Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

Hydrology and Earth System Sciences
Discussions

Development and Evaluation of a Stochastic Daily Rainfall Model with Long Term Variability

AFM Kamal Chowdhury[1], Natalie Lockart[1], Garry Willgoose[1], George Kuczera[1], Anthony S. Kiem[2], and Nadeeka Parana Manage[1]

[1] School of Engineering, The University of Newcastle, Callaghan 2308, New South Wales, Australia
[2] School of Environmental and Life Sciences, The University of Newcastle, Callaghan 2308, New South Wales, Australia

Correspondence to: Professor Garry Willgoose (garry.willgoose@newcastle.edu.au)

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

Abstract. The primary objective of this study is to develop a stochastic rainfall generation model that can match not only the short resolution (daily) variability, but also the longer resolution (monthly to multiyear) variability of observed rainfall. This study has developed a Markov Chain (MC) model, which uses a two-state MC process with two parameters (wet-to-wet and dry-to-dry transition probabilities) to simulate rainfall occurrence and a Gamma distribution with two parameters (mean and standard deviation of wet day rainfall) to simulate wet day rainfall depths. Starting with the traditional MC-Gamma model with deterministic parameters, this study has developed and assessed four other variants of the MC-Gamma model with different parameterisations. The key finding is that if the parameters of the Gamma distribution are randomly sampled from fitted distributions prior to simulating the rainfall for each year, the variability of rainfall depths at longer resolutions can be preserved, while the variability of wet periods (i.e. number of wet days and mean length of wet spell) can be preserved by decade-varied MC parameters. This is a straightforward enhancement to the traditional simplest MC model and is both objective and parsimonious.

1. Introduction

Observed rainfall data generally provides a single realisation of a short record, often not more than a few decades. The direct application of these data in hydrological and agricultural systems may not provide the necessary robustness in identification and implication of extreme climate conditions (e.g. droughts, floods). In particular, for urban water security analysis of reservoirs, long-term hydrologic records are required to sample extreme droughts that drive the security of the urban system (Mortazavi et al., 2013). However, the observed data may still be suitable to calibrate stochastic rainfall models that can, in turn, be used to generate long stochastic streamflow sequences for use in reservoir reliability modelling. In addition to historical and current scenarios, the stochastic models are useful to evaluate the climate and hydrological characteristics of future climate change scenarios (Glenis et al., 2015).

There is a major issue in the use of stochastic daily rainfall models. The daily models generally efficiently preserve the short-term daily rainfall variability (since they are calibrated to the daily resolution data) but tend to underestimate the longer-term rainfall variability of monthly and multiyear resolutions. Such underestimation is critically important for the application of these models in hydrological planning and design. Preserving the long-term variability is important for drought security analysis of reservoirs because the reservoir water levels usually vary at monthly to multiyear resolutions. The underestimation of longer-term variability of rainfall may cause an overestimation of reservoir reliability in urban water planning (Frost et al., 2007). Therefore, preserving key statistics of wet and dry spells, and rainfall depths in daily to multiyear resolutions is important in stochastic rainfall simulation.

Markov Chain (MC) models are very common for stochastic rainfall generation. A typical MC rainfall model is composed of two parts: a rainfall occurrence model that uses a transition probability between wet and dry days, and a rainfall magnitude model that uses a probability distribution of wet day rainfall depths (commonly a Gamma distribution) fitted to the observed data. The two-part MC-Gamma model is one of the most popular parametric models for daily rainfall simulation, primarily proposed by Richardson (1981) and known as Weather Generator (WGEN). In addition to rainfall, the WGEN also simulates temperature and solar radiation.

The first component of the MC model defines wet and dry days. This is determined by the state and order of the Markov process. Most MC models (Richardson, 1981; Dubrovský et al., 2004) use a simple two-state, first-order approach, that is, a day can be either 'wet' or 'dry' (two-state) and the state of the current day is only dependent on the state of the preceding day (first-order). Other models use higher states and orders – examples include, the four-state model (Jothityangkoon et al., 2000), alternating renewal process model with negative binomial (Wilby et al., 1998) distribution of wet and dry spell lengths, bivariate mixed distribution model (Li et al., 2013), and multi-order model (Lennartsson et al., 2008). These models are more complex as the number of parameters required in the model increases with the number of states and orders of the Markov process. However, the two-state, first-order MC model can often reproduce the statistics of wet and dry periods as well as these higher state/order models (Chen and Brissette, 2014). Dubrovský et al. (2004) recommended that, rather than trying an increased order MC, one should consider other approaches for better reproduction of wet and dry days. Mehrotra and Sharma (2007) proposed a modified MC process using memory of past wet periods, which has been found to reproduce the wet and dry spell statistics reasonably well. They also tested a first-order and a second-order process in their modified MC model and found that the second-order process provided marginal improvements over the first-order process. Another important finding of Dubrovský et al. (2004) was that the order of MC generally had no effect on the variability of monthly rainfall depths.

The second component of the MC model is the probability distribution for the wet day rainfall. As the distribution of wet day rainfall is generally right-skewed (Hundecha et al., 2009), it is common practice to use right-skewed exponential type distributions. Common distributions include the Gamma distribution (Wang and Nathan, 2007; Chen et al., 2010), Weibull distribution (Sharda and Das, 2005), truncated normal distribution (Hundecha et al., 2009), and kernel-density estimation techniques (Harrold et al., 2003). A number of other studies fitted a mixture of two or more distributions, for example mixed exponential distribution (Wilks, 1999a; Liu et al., 2011), Gamma and generalised Pareto distribution (Furrer and Katz, 2008), and transformed normal and generalised Pareto distribution (Lennartsson et al., 2008). However, the Gamma distribution is the most commonly used distribution, because it has two simple parameters calculated from the mean and standard deviation (SD) of wet day rainfall and adequately represents the rainfall distribution functions. The parameterisation and application of the distribution in the model is straightforward. Although the Gamma distribution has been found to be appropriate for simulating most of the variability of rainfall depth (Bellone et al., 2000), the major drawback of using a Gamma distribution is that its tail is too light to capture heavy rainfall intensities (Vrac and Naveau, 2007). Therefore, direct use of a Gamma distribution usually causes an underestimation of SD and extreme rainfall depths at monthly to multiyear resolutions.

A number of methods have been developed in an attempt to resolve the underestimation of long-term variability. The major approaches for resolving this issue include (i) Models with mixed distributions, (ii) Nesting-type models, (iii) Models with rainfall-climate index correlation, and (iv) Models with modified Markov Chain.

The models with mixed distributions concentrate on the upper tail behaviour of the probability distribution of wet day rainfall depths. Since a single component distribution cannot incorporate the extreme rainfall depths well, a mixture of distributions is introduced. In these models, rainfall above a threshold depth is defined as 'extreme' and two separate distributions are used to simulate the 'extreme' and 'small' rainfall amounts. Wilks (1999a) proposed a mixture of two exponential distributions with one shape parameter, but two scale parameters which are used to incorporate the extreme and small rainfall depths respectively. In other models, the 'extreme' rainfall depths are modelled by a generalised Pareto distribution (Vrac and Naveau, 2007) or stretched exponential distribution (Wilson and Toumi, 2005), while small rainfall depths are modelled by a Gamma distribution. Nonetheless, these models have difficulty in defining the threshold corresponding to the 'extreme value'. Wilson and Toumi (2005) defined extreme rainfall as daily totals with exceedance probability less than 5%. Although Vrac and Naveau (2007) used a dynamic mixture to avoid choosing a threshold for 'extreme', Furrer and Katz (2007) described the method as over-parameterised. Recently, Naveau et al. (2016) proposed a new model with smooth transition between the 'small rainfall' and 'extreme rainfall' simulation process to generate low, moderate and heavy rainfall depths without selecting a threshold.

Nesting models adjust the daily rainfall series at different temporal resolutions to obtain statistics that are optimal for all resolutions. These models initially generate a daily rainfall series, which is then modified to adjust the monthly and yearly statistics. Several models (Dubrovský et al., 2004; Wang and Nathan, 2007; Srikanthan and Pegram, 2009; Chen et al., 2010) use the nesting method. They generally generate a daily rainfall series, then the generated daily rainfall data are aggregated to monthly rainfall values, and these monthly values are modified by a lag–1 autoregressive monthly rainfall model. The modified monthly rainfall values are aggregated to annual rainfall values and these values are then modified by a lag–1 autoregressive annual model (Srikanthan and Pegram, 2009). The nesting-type models generally perform well to reproduce the rainfall variability at all resolutions. Dubrovský et al. (2004) also showed satisfactory performance of their nesting-type model to reproduce the variability of monthly streamflow characteristics and the frequency of extreme streamflow. Although the nesting-type models preserve the daily, monthly and yearly statistics, they are generally based on statistical adjustments and thus have a weak physical basis.

Some parametric models introduced the influence of the large-scale climate mechanisms such as El Niño/Southern Oscillation (ENSO) in parameterisation (Hansen and Mavromatis, 2001; Furrer and Katz, 2007). Bardossy and Plate (1992) used the correlation between atmospheric circulation patterns and rainfall in a transformed conditional multivariate autoregressive AR (1) model for daily rainfall simulation. Katz and Parlange (1993) developed a model with parameters conditional on the ENSO indices. Yunus et al. (2016) developed a generalised linear model for daily rainfall by using ENSO indices as predictors. Although the climate indices were often not strongly correlated to the rainfall, Katz and Zheng (1999) described it as a diagnostic element to detect a 'hidden' (i.e. unobserved) index which could be used to obtain long-term variability. Thyer and Kuczera (2000) developed a hidden state MC model for annual data, while Ramesh and Onof (2014) developed a hidden state MC model for daily data. The major drawback of this model approach is that the 'hidden' index is unobserved and its origin is unknown.

