and annual statistics (see illustration in Fig. 4-1(c)). The comparison of errors from the splice test forms the basis of interpretation for hydrological insights and their relationship to the rainfall model.

Comparing the splice test across all 12 influencing months with the integrated test can lead to several possible interpretations of where streamflow errors in the evaluation month originate. Differentiating the cases allows for rainfall model improvements to be targeted in terms of their ultimate impact on streamflow statistics. For example, streamflow errors might originate from (i) rainfall model deficiencies mostly in the evaluated month, (ii) rainfall model deficiencies over a contiguous block of months including and preceding the evaluated month, or (iii) rainfall model deficiencies in a preceding month more so than in the evaluated month. Examples for each of case are presented in Section 4.4.

4.3 Case Study

The Onkaparinga catchment in South Australia is used as a case study (Fig. 4-2). The catchment lies 25km to the south of the Adelaide metropolitan area and contains the largest reservoir in the Adelaide Hills supplying the region (Mount Bold Reservoir). The catchment rainfall has a strong seasonal cycle with the majority occurring in winter and spring (June to November) and with a predominantly dry summer season (December to February). There is a strong rainfall gradient (Table 4-1), with average annual rainfall ranging from approximately 500 mm on the coast (Site No. 19) to over 1000 mm in the region of highest elevations (Site No. 20).

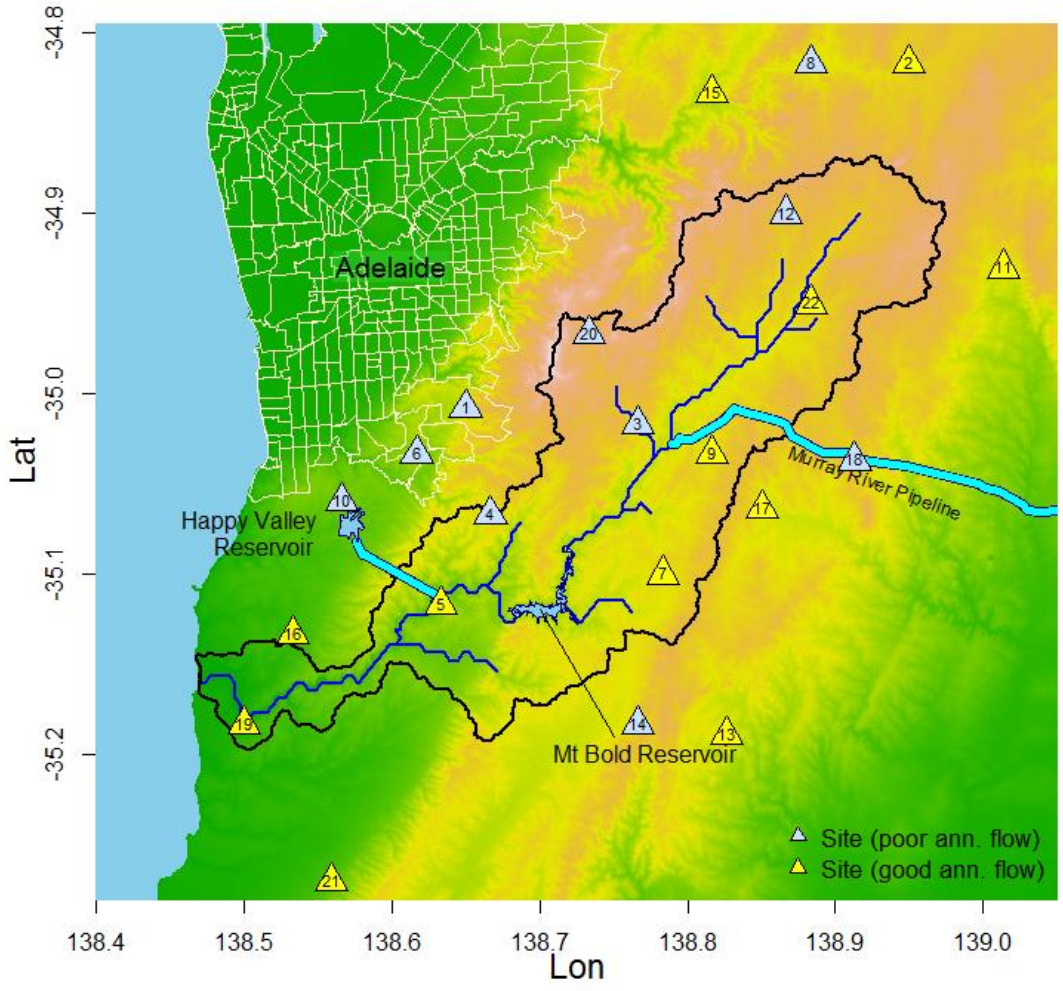


Fig. 4-2 Onkaparinga catchment, South Australia.

Table 4-1 Site names and locations.

Site No	Site Name	Elev (m)	Ann. Ave. Rain (mm)	Site No	Site Name	Elev (m)	Ann. Ave. Rain (mm)
1	Belair	386	790	12	Lobethal	470	880
2	Birdwood	385	720	13	Macclesfield	302	730
3	Bridgewater	376	1050	14	Meadows	384	870
4	Cherry gardens	345	920	15	Cudlee Creek	311	830
5	Clarendon	223	820	16	Morphett Vale	90	560
6	Coromandel Valley	234	710	17	Mount Barker	349	770
7	Echunga	375	805	18	Nairne	403	680
8	Gumeracha	346	790	19	Old Noarlunga	7	520
9	Hahndorf	347	850	20	Uraidla	499	1090
10	Happy Valley	148	640	21	Willunga	158	640
11	Harrogate	335	550	22	Woodside	387	800

The simulated rainfall was determined from the latent variable autoregressive daily rainfall model of Bennett et al. (2016) using at-site calibrated parameters. The rainfall model was calibrated and simulated at 22 locations throughout the catchment that have long, high-quality records (Table 4-1). 10,000 replicates of simulated rainfall covering a 73 year period (1914-1986) were used.

In this paper annual total flow volumes were designated as the primary streamflow characteristic to narrow the number of sites investigated (Step 1).

The hydrological model, GR4J (Perrin et al., 2003) was used to simulate virtual-observed streamflow (Step 2). GR4J is a daily lumped hydrological model that simulates daily streamflow in a parsimonious manner using four parameters. The GR4J model was calibrated to simulate streamflow for the Onkaparinga catchment (Westra et al., 2014) and was good fit to the observed streamflow, with a Nash-Sutcliffe efficiency of 0.8. The model has also been used for other virtual evaluation studies (Li et al. 2014; 2016).

4.4 Results

Following the selection of annual total flow as the primary streamflow characteristic (Step 1) and selection of the hydrological model (Step 2), Step 3 was undertaken to evaluate the simulated rainfall at the 22 sites.

The annual total flow distribution was used to give a broad indication of performance (see Step 4). This step categorised 12 of the 22 sites as ‘good’, which is in strong contrast to earlier evaluation efforts using observed-rainfall (Bennett et al. 2016) that

categorised the majority of sites and statistics as ‘good’ (see Section 4.2.2 for category definitions). Following Step 4, the 10 sites categorised ‘poor’ are the focus of subsequent steps in the hydrologic evaluation framework. The chosen subset of ‘poor’ sites is indicated by the blue triangles in Fig. 4-2.

4.4.1 Integrated test

Fig. 4-3 compares the rainfall model’s performance for the 10 sites in terms of both observed-rainfall evaluation and the integrated test (Step 5).

Although the rainfall model performed well in simulating the total rainfall means and standard deviations for the majority of months and sites, it is clear from Fig. 4-3 that ‘poor’ performance in rainfall did not necessarily translate to ‘poor’ performance is simulating streamflow. For example, in January the rainfall model’s ability to simulate variability in the number of wet days, *sd(nwet),* was ‘poor’ for all sites. However, Fig. 4-3 shows that despite this, most sites had ‘good’ performance in simulating the January distribution of monthly total flow. Similar results were seen in months February, May and November.

Conversely, ‘good’ performance in the observed-rainfall evaluation does not necessarily translate to ‘good’ performance in the simulation of monthly total flows. For example, June and August (Fig. 4-3) have large percentages of ‘poor’ sites for simulating monthly total flow. This deficiency would have been difficult to infer using the observed-rainfall evaluation (Bennett et al., 2016) due to the overall ‘good’ performance of rainfall in these months. The observed-rainfall evaluation had unbiased total annual rainfall (100% ‘good’) and yet the mean annual total flows showed only 10% of sites as ‘good’. Discussion of the splice tests in the following section will investigate reasons why apparently ‘good’ rainfall can yield ‘poor’ flow

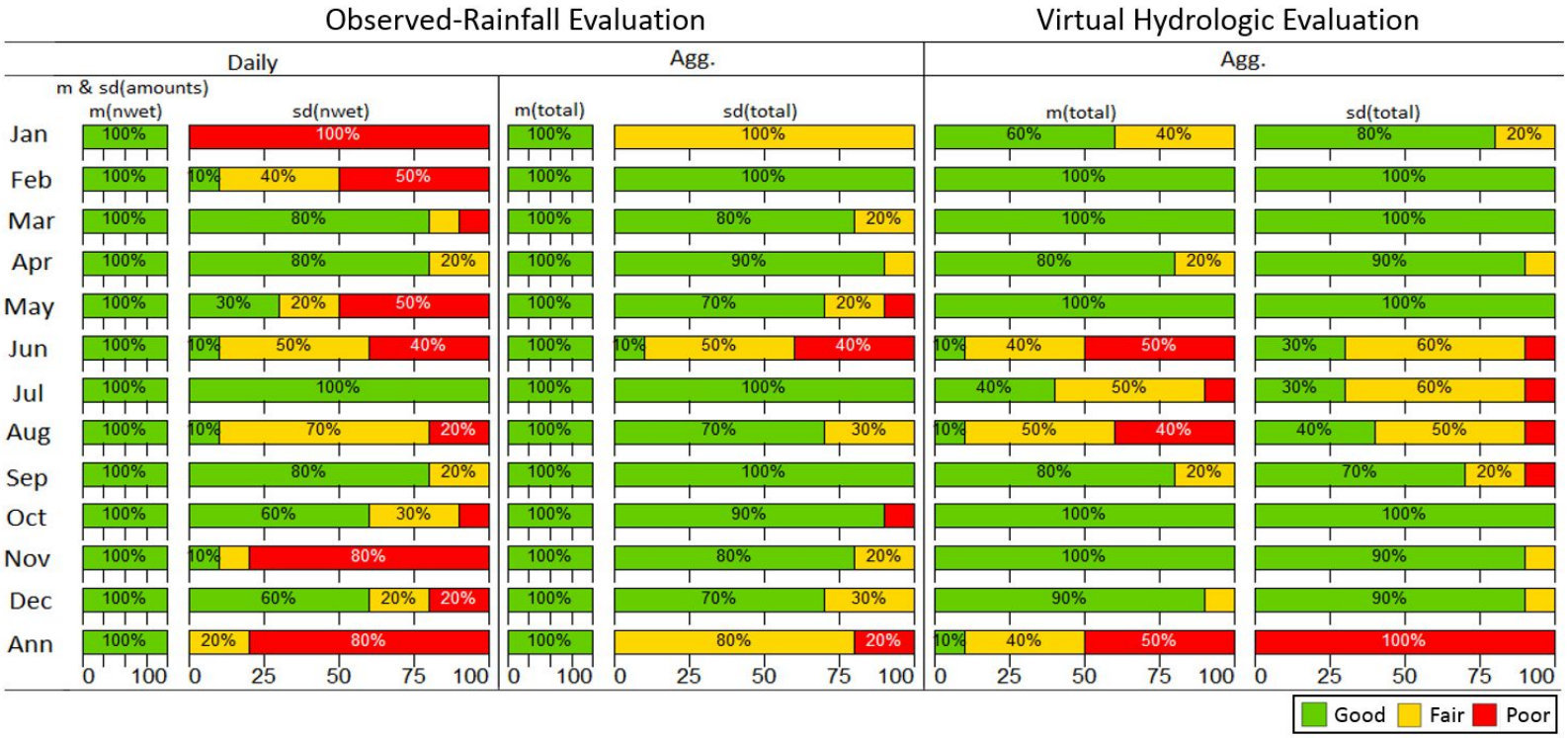


Fig. 4-3 Integrated test, comparing observed-rainfall evaluation (left) with the virtual hydrologic evaluation (right). Comparison of daily and aggregate ('Agg.') rainfall statistics against aggregate flow statistics for individual months and years. Daily rainfall statistics presented include the means and standard deviations of amounts (m(amounts), sd(amounts)), means and standard deviations of the number of wet days (m(nwet), sd(nwet)). The aggregate rainfall statistics presented include means and standard deviations of total rainfall volumes (m(total), sd(total)). The aggregate streamflow statistics presented are the means and standard deviations of total flow volumes (m(total), sd(total)).

4.4.2 Splice test – influence of individual monthly rainfall isolated

The splice test was run for the subset of 10 sites (Step 6) and the results of the integrated and splice tests were compared (Step 7). This splice test results and comparison with integrated test results are illustrated for two locations, Site 12 and Site 10 (see Fig. 4-2).

Streamflow errors mostly originate from rainfall model deficiencies in the evaluated month

A common and obvious case for streamflow errors is that they originate from rainfall in the same month. This case is illustrated in Fig. 4-4 for Site 12 in September where left-side panels show results for the mean and right-side panels show the standard deviation and where panels (a) and (b) summarise the observed-rainfall evaluation, (c) and (d) summarise the integrated test, (e) and (f) summarise the splice test for September. From panels (a) and (b) the simulated monthly rainfall is unbiased but from (c) and (d) the simulated streamflow is biased low and underdispersed, including for September. The error due to splicing the simulated September rainfall is greatest and closest to the error for the integrated test for both the mean monthly total flow (Fig. 4-4 (e)) and standard deviation of monthly total flow (Fig. 4-4 (f)). For the example of the standard deviation, whereas the September splice results in a median error of about 16%, the July splice results in a median error of less than 2% (Fig. 4-4 (f)). Therefore, to improve September flows, September rainfall should be improved in preference to all other months.

Analysing other sites and months suggests that over 50% of the evaluations correspond to this case, and they typically occur in spring and summer months when the catchment is drying out. Examination of the simulated flow duration curves instead of monthly totals (not shown) yields a similar conclusion.

Streamflow errors originate from rainfall model deficiencies over a contiguous block of months

An illustration of the case where streamflow errors originate from rainfall model deficiencies over a contiguous block of months is provided by Site 12 in July. Comparison of the July performance in the integrated and splice tests (Fig. 4-4 (g) and (h)) demonstrates that the errors in July streamflow do not originate in the July

rainfall alone (unlike the case for September – see above). Although the largest percentage error in flow is attributable to July (a median error of 8% in mean monthly total flow and 25% in the standard deviation of monthly total flow) a significant proportion of the error for July streamflow originates in prior months. June and May rainfall have a significant influence on the July flow with percentage errors of up to 15% in July flow when the months of June or May are spliced. Therefore, to improve July flows, it is not just the July rainfall that should be improved but also the preceding two months. Typically autumn (transitional) and winter months (May to August) fall in this case where streamflow errors originate from rainfall model deficiencies over a contiguous block of months, approximately 40% of the site/month combinations.

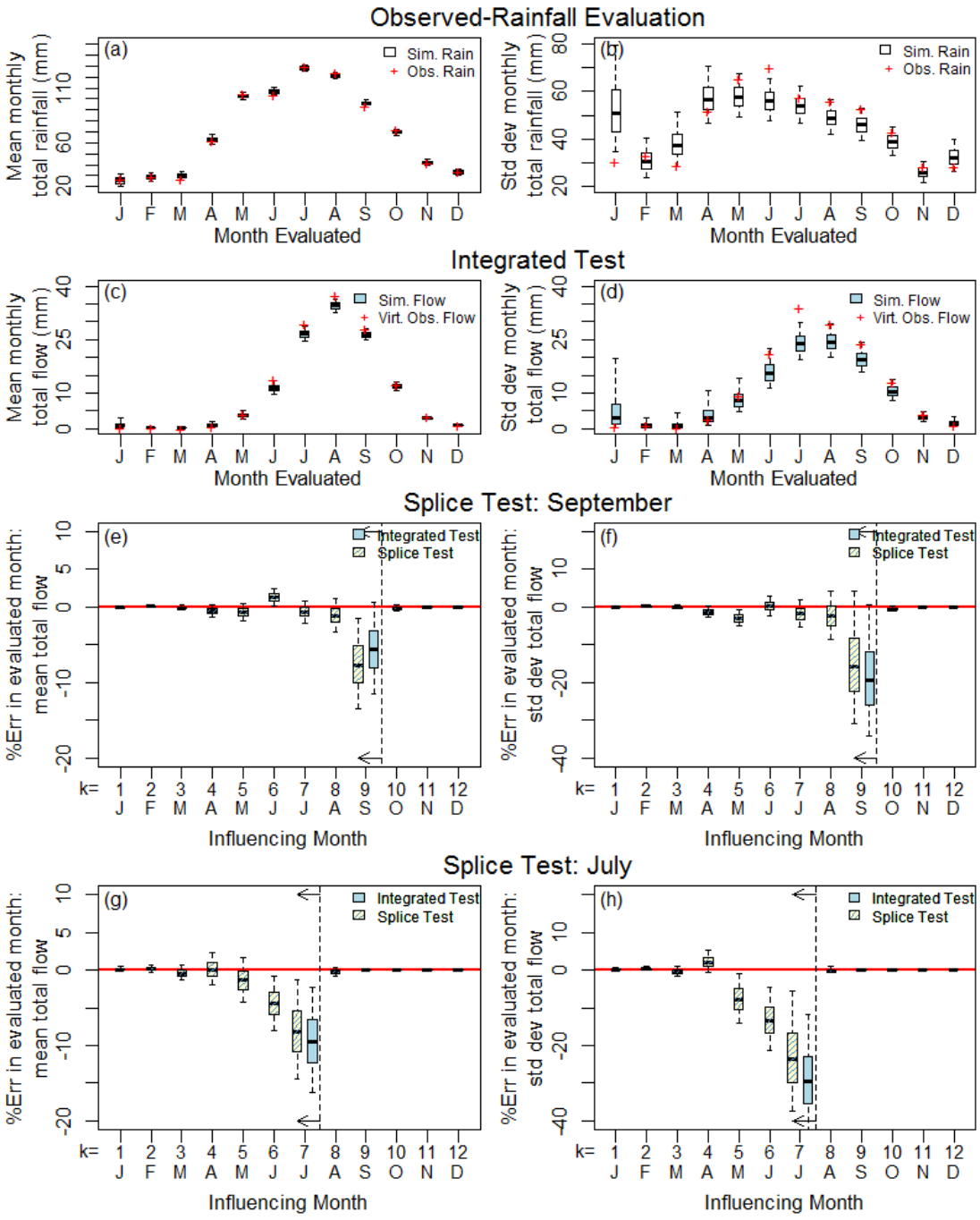


Fig. 4-4 Lobethal, Site 12 (90% limits shown) (a) observed-rainfall evaluation mean monthly total rainfall, (b) observed-rainfall evaluation standard deviation of monthly total rainfall, (c) integrated test mean monthly total rainfall, (d) integrated test standard deviation of monthly total rainfall (e) splice test error in mean monthly flow (September), (f) splice test error in standard deviation of monthly flow (September), (g) splice test error in mean monthly total flow (July), and (h) splice test error in standard deviation of monthly total flow (July).

