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

A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model

Journal of Hydrology, 2018; 556:1123-1138

Final publication at http://dx.doi.org/10.1016/j.jhydrol.2016.12.043

A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model

Bree Bennett[1], Mark Thyer[1], Michael Leonard[1], Martin Lambert[1], Bryson Bates[2]

[1]School of Civil, Environmental and Mining Engineering,

University of Adelaide North Terrace Campus

SA 5005

Australia

Email: bree.bennett@adelaide.edu.au

Telephone: +61 8 8313 1113

Fax: +61 8 8303 4359

[2]CSIRO Oceans and Atmosphere

Underwood Ave

Floreat

WA 6014

Australia

Keywords

Rainfall generation, spatial rainfall simulation, continuous simulation, rainfall intensity, latent variable approach.

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, the first goal of this study is to present a new parsimonious stochastic model that produces daily rainfall fields across the catchment. To achieve parsimony, the model used the latent-variable approach (because this parsimoniously simulates rainfall occurrences as well as amounts) and several other assumptions (including contemporaneous and separable spatiotemporal covariance structures). The second goal was to develop a comprehensive and systematic evaluation (CASE) framework to identify model strengths and weaknesses. This included quantitative performance categorisation that provided a systematic, succinct and transparent method to assess and summarise model performance over a range of statistics, sites, scales and seasons. 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 CASE framework was that it was able to 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, including drought and flood risk.

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 all practically relevant temporal and spatial characteristics over a broad range of scales. Despite the significance of this need, as yet, there are 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 (Kleiber et al. 2012; Wilks 2009) 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).

Box and Jenkins (1976 p. 17) note the general importance of parsimony in the development of stochastic models. The particular applications of spatial rainfall field models also require a parsimonious approach. Continuous hydrological simulation for applications such as flood and drought risk typically require long-term sequences of rainfall. 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 applications assessing 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 their focus is typically restricted to single events. While these complex spatial-temporal rainfall models provide insight into the spatiotemporal structure of individual rainfall events, it remains to be demonstrated how they can be used to generate long-term rainfall sequences suitable for continuous hydrological simulation of a catchment.

This paper describes a parsimonious model for spatial rainfall fields and evaluates its performance over a range of spatial and temporal scales. The rainfall field model is based on the multi-site model of (Rasmussen 2013) and uses a Gaussian latent-variable approach that simulates rainfall occurrence and amounts using a simple power transformation, taking full advantage of the parsimonious nature of the transformed latent-variable approach. There are numerous enhancements on (Rasmussen 2013) model to enable parsimonious simulation of rainfall fields. These include adopting a contemporaneous and separable covariance structure. Kriging is used to produce parameter surfaces since the Gaussian latent-variable representation is well 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 areal 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 hydrological models.

While there are some 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 selected statistics, sites and months using adhoc, descriptive performance assessment (e.g. words such as 'adequate', or 'suitable'). In this paper, a comprehensive and systematic approach to model evaluation is presented. It is comprehensive because it clearly summarises model performance over a wide range of spatial (all sites/fields) and temporal (days/seasons/years) scales. It is systematic because it includes a transparent performance categorisation scheme, which enables comparison of performance over a range of model properties and hence provides a mechanism to clearly identify model strengths and weakness. Furthermore, in previous studies, cross-validation was typically undertaken for only a few select sites. A further benefit of a systematic approach is that it enables evaluation on the basis of full leave-one-out cross-validation across all sites within the region.

This paper has two objectives: (1) to present a parsimonious latent-variable rainfall model to generate spatial rainfall fields continuously; and (2) to present and apply a comprehensive and systematic evaluation framework of model performance across a range of spatial and temporal scales. The remaining paper is divided into the following sections. Section 2 describes the development of the parsimonious rainfall field model, while Section 3 sets out the calibration procedure. Section 4 introduces the comprehensive and systematic performance evaluation framework. Section 5 presents the case study while Section 6 provides the results of applying the rainfall field model to the case study, using the comprehensive and systematic evaluation framework. The discussion in Section 7, interprets the performance results and compares to other rainfall field models in the literature. Section 8 summarises the key conclusions.

2 Stochastic Rainfall Field Model Development

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).

A new stochastic daily rainfall field model is presented here that is parsimonious and simulates daily rainfall continuously in space as a field. Hereafter this new model will be referred as the parsimonious rainfall field (PRF) model and it is based on the multisite latent-variable model of Rasmussen (2013). The presentation begins by summarising a general form of latent-variable models for rainfall (Section 2.1), then summarises the multisite model of Rasmussen (2013) in Section 2.2. The extension of the latent-variable approach for simulation of spatial fields is presented in Section 2.3 with specific discussion of the temporal and spatial modelling components.

2.1 General Set-up for a Daily Multivariate Latent-Variable Model

The latent-variable concept simulates rainfall by sampling from a normally distributed ‘hidden’ variable. Where values lie below zero, the distribution is truncated and assigned a value of zero, representing a dry day. Where the values of the latent-variable are positive the latent-variable undergoes a transformation, in this case a power transform, so that the skewed distribution of observed rainfall can be reproduced. This procedure can be defined as follows: where r_t^i is 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_t^i via truncation at zero and a power transformation,

$$r_t^i = \begin{cases} (l_t^i)^{\beta_t^i} & l_t^i > 0 \\ 0 & \text{otherwise} \end{cases}, \qquad (1)$$

where β_t^i is the power transformation parameter. Note that, in general, transformations other than the power transformations (e.g. Baxevani and Lennartsson 2015) could also be used.

To enable simulation at multiple sites or spatial fields of rainfall, multivariate specifications are required where the latent-variable is specified as a multivariate normal. Let $\boldsymbol{L} = [l_t^i;\ i = 1, \dots, N;\ t = 1, \dots, T]$ be the latent variable at all spatial locations (all sites in multi-site implementation or all points in the entire field for spatial field implementation), N, and timesteps, T, the multi-variate representation becomes

$$\boldsymbol{L} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad (2)$$

where $\boldsymbol{\mu}$ is the mean at all locations and timesteps and $\boldsymbol{\Sigma}$ is the covariance matrix between all locations and timesteps. To simplify the spatial simulation of the latent-variables at all locations, for a given time step, $\boldsymbol{l}_t = [l_t^i;\ i = 1, \dots, N]$ is conditioned on the previous timestep according to:

$$\boldsymbol{l}_t | \boldsymbol{l}_{t-1} \sim \mathrm{MVN}(\boldsymbol{\mu}', \boldsymbol{\Sigma}') \qquad (3)$$

$$\begin{aligned} \boldsymbol{\mu}' &= \boldsymbol{\mu}_t + \boldsymbol{\Sigma}_{t,t-1} \boldsymbol{\Sigma}_{t-1,t-1}^{-1} (\boldsymbol{l}_{t-1} - \boldsymbol{\mu}_{t-1}) \\ \boldsymbol{\Sigma}' &= \boldsymbol{\Sigma}_{t,t} - \boldsymbol{\Sigma}_{t,t-1} \boldsymbol{\Sigma}_{t-1,t-1}^{-1} \boldsymbol{\Sigma}_{t-1,t} \end{aligned} \qquad (4)$$

where $\boldsymbol{\mu}_t$ and $\boldsymbol{\mu}_{t-1}$ are the means at respective time steps, $\boldsymbol{\Sigma}_{t,t}$ and $\boldsymbol{\Sigma}_{t-1,t-1}$ 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. Following the specification of the lag-0 and lag-1 covariance matrices time series of rainfall at multiple locations can be simulated.

2.2 Multisite Latent-Variable Model

The following description is a summary of the relevant components of the multi-site latent-variable model from Rasmussen (2013), hereafter referred to as R2013 model. To incorporate seasonality in rainfall, parameters in the R2013 model parameters are different for each month, but constant within a particular month. The implications of this are that for a particular month, $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1}$, $\boldsymbol{\Sigma}_{t,t} = \boldsymbol{\Sigma}_{t,t}$ and $\boldsymbol{\Sigma}_{t,t-1} = \boldsymbol{\Sigma}_{t-1,t}$. This simplifies equation (4) such that:

$$\begin{aligned} \boldsymbol{\mu}' &= \boldsymbol{\mu}_t + \boldsymbol{\Sigma}_{t,t-1}\boldsymbol{\Sigma}_{t,t}^{-1}(\boldsymbol{l}_{t-1} - \boldsymbol{\mu}_t) \\ \boldsymbol{\Sigma}' &= \boldsymbol{\Sigma}_{t,t} - \boldsymbol{\Sigma}_{t,t-1}\boldsymbol{\Sigma}_{t,t}^{-1}\boldsymbol{\Sigma}_{t,t-1} \end{aligned} \tag{5}$$

To preserve the spatial-temporal properties of rainfall at multi-sites the R2013 model used a full multivariate first order autoregressive model. This means that all the lag-0 and lag-1 cross-covariances between modelled sites are explicitly specified for all pairs of locations i and $j = 1, \dots, N$,

$$\boldsymbol{\Sigma}_{t,t} = \begin{bmatrix} \Sigma_{t,t}^{1,1} & \cdots & \Sigma_{t,t}^{1,j} \\ \vdots & \ddots & \vdots \\ \Sigma_{t,t}^{i,1} & \cdots & \Sigma_{t,t}^{i,j} \end{bmatrix} \quad \boldsymbol{\Sigma}_{t,t-1} = \begin{bmatrix} \Sigma_{t,t-1}^{1,1} & \cdots & \Sigma_{t,t-1}^{1,j} \\ \vdots & \ddots & \vdots \\ \Sigma_{t,t-1}^{i,1} & \cdots & \Sigma_{t,t-1}^{i,j} \end{bmatrix} \tag{6}$$

This model specification requires the estimation of a large number of parameters ($3N + 2N^2$) based on N sites and where a site here refers to a location within a region with observed rainfall data. This large number of parameters makes it infeasible to apply for the simulation of spatial fields (e.g. for 100 km^2 field with grid size 1 km^2 this would require over 20,000 parameters). Hence model enhancements are required to enable spatial field simulation. The requirement for model parsimony to enable spatial field simulation is discussed in Section 2.3.4.

