Continuous rainfall simulation: 2. A regionalized daily rainfall generation approach

Rajeshwar Mehrotra,[1] Seth Westra,[2] Ashish Sharma,[1] and Ratnasingham Srikanthan[3]

Received 28 January 2011; revised 17 November 2011; accepted 28 November 2011; published 25 January 2012.

[1] This paper is the second of two in the current issue that presents a framework for simulating continuous (uninterrupted) rainfall sequences at both gaged and ungaged locations. The ultimate objective of the papers is to present a methodology for stochastically generating continuous subdaily rainfall sequences at any location such that the statistics at a range of aggregation scales are preserved. In this paper we complete the regionalized algorithm by adopting a rationale for generating daily sequences at any location by sampling daily rainfall records from "nearby" gages with statistically similar rainfall sequences.The approach consists of two distinct steps: first the identification of a set of locations with daily rainfall sequences that are statistically similar to the location of interest, and second the development of an algorithm to sample daily rainfall from those locations. In the first step, the similarity between all bivariate combinations of 2708 daily rainfall records across Australia were considered, and a logistic regression model was formulated to predict the similarity between stations as a function of a number of physiographic covariates. Based on the model results, a number of nearby locations with adequate daily rainfall records are identified for any ungaged location of interest (the "target" location), and then used as the basis for stochastically generating the daily rainfall sequences. The continuous simulation algorithm was tested at five locations where long historical daily rainfall records are available for comparison, and found to perform well in representing the distributional and dependence attributes of the observed daily record. These daily sequences were then used to disaggregate to a subdaily time step using the rainfall state-based disaggregation approach described in the first paper, and found to provide a good representation of the continuous rainfall sequences at the location of interest.

Citation: Mehrotra, R., S. Westra, A. Sharma, and R. Srikanthan (2012), Continuous rainfall simulation: 2. A regionalized daily rainfall generation approach, *Water Resour. Res.*, *48*, W01536, doi:10.1029/2011WR010490.

1. Introduction

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

[3] Generation of daily rainfall is usually undertaken in two distinct stages: first, the generation of rainfall occurrence, and second, the generation of rainfall amounts on the "wet" days. One of the earliest, and still most widely used, rainfall occurrence models is the first-order Markov model developed by *Gabriel and Neumann* [1962], in which the probability of a wet or dry day is defined conditional only on the previous day's rainfall state. Deficiencies of such "short memory" process models (in which precipitation is only dependent on the past through the most recent day's rainfall occurrence) include undersimulation of both long dry spells as well as variability at the interannual timescale, with these issues being addressed in more recent work using higher-order Markov models and Markov models that consider exogenous climate variables as additional predictors [*Wilks and Wilby*, 1999]. To generate precipitation amounts, *Todorovic and Woolhiser* [1975] used an exponential model to simulate the rainfall amount for each wet day, with two-parameter gamma distributions and mixed exponential distributions also commonly used. An

[1]School of Civil and Environmental Engineering, University of New South Wales, Sydney, New South Wales, Australia.
[2]School of Civil, Environmental, and Mining Engineering, University of Adelaide, South Australia, Australia.
[3]Water Division, Australian Bureau of Meteorology, Melbourne, Victoria, Australia.

0043-1397/12/2011WR010490

alternative that does not need to assume the probability distribution associated with the rainfall is presented in the nonparametric weather generation literature [*Brandsma and Buishand*, 1998; *Buishand and Brandsma*, 2001; *Harrold et al.*, 2003a, 2003b; *Lall et al.*, 1996; *Mehrotra and Sharma*, 2007a, 2007b; *Rajagopalan and Lall*, 1999; *Rajagopalan et al.*, 1996; *Sharma and O'Neill*, 2002; *Sharma et al.*, 1997]. In addition to the abovementioned papers, a detailed review of stochastic generation of rainfall for current as well as climate change conditions is presented in *Sharma and Mehrotra* [2010].

[4] The aim of this paper is to present a methodology that allows the generation of daily rainfall at locations where historical daily rainfall records are not available. Traditionally, such regionalized extensions have been achieved via the use of spatial interpolation or extrapolation of model parameters [*Guenni and Hutchinson*, 1998; *Johnson et al.*, 2000; *Kyriakidis et al.*, 2004; *Wilks*, 2008]. This paper describes an alternative approach in which sequences are developed using daily rainfall records at locations which are "nearby" the location of interest (henceforth referred to as the "target" location), for which the rainfall data is presumed to be statistically consistent with the target rainfall.

[5] The regionalized procedure presented here uses the modified Markov model (MMM)—kernel density estimate (KDE) modeling framework for stochastic generation of daily rainfall as presented by *Mehrotra and Sharma* [2007a], in which the occurrence model comprises a Markov chain conditional on the previous day's rainfall occurrence as well as aggregate rainfall occurrences over a number of prior days (e.g., aggregate number of wet days over the previous 365 days) to account for low-frequency persistence. The amounts model uses a kernel-density estimation procedure with conditional dependence on the previous day's rainfall. Finally, to convert the daily rainfall into continuous (subdaily) rainfall data, the daily sequences generated using the regionalized daily model are disaggregated based on the approach presented in the first part of this series [*Westra et al.*, 2011].

[6] The regionalized procedure presented here was developed using 2708 daily rain gage locations across Australia as discussed in section 2. In section 3 we summarize the proposed algorithm, and describe the basis for identifying stations which are statistically "similar" to the rainfall at the location of interest. The model was evaluated by comparing the simulated results with observed daily and subdaily rainfall, with results presented in section 4. Finally, conclusions are provided in section 5.

2. Data

[7] Daily rainfall data were obtained from the Australian Bureau of Meteorology at 17,451 gaging stations, with a maximum of about 8000 daily rain gages recording rainfall in any given year. The distribution of the daily rainfall network is illustrated in Figure 1, in which the number of recording rain gages is plotted as a time series from 1850 to 2007, with low numbers of stations recording in the mid 1800s, and a build-up of rainfall gages in the decades surrounding 1900 to the approximately present levels. This can be contrasted with the series of subdaily rainfall presented as Figure 2 in the work of *Westra et al.* [2011], in which there are a maximum of only around 600–700 subdaily rainfall stations recording at any time, and with very few recording prior to the 1960s.

[8] Of these daily gaging stations, we selected a subset of 2708 locations (Figure 2) that have more than 25 yr of

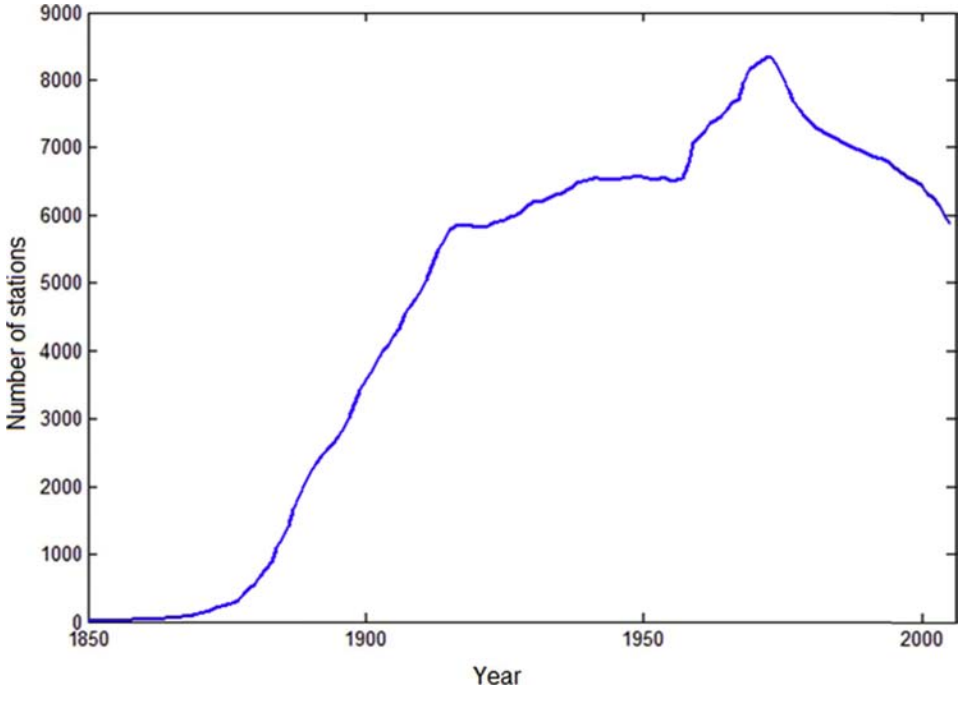

Figure 1. Number of Australia-wide daily rainfall records against year of record, plotted from 1850, considering only stations with <1% data missing.

continuous record and less than 1% of the record classified as "missing" for developing the similarity metric discussed in the section 3. Of these stations, 940 have less than 40 yr of record, 1437 have between 40 and 100 yr, and a further 331 stations have records of more than 100 yr. In spite of large network of rain gages, the spatial distribution of the gaging stations is not homogeneous, with a higher density of gages in the populated regions particularly along the eastern part of Australia. For the remaining analysis we only focus on this set of 2708 stations and fill in the small percentage of missing data using the records from nearby stations.

3. Methodology

[9] In this paper we present a regionalized approach for generating daily rainfall data at any location of interest, irrespective of whether gaged data at the location are available or not. This is achieved by sampling the daily rainfall from a number of nearby rain gages in which the rainfall sequences are expected to be statistically "similar" to the rainfall at the target location. The methodology uses a scaling approach which is similar to that used by *Westra et al.* [2011] to identify and define similarity, except that in this case the relationship being investigated concerns the scaling from daily to annual timescales. Prior to describing the regionalization approach, we will briefly summarize the daily rainfall generator which is based on the MMM-KDE modeling framework of *Mehrotra and Sharma* [2007a], and which was developed to preserve variability across multiple timescales.

3.1. MMM-KDE Model for Generation of Daily Rainfall Sequences

[10] As in the work of *Westra et al.* [2011], we denote R_t as the rainfall amount on day t (where $t = 1, \ldots, 365$ represents the calendar day) at a given station, and a rainfall occurrence as $\mathrm{I}(R_t) = 1$ if $R_t \geq 0.3$ mm and $\mathrm{I}(R_t) = 0$ otherwise, with I() representing the indicator function. In a traditional Markov order one model, we can express the transition probabilities via $Pr(\mathrm{I}(R_t) \mid \mathrm{I}(R_{t-1}))$, with transition probabilities for each day t estimated separately over a sliding moving window of 15 d on either side of t.

[11] Such a Markov order one model is limited in that it is only dependent on rainfall occurrence on the previous day, and thus cannot represent low-frequency variability which is known to exist in precipitation data [*Buishand*, 1978]. *Mehrotra and Sharma* [2007a] showed that inclusion of additional predictors as a conditioning variables improves the representation of low-frequency variability. These predictors may include either aggregated rainfall statistics over a defined number of prior time steps, exogenous predictors such as climate indices representing the El Niño Southern Oscillation, or both. For the present study we focus on a single predictor, namely the aggregate number of rainfall occurrences over the previous 365 d, defined as

$$Z_t = \sum_{j=1}^{365} I(R_{t-j}), \tag{1}$$

with the implicit understanding in this notation that the summation continues into the previous year for $t \leq j \leq 365$.