Streamflow errors originate from rainfall model deficiencies in a preceding month more so than evaluated month

An example of the case where streamflow errors originate from rainfall model deficiencies in a preceding month is provided by Site 10. July is selected as an illustrative case for application of the splice test since it shows biased flow (See Fig. 4-5 (c) and (d)), but did not show any bias in the simulated rainfall (See Fig. 4-5 (a) and (b)). Comparison of July performance for the integrated and monthly splice tests demonstrates that the largest contributor to error in July flow is not July rainfall but June rainfall (Fig. 4-5 (e) and (f)). That is, the largest errors occur when there is observed rainfall for July spliced with simulated rainfall for June. In contrast, simulated July rainfall spliced with observed rainfall in other months, yields a smaller median error.

While improving the July rainfall will improve the simulation of July flow, a more significant improvement will be obtained by focusing on improving the June rainfall. The category where streamflow errors originate from rainfall model deficiencies in a preceding month represents about 10% of the site/month combinations for this case study.

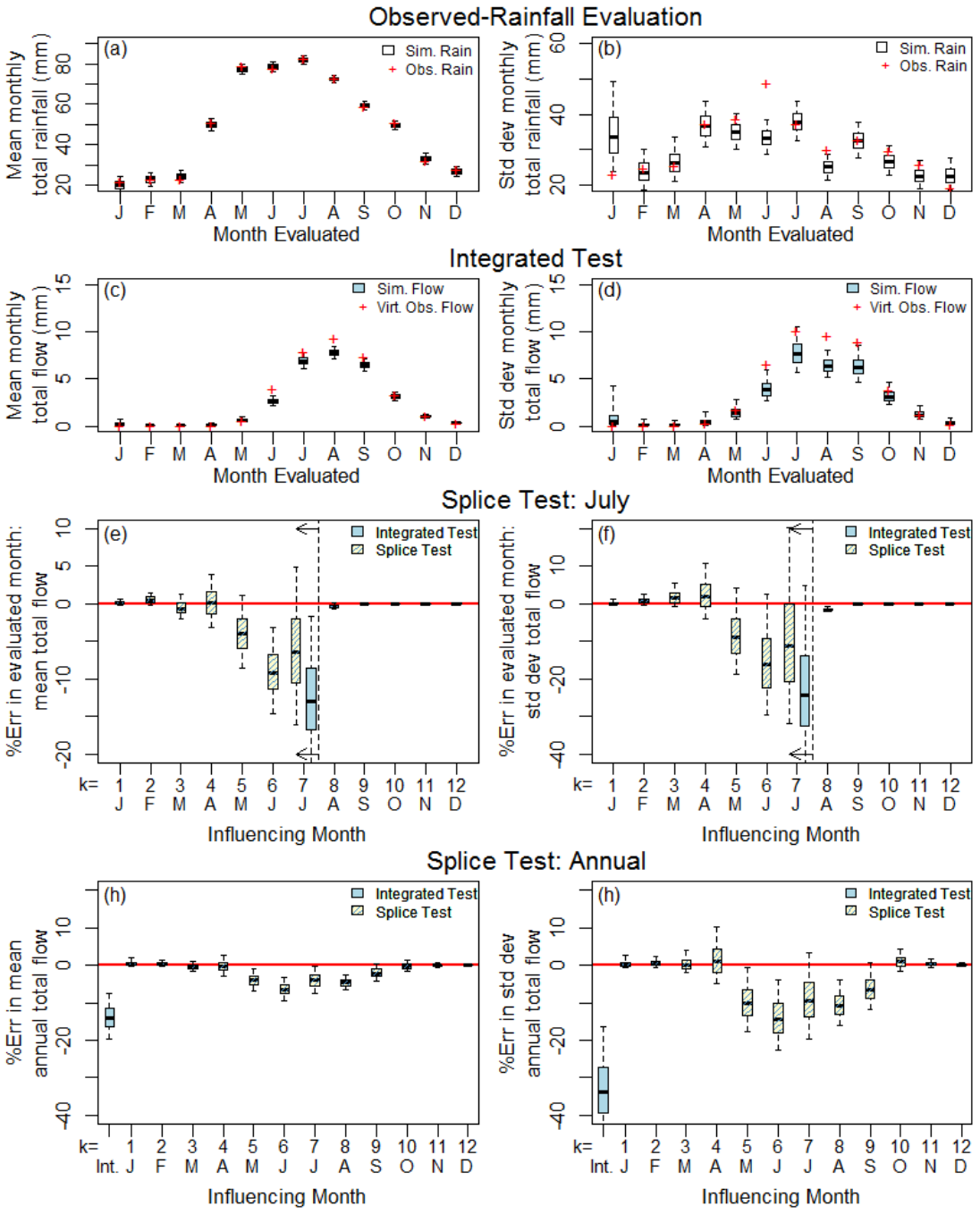


Fig. 4-5 Happy Valley (Site 10) (90% limits shown) (a) observed-rainfall evaluation mean monthly total rainfall, (b) observed-rainfall evaluation standard deviation of monthly total rainfall, (c) integrated test mean monthly total rainfall, (d) integrated test standard deviation of monthly total rainfall (e) splice test error in mean monthly flow (July), (f) splice test error in standard deviation of monthly flow, (g) splice test error in mean annual total flow, and (h) splice test error in standard deviation of annual total flow.

4.4.3 Influence of monthly rainfall on annual flow volumes

While annual simulated rainfall was unbiased, annual simulated streamflow was biased. An illustration of how errors in annual total streamflow arise from rainfall is shown for Site 10. Fig. 4-5 (g) and (h) show that when the months of May to August are spliced they produce the largest errors in distribution of annual total flow for Site 10. Splices of other months do not significantly degrade the simulation of total annual flow. Improvements to the simulation of annual total flow will therefore come from improving the rainfall model in the 'wetting-up' months of May to August in late-autumn and winter. This insight from the use of splice testing would be difficult to obtain using other evaluation strategies (further discussed in Section 4.5.3).

4.5 Discussion

4.5.1 The importance of streamflow-based evaluation

Streamflow arises from the integration of rainfall processes (e.g. rainfall amounts, occurrences and wet-dry patterns) over a catchment. Features of the catchment, such as catchment storage, thresholds and nonlinearities in the rainfall-streamflow response function, can either act to amplify or dampen the rainfall errors at different times of the year.

In terms of amplification, the elasticity of the rainfall-streamflow response (Chiew, 2006) suggests that catchments can have strong sensitivities to discrepancies in rainfall. Given that the rainfall elasticity of streamflow to rainfall is a factor of 2 to 3.5, using the basic principles of error propagation of linear functions (Ang and Tang, 2007) it follows that a 10% error in mean/standard deviation of rainfall could be amplified to 20-35% error in the mean/standard of streamflow. This indicates that streamflow-based evaluation of rainfall models provides a stronger test than observed-rainfall evaluation in terms of the sensitivity of the statistics. For example, Fig. 4-4 shows that July rainfall statistics were classified as 'good', yet despite this, the streamflow response was 'poor' (see Section 4.5.3 for further discussion). It could be argued that the rainfall results presented in Fig. 4-4 were classified as 'good' because the observed-rainfall evaluation was limited, but the evaluation was methodical and used a comprehensive range of daily and monthly statistics (Bennett et al., 2016). While many rainfall statistics were preserved (means, standard

deviation, extremes, marginal distributions of daily rainfall) the rainfall-streamflow response of the catchment exposes that there are deficiencies in the rainfall model not clearly identified by the observed-rainfall evaluation (Bennett et al., 2016).

In terms of dampened influence, catchment storages and high evapotranspiration can also act to suppress errors in the rainfall simulations. For example, Fig. 4-3. showed that the variability in the number of wet days, *sd*(*nwet*), was 'poor' for all sites in January, yet this did not result in 'poor' streamflow. The high potential evapotranspiration in January indicates that the majority of rainfall in January is converted into actual evapotranspiration yielding little streamflow. Hence, any errors in rainfall do not noticeably impact on January streamflow.

It is clear that streamflow-based evaluation is beneficial in addition to traditional observed-rainfall evaluation. However, a hydrological framework is not without its own challenges (see Sections 4.5.2 and 4.5.3 below).

4.5.2 The importance of virtual evaluation

A benefit of virtual hydrologic evaluation is the ability to undertake streamflow-based evaluation at any site where rainfall is observed and simulated. In contrast, observed-streamflow evaluation relies on having measurements of streamflow – which is only undertaken at limited number of locations (e.g. across Australia there are over 8, 000 observed daily rainfall sites, typically greater than 50 years length, compared to approximately 3,300 observed streamflow sites, typically 20 years length). Virtual evaluation can be undertaken at a much wider number of locations. For example, annual total flow volume was compared at 22 virtual-observed streamflow sites and identified that only 10 sites showed 'poor' performance (Fig. 4-2). While there are several streamflow gauges in the Onkaparinga catchment, their records are shorter and complicated by inter-catchment pipeline transfers (Fig. 4-2) and nonstationarities (Westra et al, 2014) making observed-streamflow evaluation difficult. Furthermore, the process of comparing to observed streamflow requires estimates of the catchment rainfall, which is constructed from multiple rainfall sites and obscures the ability to evaluate rainfall at individual sites.

A potential limitation of the virtual hydrologic evaluation framework is that it is reliant on the use of a hydrological model – which is common to both the observed

and simulated rainfall. There is the potential for hydrological structural errors to skew interpretation of the rainfall model evaluation if the hydrological model poorly represents the catchment processes. To reduce these impacts it is recommended to use a well-tested hydrological model that has demonstrated good performance on a wide range of catchments. In this study GR4J was used, which has been tested on hundreds of catchment in Europe and Australia (Coron et al., 2012; Perrin et al., 2003). It is further recommended to use a hydrological model that has been calibrated and evaluated using observed data located close to the observed rainfall sites. For example, in this study the GR4J model was calibrated to the Onkaparinga catchment (Westra et al., 2014) – see Fig. 4-2. Even where a well-tested hydrological model is calibrated, there is the remaining question of whether or not the conclusion of a virtual evaluation is dependent on the choice of the hydrological model. To address this limitation, multiple hydrological models could be used as part of the virtual framework. As this is the first proof-of-concept study that has introduced the virtual hydrologic evaluation framework, the use of multiple hydrological models is outside the scope of this paper and is left for future research.

4.5.3 Splice Testing: A targeted evaluation of rainfall model performance

Previous discussion has highlighted the benefits of the streamflow-based and virtual elements of an evaluation framework of rainfall models when compared to existing evaluation methods (observed-rainfall evaluation and observed-streamflow evaluation). However, these innovations alone lack the ability to target which periods of the rainfall model produce poor streamflow simulations. For example, it would be possible to identify a discrepancy in July for Site 10, but it would be impossible to determine whether this was due to a deficiency in rainfall in July, or a prior month such as May or June (Fig. 4-4 (g) and (h)).

The use of a virtual hydrological framework for evaluation provides the unique opportunity to develop innovative tests that can target specific aspects of the stochastic rainfall model. A splice test was introduced as a method for isolating the influence of rainfall in a month (i.e. the influencing month) on streamflow in an evaluated month while excluding the possibility of deficiencies from other rainfall months. The test enables a procedure for targeting months that are influential in terms of streamflow production rather than interpret model performance based on

blunt evaluation of rainfall or streamflow.

Without the splice test, identifying which months have deficiencies in the modelled rainfall that produce poor streamflow predictions would not have been possible. If the modeller, had focussed on improving the rainfall model by focusing on months with the highest contribution to annual total flow, June to September would have been identified as important, whereas the splice test identifies a different focus (May-August). For example, May and June combined contribute 13% in the annual total flow (Fig. 4-5 (c)), compared to 11% error in the mean (Fig. 4-5 (g)) and 24% error in the standard deviation (Fig. 4-5 (h)). By contrast, September is a high flow month contributing 21% of the annual total flow, but only 2% error in the mean and 6% error in the standard deviation. Without the splice test, it would have been less clear that the 'wetting-up' months such as May and June were a more important focus for rainfall model improvement than a high-flow month such as September.

Although it is possible to isolate which months were the main contributors to errors in the annual total flow volume, questions remain surrounding the influence of longer term wet-dry patterns, month-to-month correlations, and inter-annual correlations on annual flow. Future extensions of the framework will be developed to address these questions.

4.6 Conclusions

This paper has introduced the virtual hydrologic evaluation framework to overcome the problems of existing evaluation methods which are not able to identify which rainfall characteristics are important for stream flow prediction. The framework has two key innovations, an integrated test and a splice test. The integrated and splice tests enabled different conclusions to be reached in terms of priorities for improving the rainfall model. These conclusions would not otherwise have been possible with traditional evaluation methods that focus either on rainfall statistics, or on high streamflow months. The integrated test demonstrated that while large discrepancies were identified in low rainfall months these did not translate to deficiencies in streamflow due to the dry state of the catchment. The test also indicated instances where modelled rainfall categorised as 'good' translated to 'poor' flow due to the influence of catchment memory and rainfall from prior months. The splice test identified the importance of transition months May and June (late autumn/early

winter) in the ‘wetting-up’ phase of the catchment cycle. By contrast, a traditional approach focusing on high flows would have emphasised the importance of correctly simulating high flow months whilst a prior observed-rainfall evaluation (Bennett et al., 2016) emphasised the importance of persistence in the low-rainfall months. The virtual hydrologic evaluation provides insights not available through traditional approaches and should be an essential step in the development and application of stochastic rainfall models.

4.7 Acknowledgements

This work was supported by an Australian Research Council Discovery grant: A new flood design methodology for a variable and changing climate DP1094796. Additional support was provided by the CSIRO Climate Adaptation Flagship.

4.8 References

Andreassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., Lavabre, J., 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. J Hydrol, 250(1-4): 206-223.

Ang, A., Tang, W., 2007. Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering (v. 1), 52 pp.

Ball, J.E., 1994. The influence of storm temporal patterns on catchment response. J. Hydrol., 158(3-4): 285-303. DOI:Doi: 10.1016/0022-1694(94)90058-2

Baxevani, A., Lennartsson, J., 2015. A spatiotemporal precipitation generator based on a censored latent Gaussian field. Water Resources Research.

Bennett, B., Thyer, M., Leonard, M., Lambert, M., Bates, B., 2016. Comprehensive evaluation of a latent variable approach to continuously simulate daily rainfall fields J. Hydrol., currently under review.

Blazkova, S., Beven, K., 2002. Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty). Water Resour Res, 38(8): 14-1-14-14. DOI:10.1029/2001WR000500

Blazkova, S., Beven, K., 2009. A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. Water Resources Research, 45(12): W00B16.

Camici, S., Tarpanelli, A., Brocca, L., Melone, F., Moramarco, T., 2011. Design soil moisture estimation by comparing continuous and storm-based rainfall-runoff modeling. Water Resour. Res., 47(5): W05527. DOI:10.1029/2010wr009298

Chiew, F.H., 2006. Estimation of rainfall elasticity of streamflow in Australia. Hydrological Sciences Journal, 51(4): 613-625.

Clark, M. et al., 2008. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resour. Res., 44. DOI:10.1029/2007wr006735

Coron, L. et al., 2012. Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. Water Resources Research, 48(5).

Cowpertwait, P.S.P., 2006. A spatial-temporal point process model of rainfall for the Thames catchment, UK. J. Hydrol., 330(3-4): 586-595.

Evin, G., Thyer, M., Kavetski, D., McInerney, D., Kuczera, G., 2014. Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. Water Resources Research, 50(3): 2350-2375.

Henley, B.J., Thyer, M.A., Kuczera, G., 2013. Climate driver informed short- term drought risk evaluation. Water Resour Res, 49(5): 2317-2326. DOI:Doi 10.1002/Wrcr.20222

Kuczera, G., Williams, B.J., 1992. Effect of rainfall errors on accuracy of design flood estimates. Water Resources Research,, 28(4): 1145-1154.

Li, J., Thyer, M., Lambert, M., Kuczera, G., Metcalfe, A., 2014. An efficient causative event-based approach for deriving the annual flood frequency distribution. J Hydrol, 510: 412-423. DOI:DOI 10.1016/j.jhydrol.2013.12.035

Li, J., Thyer, M., Lambert, M., Kuzera, G., Metcalfe, A., 2016. Incorporating seasonality into event-based joint probability methods for predicting flood frequency: A hybrid causative event approach. J Hydrol, 533: 40-52. DOI:10.1016/j.jhydrol.2015.11.038

Li, Z., Lü, Z., Li, J., Shi, X., 2015b. Links between the spatial structure of weather generator and hydrological modeling. Theoretical and Applied Climatology: 1-9. DOI:10.1007/s00704-015-1691-8

McMillan, H., Freer, J., Pappenberger, F., Krueger, T., Clark, M., 2010. Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. Hydrological Processes, 24(10): 1270-1284. DOI:10.1002/hyp.7587

McMillan, H.K., Brasington, J., 2008. End-to-end flood risk assessment: A coupled model cascade with uncertainty estimation. Water Resources Research, 44(3): W03419. DOI:10.1029/2007WR005995

Mortazavi-Naeini, M. et al., 2015. Robust optimization to secure urban bulk water supply against extreme drought and uncertain climate change. Environmental Modelling & Software, 69: 437-451. DOI:http://dx.doi.org/10.1016/j.envsoft.2015.02.021

Nicótina, L., Alessi Celegon, E., Rinaldo, A., Marani, M., 2008. On the impact of rainfall patterns on the hydrologic response. Water Resources Research, 44(12). DOI:10.1029/2007WR006654

Paschalis, A., Molnar, P., Fatichi, S., Burlando, P., 2013. A stochastic model for high-resolution space-time precipitation simulation. Water Resources Research, 49(12): 8400-8417.

Paton, F.L., Maier, H.R., Dandy, G.C., 2013. Relative magnitudes of sources of uncertainty in assessing climate change impacts on water supply security for the southern Adelaide water supply system. Water Resour Res, 49(3): 1643-1667. DOI:10.1002/wrcr.20153

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol., 279(1): 275-289.