2.3 Enhancements to enable parsimonious spatial field modelling

This section outlines the enhancements made to the R2013 model to develop the PRF model. To extend the R2013 model to be continuous in space each parameter must be specified across the whole simulation region, rather than just for selected sites. Therefore the latent-variable is specified as a Gaussian Random Field. To represent seasonality, the approach of R2103 model is used, where the parameters are specified on monthly basis.

To develop the PRF, by extending the R2013 model to space, the following assumptions have been made:

1. A contemporaneous approach is used, in that only the lag-0 cross-covariances are explicitly modelled. Therefore, $\boldsymbol{\Sigma}_{t,t-1}$ preserves the diagonal covariances (related to at-site autoregressive parameters) and off-diagonals are zero.
2. The use of a separable cross-covariance $\boldsymbol{\Sigma}_{t,t-1} = \varphi_t \boldsymbol{\Sigma}_{t,t}$ where the temporal component is denoted by a scalar autoregressive parameter, φ_t, is separate from the spatial component $\boldsymbol{\Sigma}_{t,t}$ (Genton 2007).
3. The use of a single scalar autoregressive parameter, φ_t**,** across the entire field.
4. The use of a spatial correlation function to model the lag-0 cross-covariances.
5. The use of spatial interpolation approach to specify the latent-variable parameters for all locations over the entire field.

The PRF modelling specification that results from these assumptions is outlined in the following sections, first considering the temporal component and then the spatial modelling components. During this description, the PRF model will be compared against R2013 model to clearly identify the differences.

2.3.1 Temporal modelling component

Assumptions 1, 2 and 3 above mean that the temporal modelling component reduces to an multivariate AR(1) model, specified as follows:

$$\begin{aligned} \boldsymbol{\mu}' &= \boldsymbol{\mu}_t + \varphi_t(\boldsymbol{l}_{t-1} - \boldsymbol{\mu}_{t-1}) \\ \boldsymbol{\Sigma}' &= (1 - \varphi_t^2)\, \Sigma_{t,t}^2 \end{aligned} \tag{7}$$

where φ_t is the autoregressive parameter. This assumption of a spatially constant autoregressive parameter across the region enables the temporal correlation structure to be continuous over the simulation region. In contrast, the R2013 model has individual auto-correlations for each site. While this may improve the fit of the R2013 model to empirical auto-correlations at individual sites (Rasmussen 2013), it requires many additional parameters to be estimated and, because it does not have a mechanism to spatially interpolate the at-site auto-correlations, it is not possible to simulate continuous spatial fields with a temporal correlation structure. Whether the assumption of a spatially constant autoregressive parameter in the PRF model will provide an adequate fit to the observed data will depend on how spatially homogeneous a study region is in terms of observed daily auto-correlations. This assumption is tested in Section 6.1.1.

2.3.2 Spatial modelling component

The use of a separable covariance structure and contemporaneous approach (assumptions 1 and 2) enables the specification of a spatial correlation function (assumption 3) to model the lag-0 covariances. The main reason for this assumption is to enable the PRF model to be continuous in space, which requires a continuous positive definite spatial correlation structure for simulation of the latent variable.

For this spatial correlation function an isotropic, powered exponential function was chosen (Gneiting 2002).This is specified by considering that the lag-0 covariance matrix $\mathbf{\Sigma}_{t,t}$ has elements $\Sigma_{t,t}^{ij} = \sigma_t^i \sigma_t^j \rho\left(d^{ij} \middle| \nu_t, \alpha_t, \lambda_t\right)$ for all pairs of locations i and $j = 1, \ldots, N$. Where $\sigma^i$and $\sigma^j$are the standard deviations at each location, $d^{ij}$is the distance between the locations, and ν_t, α_t and λ_t are the parameters of an isotropic powered-exponential correlation function defined by

$$\rho\left(d^{ij} \middle| \nu_t, \alpha_t, \lambda_t\right) = \begin{cases} 1 & d^{ij} = 0 \\ (1 - \nu_t)\exp\left(-\left(\frac{d^{ij}}{\alpha_t}\right)^{\lambda_t}\right) & d^{ij} > 0 \end{cases} \tag{8}$$

where for time step t, α_t is the range parameter, λ_t is the power term and ν_t is the nugget.

In contrast multisite models, such as R2013, are not required to be continuous in space and are therefore more flexible, because they can fit individual lag-0 covariances for all pairs of sites. This may lead to a better fit to observed spatial correlations, but it also leads to a higher number of parameters to represent the observed spatiotemporal correlation structure (Rasmussen 2013). The PRF model's approach of adopting an isotropic, powered exponential function spatial correlation function is a less flexible but more parsimonious assumption which will be tested in Section 6.2.

2.3.3 Spatial parameter interpolation

Based on the enhancements outlined above, the full parameter specification for the PRF is $\boldsymbol{\mu}_t = \{\mu_t^1, \dots, \mu_t^N\}$, $\boldsymbol{\sigma}_t = \{\sigma_t^1, \dots, \sigma_t^N\}$, $\boldsymbol{\beta}_t = \{\beta_t^1, \dots, \beta_t^N\}$, $\bar{\varphi}_t$, ν_t, α_t and λ_t, where the parameter values remain constant for all time steps t in a given month. 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 a technique is required to interpolate these parameter surfaces from the observed sites. A kriging approach is used to produce the parameter surfaces by independently kriging each parameter surface using dependent variables of distance between locations and elevation.

2.3.4 Impact on Model Parsimony

Model parsimony is important for efficient parameter estimation and simulation of spatial rainfall fields to estimate engineering design risks (Section 1). For the PRF model the use of a contemporaneous approach and separable covariance structure, among other assumptions (e.g. spatial correlation function - see Section 2.3 for full list) significantly reduces the number of model parameters required for simulation compared with R2013 model (see Table 1). The PRF model has a major advantage over the R2013 model because it simulates a field continuously across all locations within a given region, whereas the R2013 only simulates at specific sites with observed rainfall. This analysis demonstrates that the PRF model is relatively parsimonious and the model complexity is further discussed in Section 7.

It should be noted that there are alternative methodologies for simulating spatial rainfall fields (as mentioned in Section 1). These alternate modelling approaches are elaborated on in Sections 7.1 and 7.3 and, where feasible, a comparison to the PRF model is undertaken.

3 Model Calibration

Calibration of the model proceeds in a step-wise manner. The first step is the estimation of the marginal distribution parameters (μ, σ and β) 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 to regionalise the latent parameters.

In the first step for estimating the marginal distribution parameters, both the method of moments and the maximum likelihood method are valid. 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 an observed time series of daily rainfall at a site that is partitioned according to a number, n_d, of 'dry' zero values and n_w truncated '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

$$\mathrm{E}[\boldsymbol{L}_w] = \frac{1}{n_w}\sum_{t=1}^{n_w} r_t^{1/\beta} \tag{9}$$

$$\mathrm{E}[\boldsymbol{L}_w^2] = \frac{1}{n_w}\sum_{t=1}^{n_w} \left(r_t^{1/\beta}\right)^2. \tag{10}$$

Consider the left-truncated normal distribution with known truncation point. The parameter $\hat{\beta}$ can be estimated by solving the following two equations according to Johnson et al. (1994, pages 161-2).

$$\frac{\hat{\sigma}}{\mathrm{E}[\boldsymbol{L}_w]} = \left(\frac{\phi(\hat{\delta})}{1-\Phi(\hat{\delta})} - \hat{\delta}\right)^{-1} \tag{11}$$

$$\left(\frac{\hat{\sigma}}{\mathrm{E}[\boldsymbol{L}_w]}\right)\left(\frac{\hat{\sigma}}{\mathrm{E}[\boldsymbol{L}_w]} - \hat{\delta}\right) = \frac{\mathrm{E}[\boldsymbol{L}_w^2]}{\mathrm{E}[\boldsymbol{L}_w]^2} \tag{12}$$

where $\mathrm{E}[\boldsymbol{L}_w]$ and $\mathrm{E}[\boldsymbol{L}_w^2]$ are defined in Eq. (9) and Eq. (10), $\phi(\,)$ is the probability density of the standard normal distribution, $\Phi(\,)$ is the normal cumulative distribution function, and $\hat{\delta} = \hat{\mu}/\hat{\sigma}$, which represents 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. (11). Following this, the left hand side of Eq. (12) is reduced to a constant (see Eq (13)), whilst the right hand side is dependent on the $\hat{\beta}$ parameter through Eq. (9) and Eq. (10).

$$\left(\frac{\phi(\hat{\delta})}{1-\Phi(\hat{\delta})} - \hat{\delta}\right)^{-2} - \hat{\delta}\left(\frac{\phi(\hat{\delta})}{1-\Phi(\hat{\delta})} - \hat{\delta}\right)^{-1} = \frac{\mathrm{E}[\boldsymbol{L}_w^2]}{\mathrm{E}[\boldsymbol{L}_w]^2} \tag{13}$$

