Probability Structure and Return Period of Multiday Monsoon Rainfall

Nur S. Muhammad[1]; Pierre Y. Julien, M.ASCE[2]; and Jose D. Salas, M.ASCE[3]

Abstract: The daily monsoon rainfall data recorded at Subang Airport, Malaysia, from 1960 to 2011 is examined in terms of probability structure for the estimation of extreme daily rainfall precipitation during the Northeast (NE) and Southwest (SW) Malaysian monsoons. The discrete autoregressive and moving average [DARMA(1,1)] model is preferable to the first-order Markov chain [DAR(1)] model. The conditional probabilities of t consecutive rainy days are time dependent. Nevertheless, a simple two-parameter gamma distribution appropriately fits the frequency distribution of multiday rainfall amounts. An algorithm is developed by combining the DARMA(1,1) and gamma models to estimate the return period of multiday rainfall. Extensive comparisons showed that the DARMA(1,1)-gamma model gives a reliable estimate of the return period of rainfall for both NE and SW monsoons at Subang Airport. Furthermore, values generated from the models enable the analysis of the frequency distribution of extreme rainfall events. **DOI: 10.1061/(ASCE)HE.1943-5584.0001253.**

Author keywords: Multiday rainfall; Monsoon rainfall precipitation; Return period; Conditional probability; Stochastic modeling.

Introduction

The planning and design of water resources projects require the analysis of reliable, long-term hydrological data such as rainfall and streamflow. The stochastic point process was first introduced by Todorovic (1968) and subsequently used by Todorovic and Yevjevich (1969) and Eagleson (1978) for modeling short-term rainfall. Kavvas and Delleur (1981) were successful in modeling the sequence of daily rainfall in Indiana using the Neyman-Scott (NS) cluster process. They assumed that a rainfall event occurs in midday, to comply with the model order. Multiday rainfall events were treated as a group instantaneous rainfall that occurs once a day, with a 1-day interval. Rodriguez-Iturbe et al. (1987) tested the performance of different types of point process models, i.e., the Poisson and cluster-based models using hourly rainfall data from Denver, Colorado. They concluded that the white-noise Poisson model was unable to produce satisfactory results. Instead, cluster-based models, namely the Neyman-Scott (NS) and Bartlett-Lewis (BL) processes, are more flexible, reliable, and able to represent the actual rainfall scenarios. Since then, other researchers have improved the NS and BL processes to model fine-scale rainfall. Examples of the use of a modified BL model can be found in Rodriguez-Iturbe et al. (1988), Glasbey et al. (1995), Khaliq and Cunnane (1996), Cowpertwait et al. (2007), and Verhoest et al. (2010). In addition, Cowpertwait (1995), Cowpertwait et al. (1996), and Burton et al. (2008, 2010) are studies that utilize the modified NS to model rainfall. However, Rodriguez-Iturbe et al. (1988), Cowpertwait et al. (1996, 2007), and Burton et al. (2008) show that the point-process models were unable to produce extreme values with good accuracy. Furthermore, Obeysekera et al. (1987) applied various types of point-process models for hourly rainfall considering the diurnal cycle that is characteristic at certain locations during some months of the year.

Low-order discrete autoregressive family models, such as the discrete autoregressive [DAR(1)] and discrete autoregressive and moving average [DARMA(1,1)] models, are frequently used for simulating daily rainfall sequences. The DAR(1) model is also equivalent to a first-order Markov chain model. This model assumes that the probability of rain depends only on the current state (wet or dry) and will not be influenced by its past behavior. Haan et al. (1976), Katz (1977), Roldán and Woolhiser (1982), Small and Morgan (1986), Jimoh and Webster (1996), Sharma (1996), Tan and Sia (1997), and Wilks (1998) are among the studies that were successful in modeling the sequence of rainy and dry days using first-order Markov chains. Wilks (1998) used the first-order Markov chain to simulate the occurrence of daily rainfall based on data from 1951 to 1996 from 25 stations in New York State, USA. The statistical properties such as the joint probabilities for both rainy and dry days, mean monthly rainfall, and standard deviations of monthly rainfall indicate that the simulated rainfall data reproduce the rainfall data statistics really well. It was concluded that the model was successful in preserving the dependence nature of daily rainfall at these stations. First-order Markov chains are simple and do not require a lot of computational effort. However, Feyerherm and Bark (1965) found that first-order Markov chains are unable to model the scenario of strong dry day persistence. Similar findings were reported by Wallis and Griffiths (1995) and Semenov et al. (1998). The order of a Markov chain may be influenced by seasonal change and location (Chin 1977; Cazacioc and Cipu 2005; Deni et al. 2009). Chin (1977) found that the seasonal change has a significant impact in determining the suitable order of a Markov chain in more than 200 stations located throughout the USA. High-order Markov chains are suitable to model the sequence of daily precipitation during winter at most stations, and first-order Markov chains are appropriate for summer.

[1]Lecturer, Dept. of Civil and Structural Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Selangor 43600, Malaysia (corresponding author). E-mail: shazwani.muhammad@ukm.edu.my

[2]Professor, Dept. of Civil and Environmental Engineering, Colorado State Univ., Fort Collins, CO 80523.

[3]Professor Emeritus, Dept. of Civil and Environmental Engineering, Colorado State Univ., Fort Collins, CO 80523.

Note. This manuscript was submitted on August 2, 2013; approved on April 21, 2015; published online on June 17, 2015. Discussion period open until November 17, 2015; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, © ASCE, ISSN 1084-0699/04015048(11)/$25.00.

The physical environmental causes and geography can influence the order of Markov chains. Similar findings were reported by Cazacioc and Cipu (2005) for the simulation of rainfall sequences at several stations in Romania.

For tropical regions, a different approach was used by Deni et al. (2009) in the analysis of Malaysian daily rainfall data based on the Markov chain model. The objective of their study was to find the optimum order of a Markov chain for daily rainfall during the Northeast (NE) and Southwest (SW) monsoons using two different thresholds, i.e., 0.1 and 10.0 mm, where NE and SW are the directions from which the monsoons are coming. The Akaike information criteria (AIC) and Bayesian information criteria were used to determine the appropriate order of the Markov chain models. The study used the available data from 18 rainfall stations located in various parts of Peninsular Malaysia. They concluded that the optimum order of a Markov chain varies with the location, monsoon season, and the level of threshold. For example, the occurrence of rainfall (threshold level 10.0 mm) for the NE and SW monsoons at stations located in the northwestern and eastern regions of Peninsular Malaysia can be represented using a first-order Markov chain. Additionally, higher order Markov chain models are suitable to represent rainfall occurrence, especially during the NE monsoon, for both levels of threshold. Other examples of the use of a high-order Markov chain to simulate the rain and dry day sequence are reported by Mimikou (1983), Dahale et al. (1994), Katz and Parlange (1998), and Dastidar et al. (2010).

Even though higher order Markov chain models may be used to overcome the lack of persistence of the simple Markov chain, more parameters have to be used, which increases the model uncertainty (Jacobs and Lewis 1983) and also makes the calculations more complex. Jacobs and Lewis (1978) and Kedem (1980) discuss the concept of the stationary DARMA model, which is intended to be simpler for modeling stationary sequences of dependent discrete random variables with specified marginal distribution and correlation structure. Buishand (1977, 1978) modeled the sequence of daily rainfall using DARMA(1,1) at several stations in the Netherlands, Suriname, India, and Indonesia. Since DARMA(1,1) is a stationary model, the data for each station were divided into their respective seasons in order to consider the rainfall seasonal variations. The results have shown that the DARMA(1,1) model is successful in simulating the daily rainfall in tropical and monsoon areas, where prolonged dry and wet seasons may occur. The DARMA(1,1) model provides longer persistence than the DAR(1) model does. Other studies that use the DARMA(1,1) model to simulate sequences of daily rainfall include Chang et al. (1982, 1984b, a), Delleur et al. (1989), and Cindrić (2006). In addition, DARMA models have been applied for the analysis of droughts (Chung and Salas 2000; Salas et al. 2005; Cancelliere and Salas 2010). For example, Chung and Salas (2000) analyzed the annual streamflow time series of the Niger River in Africa and concluded that the drought occurrence can be successfully simulated using the DARMA(1,1) model. The results showed long periods of low flows (drought) and high flows, and the DARMA (1,1) model was suitable for simulating streamflows with a longer memory as compared to the DAR(1) model.