Rasmussen, P., 2013. Multisite precipitation generation using a latent autoregressive model. Water Resources Research, 49(4): 1845-1857.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resour. Res., 46(5): W05521. DOI:10.1029/2009wr008328

Renard, B. et al., 2011. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. Water Resour. Res., 47(11): W11516. DOI:http://dx.doi.org/10.1029/2011wr010643

Shah, S.M.S., O'Connell, P.E., Hosking, J.R.M., 1996. Modelling the effects of spatial variability in rainfall on catchment response. 2. Experiments with distributed and lumped models. J. Hydrol., 175(1-4): 89-111. DOI:10.1016/s0022-1694(96)80007-2

Smith, P.J., Beven, K.J., Tawn, J.A., 2008. Detection of structural inadequacy in process-based hydrological models: A particle-filtering approach. Water Resour Res, 44(1): W09403. DOI:10.1029/2006wr005205

Srikanthan, R., Pegram, G.G.S., 2009. A nested multisite daily rainfall stochastic generation model. J. Hydrol., 371(1-4): 142-153. DOI:10.1016/j.jhydrol.2009.03.025

Westra, S., Thyer, M., Leonard, M., Kavetski, D., Lambert, M., 2014. Impacts of climate change on surface water in the Onkaparinga catchment-Final report volume 1: hydrological model development and sources of uncertainty. 1839-2725.

Wilks, D.S., 2008. High-resolution spatial interpolation of weather generator parameters using local weighted regressions. Agricultural and Forest Meteorology, 148(1): 111-120.

Wilson, C.B., Valdes, J.B., Rodriguez-Iturbe, I., 1979. On the influence of the spatial distribution of rainfall on storm runoff. Water Resources Research, 15(2): 321-328. DOI:10.1029/WR015i002p00321

Chapter 5

5.1 Research contribution

This thesis has concentrated on developing methods for improving the representation of spatial features in rainfall modelling and hydrological design. These methods have only recently become feasible, since spatial rainfall modelling relies on advances in computational power, data availability and scientific methods. To date, hydrological design has primarily focussed on the time domain, such as the sequence of storm events, the wet-dry pattern of intermittent rain, the cycle of seasonal totals and the variability of inter-annual totals. Future hydrological design will increasingly allow for both spatial and temporal rainfall features and this thesis provides a number of contributions towards that goal.

Objective 1 – Direct spatial rainfall: The current approach for extreme spatial rainfall estimation is inconsistent as it relies on pointwise estimates of extreme rainfall, which differ from the required areal extreme rainfall. Real storm events are not equally extreme at all locations. To account for this limitation the traditional approach uses a statistical correction factor, referred to as an areal reduction factor, to approximate the extreme spatial rainfall intensity. This factor has no physical basis and does not directly relate to real storm events. It is instead a statistical relationship to match point and areal rainfall frequencies for a given duration.

The IFDA approach developed in the first paper simplifies the procedure for estimating and using extreme spatial rainfall in hydrological design. With the advent of gridded rainfall products it is possible to directly estimate the volume of rainfall on a catchment and its corresponding frequency in the form of an IFDA. The approach avoids assumptions of homogeneity that are introduced by areal reduction factors, therefore making it more flexible and less biased. The approach is generic and can be applied to any gridded rainfall dataset, and can also be constructed for stochastic rainfall data sources. The method directly estimates the spatial extreme rainfall at a location for any event and assigns a frequency. For this reason it can also be used as a metric to evaluate spatial features of a rainfall model.

Objective 2 – Stochastic spatial rainfall model: Continuous simulation is important for many applications such as droughts and floods because it allows for the interaction between long-term sequences of rainfall and catchment conditions. Most continuous rainfall models simulate at a single point, while most spatial rainfall models are for individual events. For these reasons, a new spatial rainfall model was developed that is capable of simulating rainfall fields over a region, at the daily timescale, and for long periods of time. The latent-variable framework was adopted because it allows for complex patterns of spatial rainfall to be generated while retaining the benefit of parsimony. The new model relies on the physically based assumption that the spatial correlation should be a smoothly varying function in space. This assumption makes the model much simpler than equivalent multisite models (Rasmussen, 2013), but along with all model assumptions, requires a rigorous evaluation of performance.

The observed-rainfall evaluation method applied to the spatial rainfall model is one of the most rigorous presented in literature. While past studies have made important advances in spatial rainfall modelling there has been a tendency to focus on presenting model developments over formal evaluation. To improve rainfall models it is necessary to also improve the methods used to evaluate them. To this end a comprehensive rainfall evaluation approach was developed. The key features of the approach were the use of a systematic approach and a classification system which is specified a priori. These features are seldom present in other rainfall evaluations and the paper serves as a demonstration of the approach's efficacy. For example, it was possible to objectively state that the variability in annual total rainfall was under predicted and this was due to difficulties in simulating variability in the number of wet days. The framework also emphasised the importance of cross-validation in comprehensive evaluation.

Objective 3 – Hydrological evaluation: The theme of evaluation from the second paper is expanded upon in the third paper. An entirely new approach was developed which evaluated the performance of a stochastic rainfall model against resulting streamflow, which is often the intended goal of rainfall modelling. In addition to the requirement for streamflow evaluation, the key innovations of the framework are (i) the virtual nature of the evaluation, which allows for flexible evaluation of the

rainfall model and removes confounding errors when using observed streamflow; and (ii) the targeted nature of the assessment, which isolates sources of error in the rainfall model according to sites and months.

These innovations allow the framework to explicitly identify where errors in modelled streamflow arise from within the simulated rainfall. The framework provides insight for model development not otherwise available via approaches which focus on observed-rainfall evaluation or on the performance in high flow-volume months.

5.2 Limitations

Three limitation themes occur throughout this thesis, which are the assumption of spatial homogeneity and issues surrounding data dependency and model dependency. Limitations are discussed with respect to each objective:

Objective 1 – Direct spatial rainfall: Spatial homogeneity is a typical assumption in rainfall modelling. It is often a simplifying assumption and allows for the regionalisation of approaches and parameters (Hosking and Wallis., 1997). However, many regions and catchments are spatially heterogeneous and it is important that spatial methods can accommodate this reality.

Currently, the IFDA approach allows for some heterogeneity. It is location specific and allows the relationship between point and areal rainfall to be spatially heterogeneous within a region. However, the IFDA is derived from spatially interpolated rainfall grids. Therefore there is the potential for the interpolation algorithm to introduce artefacts that obscure or emphasise the spatial heterogeneity/homogeneity of the IFDA relationship.

The IFDA approach relied on data at the daily timescale. This is both a strength and a weakness. It is a strength because daily rainfall data is relatively abundant and has long record lengths. However, it is a weakness because sub-daily variation and small-scale spatial variability of rainfall is neglected. Modelling spatial variability at shorter time and spatial scales is a key step in developing a better event-scale understanding of catchment dynamics. For example, the daily time step is an insufficient temporal resolution for applications involving flash floods or urban catchments. The methods would require extension to the sub-daily timescale for use

in these cases.

The IFDA approach is reliant on gridded spatial rainfall produced by an interpolation algorithm. The IFDA approach was presented relying on the spatially gridded dataset, AWAP (Raupach et al., 2009; Raupach et al., 2012). AWAP relies on daily point rainfall gauges, the density of which changes with time and leads to non-stationarity in the standard deviation and smoothing of extreme rainfall intensities (King et al., 2013). Although the IFDA approach was repeated for different time spans with different gauge densities to check the generality of conclusions further testing with alternate interpolated data sources would be required to test for issues of model artefacts.

Objective 2 – Stochastic spatial rainfall model: The latent variable spatial rainfall model relies on spatial interpolation (universal kriging) to estimate parameter surfaces for the simulated region. It also assumes a regionalised AR-1 parameter. Due to the small size of the simulated region, the use of a regionalised AR-1 parameter and the use of universal kriging to provide parameter surfaces was an adequate approach. However, different and larger regions may have more complex spatial relationships that challenge the use of linear relationships for the parameter surfaces and that challenge the homogeneity requirement of the AR-1 parameter.

The stochastic spatial rainfall model relies on and simulates daily rainfall sequences. The daily scale of the model precludes it from application to sub-daily problems, such as urban flooding. Additionally, the reliance on point rainfall gauges means that the spatial rainfall features are relatively smooth compared to those observed from weather radar.

Objective 3 – Hydrological evaluation: The main limitation of the virtual hydrologic evaluation framework, as presented in **Chapter 4**, is that the framework was demonstrated using a single hydrological model, GR4J. The framework is reliant on the streamflow simulated via the selected hydrological model, therefore it is possible that conclusions are model dependent and may vary according to the hydrological model. Another limitation of the approach is that while it can isolate sites and months, it cannot attribute streamflow deficiencies to specific features of the rainfall (such as the wet-dry pattern, the correlation structure or the skewness).

5.3 Future work

Aforementioned limitations of the research present opportunities for future development:

- *Extension to sub-daily time scale:* Extension of the three approaches to sub-daily timescales would greatly widen their range of potential applications. Extension of the IFDA approach (Objective 1) would require a source of sub-daily gridded rainfall but minimal adaption of the generic IFDA framework. Similarly, the virtual hydrological evaluation approach (Objective 3) is easily extendable to the sub-daily setting by using a sub-daily hydrological model. The extension of the stochastic spatial rainfall model (Objective 2) is more challenging, because it would require the incorporation of sub-daily data into model calibration whilst retaining good predictive performance at the daily scale. This extension may require the use of a hierarchical simulation framework to account for the growth, decay and sequencing of storm events.

- *Improve simulated spatial rainfall features:* To improve the representation of spatial rainfall features simulated by the stochastic spatial rainfall model (Objective 2) spatial data sources could be used to evaluate and calibrate the model. For example, the stochastic rainfall model's spatial features might be improved by comparing against radar or satellite observed spatial rainfall patterns.

- *Extension of virtual hydrological evaluation framework:* The virtual hydrological evaluation framework (Objective 3) can be extended to become a more comprehensive diagnostic tool. For example, it might considering internal hydrological model states or use distributed hydrological models to allow for comparisons against multiple internal virtual-observed streamflow sequences. The framework can also be augmented through the development of better methods to target specific relationships between the evaluated streamflow and the evaluated stochastic rainfall model's configuration.

- *Demonstration of approaches using diverse case studies:* The rainfall model and evaluation techniques would benefit from demonstration across further case studies covering a diversity of rainfall regimes, catchment conditions

and hydrological models. Similarly the evaluation techniques can be used to compare the performance of different stochastic rainfall models for use in a given application.

5.4 Final recommendations

This body of work prompts the following recommendations for future hydrological modelling.

- The demonstrated bias in current ARF-based estimates of spatial extreme rainfall support the recommendation that spatial extreme rainfall should be directly estimated via the IFDA approach (*Paper 1*).

- An approach of minimum necessary complexity should be adopted with respect to rainfall models. It is likely that spatial rainfall models are more parsimonious and reliable under cross-validation than multisite rainfall models (*Paper 2*).

- Evaluation that is comprehensive and systematic (*Paper 2*) as well as using a virtual hydrologic framework (*Paper 3*) should form an essential part of all stochastic rainfall model development, refinement and application.

References

Bardossy, A., Plate, E.J., 1992. Space-time model for daily rainfall using atmospheric circulation patterns. Water Resources Research, 28(5): 1247-1259. DOI:10.1029/91wr02589

Bennett, B., Lambert, M., Thyer, M., Bates, B.C., Leonard, M., 2016a. Estimating Extreme Spatial Rainfall Intensities. Journal of Hydrologic Engineering, 21(3): 04015074. DOI:10.1061/(ASCE)HE.1943-5584.0001316

Bennett, B., Thyer, M., Leonard, M., Lambert, M., Bates, B., 2016b. Comprehensive evaluation of a latent variable approach to continuously simulate daily rainfall fields J. Hydrol., currently under review.

Bennett, B., Thyer, M., Leonard, M., Lambert, M., Bates, B., 2016c. A Virtual Hydrological Framework to Evaluate Stochastic Rainfall Models. Water Resources Research, submitted for review.

Brocca, L., Camici, S., Tarpanelli, A., Melone, F., Moramarco, T., 2011. Analysis of Climate Change Effects on Floods Frequency Through a Continuous Hydrological Modelling. In: Baba, A. et al. (Eds.), Climate Change and Its Effects on Water Resources: Issues of National and Global Security. NATO Science for Peace and Security Series C-Environmental Security. Springer, Dordrecht, pp. 97-104. DOI:10.1007/978-94-007-1143-3_11

Davison, A.C., Padoan, S.A., Ribatet, M., 2012. Statistical Modeling of Spatial Extremes. 161-186. DOI:10.1214/11-STS376

Faures, J.-M., Goodrich, D.C., Woolhiser, D.A., Sorooshian, S., 1995. Impact of small-scale spatial rainfall variability on runoff modeling. J. Hydrol., 173(1-4): 309-326. DOI:10.1016/0022-1694(95)02704-s

Frost, A.J., Thyer, M.A., Srikanthan, R., Kuczera, G., 2007. A general Bayesian framework for calibrating and evaluating stochastic models of annual multi-site hydrological data. J. Hydrol., 340(3–4): 129-148. DOI:http://dx.doi.org/10.1016/j.jhydrol.2007.03.023

Groppelli, B., Bocchiola, D., Rosso, R., 2011. Spatial downscaling of precipitation from GCMs for climate change projections using random cascades: A case study in Italy. Water Resources Research, 47(3): W03519. DOI:10.1029/2010WR009437

Haberlandt, U., 2007. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. J. Hydrol., 332(1-2): 144-157. DOI:10.1016/j.jhydrol.2006.06.028

Heneker, T.M., Lambert, M.F., Kuczera, G., 2001. A point rainfall model for risk-based design. J. Hydrol., 247(1-2): 54-71. DOI:Doi: 10.1016/s0022-1694(01)00361-4

Hosking, J.R.M., Wallis., J.R., 1997. Regional Frequency Analysis. Cambridge University Press.

Hutchinson, M.F., 1995. Interpolating mean rainfall using thin plate smoothing splines. International Journal of Geographical Information Systems, 9(4): 385-403. DOI:10.1080/02693799508902045

Kim, S., Tachikawa, Y., Sayama, T., Takara, K., 2009. Ensemble flood forecasting with stochastic radar image extrapolation and a distributed hydrologic model. Hydrological Processes, 23(4): 597-611. DOI:10.1002/hyp.7188

King, A.D., Alexander, L.V., Donat, M.G., 2013. The efficacy of using gridded data to examine extreme rainfall characteristics: a case study for Australia. International Journal of Climatology, 33(10): 2376-2387.

Lebel, T., Laborde, J.P., 1988. A geostatistical approach for areal rainfall statistics assessment. Stochastic Hydrology and Hydraulics, 2(4): 245-261.

Leonard, M., Lambert, M.F., Metcalfe, A.V., Cowpertwait, P.S.P., 2008. A space-time Neyman-Scott rainfall model with defined storm extent. Water Resources Research, 44(9): W09402.

Li, J., Thyer, M., Lambert, M., Kuczera, G., Metcalfe, A., 2014. An efficient causative event-based approach for deriving the annual flood frequency distribution. J. Hydrol., 510: 412-423.

McMahon, T.A., Finlayson, B.L., 2003. Droughts and anti-droughts: the low flow hydrology of Australian rivers. Freshwater Biology, 48(7): 1147-1160. DOI:10.1046/j.1365-2427.2003.01098.x

Northrop, P., 1998. A clustered spatial-temporal model of rainfall. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 454(1975): 1875-1888. DOI:10.1098/rspa.1998.0238

Onof, C., Wheater, H.S., 1993. Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulse Model. J. Hydrol., 149(1): 67-95. DOI:http://dx.doi.org/10.1016/0022-1694(93)90100-N

Pegram, G.G.S., Clothier, A.N., 2001. High resolution space-time modelling of rainfall: the "String of Beads" model. J. Hydrol., 241(1-2): 26-41. DOI:10.1016/s0022-1694(00)00373-5

Qin, J., 2010. A High-Resolution Hierarchical Model for Space-time Rainfall. PhD Thesis, University of Newcastle, Newcastle.

Rasmussen, P., 2013. Multisite precipitation generation using a latent autoregressive model. Water Resources Research, 49(4): 1845-1857.

Raupach, M. et al., 2009. Australian Water Availability Project (AWAP): CSIRO Marine and Atmospheric, Research Component: Final Report for Phase 3.

Raupach, M. et al., 2012. Australian Water Availability Project. In: Research, C.M.a.A. (Ed.), Canberra, Australia.

Rodriguez-Iturbe, I., Cox, D.R., Isham, V., 1988. A Point Process Model for Rainfall: Further Developments. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 417(1853): 283-298. DOI:10.1098/rspa.1988.0061

Sanso, B., Guenni, L., 2000. A nonstationary multisite model for rainfall. Journal of the American Statistical Association, 95(452): 1089-1100. DOI:10.2307/2669745

Seed, A.W., Pierce, C.E., Norman, K., 2013. Formulation and evaluation of a scale decomposition-based stochastic precipitation nowcast scheme. Water Resources Research, 49(10): 6624-6641. DOI:10.1002/wrcr.20536

Seed, A.W., Srikanthan, R., Menabde, M., 1999. A space and time model for design storm rainfall. Journal of Geophysical Research: Atmospheres, 104(D24): 31623-31630. DOI:10.1029/1999JD900767

Segond, M.-L., Wheater, H.s., Onof, C., 2007. The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment, UK. J. Hydrol., 347: 116-131.

Srikanthan, R., Pegram, G.G.S., 2009. A nested multisite daily rainfall stochastic generation model. J. Hydrol., 371(1-4): 142-153. DOI:10.1016/j.jhydrol.2009.03.025

Thiessen, A.H., 1911. Precipitation averages for large areas. Monthly weather review, 39(7): 1082-1089.

Wilks, D.S., 1998. Multisite generalization of a daily stochastic precipitation generation model. J. Hydrol., 210(1–4): 178-191. DOI:http://dx.doi.org/10.1016/S0022-1694(98)00186-3

Zhang, Z., Switzer, P., 2007. Stochastic space-time regional rainfall modeling adapted to historical rain gauge data. Water Resources Research, 43(3): W03441. DOI:10.1029/2005WR004654

Appendix A

Copy of paper from Chapter 2.