Modified MC models concentrate not only on the order of MC, but also introduce modifications to the parameterisation of the MC process to better reproduce the rainfall variability. The transition probabilities are generally modified by considering their long-term variability (i.e. memory of past wet and dry periods), and the wet day rainfall depth is modelled using a nonparametric kernel-density simulator conditional on previous day rainfall (Lall et al., 1996; Harrold et al., 2003). The nonparametric kernel-density techniques usually used resampling of observed data (Rajagopalan and Lall, 1999). While these models perform reasonably well, they usually cannot generate extreme values higher than the observed extremes, because only the original observations are resampled in the model (Sharif and Burn, 2006). Mehrotra and Sharma (2007) proposed a semi-parametric Markov model, which was further evaluated by Mehrotra et al. (2015). To incorporate the long-term variability, they modified the transition probabilities of the MC process by taking the memory of past wet periods into account, while the wet day rainfall depths were simulated by a nonparametric kernel-density process. For raingauge data around Sydney, the semi-parametric model preserved the rainfall variability at daily to multiyear resolutions (Mehrotra et al., 2015).

The MC models that focus specifically on resolving the underestimation of long-term variability involve subjective assumptions and limitations. In the models with mixed distributions, defining a certain rainfall depth as an extreme value has a weak physical basis. The nesting type models used empirical adjustment factors, generally without physical foundation. The hidden indices of hidden state MC models are unobserved. The models with modified MC parameters modified the transition probabilities of wet and dry periods to obtain long term variability, but used the kernel density technique to resample wet day rainfall depths from observed records. Therefore, they usually cannot generate extreme values higher than the observed extremes.

The overarching objectives of the research, that this paper forms part of, is to develop a stochastic rainfall generator that can be calibrated to daily rainfall data derived from dynamically downscaled global climate simulations (Evans et al., 2014). A common problem with these simulations is that typical computational CPU limits mean that the length of the simulation is rarely more than a few decades, not long enough to facilitate stochastic assessment of the reliability of water supply reservoirs (e.g. Lockart et al., 2016). Accordingly we need a rainfall climate simulator that can be calibrated and run at the daily timescale (to be used as input into a hydrology model at the daily resolution), but which has the right statistical properties (specifically variability about the mean) when averaged over periods out to a decade. In this paper we develop and test five models against observed rainfall at two sites in Australia with contrasting climate. In subsequent papers we will look at the use of these models with dynamically downscaled climate data.

Accordingly, this study details the development of a MC model for stochastic generation of daily rainfall. This MC model uses a two-state MC process with two parameters (wet-to-wet and dry-to-dry transition probabilities) for simulating rainfall occurrence and a two parameter Gamma distribution (mean and SD of wet day rainfall) for simulating wet day rainfall depths. Five variants of the MC model, with gradually increasing complexity of parameterisation, are developed and assessed. Starting with a very simple model against which the performances of the other models will be compared, each of the successive models provides better performance in reproducing the variability and dependence of observed rainfall over the range of resolutions from day to decade.

2. Data and Study Sites

In the development and assessment of the stochastic rainfall models, this study has used daily raingauge data from Sydney Observatory Hill and Adelaide Airport stations (station number 66062 and 023034 respectively) obtained from the Bureau of Meteorology (BoM), Australia (Figure 1) for 1979-2008 (BoM, 2013). These two stations have been used because they provide a contrast between a highly seasonal Mediterranean climate with low inter-decadal variability in Adelaide and a relatively non-seasonal climate with high inter-decadal variability in Sydney (see Figure 2). This paper has also used Oceanic Nino Index (ONI) and Inter-Decadal Pacific Oscillation (IPO) index at monthly resolution for 1979–2008 period (Folland, 2008; NOAA, 2014). These climate indices are used to develop two variants of the MC models discussed in section 4.2.2.

3. Model Assessment Procedures

3.1. Statistics for Assessment

Each model developed in this study has been assessed to understand its ability to reproduce the distribution and autocorrelation of observed rainfall. Assessment of the distribution and autocorrelation are generally used to inform the suitability of the model to be used in urban drought security assessment. The assessment criteria of each model considers its ability to reproduce (i) mean and SD of rainfall depths, number of wet days and mean length of wet spells at daily to multiyear resolutions, and (ii) month-to-month autocorrelations of monthly rainfall depths and monthly number of wet days. The performances of the MC models for dry period statistics are found to be complementary to the wet period statistics ('wet period statistics' will refer the number of wet days and mean length of wet spell hereafter), and hence, not shown.

At daily and monthly resolutions, the distribution statistics are assessed for each month starting from January, while at multiyear resolutions, the distribution statistics are assessed for 1 to 10 overlapping years. Mean length of wet spells are calculated at monthly and annual resolution by extracting the *1*, *2*, …, *n* consecutive wet days and using equation 1:

$$\text{mean length of wet spell} = \frac{\sum(\text{length of wet spells})}{\sum(\text{occurrences of wet spells})} \quad (1)$$

3.2. Bootstrapping and Calculation of Z Scores

For the distribution statistics (i.e. mean and SD) of rainfall depths and wet periods (number of wet days and mean length of wet spells), this study has used bootstrapping to calculate the expected value and 95% confidence limit (2SD) and then the Z score of a model simulation. The bootstrapping and calculation of Z score are as follows:

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

- Run the model using parameters calibrated to the observed data 1000 times, with each run being the same length as the observed data.
- Calculate the desired statistics (e.g. mean and SD of the daily rainfall depths) in each run, which gives 1000 realisations of each statistic.
- For each statistic, calculate the mean (expected value) and SD (error limit) of the 1000 realisations.
- Calculate the Z Score of a statistic by comparing the expected value with the respective observed value (calculated from the observed data), as follows:

$$\text{Z Score} = \frac{\text{Observed value - Expected value}}{\text{SD}} \tag{2}$$

A Z Score between –2 and +2 for a statistic indicates that the observed value falls within the 95% confidence limits of the simulated rainfall assuming a normal distribution approximates the sampling distribution of Z. A Z Score less than –2 or greater than +2 suggests that the statistic is over- or under-estimated respectively in the model simulation.

4. Markov Chain (MC) Models

This study has developed and assessed the following five variants of a Markov Chain (MC) model:

- Model 1: Average Parameter Markov Chain (APMC) model,
- Model 2: Decadal Parameter Markov Chain (DPMC) model,
- Model 3: Compound Distribution Markov Chain (CDMC) model,
- Model 4: Hierarchical Markov Chain (HMC) model,
- Model 5: Decadal and Hierarchical Markov Chain (DHMC) model.

4.1. Model 1: Average Parameter Markov Chain (APMC) model

The first MC model – the APMC – is a traditional two-part MC-Gamma distribution model. This is similar to the rainfall generator proposed by Richardson (1981), widely known as the Weather Generator (WGEN) model. The exception is that the parameters in WGEN were smoothed with Fourier harmonics, which has not been done in the case of APMC parameters. Although APMC is not the final model of this study, it forms the baseline of the modelling approaches against which the more sophisticated models developed in this study are compared.

The APMC simulates the daily rainfall in two steps: daily rainfall occurrence (i.e. wet and dry day) simulation by first-order Markov Chain, and wet day rainfall depth simulation by Gamma distribution. To incorporate the seasonal variability in the model, the APMC uses a separate set of parameters for each month, where the first month of the simulation is January.

4.1.1. Rainfall occurrence simulation

The APMC uses 24 (2 × 12) MC parameters, transition probabilities of dry-to-dry day (P_{00}) and wet-to-wet day (P_{11}), for wet and dry day occurrence simulation. In addition, the unconditional probability of a dry day (π_0) in January is used to simulate rainfall occurrence for the first day of the series. In the model calibration, these deterministic MC parameters are calculated from the observed daily rainfall data. To calculate these parameters, a day with rainfall depth of 0.3 mm and above has been considered a wet day, otherwise it was considered a dry day (similar to Mehrotra et al, 2015). In simulation, the MC parameters are used in a Monte-Carlo process to simulate the occurrences of wet and dry days.

4.1.2. Rainfall depth simulation

After simulation of the rainfall occurrence using MC parameters, the next step is to generate rainfall depths for the wet days. The rainfall depth for dry days is zero. The APMC rainfall depth simulation process assumes that (i) daily rainfall depth for wet days follows a Gamma distribution, and (ii) the rainfall depth for a wet day is independent of the rainfall depth of the preceding day.

The Gamma distribution has two parameters α (shape parameter) and β (scale parameter) with mean $\mu = \alpha\beta$ and variance $\sigma^2 = \alpha\beta^2$. Since both α_i and β_i are directly proportional to and can be derived from μ_i and σ_i of wet day rainfall of the month i, during calibration of the model it is convenient to calculate μ_i and σ_i values from the daily rainfall observed data. The appropriate ratios of μ_i and σ_i can then be used in the rainfall depth generation process using the Gamma distribution. Therefore, μ_i and σ_i will be referred to as the Gamma distribution parameters in further discussions of this paper.