Return periods are useful in hydrology to measure the severity of an event. Various definitions of return period have been reported in the literature, such as first arrival time and interarrival time or recurrence interval. These definitions give different values when the events are dependent in time. However, for single and independent events, the first arrival time and recurrence interval give the same value (Fernández and Salas 1999a). Extensive theories and applications on the return period definitions and serial dependence are discussed in Fernández and Salas (1999a, b). Woodyer et al. (1972), Kite (1978), Lloyd (1970), Loaiciga and Mariňo (1991), and Şen (1999) defined recurrence interval as the average elapsed time between the occurrences of critical events, such as earthquakes of high magnitude and extreme floods or droughts. In addition, Vogel (1987) and Douglas et al. (2002) used the return period as the average number of trials required to the first occurrence of a critical event. This definition may be more useful in relation to reservoir operation because knowing the first time that the reservoir is at risk of failure is of greater interest than the average time between failures (Douglas et al. 2002). Furthermore, Goel et al. (1998), Shiau and Shen (2001), Kim et al. (2003), González and Valdéz (2003), Salas et al. (2005), and Cancelliere and Salas (2004, 2010) reported studies on the calculation of return period and risk that include both the amount and duration of extreme hydrological events.

This study concentrates on the occurrence of multiday rainfall events in Malaysia. The country experiences two major seasons classified as the Northeast (NE) and Southwest (SW) monsoons. The NE monsoon typically occurs from November to March, while the SW monsoon is from May to September. April and October are known as intermonsoon months. Both monsoons bring lots of moisture and as a result, Malaysia receives between 2,000 to 4,000 mm of rainfall with 150 to 200 rainy days annually (Suhaila and Jemain 2007). One of the most devastating recent multiday rainfall events resulted in the Kota Tinggi flood in December 2006 and January 2007. These two extreme monsoon events resulted in more than 350 and 450 mm of cumulative rainfall in less than a week. The estimated economic loss reached half a billion US dollars and more than 100,000 local residents had to be evacuated (Abu Bakar et al. 2007). Even though it is well known that multiday events are the main cause of flooding in Malaysia, the topic has received little attention from local researchers.

This paper discusses various aspects of Malaysian monsoons, including the probability distribution and probability structure of multiday monsoon rainfall events, the modeling and simulation of daily rainfall sequences, the estimation of extreme rainfall quantiles, and the estimation of the return period of multiday rainfall. The occurrences of daily rainfall are characterized and simulated using the discrete autoregressive and moving average [DARMA (1,1)] model. These approaches were tested using the observed daily rainfall measurements collected from Subang Airport near Kuala Lumpur, Malaysia.

Summary of DAR(1) and DARMA(1,1) Models

This study uses the DAR(1) and DARMA(1,1) models to simulate the occurrence of daily rainfall. The DAR(1) model is represented as (Jacobs and Lewis 1978)

$$A_t = V_t A_{t-1} + (1 - V_t) Y_t$$

$$\text{with} \quad A_t = \begin{cases} A_{t-1} \text{ with probability } \lambda \\ Y_t \text{ with probability } (1-\lambda) \end{cases} \tag{1}$$

where V_t is an independent random variable taking values of 0 and 1 such that

$$P(V_t = 1) = \lambda = 1 - P(V_t = 0) \tag{2}$$

and λ is a parameter. The variable Y_t is another independent and identically distributed (i.i.d.) random variable, with a common probability $\pi_k = P(Y_t = k), k = 0,1$.

It should be noted that A_t is a first-order Markov chain and the process of simulation is assumed to start at A_{-1}

(Buishand 1978). The autocorrelation function of the DAR(1) model is (Jacobs and Lewis 1978)

$$\mathrm{corr}(A_t, A_{t-k}) = r_k(A) = \lambda^k, k \geq 1 \tag{3}$$

where r_k is the lag-k (days) autocorrelation function.

The autocorrelation function (r_k) is estimated based on the sequences of dry and rainy days, i.e., 0 s and 1 s, and not the rainfall amounts (Delleur et al. 1989) as

$$r_k = \left[\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})\right]\left[\sum_{t=1}^{N}(x_t - \bar{x})^2\right]^{-1} \tag{4}$$

$$\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t \tag{5}$$

where r_k is the lag-k autocorrelation coefficient and N is the sample size.

There are two parameters associated with the DAR(1) model, i.e., π_0 (or π_1) and λ. The parameter λ may be estimated from the lag-1 autocorrelation coefficient as given in Eqs. (3) and (4). The parameters π_0 and π_1 are based on the dry and wet run lengths that are obtained from the observed daily rainfall data set. They are estimated using Eqs. (6) and (7) (Buishand 1978) as follows:

$$\pi_0 = \frac{\overline{T_0}}{\overline{T_0} + \overline{T_1}} \tag{6}$$

$$\pi_1 = 1 - \pi_0 \tag{7}$$

where $\overline{T_0}$ = mean run length for dry days and $\overline{T_1}$ = mean run length for wet days.

The one-step transitional probability, $p(i,j) = P(A_{t+1} = j|A_t = i)$ is given by (Jacobs and Lewis 1978) as follows:

$$p(i,j) = \begin{cases} \lambda + (1-\lambda)\pi_j, & \text{if } i = j \\ (1-\lambda)\pi_j, & \text{if } i \neq j \end{cases} \qquad i,j = 0,1 \tag{8}$$

Eq. (8) can also be represented in terms of the transitional probability matrix, as shown in Eq. (9)

$$P = \begin{bmatrix} \lambda + (1-\lambda)\pi_0 & (1-\lambda)\pi_1 \\ (1-\lambda)\pi_0 & \lambda + (1-\lambda)\pi_1 \end{bmatrix} \tag{9}$$

The transitional probability matrix simplifies the calculation of run length. The concept of run length is important, especially in modeling the sequence of daily rainfall. The run length is defined as the succession of events of the same kind, and it is bounded at the beginning and the end by events of a different kind. For the DAR(1) model, the probability distribution of wet and dry run lengths can be obtained from Eqs. (10) and (11) as derived by Chang et al. (1984b)

$$P(T_1 = n) = p^{n-1}(1,1)[1 - p(1,1)] \tag{10}$$

$$P(T_0 = n) = p^{n-1}(0,0)[1 - p(0,0)] \tag{11}$$

The DARMA(1,1) model is represented as (Jacobs and Lewis 1978)

$$X_t = U_t Y_t + (1 - U_t)A_{t-1}$$
$$\text{with} \quad X_t = \begin{cases} Y_t \text{ with probability } \beta \\ A_{t-1} \text{ with probability} (1-\beta) \end{cases} \tag{12}$$

where U_t is an independent random variable taking values of 0 or 1 only such that

$$P(U_t = 1) = \beta = 1 - P(U_t = 0) \tag{13}$$

Y_t is another i.i.d. random variable having a common probability $\pi_k = P(Y_t = k)$, $k = 0{,}1$, and A_t is an autoregressive component given by

$$A_t = \begin{cases} A_{t-1} \text{ with probability } \lambda \\ Y_t \text{ with probability } (1-\lambda) \end{cases}$$

The variable A_t has the same probability distribution as Y_t but is independent of Y_t. It should be noted that X_t is not Markovian, but (X_t, A_t) forms a first-order bivariate Markov chain.

The autocorrelation function of the DARMA(1,1) model is (Buishand 1978)

$$\mathrm{corr}(X_t, X_{t-k}) = r_k(X) = c\lambda^{k-1}, \qquad k \geq 1 \tag{14}$$

$$\text{where } c = (1-\beta)(\beta + \lambda - 2\lambda\beta) \tag{15}$$

The three parameters of the DARMA(1,1) model need to be estimated, namely π_0 or π_1, λ, and β. The parameters π_0 or π_1 may be estimated from Eqs. (6) to (7). The estimation of λ may be determined by minimizing Eq. (16) using the Newton-Raphson iteration techniques. Buishand (1978) suggested using the ratio of the second to the first autocorrelation coefficients as an initial estimate for λ, as shown in Eq. (17)

$$\phi(\lambda) = \sum_{k=1}^{M}[r_k - c\lambda^{k-1}]^2; \qquad k \geq 1 \tag{16}$$

$$\hat{\lambda} = \frac{r_2}{r_1} \tag{17}$$

in which M is the total number of lags considered, c can determined from the lag-1 autocorrelation coefficient of the DARMA(1,1) model, and β can be estimated from Eq. (18)

$$\hat{\beta} = \frac{(3\hat{\lambda} - 1) \pm \sqrt{(3\hat{\lambda} - 1)^2 - 4(2\hat{\lambda} - 1)(\hat{\lambda} - \hat{c})}}{2(2\hat{\lambda} - 1)} \tag{18}$$

The probability distributions of the wet and dry run lengths for the DARMA(1,1) model are well known in the literature (e.g., Jacobs and Lewis 1978), see the Appendix for more details.