Bennett, B., Lambert, M., Thyer, M., Bates, B.C., Leonard, M., 2016. Estimating Extreme Spatial Rainfall Intensities. Journal of Hydrologic Engineering, 21(3): 04015074. DOI: 10.1061/(ASCE)HE.1943-5584.0001316

Estimating Extreme Spatial Rainfall Intensities

Bree Bennett[1]; Martin Lambert, A.M.ASCE[2]; Mark Thyer[3]; Bryson C. Bates[4]; and Michael Leonard[5]

Abstract: Determining the impact of catchment flooding requires an estimate of extreme spatial rainfall intensity. Current flood design practice typically converts a point estimate of rainfall intensity into a spatial rainfall intensity using an areal reduction factor, assumed constant across an entire region. Areal reduction factors do not explicitly consider regional variations in extreme rainfall. Here, a new approach for spatial estimates of extreme rainfall is introduced that directly incorporates the spatial area (A) into an intensity-frequency-duration relationship (IFD). This IFDA approach uses spatial rainfall fields to overcome shortcomings of the areal reduction factor by explicitly incorporating spatial variations in the extreme rainfall intensity. The IFDA approach is evaluated for 11 case study regions in Australia, across climates (tropical to Mediterranean), areas (25–7,225 km^2), durations (1–4 days), and average recurrence intervals (ARI 2–100 years). The change in extreme spatial rainfall with respect to area varies markedly within each region suggesting that constant areal reduction factors for a region are inappropriate. Constant areal reduction factors are shown to underestimate extreme spatial rainfall intensities by 5–15%. The IFDA approach avoids these biases and is a promising new technique for use in design flood estimation. **DOI: 10.1061/(ASCE)HE.1943-5584.0001316.**

Introduction

Of all natural disasters, floods have the highest global cost and affect the most people (Kousky and Walls 2014; Miller et al. 2008; Strömberg 2007). A key input for estimating flood risk is the spatial intensity of extreme rainfall events over a catchment. Current techniques for estimating extreme spatial rainfall rely on the use of an areal reduction factor (ARF) to convert intensity estimates of extreme point rainfall to extreme spatial rainfall. It is common practice to ignore the spatial variation in rainfall intensity and assume a fixed ARF applies over large regions. The aim of this paper is to introduce a new approach that explicitly incorporates the area and variation of spatial rainfall, referred to as intensity-frequency-duration area (IFDA).

The IFDA approach uses spatially interpolated rainfall grids to directly provide an estimate of how spatial rainfall intensities vary with duration, frequency, and area for a location. An IFDA adds the extra dimension of area (A) to an IFD curve to account for spatial variation in intensity over a catchment.

Existing design methods rely on interpolated maps of intensity-frequency-duration (IFD) rainfall calculated from point rainfall data. The ARF is then used to determine the areal average rainfall intensity from the point rainfall. The factor is defined as the ratio of extreme rainfall at a point to the extreme rainfall over an area for a given frequency (Asquith and Famiglietti 2000). In brief, the ARF is a spatial correction factor used to fix limitations of a design methodology focused around pointwise rainfall estimates (IFDs).

The validity of assuming a fixed ARF for a region has been previously questioned (Catchlove and Ball 2003; Durrans et al. 2002). However, these evaluations of ARF spatial variation have typically been limited to a single location (Catchlove and Ball 2003) or climatic region (Durrans et al. 2002). In contrast, this study will evaluate these impacts across multiple climate regions.

This paper proposes that estimates of extreme spatial rainfall are better and more efficiently obtained by directly using gridded spatial rainfall in preference to scaling pointwise extreme rainfall. IFDAs, as direct spatial rainfall estimates, overcome the assumption that a single scaling relationship is sufficient for a region.

The key objectives of the paper are:

1. To demonstrate the utility of IFDAs and evaluate their spatial variation for 11 study regions across a range of climates within Australia.
2. To evaluate the differences in spatial rainfall estimates obtained using the ARF-based approach and the "true" spatial rainfall, the IFDA.
3. To evaluate the characteristics of differences between the IFDA and the ARF-based approach with respect to properties of the extreme spatial rainfall such as area, frequency, duration and region.

IFDA Approach

The method of constructing IFD curves has been modified to directly incorporate spatial extent. In this approach, each IFDA relationship corresponds to a grid point of specific longitude and latitude. For each grid cell, a set of IFDAs are produced for a range of areas and durations. The approach is generic and is able to be applied to any gridded spatial rainfall data source such as spatially interpolated point rainfall (Bárdossy and Pegram 2013), radar rainfall, and downscaled rainfall from climate model simulations.

An IFDA curve is constructed by first designating a grid cell within the subject region [Fig. 1(a)]. This grid cell of specific latitude and longitude is the central point around which the spatial variation in rainfall is considered. A fixed area is designated around the center grid cell [Fig. 1(b)]. This designated area is illustrated as

[1]School of Civil, Environmental and Mining Engineering, Univ. of Adelaide North Terrace Campus, SA 5005, Australia (corresponding author). E-mail: bree.bennett@adelaide.edu.au

[2]School of Civil, Environmental and Mining Engineering, Univ. of Adelaide North Terrace Campus, SA 5005, Australia.

[3]School of Civil, Environmental and Mining Engineering, Univ. of Adelaide North Terrace Campus, SA 5005, Australia.

[4]CSIRO Oceans and Atmosphere Flagship, Underwood Ave., Floreat, WA 6014, Australia.

[5]School of Civil, Environmental and Mining Engineering, Univ. of Adelaide North Terrace Campus, SA 5005, Australia.

Note. This manuscript was submitted on May 19, 2015; approved on September 11, 2015; published online on December 21, 2015. Discussion period open until May 21, 2016; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699.

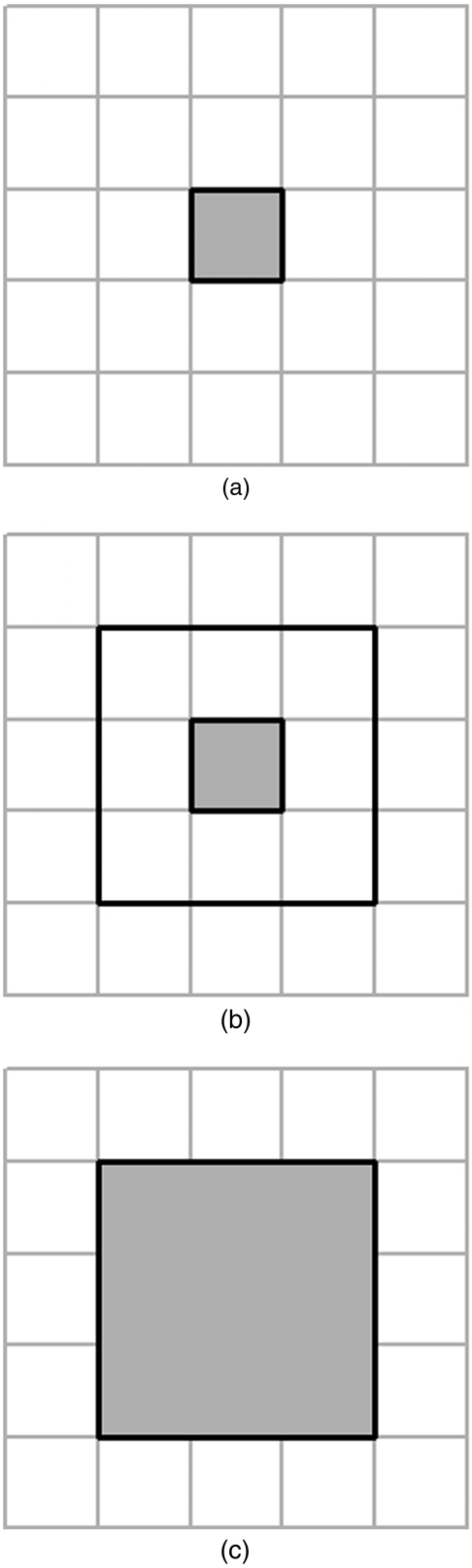


Fig. 1. Schematic of IFDA area designation: (a) centroidal cell identified; (b) area designated around centroidal cell; (c) area consisting of aggregated contributing cells

square in Fig. 1, but the procedure is general and can accommodate areas of any shape.

For each time step the total rainfall in the designated area is summed over the contributing grid cells [Fig. 1(c)]. The increments of spatial rainfall are then summed over the duration and converted to intensity. This results in a time series of spatial rainfall intensities for the designated location, duration, and spatial extent. From this time series the annual maximums are extracted and a generalized extreme value distribution is fitted (Green et al. 2012; Jordan et al. 2011; Siriwardena and Weinmann 1996). The fitted distribution forms the IFDA corresponding to the center grid cell for the designated area and duration.

The IFDA procedure for a single location can be summarized mathematically as follows: Let R_i denote the set of point rainfall intensities for a given duration for all points x in a spatial domain Ω and time increments t in Y_i, the ith year ($i = 1, \ldots, n$), such that $R_i = \{R_i(x,t)|x \in \Omega, t \in Y_i\}$. Let annmax[] be a function that takes the maximum value across all the time increments t in Y_i. For the ith year, the spatial annual maximum rainfall intensity for a catchment domain Ω with area A is defined as

$$C_i(A,t) = \frac{1}{A}\text{annmax}\left[\int_{\Omega} R_i(x,t)dx\right] \quad (1)$$

If the set of extracted events, $C_i = \{C_1, \ldots, C_n\}$ is ordered in terms of magnitude and assigned a frequency, then $C_F(A,t)$ denotes the spatial rainfall intensity for the catchment of area A and frequency F. The rainfall intensities of defined frequency, duration, and area compose the IFDA relationship for that location. The repetition of this process at all locations throughout the domain creates the field of IFDAs and each cell within the region has its own set of IFDA relationships.

Case Study Data

The Australian water availability project (AWAP) gridded rainfall database provides daily rainfall depths on a 5 km square grid across Australia (Raupach et al. 2012). The grids are an interpolated product based on daily point rainfall records and are available from 1900 onwards. This study uses two subsets, 1900–2011 and 1973–2011, where the shorter period provides a check against any changes in gauge density over the period of the longer record.

A conversion factor of 1.15 is applied in this study to account for the restricted time period of the daily observations (e.g., 9 a.m.–9 a.m.). Over restricted periods the maximum is lower than 24 h totals that have been aggregated over an unrestricted time period (Boughton and Jakob 2008; Jakob et al. 2005; van Montfort 1990). For studies that focus on shorter durations (Catchlove and Ball 2003) there is also an implicit need to apply the conversion factor to any daily observations that are used.

The IFDA approach is generic and can be applied to any spatial rainfall data set. For the purposes of this paper the gridded AWAP dataset is regarded as the best estimate of spatial "truth." While spatial interpolation can introduce artifacts (King et al. 2013), the development of better spatial interpolation procedures is a separate (and important) research area. An advantage of the IFDA is that it can easily be updated to take advantage of different or better data products, whereas the current ARF approach cannot (further discussion in "Advantages and Limitations of IFDAs").

Case study regions were selected for detailed analyses based on their gauge density and to ensure coverage of different climates (Tropical: Broome, Cairns, Darwin. Sub-tropical: Brisbane, Sydney. Temperate: Melbourne, Tasmania, Australian Capital Territory (ACT). Mediterranean: Adelaide, Perth. Arid: Alice Springs) (Fig. 2). This paper focuses on the regions of Sydney and Melbourne to illustrate the IFDA relationships. Results from additional regions are included to illustrate that the same general behaviors of

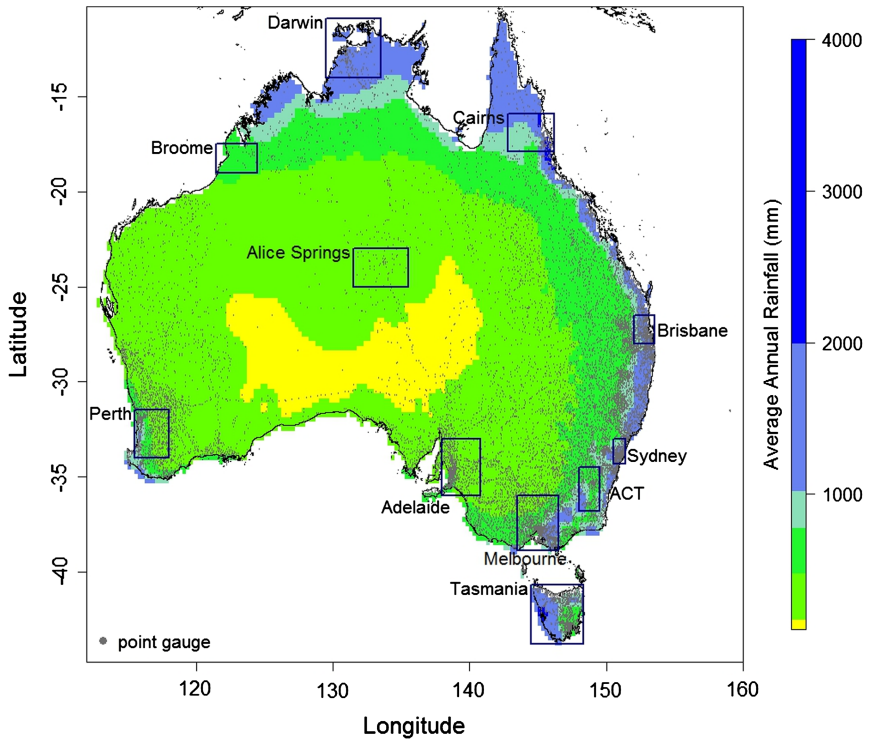


Fig. 2. Location of study regions with annual rainfall isohyets (mm)

the IFDA relationship are exhibited at other locations. The analyses demonstrate the inappropriateness of fixed regional ARFs for estimating extreme spatial rainfall intensities. Melbourne and Sydney have especially dense observation networks, therefore the impact on the results of the AWAP interpolation procedure is minimized.

Methodology

To achieve objective 1, IFDA relationships are derived for all study regions and their behavior is examined to evaluate variations in extreme spatial rainfall. To evaluate the differences between an ARF-based and IFDA approaches (objective 2), the distribution of at-site ARFs calculated using the gridded rainfall are compared with currently recommended fixed regional ARFs (ARF_R). Furthermore, the error induced by using ARF_R is evaluated as a percentage error of the direct estimate of extreme spatial rainfall. The characteristics of the resultant percentage errors are assessed against extreme spatial rainfall properties such as area, frequency, duration, and region (objective 3).

IFDA Calculation

IFDAs were calculated following the method previously mentioned for all study regions and for areas of 25; 225; 625; 1,225; 2,025; 3,025; 5,625; and 7,225 km^2. Durations of 1, 2, 3, and 4 days were investigated and results are presented for average recurrence intervals (ARIs) of 2–100 years. The IFDAs are presented as both location specific relationships and intensity fields. These intensity fields show the rainfall intensity for the specific IFDA area and duration plotted at the center pixel. Therefore, as the area around the center grid increases the boundaries of the region contract to ensure no points outside the original region are sampled in the calculation of intensities for larger areas.

Evaluation of Fixed Area ARFs against the IFDA Approach

Objective 2 is to evaluate the differences in spatial rainfall estimates obtained from the ARF-based approach and the direct estimate of spatial rainfall obtained using the IFDA approach. Therefore, it assesses whether current practice is appropriate for providing robust estimates of extreme spatial rainfall.

Extreme spatial rainfall intensities are traditionally estimated using a representative extreme point rainfall and a regional ARF. The representative point rainfall of a certain frequency is denoted here as R_F, and there are two general approaches for defining it. The first is where the representative extreme point rainfall intensity is taken from a single location, typically the center of the catchment area, $\hat{R}_F^c$. The second is where the spatial average of the extreme point rainfall values within the catchment domain is taken as the representative extreme point rainfall, $\hat{R}_F^{\overline{pt}}$. That is

$$\hat{R}_F^{\overline{pt}}(t) = \frac{1}{A}\int_{\Omega} R_F(x,t)dx \tag{2}$$

This second approach is used in preference to the center extreme point rainfall intensity when there is a strong spatial rainfall trend (e.g., attributable to elevation gradient) within a catchment. In this study, the impact of using both approaches is evaluated.

To evaluate whether the use of fixed regional ARF values, ARF_R, are appropriate for providing robust estimates of extreme spatial rainfall intensities, ARFs were calculated for all locations within each study region. The calculated ARFs are the values required at each sampled location to scale the representative extreme pointwise rainfall to obtain the true extreme spatial rainfall intensity of equivalent ARI. These at-site ARFs were then evaluated against currently recommended ARF_R values.

ARFs are a ratio of spatial rainfall to a representative point rainfall of equal recurrence interval and can be defined as

$$\mathrm{ARF}_{(A,F,t)} = \frac{C_F(A,t)}{\hat{R}_F(t)} \tag{3}$$

where $C_F(A,t)$ is the spatial rainfall intensity of specified area, A, frequency of F, and duration t. This definition is not a storm-centerd ARF that is based on the point and spatial rainfall of an individual storm (Svensson and Jones 2010). Rather it is a statistical ARF that relates representative extreme point rainfall to extreme spatial rainfall with equal recurrence interval (Allen and DeGaetano 2005; Durrans et al. 2002; Myers and Zehr 1980; Sivapalan and Blöschl 1998). ARFs are constructed by using Eq. (3) with a representative extreme point rainfall intensity for the two different cases outlined above. Where the center extreme point rainfall intensity, $\hat{R}_F^c$, was used in Eq. (3), the resulting at-site ARF has been denoted as ARF_c, where the spatially averaged extreme point rainfall, $\hat{R}_F^{\overline{pi}}$, is used the at-site ARF is denoted as $\mathrm{ARF}_{\overline{pi}}$. For this analysis the single grid cell (25 km^2 area) was nominated as the highest resolution approximation of point rainfall.

The at-site ARF_c and $\mathrm{ARF}_{\overline{pi}}$ values are presented as boxplots to demonstrate the range and distribution of calculated ARF values throughout a region, with a comparison against the fixed regional ARF_R. These fixed regional ARF_R include values published in Pilgrim (1987) (following Myers and Zehr 1980), updated values for New South Wales (Jordan et al. 2011), and updated values for Victoria (Siriwardena and Weinmann 1996). The values used for reference in this study are for the areal rainfall intensity at 25 km^2.