[12] This approach was preferred over the use of climate indices because it only relies on information contained within the rainfall record itself, and hence does not require an additional model to generate the covariates. Furthermore, there remains considerable uncertainty in identifying climate drivers which force interannual and longer-scale variability in rainfall [e.g., *Westra and Sharma*, 2010]. Finally, the relative role of different large-scale climatic drivers would be expected to vary over our study domain (the Australian continent) which would preclude us from using the same algorithm everywhere.

[13] Taking into account the added exogenous predictor, the resulting transition probabilities (Pr^*) can then be written as,

$$Pr^*_{j,i} = Pr(\mathrm{I}[R_t] = j \mid \mathrm{I}[R_{t-1}] = i, \mathrm{Z}_t), \tag{2}$$

where $i, j \in \{0, 1\}$ represent the case where a day is dry or wet. Expansion of this equation based on conditional probabilities and rearrangement of the terms of equation (2) (see *Mehrotra and Sharma* [2007a] for details) leads to the following:

$$Pr^*_{j,i} = \frac{P(I[R_t] = j, I[R_{t-1}] = i)}{P(I[R_{t-1}]i)} \times \frac{f(Z_t|I[R_t] = j, I[R_{t-1}] = i)}{f(Z_t|I[R_{t-1}] = i)}, \tag{3}$$

where $f(Z_t|I[R_t] = j, I[R_{t-1}] = i)$ and $f(Z_t|I[R_{t-1}] = i)$, respectively, are conditional probability densities of Z_t given the current and previous days rainfall state, and given the previous day's rainfall state alone. Following *Mehrotra and Sharma* [2007a], we draw on the Central Limit Theorem for Z_t as a summation of random numbers and calculate the conditional probabilities using a parametric multivariate Gaussian model. This leads to the following expression for the modified Markovian transition probability, $Pr^*_{1,i}$ for a wet day:

$$Pr^*_{1,i} = Pr_{1,i} \frac{\frac{1}{(V_{1,i})^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(Z_t - \mu_{1,i})\, V_{1,i}{}^{-1} (Z_t - \mu_{1,i})'\right\}}{\left[\frac{1}{(V_{1,i})^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(Z_t - \mu_{1,i})\, V_{1,i}{}^{-1} (Z_t - \mu_{1,i})'\right\} Pr_{1,i}\right] + \left[\frac{1}{(V_{0,i})^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(Z_t - \mu_{0,i})\, V_{0,i}{}^{-1} (Z_t - \mu_{0,i})'\right\} (1 - Pr_{1,i})\right]}, \tag{4a}$$

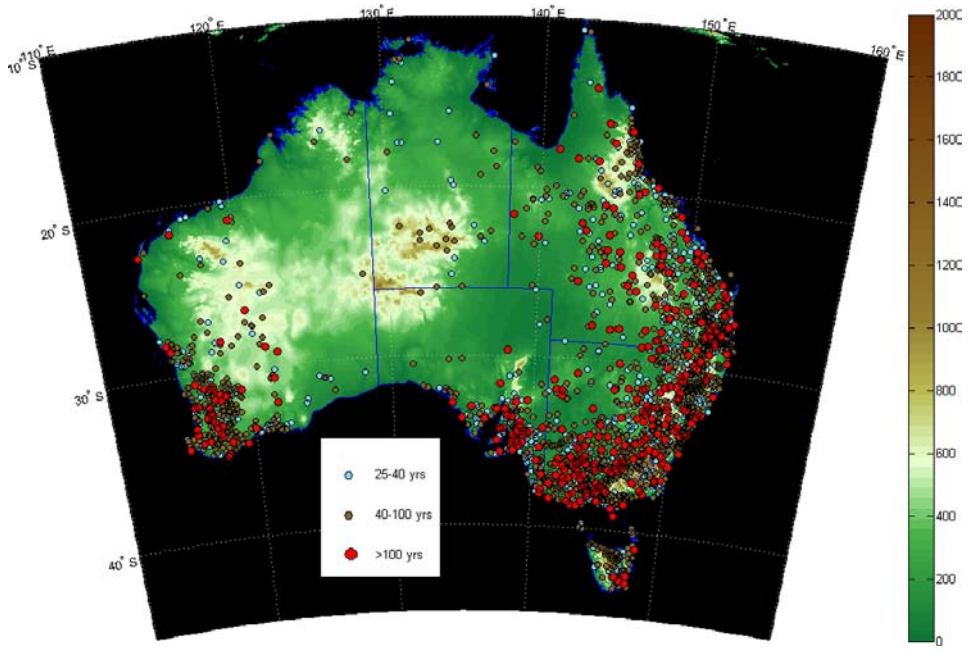

Figure 2. Spatial coverage and record length of the Australian daily rainfall record. Only locations with <1% data missing and length >25 yr are presented, totaling 2708 stations.

where the $\mu_{1,i}$ is conditional mean and $V_{1,i}$ is the corresponding conditional variance of Z when $(I[R_t] = 1)$ and $(I[R_{t-1}] = i)$. Similarly, $\mu_{0,i}$ and $V_{0,i}$ represent, respectively, the mean and the variance of Z when $(I[R_t] = 0)$ and $(I[R_{t-1}] = i)$, and $Pr_{1,i}$ represents the baseline transition probability of the first order Markov model $Pr(I[R_t] = 1|I[R_{t-1}] = i)$. The conditional means and variances are estimated as,

$$\mu_{j,i} = \frac{1}{N_{j,i}} \sum_{k=1}^{N_{j,i}} Z_{(t)k} | [I(R_t) = j, I(R_{t-1}) = i] \tag{4b}$$

and

$$V_{j,i} = \frac{1}{N_{j,i}} \sum_{k=1}^{N_{j,i}} \{Z_{(t)k} | [I(R_t) = j, I(R_{t-1}) = i] - \mu_{j,i}\}^2, \tag{4c}$$

where $N_{j,i}$ represents the number of observations in a moving window centered at day (t), and $Z_{(t)k}$ is the k-th observation in the moving window, ascertained conditional to $[I(R_t) = j|I(R_{t-1}) = i]$.

[14] The model requires estimation of the empirical wet day transition probabilities $Pr(I(R_t) = 1|I(R_{t-1}) = i)$, the sample means $\mu_{j,i}$, and the sample variances $V_{j,i}$ of Z for the four combinations of the current and previous day being wet or dry. To preserve seasonality, separate values of each of these parameters are calculated for each calendar day, using observed data within a moving window centered at that day so as to maintain a sufficient sample size.

[15] Having developed the methodology for the binary sequence of wet and dry days, it is now necessary to generate rainfall amounts R_t for each wet day. Following *Mehrotra and Sharma* [2007a, 2010] and using only the previous day's rainfall depth, R_{t-1}, as the predictor of current-day rainfall depth, the rainfall amounts are generated by sampling from a kernel density estimate of the conditional probability density function $f(R_t|R_{t-1})$. This simplification makes the implicit assumption that low-frequency variability in rainfall can be fully accounted for by simulating low-frequency variability in rainfall occurrences. Evidence to support this assumption comes from a related study which finds that low frequency climate modes such as the El Niño Southern Oscillation tend to influence rainfall occurrences [*Harrold et al.*, 2003a] much more strongly than rainfall amounts on wet days [*Pui et al.*, 2011].

[16] A Gaussian kernel density estimate (KDE) [*Sharma and O'Neill*, 2002; *Sharma et al.*, 1997] is used to define $f(R_t|R_{t-1})$ as follows:

$$f(R_t|R_{t-1}) = \sum_{k=1}^{N} \frac{1}{(2\pi\Sigma')^{0.5}\lambda} \Lambda_k \exp\left(-\frac{(R_t - b_k)^2}{2\lambda^2\Sigma'}\right). \tag{5}$$

[17] Here $f(R_t|R_{t-1})$ is the probability density estimate of rainfall amount conditional on the previous day's rainfall amount, λ is the bandwidth, and N is total number of data points falling within the sliding window and satisfying the condition $(I[R_k] = 1,\ k = 1,\ N)$ and, Σ' is a measure of spread of the conditional density, estimated as

$$\Sigma' = \Sigma_{R_t R_t} - \Sigma^T_{R_t R_{t-1}} \Sigma^{-1}_{R_{t-1} R_{t-1}} \Sigma_{R_{t-1} R_t}, \tag{6}$$

where Σ represents the variance-covariance matrix of R_t and R_{t-1} and superscripts T and -1 denote the transpose and inverse of a matrix. The contribution of each kernel k, in forming the conditional probability density is expressed as Λ_k and represents the weight associated with each kernel. This weight is estimated as

$$\Lambda_k = \frac{\exp\left(-\frac{[R_{t-1} - R_{k-1}]^T \Sigma^{-1}_{R_{t-1}R_{t-1}} [R_{t-1} - R_{k-1}]}{2\lambda^2}\right)}{\sum_{k=1}^{N} \exp\left(-\frac{[R_{t-1} - R_{k-1}]^T \Sigma^{-1}_{R_{t-1}R_{t-1}} [R_{t-1} - R_{k-1}]}{2\lambda^2}\right)}. \tag{7}$$

[18] The conditional mean associated with each kernel slice, b_k is expressed as

$$b_k = R_k + (R_{t-1} - R_{k-1})\Sigma^T_{R_{t-1}R_t}\Sigma^{-1}_{R_{t-1}R_{t-1}}, \quad (8)$$

where all notations are as described before.

[19] The bandwidth λ is adopted here is the Gaussian reference bandwidth λ_{ref} following [*Scott*, 1992] and is expressed as $\lambda_{\text{ref}} = (4/3)^{0.2}N^{-0.2}$, this bandwidth being optimal for the estimation of the probability density for a univariate response.

3.2. Identifying "Nearby" Daily Rainfall Stations

[20] The regionalized approach relies on using data from nearby rainfall stations (in this case daily-read stations) as a substitute for at-site data for cases where at-site data is either unavailable or too short. As such, it is necessary to identify metrics to determine whether two stations are "similar," and to predict the probability that stations within a "neighborhood" of the target site are similar by regressing against physiographic indicators such as the difference in latitude, longitude, elevation, and relative distance to coast between station pairs. Similar to the first paper of this two paper series, the relative distance to the coast is obtained by dividing the difference in the distance to the coast between two stations by the distance to the coast of the target site. This is done to account for the fact that the relative influence of the distance to the coast is likely to be greater for two stations having greater proximity to the coastline.