Probability Distribution and Return Period of Multiday Rainfall Events

In this section, the probability distribution and return period for multiday rainfall events are investigated considering that the occurrence of daily rainfall is correlated. The return period of multiday rainfall events is based on the number of trials between two successive occurrences of the same event. Multiday rainfall events occur frequently during the NE and SW monsoons. Therefore, it is appropriate to estimate the return period as the average time (in days) between the occurrences of specific events. It may also be referred to as the recurrence interval. The most important parameters that hydrologists and water resources specialists are concerned about when analyzing a multiday rainfall event are the duration and the amount of cumulative rainfall. Hence, this study considers both parameters in formulating the estimation of the return period.

Muhammad (2013) found that the two-parameter gamma function is most suitable for representing the rainfall amount distribution for specified durations at Subang Airport. The method of moments is used to estimate the parameters and the formulas can be found in Mood et al. (1974) or Yevjevich (1984). It should be noted that (e.g., Mood et al. 1974) if two independent gamma variables are added, for example, $X = R_1 + R_2$ where both R_1 and R_2 are gamma(α, β), then X is also gamma(α, 2β). Likewise, if you add t independent gamma(α, β) variables then $X = R_1 + R_2 + \ldots + R_t$ is gamma(α, $t\beta$). The data analysis showed that the two-parameter gamma distribution was best for describing the distribution of 1-day and multiday rainfall events at Subang Airport. The empirical representation of rainfall amount distribution in t consecutive rainy days is given in Eq. (19)

$$f(x) \cong \frac{1}{|24.0|\Gamma(0.6t)}\left(\frac{x}{24.0}\right)^{0.6\,t-1}\exp\left(-\frac{x}{24.0}\right) \qquad (19)$$

where x = total amount of rainfall for t consecutive rainy days (mm) and t = number of consecutive rainy days.

Thus, for a given rainfall duration, t, Eq. (19) enables one to determine the probability of any rainfall event exceeding say x_0. Denoting such a probability as $P(E|t)$, it may be determined as follows:

$$P(E|t) = \int_{x_0}^{\infty} f(x)dx \qquad (20)$$

To determine the return period of rainfall events, E, of a given duration, t, an approach used previously for determining the return period of droughts (e.g., Cancelliere and Salas 2002; Gonzáles and Valdés 2003; Salas et al. 2005) were followed. It follows that the return period of multiday rainfall events can be determined as

$$T = \frac{\overline{T_1} + \overline{T_0}}{P(E|t)} \qquad (21)$$

where $\overline{T_1}$ = mean run length for wet days; $\overline{T_0}$ = mean run length for dry days; and $P(E|t)$ = probability of a rainfall event given by Eq. (20).

Results and Discussion

Probability Distribution of Daily Rainfall

The daily rainfall measurements at Subang Airport (3°7′1.20″N, 101°33′0.00″E) were used in this study. A long and reliable record of 52 years for the period 1960 to 2011 was provided by the Department of Meteorology, Malaysia.

Fig. 1 shows the cumulative distribution function (CDF) for 1-day and multiday rainfall at Subang Airport. The figure shows that the two-parameter gamma distribution function given by Eq. (19) fits reasonably well the historical CDF for 1 through 6 days of rainfall duration. The CDF plot shows that for a single rainy day, there is about 60% chance that the rainfall amount will be less than 10 mm, and there is a less than 5% chance that the rainfall amount will exceed 50 mm. The multiday rainfall events resulted in a significant amount of rainfall to the study area. The CDF plot also shows that there is a nonnegligible probability that 2 and 3 consecutive rainy days may produce more than 100 mm of rain. Further, there is 50% of chance of 4, 5, and 6 consecutive rainy days yielding more than 55, 65, and 85 mm of rainfall, respectively. The probability of rainfall events with more than 100 mm of rain increases as the number of consecutive rainy days increases. These results illustrate the important need for a detailed analysis of both duration and magnitude of multiday rainfall events.

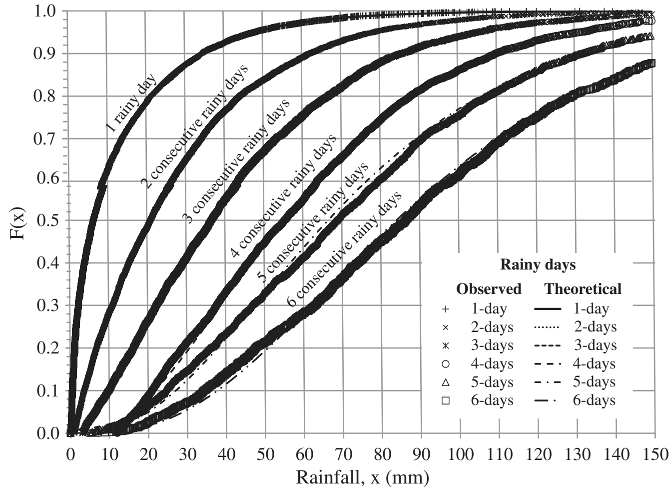

Fig. 1. Cumulative distribution function of 1-day and multiday rainfall events at Subang Airport

Probability Structure of Rainfall Occurrence

When rain occurs on a given day, it is called a wet day while the absence of rain on a given day is called a dry day. In this study, a wet day is indicated for rainfall amounts of more than 0.1 mm, while a dry day is assumed for amounts less than or equal to 0.1 mm. The threshold amount was determined based on the Von Neumann (1941) ratio and a more detailed analysis is presented in Muhammad (2013).

The analysis of the probability structure of rainfall events at Subang Airport shows that more than 50% of the events observed at the study site are rainy days. The estimated probability of rain on any given day is 0.53. If day-to-day rainfall events were independent, the probability of rain on any day would remain constant at 0.53 [shown by a triangle in Fig. 2(a)]. However, Fig. 2(a) shows that the field rainfall measurements of the conditional probability increases significantly as the number of consecutive rainy day increases, i.e., from 0.53 for a single rainy day to about 0.80 for 15 consecutive rainy days. For example, the estimated conditional probability of a fourth rainy day, given that it follows 3 consecutive rainy days, is 0.68. This probability is far greater than the estimated probability of the first day of rain, i.e., 0.53. The occurrence of rain on a given day affects the probability of rain the following days. Thus, the conditional probabilities estimated from the historical data show that the events are dependent.

Likewise, Fig. 2(b) gives the estimated conditional probabilities of n consecutive dry days at Subang Airport. The estimated probability that any given day is dry is 0.47, which increases significantly to 0.72 after 15 consecutive dry days. For example, the estimated conditional probability for a second consecutive dry day is 0.58, and the estimated probability for the third dry day increases to 0.63. Thus, the probability structure of n consecutive dry days is also dependent as is the case for rainy days. Table 1 gives the details of the frequency and the estimated conditional probabilities of 1 to 15 consecutive wet and dry days.

Modeling the Occurrence of Daily Rainfall

In this study, the NE and SW monsoons are considered as the daily rainfall recorded during the months of October to March and April

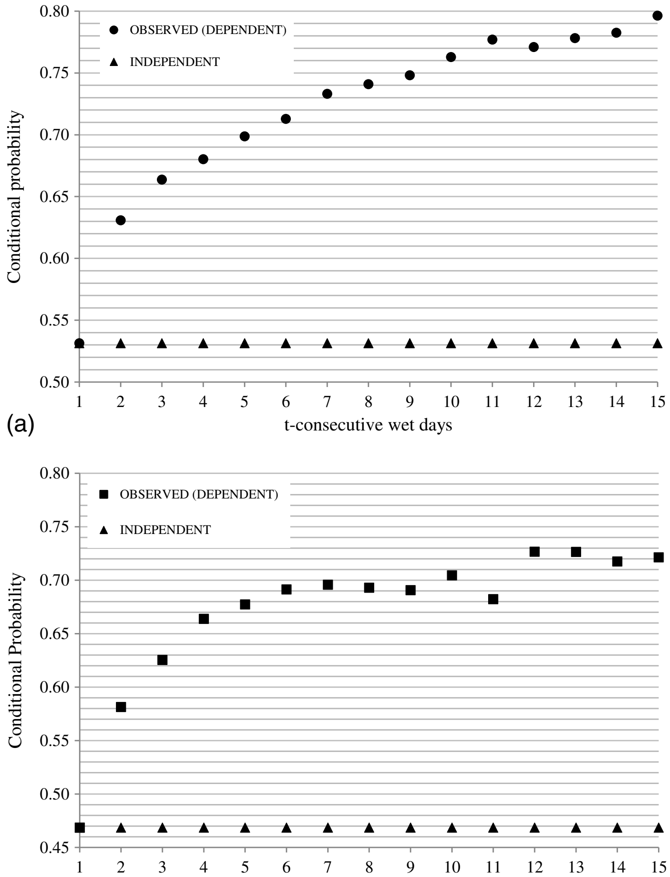

Fig. 2. Plot of conditional probability of t consecutive (a) wet days and (b) dry days

to September, respectively. Separate analyses are conducted for NE and SW monsoons. The sample mean wet ($\overline{T_1}$) and dry ($\overline{T_0}$) run lengths are estimated from the observed daily rainfall data set. The results from the field measurements give $\overline{T_1} = 3.00$ days (2.43 days) and $\overline{T_0} = 2.19$ days (2.58 days) for the NE (SW) monsoon. The probabilities of a day being dry or wet were estimated using Eq. (6) or Eq. (7), respectively. The analysis gives $\widehat{\pi_1} = 0.5781$ (0.4851) and $\widehat{\pi_0} = 0.4219(0.5149)$, respectively, for the NE (SW) monsoons.