To evaluate the bias present in using the ARF-based approach, the percentage error resulting from using this estimate compared against the spatial rainfall total is calculated. The percentage error is calculated as

$$\%\mathrm{Err} = \frac{\hat{C}_F - C_F}{C_F} \times 100 \tag{4}$$

where C_F is the directly obtained spatial rainfall intensity (i.e., the IFDA value at the required location with corresponding frequency

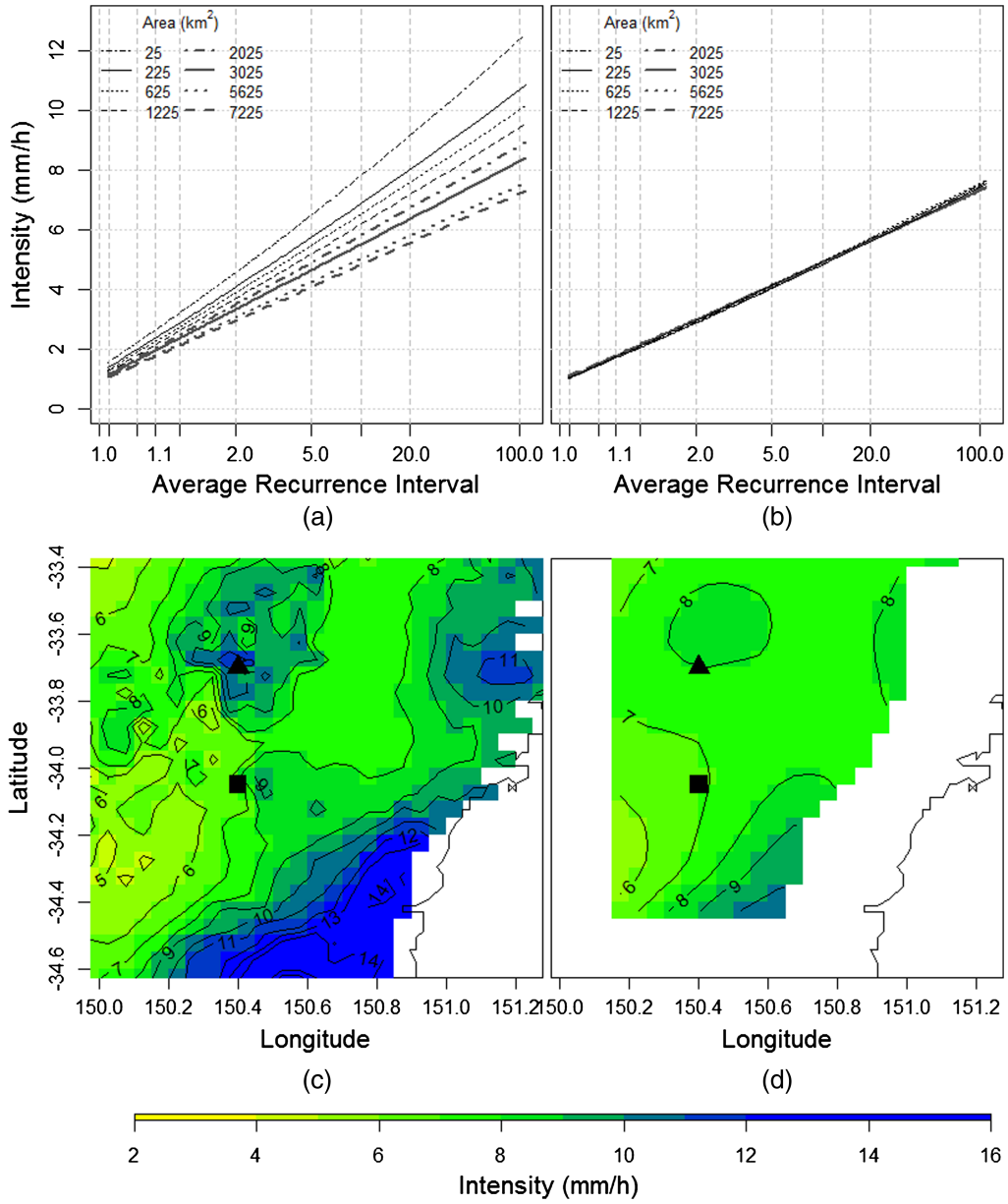


Fig. 3. Sydney region IFDA 1-day: (a) 150.400°E 33.700°S; (b) 150.400°E 34.050°S; (c) IFDA field 25 km^2, 50 year ARI, 1-day; (d) IFDA field 2,025 km^2, 50 year ARI, 1-day

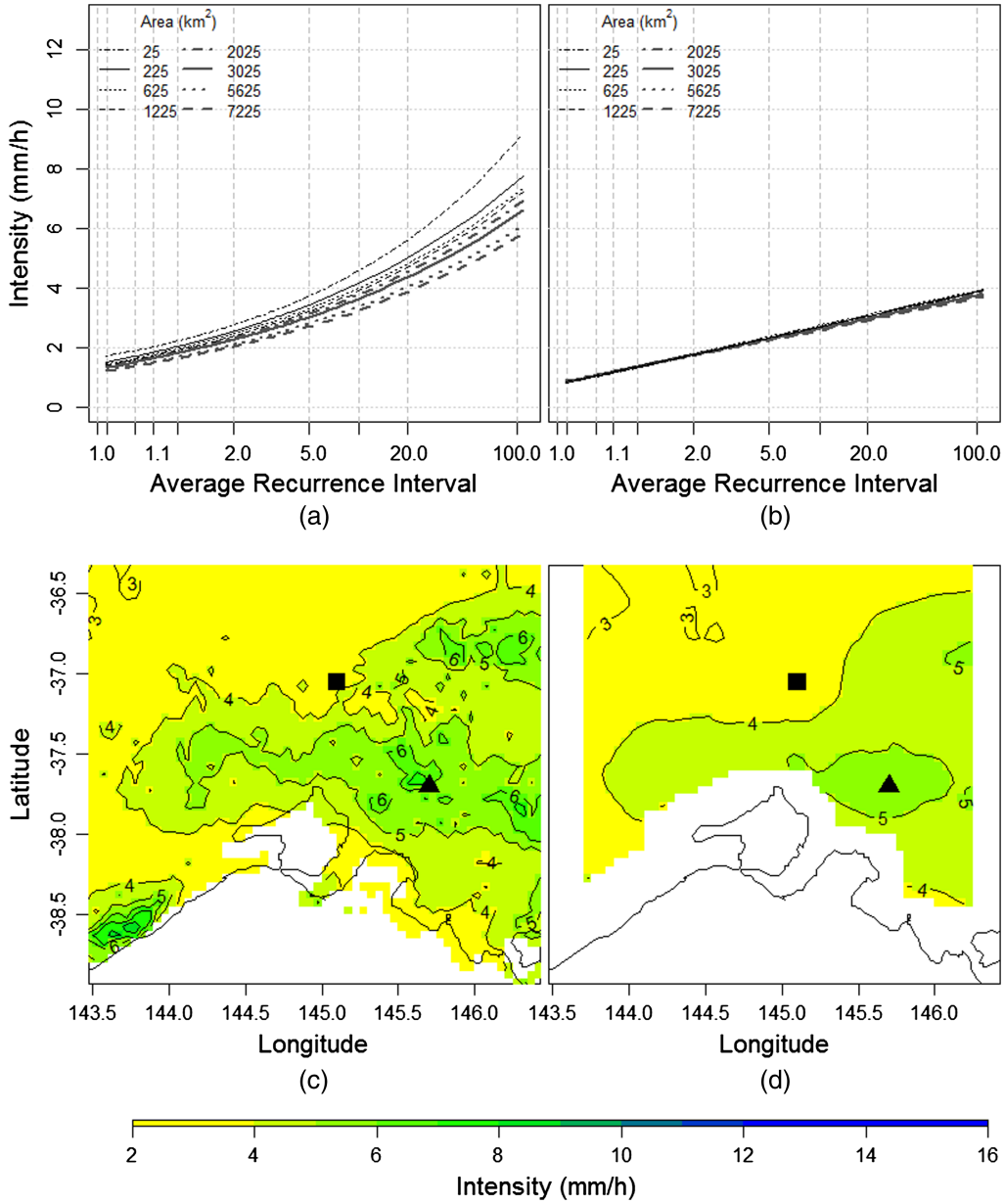


Fig. 4. Perth region IFDA 1-day: (a) 145.700°E 37.700°S; (b) 145.100°E 37.050°S; (c) IFDA field 25 km^2, 50 year ARI, 1-day; (d) IFDA field 3,025 km^2, 50 year ARI, 1-day

F); and $\hat{C}_F$ is the spatial rainfall intensity estimated using a representative extreme point rainfall via Eq. (5)

$$\hat{C}_F = \mathrm{ARF}_F \hat{R}_F(t) \tag{5}$$

For the case $\hat{R}_F^c$, the estimated extreme spatial rainfall and resulting percentage error are denoted as $\hat{C}_F^c$ and $\%\mathrm{Err}_c$, respectively. For the case $\hat{R}_F^{\overline{pt}}$, the estimated extreme spatial rainfall and percentage error are denoted as $\hat{C}_F^{\overline{pt}}$ and $\%\mathrm{Err}_{\overline{pt}}$, respectively.

Identification of Extreme Spatial Rainfall Properties Influencing Errors in the ARF-Based Approach

To evaluate the differences between IFDA and ARF-based approaches four densely gauged regions are considered (Sydney, Brisbane, Melbourne, and Perth). The comparison considers the two cases of $\hat{R}_F^c$ and $\hat{R}_F^{\overline{pt}}$. Specifically the coefficient of variation between the two cases is constructed for ARFs and percentage errors. Changes with respect to properties such as catchment area, frequency, and location are assessed to evaluate which has the largest influence.

Results

IFDA Analysis

At each location within a region the IFDA relationship showed different trends with area. In all regions there were some locations with a large increase in intensity with increasing area, while other locations showed little or no increase in intensity with increasing area. For the Sydney region, with an ARI of 50 years and duration of 1 day, Fig. 3(a) indicates a change of 41% between an area of 25 and 7,225 km^2 whereas Fig. 3(b) illustrates little change in intensity with area. These locations are indicated on Figs. 3(c and d) as triangular and square points. The intensity patterns over the

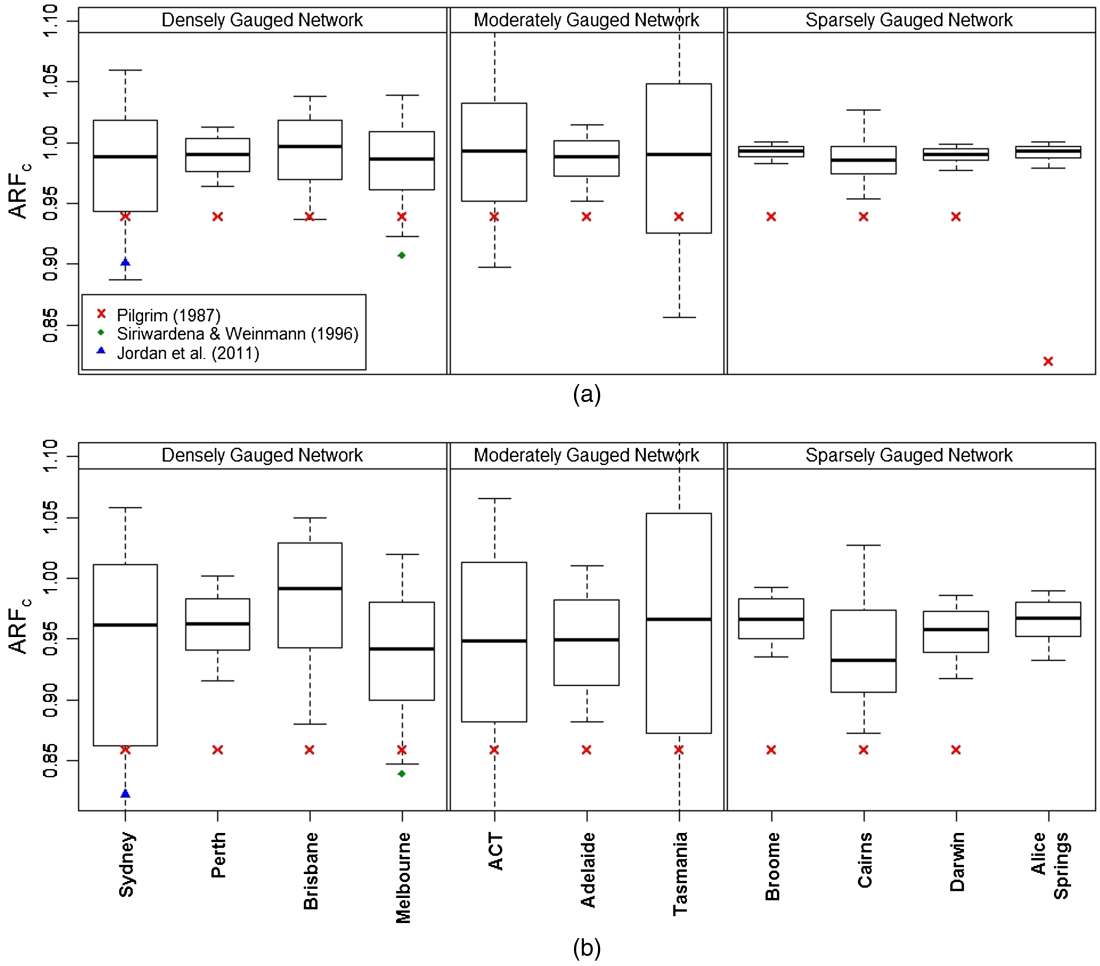


Fig. 5. ARF_c boxplots for 1-day, 50 year ARI rainfall: (a) 625 km^2; (b) 3,025 km^2 (10 and 90% limits shown)

region change as the IFDA area increases. It is this spatial variation in IFDAs that produces the different IFDA curve behaviors exhibited in Figs. 3(a and b).

Fig. 4 similarly shows the change in IFDA with area for the Melbourne region. It similarly illustrates that extreme spatial rainfall intensity patterns change as IFDA areas increase. In Fig. 4 the triangle and square indicate locations at which intensity is highly variable or less variable with area, respectively. Spatial variation in IFDA curves was similarly observed for all other regions (not shown).

The changing spatial patterns of rainfall intensity for different areas can be clearly seen [Figs. 3(c and d), 4(c and d)]. The relationship between intensity and area is spatially heterogeneous and this leads to spatially heterogeneous fields of ARF values.

Evaluation of the Differences between Estimates of Extreme Spatial Rainfall Obtained Using the IFDA and ARF-Based Approaches

Whereas typical flood designs rely on a single ARF value for a given region, there is often considerable variation. To demonstrate this, Fig. 5 summarizes the spatially heterogeneous ARF_c fields as boxplots for all regions, an ARI of 50 years, a 1-day duration, and areas of 625 and 3,025 km^2. The boxplot whiskers extend to the 10 and 90% limits. The derived ARF_c distributions are compared with currently recommended ARF_R values for the Australian regions. The spread of these distributions varies from small to large and is approximately symmetrical. Values greater than 1 occur frequently. An ARF greater than 1 implies that the surrounding spatial rainfall intensity was greater than the individual center point rainfall intensity for that frequency and duration.

Fig. 5 shows that although the regional ARF_R values lie within the range of the derived ARF_c distribution, the fixed values lie below the mean, often significantly so. For example, in Fig. 5(b) the mean ARF_c sits above the ARF_R value of Jordan et al. (2011) by approximately 14% for the Sydney region. The ARF_c distributions exhibit a number of distinctive features. They may be highly variable [Fig. 5(a) Sydney, Tasmania; and Fig. 5(b) Sydney, Melbourne, ACT, Tasmania], indicating that ARF_R may differ significantly from ARF_c values for the majority of locations throughout a region. Alternatively the ARF_c distribution may possess a narrow range of values centering on one [Fig. 5(a) Perth, Broome, Darwin] implying that the ARF_R are biased. Additionally, many ARF_c distributions exhibited values significantly greater than 1 (Fig. 5), indicating that it is common for extreme spatial rainfall surrounding a point to exceed the point rainfall. This behavior was observed for all studied durations.

Fig. 6 presents similar results, using $ARF_{\overline{pi}}$ values. The intention of this alternative is to better account for the spatial variation in point rainfall over the area. However, this modification still leads to significant differences from the recommended ARF_R values. Notably, the $ARF_{\overline{pi}}$ are less variable and have fewer values above 1.

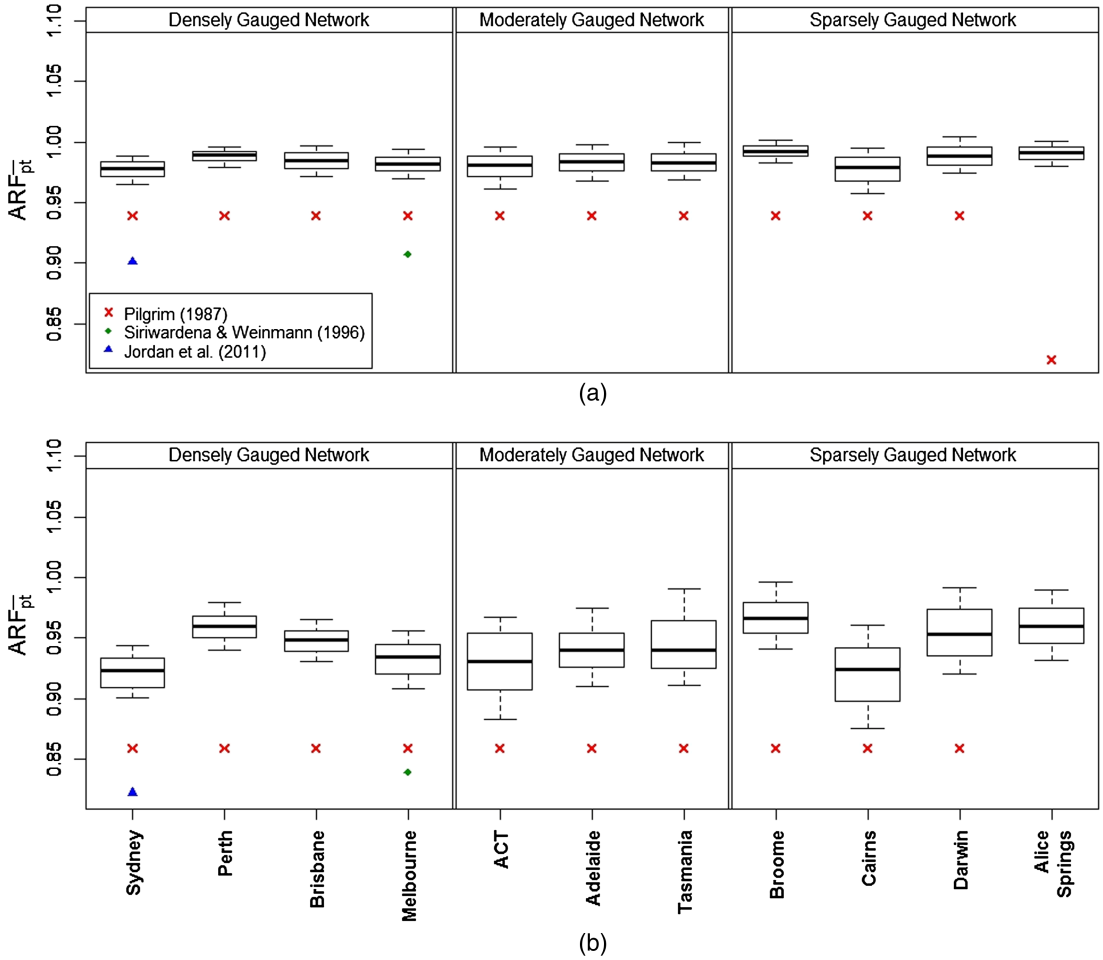


Fig. 6. $ARF_{\overline{pt}}$ boxplots for 1-day, 50 year ARI rainfall: (a) 625 km^2; (b) 3,025 km^2 (10 and 90% limits shown)

Fig. 7 shows percentage errors ($\%Err_c$) resulting from the central point rainfall combined with ARF_R values across areas of 625 and 3,025 km^2. As a result, the extreme spatial rainfall is consistently underestimated for the majority of locations in all regions. Fig. 8 similarly shows the percentage error for the point-averaged rainfall method ($\%Err_{\overline{pt}}$) and this technique underestimates the extreme spatial rainfall for 95–100% of locations in the region. Therefore, the method of averaging point rainfall does not improve ARF-based techniques.