In calibration of APMC, deterministic values of μ_i and σ_i are calculated from the entire period of data record for each month. This gives 12 values of μ and σ each. In simulations, the rainfall depth for each wet day of a month i is generated using the μ_i and σ_i values of the respective month using the Gamma distribution. In generating the rainfall depth for a wet day, if a random sample from the Gamma distribution gives a rainfall depth less than 0.3 mm then the rainfall for that day is set to 0.3 mm (i.e. the threshold rainfall depth), while the rainfall depths for dry days are set to 0.0 mm.

4.1.3. Independence of rainfall depths in successive wet days

The APMC assumes that the rainfall depth for a particular day is independent of the rainfall depth of the preceding day. To validate this assumption, this study examined the autocorrelation of wet day rainfall depths, and only found very weak lag-1 autocorrelations ($r^2 < 0.1$) for both Sydney and Adelaide. This finding is consistent irrespective of seasonal variations. The conclusion is that the underlying assumption of daily independence of the APMC is consistent with the respective characteristic of the observed data.

4.2. Model 2: Decadal Parameter Markov Chain (DPMC) Model

Section 6 will show that the APMC significantly underestimates the rainfall variability at monthly to multiyear resolutions. The DPMC assumes that the inter-annual rainfall variability can be captured by the decade-to-decade variability of the parameters that APMC failed to capture. The idea is to divide the observed rainfall sample into sub-samples of 10-year lengths (similar models with climate-based sub-samples are discussed in section 4.2.2). For example, a 30-year rainfall sample is divided into three sub-samples of 10-years in length. Then, 4 × 12 parameters of P_{00}, P_{11}, μ, and σ (one set of four parameters for each of the 12 months) are calculated from each of the sub-samples. The simulation proceeds in a way similar to the APMC, except that the deterministic parameters of DPMC are varied from decade to decade.

4.2.1. Decadal variability of DPMC parameters

Figure 2 shows the DPMC values of P_{11} and μ for each decade along with APMC values (i.e. the 30-year averages) for Sydney and Adelaide. For Sydney, DPMC values of P_{11} and μ show clear variabilities among the three decadal samples and deviations from the APMC values. However, DPMC values of P_{11} and μ for Adelaide show less variability among the decadal samples.

The use of decade-varied parameters in DPMC is subject to the question of how significant the decadal variability of these parameters is – is the decadal variability statistically significant or just sampling variability? Therefore, the statistical significance of the decadal variability of DPMC parameters are examined by parametric bootstrapping as per section 3.2.

The parametric bootstrapping of the DPMC parameters indicates that the sampling variability of these parameters in decadal samples is mostly within the sampling variability of their corresponding APMC values (not shown). This suggests that the decadal variability of DPMC parameters is not statistically significant.

4.2.2. Potential impact of climate modes

This study has also investigated other sub-sampling approaches of the MC-Gamma parameters similar to the DPMC. In these models, this study has calibrated the MC-Gamma parameters to sub-samples of rainfall timeseries divided according to the phases of IPO (e.g. positive and negative) and ENSO (La Niña, Neutral and El Niño). Since previous studies (Verdon-Kidd et al., 2004) found that the inter-annual variabilities of East-Australian rainfall are influenced by these large-scale climate drivers, the idea behind these models was to introduce more inter-annual variability to the model by simulating rainfall for different phases of climate drivers with parameters calibrated to respective phases. These climate-based models are very similar to DPMC, except that the sub-samples are different. The following two types of climate-based models have been tested:

- For model according to IPO phases, the observed data for every month was divided into two sub-samples according to the positive and negative values of monthly IPO index (e.g. for January, data of the years with positive IPO index and data of the years with negative IPO index are separated). Then, for each month, the MC-Gamma parameters (P_{00}, P_{11}, μ, and σ) are calibrated to each sub-sample. In simulation, the rainfall for the months of each IPO phase were modelled by using parameters of respective phase.
- For model according to ENSO phases, the observed data for every month was divided into three sub-samples according to monthly ONI index: La Niña (ONI ≤ –0.5), Neutral (–0.5 < ONI < 0.5), and El Niño (ONI ≥ 0.5). Then, the MC-Gamma parameters are calibrated to each sub-sample and the rainfall for the months of each ENSO phase were modelled by using parameters of the respective phase.

4.3. Model 3: Compound Distribution Markov Chain (CDMC) Model

The results in section 6 will show that the APMC and DPMC cannot satisfactorily reproduce the SD of rainfall depths for monthly to multiyear resolutions. Therefore, in the third MC model – the CDMC – this study has incorporated the long-term variability of rainfall depths by introducing random variability in μ and σ. However, for wet and dry period simulation, the CDMC still uses the deterministic parameters of P_{00} and P_{11}, as in the APMC.

In the CDMC, μ_i and σ_i are randomly sampled for each month of each year. The random sampling was done independently of the sampling for the preceding month/s. To estimate the distribution of μ_i and σ_i, this study has calculated μ_i and σ_i for every month of every year from the observed data. For example, from the 30-year observed data, for January (i = 1), this study has calculated 30 samples of μ_1 and σ_1 values each.

By testing the probability distributions of μ_i and σ_i values for each month, this study has found that both μ_i and σ_i values for each month are lognormally distributed (i.e. best suitable distribution). Figure 3 shows the lognormal probability plots of μ_i and σ_i values for July (i = 7) as a representative month. The r^2 for log μ_i and log σ_i are generally above 0.90, indicating a very good fit of the lognormal distributions. Additionally, the hypothesis that log μ_i and log σ_i are normally distributed is supported by the Kolmogorov-Smirnov test at 5% significance level. In addition to the lognormally distributed μ_i and σ_i values, this study has also found that the log μ_i and log σ_i values for each month are strongly correlated with each other with correlation coefficient $r_{c,i}$ around 0.90 (Figure 4). Therefore, for each month i, this study has fitted a bivariate-normal distribution to the log μ_i and log σ_i values with parameters ($\lambda_{\mu,i}$, $\zeta_{\mu,i}$), ($\lambda_{\sigma,i}$, $\zeta_{\sigma,i}$) and $r_{c,i}$. The λ and ζ parameters denote the mean and SD of the log variate, while r_c is the correlation coefficient between log μ and log σ.

At the start of each month of each year of the simulation, the log μ_i is sampled from its fitted normal distribution log $\mu_i \sim N(\lambda_{\mu_i},\ \zeta^2_{\mu_i})$ for month *i*. Then, the log σ_i is sampled from the fitted conditional normal distribution:

$$\log \sigma_i \,|\, \log \mu_i \ \sim\ N\left(\lambda_{\sigma_i} + \frac{\zeta_{\sigma_i}}{\zeta_{\mu_i}} r_{c,i}\left(log\mu_i - \lambda_{\mu_i}\right),\quad \left(1 - r^2_{c,i}\right)\left(\zeta_{\sigma_i}\right)^2\right) \tag{3}$$

These stochastically sampled μ_i and σ_i values are then used to generate rainfall in the wet days for the month in question, while the sequence of wet and dry days is determined using the deterministic APMC values of $P_{00,i}$ and $P_{11,i}$. However, the sampled μ_i and σ_i values of a month (i) are not correlated to the μ_{i-1} and σ_{i-1} of the preceding month (i–1) as this study has found that the month-to-month autocorrelations of μ and σ values are not significantly strong (Figure 5).

Calibration of CDMC to gridded data of New South Wales/Australian Capital Territory Regional Climate Modelling project and Australian Water Availability Project in a separate case study site in East-Australia was previously published in Chowdhury et al. (2015).

4.4. Model 4: Hierarchical Markov Chain (HMC) Model

The results in section 6 will show that the CDMC cannot satisfactorily reproduce the SD of wet periods for monthly to multiyear resolutions. Therefore, in the fourth MC model – the HMC – this study has introduced stochastic parameterisation of both MC and the Gamma distribution to incorporate long-term variability of rainfall depths as well as wet and dry periods. In calibration, for month *i*, the $P_{00,i}$ and $P_{11,i}$ are calculated for each month of each year from the observed data. For month *i*, these $P_{00,i}$ and $P_{11,i}$ values (e.g. 30 $P_{11,7}$ values for July from the 30-year observed data) are found to be normally distributed with values between 0 and 1 (Figure 6). Therefore, this study has fitted a truncated normal distribution, bounded by 0 and 1 to the calculated P_{00} and P_{11} values. In simulation, for each year, the $P_{00,i}$ and $P_{11,i}$ are sampled from their truncated normal distributions. This procedure is similar to what was done for μ_i and σ_i to sample from bivariate-lognormal distribution. However, it does not include a bivariate distribution because the correlation between $P_{00,i}$ and $P_{11,i}$ was weak.