The λ parameter of the DAR(1) model is calculated based on Eqs. (3) and (4). The estimated values for $\hat{\lambda}$ are 0.196 (0.192) for NE (SW) monsoon. It can be observed that there are no significant differences in the estimated values of $\hat{\lambda}$ for the two monsoon seasons. For the DARMA(1,1) model, the parameter λ is estimated based on the Newton-Raphson iteration techniques using Eqs. (16) and (17). Then, Eq. (18) was applied to estimate the parameter β. For the NE (SW) monsoon, the estimated model parameters for DARMA(1,1) are $\hat{\lambda} = 0.7339(0.7827)$ and $\hat{\beta} = 0.5775(0.5789)$.

The autocorrelation functions (ACFs) for the DAR(1) and DARMA(1,1) models are determined using Eqs. (3) and (14), respectively. Figs. 3(a and b) show the comparisons between the observed and theoretical ACFs of the DAR(1) and DARMA(1,1) models for the NE and SW monsoon, respectively. The ACFs estimated using the DAR(1) model decays to zero after day 2 and departs from the sample autocorrelation. On the other hand, the fitted DARMA(1,1) models' ACFs for NE (SW) monsoon decay slowly and eventually reach zero at day 15. Thus, close agreement between the observed and theoretical ACFs for the estimated DARMA(1,1) model are shown for both monsoon seasons. This suggests that the DARMA(1,1) model may be suitable for representing the occurrence of daily rainfall for any season at Subang Airport.

In addition, Chang et al. (1984b, a) suggested minimizing the sum of squared errors between the observed and theoretical probability distributions of wet and dry run lengths for further assessing and comparing alternative models. The transitional probabilities are used to calculate the probability distribution function of wet and dry run lengths. The transitional probabilities for the DAR(1) model during NE and SW monsoons are estimated using Eq. (9) as $\begin{bmatrix} 0.5352 & 0.4648 \\ 0.3392 & 0.6608 \end{bmatrix}$ and $\begin{bmatrix} 0.6079 & 0.3921 \\ 0.4161 & 0.5839 \end{bmatrix}$, respectively.

The probability distributions of wet and dry run lengths for the DARMA(1,1) model were determined based on the transitional probability matrices, H_0 and H_1. The details are given in the Appendix. The transitional probability matrices, H_0 and H_1, for the NE and SW monsoons are $H_0 = \begin{bmatrix} 0.6012 & 0.0650 \\ 0.0648 & 0.1788 \end{bmatrix}$;

Table 1. Frequencies and Estimated Conditional Probabilities of t Consecutive Wet and Dry Days

Wet			Dry		
t consecutive wet days	Frequency	Estimated conditional probability	t consecutive dry days	Frequency	Estimated conditional probability
1	10,092	0.53	1	8,901	0.47
2	6,366	0.63	2	5,174	0.58
3	4,226	0.66	3	3,236	0.63
4	2,875	0.68	4	2,148	0.66
5	2,009	0.70	5	1,455	0.68
6	1,432	0.71	6	1,006	0.69
7	1,050	0.73	7	700	0.70
8	778	0.74	8	485	0.69
9	582	0.75	9	335	0.69
10	444	0.76	10	236	0.71
11	345	0.78	11	161	0.68
12	266	0.77	12	117	0.73
13	207	0.78	13	85	0.73
14	162	0.78	14	61	0.72
15	129	0.80	15	44	0.72

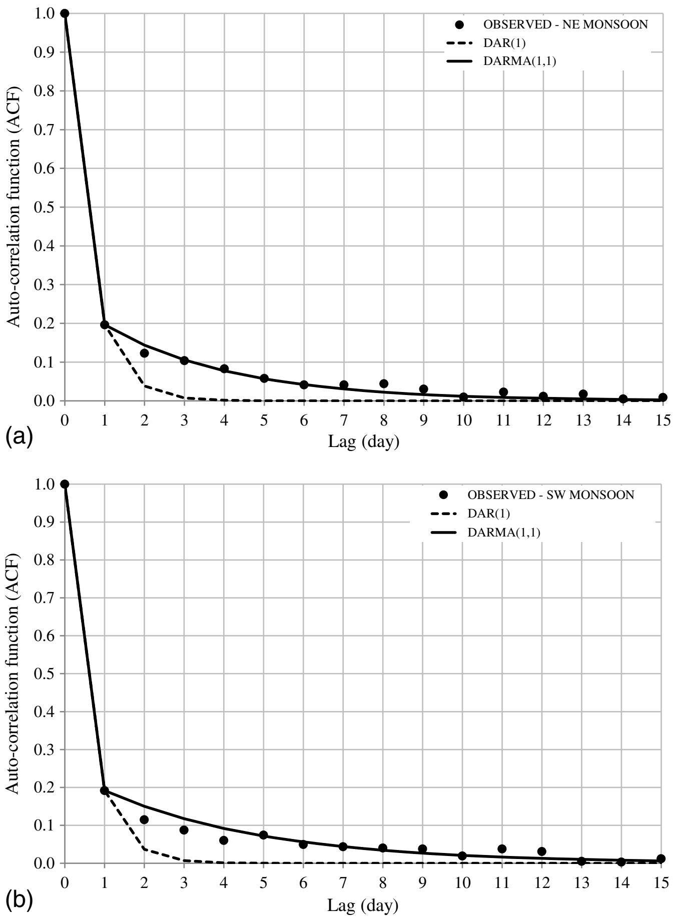

Fig. 3. Observed (dependent) and theoretical ACF for (a) NE monsoons; (b) SW monsoons

$$H_1 = \begin{bmatrix} 0.2450 & 0.0888 \\ 0.0474 & 0.6548 \end{bmatrix} \quad \text{and} \quad H_0 = \begin{bmatrix} 0.6748 & 0.0444 \\ 0.0648 & 0.2333 \end{bmatrix};$$

$$H_1 = \begin{bmatrix} 0.2198 & 0.0610 \\ 0.0471 & 0.6548 \end{bmatrix}, \text{ respectively.}$$

The analysis of run length distributions shows that the DARMA (1,1) model is able to generate the probabilities with the least amount of error, as compared to the DAR(1) model. For example, during the NE (SW) monsoon, the theoretical probability distribution for 2 consecutive wet days estimated for the DARMA (1,1) is 0.197 (0.206), while the observed rainfall data give a probability of 0.207 (0.218). On the other hand, the corresponding probabilities for the DAR(1) model are 0.224 (0.243) for 2 consecutive wet days. Likewise, the sum of squared errors for NE (SW) monsoons for the DARMA(1,1) model is 0.0015 (0.0021), as compared to 0.0091 (0.0079) for the DAR(1) model. In addition, Fig. 4 gives the wet run length distributions during NE monsoons for wet run lengths varying from 1 to 14 days.

Simulating Sequences of Daily Rainfall Using DARMA (1,1) Models

Sequences of daily rainfall were generated separately for the NE and SW monsoons. For each monsoon season, two simulations were performed; simulation A consists of 100 samples of 9,600 days while simulation B consists of a very long sequence, i.e. 1,000,000 days. The sample size of 9,600 days was chosen because this size is about the same as that of the observed data for each monsoon. The main purpose of simulation A is to assess the

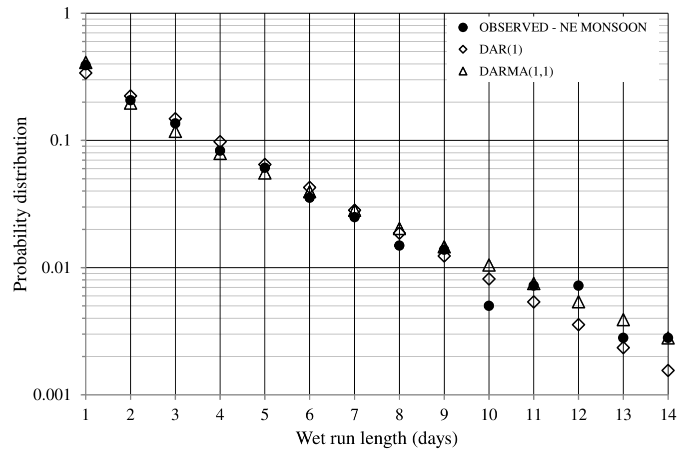

Fig. 4. Probability distribution of wet run lengths for NE monsoons

variability of a number of statistics that are derived from the fitted DARMA(1,1) model and to compare them to those of the historical sample measured at Subang Airport.