3.2.1. Annual and Within-Year Daily Rainfall Characteristics

[21] To enable sampling of the daily rainfall series from stations within a neighborhood of the target location, one needs to consider the equivalence not only of the marginal distributions of annual and within-year rainfall but also the joint relationship between the annual and within-year rainfall. Let the various rainfall attributes at the target site be indicated by superscript "o," and at nearby station by the superscript "*s*." This equivalence can be expressed as

$$f(R^s_{y,t}, A^s_y) = f(R^o_{y,t}, A^o_y), \quad (9)$$

with $R_{y,t}$ representing daily rainfall amount for year y, A_y representing the total annual rainfall for that same year, and $f()$ representing the joint probability density function relating the two variables. (For convenience, the subscript y will be omitted from subsequent notation; however, when referring to conditional or joint probabilities between annual and daily rainfall, it is implicit that the daily rainfall is sampled from the same year as the aggregate annual rainfall.)

[22] A difficulty with this formulation is that R^s_t and R^o_t represent a daily time series for each year of record ($t = 1, \ldots, 365/6$) whereas A^s and A^o represents the total rainfall amount for that year and is therefore a scalar. We therefore modify equation (9) to give:

$$f(Y^s, A^s) = f(Y^o, A^o), \quad (10)$$

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

[23] Maximum daily intensity attributes: for each year, the maximum daily rainfall in each season,

[24] Maximum wet spells: for each year, the maximum length of sequence of wet days in each season,

[25] Maximum dry spells: for each year, the maximum length of sequence of dry days in each season,

[26] Rainfall in maximum wet spells: for each year, the total rainfall in the maximum length of sequence of wet days in each season,

[27] Amount per wet day: for each year, the average rainfall amount per wet day for each season,

[28] Seven days rainfall totals: for each year, the maximum 7 d rainfall amount for each season,

[29] Seasonal rainfall: for each year, the total rainfall amount for each season,

[30] Seasonal wet days: for each year, the total number of wet days for each season, and

[31] Annual wet days: for each year, the total number of wet days.

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

[33] To illustrate these concepts, we present in Figure 3 a bivariate scatterplot of annual and summer rainfall at five locations in Australia: Sydney, Perth, Alice Springs, Cairns, and Hobart. These locations are selected as they have a distinctly different climatology, with Hobart located in the south of Tasmania representing one of the southernmost records with temperate climate and Cairns in the north of Queensland representing a location having a moist tropical climate. Similarly, Alice Springs is located in the center of Australia with a semiarid climate, Perth in western Australia representing one of the westernmost records with a mixture of Californian and Mediterranean climates, and Sydney in eastern Australia representing intermediate latitudes. These locations are also the ones used for evaluating the disaggregation model by *Westra et al.* [2011], and thus have been maintained for consistency.

[34] As can be seen from Figure 3, the relationships between seasonal and annual rainfall at each station are distinctly different. Cairns has high annual and summer rainfall amounts whereas Hobart and Alice Springs have relatively little annual and summer rainfall, with summer rainfall being ~25% of annual for Hobart and ~40% of annual for Alice Springs. Sydney and Perth have intermediate values of annual rainfall, although a much lower fraction of annual rainfall occurs in summer in Perth compared to Sydney. It is this relationship between annual average rainfall and various subannual attributes which is of interest for this study, as it enables a range of climate regimes to be clearly distinguished. Although figures are not provided here, similar conclusions can be drawn from consideration of other within-year rainfall attributes.

[35] Another important consideration when dealing with rainfall regionalization relates to the high-spatial variability in rainfall. To highlight this aspect, consider rainfall observations at Sydney Observatory Hill. The observed average annual rainfall at the station, on the basis of a 150-yr long record, is 1216 mm, while the observed average annual rainfall at locations within a 20 km radius of Sydney Observatory Hill varies significantly (e.g., Sydney airport, 1087 mm [79 yr], Concord golf club, 1135 mm [69 yr], and Potts Hill reservoir 917 mm [113 yr]). The best estimate of average annual rainfall from nine nearby stations is 1096 mm, which is 10% below the estimate of the Sydney Observatory Hill annual average rainfall. It

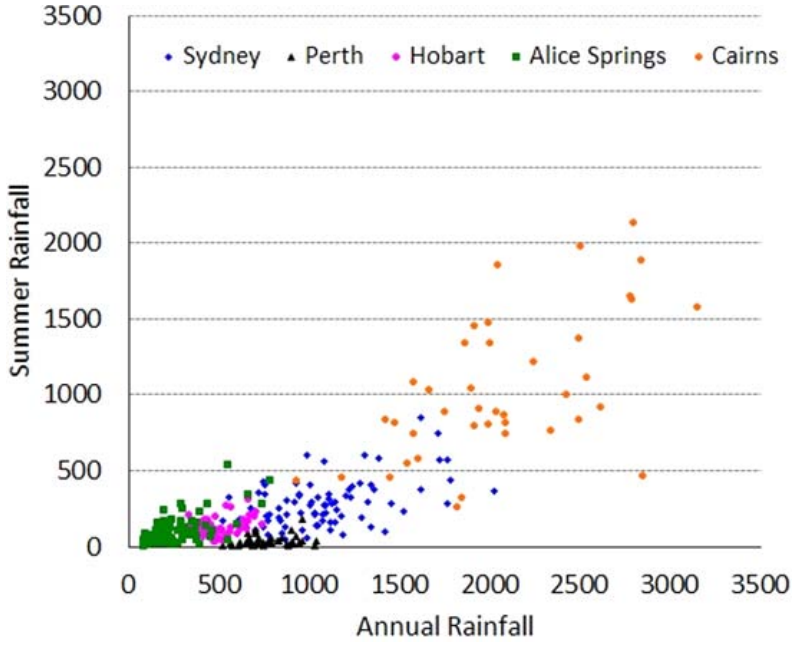

Figure 3. Plot of annual rainfall amount and an attribute of within-year rainfall (the summer rainfall amount) at five locations in Australia.

is therefore quite likely that identified nearby stations, in spite of having similarity in other rainfall attributes, such as seasonality and wet spell characteristics, might contain a bias in annual rainfall at the target site. In the following discussions we assume that a good estimate of long-term average annual rainfall at the target site is known from some other reliable sources, for example, from the long-term relationships that have been developed by the Australian Bureau of Meteorology for annual rainfall across Australia (available at http://www.bom.gov.au/jsp/ncc/climate_averages/rainfall/index.jsp). Although errors remain in annual rainfall estimates, particularly for regions which are sparsely gaged, such products are likely to be the best source of information on annual average rainfall, with a discussion of two Australia-wide rainfall products and their associated errors given by *Beesley et al.*, [2009]. These estimates then can be used to scale the generated daily rainfall following a scaling procedure as described in section 3.3.

3.2.2. Defining the Neighborhood

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

[37] Consistent with *Westra et al.* [2011], for all pairs of daily rainfall stations across Australia, we first examine the bivariate distributions $f(Y^s, A^s) = f(Y^o, A^o)$ for annual rainfall and each of the subannual rainfall attributes described in section 3.2.1, and test whether they are statistically similar using the two-dimensional, two-sample Kolmogorov-Smirnov (K-S) test described in [*Westra et al.*, 2011]. In total, 2708 separate rain gage stations with at least 25 yr of data were used to formulate this relationship, totaling 3,665,278 station pairs. We classify a station pair to be statistically similar based on the K-S test using a 5% significance level, and thus have a vector of length of 3,665,278 with all of the classifications whether or not the stations are statistically similar or different.

[38] Presented in Figure 4 are changes in the percentage of station pairs which are statistically "similar," with increases in absolute difference in latitude and longitude between station pairs based on a frequency binning approach. The percentage of significant stations is calculated by counting the number of statistically similar station pairs in each bin (using a total of 50 bins), and are presented for seven attributes of within-year rainfall: maximum summer wet spell, maximum winter dry spell, daily maximum summer rainfall, 7 d cumulative summer rainfall, rainfall in maximum summer wet spell, summer total rainfall, and number of winter wet days.

[39] Some interesting conclusions can be derived from Figure 4. First, with the exception of the number of wet days in winter, there is between a 35% and 65% chance that the joint distribution of annual rainfall and each of the within-year rainfall attributes are statistically similar, provided the difference in latitude or longitude is small, with the probability decreasing rapidly as the difference in latitude or longitude increases. This is interesting, as in Figure 4a, no account is made of any other physiographic information, such as longitude, elevation, and distance to coast, such that stations may be located in opposite sides of the continent, or at very different elevations, and yet still have close to a 50% chance of having the same scaling between annual and within-year rainfall provided they are at similar latitudes.

[40] Second, while the probability that two stations are similar decreases with increasing difference in longitude