A further comparison was conducted as check for any influence of changing gauge density over the length of the record. Therefore, the analyses were repeated using a shorter 36 year series of daily grids corresponding to the latter third of the record, which has the highest gauge density. Fig. 9 provides an example summary of calculated ARF and percentage error values for the Sydney region for an ARI of 25 years. The analyses did not find any significant effect of changing gauge density on the results.

Evaluation of Extreme Spatial Rainfall Properties Influencing the Differences between the ARF-Based and IFDA Approaches

Given the variability of the ARF values over the region, it is important to understand situations that are most affected. To assess this, the change in the coefficient of variation of the at-site ARFs against catchment area and ARI was evaluated. Fig. 10 shows typical results for both ARF_c and $ARF_{\overline{pt}}$ for four densely gauged regions (Perth, Melbourne, Brisbane, and Sydney). In all cases the coefficient of variation increases with catchment area and $ARF_{\overline{pt}}$ is less variable than ARF_c. This observation was consistent for all ARIs and durations (not shown). This demonstrates that there is greater variability in estimates of extreme spatial rainfall produced using a fixed regional ARF for larger catchment areas.

Fig. 11 compares the mean percentage error $\%Err_c$ against ARI for areas of 625 and 3,025 km^2. There is a bias towards underestimation for all locations and frequencies. For the 3,025 km^2 area [Fig. 11(b)] the errors are larger and the differences between regions are more pronounced, especially for less frequent events. These observations are consistent over other catchment areas, regions, and for $\%Err_{\overline{pt}}$ (not shown).

Discussion and Practical Implications

This paper has demonstrated that current ARF-based approaches for estimating extreme spatial rainfall are biased, underestimating rainfall for the majority of locations. This motivates the need for more direct approaches that do not rely on fixed regional ARFs. Underestimating extreme spatial rainfall will typically lead to an underestimate of streamflow. There is a nonlinear relationship between the flow and the cost of damage to infrastructure. Therefore, the resulting cost of damage from an underestimate may outweigh the alternative cost of adopting a larger design flood (Botto et al. 2014). By improving design practice the economic impact and cost to society can be significantly reduced.

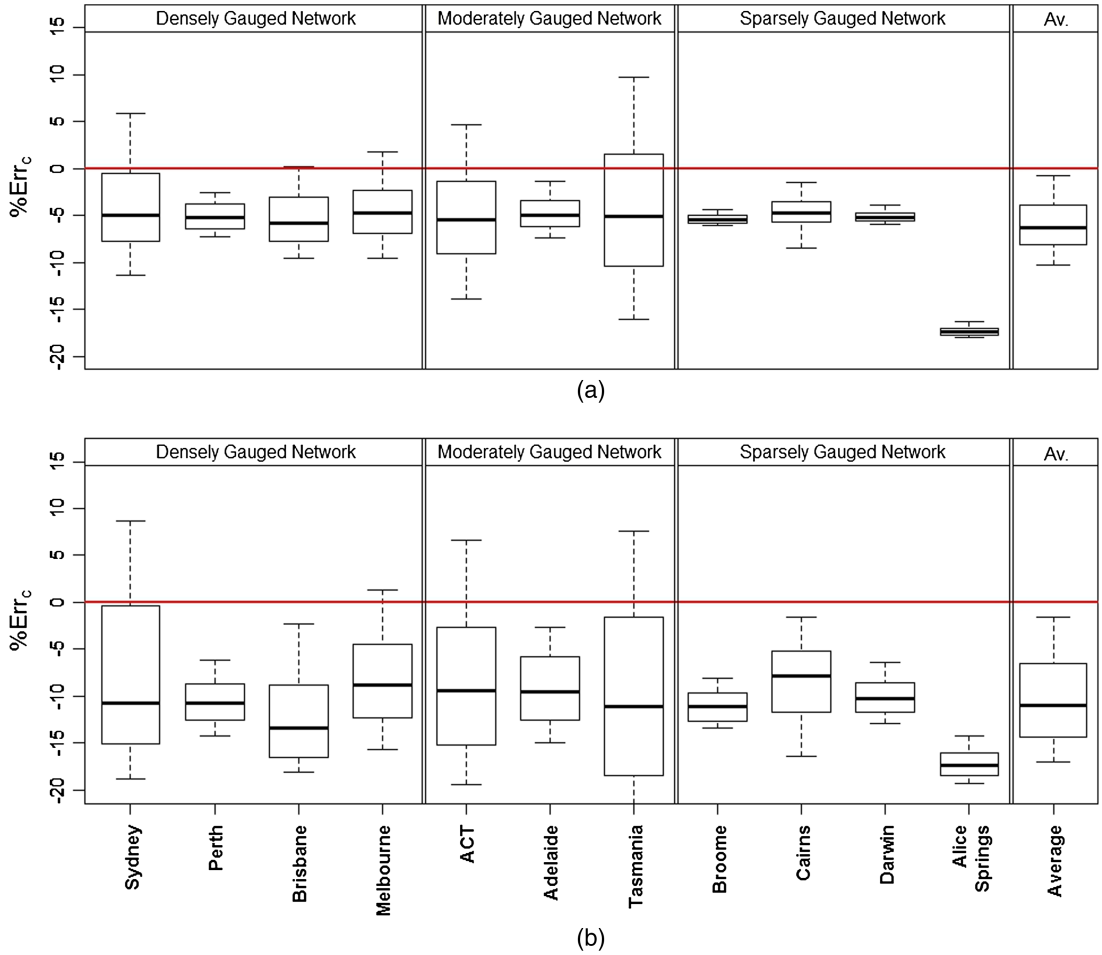


Fig. 7. %Err$_c$ for 1-day, 50y ARI rainfall: (a) 625 km^2; (b) 3,025 km^2 (10 and 90% limits shown)

Sources of Bias

Within a region, the relationship between point and areal rainfall is spatially heterogeneous. For two different locations within a region the scaling of intensity with area can be vastly different. For both ARF approaches, ARF$_c$ and ARF$_{\overline{pt}}$, the derivation of ARFs for each coordinate showed that ARFs varied spatially, leading to a range possible values for a given frequency, duration and spatial extent. The use of a fixed ARF$_R$ contradicts this observed spatial variation. Applying a fixed regional ARF$_R$ introduces bias into estimates of extreme spatial rainfall for the majority of locations (e.g., Fig. 8).

The practice of using a fixed regional ARF$_R$ is attributable to the historical legacy of poor spatial estimates of rainfall. Typically, ARFs have been derived using smaller catchments with higher gauge density and by pooling information from a range of catchments over larger regions (Jordan et al. 2011; Niemczynowicz 1982; Pilgrim 1987; Siriwardena and Weinmann 1996; Yoo et al. 2007). More recently, remotely sensed estimates of rainfall have provided an alternative source to suggest that ARF values are not constant over a region (Durrans et al. 2002). However, the limited length of remotely sensed records prevents definitive conclusions.

This study has relied on spatially interpolated rainfall and was able to show significant spatial variation in ARF values. The challenge with using interpolated data is the inherent spatial smoothing and artifacts introduced by the algorithm used (King et al. 2013). One of the potential impacts of spatial smoothing is that the true at-site ARF values may exhibit more distinct spatial variations than observed in the current study. As a result, the bias in at-site estimates of extreme spatial rainfall are likely to be greater than shown in Figs. 7 and 8.

Few acknowledge that gauge-based estimates of ARFs also rely on an algorithm for constructing the spatial rainfall estimate. Therefore, no matter what method is used there is the potential for influence by an interpolation algorithm. To mitigate this influence the authors focused on densely gauged regions in this study. One of the advantages of the IFDA approach is that it is generic and independent of the interpolation algorithm. Hence it can take advantage of advances in spatial rainfall interpolation techniques.

Both derivation approaches, ARF$_c$ and ARF$_{\overline{pt}}$, exhibited values exceeding 1, indicating the sampled point rainfall was less than the surrounding rainfall. This contrasts with the widely used recommendation that ARF are upper bounded at 1. Only a few previous studies (e.g., Catchlove and Ball 2003) have reported ARF values greater than 1 for three main reasons. Firstly, calculation of an ARF starts with the assumption that extreme point rainfall is more intense than extreme spatial rainfall. This is not always the case, for instance in the presence of a dominant storm path. Secondly, the focus on obtaining a regional ARF value has meant that spatial variations in extreme spatial rainfall are smoothed out. Thirdly, point gauge data is underrepresented in hard to access areas with strong rainfall gradients (Prudhomme and Reed 1999; Svensson and Jones 2010).

Conditions Where Bias Has the Greatest Impact

The paper demonstrated that a significant change in mean percentage error in estimates of extreme spatial rainfall occurs with

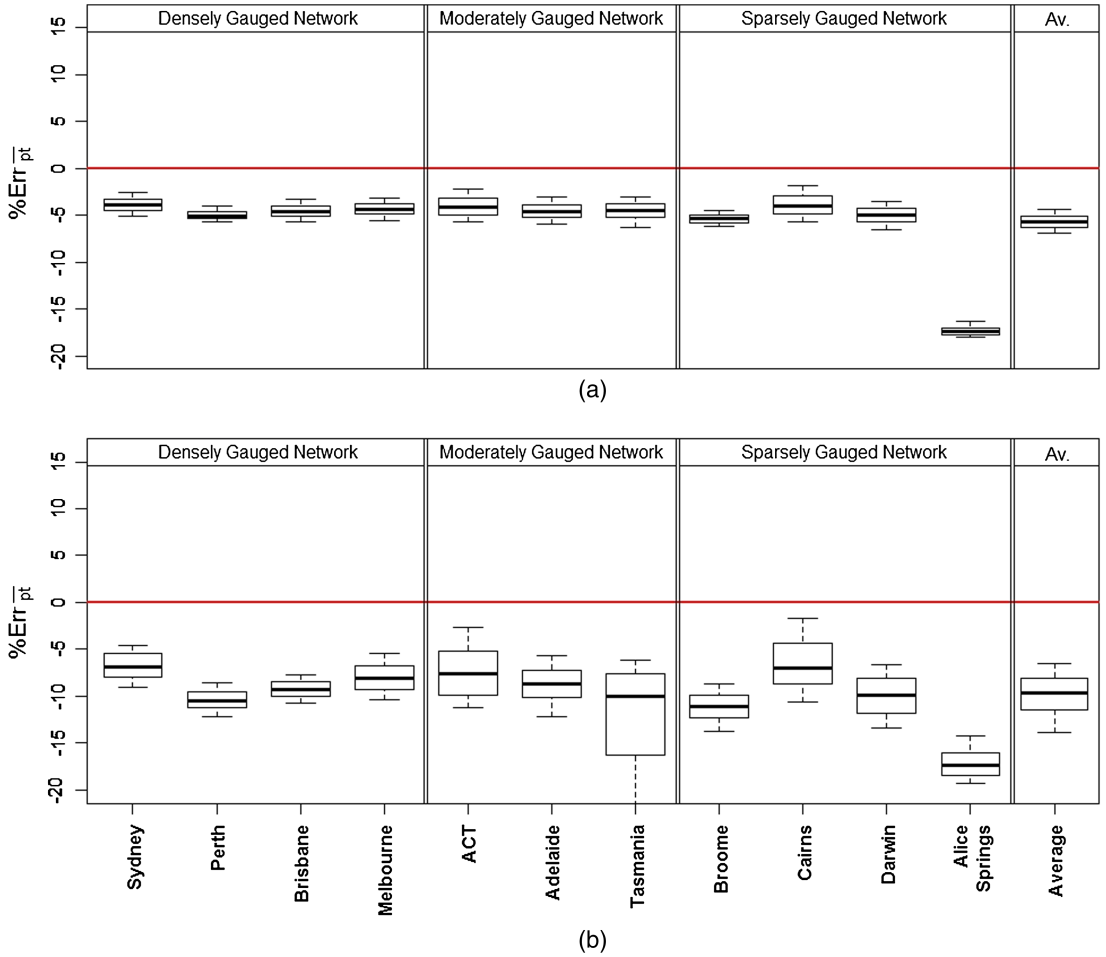


Fig. 8. $\%Err_{\tilde{pt}}$ for 1-day, 50y ARI rainfall: (a) 625 km^2; (b) 3,025 km^2 (10 and 90% limits shown)

increasing area. The largest mean percentage errors were observed for the largest catchments, suggesting that current practice is least effective for large catchments. The poor performance for large catchments is likely attributable to the typically long distances over which rainfall events are correlated at a daily time step. Because of the spatial correlation structure of large rainfall events, the areal reduction of a sampled point rainfall using an ARF_R is likely to underestimate extreme spatial rainfall intensity.

Significant differences in the percentage errors of the extreme spatial rainfall estimates were observed between each region. The difference between the mean percentage errors at different case study locations is greater than the change in mean percentage error across different frequencies (Fig. 11). Thus the frequency of the event seems to be a less decisive factor than location in producing errors in extreme spatial rainfall estimates. The poor performance of the ARF-based approach across all studied regions implies that further investigation into the governing extreme spatial rainfall properties specific to each region is required. These extreme spatial rainfall properties may include seasonality, topography, and rainfall mechanisms contributing to extreme events.

Advantages and Limitations of IFDAs

To summarize advantages and limitations Table 1 presents a comparison of the IFDA and ARF-based approaches across a range of criteria.

ARF values calculated using point rainfall data are assumed to be spatially homogeneous within a general climatic zone (Jordan et al. 2011; Pilgrim 1987; Siriwardena and Weinmann 1996). As IFDAs are location specific estimates of extreme spatial rainfall, any impact of location or climatic region is directly accounted for in the IFDA.

Current practice relies on an ARF and IFD that have been developed separately. This concern was also raised by Panthou et al. (2014) who, in a different approach, sought statistical consistency between ARF and IFD models rather than adopting direct extreme spatial rainfall estimates. The derivation of IFDAs provides unbiased design rainfall estimates by incorporating the spatial variation in spatial rainfall intensity directly. Furthermore, by estimating extreme spatial rainfall directly the IFDA approach ensures statistical consistency in the extreme spatial rainfall estimate. However, they are not without limitation. For example, the application of the approach to irregular catchment areas and near-miss storms requires further investigation.

An ARF is a ratio of spatial rainfall to a representative point rainfall of equal recurrence interval. Its derivation requires an estimate of spatial extreme rainfall, but the limitations of this interpolation are rarely separated from the limitations of the ARF. While the quality of the spatial interpolation method will influence the derivation of IFDAs as they are direct estimates of rainfall intensity, it is nonetheless preferable to separate out the method of interpolation. Ultimately, the use of IFDAs depends on the availability of

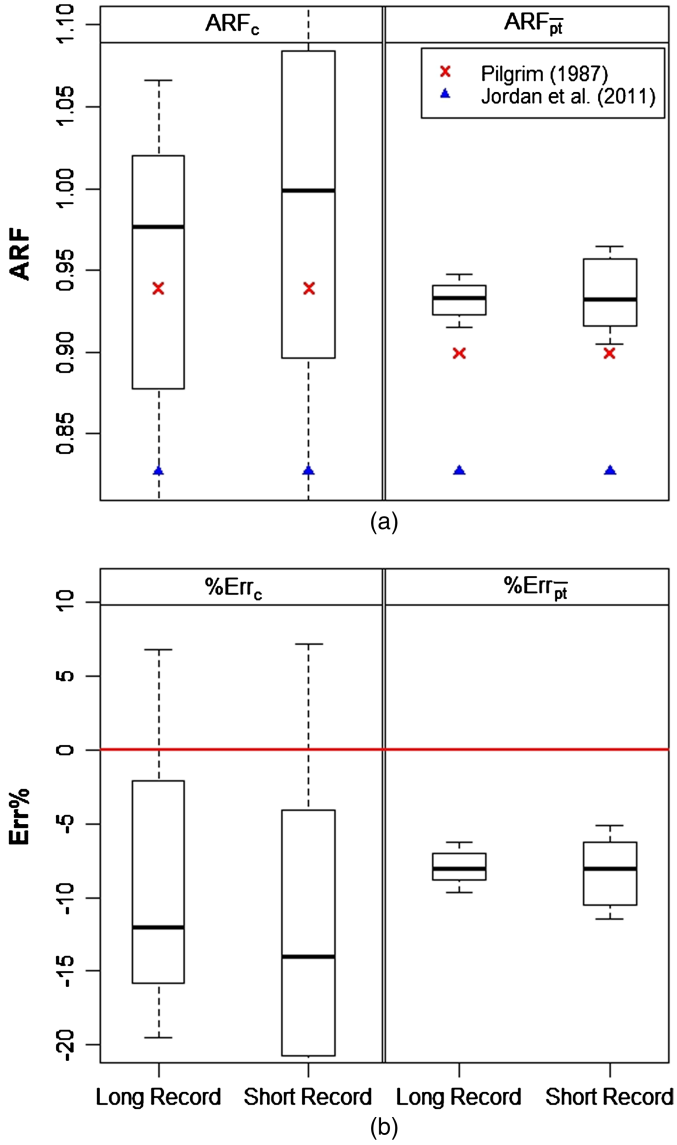


Fig. 9. Sydney ARF and percentage error boxplots for 25y ARI, 1-day rainfall comparing results of the long and short series analysis for an area of 3,025 km^2 (10 and 90% limits shown): (a) ARF$_c$ and ARF$_{\bar{pt}}$; (b) %Err$_c$ and %Err$_{\bar{pt}}$

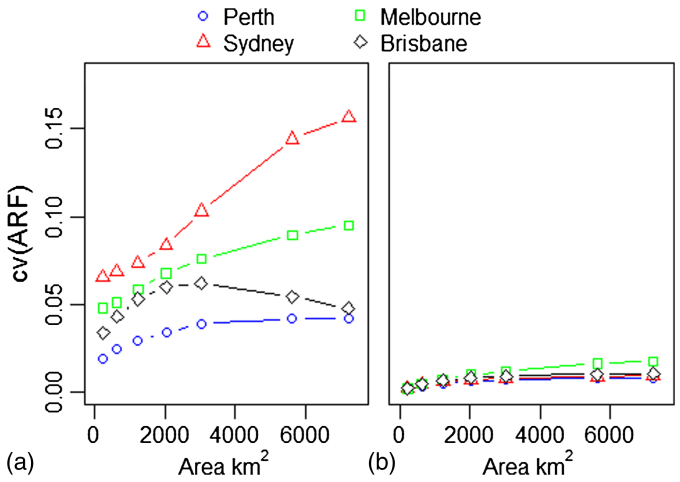


Fig. 10. Comparison of coefficient of variation of at-site ARFs against catchment area for 10y ARI, 1-day rainfall: (a) ARF$_c$; (b) ARF$_{\bar{pt}}$

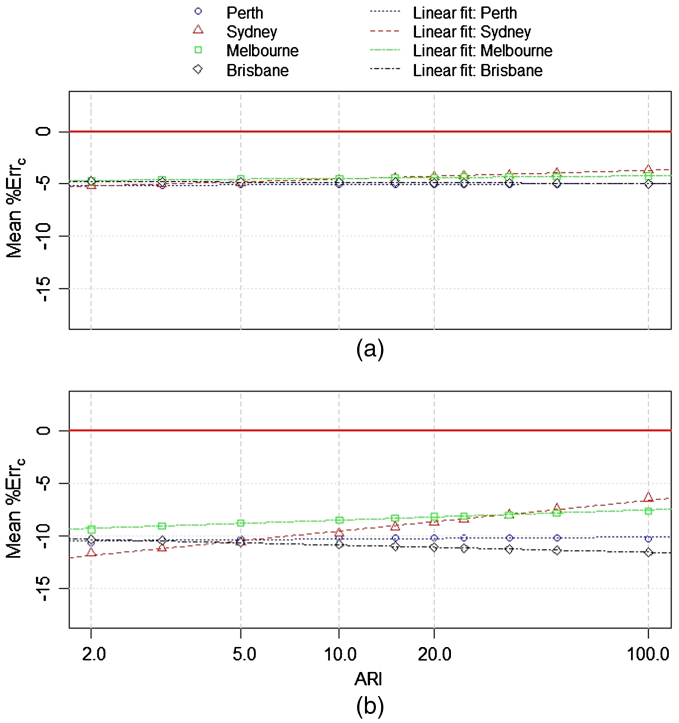


Fig. 11. Mean %Err$_c$ against ARI for 1-day rainfall: (a) 625 km^2; (b) 3,025 km^2

high-quality spatial rainfall data sources. The application of IFDAs to real-world catchments can be improved with further development of continuous spatial rainfall models (Kleiber et al. 2012; Leonard et al. 2008), the continued collection and processing of radar data, and continued development of interpolated spatial rainfall algorithms (Bárdossy and Pegram 2013; Beesley et al. 2009).