Subsequently, Eq. (13) 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

$$min\ ((\mathrm{E}[\boldsymbol{R}_w] - \widehat{m}_w)^2 + (\mathrm{VAR}[\boldsymbol{R}_w] - \hat{s}_w^2)^2) \tag{14}$$

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

$$\widehat{m}_w = \frac{1}{n_w}\sum_{t=1}^{n_w} r_t \tag{15}$$

$$\hat{s}_w^2 = \frac{1}{n_w - 1}\sum_{t=1}^{n_w}(r_t - \widehat{m}_w)^2 \tag{16}$$

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,

$$\mathrm{E}[\boldsymbol{R}_w] = (1 - \hat{p}_d)^{-1}\int_0^\infty r f_R(r)dr \tag{17}$$

$$\mathrm{VAR}[\boldsymbol{R}_w] = (1 - \hat{p}_d)^{-1}\int_0^\infty (r - E[\boldsymbol{R}_w])^2 f_R(r)dr. \tag{18}$$

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

$$f_R(r) = \left(2\pi\hat{\sigma}^2\hat{\beta}^2\right)^{-1/2} r^{(-1+1/\hat{\beta})} exp\left(-\hat{\sigma}^{-2}/2\left(r^{1/\hat{\beta}} - \hat{\mu}\right)^2\right),\ r > 0 \tag{19}$$

In the second step, the lag-1 autocorrelation (l_{t+1}, l_t) is estimated for all pairs of points above the zero threshold and relies on the at-site marginal distribution parameters ($\hat{\mu}$ and $\hat{\sigma}$) for each site and month (from Step 1). This estimate corresponds to an estimate of correlation in the left-truncated bivariate normal distribution (wet day amounts) and can be related to the underlying autocorrelation parameter $\hat{\varphi}_i$ of the non-truncated bivariate distribution (latent-variable) at site i (Weiler 1959). The 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 pairwise 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 pairwise sample spatial correlation in a left-truncated bivariate normal distribution for each pair of sites. The $\hat{\Sigma}^{ij}$ are found by solving the covariance relationship between the truncated and non-truncated Gaussians (Weiler 1959) with known $\hat{\mu}$ and $\hat{\sigma}$. A sum of squared errors approach is then used to minimise the differences between the pairwise sample spatial correlations and the spatial correlation function, Eq. (8), to obtain the parameters $\hat{\nu}, \hat{\alpha}$ 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 marginal parameter estimates ($\hat{\mu}$, $\hat{\sigma}$ and $\hat{\beta}$), obtained from step 1, with dependent variables of distance between locations and elevation. 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. The ability of this kriging approach to produce realistic parameters is tested by comparing the results from full calibration versus via leave-one-out cross-validation (Section 6.3). For the lag-one correlation, a single representative $\hat{\bar{\varphi}}$ for each month is specified as a weighted average of the $\hat{\varphi}_i$ at-site estimates.

4 Comprehensive and Systematic Evaluation Framework

This section describes a comprehensive and systematic evaluation (CASE) framework. It is designed to systematically evaluate the performance of rainfall models against a comprehensive range of observed statistics at multiple spatial scales (individual sites to entire fields) and temporal scales (daily to monthly to annual).

4.1 Description of framework

The CASE framework consists of four steps:

1) Determine a comprehensive range of key observed statistics of interest

For a spatial field or multisite model this range of statistics should assess both the temporal and spatial properties. For example, daily statistics, annual totals, extremes and spatial rainfall gradient may be targeted for evaluation. The range of statistics evaluated in this paper is outlined in Section 4.2.1.

2) Systematically categorise performance at specific spatial and temporal scales using quantitative criteria for each statistic

For example, this paper applies a three level categorisation system, where model performance of a single statistic for given spatial or temporal scale was placed into one of three categories; 'good', 'fair' and 'poor' performance (see Section 4.2.2).

3) Systematically categorise 'aggregate' performance over multiple spatial and/or temporal scales using quantitative criteria, informed by Step 2, for each statistic

This enables an assessment of common strengths and/or weaknesses in the models ability to reproduce a particular statistic over multiple spatial and temporal scales. For example, this paper the 'aggregate' performance is based on the percentage of cases across multiple sites or months which are classified as 'good', 'fair' or 'poor' in the first stage (see Section 4.2.3).

4) Comprehensively evaluate performance in both calibration and LCV

Both evaluations are essential because calibration will identify deficiencies in model structure where the model has access to the full set of available data. Whereas a comparison of calibration and LCV performance will enable the identification of model overfitting. The quantitative categorisation approach of Steps 2 and 3 enables for easy side-by-side comparison of performance of the comprehensive set of statistics in calibration and LCV.

One of the main advantages of the CASE framework is that categorising performance for each statistic for a range of spatial and temporal scales means that it provides a systematic and transparent method to analyse the multitude of results. By using a quantitative approach to categorise model performance it reduces the often used, ad-hoc nature of descriptive assessments, (words such as 'adequate', 'suitable', or 'reasonable'). The disadvantage of performance categories is that even with quantitative criteria to define categories, an element of judgement/subjectivity is required when choosing the number of categories and defining the difference between the categories (e.g. Evin et al. 2014). For example, in this paper, we use three categories, 'good', 'fair' and 'poor' performance in Step 2 (see Section 4.2.2). The difference between 'fair' and 'poor' is somewhat subjective (see Section 7.2 for a discussion of the impact of this subjectivity). Ultimately, what constitutes the differences the types of performance will depend on the user preferences and/or on the practical application of the model (this is further discussed in Section 7.2). Despite this element of subjectivity, this categorization of performance is far more transparent, consistent and less subjective than the usual ad-hoc descriptive assessment, commonly employed in other studies (Baxevani and Lennartsson 2015; Leonard et al. 2008; Rasmussen 2013).

4.2 Implementation of Framework

This section describes the implementation of the CASE framework to assess the PRF model. Other models developed for different contexts can follow the same framework steps but, depending on the practical application, may need to rely on different choices in the detailed implementation of each step.

4.2.1 Determining Key Observed Statistics of Interest

The first step of implementing the CASE framework in the context of evaluating the PRF involves choosing the key observed statistics of interest. This case study uses a range of statistics at different spatial and temporal scales. At the individual site scale, the temporal scales include: daily, monthly and annual time scales. At the regional scale, both daily and annual scales were evaluated.

The evaluation of individual site scale performance focuses primarily on the following temporal statistics listed below. At the daily scale the following statistics are evaluated:

- Mean and standard deviations of number of wet days and wet day amounts as well as skewness of amounts to evaluate if the marginal distribution of daily rainfall occurrences and amounts are being preserved.
- Wet and dry spell length distributions – where a ‘spell’ is a block of consecutive time steps having the same ‘wet’ or ‘dry’ state – to evaluate if the wet and dry rainfall intermittence/ auto-correlation is being preserved.

Monthly and annual scale statistics are important for preserving seasonal characteristics and inter-annual variability. Whilst rainfall extremes are an important feature to reproduce for flood frequency applications. At the monthly and annual scale the following statistic are evaluated:

- Distributions of monthly total rainfall, annual total rainfall and the number of wet days annually are presented. 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 - 5th percentile, upper tail indicator - 95th percentile).

- Monthly temporal correlations (e.g. January-to-February) are evaluated to further understand the structure of variability in the annual rainfall.
- Annual temporal correlations are evaluated to assess whether the model reproduces inter-annual variability.
- The distribution of daily annual maxima are assessed which are an emergent property of the model.

There are three parts to the evaluation at the regional scale. The following tests are presented:

- The distributions of the number of jointly wet sites for each month are compared.
- 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 aggregated series for the observed rainfall and the simulated rainfall are compared using the aforementioned metrics to assess at-site rainfall statistics (see Section 4.2.4).
- The spatial rainfall gradient produced by the model is evaluated by comparing the field of average annual total rainfalls produced by the model against interpolated observed annual rainfalls.

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.

4.2.2 Selection of Performance Categories at Specific Temporal and Spatial Scales

To implement Step 2 of the framework this paper categorises the performance of each evaluated statistic as one of three categories; ‘good’, ‘fair’ and ‘poor’ performance.

Table 2 summarises the quantitative tests for each performance category with accompanying examples. ‘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. 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. 1, case (2)), or the absolute relative difference between the observation and the simulated mean is less than 5% (Fig. 1, case (3)). The absolute relative difference is calculated as

$$RD = |100\,(x_{obs} - E[X_{sim}])/x_{obs}| \tag{20}$$

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. 1, case (4)).

4.2.3 Selection of ‘Aggregate’ Performance Categories Over Multiple Temporal and Spatial Scales

To implement Step 3 of the framework this paper categorises the aggregate performance of each evaluated statistic as one of six categories. Table 3 details the aggregate performance categories, which range from ‘Overall Good’ to ‘Overall Poor’, and the quantitative tests used to determine them. The tests are based on the percentage of cases (e.g. sites/months) which are categorised as ‘good’, ‘fair’ or ‘poor’. For example, ‘Overall Variable’ occurs when the percentage of cases classified as ‘good’ and ‘poor’ is greater than the percentage of cases deemed ‘fair’.

4.2.4 Comparison of Calibration and LCV Performance

A LCV of both the parameters predicted by the kriging and at-site model performance is conducted to assess the error associated with spatial interpolation. The LCV was performed by calibrating to all sites, except for the one validation site. Kriging was then used to estimate the parameters at the validation site. The estimated surfaces (μ^*, σ^*, β^*) at the validation site are compared against the calibrated parameter values (μ, σ, β). Rainfall time series simulated at the validation site is then evaluated for all at-site statistics listed in Section 4.2.1. This process was then repeated so that each site in turn was treated as the validation site.