Simulation B was conducted to estimate the population properties of the fitted DARMA(1,1) model, particularly those that cannot be derived in analytical form. This analysis may also indicate whether some statistics obtained from the limited historical sample may show some evidence of departures or bias.

The statistics estimated in both simulations include the mean and standard deviation of the amount of rainfall, maximum rainfall in a day, lag-1 autocorrelation coefficient, and the maximum wet and dry run lengths. The mean, standard deviation, and the maximum daily rainfall were included to evaluate the statistics of the generated rainfall amounts, while the lag-1 autocorrelation and maximum wet and dry run lengths are used to evaluate the statistics of the simulated sequences of the occurrence of daily rainfall. Table 2 summarizes the statistics of the observed and simulated daily rainfall events at Subang Airport during the NE and SW monsoons.

Generally, the statistics of generated rainfall from simulation A show good results for both monsoons. For NE monsoons, all statistics derived from the observed data fall within two standard deviations relative to the mean calculated from the simulated samples. The results are similar for SW monsoon except for the mean, which falls within three standard deviations. The coefficient of variation obtained from the generated samples are generally small, i.e., on the order of 0.02 for the mean and standard deviation, 0.065 for the lag-1 correlations, and about 0.16 for the statistics related to the maximums.

For the NE and SW monsoons, the results obtained from simulation B regarding the mean, standard deviation, and lag-1 correlation are similar to those obtained from simulation A. The maximum rainfall obtained from simulation B is about 70% higher than that obtained from the historical sample for NE monsoon and about 100% higher for SW monsoon. The higher results are expected because of the longer sample considered for simulation B. Likewise, the maximum wet run length and the maximum dry run length obtained from simulation B are higher than those obtained from the historical data and the differences are more noticeable for the SW monsoon.

Further verification was made by comparing the probabilities of wet and dry run lengths obtained from the observed data and from simulations A and B. Overall, the analyses show reasonable results. For example, for the NE (SW) monsoon the estimated probabilities for 5 consecutive rainy days derived from the observed data and

Table 2. Statistics for Observed and Simulated Daily Rainfall during NE (SW) Monsoon

Statistics	Observed data NE monsoon (SW monsoon)	Simulation A—Simulated daily rainfall (based on 100 samples, each 9,600 days)		Simulation B—Simulated daily rainfall (based on one sample of 1,000,000 days)
		Mean	Standard deviation	
Mean (mm)	13.4 (12.0)	12.9 (12.9)	0.3 (0.3)	12.9 (12.9)
Standard deviation (mm)	17.6 (16.8)	17.2 (17.2)	0.4 (0.4)	17.3 (17.2)
Maximum rainfall in a day (mm)	171.5 (158.3)	178.9 (173.6)	24.6 (26.4)	292.2 (325.1)
Lag-1 correlation	0.196 (0.192)	0.179 (0.181)	0.012 (0.011)	0.181 (0.180)
Maximum wet run length (days)	31 (17)	24 (20)	4 (3)	34 (27)
Maximum dry run length (days)	21 (20)	16 (20)	3 (3)	25 (28)

simulations A and B, respectively, are 0.0588 (0.0469), 0.0550 (0.0458), and 0.0519 (0.0434). Likewise, the estimated probabilities for 7 consecutive dry days obtained from the observed data and simulations A and B, respectively, are 0.0130 (0.0216), 0.0134 (0.0216), and 0.0135 (0.0209). In addition, Fig. 5(a) gives the plot of wet probability distributions obtained from the observed data and simulations A and B for NE monsoons and Fig. 5(b) show results for dry run lengths. Therefore, the results obtained suggest that the DARMA(1,1) model for representing the rainfall occurrence and the two-parameter gamma model for representing the distribution of the rainfall amount for a given rainfall duration give reasonable results for simulating the sequences of daily rainfall for the monsoons at Subang Airport.

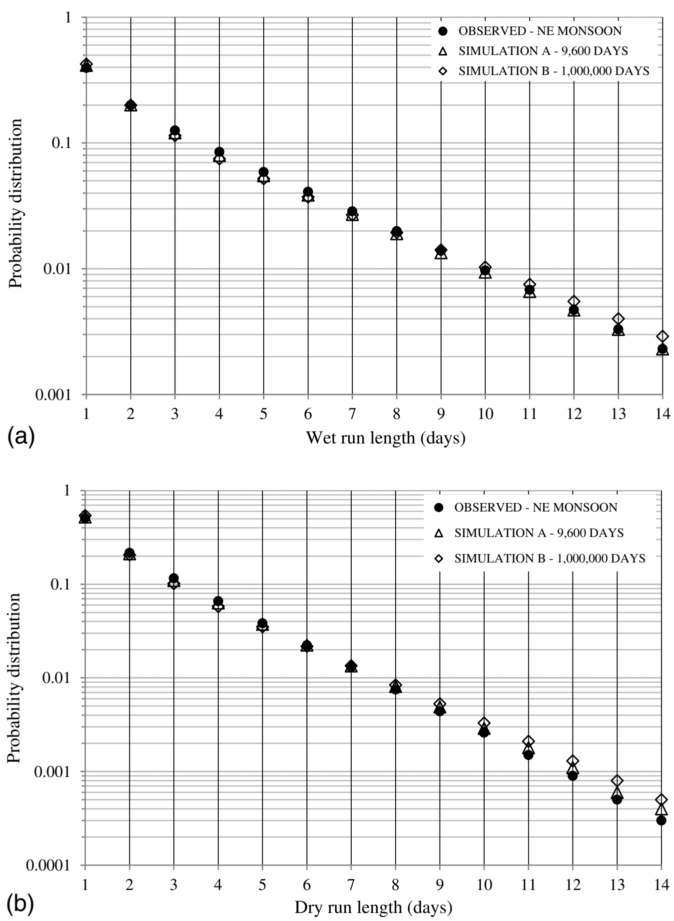

Fig. 5. Probability distribution of (a) wet run lengths for NE monsoons generated from simulations A and B; (b) dry run lengths for NE monsoons generated from simulations A and B

Return Period Curves

Eqs. (19)–(21) were used to calculate the theoretical return periods, which were then compared with those obtained from the observed data. The threshold amounts of rainfall, x_0 (in mm), considered for this analysis are 1, 13, 30, 60, 90, 120, and 150. The smallest amount of 1 mm is selected to represent the majority of rainfall events and 13 mm is the average daily rainfall. The other amounts are selected because they represent significant values of rainfall, especially during multiday events. The return periods were estimated separately for the NE and SW monsoons.

The observed return period, T, was calculated for a given duration, e.g. 2 days, for precipitation exceeding a specified threshold, say x_0 (in mm). Empirical frequency analysis was made using the Weibull (plotting position) formula and $T = 1/p$, where p is the exceeding probability.

The comparison of the estimated return periods for NE monsoons is shown in Fig. 6(a). In general, the theoretical return period curves obtained from the DARMA(1,1)-gamma model show good agreement with those of the observed data. Fig. 6(a) shows the complex behavior of the return period curves for various rainfall durations and rainfall threshold amounts of the corresponding events. For the smallest amounts, the return periods increase as the rainfall durations increase (e.g., the estimated return periods for multiday rainfall events for amounts >1 mm are higher as compared to the 1-day event). It can be said that the 1-day events occur more often as compared with 2 consecutive days or more. However, for higher rainfall amounts (in mm), e.g, 30, 60, 90, 120, and 150, the return periods decrease as the rainfall duration t increases, reach a minimum, then increase steadily after that. Similar patterns of return period curves are shown in Fig. 6(b) for the case of SW monsoons. Both Figs. 6(a and b) indicate that for the frequent events where there are a large number of historical observations, the return periods estimated from the fitted DARMA(1,1)-gamma models (theoretical) correspond reasonably well with those estimated from the observations. However, in those cases where a very small number of observations are available because of the rarity of the events (e.g., for large rainfall thresholds or large rainfall durations) some significant departures may occur. In addition, the comparison of Figs. 6(a and b) show that while the return period patterns for the NE and SW monsoons are similar, some differences can be noted particularly for rainfall events longer than 4 days.