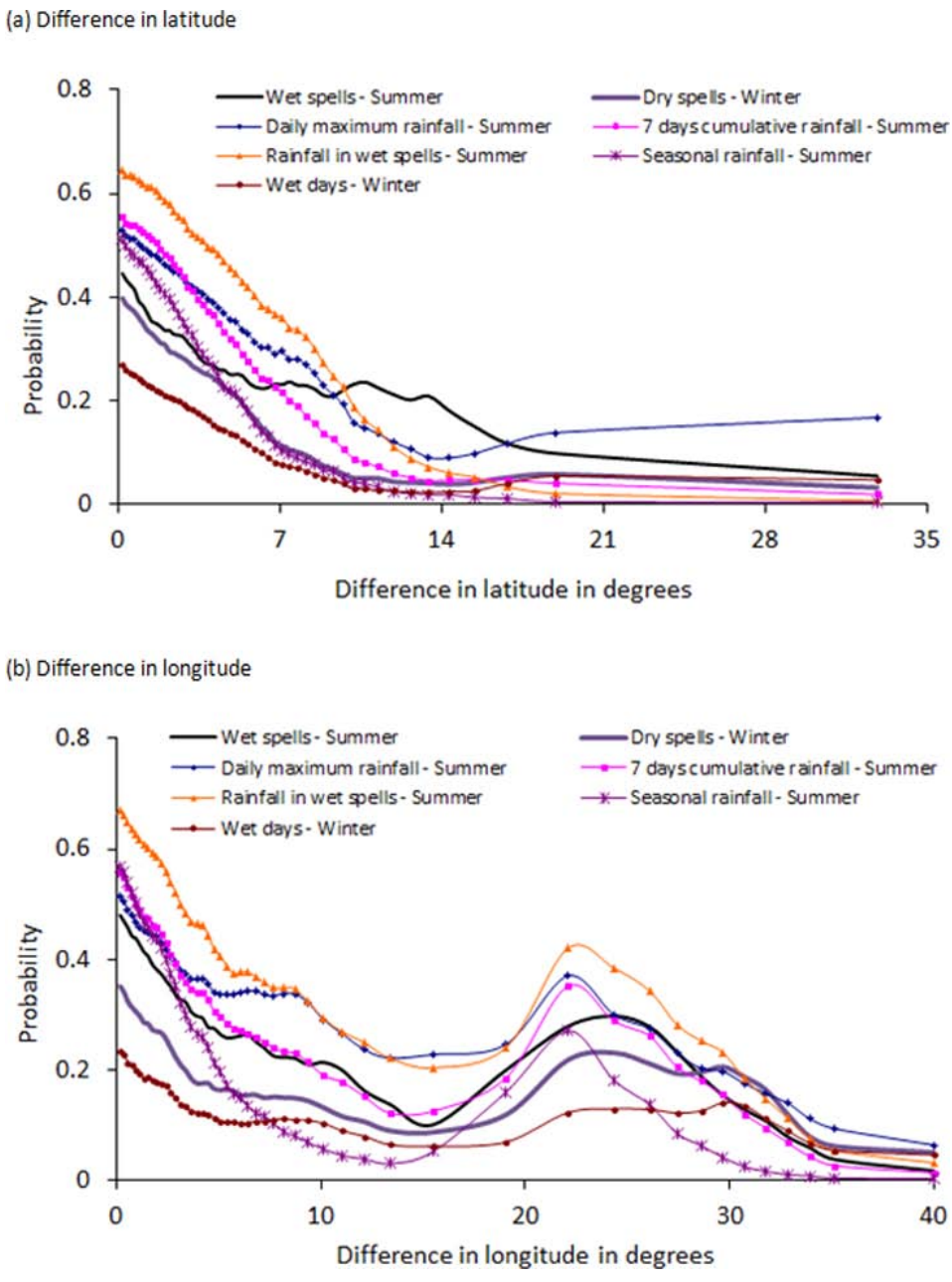

Figure 4. Probability that the annual and within-year rainfall attributes at two stations are statistically similar (using Kolmogorov-Smirnov test) is plotted against a single predictor. (a) The difference in latitude and (b) longitude between station pairs, and seven responses representing different within-year rainfall attributes.

for small differences, the probability increases once the difference in longitude reaches ~20°–25°. This result is because of the clustering of stations as shown in Figure 2, with groups of stations in the southwest and southern parts of the continent showing similar climatology. For subsequent analyses we only consider predictors for station pairs with a difference in latitude <15°, a difference in longitude <10°, and a difference in elevation <350 m, with a total of 1,646,664 station pairs meeting these criteria. Although somewhat arbitrary, these thresholds of differences in latitude, longitude, and elevation were selected to ensure that the probability that two stations are similar decreases smoothly as the magnitude of each of the predictors increases, while still ensuring that all nearby station pairs were included.

[41] We now use a logistic regression model to find the probability that any two stations are similar conditional on a range of physiographic factors, such as the difference in latitude, longitude, elevation, and the distance to the coast between each station pair. This formulation is equivalent to the formulation specified in equation (9) of *Westra et al.* [2011] and the reader is referred to that paper for a detailed mathematical description. The logistic regression model formulation was selected as it enables simulation of a binomial response (i.e., if any station pair is classified as "similar" it is represented by a 1′, and if they are not similar it is represented by a 0′), as a function of a range of predictors, namely, all of the above physiographic metrics.

[42] The results of this multivariate regression for all key rainfall attributes are presented in Table 1, and plotted for the selected rainfall attributes in Figure 5, against an amalgamated variable comprising the mean of all of the predictors when expressed as a percentage of their range. As can be seen, the results show notable improvements in the probability that two stations are similar compared to Figure 4, since we are now considering the influence of all predictors simultaneously. In fact, with the exception of the number of wet days and maximum dry spells in the winter season, the results show that for small values of each of the predictors there is more than 80% probability that the annual-to-within-year joint probability distributions are statistically similar. This forms the basis for our assertion that, provided an adequate estimate of annual rainfall is available at the location of interest, it is possible to draw data from daily-read gages within a neighborhood of that location.

3.2.3. Implementation

[43] On the basis of the methodology described in section 3.2.2, multivariate logistic relations are developed for all key rainfall attributes, with regression coefficients shown in Table 1. Owing to a large pool of rainfall attributes, the developed relationships are examined closely and a few important rainfall attributes are selected encompassing the full distribution of relationships as well as capturing the overall seasonal variations. The finally selected rainfall attributes include: rainfall in maximum wet spells: winter, rainfall in maximum wet spells: summer, number of wet days: winter, number of wet days: summer, total rainfall amount: winter, total rainfall amount: summer, and maximum wet spells: summer, totaling seven rainfall attributes.

[44] The approach to identifying "nearby" stations is as follows:

[45] 1. For any location of interest (the "target" location), identify the probability (u) that each of the 2708 daily rain gage stations in Australia is statistically similar using the logistic regression coefficients provided in Table 1.

[46] 2. For each attribute, rank each of the 2708 stations from highest to lowest in terms of the probability that the rainfall at both stations are statistically similar, and calculate the average rank, $r_{s,}$ for each station across all rainfall attributes. Low values of the rank r_s therefore represent stations with a high probability of having similar rainfall patterns to the target site.

[47] 3. Select the S lowest-ranked stations to represent the set of "statistically similar nearby stations" for inclusion in the daily rainfall generation model.

[48] 4. Calculate a weight associated with each nearby station using the following:

$$w_s = \frac{1/r_s}{\sum_{i=1}^{S} 1/r_i}, \tag{11}$$

Table 1. Logistic Regression Coefficients[a]

		Logistic Regression Coefficients					
Season	Within-Year Rainfall Attribute	Intercept	Latitude (Degrees)	Longitude (Degrees)	Elevation (m)	Distance_Coast (Dimensionless)	Latitude × Longitude
DJF	Maximum daily rainfall	1.94017	−0.31149	−0.21721	−0.00006	−0.99617	0.02921
DJF	Maximum wet spells	1.57173	−0.12363	−0.23577	−0.00097	−1.84297	0.01944
DJF	Maximum dry spells	1.27269	−0.08153	−0.29881	−0.00067	−1.98523	0.02596
DJF	Maximum 7 d cumulative rainfall	2.10002	−0.3594	−0.23293	0.00022	−0.52836	0.01663
DJF	Rainfall in maximum wet spell	2.69367	−0.33091	−0.22776	0.00061	−0.66419	0.01488
DJF	Amount per wet day	0.71517	−0.15911	−0.14189	−0.00092	−2.02995	0.02403
DJF	Total rainfall	2.25955	−0.42135	−0.3508	0.00035	−0.27985	0.01459
DJF	Wet days	0.68726	−0.10239	−0.28291	−0.00129	−1.77739	0.02196
MAM	Maximum daily rainfall	1.80769	−0.1338	−0.17474	0.00103	−1.21653	0.01686
MAM	Maximum wet spells	1.41418	−0.09489	−0.0953	−0.00044	−2.88061	0.01235
MAM	Maximum dry spells	1.5952	−0.09892	−0.13585	−0.00087	−3.17255	0.01828
MAM	Maximum 7 d cumulative rainfall	2.44975	−0.17271	−0.1589	0.0006	−1.29365	0.01679
MAM	Rainfall in maximum wet spell	3.1261	−0.21213	−0.19391	0.00052	−1.29154	0.02174
MAM	Amount per wet day	0.74816	−0.14191	−0.16185	−0.0007	−1.92763	0.02497
MAM	Total rainfall	3.42643	−0.1453	−0.10949	−0.00028	−2.57844	0.00766
MAM	Wet days	0.70431	−0.16809	−0.12329	−0.00067	−2.4662	0.02275
JJA	Maximum daily rainfall	1.82321	−0.13547	−0.22674	0.00107	−1.71183	0.01781
JJA	Maximum wet spells	0.65472	−0.26519	−0.11999	−0.0001	−0.81665	0.01826
JJA	Maximum dry spells	0.74039	−0.32551	−0.16744	0.00036	−0.52668	0.01107
JJA	Maximum 7 days cumulative rainfall	1.93608	−0.2260	−0.22303	0.00047	−0.89708	0.01153
JJA	Rainfall in maximum wet spell	2.14629	−0.18896	−0.18134	0.00044	−0.99356	0.01146
JJA	Amount per wet day	0.47694	−0.1646	−0.18979	−0.00026	−1.43559	0.02205
JJA	Total rainfall	1.53271	−0.33745	−0.27651	0.00021	−0.23115	0.0059
JJA	Wet days	0.03531	−0.31827	−0.13876	0.00015	−0.50433	0.0136
SON	Maximum daily rainfall	2.14753	−0.1352	−0.1751	−0.00038	−2.22906	0.01969
SON	Maximum wet spells	1.13659	−0.1740	−0.15363	0.00015	−1.7462	0.02285
SON	Maximum dry spells	1.1662	−0.38858	−0.1697	0.00058	−1.00345	0.03367
SON	Maximum 7 d cumulative rainfall	2.68549	−0.15744	−0.15262	0.00052	−2.14849	0.01239
SON	Rainfall in maximum wet spell	3.37873	−0.19688	−0.09432	0.00065	−1.63498	0.01413
SON	Amount per wet day	0.60215	−0.14703	−0.1158	−0.00085	−2.14895	0.02161
SON	Total rainfall	2.55865	−0.22178	−0.13762	0.00025	−1.62241	0.01607
SON	Wet days	0.32418	−0.21207	−0.16741	0.00014	−1.31526	0.02662
Annual	Annual wet days	−0.49974	−0.14735	−0.09808	−0.00039	−1.61963	0.01992

[a]All predictors were found to be statistically significant (usually with a p-value < 0.001 level).

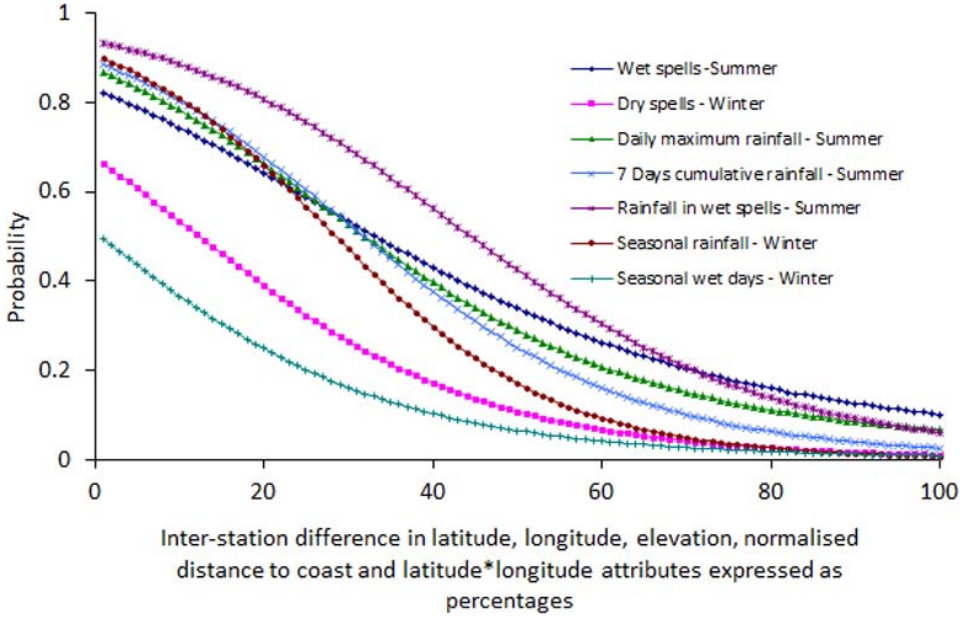

Figure 5. Multivariate logistic regression results for different rainfall attributes. The probability that annual and within-year rainfall attributes are statistically similar is plotted against percent differences in latitude, longitude, elevation, relative distance to coast, and latitude times longitude with 100% representing a 15° difference in latitude, a 10° difference in longitude, 350 m difference in elevation, 1 unit of scaled relative difference in distance to coast, and $75^{\circ 2}$ latitude times longitude.

where the w_s represents the weight associated with the s-th station, and is used as the basis for probabilistically selecting nearby stations in the modified Markov model. Lower-ranked stations, which by definition have rainfall attributes which are most statistically similar to the target site, will have higher weight and therefore have a higher probability of being selected in the rainfall generation algorithm.