5 Case Study

The Onkaparinga catchment in South Australia is used as a case study (Fig. 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.

There are 22 daily rainfall gauges within and surrounding the Onkaparinga catchment (Fig. 2 and Table 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 erroneous trends and data inhomogeneities (Westra et al. 2014) by comparing against the Happy Valley site (23721) which is part of the high quality network of gauges. None of the sites showed strong evidence of erroneous trends or data in-homogeneities. From Fig. 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 19) at elevation 7 m up to 1088 mm at Uraidala (Site 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).

Nineteen rainfall gauges lie inside the boundary of the Onkaparinga catchment and are used for model calibration and evaluation. The three gauges that lie outside the catchment are used in calibration only to reduce edge effects due to the spatial interpolation of parameters. The simulation experiment consisted of 100 replicates using 0.88 km square grids over the case study region.

6 Results

The results present a wide range of statistics as described in Section 4.2.1, and uses the performance categorisation system from Section 4.1. Section 6.1 assesses at-site performance of the model, Section 6.2 assesses spatial field performance of the model and Section 6.3 presents the performance in LCV. Table 5 shows a summary of results across all rainfall sites in calibration and LCV. 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 located in Supplementary Material A—F.

6.1 At-site performance in calibration

6.1.1 Daily rainfall occurrence and amounts

'Overall Good' performance is shown for the mean, standard deviation and skewness of wet day rainfall amounts, and the mean number of wet days (Table 5 and Fig. 3). This shows the model reproduces the observed daily marginal rainfall statistics. However, the model under-predicts standard deviation 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. 3d).

Table 5 shows model performance in simulating the wet spell and dry spell distributions to be 'Overall Fair - Good' and 'Overall Good' respectively. This suggests the use of a spatially constant autoregressive parameter for each month yields 'Overall Fair– Good' performance in producing realistic temporal patterns and rainfall persistence. Fig. 4 shows the model performance in simulating wet spell and dry spell length distributions on a seasonal basis (Fig. 4 a-b) and for illustrative sites/months (Fig. 4 c-f). Specifically, the model shows 'Overall Good' performance in simulating the wet spell length distribution for the autumn and winter months (MAMJJA) and 'Overall Fair – Good' performance for the spring and summer months (SONDJF) (Fig. 4a). The model shows 'Overall Good' performance in simulating dry spell lengths (Fig. 4b).

The majority of instances in which performance was categorised as 'poor' occurred in February or November when the catchment has very little rainfall (e.g. Fig. 4e). The 'poor' categorisation typically resulted from an over-estimation of short duration wet spells and an under-estimation for longer durations. The potential reasons for this weakness are discussed in Section 7.2.

6.1.2 Monthly and annual statistics

Table 5 shows the model performance to be 'Overall good' in simulating the distribution of total monthly rainfall amounts (mean, standard deviation, lower and upper tails). Quantile-quantile plots of performance are shown in Fig. 5. Although, the simulated standard deviation of monthly rainfall totals is 'poor' for some sites with higher monthly standard deviation in June (Fig. 5b).

Table 5 shows that the model exhibits 'Overall Good' performance in simulating the mean and upper tail of the total annual rainfall distribution (see Fig. 6a and d). However, the model underestimates the variability of the total annual rainfall, exhibiting 'Overall Fair – Poor' performance for the standard deviation (Fig. 6b). This is because model does not reproduce the rainfall in drier years. This is seen in the 'Overall Fair – Poor' performance in simulating the lower tail of the total annual rainfall, with the simulated rainfall being larger than the observed by on average 15% (Fig. 6c). Whilst the upper tail performance is 'Overall Good'.

The simulation demonstrates 'Overall Good' performance annually in simulating wet day amounts (means and standard deviations) and the mean annual number of wet days (Table 5). The model shows 'Overall Poor' performance in simulating the variance in the number of wet days annually because the annual variance is underestimated (Table 5 and Supplementary Material A). This suggests that the over-estimation of the lower tail of the annual total rainfall distribution (Fig. 6c) is due to the deficiency that the variance in the number of wet days annually is under-estimated, rather than a problem with rainfall amount generation.

6.1.3 Temporal correlation of annual and monthly totals

At the annual scale, the observed correlation between consecutive annual rainfall totals has 'Overall Good' performance (Table 5) as there is little inter-annual persistence for rainfall in this catchment. The model does not include month-to-month correlation, thus the simulations are centred on zero. As there is low monthly persistence in this catchment, the correlations between monthly consecutive rainfall totals show 'Overall Good' performance for all sites and months (Table 5). However, correlations in the consecutive totals for June and July show a number of sites that are deemed 'fair' (Supplementary Material B). 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.

6.1.4 Daily rainfall extremes

The model exhibited 'Overall Good' performance in reproducing the daily annual maximas (Table 5). Fig. 7 shows a comparison of the observed and simulated daily annual maximum rainfall for example sites Coromandel Valley (site 6, 'fair'), Cherry Gardens (site 4, 'poor'), and summarises the aggregate performance. Fourteen sites (53%) across the catchment the model showed 'good' or 'fair' performance in reproducing the distribution of daily annual maxima (Fig. 7).

6.2 Spatial field performance

6.2.1 Multi-site occurrences

The top panel of Fig. 8 illustrates three categories of spatial rainfall coverage: 'sparse' rain (Fig. 8a), 'patchy' rain (Fig. 8b) and 'dense' rain (Fig. 8c). 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. 8a and b) but is deemed 'Overall Variable' for the 'dense' rain category (Fig. 8c) due to the model over-predicting the number of instances in which all 19 sites were wet.

6.2.2 Catchment rainfall

The domain aggregated behaviour of the sites can be assessed using the same statistics as for individual sites. This approach showed that the domain aggregated behaviour had 'Overall Fair – Good' performance when reproducing the same statistics evaluated in the at-site analysis (see Section 4.2.1). As with the individual sites, the model showed poorer performance in reproducing the lower-tail of the annual total rainfall distributions (see Supplementary Material C), but this was to be expected.

The interpolated mean observed annual rainfall fell within the 90% limits of the simulated mean annual rainfall for 78% of the region indicating 'Overall 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 suggest a potential limitation of the interpolation approach.

6.3 Leave-one-out cross-validation performance

The model was evaluated using a LCV approach (Section 4.2.4). There was minimal difference between the observed and predicted parameters over the region (see Supplementary Material D), suggesting that the regression against elevation and the variogram parameters are appropriate. This is further assessed by comparing the model's at-site performance calibrated using all data against the LCV at-site performance. Table 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 LCV shows some decrease in performance, but this decrease typically occurs when sites nearer the boundary (e.g. Site 11) are left out and there is little other nearby information to assist the interpolation. This issue is a property of the spatial interpolation component of 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'.

Several statistics are worth noting in which changes in performance were observed between calibration and LCV. Mean annual rainfall total performance changed from 'Overall Good' in calibration to 'Overall Poor' in LCV. However, the relative difference between the simulated and observed mean annual total rainfall is within 10% for 14 sites (Table 6). This issue is due to the small uncertainty in the simulated mean annual total rainfall, such that changes to the interpolated mean can easily lie outside the 90% limits (see Supplementary Material E). Likewise, the same issue occurred for the simulated mean annual number of wet days, which dropped from 'Overall Good' to 'Overall Fair – Good', and mean monthly total rainfall, which dropped from 'Overall Good' to 'Overall Variable' between calibration and LCV (Table 6 and Supplementary Material F).

The simulation of the variability in annual total rainfall changes from 'Overall Fair – Poor' to 'Overall Variable' 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 daily annual maximas.

7 Discussion

The latent-variable approach used in this study has a number of features that make it more parsimonious than existing approaches (see Section 2.3.4 for a full description). Firstly, it implicitly accounts for temporal correlations in the wet-dry pattern (Section 6.2) as well as the rainfall amounts (Section 6.2.2) 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 (see Section 2.3.2 for discussion).

7.1 Differences with existing spatial rainfall field models using latent-variables

Other latent-variable (LV) models have been used to generate spatial rainfall fields suitable for continuous hydrological simulation of a catchment. A key difference between existing LV approaches and the PRF model is that a different transformation approach is used. For example, Baxevani and Lennartsson (2015) have adopted a composite transformation that applies two different transformation functions to extreme and non-extreme rainfall amounts. In contrast the PRF model, uses a single power transformation function across the entire range of rainfall (e.g. Eq (1)), which requires less parameters to be estimated than the composite transformation. Another key differences is that some LV approaches use a different approach to handle seasonality in rainfall than the PRF's approach by of monthly parameters. For example, contemporary models have used parameters that vary by day achieved by defining a cyclic relationship between the parameters and the day of the year (Baxevani and Lennartsson 2015; Kleiber et al. 2012). Depending on how many parameters are required for the cyclic relationship this could potentially lead to fewer parameters than the vary by month approach adopted in the PRF model. Whether these key differences result in better reproduction of rainfall statistics is difficult to determine without a comparison using the same catchment.

7.2 Interpretation of performance results.

This section interprets the performance results and discusses model strengths and weaknesses with respect to the model assumptions.