Extreme Rainfall Events

A sample of 1,000,000 days (more than 2,000 years) of daily rainfall was generated to further verify the applicability of Eqs. (19)–(21) for estimating the return period. In this case, the estimations were performed for significant rainfall threshold values, i.e. 50, 100, 150, 200, 250, 300, and 350 mm. Values in

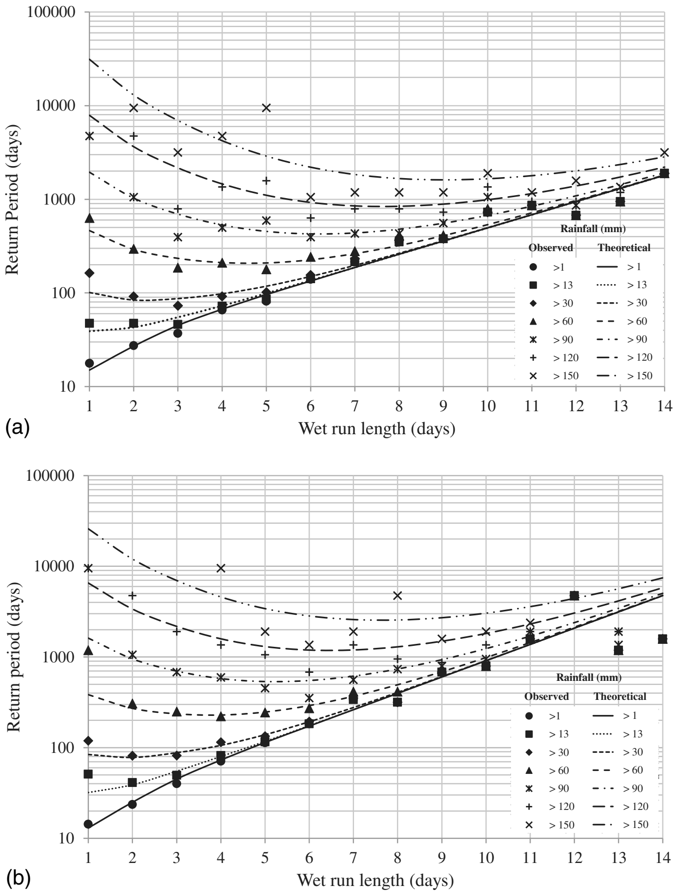

Fig. 6. Observed and theoretical return periods for (a) NE monsoons and (b) SW monsoons

excess of 150 mm are considered to represent extreme to rare events and can cause devastating floods on large watersheds.

Fig. 7 shows the comparison between calculated return periods based on the generated sample and the theoretical equations corresponding to NE monsoons. The 2,000-year rainfall was generated using the fitted DARMA(1,1)-gamma models as described earlier. Generally, the return period curves for most rainfall

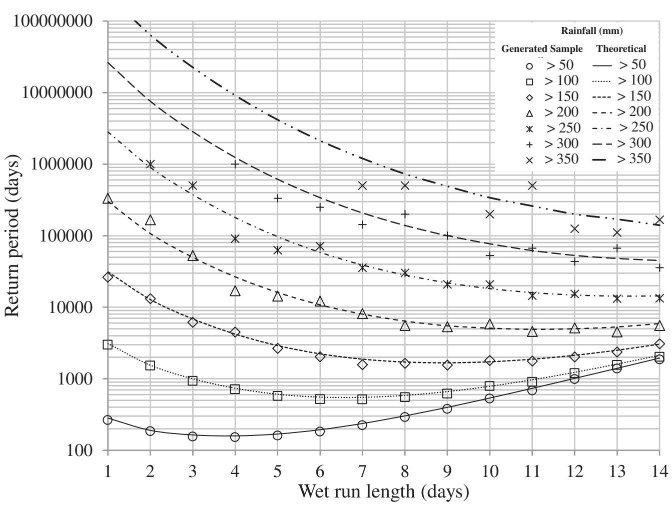

Fig. 7. Return periods calculated from generated daily rainfall sequence (1,000,000 days) and theoretical equations for NE monsoons

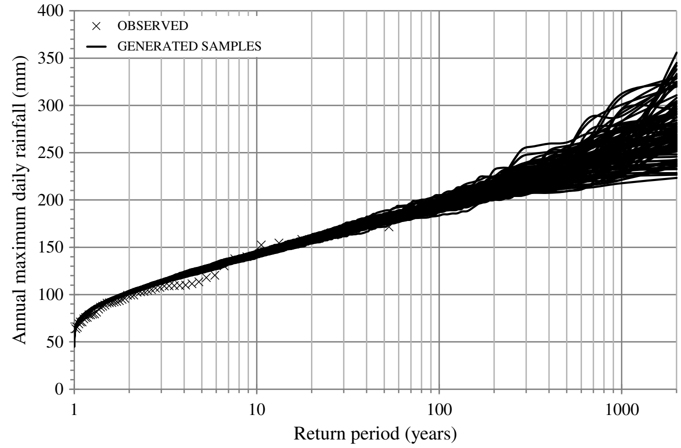

Fig. 8. Frequency distributions of maximum daily/annual rainfall obtained from the 52-year record and from the 100 samples of annual rainfall derived from the generated daily rainfall based on the fitted DARMA(1,1)/gamma models

threshold amounts show excellent agreement, which further verifies that Eqs. (19)–(21) are reliable to estimate the return periods for multiday events. Some departures are noted for the very extreme amounts, i.e. 300 and 350 mm, which is attributed to the variability of the generated samples.

Furthermore, it was desired to illustrate the applicability of the fitted DARMA(1,1)-gamma models for obtaining via stochastic rainfall generation the variability of the T-year rainfall quantiles, i.e., the annual frequency distribution of the maximum daily rainfall at Subang Airport. Fig. 8 shows the computed annual frequency distribution of maximum daily rainfall obtained from the 52-year historical records and from the 100 samples of 2,000 years of rainfall derived from the generated daily rainfall based on the fitted DARMA(1,1)-gamma models. This result may be particularly useful in cases of short records that may be available at a given site where the occurrence of daily rainfall and the variability of rainfall amount can be modeled using the procedures described in this paper and the variability of the annual frequency distribution obtained based on data generation.

Summary and Conclusions

The analysis of both the Northeast (NE) and Southwest (SW) monsoon rainfall precipitation events at Subang Airport in Malaysia from 1960–2011 demonstrates the following:

1. The majority (57%) of rainfall events are multiday events;
2. The distribution of daily rainfall is well reproduced (Fig. 1) with a gamma distribution. Likewise, the distribution of multiday rainfall events is also well reproduced with a gamma distribution. Considering the well-known properties of the sum of independent gamma variables enables the derivation of a simple two-parameter gamma distribution to fit the distribution of daily and multiday rainfall;
3. As expected, the probability of rainfall occurrence (or nonoccurrence) on a given day is not independent, but depends on whether the previous day was dry or wet. The conditional probabilities increase with the number of consecutive rainy (or dry) days (Fig. 2);
4. The rainfall occurrence for both NE and SW monsoons at Subang Airport can be well represented by the DARMA(1,1) model. It reproduces reasonably well a number of key statistics, such as the autocorrelation function (Fig. 3) and run lengths;

5. A simple algorithm has been suggested for estimating the return period for multiday rainfall events defined by combining the DARMA(1,1) model and the gamma distribution. The resulting DARMA(1,1)-gamma model yields good agreement (Fig. 6) between the return periods estimated from the observed historical sample and those estimated by the proposed method. As expected, some departures occur in cases of rare rainfall extremes for which very few observations are available; and
6. The proposed DARMA(1,1)-gamma model also enables the estimation of the variability in T-year daily maximum rainfall, which could be especially useful for the analysis of extreme rainfall precipitation in areas with short historical records.

Appendix. Probability Distributions of the Wet and Dry Run Lengths for DARMA(1,1) Model

The procedures to estimate the probability distributions of wet and dry run lengths are given in this section.

The one-step transitional probabilities $H_k(u,v)$ can be written as (Jacobs and Lewis 1978)

$$H_k(u,v) = P(X_{t+1}=k, A_{t+1}=v|X_t=m, A_t=u)$$
$$= P(X_{t+1}=k, A_{t+1}=v|A_t=u)$$

where (X_{t+1}, A_{t+1}) is independent of X_t and u, v, k, and m are 0,1 values.