[49] The selection of the size of S is somewhat subjective, as larger values of S result in a decrease to the probability of selecting stations which are statistically similar to the target station, whereas smaller values of S will result in small sample sizes. For this study we selected $S = 5$, resulting in an average of ~125–200 yr of data distributed over the five stations.

3.3. Regionalized Extension of the Daily Rainfall Generation Model

[50] The regionalized extension of the daily rainfall generation model is different from the regionalized subdaily rainfall disaggregation model described by *Westra et al.* [2011]. Rather than resampling daily rainfall from neighboring stations, we use the information from nearby daily stations to estimate the parameters for the rainfall occurrence model and form the nonparametric kernel density estimate for the rainfall-amounts model. The algorithm is described below:

[51] 1. Identify the S nearby stations following the procedure outlined in section 3.2.3. Calculate the weight w_s associated with each nearby station s using equation (11). Transform these weights to probabilities (Pr_s) and cumulative probabilities (F_{ws}) using the following:

$$Pr_s = \frac{w_s}{\sum_{i=1}^{S} w_i} \quad \text{and} \quad F_{ws} = F_{ws-1} + Pr_s \quad \text{for} \quad s > 1; F_{w1} = Pr_1. \tag{12}$$

[52] 2. Calculate the average annual rainfall, $\overline{A}^s$, at these stations and, the average annual rainfall at the target station, $\overline{A}^o$, using a spatially interpolated map of total annual rainfall across Australia.

[53] 3. At each identified nearby location s, for all calendar days of the year, calculate the transition probabilities of the standard first-order Markov model and conditional means and variances of the higher timescale predictor variable Z (previous 365 d wetness state) using equations (4b) and (4c). Also, for each calendar day (t), look for wet days ($I[R_k] = 1$) within the same moving window and form a series of current day rainfall amount ($R_k[s]$) and associated previous day's rainfall value ($R_{k-1}[s]$). Let N represent the total number of such observations. Calculate the variances and covariances (Σ) of $R_k(s)$ and $R_{k-1}(s)$ series.

[54] 4. Before the start of the simulation, select at random a nearby station. Pick a short segment (1 yr) of the historical sequence at this station to use for the initial specification of Z_t. The first day in the generated sequence is the day immediately after the end of this start-up sequence.

[55] 5. At a given day t, generate a uniformly distributed random number u and identify the position s^* such that $F_{ws^*-1} < u \leq F_{ws^*}$, thereby selecting a nearby station s. Assign appropriate transition probability to the day t based on previous day's rainfall state of the generated series at the target station. If the previous day is wet, assign probability, Pr as $Pr_{11}(s)$, otherwise assign $Pr_{10}(s)$.

[56] 6. Calculate the value of the previous 365 d wetness state (number of wet days) prior to the day t using equation (1) and the available generated sequence $I(R^*)$, where R^* defines the generated rainfall series at the target station. Estimate the modified transition probability of a wet day, Pr^* using equation (4a). Generate a uniformly distributed random number v and compare it with Pr^*. If v is $\leq Pr^*$, assign

a rainfall occurrence, $I(R_t^*)$ for the day t as wet (1) otherwise dry (0). If the day is simulated as dry, move on to the next day ignoring the rainfall amount generation steps.

[57] 7. Estimate the weights Λ_k for the kernel slices for all N data points that are associated with each data pair (R_k, R_{k-1}) and R_{t-1}^* using equation (7). Note that in the simulation scheme one does not need to estimate explicitly the conditional density in equation (5). Since the conditional density function is the sum of N Gaussian kernel slices that each contribute weight Λ_k (the weights sum to 1, equation (7)), simulation can be achieved by first picking a slice with probability Λ_k and then selecting $R_t(s)$ as a random variate from that kernel slice with mean b_k and variance equal to $(\lambda^2\Sigma')$,

$$R_t^* = b_k + \lambda(\sqrt{\Sigma'})W_t, \tag{13}$$

where W_t is a random variate from a normal distribution with mean of 0 and variance of 1, Σ' is a measure of spread of the conditional density given by equation (6), b_k is the conditional mean associated with the kernel slice k, calculated using equation (8), and R_t^* is the generated rainfall at a day t. Generate another W_t if the generated rainfall is less than the minimum rainfall threshold of 0.3 mm, or else move on to the next step.

[58] 8. Rescale the generated daily rainfall by multiplying it with the ratio $\overline{A}^o/\overline{A}^s$.

[59] 9. Move to the next day in the generated sequence and repeat the above steps, starting from step 5, until the desired length of generated sequence is obtained.

4. Results

[60] We now test the applicability of the logic outlined in section 3. Specifically, we assessed the capability of the regionalized daily simulation model (not using the observed record for the location being modeled) in representing the attributes derived from the observed daily record, followed by an assessment of the continuous rainfall sequences derived through disaggregation from the generated daily sequence. Our assessment is based on daily and subdaily rainfall data at five climatologically different locations in Australia (Sydney, Perth, Alice Springs, Cairns, and Hobart). The assessment results in subsections 4.1 and 4.2 are based on 100 realizations, each equaling the record length of the historical data available at each location.

4.1. Monthly, Seasonal, and Annual Statistics

[61] The seasonal and annual means and standard deviations of wet days and rainfall amounts from the simulated and observed daily rainfall time series are presented in Table 2.

Table 2. Observed and Simulated Rainfall Statistics for Five Selected Locations

	Wet Days				Seasonal/Annual Total Rainfall Amount (mm)			
	Mean		SD		Mean		SD	
Season/ Station	Observed	Simulated (5% and 95%)	Observed	Simulated (5% and 95%)	Observed	Simulated (5% and 95%)	Observed	Simulated (5% and 95%)
				Sydney (066037)				
Autumn	33.1	34.22 (32.8–35.7)	8.4	7.31 (6.3–8.3)	319.9	335.1 (316.6–353.9)	159.0	140.8 (122.3–161.4)
Winter	28.4	27.7 (26.6–29.0)	8.3	6.8 (6.1–7.8)	267.2	247.8 (232.3–268.1)	147.3	120.8 (98.6–141.7)
Spring	30.0	30.1 (28.7–31.6)	7.9	6.5 (5.6–7.4)	213.9	223.3 (207.3–239.0)	109.7	95.6 (82.2–111.4)
Summer	31.9	32.7 (31.3–34.2)	8.4	6.7 (5.9–7.7)	285.1	280.6 (259.5–299.3)	158.1	118.7 (99.5–142.4)
Annual	123.0	124.8 (120.0–129.4)	21.0	17.1 (15.0–19.7)	1085.7	1087 (1087–1087)	317.0	251.4 (222.0–285.9)
				Perth (009021)				
Autumn	23.5	23.2 (21.9–24.6)	5.4	6.0 (5.3–6.9)	160.5	169.1 (157.7–182.3)	62.7	67.4 (56.3–77.0)
Winter	49.5	49.0 (47.5–51.2)	7.4	7.7 (6.7–8.7)	437.8	428.2 (416.0–441.0)	89.1	97.2 (83.0–112.7)
Spring	28.2	29.7 (28.5–31.0)	6.9	6.3 (5.5–7.3)	143.6	152.1 (143.9–161.3)	47.0	45.6 (39.5–52.9)
Summer	8.3	8.5 (7.7–9.2)	4.1	3.5 (2.9–4.1)	34.5	32.2 (27.2–38.0)	34.5	24.8 (19.5–31.9)
Annual	109.9	110.4 (107.5–113.4)	15.8	13.1 (10.9–14.8)	781.2	781.4 (781.4–781.5)	143.0	135.0 (115.5–153.3)
				Alice Springs (015590)				
Autumn	7.8	5.6 (4.8–6.7)	5.1	3.8 (3.1–5.3)	67.2	66.8 (56.5–76.2)	75.5	62.6 (51.4–79.4)
Winter	6.9	3.6 (3.1–4.2)	5.4	2.6 (2.1–3.2)	38.2	30.3 (24.0–38.3)	45.4	30.0 (23.1–39.3)
Spring	11.6	8.0 (6.9–9.6)	5.5	4.4 (3.4–6.3)	57.7	61.2 (54.1–67.2)	41.6	43.9 (34.0–57.6)
Summer	14.2	10.5 (9.1–11.8)	6.0	5.3 (4.2–7.0)	116.7	120.3 (111.5–131.7)	101.8	87 (71.8–107.0)
Annual	40.7	27.7 (24.8–31.9)	12.9	10.6 (8.0–16.1)	279.4	278.9 (278.8–278.9)	151.7	138.9 (113.7–168.5)
				Cairns (031011)				
Autumn	49.2	48.4 (46.5–50.1)	8.2	8.0 (6.9–9.3)	722.1	748.4 (711.7–794.3)	326.6	272.6 (226.1–325.6)
Winter	25.0	26.2 (24.3–27.7)	8.3	7.2 (6.2–8.2)	104.8	148.3 (135.5–163.1)	51.2	69.6 (56.5–85.6)
Spring	24.3	23.8 (22.5–25.3)	8.9	7.1 (6.0–8.4)	165.4	182.0 (163.6–202.2)	109.7	99.7 (75.1–123.9)
Summer	49.4	45.4 (43.3–47.7)	9.0	8.0 (6.8–9.3)	1007.7	912.0 (862.8–955.5)	413.5	334.4 (292.9–384.0)
Annual	147.6	143.8 (139.6–148.5)	17.8	17.3 (15.2–19.3)	1992.4	1991 (1991–1991)	554.7	459.3 (395.3–528.4)
				Hobart (094008)				
Autumn	29.4	26.9 (25.4–28.9)	6.7	5.9 (4.9–7.0)	114.6	111.3 (104.1–121.1)	53.3	44.4 (36.2–57.0)
Winter	34.7	32.4 (30.1–34.6)	8.2	6.6 (5.6–7.9)	119.5	125.6 (118.1–137.6)	42.3	41.9 (34.3–49.9)
Spring	35.9	33.8 (31.5–36.0)	7.2	7.2 (6.2–8.6)	131.2	134.8 (124.3–144.7)	45.6	44.5 (35.5–55.4)
Summer	26.2	24.0 (22.5–26.0)	5.8	5.7 (4.8–6.8)	131.3	124.3 (113.7–133.9)	60.4	52.1 (42.7–63.7)
Annual	126.6	117.1 (110.6–123.5)	19.8	17.6 (14.5–21.1)	493.6	495.5 (495.3–495.6)	110.5	100.9 (83.7–119.9)

The means of both the number of wet days and rainfall amounts are reproduced reasonably well, with the simulated results generally within 10% of the observed data. The primary exception to this is for Alice Springs, in which the simulated mean number of wet days is between 26% and 48% below the observed number of wet days, with the rainfall amount also underestimated by 21% for the winter season. The reason for this discrepancy is likely to be the sparse sampling of rainfall in the vicinity of Alice Springs, leading to the selection of "nearby" gages which are not reflective of at-site daily rainfall. Furthermore, the arid nature of the Alice Springs climate may also contribute to the results, with much of the rainfall being contained in a small number of wet years potentially leading to less consistent results. In all cases, the average annual observed and simulated rainfall amounts correspond almost exactly, as each simulated series is scaled to the observed rainfall amounts. As already discussed, in settings where observed data is not available, such scaling will be achieved using a spatially interpolated total annual rainfall product, therefore introducing an additional source of uncertainty [*Beesley et al.*, 2009]. Unlike the mean rainfall, the annual standard deviations are generally undersimulated, by an average of ~14% for the number of wet days and by an average of 12% for the rainfall amounts.