The at-site performance evaluation showed the wet/dry occurrences, rainfall amounts, and extremes evaluations to be 'Overall Good' (Table 5). This indicates that the underlying latent-variable model is sufficient for reproducing the marginal statistics. In this study a power transformation was adopted, while more complex composite transformations have been used (Baxevani and Lennartsson 2015), the 'Overall Good' performance seen in the comprehensive evaluation did not suggest that a more complicated composite transformation was needed. The 'Overall Good' performance in simulating extremes is a benefit of the modelling approach as 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).

The model exhibited 'Overall Fair – Good' performance in simulating wet spell durations for some sites. However, the model exhibited 'poor' performance in drier months, November and February (Section 6.1.1). The 'poor' categorisation typically arose due to the over-estimation of short duration wet spells and under-estimation for longer durations. This difficulty in reproducing the wet spell distribution for drier months may be a limitation of the AR(1) model and/or a consequence of applying a single homogeneous AR(1) parameter, φ_t, for both dry and wet spells. The model also under-estimated the variability in the number of wet days simulated in these months (November and February). Whether the difficulties matter in terms of hydrological model performance is another question, because these months contribute very little to annual total flow in this catchment.

The model shows 'Overall Good' performance in simulating the number of jointly wet sites. Although the model over-predicted the frequency of days where rainfall is observed at all sites. Baxevani and Lennartsson (2015) similarly noted the higher probability of observing rainfall at all sites (right hand side of Fig. 8c) compared to partial coverage of the region, but their model under-predicted instances where the sites were either all dry or all wet. The 'Overall Good' performance suggests applied spatial correlation function is sufficient.

A lack of variability at annual scales was observed and identified as a model 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 evaluation showed that inter-annual correlations were 'Overall Good' (Section 6.1.3) 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 likely cause of a lack of variability in annual total rainfall amounts (See Section 6.1.2). Specifically, the model showed 'poor' performance in simulating drier years.

7.3 Benefits of comprehensive and systematic evaluation framework

As demonstrated above the CASE framework enabled the identification of model deficiencies and the attribution of these deficiencies to specific model features. Specifically, the CASE framework demonstrated that a difficulty in simulating variability in the number of wet days annually was a likely cause of the lack of variability in annual total rainfall amounts. 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' 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 CASE framework has been a direct comparison between the performance in calibration and spatial LCV. This is rare in studies that present spatial continuous simulation approaches. Overall there was not a large change in performance for the spatial LCV, which provides greater confidence in model performance ability. The largest differences in LCV were for locations with less adjacent surrounding gauges or higher elevations - approaches to remedy this are discussed in the following section. The CASE framework could be extended to undertake a temporal LCV analysis (Wang and Robertson 2011; Wang et al. 2009) in addition to the specified spatial LCV. However, this was not undertaken here as the focus of this study was to evaluate the parameter interpolation scheme, a key new component of the PRF model.

A future benefit of the CASE framework will be the comparison of different rainfall field models at a given catchment of interest. There are many varied approaches to simulating spatial rainfall fields, ranging from cluster point processes (Burton et al. 2010; Leonard et al. 2008) to random field models (Paschalis et al. 2013) and disaggregation based models (Jothityangkoon et al. 2000). Each relies on different mechanisms to simulate precipitation in space and time. Due to these stark differences it is difficult to compare the models based on model structure alone. Therefore a better approach is to compare them based on their ability to reproduce the key observed rainfall statistics for the same catchment, however, these comparisons are rarely undertaken. The key issues being that, until now, there has been no systematic approach to achieve this on a comprehensive range of key statistics. The CASE framework overcomes this issue and enables future studies to be undertaken to compare and evaluate spatial rainfall field models.

7.4 Future PRF model developments

The CASE framework identified the variance in annual totals and occurrences as being a limiting feature of the PRF model for the given case study. Future versions of the PRF 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 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 PRF model across different regions and 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.

Additionally, the model will be compared against contemporary spatial rainfall field models (e.g. cluster point processes, (Burton et al. 2010; Leonard et al. 2008) using the CASE framework to systematically identify model strengths and weakness on a wide range of catchments.

8 Conclusions

The first 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 simulation of the temporal and spatial correlation structure. The second goal was to develop a comprehensive and systematic evaluation framework. The framework was developed using a performance categorisation system to provide a systematic, succinct and transparent method to assess and summarise model performance over a comprehensive range of statistics, sites, scales and seasons. Importantly it was able to identify and diagnose PRF 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 observed rainfall characteristics, with the majority of statistics categorised 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. The CASE framework was able to identify that the source of this over-estimation was poor representation of the annual variability of rainfall occurrences.

Further research will address model weaknesses, and then apply the model in different regions using the comprehensive and systematic evaluation framework to identify if further enhancements are required. Given the strengths of the 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.

9 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 (Rasmussen 2013).

10 References

Allcroft, D. J., and 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.

Bardossy, A., and Plate, E. J. (1992). "Space-time model for daily rainfall using atmospheric circulation patterns." *Water Resources Research*, 28(5), 1247-1259.

Baxevani, A., and 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.

Bennett, B., Lambert, M., Thyer, M., Bates, B., and Leonard, M. (2015). "Estimating Extreme Spatial Rainfall Intensities." *Journal of Hydrologic Engineering*, 0(0), 04015074.

Box, G., and Jenkins, G. (1976). *Time series analysis: forecasting and control*, Holden-Day Inc., California.

Burton, A., Fowler, H. J., Kilsby, C. G., and O'Connell, P. E. (2010). "A stochastic model for the spatial-temporal simulation of nonhomogeneous rainfall occurrence and amounts." *Water Resources Research*, 46(11), n/a-n/a.

Candela, L., Tamoh, K., Olivares, G., and 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.

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

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). "Statistical Modeling of Spatial Extremes." 161-186.

Durban, M., and Glasbey, C. A. (2001). "Weather modelling using a multivariate latent Gaussian model." *Agricultural and Forest Meteorology*, 109(3), 187-201.

Evin, G., Thyer, M., Kavetski, D., McInerney, D., and 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.

Gneiting, T. (2002). "Compactly Supported Correlation Functions." *Journal of Multivariate Analysis*, 83(2), 493-508.

Groppelli, B., Bocchiola, D., and 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.

Heneker, T. M., Lambert, M. F., and Kuczera, G. (2001). "A point rainfall model for risk-based design." *J. Hydrol.*, 247(1-2), 54-71.

Hundecha, Y., Pahlow, M., and 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., and Beswick, A. R. (2001). "Using spatial interpolation to construct a comprehensive archive of Australian climate data." *Environ. Modell. Softw.*, 16(4), 309-330.

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

Jothityangkoon, C., Sivapalan, M., and Viney, N. R. (2000). "Tests of a space-time model of daily rainfall in southwestern Australia based on nonhomogeneous random cascades." *Water Resources Research*, 36(1), 267-284.

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

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

Kleiber, W., Katz, R. W., and Rajagopalan, B. (2012). "Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes." *Water Resources Research*, 48, 17.

Kwon, H.-H., Sivakumar, B., Moon, Y.-I., and 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.

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 Resources Research*, 44(9), W09402.

Li, C., Singh, V. P., and 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., and Metcalfe, A. (2014). "An efficient causative event-based approach for deriving the annual flood frequency distribution." *J. Hydrol.*, 510, 412-423.

Mehrotra, R., and 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.

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.

Onof, C., and Wheater, H. S. (1993). "Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulse Model." *J. Hydrol.*, 149(1), 67-95.

Paschalis, A., Molnar, P., Fatichi, S., and Burlando, P. (2013). "A stochastic model for high-resolution space-time precipitation simulation." *Water Resources Research*, 49(12), 8400-8417.
Paschalis, A., Molnar, P., Fatichi, S., and Burlando, P. (2013). "A stochastic model for high-resolution space-time precipitation simulation." *Water Resources Research*, 49(12), 8400-8417.
Qin, J. (2010). "A High-Resolution Hierarchical Model for Space-time Rainfall." Doctor of Philosophy PhD, University of Newcastle, Newcastle.

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

Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., and Franks, S. W. (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., and 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.

Sanso, B., and Guenni, L. (2000). "A nonstationary multisite model for rainfall." *Journal of the American Statistical Association*, 95(452), 1089-1100.

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

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

Sharma, A., and Mehrotra, R. (2013). "Rainfall Generation." *Rainfall: State of the Science*, American Geophysical Union, 215-246.

Srikanthan, R., and 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., and Pegram, G. G. S. (2009). "A nested multisite daily rainfall stochastic generation model." *J. Hydrol.*, 371(1–4), 142-153.

Wang, Q., and Robertson, D. (2011). "Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences." *Water Resources Research*, 47(2).

Wang, Q., Robertson, D., and Chiew, F. (2009). "A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites." *Water resources research*, 45(5).

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., and 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 Technical Report Series* 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.

Wilks, D. S. (1999). "Interannual variability and extreme-value characteristics of several stochastic daily precipitation models." *Agricultural and Forest Meteorology*, 93(3), 153-169.

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

Zhang, Z., and Switzer, P. (2007). "Stochastic space-time regional rainfall modeling adapted to historical rain gauge data." *Water Resources Research*, 43(3), W03441.