4.4.1. Impact of autocorrelations on stochasticity of MC parameters

In the HMC, the sampled MC parameters of each month are independent of the parameters of preceding month. However, this study has found strong month-to-month autocorrelations of the P_{00} and P_{11} for Adelaide (Figure 5a), although the autocorrelations are weak for Sydney (Figure 5b). Therefore, this study has tested an alternative to the HMC by using a lag–1 autocorrelation equation (similar equation was used by Wang and Nathan (2007) in their rainfall depth model) in the stochastic sampling of $P_{00,i}$ and $P_{11,i}$ from the truncated normal distribution. The following lag–1 autocorrelation equation has been used to modify the randomly sampled $P_{00,i}$ (same method used for $P_{11,i}$) for month *i* by correlating with the $P_{00,i-1}$ of month *i*–1 (preceding month):

$$\frac{\overline{\overline{P_{00,i}}} - mean(P_{00,i})}{sd(P_{00,i})} = r\times \frac{\overline{\overline{P_{00,i-1}}} - mean(P_{00,i-1})}{sd(P_{00,i-1})} + (1 - r^2)^{1/2}\, \frac{\overline{P_{00,i}} - mean(P_{00,i})}{sd(P_{00,i})} \tag{4}$$

where, for a month *i* (e.g. January),

- $\overline{\overline{P_{00,i}}}$ is auto-correlated parameter (which is used in simulation) for month *i*,

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

Hydrology and Earth System Sciences
Open Access
Discussions

- $\overline{P_{00,i}}$ is parameter value sampled from the truncated normal distribution for month *i*,
- *r* is lag–1 autocorrelation coefficient (constant for all month),
- $mean(P_{00,i})$ is mean of the parameter values calculated from observed data for month *i*,
- $sd(P_{00,i})$ is SD of the parameter values calculated from observed data for month *i*,
- $\overline{\overline{P_{00,i-1}}}$ is auto-correlated parameter for month *i*–1 (preceding month),
- $mean(P_{00,i-1})$ is mean of the parameter values calculated from observed data for month *i*–1,
- $sd(P_{00,i-1})$ is SD of the parameter values calculated from observed data for month *i*–1.

4.4.2. Impact of cross-correlations on stochasticity of MC parameters

This study has also observed strong positive correlations of $P_{11,i}$ with the log μ_i and log σ_i, although the correlations of $P_{00,i}$ with the log μ_i and log σ_i are weak. Therefore, another alternative to HMC is tested by using a multivariate sampling system for the $P_{11,i}$, μ_i and σ_i, while $P_{00,i}$ remains independent.

4.5. Model 5: Decadal and Hierarchical Markov Chain (DHMC) Model

Section 6 will show that the CDMC, with APMC values of MC parameters, significantly underestimates the wet period variability at multiyear resolutions, while the HMC (including the two alternatives) with stochastic MC parameters, significantly overestimates the wet period variability at monthly resolution. However, this study has found that the DPMC can satisfactorily preserve the wet period variability at both monthly and multiyear resolutions, although it underestimates the rainfall depths variability. Therefore, in the DHMC model, this study has used the DPMC values of MC parameters (the parameter values vary for each decade of simulation) for simulation of wet and dry days, while the stochastic parameters of Gamma distribution (same as CDMC) are used for simulation of wet day rainfall depths.

5. Methodological Comparison of Five MC Models

The following points discuss the key common features in the five MC models of this study, while other key methodological comparisons are shown in Table 1.

- All models use first-order MC parameters to simulate the rainfall occurrences and Gamma distribution to simulate rainfall depths in wet days.
- Simulation of rainfall depth for each wet day is independent of the rainfall depth of the preceding day.
- Separate sets of parameters are used for each month (e.g. 12 sets of MC and Gamma parameters) to incorporate seasonal variability.

6. Model Comparison for Distribution Statistics

This section compares the performances of five MC models for the mean and SD of rainfall depths and wet period statistics (i.e. number of wet days and mean length of wet spell).

6.1. Mean and SD of Rainfall Depths

Figure 7 and 8 compare the five MC models for the mean and SD of rainfall depths at monthly and multiyear resolutions respectively. For mean and SD of rainfall depths, the performances of APMC and DPMC are similar. The performances of CDMC, HMC and DHMC are also similar, but different from APMC and DPMC. All five models preserve the mean (i.e.

satisfactorily reproduce the observed mean) rainfall depths at all resolutions. However, the CDMC, HMC, and DHMC show a tendency to underestimate the mean with Z scores mostly between 0 and +2. The APMC and DPMC significantly underestimate the SD of rainfall depths at monthly and multiyear resolutions for Sydney but preserve the SDs for Adelaide (Figure 7 and 8), while the inter-decadal variabilities of parameters are less in Adelaide and high in Sydney (Figure 2). We conclude that those models with stochastic parameters for the Gamma distribution (i.e. CDMC, HMC, and DHMC) best preserve SDs at all resolutions for both stations.

6.2. Mean and SD of Number of Wet Days

Figure 9 and 10 compare the five MC models for the mean and SD of number of wet days at monthly and multiyear resolutions respectively. All five models preserve the mean of number of wet days at both monthly and multiyear resolutions, while the HMC tends to overestimate the statistic. For SD of monthly number of wet days, all models except HMC can satisfactorily reproduce the SD, while the HMC tends to overestimate the statistic (Figure 9). For SD of multiyear number of wet days, the APMC and CDMC significantly underestimate the SD for Sydney but preserve the statistic for Adelaide. The DPMC and DHMC perform similarly and satisfactorily to preserve the SD of multiyear number of wet days for both Sydney and Adelaide, while HMC also preserves the statistic for both stations. We conclude that the models with stochastic, yearly varied, parameters for the MC part of the model (i.e. HMC) perform relatively poorly at reproducing the variability of the number of wet days per month.

6.3. Mean and SD of Mean Length of Wet Spells

The comparative performances of the five MC models for the mean and SD of mean length of wet spells at monthly and annual resolutions are mostly consistent with their respective performances for mean and SD of number of wet days. All models except HMC preserve the mean and SD of mean length of wet spells, while the HMC tends to overestimates the SD (Figure 11). We conclude that models with stochastic, yearly varied, parameters for the MC part of the model (i.e. HMC) perform relatively poorly at reproducing the variability of the length of wet spells.

6.4. Potential Impact of Climate Modes

Since the DPMC significantly underestimates the SD of rainfall depths at monthly and multiyear resolutions, the major target of the models with sub-samples according to climate modes such as IPO and ENSO indices (discussed in section 4.2.2) was to preserve the SD of rainfall depths at monthly and multiyear resolutions. However, these climate-based models also significantly underestimate the SD of rainfall depths at month and multiyear resolutions with performances similar to the DPMC.

6.5. Impact of Stochasticity of MC Parameters

Since the HMC significantly overestimates the SD of monthly wet periods, the major target of the HMC-like models with a lag–1 autocorrelation equation and a multivariate sampling system (see section 4.4.1) was to preserve the SD. However, these models also significantly overestimate the SD of monthly wet periods with performances similar to the HMC (negative Z scores less than -2 for all months).

7. Model Comparison for Autocorrelations

Figure 12 compares how the five MC models reproduce the month-to-month autocorrelations of the monthly number of wet days and monthly rainfall depths. For Adelaide (Figure 12a), the lag–1 and lag–12 autocorrelations are strong with systematic seasonal variation, which have been reproduced very well in the corresponding APMC, DPMC, CDMC and

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

DHMC simulations, while the HMC (the model with stochastic MC parameters) tends to underestimate the autocorrelations. For Sydney (Figure 12b), the month-to-month autocorrelations of monthly number of wet days and monthly rainfall depths are weak and all models perform well.

8. Discussion

The primary motivation of this study is to develop a stochastic rainfall generation model that can match not only the short resolution (daily) variability, but also the longer resolution (monthly to multiyear) variability of observed rainfall. Preserving long-term variability in rainfall models has been a difficult challenge for which a number of solutions have been proposed in the stochastic rainfall generation literature. The solutions developed and tested by this study are relatively simple MC models: two MC parameters (P_{00} and P_{11}) of two-state, first-order processes defining the wet and dry days, and two Gamma-distribution parameters (μ and σ) defining the rainfall depths in wet days. For seasonal variability, the models operate at daily time step with monthly varying parameters for each of 12 months. Starting with the simplest MC-Gamma modelling approach with deterministic parameters (similar to Richardson, 1981), this study has developed and assessed four other variants of the MC-Gamma modelling approach with different parameterisations. The key finding is that if the parameters of the Gamma distribution are randomly sampled from fitted distributions prior to simulating the rainfall for each year, the variability of rainfall depths at longer resolutions can be preserved, while the variability of wet periods (i.e. number of wet days and mean length of wet spell) can be preserved by decade-varied parameters. This is a straightforward enhancement to the traditional simplest MC model, and the enhancement is both objective and parsimonious.