[62] Box plots of observed and simulated wet days and rainfall totals at the monthly timescale are presented in Figure 6. The simulated statistics generally follow the observed monthly trends at all of the stations except at Alice Springs, where the model undersimulates the means of monthly wet days and rainfall totals. Undersimulation of the standard deviation is also evident for some months at several locations.

[63] Figure 7 presents the year-to-year distribution of the annual rainfall amounts and the annual number of wet days across a range of exceedance probabilities. As can be seen,

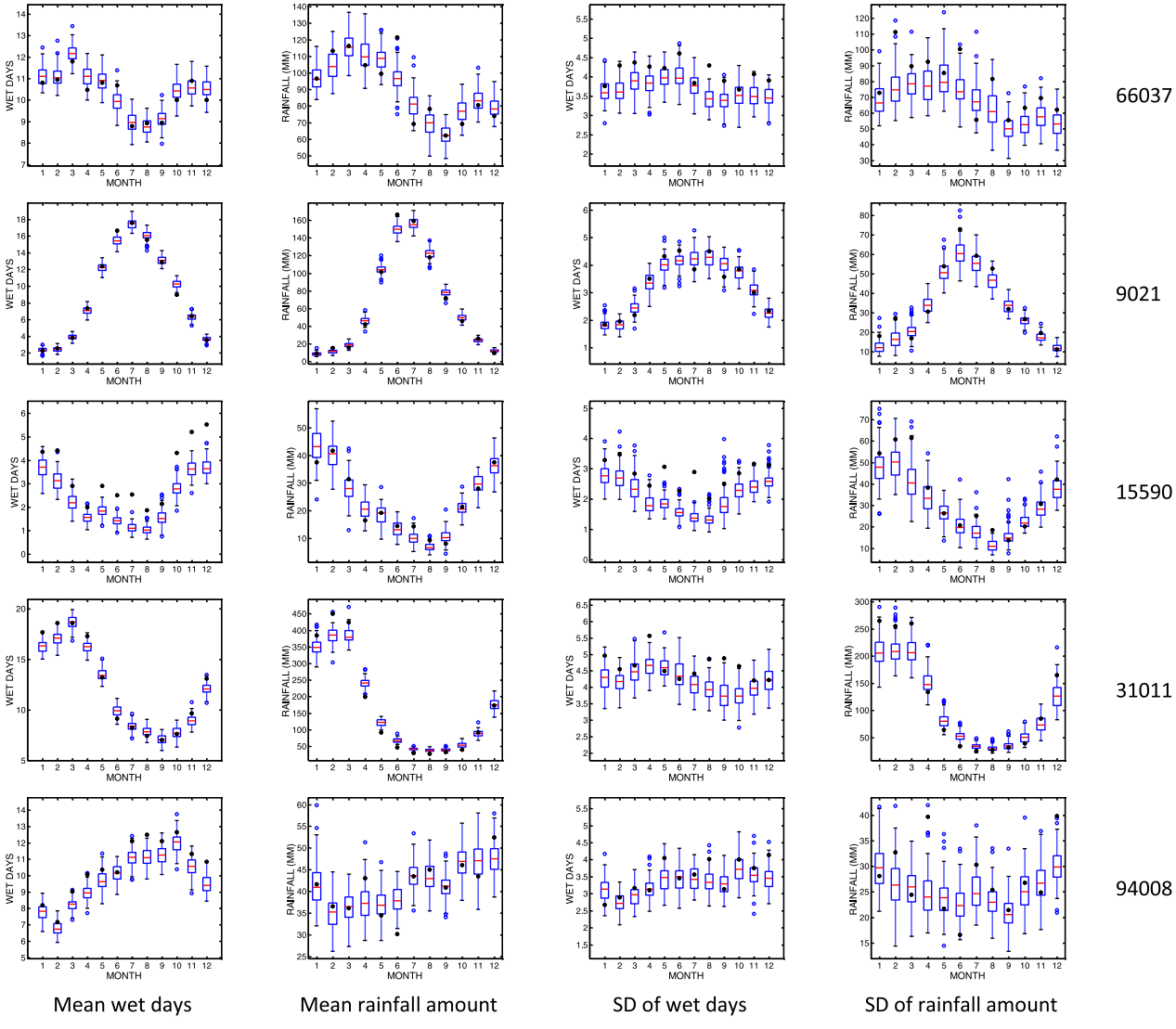

Figure 6. Distribution of means and standard deviation (SD) of observed and model simulated monthly wet days and rainfall amount for five selected test stations. Solid circles represent observed statistic while boxes are for lower quartile, median, and upper quartile values of the simulated statistics drawn from 100 realizations.

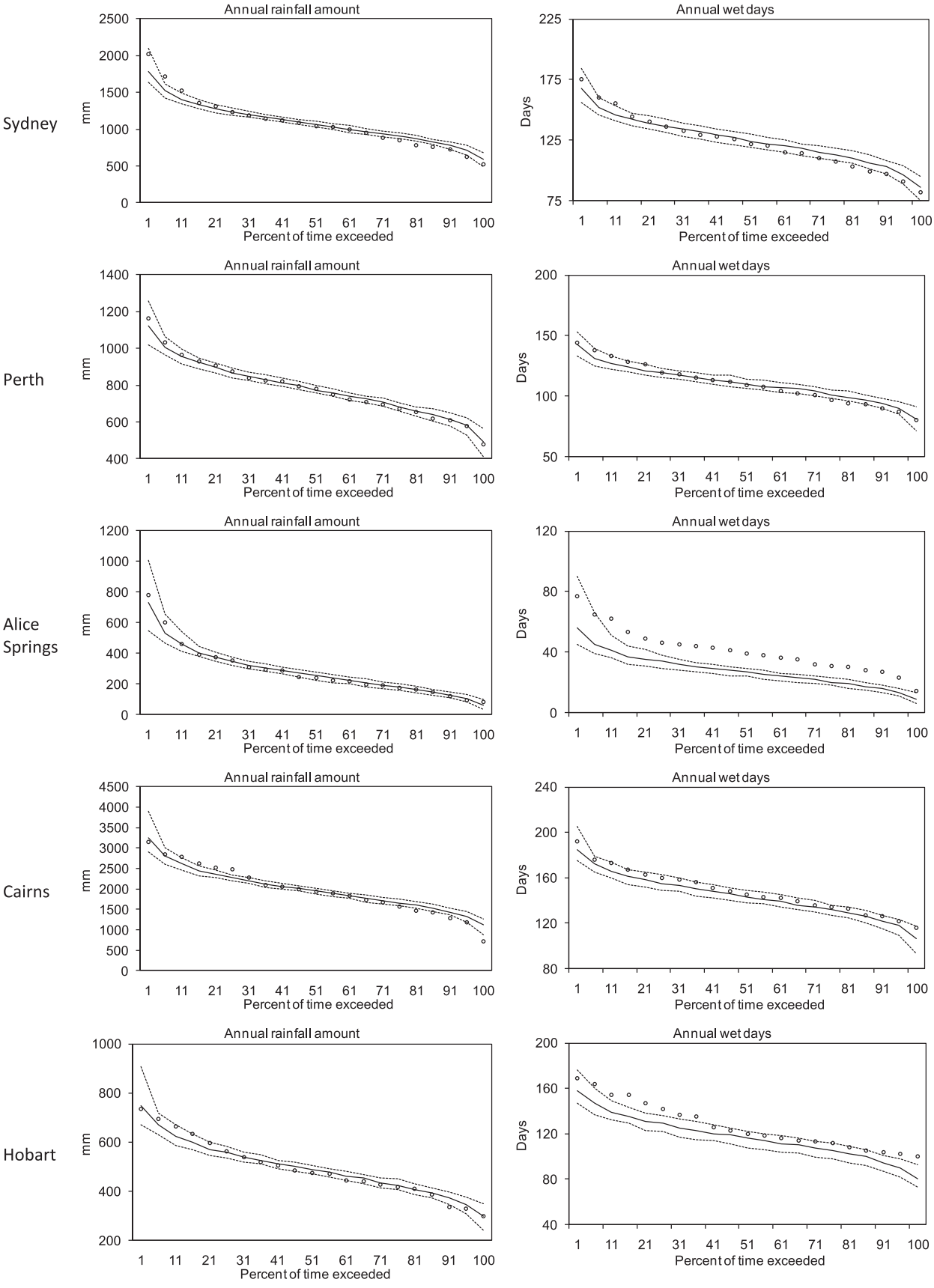

Figure 7. Distribution plots of observed and model-simulated annual number of wet days and rainfall amount for five selected locations.