11 Tables

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

No. locations modelled	R2013 multi-site model	PRF Spatial rainfall field model	Reduction in parameters
N	$(3N+2N^2)$	16	$100[(3N+2N^2)-16]/(3N+2N^2)$
25	1,325	16	99%
100	20,300	16	~100%
2048*	8,394,752	16	~100%

* case study Onkaparinga catchment

Table 2 Performance categorisation criteria

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

Table 3 Aggregate performance categorisation criteria

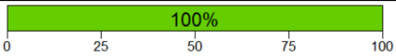
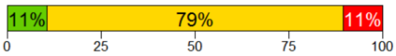
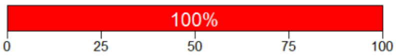
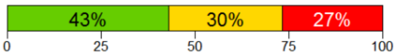
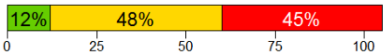
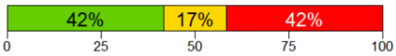
Overall Performance Categorisation	Test	Example
'Overall Good'	'good' > 50 %	
'Overall Fair'	'fair' > 50 %	
'Overall Poor'	'poor'> 50%	
'Overall Fair – Good'	'fair' & 'good' > 'poor'	
'Overall Fair – Poor'	'fair' & 'poor'> 'good'	
'Overall Variable'	'good' & 'poor' > 'fair'	

Table 4 Site names and locations

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

Table 5 Comparison of calibration and LCV performance. Aggregate performance measure summarised to the right of each bar using the Table 3 categorisation scheme.

Metric	Calibration: Percent of cases	Calibration: Aggregate performance	LCV: Percent of cases	LCV: Aggregate performance
Monthly				
Wet day amounts – means	100%	Overall Good	65% / 19% / 15%	Overall Good
Wet day amounts – std dev	100%	Overall Good	75% / 18%	Overall Good
Wet day amounts – skew	77% / 20%	Overall Good	77% / 18%	Overall Good
No. wet days – means	100%	Overall Good	64% / 25% / 11%	Overall Good
No. wet days – std dev	44% / 24% / 32%	Overall Variable	43% / 27% / 30%	Overall Variable
Wet spell distribution	43% / 46% / 11%	Overall Fair – Good	43% / 46% / 11%	Overall Fair – Good
Dry spell distribution	65% / 35%	Overall Good	65% / 35%	Overall Good
Total rainfall - means	100%	Overall Good	44% / 21% / 35%	Overall Variable*
Total rainfall – std dev	69% / 28%	Overall Good	61% / 31%	Overall Good
Total rainfall – lower tail	80% / 18%	Overall Good	79% / 15%	Overall Good
Total rainfall – upper tail	78% / 20%	Overall Good	65% / 25%	Overall Good
Annual				
Total rainfall - means	100%	Overall Good	11% / 32% / 58%	Overall Poor*
Total rainfall – std dev	16% / 47% / 37%	Overall Fair – Poor	37% / 16% / 47%	Overall Variable
Total rainfall – lower tail	21% / 47% / 32%	Overall Fair – Poor	26% / 26% / 47%	Overall Fair – Poor
Total rainfall – upper tail	53% / 47%	Overall Good	53% / 32% / 16%	Overall Good
Wet day amounts - mean	100%	Overall Good	16% / 42% / 42%	Overall Fair – Poor*
Wet day amounts – std dev	100%	Overall Good	84% / 16%	Overall Good
No. wet days – means	100%	Overall Good	37% / 42% / 21%	Overall Fair – Good*
No. wet days – std dev	100%	Overall Poor	5% / 95%	Overall Poor
Correlations				
Monthly total rainfall	89% / 11%	Overall Good	85% / 15%	Overall Good
Annual total rainfall	100%	Overall Good	100%	Overall Good
Extremes				
Daily annual maxima	53% / 21% / 26%	Overall Good	37% / 16% / 47%	Overall Variable*

* The statistics are very sensitive to the choice of relative difference for the categorisation system. This is discussed in Section 6.3 and further explored in Table 6.

Table 6 Comparison of LCV aggregate performance with relative distance set at 5% and 10%.

	LCV (Percent of cases)			
Metric	**Fair relative difference 5%**	**Aggregate performance**	**Fair relative difference 10%**	**Aggregate performance**
Monthly total rainfall - means	44% 21% 35%	Overall Variable	44% 37% 19%	Overall Fair - Good
Annual total rainfall - means	11% 32% 58%	Overall Poor	11% 64% 25%	Overall Fair
Annual daily rainfall amounts - means	16% 42% 42%	Overall Fair-Poor	16% 73% 11%	Overall Fair

12 Graphics

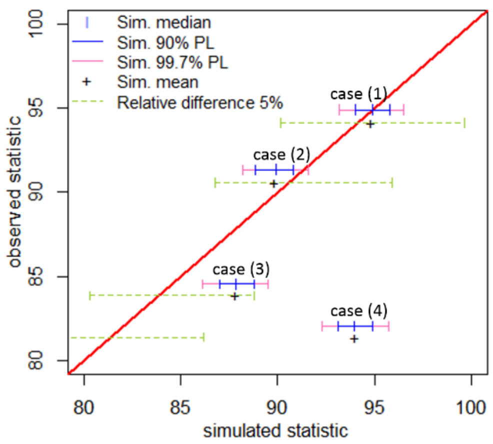


Fig. 1 Illustration of performance categorisation, case (1) shows 'good' performance, cases (2) and (3) show 'fair' performance and case (4) shows 'poor' performance.

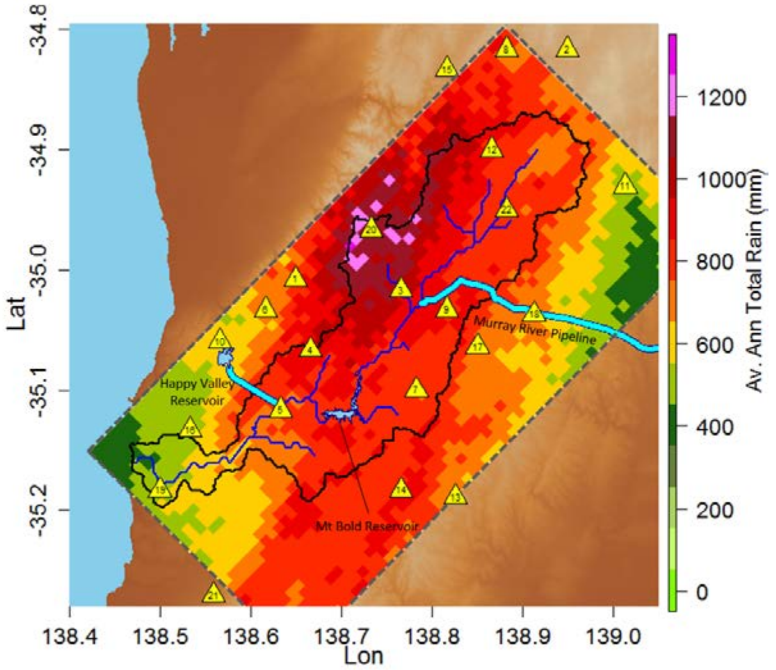


Fig. 2 Locations of rainfall observation sites, Onkaparinga catchment and study region.

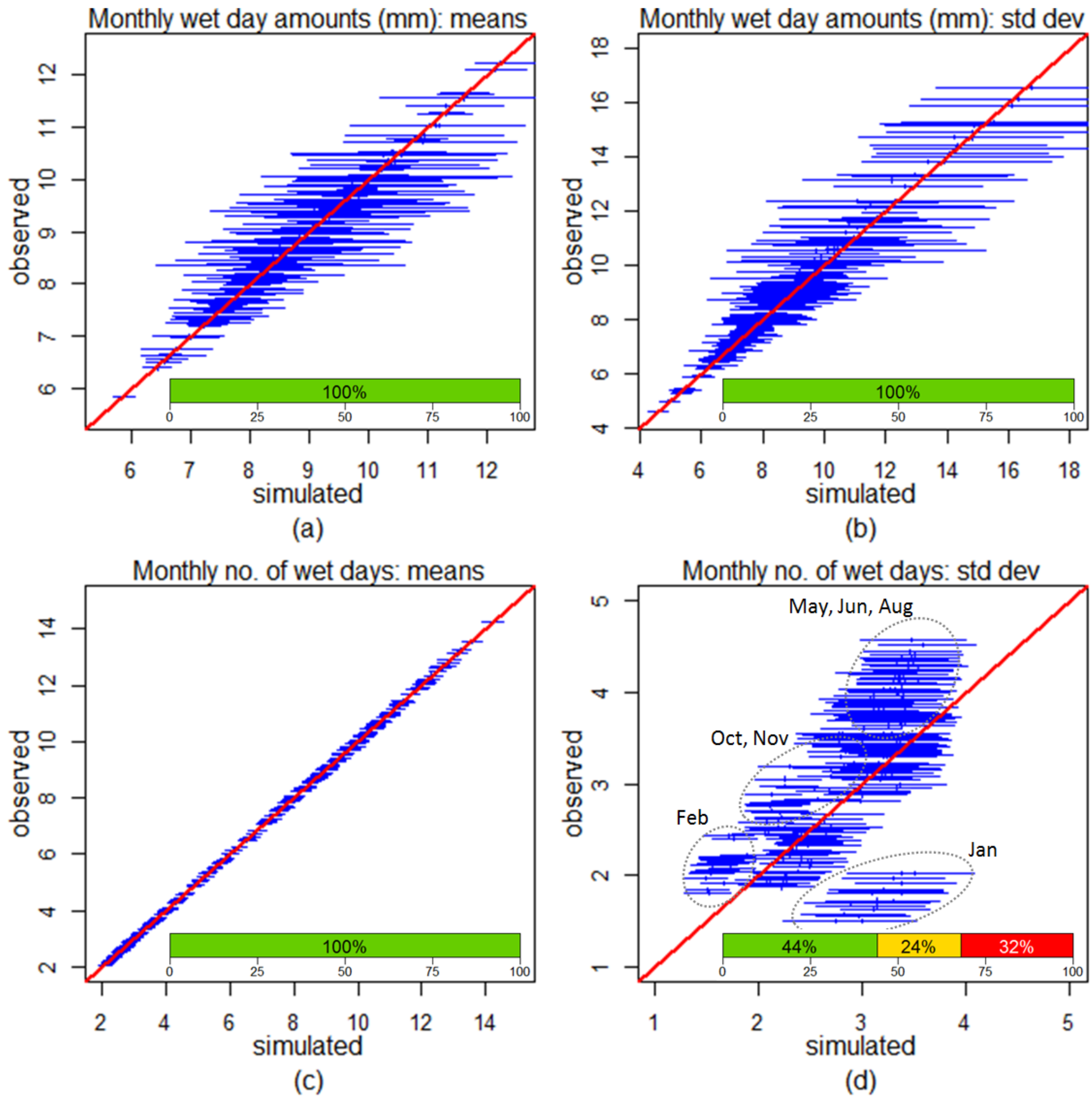


Fig. 3 At site daily statistics for all sites and months, 90% probability limits shown, overall performance shown as a percentage of all sites and months.