The transition probability matrices are

$$H_0(u,v) = \begin{bmatrix} \lambda(1-\beta)+[1-\lambda(1-\beta)]\pi_0 & (1-\beta)(1-\lambda)\pi_1 \\ \beta(1-\lambda)\pi_0 & \beta\lambda\pi_0 \end{bmatrix}$$
$$H_1(u,v) = \begin{bmatrix} \beta\lambda\pi_1 & \beta(1-\lambda)\pi_1 \\ (1-\beta)(1-\lambda)\pi_0 & \lambda(1-\beta)+[1-\lambda(1-\beta)]\pi_1 \end{bmatrix}$$

Lloyd and Salem (1979) introduced the use of *label variable* $W_t = 2X_t + A_t$ to convert the first-order bivariate Markov chain (X_t, A_t) into a four-state simple Markov chain. (X_t, A_t) can have values of 0 or 1, so there are four possibilities for the value of W_t, i.e., $\{0,1,2,3\}$. Table 3 summarizes the W_t values.

The value of 0 and 1 for W_t corresponds to the state of 0 in X_t, which implies a dry day. In the same manner, a wet day is represented as 1 in X_t, which gives the value of 2 and 3 for W_t.

The transition probabilities are given as

$$p_W(0,1) = P(X_{t+1}=0, A_{t+1}=1|X_t=0, A_t=0)$$
$$= P(X_{t+1}=0, A_{t+1}=1|A_t=0) = H_0(0,1)$$
$$p_W(0,2) = P(X_{t+1}=1, A_{t+1}=0|X_t=0, A_t=0)$$
$$= P(X_{t+1}=1, A_{t+1}=0|A_t=0) = H_1(0,0)$$
$$p_W(0,3) = P(X_{t+1}=1, A_{t+1}=1|X_t=0, A_t=0)$$
$$= P(X_{t+1}=1, A_{t+1}=1|A_t=0) = H_1(0,1)$$
$$p_W(1,0) = P(X_{t+1}=0, A_{t+1}=0|X_t=0, A_t=1)$$
$$= P(X_{t+1}=0, A_{t+1}=0|A_t=1) = H_0(1,0)$$

Table 3. Four State Markov Chain, W_t

Variable	Value			
X_t	0	0	1	1
A_t	0	1	0	1
W_t	0	1	2	3

Transition probability matrix, Q, of the univariate Markov chain W_t is

$$Q = \begin{matrix} & 0 & 1 & 2 & 3 \\ 0 & H_0(0,0) & H_0(0,1) & H_1(0,0) & H_1(0,1) \\ 1 & H_0(1,0) & H_0(1,1) & H_1(1,0) & H_1(1,1) \\ 2 & H_0(0,0) & H_0(0,1) & H_1(0,0) & H_1(0,1) \\ 3 & H_0(1,0) & H_0(1,1) & H_1(1,0) & H_1(1,1) \end{matrix}$$

and its marginal distribution is

$$P[W_i=0] = P(X_t=0, A_t=0)$$
$$= P(X_t=0, A_t=0|A_{t-1}=0)P(A_{t-1}=0)$$
$$+ P(X_t=0, A_t=0|A_{t-1}=0)P(A_{t-1}=1)$$
$$= H_0(0{,}0)\pi_0 + H_0(1{,}0)\pi_1$$
$$P[W_i=1] = P(X_t=0, A_t=1)$$
$$= P(X_t=0, A_t=1|A_{t-1}=0)P(A_{t-1}=0)$$
$$+ P(X_t=0, A_t=1|A_{t-1}=1)P(A_{t-1}=1)$$
$$= H_0(0{,}1)\pi_0 + H_0(1{,}1)\pi_1$$
$$P[W_i=2] = P(X_t=1, A_t=0)$$
$$= P(X_t=1, A_t=0|A_{t-1}=0)P(A_{t-1}=0)$$
$$+ P(X_t=1, A_t=0|A_{t-1}=1)P(A_{t-1}=1)$$
$$= H_1(0{,}0)\pi_0 + H_1(1{,}0)\pi_1$$
$$P[W_i=3] = P(X_t=1, A_t=1)$$
$$= P(X_t=1, A_t=1|A_{t-1}=0)P(A_{t-1}=0)$$
$$+ P(X_t=1, A_t=1|A_{t-1}=1)P(A_{t-1}=1)$$
$$= H_1(0{,}1)\pi_0 + H_1(1{,}1)\pi_1$$

Probability distributions of wet and dry run lengths of t consecutive days for the DARMA(1,1) model, denoted by $P(T_1=t)$ and $P(T_0=t)$, respectively, can be calculated using conditional probabilities, as given by Chang et al. (1984b)

$$P(T_1=t) = P(X_0=0, X_1=1, \ldots, X_t=1,$$
$$X_{t+1}=0|X_0=0, X_1=1); t=1,2,\ldots$$
$$= \frac{P(X_0=0, X_1=1, \ldots, X_t=1, X_{t+1}=0)}{P(X_0=0, X_1=1)}$$

Note that

$$P(X_0=0, X_1=1, \ldots, X_t=1, X_{t+1}=0) = \{P[W_0=0][H_1^{(n)}(0)$$
$$- H_1^{(n+1)}(0)] + P[W_0=1][H_1^{(n)}(1) - H_1^{(n+1)}(1)]\}$$

where $H_1^{(n)}(j) = H_1^{(n)}(j,0) + H_1^{(n)}(j,1); j=0{,}1$

Both $H_1^{(n)}(j,0)$and$H_1^{(n)}(j,1)$ are elements of the n step transition probability matrix

$$P(X_0=0, X_1=1) = \sum_{k=0}^{1}\sum_{j=0}^{1} H_1(j,k)\left[\pi_j - \sum_{l=0}^{1} H_1(l,j)\pi_l\right]$$

where $H_1(j,k), H_1(l,j)$ = elements of the n step transitional probability matrix

$$P(T_0 = t) = P(X_0 = 1, X_1 = 0, \ldots, X_t = 0, X_{t+1} = 1 | X_0 = 1, X_1 = 0); t = 1,2, \ldots$$
$$= \frac{P(X_0 = 1, X_1 = 0, \ldots, X_t = 0, X_{t+1} = 1)}{P(X_0 = 1, X_1 = 0)}$$
$$P(X_0 = 1, X_1 = 0, \ldots, X_t = 0, X_{t+1} = 1) = \{P[W_0 = 2][H_0^{(n)}(0) - H_0^{(n+1)}(0)] + P[W_0 = 3][H_0^{(n)}(1) - H_0^{(n+1)}(1)]\}$$

where $H_0^{(n)}(j) = H_0^{(n)}(j,0) + H_0^{(n)}(j,1), j = 0,1$

Both $H_0^{(n)}(j,0) and H_0^{(n)}(j,1)$ are elements of the n step transition probability matrix

$$P(X_0 = 1, X_1 = 0) = \sum_{k=0}^{1}\sum_{j=0}^{1} H_0(j,k)\left[\pi_j - \sum_{l=0}^{1} H_0(l,j)\pi_l\right]$$

where $H_0^{(n)}(j,k), H_0^{(n)}(l,j)$ are elements of the n step transition probability matrix.

Acknowledgments

This study has been carried out at Colorado State University during the Ph.D. studies of the first author. Financial support for the first author from the Ministry of Education, Malaysia and Universiti Kebangsaan Malaysia is gratefully acknowledged.

References

Abu Bakar, S., Yusuf, M. F., and Amly, W. S. (2007). "Johor almost paralysed (. . .)." ⟨http://www.utusan.com.my/utusan/ info.asp?y=2007&dt=0115&pub=Utusan_Malaysia&sec=Muka_Hadapan&pg=mh_01.htm⟩ (Feb. 11, 2013).

Buishand, T. A. (1977). *Stochastic modelling of daily rainfall sequences*, Veenman & Zonen, Wageningen, the Netherlands.

Buishand, T. A. (1978). "The binary DARMA (1, 1) process as a model for wet-dry sequences." *Technical Note*, Agricultural Univ., Wageningen, the Netherlands.

Burton, A., Fowler, H. J., Blenkinsop, S., and Kilsby, C. G. (2010). "Downscaling transient climate change using a Neyman-Scott rectangular pulses stochastic rainfall model." *J. Hydrol.*, 381(1–2), 18–32.

Burton, A., Kilsby, C. G., Fowler, H. J., Cowpertwait, P. S. P., and O'Connell, P. E. (2008). "RainSim: A spatial-temporal stochastic rainfall modelling system." *Environ. Modell. Software*, 23(12), 1356–1369.

Cancelliere, A., and Salas, J. D. (2002). "Characterizing the recurrence of hydrologic droughts." *AGU Fall Meeting San Francisco*, American Geophysical Union, Washington, DC.

Cancelliere, A. and Salas, J. D. (2004). "Drought length properties for periodic-stochastic hydrologic data." *Water Resour. Res.*, 40(2), 1–13.

Cancelliere, A., and Salas, J. D. (2010). "Drought probabilities and return period for annual streamflow series." *J. Hydrol.*, 391(1–2), 77–89.