The overall comparative performances of the models to reproduce the distribution and autocorrelation characteristics of observed rainfall are as follows:

- For simulation of the distribution of rainfall depths, the trend of performances of the APMC and DPMC with deterministic Gamma parameters are similar, although DPMC with more (e.g. three times more) parameters performs slightly better. The performances of CDMC, HMC and DHMC are similar as they use the same stochastic sampling for the parameters of the Gamma distribution.
- For mean and SD of daily rainfall depths, all five models perform satisfactorily. Good reproduction of daily statistics is expected as the model parameters are calibrated to daily timeseries. While the APMC and DPMC reproduce the statistics almost exactly, the CDMC, HMC and DHMC show a slight tendency to underestimate the statistics. This indicates that the stochastic parameters of these three models slightly affected their performances at daily resolution compared to the APMC and DPMC with deterministic parameters.
- At monthly to multiyear resolution, the APMC and DPMC reproduce the mean of rainfall depths well, but significantly underestimate the SD of rainfall depths. The underestimation of rainfall variability at monthly to multiyear resolutions by APMC-like models with deterministic parameters is a well-known limitation of deterministic parameter (i.e. APMC-like) models (Wang and Nathan, 2007). Although the DPMC uses more parameters than the APMC, the DPMC has not significantly improved performance in reproducing the SD of rainfall depths at monthly to multiyear resolutions. Other models similar to DPMC (e.g. models with parameters varying for phases of IPO or ENSO) show similar performances to the DPMC and systematically underestimate the SD of rainfall depths at monthly to multiyear resolutions. This suggests that the use of deterministic parameters in DPMC-like models might not be adequate to satisfactorily reproduce the SD of rainfall depths at longer resolutions.
- While the APMC and DPMC, with deterministic parameters for the Gamma distribution, significantly underestimate the SD of rainfall depths at monthly to multiyear resolutions, the CDMC, HMC and DHMC, with

stochastic parameters for the Gamma distribution, preserve the SD of rainfall depths at monthly to multiyear resolutions. This indicates that the stochastic parameters for the Gamma distribution are useful to incorporate the longer-term variability of rainfall depths. However, these three models show a tendency to underestimate the mean of rainfall depths at all resolutions.

- For simulation of the distribution of wet periods, the performances of the APMC and CDMC are similar as both models use the same deterministic MC parameters. With a similar trend, the DPMC and DHMC perform better than the APMC and CDMC, while DPMC and DHMC use more deterministic MC parameters. The performances of the HMC, with stochastic MC parameters, is different (discussed below) from the other four models with deterministic MC parameters.
- For mean of wet period statistics (e.g. number of wet days and mean length of wet spells) at monthly to multiyear resolutions, all models except HMC perform similarly and satisfactorily, while the HMC tends to overestimate the mean. We conclude that introducing stochasticity from year to year into the MC parameters, as in HMC, degrades the performance.
- For SD of monthly wet period statistics, all models except HMC perform similarly and satisfactorily, while the HMC significantly overestimates the SD. This indicates that the stochastic MC parameters of the HMC introduce excessive variability in the wet period simulation at monthly resolution. This study has further examined two other variants of the HMC with different stochastic parameterisation of the MC process, but they did not perform better than the HMC. We conclude that introducing stochasticity from year to year into the MC parameters, as in HMC, degrades the ability to reproduce the variability about the mean of all of the wet period statistics.
- For SD of wet period statistics at annual and multiyear resolutions, the APMC and CDMC tend to underestimate the SDs. This suggests that the APMC values of MC parameters (same monthly parameter values for each year of simulation) limits the reproduction of the wet period variability at multiyear resolutions. However, the APMC and CDMC preserved the multiyear SDs for Adelaide, where the inter-decadal variability of MC parameters is less variable. This suggests that for locations with less variability of wet-to-wet and dry-to-dry day transitions, a single set of deterministic MC parameters is adequate, however for locations with more transition variability, a single set of MC parameters is insufficient, as it cannot introduce enough variability.
- The DPMC and DHMC with decade-varied MC parameters show better and satisfactory ability to reproduce the SD of annual mean length of wet spells and SD of multiyear number of wet days. This suggests that the decade-varied MC parameters can significantly improve the simulation of wet period variability, although the decade-varied Gamma parameters cannot improve the simulation of rainfall depths variability. However, the HMC preserves the SD of multiyear number of wet days but overestimates the SD of annual mean length of wet spells. This suggests that the monthly and annually varying stochastic MC parameters can improve the simulation of wet period (i.e. number of wet days and mean length of wet spell) variability at multiyear resolutions, although they significantly overestimate the wet period variability at monthly and annual resolutions (i.e. they introduce too much variability).
- The autocorrelation assessments have shown that the APMC, DPMC, CDMC and DHMC can satisfactorily reproduce the strong lag–1 and lag–12 autocorrelations of monthly number of wet days and monthly rainfall depths. However, the HMC (the only model with monthly and annually varying MC parameter values) tends to underestimate the autocorrelations, which is possibly due to excessive variability in wet period simulation at monthly resolution.

9. Conclusions

Each model developed in this study has advantages and disadvantages. The APMC and DPMC with deterministic parameters significantly underestimate the variability of rainfall depths at monthly to multiyear resolutions. This systematic underestimation of the rainfall depths variability at monthly to multiyear resolutions is critical for using the models in urban water security assessment as the reservoir water levels usually vary at these longer resolutions. The CDMC, HMC and DHMC with stochastic parameters of the Gamma distribution preserve the rainfall depths variability at all resolutions, but the CDMC and HMC have limitations in reproducing the variability of wet periods. The CDMC with APMC values of MC parameters tends to underestimate the multiyear variability of wet periods, while the HMC with stochastic MC parameters tends to overestimate the monthly variability of wet periods. However, the DHMC with decade-varied MC parameters (same as DPMC) performs better than the CDMC and HMC, and preserves the wet period variability at monthly to multiyear resolutions.

Among the five MC models of this study, the overall performance of the DHMC is better than the other four MC models. In summary the DHMC model has (1) monthly varying MC parameters that vary from decade to decade, and (2) stochastic parameters for the Gamma rainfall distribution where the parameters are randomly varied from year to year using a probability distribution function that is derived for each month of the year. While the DHMC has great potential to be used in hydrological and agricultural impact studies (e.g. urban drought security assessment), there are two important limitations of the DHMC:

- The DHMC tends to underestimate the mean of multiyear rainfall depths, which is probably linked to the use of stochastic Gamma parameters. Therefore, the stochastic sampling of Gamma parameters might be improved to overcome this limitation.
- The performance of the DHMC suggests that the use of decade-varied MC parameters are effective to incorporate the long-term variability of wet periods (although the use of decade-varied Gamma parameters in DPMC were not effective to incorporate the long-term variability of rainfall depths). However, other climate-based sub-samples (e.g. according to the ENSO phases) instead of decadal samples can be used for parameter calibration. This study tested the sub-samples according to the phases of IPO and ENSO climate modes with a focus on incorporating the long-term variability of rainfall depths, but has not incorporated climate-based sub sampling into DHMC because DHMC had not been developed at the time this analysis was performed. A more comprehensive assessment of such ideas might improve the wet period simulation of the DHMC.

In a subsequent paper, the performances of the CDMC, HMC and DHMC will be compared against the semi-parametric model of Mehrotra and Sharma (2007) using raingauge data from 30 stations around Sydney (those used in Mehrotra et al., 2015) and the 12 stations (Figure 1) around Australia.

10. Data Availability

- Daily rainfall data used in this study can be obtained from the Bureau of Meteorology, Australia website link http://www.bom.gov.au/climate/data/index.shtml by using weather station number 66062 and 023034 for Observatory Hill and Adelaide Airport stations respectively.
- ONI and IPO indices used in this study can be obtained from the National Oceanic and Atmospheric Administration website link https://www.esrl.noaa.gov/psd/data/climateindices/list/ and Folland (2008) respectively.

11. Code Availability

Python codes for modelling and statistical analysis of this study are available from the corresponding author.

12. Author Contributions

AFM Kamal Chowdhury has conducted the model development and statistical analysis of this study. Natalie Lockart and Garry Willgoose were the primary supervisors of this work and provided scientific oversight for the model development and statistical analysis. George Kuczera and Anthony Kiem provided more focussed advice on statistics and climatology. Nadeeka Parana Manage was involved scientific discussions as a team member of our project team.

13. Acknowledgments

Funding for this project was provided by an Australian Research Council Linkage Grant LP120200494, the NSW Office of Environment and Heritage, NSW Department of Financial Services, NSW Office of Water, and Hunter Water Corporation.

14. References

Bardossy, A. and Plate, E. J.: Space-Time Model for Daily Rainfall Using Atmospheric Circulation Patterns, Water Resources Research, 28(5), 1247-1259, doi:10.1029/91wr02589, 1992.

Bellone, E., Hughes, J. P. and Guttorp, P.: A Hidden Markov Model for Downscaling Synoptic Atmospheric Patterns to Precipitation Amounts, Climate Research, 15(1), 1-12, doi:10.3354/cr015001, 2000.

BoM: Daily Rainfall Data, http://www.bom.gov.au/climate/data/index.shtml, Bureau of Meteorology (BoM), Australia, visited on 11/2013, 2013.

Chen, J. and Brissette, F. P.: Comparison of Five Stochastic Weather Generators in Simulating Daily Precipitation and Temperature for the Loess Plateau of China, International Journal of Climatology, 34(10), 3089-3105, doi:10.1002/joc.3896, 2014.

Chen, J., Brissette, F. P. and Leconte, R.: A Daily Stochastic Weather Generator for Preserving Low-Frequency of Climate Variability, Journal of Hydrology, 388(3-4), 480-490, doi:10.1016/j.jhydrol.2010.05.032, 2010.

Chowdhury, A.F.M.K., Lockart, N., Willgoose, G., Kuczera, G., Kiem, A.S. and Parana Manage, N.: Modelling daily rainfall along the east coast of Australia using a compound distribution Markov Chain model, 36th Hydrology and Water Resources Symposium, December 2015, Hobart, Australia, 2015.

Dubrovský, M., Buchtele, J. and Žalud, Z.: High-Frequency and Low-Frequency Variability in Stochastic Daily Weather Generator and Its Effect on Agricultural and Hydrologic Modelling, Climatic Change, 63(1), 145-179, doi:10.1023/b:clim.0000018504.99914.60, 2004.