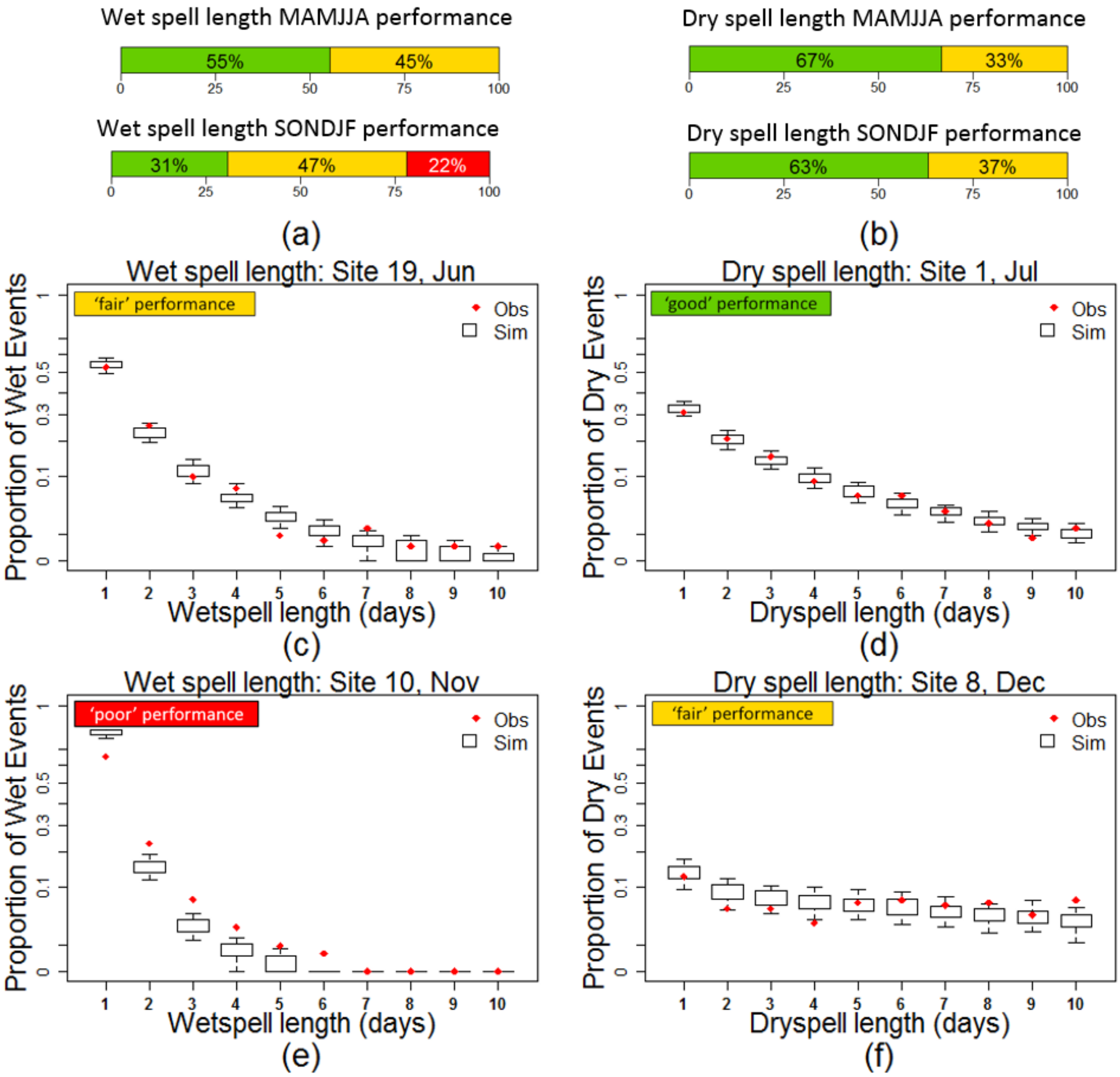


Fig. 4 Distribution of event lengths (a) wet spell length distribution summary, (b) dry spell length distribution summary, (c) wet spell length distribution Site 19 June, (d) dry spell length distribution Site 1, July, (e) wet spell length distribution Site 10 November, and (f) dry spell length distribution Site 8, December, 90% probability limits shown.

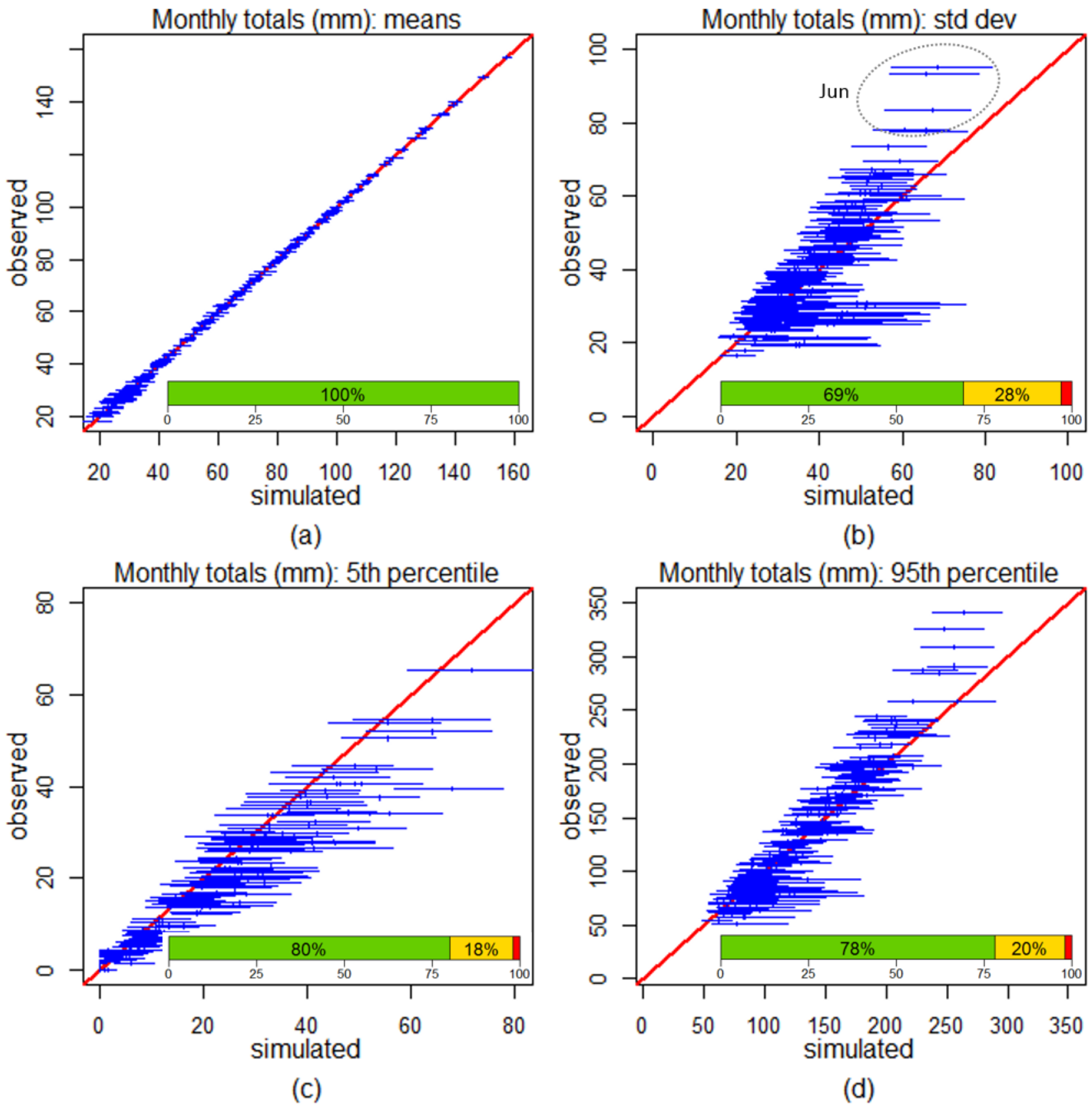


Fig. 5 At site monthly totals for all sites and months (a) means, (b) standard deviations, (c) lower 5th percentile and (d) upper 95th percentile, 90% probability limits shown, overall performance as a percentage of all sites and months.