Table 3. Observed and Simulated Rainfall Transition Probabilities for Five Selected Locations[a]

Station/ Probability	Sydney		Perth		Alice Springs		Cairns		Hobart	
	Observed	Simulated	Observed	Simulated	Observed	Simulated	Observed	Simulated	Observed	Simulated
p_{10}	0.15	0.154 (0.4%)	0.12	0.115 (−1.5%)	0.06	0.045 (−25%)	0.12	0.12 (0.2%)	0.18	0.17 (−4.5%)
p_{11}	0.18	0.188 (2.2%)	0.18	0.187 (1.7%)	0.05	0.03 (−39.9%)	0.28	0.274 (−3.8%)	0.17	0.15 (−10.7%)
p_{111}	0.10	0.104 (1.6%)	0.12	0.121 (4.3%)	0.02	0.013 (−42.4%)	0.21	0.193 (−7%)	0.08	0.071 (−12.9%)
p_{110}	0.08	0.084 (2.9%)	0.07	0.066 (−2.7%)	0.03	0.018 (−37.9%)	0.08	0.081 (4.8%)	0.09	0.079 (−8.6%)
p_{010}	0.07	0.07 (−2.5%)	0.05	0.049 (0.1%)	0.03	0.027 (−13.5%)	0.04	0.039 (−8%)	0.09	0.092 (−0.7%)
P_{011}	0.08	0.084 (2.9%)	0.07	0.066 (−2.7%)	0.03	0.018 (−37.9%)	0.08	0.081 (4.8%)	0.09	0.079 (−8.6%)

[a]Also shown are the percent differences in the brackets.

for total annual rainfall amounts, although the median is well simulated, the variability is low for most locations, with the upper and lower bounds of the extremes being underestimated. In contrast, the number of wet days is generally well reflected. The exception to this is once again Alice Springs, where the distribution of annual rainfall is accurately represented whereas the number of wet days is underestimated. This can be explained by the transition probability parameters provided in Table 3, which are generally within 10% of the at-site parameters for all locations except for Alice Springs.

[64] The results using the regionalized model show overall good agreement between the observed and simulated statistics at all stations. The underestimation of variability at the annual timescale is attributable more to the structure and assumptions of the daily rainfall generation model adopted here than to the regionalization procedure. The simplified structure of the daily rainfall generation model (a single predictor as an aggregate number of rainfall occurrences over the previous 365 d and use of global bandwidth in kernel-density estimation procedure) and the assumption of normal distribution in equation (4a) may result in these discrepancies in the results. To check whether the underestimation of the variability is due to the regionalization procedure adopted here, we used the same model for rainfall generation at these sites using the observed-at-site rainfall record, and obtained the similar results (not shown). Experimenting with a larger number of predictors [*Mehrotra and Sharma*, 2007a], using the local bandwidth in rainfall simulation procedure [*Sharma et al.*, 1997], and using aggregated wet day predictor(s) in the rainfall amount simulation stage [*Harrold et al.*, 2003b] or employing an empirical scaling adjustment procedure to match the target site standard deviation of the annual rainfall [*Boughton*, 1999], might help further improve the representation of observed year-to-year variability in the simulations. To obtain the annual standard deviation value at the target location, Bureau of Meteorology, Australia can be

Table 4. Comparison of Observed and Simulated Results for Median Annual Maxima for Different Storm Burst Durations and Antecedent Precipitation Prior to 1 h Storm Burst[a]

	Sydney		Perth		Alice Springs		Cairns		Hobart	
	Observed	Simulated (5 and 95%)	Observed	Simulated (5 and 95%)	Observed	Simulated (5 and 95%)	Observed	Simulated (5 and 95%)	Observed	Simulated (5 and 95%)
					Annual Maxima					
6 min	8.9	9.5 (8.95–10.14)	6.2	6.5 (5.89–6.86)	5.5	8.0 (7.28–8.78)	11.6	12.8 (12.03–13.87)	4.5	4.0 (3.63–4.54)
30 min	25.7	24.5 (23.05–26.07)	14.7	14.5 (13.36–15.67)	16.7	21.0 (19.21–23.48)	34.9	37.7 (35.97–40.44)	11.3	9.7 (8.74–10.58)
1 h	35.4	32.9 (30.47–34.62)	18.8	18.2 (16.86–19.61)	22.1	26.6 (24.14–29.81)	51.7	54.2 (50.91–58.45)	14.6	12.9 (11.82–14.06)
3 h	55.4	48.7 (45.7–52.52)	29.0	27.0 (25.3–28.86)	32.6	34.9 (30.59–38.54)	83.5	85.5 (80.94–92.67)	22.9	20.5 (18.9–22)
6 h	72.3	62.9 (59.05–67.16)	36.3	34.2 (31.92–36.28)	39.6	40.7 (35.64–44.62)	113.0	113.8 (107.99–120.87)	30.3	26.8 (25.1–28.74)
12 h	91.8	81.1 (76.23–87.81)	45.4	42.1 (39.38–45.08)	48.2	46.8 (41.9–51.83)	147.4	144.6 (137.12–155.86)	39.6	33.2 (30.71–35.76)
					Antecedent Precipitation Prior to 1-h Burst (mm)					
6 h	15.4	11.8 (8.52–15.42)	6.8	5.7 (4.28–7.51)	6.1	3.8 (2.56–5.71)	25.4	21.1 (14.76–26.87)	6.3	5.5 (3.99–7.19)
12 h	22.7	16.3 (11.44–21.75)	9.7	7.4 (5.47–9.84)	8.0	5.2 (3.39–8.03)	32.3	27.4 (19.23–34.48)	9.1	6.8 (4.97–9.25)
24 h	31.4	20.4 (15.51–27.77)	12.8	9.9 (7.53–12.8)	10.7	8.1 (5.62–10.66)	42.0	36.1 (26.55–45.29)	10.2	7.9 (5.91–10.7)
48 h	43.0	24.9 (19.14–32.9)	15.5	13.6 (11.16–16.46)	15.5	11.4 (8.61–14.96)	58.6	49.3 (38.11–59.45)	11.4	9.4 (7.18–12.28)

[a]The simulated median annual maxima represent the median of all 100 simulations.

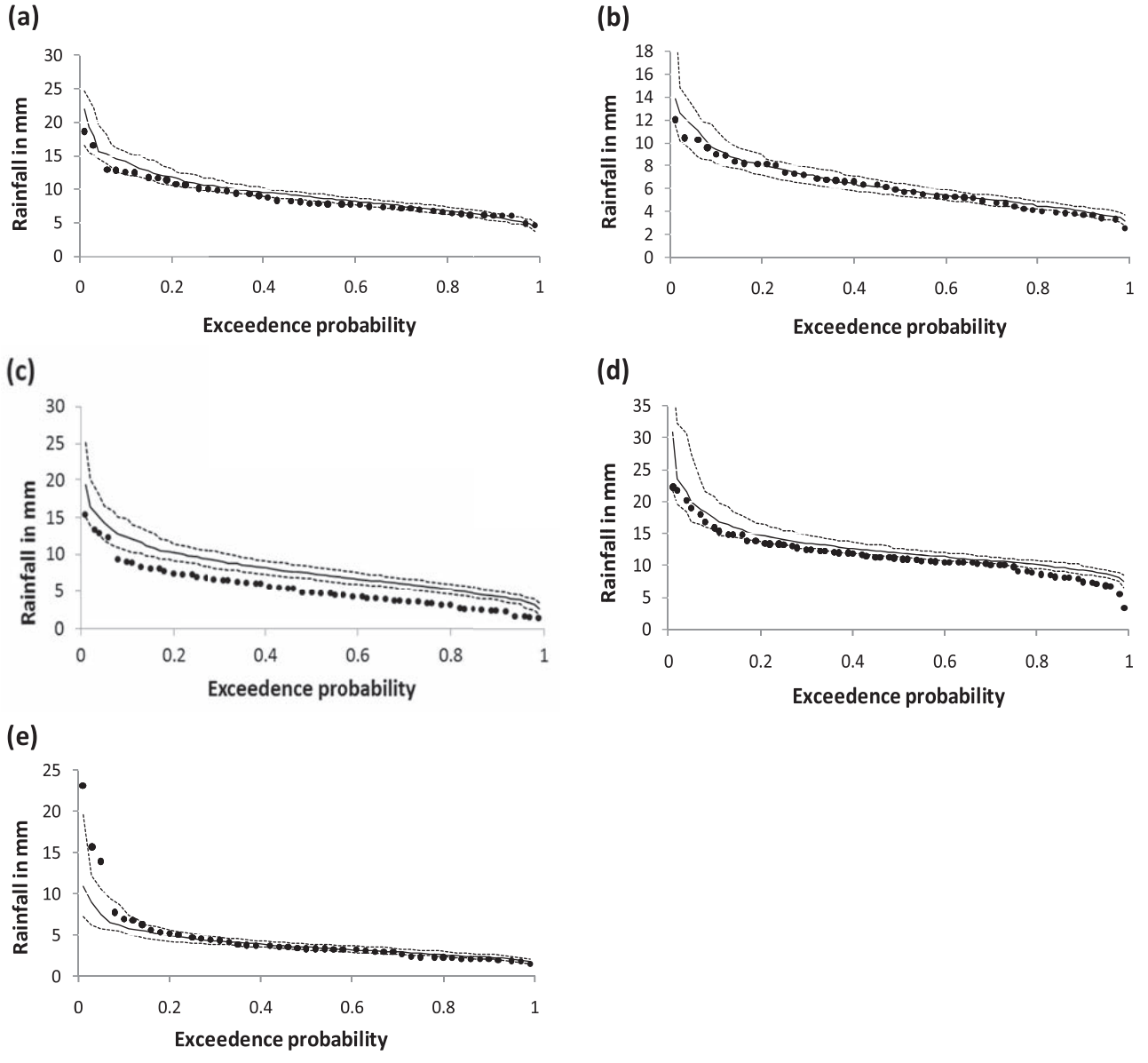

Figure 8. Six minute annual maximum rainfall against exceedance probability for (a) Sydney, (b) Perth, (c) Alice Springs, (d) Cairns, and (e) Hobart. Black dots represents observed data, black solid line represents the median of 100 simulations, and black dotted lines represent the 5th and 95th percentile simulated values.

requested to produce a spatially interpolated map of average annual standard deviation in addition to annual rainfall map across Australia.

4.2. Subdaily Statistics

[65] Results based on the disaggregation of the generated daily rainfall to a subdaily time step are presented in Table 4 and Figures 8 and 9. These results are analogous to Table 4 and Figures 9 and 10 by *Westra et al.* [2011] in which at-site daily rainfall was used but subdaily fragments were sourced from nearby pluviograph stations. Thus, the comparison of these results can be used to determine the impact on precipitation extremes and antecedent precipitation for the case when daily rainfall is also simulated using nearby station records.

[66] As can be seen, the results are very similar to those presented by *Westra et al.* [2011] for all cases, although the confidence intervals are slightly wider suggesting that sourcing daily rainfall information from a greater range of stations increases variance in both extremes and the antecedent conditions leading up to the mean. Nevertheless, these changes are minor and suggest that the regionalization of the daily rainfall model does not result in significant deterioration of simulated subdaily rainfall statistics.