Evans, J. P., Ji, F., Lee, C., Smith, P., Argüeso, D., and Fita, L. 2014. Design of a regional climate modelling projection ensemble experiment – NARCliM, *Geosci. Model Dev., 7,* 621-629, doi:10.5194/gmd-7-621-2014, 2014.

Folland, C.: Interdecadal Pacific Oscillation Time Series, Met Office Hadley Centre for Climate Change, Exeter, UK, 2008.

Frost, A. J., Srikanthan, R., Thyer, M. A. and Kuczera, G.: A General Bayesian Framework for Calibrating and Evaluating Stochastic Models of Annual Multi-Site Hydrological Data, Journal of Hydrology, 340(3-4), 129-148, doi:10.1016/j.jhydrol.2007.03.023, 2007.

Furrer, E. M. and Katz, R. W.: Generalized Linear Modeling Approach to Stochastic Weather Generators, Climate Research, 34(2), 129, 2007.

Furrer, E. M. and Katz, R. W.: Improving the Simulation of Extreme Precipitation Events by Stochastic Weather Generators, Water Resources Research, 44(12), W12439, doi:10.1029/2008wr007316, 2008.

Glenis, V., Pinamonti, V., Hall, J. W. and Kilsby, C. G.: A Transient Stochastic Weather Generator Incorporating Climate Model Uncertainty, Advances in Water Resources, 85, 14-26, doi:10.1016/j.advwatres.2015.08.002, 2015.

Hansen, J. W. and Mavromatis, T.: Correcting Low-Frequency Variability Bias in Stochastic Weather Generators, Agricultural and Forest Meteorology, 109(4), 297-310, doi:10.1016/S0168-1923(01)00271-4, 2001.

Harrold, T. I., Sharma, A. and Sheather, S. J.: A Nonparametric Model for Stochastic Generation of Daily Rainfall Amounts, Water Resources Research, 39(12), 1343, doi:10.1029/2003wr002570, 2003.

Hundecha, Y., Pahlow, M. and Schumann, A.: Modeling of Daily Precipitation at Multiple Locations Using a Mixture of Distributions to Characterize the Extremes, Water Resources Research, 45(12), W12412, doi:10.1029/2008wr007453, 2009.

Jothityangkoon, C., Sivapalan, M. and Viney, N. R.: Tests of a Space-Time Model of Daily Rainfall in Southwestern Australia Based on Nonhomogeneous Random Cascades, Water Resources Research, 36(1), 267-284, doi:10.1029/1999wr900253, 2000.

Katz, R. W. and Parlange, M. B.: Effects of an Index of Atmospheric Circulation on Stochastic Properties of Precipitation, Water Resources Research, 29(7), 2335-2344, doi:10.1029/93WR00569, 1993.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

Katz, R. W. and Zheng, X.: Mixture Model for Overdispersion of Precipitation, Journal of Climate, 12(8), 2528-2537, doi:10.1175/1520-0442(1999)012<2528:MMFOOP>2.0.CO;2, 1999.

Lall, U., Rajagopalan, B. and Tarboton, D. G.: A Nonparametric Wet/Dry Spell Model for Resampling Daily Precipitation, Water Resources Research, 32(9), 2803-2823, doi:10.1029/96wr00565, 1996.

Lennartsson, J., Baxevani, A. and Chen, D.: Modelling Precipitation in Sweden Using Multiple Step Markov Chains and a Composite Model, Journal of Hydrology, 363(1–4), 42-59, doi:10.1016/j.jhydrol.2008.10.003, 2008.

Li, C., Singh, V. P. and Mishra, A. K.: A Bivariate Mixed Distribution with a Heavy-Tailed Component and Its Application to Single-Site Daily Rainfall Simulation, Water Resources Research, 49(2), 767-789, doi:10.1002/wrcr.20063, 2013.

Liu, Y., Zhang, W., Shao, Y. and Zhang, K.: A Comparison of Four Precipitation Distribution Models Used in Daily Stochastic Models, Advances in Atmospheric Sciences, 28(4), 809-820, doi:10.1007/s00376-010-9180-6, 2011.

Lockart, N., G. R. Willgoose, G. Kuczera, A. S. Kiem, A. F. M. K. Chowdhury, N. P. Manage, L. Zhang, and C. Twomey (2016), Case study on the use of dynamically downscaled GCM data for assessing water security on coastal NSW, *Journal of Southern Hemisphere Earth Systems Science*, *66*(2), 177-202.

Mehrotra, R., Li, J., Westra, S. and Sharma, A.: A Programming Tool to Generate Multi-Site Daily Rainfall Using a Two-Stage Semi Parametric Model, Environmental Modelling & Software, 63(0), 230-239, doi:10.1016/j.envsoft.2014.10.016, 2015.

Mehrotra, R. and Sharma, A.: A Semi-Parametric Model for Stochastic Generation of Multi-Site Daily Rainfall Exhibiting Low-Frequency Variability, Journal of Hydrology, 335(1–2), 180-193, doi:10.1016/j.jhydrol.2006.11.011, 2007.

Mortazavi, M., Kuczera, G., Kiem, A. S., Henley, B., Berghout, B. and Turner, E.: Robust Optimisation of Urban Drought Security for an Uncertain Climate, 74 pp, National Climate Change Adaptation Research Facility, Gold Coast, 2013.

Naveau, P., Huser, R., Ribereau, P. and Hannart, A.: Modeling Jointly Low, Moderate, and Heavy Rainfall Intensities without a Threshold Selection, Water Resources Research, 52(4), 2753-2769, doi:10.1002/2015WR018552, 2016.

NOAA: Climate Indices: Monthly Atmospheric and Ocean Time Series: https://www.esrl.noaa.gov/psd/data/climateindices/list/, National Oceanic and Atmospheric Administration (NOAA), visited in 07/2014, 2014.

Rajagopalan, B. and Lall, U.: A K-Nearest-Neighbor Simulator for Daily Precipitation and Other Weather Variables, Water Resources Research, 35(10), 3089-3101, doi:10.1029/1999wr900028, 1999.

Ramesh, N. I. and Onof, C.: A Class of Hidden Markov Models for Regional Average Rainfall, Hydrological Sciences Journal, 59(9), 1704-1717, doi:10.1080/02626667.2014.881484, 2014.

Richardson, C. W.: Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation, Water Resources Research, 17(1), 182-190, doi:10.1029/WR017i001p00182, 1981.

Sharda, V. N. and Das, P. K.: Modelling Weekly Rainfall Data for Crop Planning in a Sub-Humid Climate of India, Agricultural Water Management, 76(2), 120-138, doi:10.1016/j.agwat.2005.01.010, 2005.

Sharif, M. and Burn, D. H.: Simulating Climate Change Scenarios Using an Improved K-Nearest Neighbor Model, Journal of Hydrology, 325(1–4), 179-196, doi:10.1016/j.jhydrol.2005.10.015, 2006.

Srikanthan, R. and Pegram, G. G. S.: A Nested Multisite Daily Rainfall Stochastic Generation Model, Journal of Hydrology, 371(1–4), 142-153, doi:10.1016/j.jhydrol.2009.03.025, 2009.

Thyer, M. and Kuczera, G.: Modeling Long-Term Persistence in Hydroclimatic Time Series Using a Hidden State Markov Model, Water Resources Research, 36(11), 3301-3310, doi:10.1029/2000wr900157, 2000.

Verdon-Kidd, D. C., Wyatt, A. M., Kiem, A. S. and Franks, S. W.: Multidecadal Variability of Rainfall and Streamflow: Eastern Australia, Water Resources Research, 40(10), W10201, doi:10.1029/2004wr003234, 2004.

Vrac, M. and Naveau, P.: Stochastic Downscaling of Precipitation: From Dry Events to Heavy Rainfalls, Water Resources Research, 43(7), W07402, doi:10.1029/2006wr005308, 2007.

Wang, Q. J. and Nathan, R. J.: A Method for Coupling Daily and Monthly Time Scales in Stochastic Generation of Rainfall Series, Journal of Hydrology, 346(3–4), 122-130, doi:10.1016/j.jhydrol.2007.09.003, 2007.

Wilby, R. L., Wigley, T. M. L., Conway, D., Jones, P. D., Hewitson, B. C., Main, J. and Wilks, D. S.: Statistical Downscaling of General Circulation Model Output: A Comparison of Methods, Water Resources Research, 34(11), 2995-3008, doi:10.1029/98wr02577, 1998.

Wilks, D. S.: Interannual Variability and Extreme-Value Characteristics of Several Stochastic Daily Precipitation Models, Agricultural and Forest Meteorology, 93(3), 153-169, doi:10.1016/S0168-1923(98)00125-7, 1999a.

Wilks, D. S.: Simultaneous Stochastic Simulation of Daily Precipitation, Temperature and Solar Radiation at Multiple Sites in Complex Terrain, Agricultural and Forest Meteorology, 96(1–3), 85-101, doi:10.1016/S0168-1923(99)00037-4, 1999b.

Wilson, P. S. and Toumi, R.: A Fundamental Probability Distribution for Heavy Rainfall, Geophysical Research Letters, 32(14), L14812, doi:10.1029/2005gl022465, 2005.