Challenges exist in determining the best method to apply IFDAs. This is because as each coordinate within a catchment possesses its own IFDA relationship, for a given ARI/area/duration there will be range of spatial rainfall intensities applicable to the catchment. This characteristic is perhaps a benefit, allowing spatial variability to be incorporated into design estimates.

Conclusions

The key objective of this paper was to introduce and evaluate the IFDA approach for extreme spatial rainfall estimation. An IFDA, which is calculated directly from spatial rainfall fields, adds the extra dimension of area to an IFD curve to account for spatial variation over a catchment. The paper demonstrated that the standard practice of using a fixed regional ARF introduces bias into estimates of extreme spatial rainfall and discussed subsequent implications. The existing approaches for estimating extreme spatial rainfall were typically in error by 5–15% (Figs. 7 and 8). Analysis of the characteristics of these errors showed that they varied with region, catchment area, and frequency.

Gridded daily rainfall from the Australian water availability project (AWAP) for 11 case study regions in Australia (spanning a wide range of climatic zones) was used. The analysis demonstrated that the IFDA relationship varied within regions. This is in contrast with current practice, which assumes that a fixed regional ARF is appropriate. The analysis also demonstrated that ARF values were frequently greater than 1 because of cases where

Table 1. Comparison of ARF-Based and IFDA Approaches

Item	Criteria	ARF-based approach	IFDA approach
1	Spatial scaling relationship	Assumed to be the same for all locations in the region. This causes complications for large catchments or catchments with steep rainfall gradients	Data driven and allowed to vary throughout the region
2	Spatial rainfall estimate	Spatial rainfall estimate has conventionally been coupled to the ARF derivation method	Made directly from spatial gridded estimates. Therefore, the spatial interpolation approach is separate from methodology
3	Conceptual approach	ARF and IFD isolate separate components of an event that are later combined but can potentially be inconsistent	Single metric directly incorporates intensity, frequency, duration, and area of design event
4	Reliance on single representative point rainfall	The ARF scales the extreme rainfall at a key location in the region to yield the spatial estimate	The spatial interpolation is estimated from all gauges in the region without the need to rely on a single location as being representative
5	Type of method for assigning frequency to design event	Indirectly approximates design intensity based on quantile matching of point and areal rainfall not necessarily derived from same event	Directly assigns frequency to design event of certain intensity, area, and duration. This is more consistent with actual events
6	Assumptions of approach	Assumes extreme point rainfall intensity is always greater than the spatial rainfall intensity (i.e., ARF < 1)	Assumes spatial interpolated estimates are appropriate

the extreme spatial rainfall exceeded the intensity of the sampled point rainfall.

The IFDA approach overcomes the shortcomings of existing approaches primarily by avoiding the need to assume a fixed regional ARF value. The IFDA methodology is proposed as a promising technique for obtaining direct and unbiased estimates of extreme spatial rainfall. The application of the method relies on robust methods for interpolating rainfall, and is therefore benefited by improvements to interpolation algorithms.

Acknowledgments

This work was supported by an Australian Research Council Discovery grant: A new flood design methodology for a variable and changing climate DP1094796. Additional financial support was provided by the CSIRO Climate Adaptation Flagship. The authors gratefully thank Mark Babister for valuable discussions and constructive comments.

References

Allen, R. J., and DeGaetano, A. T. (2005). "Areal reduction factors for two eastern United States regions with high rain-gauge density." *J. Hydrol. Eng.*, 10.1061/(ASCE)1084-0699(2005)10:4(327), 327–335.

Asquith, W. H., and Famiglietti, J. S. (2000). "Precipitation areal-reduction factor estimation using an annual-maxima centered approach." *J. Hydrol.*, 230(1–2), 55–69.

Bárdossy, A., and Pegram, G. (2013). "Interpolation of precipitation under topographic influence at different time scales." *Water Resour. Res.*, 49(8), 4545–4565.

Beesley, C. A., Frost, A., and Zajaczkowski, J. (2009). "A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets." *Proc., 18th World IMACS/MODSIM Congress*, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, Queensland, 3892–3886.

Botto, A., Ganora, D., Laio, F., and Claps, P. (2014). "Uncertainty compliant design flood estimation." *Water Resour. Res.*, 50(5), 4242–4253.

Boughton, W., and Jakob, D. (2008). "Adjustment factors for restricted rainfall." *Aust. J. Water Resour.*, 12(1), 37–48.

Catchlove, R. H., and Ball, J. E. (2003). "A hydroinformatic approach to the development of areal reduction factors." *28th Int. Hydrology and Water Resources Symp.*, Institution of Engineers, Barton, Australia, 1.9–1.15.

Durrans, S. R., Julian, L. T., and Yekta, M. (2002). "Estimation of depth-area relationships using radar-rainfall data." *J. Hydrol. Eng.*, 10.1061/(ASCE)1084-0699(2002)7:5(356), 356–367.

Green, J., Xuereb, K., Johnson, F., Moore, G., and The, C. (2012). "The revised intensity-frequency-duration (IFD) design rainfall estimates for Australia: An overview." *Hydrology and Water Resources Symp. 2012*, Engineers Australia, Barton, Australia, 808–815.

Jakob, D., Taylor, B. F., and Xuereb, K. C. (2005). "A pilot study to explore methods for deriving design rainfalls for Australia—Part 1." *HRS No. 10*, Bureau of Meteorology, Melbourne, Australia.

Jordan, P. W., Sih, K., Hill, P., Nandakumar, N., Weinmann, P. E., and Nathan, R. J. (2011). "Areal reduction factors for estimation of design rainfall intensities for New South Wales and the Australia capital territory." *Proc., 33rd Hydrology and Water Resources Symp.*, Engineers Australia, Barton, Australia, 281–225.

King, A. D., Alexander, L. V., and Donat, M. G. (2013). "The efficacy of using gridded data to examine extreme rainfall characteristics: A case study for Australia." *Int. J. Climatol.*, 33(10), 2376–2387.

Kleiber, W., Katz, R. W., and Rajagopalan, B. (2012). "Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes." *Water Resour. Res.*, 48(1), W01523.

Kousky, C., and Walls, M. (2014). "Floodplain conservation as a flood mitigation strategy: Examining costs and benefits." *Ecol. Econ.*, 104(0), 119–128.

Leonard, M., Lambert, M. F., Metcalfe, A. V., and Cowpertwait, P. S. P. (2008). "A space-time Neyman-Scott rainfall model with defined storm extent." *Water Resour. Res.*, 44(9), W09402.

Miller, S., Muir-Wood, R., and Boissonnade, A. (2008). "An exploration of trends in normalized weather-related catastrophe losses." *Climate extremes and society*, H. F. Diaz, and R. J. Murnane, eds., Cambridge University Press, Cambridge, U.K.

Myers, V. A., and Zehr, R. M. (1980). "A methodology for point to area ratios." *NOAA Technical Rep.*, National Water Service, Silver Spring, MD.

Niemczynowicz, J. (1982). "Areal intensity-duration-frequency curves for short term rainfall events in Lund." *Nordic Hydrol.*, 13(4), 193–204.

Panthou, G., Vischel, T., Lebel, T., Quantin, G., and Molinié, G. (2014). "Characterizing the space-time structure of rainfall in the Sahel with a view to estimating IDAF curves." *Hydrol. Earth Syst. Sci. Discuss.*, 11(7), 8409–8441.

Pilgrim, D. (1987). "Australian rainfall and runoff: A guide to flood estimation." Institution of Engineers, Barton, Australia.

Prudhomme, C., and Reed, D. W. (1999). "Mapping extreme rainfall in a mountainous region using geostatistical techniques: A case study in Scotland." *Int. J. Climatol.*, 19(12), 1337–1356.

Raupach, M., Briggs, P., Haverd, V., King, E., Paget, M., and Trudinger, C. (2012). *Australian water availability project*, CSIRO Marine and Atmospheric Research, Canberra, Australia.

Siriwardena, L., and Weinmann, P. (1996). "Derivation of areal reduction factors for design rainfalls in Victoria." *Cooperative Research Centre for Catchment Hydrology Rep. No. 96*, Monash Univ., Clayton, VIC, Australia.

Sivapalan, M., and Blöschl, G. (1998). "Transformation of point rainfall to areal rainfall: Intensity-duration-frequency curves." *J. Hydrol.*, 204(1–4), 150–167.

Strömberg, D. (2007). "Natural disasters, economic development, and humanitarian aid." *J. Econ. Perspect.*, 21(3), 199–222.

Svensson, C., and Jones, D. A. (2010). "Review of methods for deriving areal reduction factors." *J. Flood Risk Manage.*, 3(3), 232–245.

van Montfort, M. A. J. (1990). "Sliding maxima." *J. Hydrol.*, 118(1–4), 77–85.

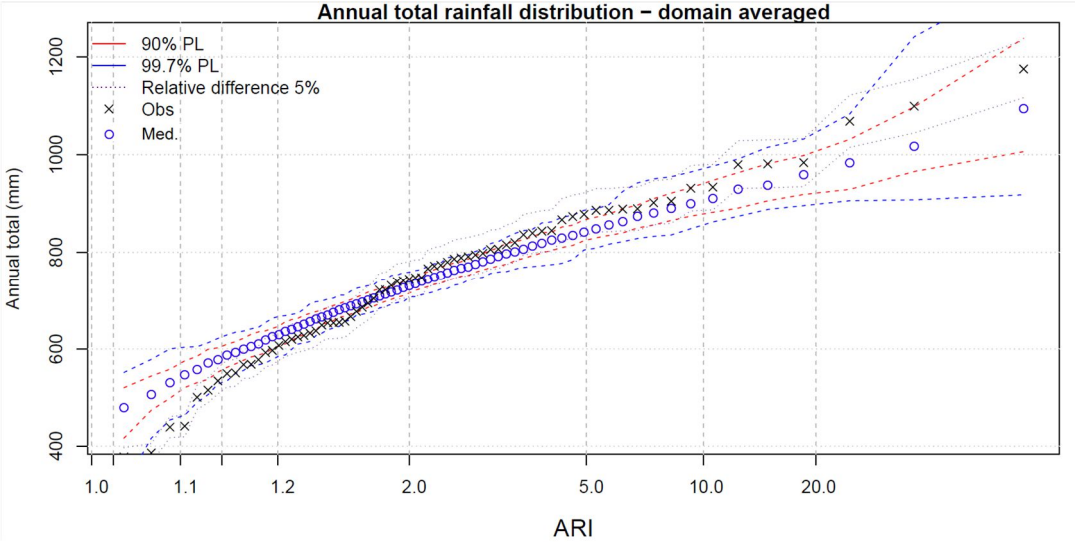
Yoo, C., Kim, K., Kim, H. S., and Park, M. J. (2007). "Estimation of areal reduction factors using a mixed gamma distribution." *J. Hydrol.*, 335(3–4), 271–284.

Appendix B

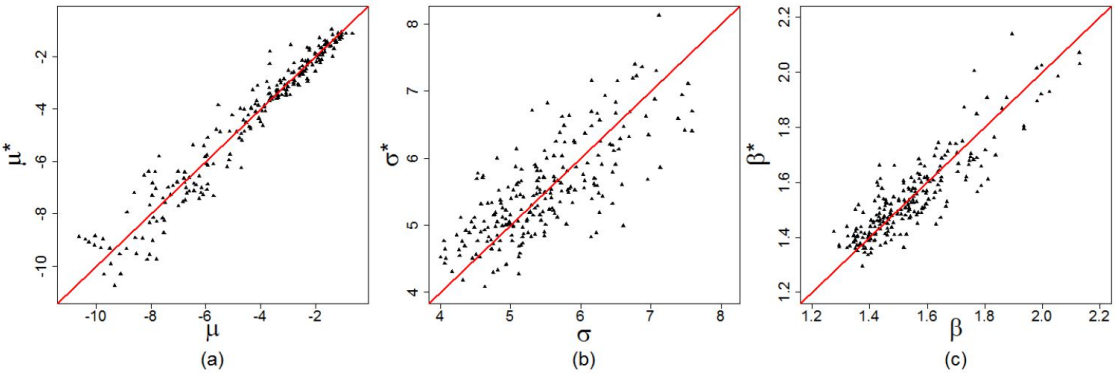
Supplementary Material:

Comprehensive evaluation of a latent variable approach to continuously simulate daily rainfall fields (Paper 2)