5. Discussion and Conclusion

[67] The objectives of this paper were to present a framework for the substitution of "nearby" daily rainfall records

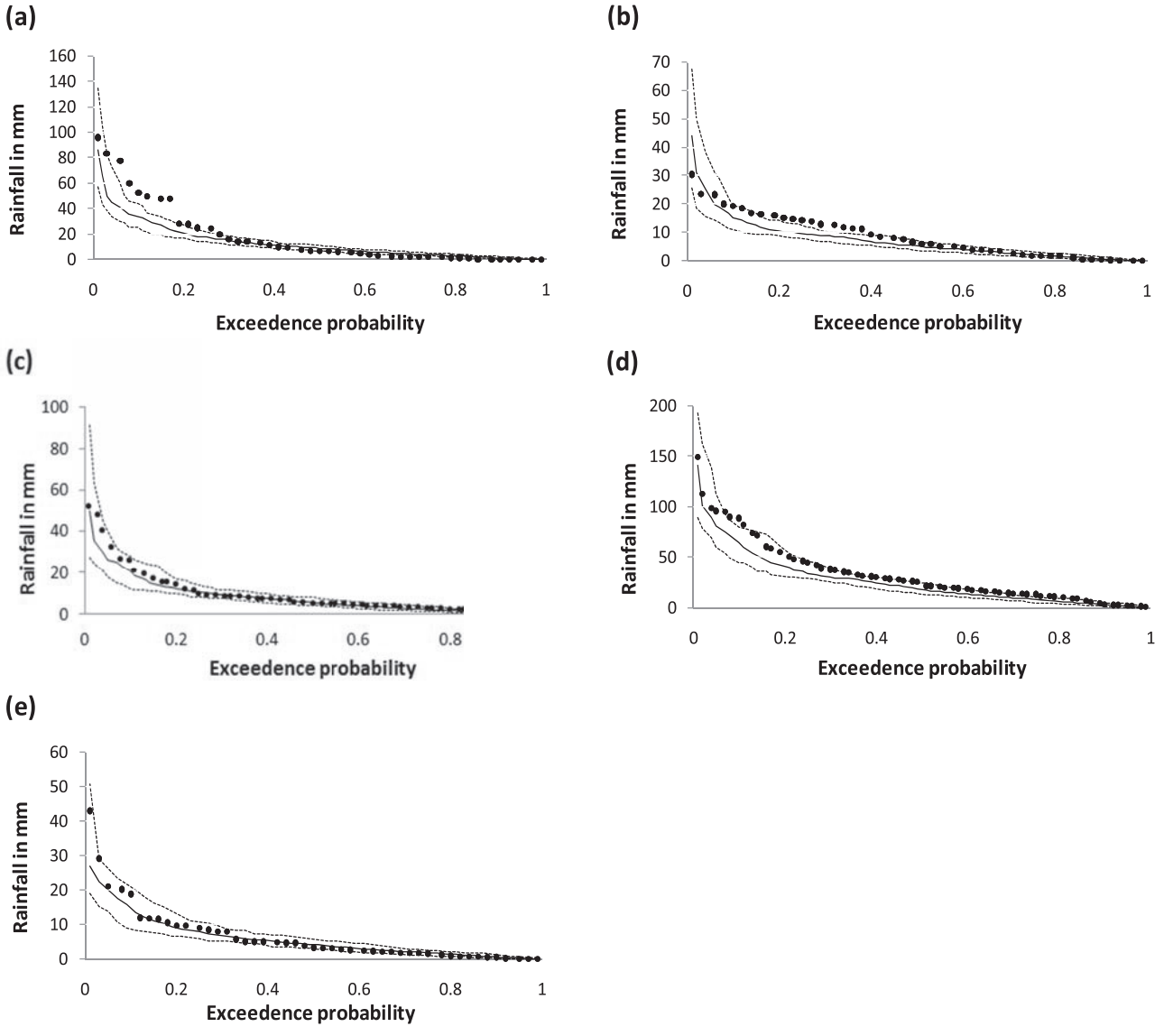

Figure 9. Six hour antecedent precipitation prior to the 6-min annual maximum storm burst plotted against exceedance probability for (a) Sydney, (b) Perth, (c) Alice Springs, (d) Cairns, and (e) Hobart. Black dots represents observed data, black solid line represents the median of 100 simulations, and black dotted lines represent the 5th and 95th percentile simulated values.

in cases where daily rainfall at the target location is either unavailable or too short, and to demonstrate the performance of the approach at a range of locations.

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

[69] The procedure was tested in a cross-validation setting, so that information from only nearby stations was used to estimate the parameters of the rainfall generation model at target locations. The results show that the method performs well in reproducing the rainfall transition probabilities, seasonal and annual number of wet days, and rainfall amounts when there are a large number of daily stations in the vicinity of the target location, although performance did deteriorate for Alice Springs, which is located in a data-sparse region of Australia. In contrast, the standard deviation of both wet days and amounts is typically undersimulated at all locations, although testing showed this was mostly due to the daily rainfall generation model rather than the regionalization procedure. The approach also captures the observed year-to-year variability of annual wet days and rainfall totals in the simulations at all locations except Alice Springs.

[70] Interestingly, the subdaily statistics, namely the annual maxima and the antecedent conditions, are well preserved, and the use of the regionalized daily model results in little deterioration in performance compared to using recorded daily data. This suggests the model is well suited for flood simulation, which requires correct representation of peak rainfall and the moisture conditions in the hours

and days leading up to the event. A range of possible refinements to the model, including adaptively selecting tuning parameters, such as the number of nearby stations S based on the number stations in the vicinity of the target site, or better improving connectivity between days, are warranted to further improve the performance of the algorithm. Nevertheless, based on the analysis in this two-paper series, it is clear that the proposed methodology represents a viable alternative regionalized methodology to generate continuous rainfall data at any location.

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

[72] **Acknowledgments.** This study was supported by an Australian Research Council Discovery grant as well as a research grant from the Institution of Engineers, Australia to help develop continuous rainfall sequences for design flood estimation. The daily and continuous rainfall records used were obtained from the Australian Bureau of Meteorology. We gratefully acknowledge the constructive comments of the anonymous reviewers and Geoff Pegram, whose inputs greatly benefited the quality of our manuscript.

References

Beesley, C. A., A. J. Frost, and J. Zajaczkowski (2009), A comparison of the BAWAP and SILO spatially interpolated daily rainfall data sets, in 18th World IMACS/MODSIM Congress, pp. 3888–3892, Cairns, Australia, 13–17 July 2009, available at http://mssanz.org.au/modsim09.

Boughton, W. C. (1999), A daily rainfall generating model for water yield and flood studies, *Rep. 99/9, CRC for Catchment Hydrology*, Monash University, Melbourne, Australia, 21 pp.

Brandsma, B., and A. T. Buishand (1998), Simulation of extreme precipitation in the Rhine basin by nearest neighbour resampling, *Hydrol. Earth Syst. Sci.*, *2*, 195–209.

Buishand, A. T. (1978), Some remarks on the use of daily rainfall models, *J. Hydrol.*, *36*, 295–308.

Buishand, A. T., and B. Brandsma (2001), Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest neighbor resampling, *Water Resour. Res.*, *37*, 2761–2776.

Gabriel, K. R., and J. Newmann (1962), A Markov chain model for daily rainfall occurrence at Tel Aviv, *Q. J. R. Meteorol. Soc.*, *88*, 90–95.

Guenni, L., and M. F. Hutchinson (1998), Spatial interpolation of the parameters of a rainfall model from ground-based data, *J. Hydrol.*, *212–213*, 335–347.

Harrold, T. I., A. Sharma, and S. J. Sheather (2003a), A nonparametric model for stochastic generation of daily rainfall occurrence, *Water Resour. Res.*, *39*(12), 1300, doi:10.1029/2003WR002182.

Harrold, T. I., A. Sharma, and S. J. Sheather (2003b), A nonparametric model for stochastic generation of daily rainfall amounts, *Water Resour. Res.*, *39*(12), 1343, doi:10.1029/2003WR002570.

Johnson, G. L., C. Daly, G. H. Taylor, and C. L. Hanson (2000), Spatial variability and interpolation of stochastic weather simulation model parameters, *J. Appl. Meteorol.*, *39*, 778–796.

Kyriakidis, P. C., N. L. Miller, and J. Kim (2004), A spatial time series framework for simulating daily precipitation at regional scales, *J. Hydrol.*, *297*, 236–255.

Lall, U., B. Rajagopalan, and D. G. Tarboton (1996), A nonparametric wet/dry spell model for resampling daily precipitation, *Water Resour. Res.*, *32*, 2803–2823.

Mehrotra, R., and A. Sharma (2007a), Preserving low-frequency variability in generated daily rainfall sequences, *J. Hydrol.*, *345*, 102–120.

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

Mehrotra, R., and A. Sharma (2010), Development and Application of a Multisite Rainfall Stochastic Downscaling Framework for Climate Change Impact Assessment, *Water Resour. Res.*, *46*, W07526, doi:10.1029/2009WR008423.

Pui, A., S. Westra, A. Santoso, and A. Sharma (2011), Impact of the El Niño Southern Oscillation, Indian Ocean Dipole, and Southern Annular Mode on daily to sub-daily rainfall characteristics in East Australia, *Monthly Weather Rev.*, in press.

Rajagopalan, B., and U. Lall (1999), A nearest neighbor bootstrap resampling scheme for resampling daily precipitation and other weather variables, *Water Resour. Res.*, *35*(10), 3089–3101.

Rajagopalan, B., U. Lall, and D. G. Tarboton (1996), A nonhomogeneous Markov model for daily precipitation simulation, *J. Hydrol. Eng.*, *1*(1), 33–40.

Scott, D. W. (1992), Multivariate density estimation: Theory, practice and visualization, John Wiley, New York.

Sharma, A., and R. Mehrotra (2010), Rainfall Generation, in *Rainfall: State of the Science*, edited by F. Testik and M. Gebremichael, p. 32, AGU, Washington, D. C.

Sharma, A., and R. O'Neill (2002), A nonparametric approach for representing interannual dependence in monthly streamflow sequences, *Water Resour. Res.*, *38*(7), 1100, doi:10.1029/2001WR000953.

Sharma, A., D. G. Tarboton, and U. Lall (1997), Streamflow simulation: a nonparametric approach, *Water Resour. Res.*, *33*(2), 291–308.

Todorovic, P., and D. A. Woolhiser (1975), A stochastic model of n-day precipitation, *J. Appl. Meteorol.*, *14*, 17–24.

Westra, S., and A. Sharma (2010), An upper limit to seasonal rainfall predictability?, *J. Clim.*, *23*, 3332–3351.

Westra, S., R. Mehrotra, A. Sharma, and R. Srikanthan (2011), Continuous Rainfall Simulation: 1. A regionalised sub-daily disaggregation approach, *Water Resour. Res.*, *48*, W01535, doi:10.1029/2011WR010489.

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

Wilks, D. S., and R. L. Wilby (1999), The weather generation game: A review of stochastic weather models, *Prog. Phys. Geogr.*, *23*(3), 329–357.

R. Mehrotra and A. Sharma, School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia. (a.sharma@unsw.edu.au)

R. Srikanthan, Water Division, Australian Bureau of Meterology, G.P.O. Box 1289, Melbourne, Victoria 3001, Australia.

S. Westra, School of Civil, Environmental and Mining Engineering, University of Adelaide, SA 5005, Australia.