Yunus, R. M., Hasan, M. M., Razak, N. A., Zubairi, Y. Z. and Dunn, P. K.: Modelling Daily Rainfall with Climatological Predictors: Poisson-Gamma Generalized Linear Modelling Approach, International Journal of Climatology, doi:10.1002/joc.4784, 2016.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

Hydrology and Earth System Sciences Discussions
Open Access
EGU

15. Table

Table 1: Methodological comparison of the five MC models.

	Wet and dry day simulation	Wet day rainfall depth simulation
APMC	Uses deterministic MC parameters, same set of parameters for each simulation year.	Uses deterministic Gamma parameters, same set of parameters for each simulation year.
DPMC	Uses decade-varied deterministic MC parameters.	Uses decade-varied deterministic Gamma parameters.
CDMC	Same as APMC.	Uses stochastic parameters (sampled from fitted bivariate-lognormal distribution) of Gamma distribution, parameters vary for each simulation year.
HMC	Uses stochastic MC parameters (sampled from fitted truncated normal distribution), parameters vary for each simulation year.	Same as CDMC.
DHMC	Same as DPMC.	Same as CDMC.

Hydrology and Earth System Sciences
Open Access
Discussions

16. Figures

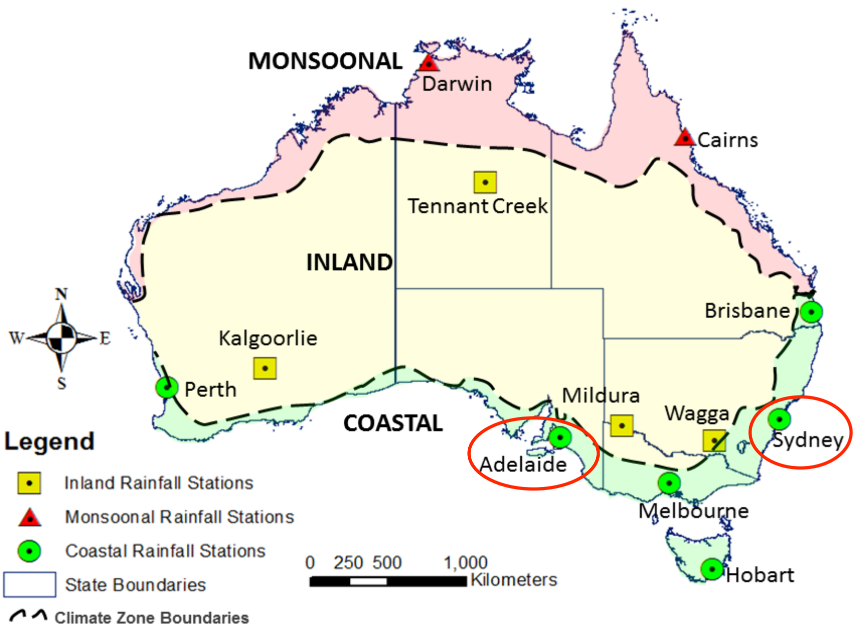


Figure 1: Location map of 12 raingauge stations around Australia. This study has presented the assessment results of the developed models for Sydney and Adelaide stations (red circled) only.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

Hydrology and Earth System Sciences
Open Access
Discussions

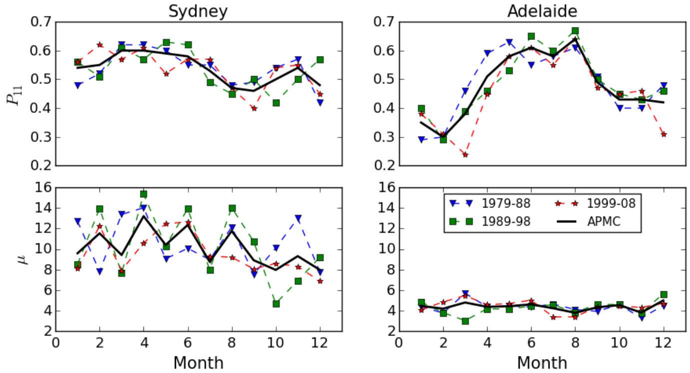


Figure 2: Comparison of the decadal variability of the DPMC parameters P_{11} and μ (mm) with the APMC parameters.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

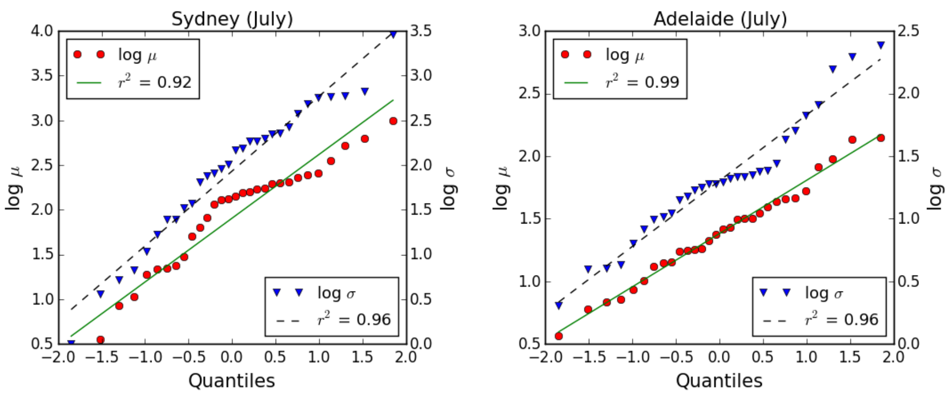


Figure 3: Lognormal probability plots of μ and σ for July (typical of other months).

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

Hydrology and Earth System Sciences
Open Access
Discussions

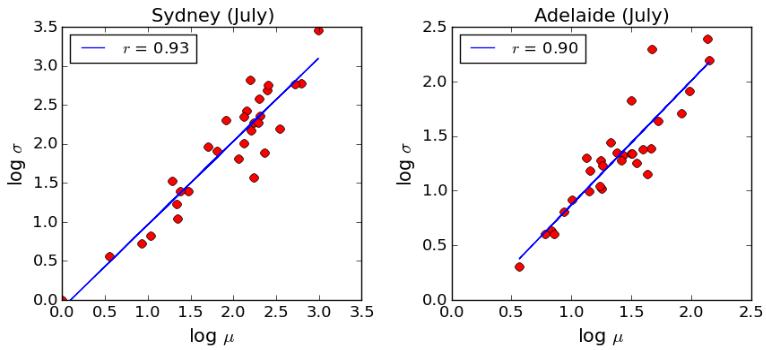
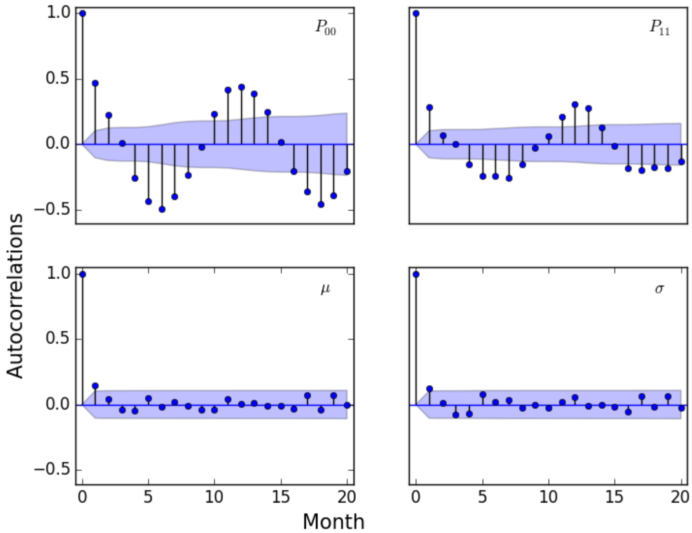


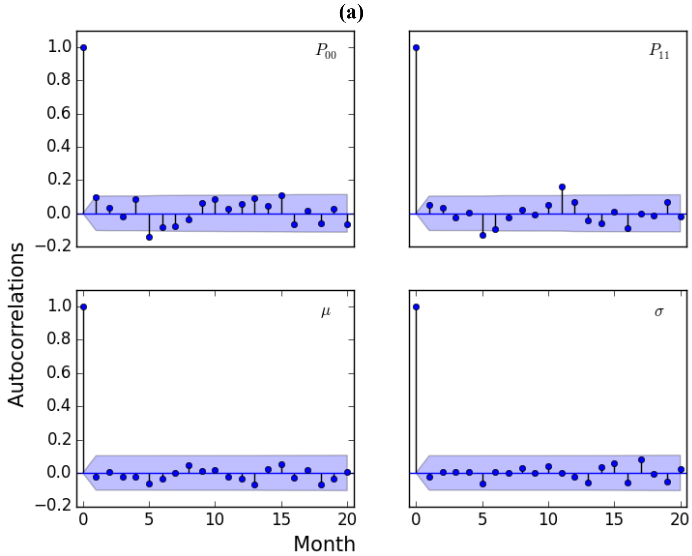
Figure 4: Correlation between *log μ* and *log σ* for July (typical of other months).