Supplementary Material A

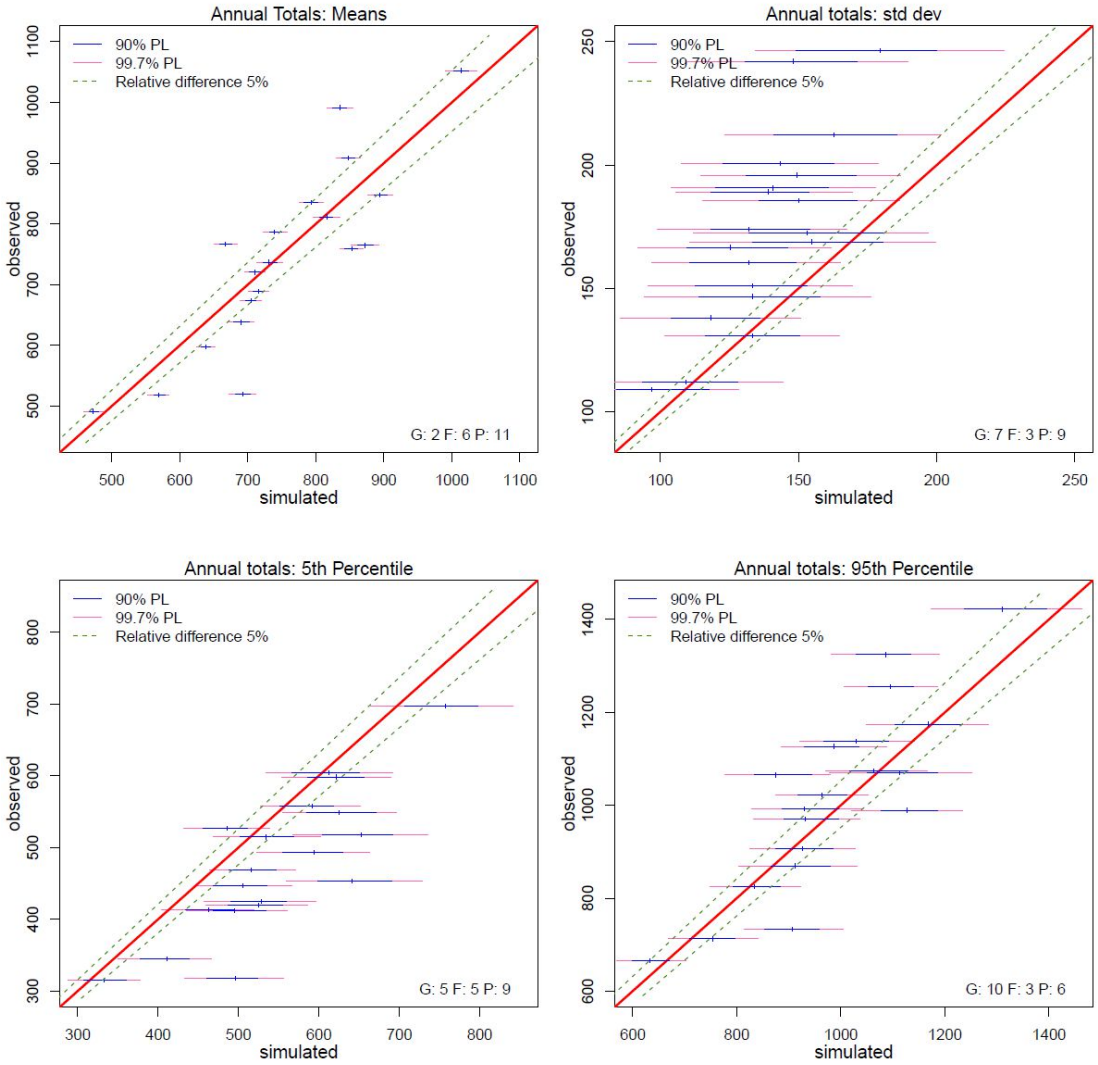


Supplementary Material B

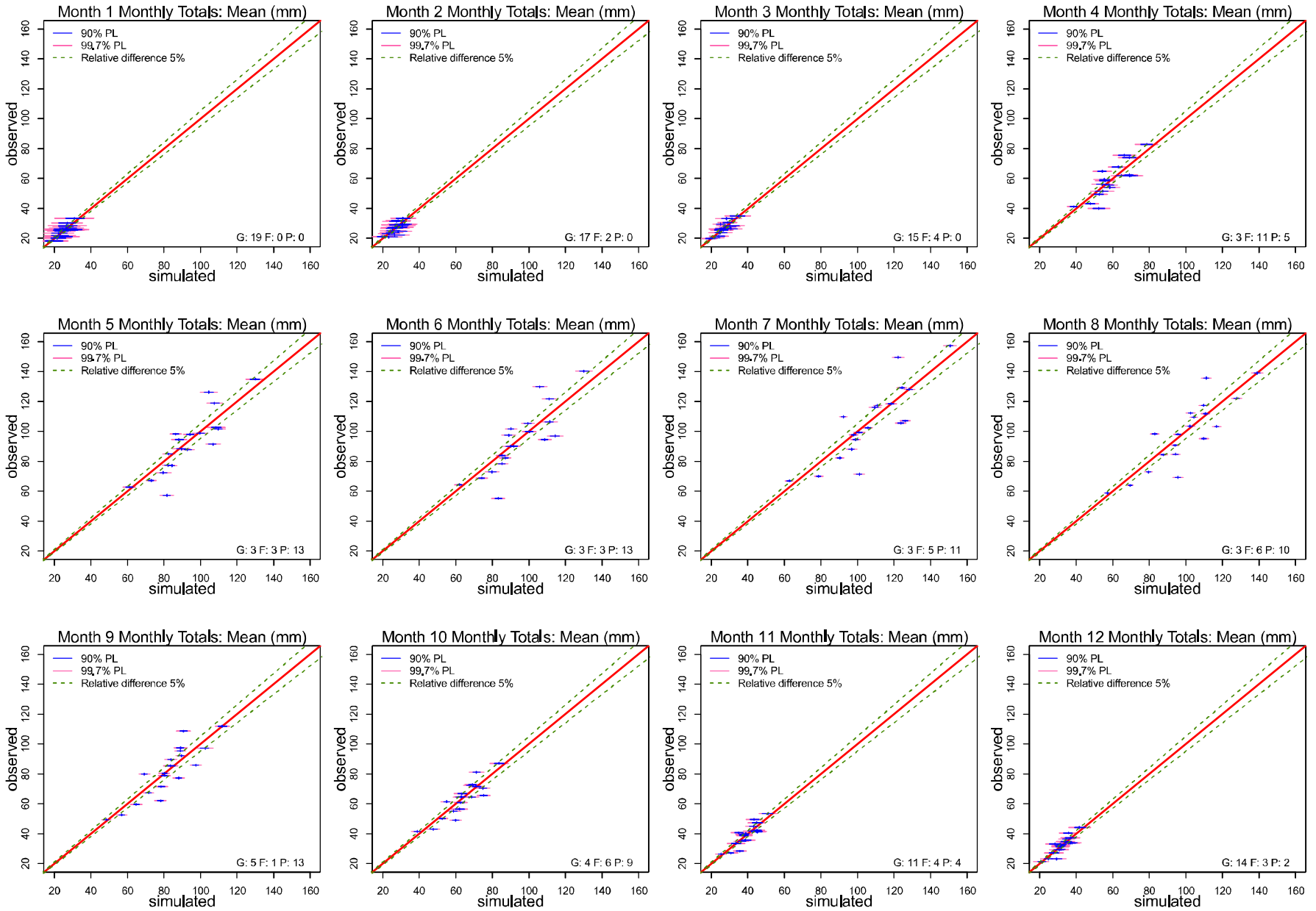


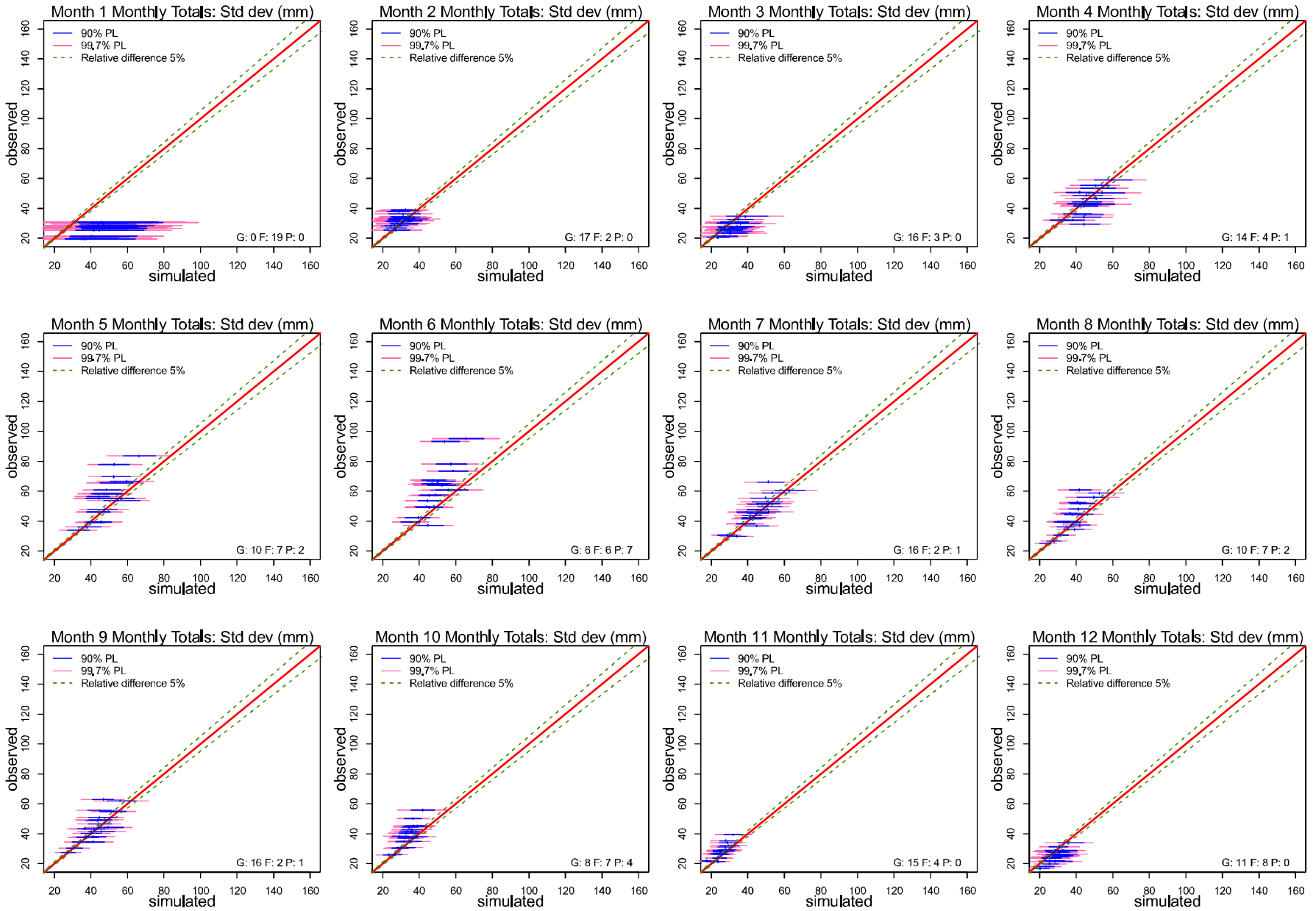
Cross-validation of kriged parameters with elevation trend pooled for all months and sites, estimated* v calibrated; (a) μ* v μ, (b) σ *v σ and, (c) β* v β

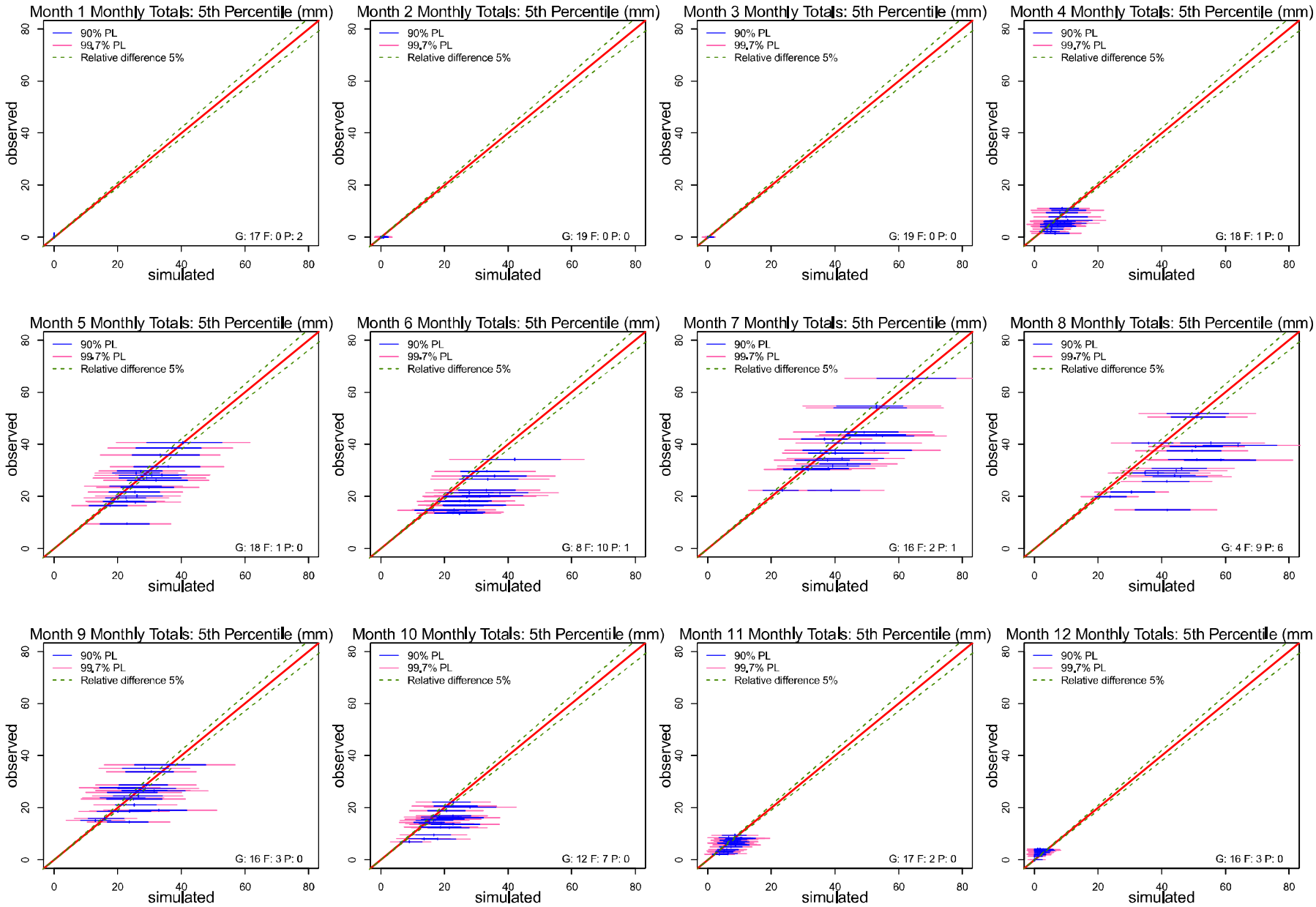
Supplementary Material C

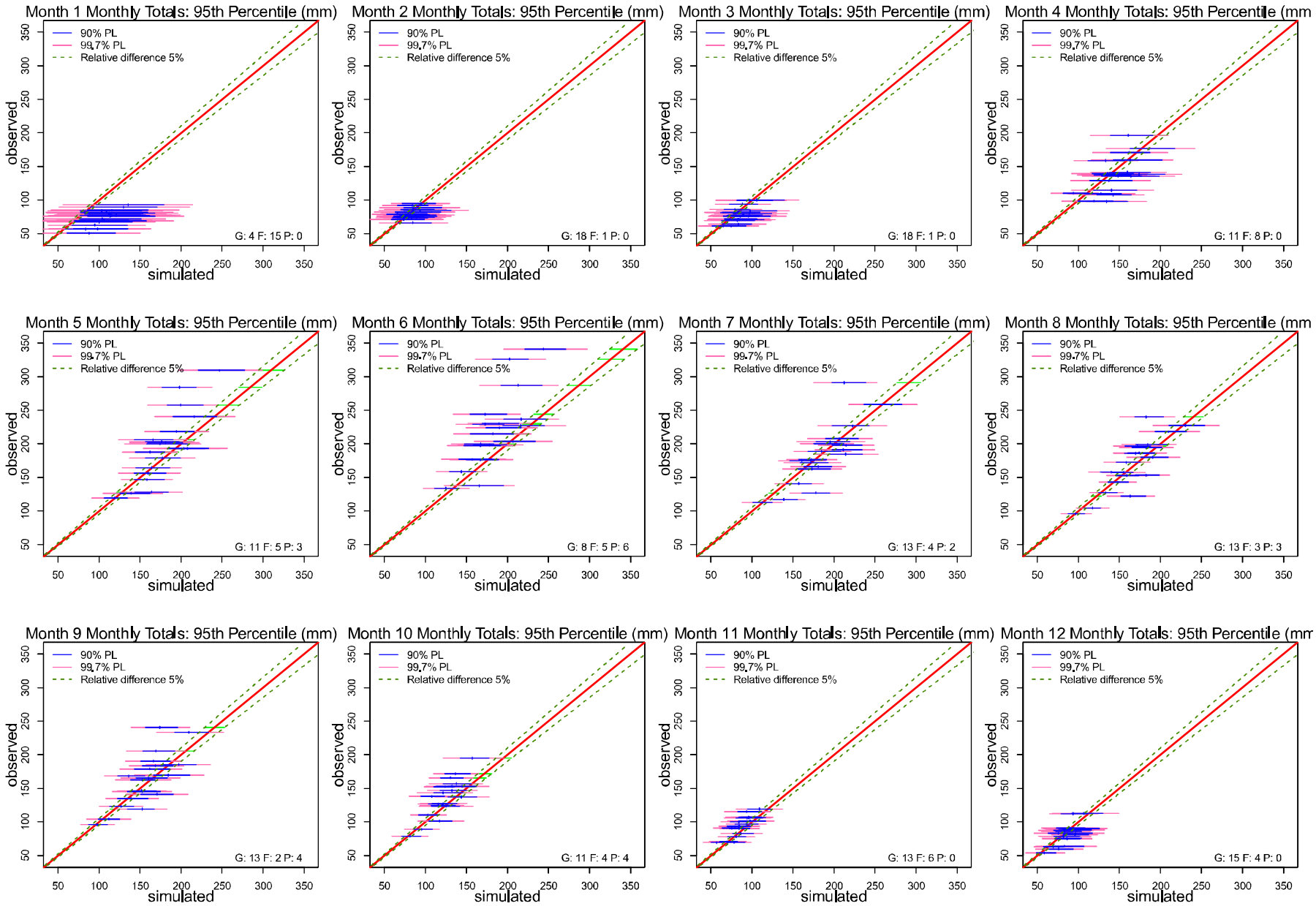


Supplementary Material D


Month 1 Monthly Totals: Mean (mm)
90% PL
99.7% PL
Relative difference 5%
G: 19 F: 0 P: 0
Month 2 Monthly Totals: Mean (mm)
G: 17 F: 2 P: 0
Month 3 Monthly Totals: Mean (mm)
G: 15 F: 4 P: 0
Month 4 Monthly Totals: Mean (mm)
G: 3 F: 11 P: 5
Month 5 Monthly Totals: Mean (mm)
G: 3 F: 3 P: 13
Month 6 Monthly Totals: Mean (mm)
G: 3 F: 3 P: 13
Month 7 Monthly Totals: Mean (mm)
G: 3 F: 5 P: 11
Month 8 Monthly Totals: Mean (mm)
G: 3 F: 6 P: 10
Month 9 Monthly Totals: Mean (mm)
G: 5 F: 1 P: 13
Month 10 Monthly Totals: Mean (mm)
G: 4 F: 6 P: 9
Month 11 Monthly Totals: Mean (mm)
G: 11 F: 4 P: 4
Month 12 Monthly Totals: Mean (mm)
G: 14 F: 3 P: 2
observed
simulated
20
40
60
80
100
120
140
160


Month 1 Monthly Totals: Std dev (mm)
90% PL
99.7% PL
Relative difference 5%
G: 0 F: 19 P: 0
Month 2 Monthly Totals: Std dev (mm)
G: 17 F: 2 P: 0
Month 3 Monthly Totals: Std dev (mm)
G: 16 F: 3 P: 0
Month 4 Monthly Totals: Std dev (mm)
G: 14 F: 4 P: 1
Month 5 Monthly Totals: Std dev (mm)
G: 10 F: 7 P: 2
Month 6 Monthly Totals: Std dev (mm)
G: 6 F: 6 P: 7
Month 7 Monthly Totals: Std dev (mm)
G: 16 F: 2 P: 1
Month 8 Monthly Totals: Std dev (mm)
G: 10 F: 7 P: 2
Month 9 Monthly Totals: Std dev (mm)
G: 16 F: 2 P: 1
Month 10 Monthly Totals: Std dev (mm)
G: 8 F: 7 P: 4
Month 11 Monthly Totals: Std dev (mm)
G: 15 F: 4 P: 0
Month 12 Monthly Totals: Std dev (mm)
G: 11 F: 8 P: 0
observed
simulated
20
40
60
80
100
120
140
160


Month 1 Monthly Totals: 5th Percentile (mm)
90% PL
99.7% PL
Relative difference 5%
G: 17 F: 0 P: 2
Month 2 Monthly Totals: 5th Percentile (mm)
G: 19 F: 0 P: 0
Month 3 Monthly Totals: 5th Percentile (mm)
G: 19 F: 0 P: 0
Month 4 Monthly Totals: 5th Percentile (mm)
G: 18 F: 1 P: 0
Month 5 Monthly Totals: 5th Percentile (mm)
G: 18 F: 1 P: 0
Month 6 Monthly Totals: 5th Percentile (mm)
G: 8 F: 10 P: 1
Month 7 Monthly Totals: 5th Percentile (mm)
G: 16 F: 2 P: 1
Month 8 Monthly Totals: 5th Percentile (mm)
G: 4 F: 9 P: 6
Month 9 Monthly Totals: 5th Percentile (mm)
G: 16 F: 3 P: 0
Month 10 Monthly Totals: 5th Percentile (mm)
G: 12 F: 7 P: 0
Month 11 Monthly Totals: 5th Percentile (mm)
G: 17 F: 2 P: 0
Month 12 Monthly Totals: 5th Percentile (mm
G: 16 F: 3 P: 0
observed
simulated
0
20
40
60
80


Month 1 Monthly Totals: 95th Percentile (mm)
Month 2 Monthly Totals: 95th Percentile (mm)
Month 3 Monthly Totals: 95th Percentile (mm)
Month 4 Monthly Totals: 95th Percentile (mm
Month 5 Monthly Totals: 95th Percentile (mm)
Month 6 Monthly Totals: 95th Percentile (mm)
Month 7 Monthly Totals: 95th Percentile (mm)
Month 8 Monthly Totals: 95th Percentile (mm
Month 9 Monthly Totals: 95th Percentile (mm)
Month 10 Monthly Totals: 95th Percentile (mm)
Month 11 Monthly Totals: 95th Percentile (mm)
Month 12 Monthly Totals: 95th Percentile (mm
90% PL
99.7% PL
Relative difference 5%
observed
simulated
G: 4 F: 15 P: 0
G: 18 F: 1 P: 0
G: 18 F: 1 P: 0
G: 11 F: 8 P: 0
G: 11 F: 5 P: 3
G: 8 F: 5 P: 6
G: 13 F: 4 P: 2
G: 13 F: 3 P: 3
G: 13 F: 2 P: 4
G: 11 F: 4 P: 4
G: 13 F: 6 P: 0
G: 15 F: 4 P: 0