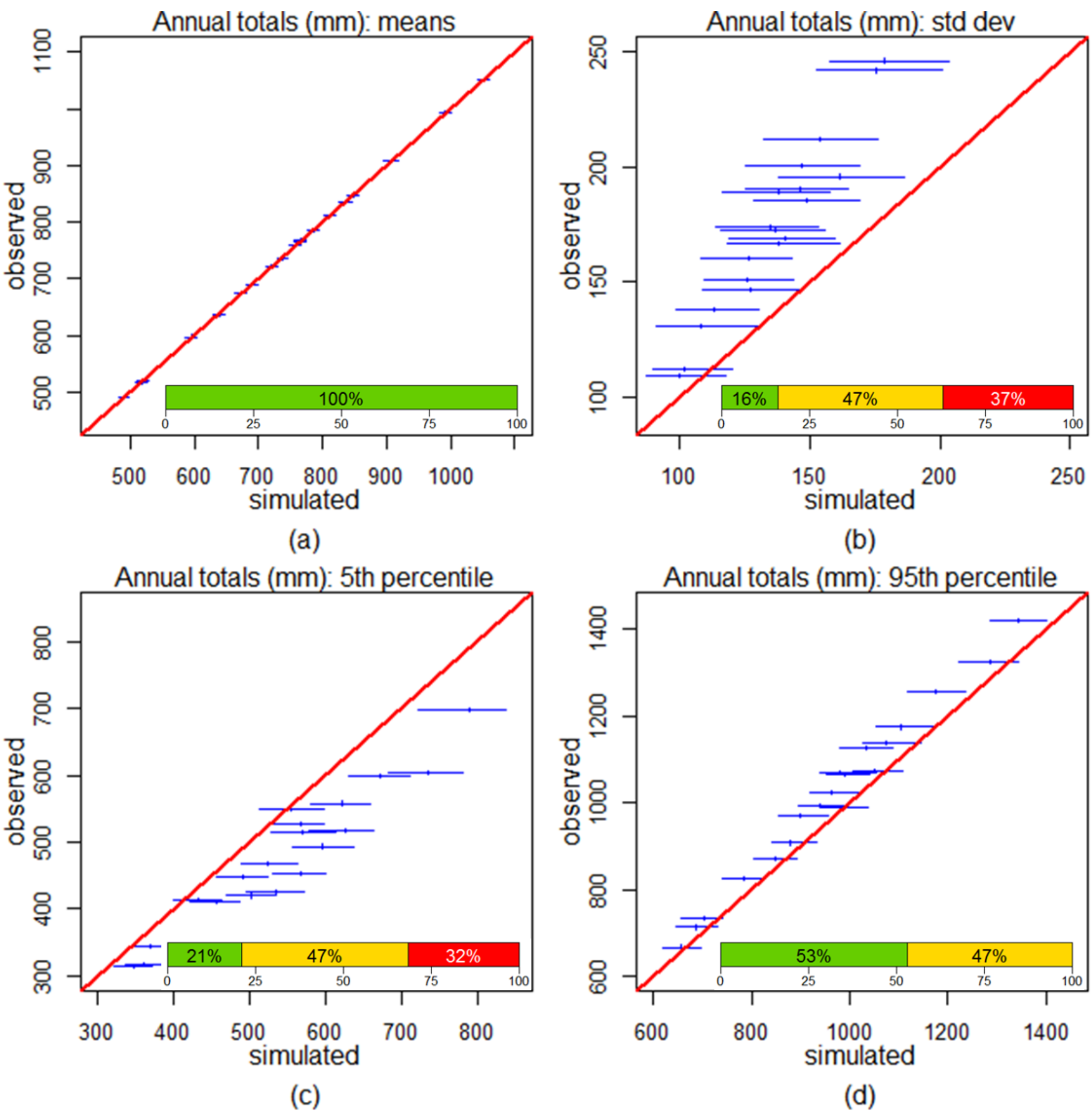


Fig. 6 At site annual totals for all sites and months (a) means, (b) standard deviations, (c) 5^{th} percentile and (d) 95^{th} percentile, 90% probability limits shown, overall performance as a percentage of sites.

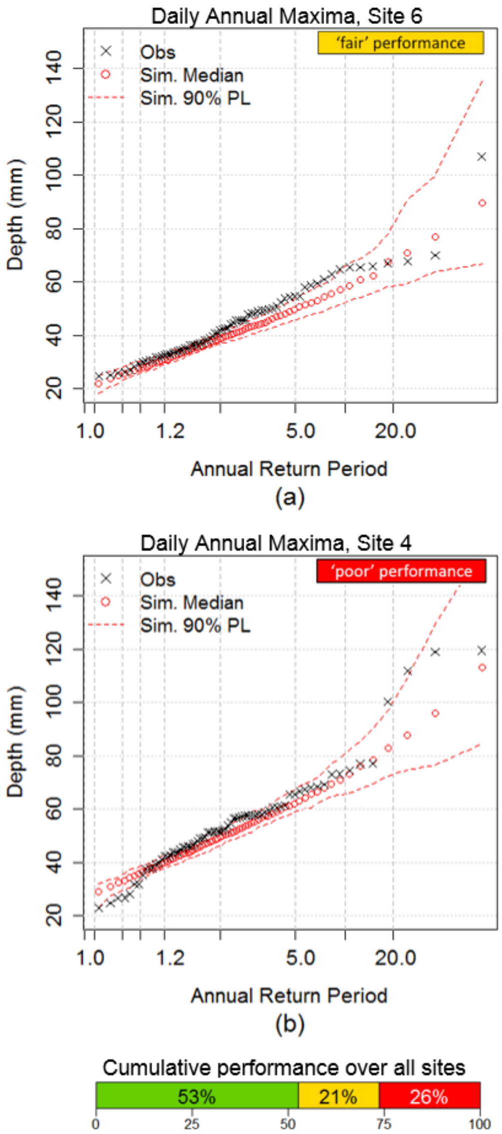


Fig. 7 Simulated and observed daily annual maxima (a) example from site 6, (b) example from site 4 and overall performance as a percentage of sites.

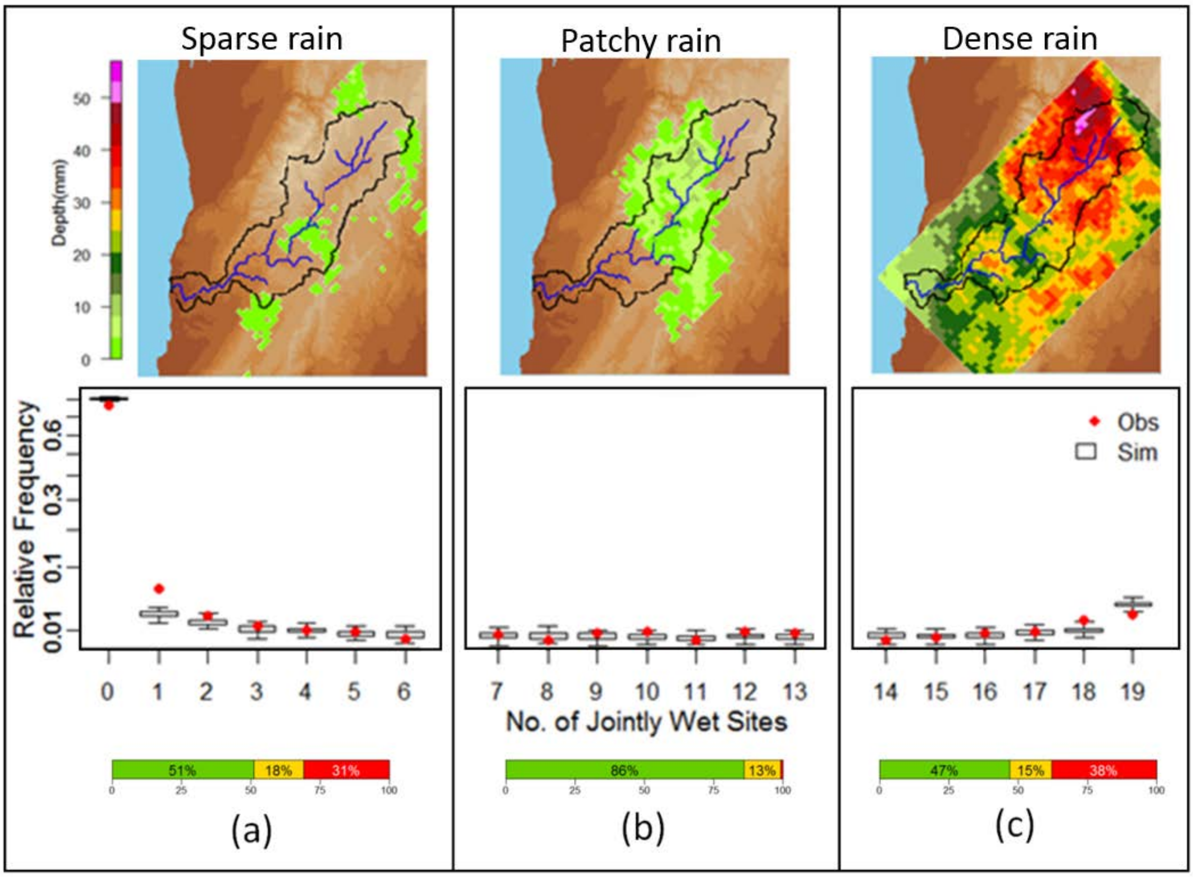


Fig. 8 Distribution of number of jointly wet sites for (a) ‘sparse’ rain, (b) ‘patchy rain’ and (c) ‘dense’ rain. Example shown for March. Overall performance shown as a percentage of months and options within a category.