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

Hydrology and Earth System Sciences
Open Access
Discussions



(a)



(b)

Figure 5: Month-to-month autocorrelations of P_{00}, P_{11}, μ and σ for (a) Adelaide and (b) Sydney.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

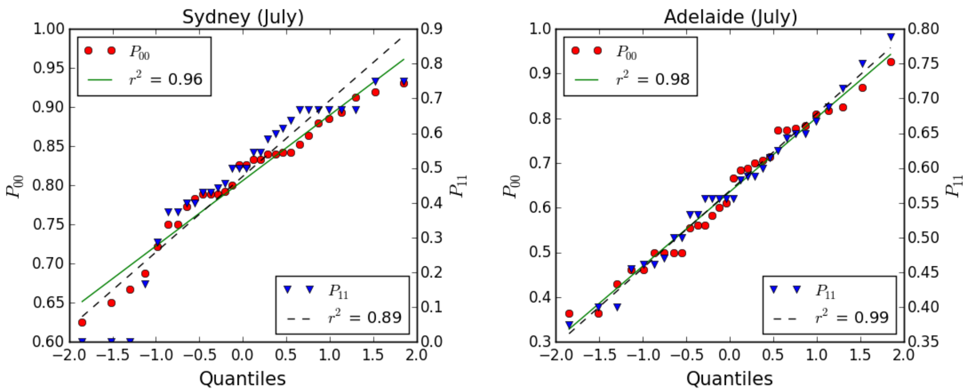


Figure 6: Normal probability plots of P_{00} and P_{11} for July (typical of other months).

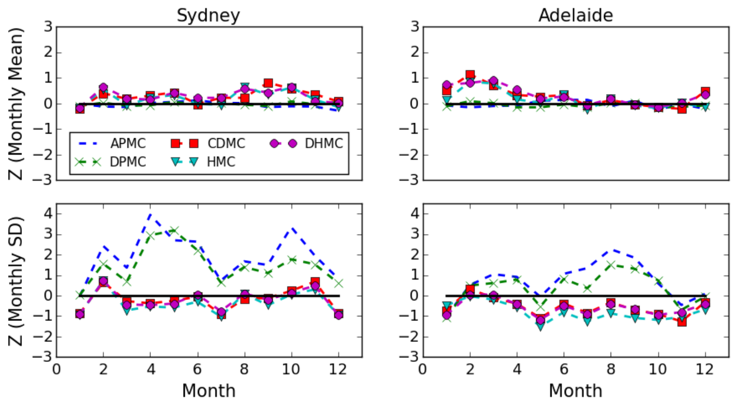


Figure 7: Comparison of the mean and SD of monthly rainfall depths for the five MC models.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

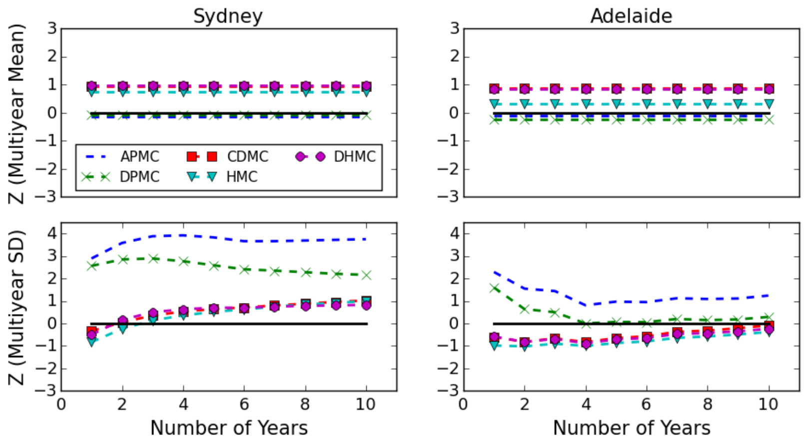
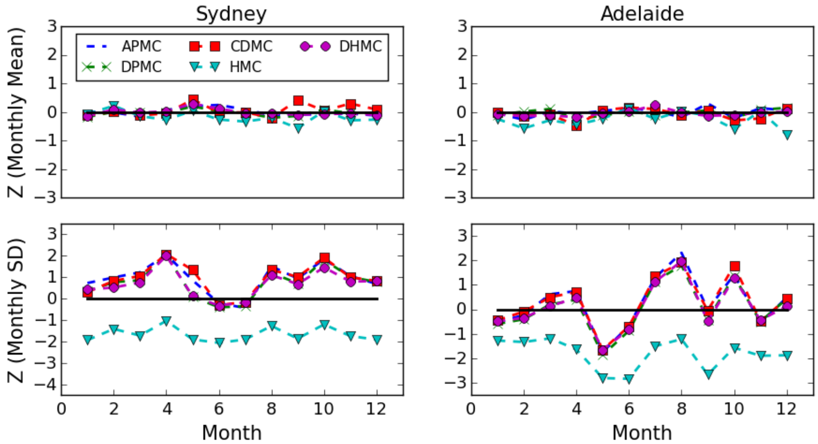


Figure 8: Comparison of the mean and SD of multiyear rainfall depths for the five MC models.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017



667

668 **Figure 9: Comparison of the mean and SD of monthly number of wet days for the five MC models.**

669

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

Hydrology and Earth System Sciences
Discussions
Open Access

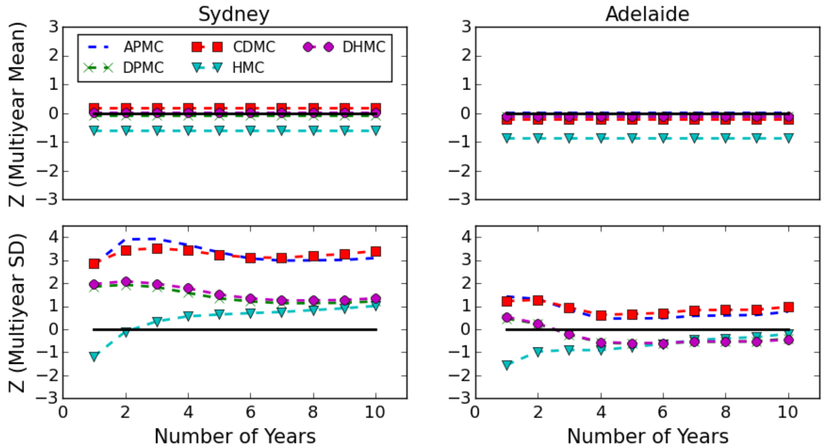


Figure 10: Comparison of the mean and SD of multiyear number of wet days for the five MC models.

Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-84, 2017
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 27 February 2017

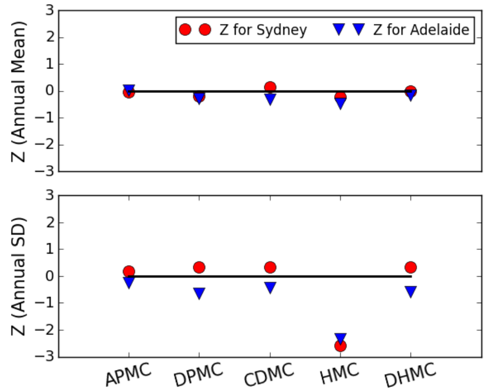
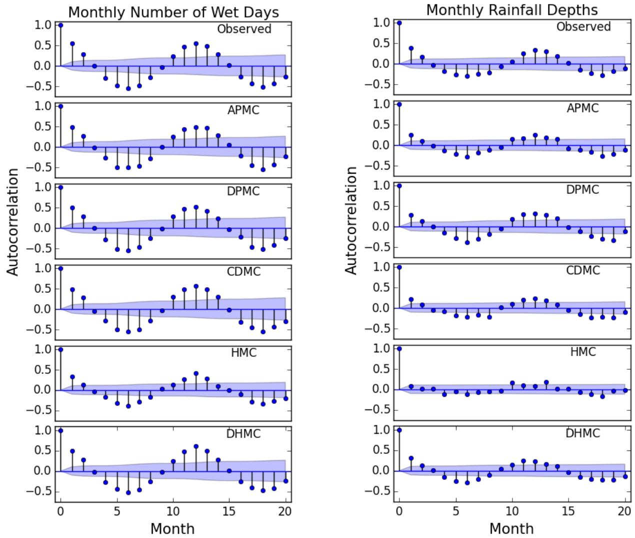
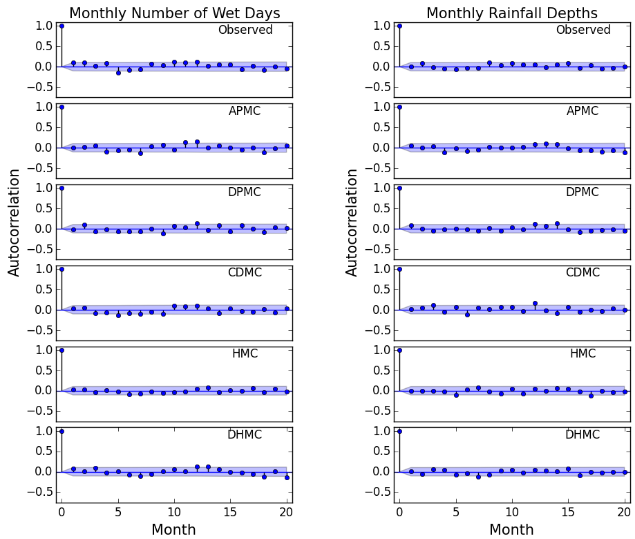


Figure 11: Comparison of the mean and SD of annual mean length of wet spells for the five MC models.



(a)



(b)

Figure 12: Comparison of the autocorrelations of monthly number of wet days and monthly rainfall depths for the five MC models for (a) Adelaide and (b) Sydney.