Cazacioc, L., and Cipu, E. C. (2005). "Evaluation of the transition probabilities for daily precipitation time series using a Markov chain model." *Mathematics in Engineering and Numerical Physics, Proc., 3rd Int. Colloquium*, Balkan Society of Geometers, Geometry Balkan Press, Bucharest, Romania, 82–92.

Chang, T. J., Kavvas, M. L., and Delleur, J. W. (1982). "Stochastic daily precipitation modeling and daily streamflow transfer processes." *Technical Rep. No. 146*, Purdue Univ. Water Resources Research Center, West Lafayette, IN.

Chang, T. J., Kavvas, M. L., and Delleur, J. W. (1984a). "Daily precipitation modeling by discrete autoregressive moving average processes." *Water Resour. Res.*, 20(5), 565–580.

Chang, T. J., Kavvas, M. L., and Delleur, J. W. (1984b). "Modeling of sequences of wet and dry days by binary discrete autoregressive moving average processes." *J. Clim. Appl. Meteorol.*, 23(9), 1367–1378.

Chin, E. H. (1977). "Modeling daily precipitation occurrence process with Markov chain." *Water Resour. Res.*, 13(6), 949–956.

Chung, C. H., and Salas, J. D. (2000). "Drought occurrence probabilities and risks of dependent hydrologic processes." *J. Hydrol. Eng.*, 10.1061/(ASCE)1084-0699(2000)5:3(259), 259–268.

Cindrić, K. (2006). "The statistical analysis of wet and dry spells by binary DARMA (1, 1) model in Split, Croatia." *BALWOIS Conf. 2006*, Balwois, Skopje, Republic of Macedonia.

Cowpertwait, P. S. P. (1995). "A generalized spatial-temporal model of rainfall based on a clustered point process." *Proc. R. Soc. London*, 450(1938), 163–175.

Cowpertwait, P. S. P., Isham, V., and Onof, C. (2007). "Point process models of rainfall: Developments for fine-scale structure." *Proc. R. Soc. London*, 463(2086), 2569–2587.

Cowpertwait, P. S. P., O'Connell, P. E., Metcalfe, A. V., and Mawdsley, J. A. (1996). "Stochastic point process modelling of rainfall. I. Single-site fitting and validation." *J. Hydrol.*, 175(1–4), 17–46.

Dahale, S. D., Panchawagh, N., Singh, S. V., Ranatunge, E. R., and Brikshavana, M. (1994). "Persistence in rainfall occurrence over tropical south-east Asia and equatorial Pacific." *Theor. Appl. Climatol.*, 49(1), 27–39.

Dastidar, A. G., Ghosh, D., Dasgupta, S., and De, U. K. (2010). "Higher order Markov chain models for monsoon rainfall over West Bengal, India." *Indian J. Radio Space Phys.*, 39(1), 39–44.

Delleur, J. W., Chang, T. J., and Kavvas, M. L. (1989). "Simulation models of sequences of dry and wet days." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)0733-9437(1989)115:3(344), 344–357.

Deni, S. M., Jemain, A. A., and Ibrahim, K. (2009). "Fitting optimum of order Markov chain models for daily rainfall occurrences in Peninsular Malaysia." *Theor. Appl. Meteorol.*, 97(1–2), 109–121.

Douglas, E. M., Vogel, R. M., and Kroll, C. N. (2002). "Impact of streamflow persistence on hydrologic design." *J. Hydrol. Eng.*, 10.1061/(ASCE)1084-0699(2002)7:3(220), 220–227.

Eagleson, P. S. (1978). "Climate, soil and vegetation: I. Introduction to water balance dynamics." *Water Resour. Res.*, 14(5), 705–712.

Fernández, B., and Salas, J. D. (1999a). "Return period and risk of hydrologic events. I: Mathematical formulation." *J. Hydrol. Eng.*, 10.1061/(ASCE)1084-0699(1999)4:4(297), 297–307.

Fernández, B., and Salas, J. D. (1999b). "Return period and risk of hydrologic events. II: Applications." *J. Hydrol. Eng.*, 10.1061/(ASCE)1084-0699(1999)4:4(308), 308–316.

Feyerherm, A. M., and Bark, L. D. (1965). "Statistical methods for persistent precipitation patterns." *J. Appl. Meteorol.*, 4(3), 320–328.

Glasbey, C. A., Cooper, G., and McGechan, M. B. (1995). "Disaggregation of daily rainfall by conditional simulation from a point-process model." *J. Hydrol.*, 165(1–4), 1–9.

Goel, N. K., Seth, S. M., and Chanra, S. (1998). "Multivariate modeling of flood flows." *J. Hydraul. Eng.*, 10.1061/(ASCE)0733-9429(1998)124:2(146), 146–155.

González, J., and Valdés, J. B. (2003). "Bivariate drought recurrence analysis using tree ring reconstruction." *J. Hydrol. Eng.*, 10.1061/(ASCE)1084-0699(2003)8:5(247), 247–258.

Haan, C. T., Allen, D. M., and Street, J. O. (1976). "A Markov chain model of daily rainfall." *Water Resour. Res.*, 12(3), 443–449.

Jacobs, P. A., and Lewis, P. A. W. (1978). "Discrete time series generated by mixtures. I: Correlational and runs properties." *J. R. Stat. Soc. Ser. B (Method.)*, 40(1), 94–105.

Jacobs, P. A., and Lewis, P. A. W. (1983). "Stationary discrete autoregressive-moving average time series generated by mixtures." *J. Time Ser. Anal.*, 4(1), 19–36.

Jimoh, O. D., and Webster, P. (1996). "The optimum order of a Markov chain model for daily rainfall in Nigeria." *J. Hydrol.*, 185(1–4), 45–69.

Katz, R. W. (1977). "Precipitation as a chain-dependent process." *J. Appl. Meterol.*, 16(7), 671–676.

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

Kavvas, M. L., and Delleur, J. W. (1981). “A stochastic cluster model of daily rainfall sequences.” *Water Resour. Res.*, 17(4), 1151–1160.

Kedem, B. (1980). *Binary time series*, Marcel Dekker, New York.

Khaliq, M. N., and Cunnane, C. (1996). “Modelling point rainfall occurrences with the modified Bartlett-Lewis rectangular pulses model.” *J. Hydrol.*, 180(1–4), 109–138.

Kim, T. W., Valdés, J. B., and Yoo, C. (2003). “Nonparametric approach for estimating return periods of droughts in arid regions.” *J. Hydrol. Eng.*, 10.1061/(ASCE)1084-0699(2003)8:5(237), 237–246.

Kite, G. W. (1978). *Frequency and risk analyses in hydrology*, 2nd Ed., Water Resources, Littleton, CO.

Lloyd, E. H. (1970). “Return periods in the presence of persistence.” *J. Hydrol.*, 10(3), 291–298.

Lloyd, E. H., and Saleem, S. D. (1979). “Waiting time to first achievement of specified levels in reservoirs subject to seasonal Markovian inflows.” *Inputs for risk analysis in water systems*, E. A. McBean, K. W. Hipel, and T. E. Unay, eds., Water Resources, Fort Collins, CO, 339–379.

Loaiciga, H. A., and Mariňo, M. A. (1991). “Recurrence interval of geophysical events.” *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1991)117:3(367), 367–382.

Mimikou, M. (1983). “Daily precipitation occurrences modeling with Markov chain of seasonal order.” *Hydrol. Sci. J.*, 28(2), 221–232.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the theory of statistics*, 3rd Ed., McGraw-Hill, Tokyo.

Muhammad, N. S. (2013). “Probability structure and return period calculations for multi-day monsoon rainfall events at Subang, Malaysia.” Ph.D. dissertation, Dept. of Civil and Environmental Engineering, Colorado State Univ., Fort Collins, CO.

Obeysekera, J., Tabios, G., and Salas, J. D. (1987). “On parameter estimation of temporal rainfall models.” *Water Resour. Res.*, 23(10), 1837–1850.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1987). “Some models for rainfall based on stochastic point process.” *Proc. R. Soc. London*, A410 (1839), 269–288.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V. (1988). “A point process model for rainfall: Further developments.” *Proc. R. Soc. London*, A417 (1853), 283–298.

Roldan, J., and Woolhiser, D. A. (1982). “Stochastic daily precipitation models. 1: A comparison of occurrence process.” *Water Resour. Res.*, 18(5), 1451–1459.

Salas, J. D., et al. (2005). “Characterizing the severity and risk of drought in the Poudre River, Colorado.” *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2005)131:5(383), 383–393.

Semenov, M. A., Brooks, R. J., Barrow, E. M., and Richardson, C. W. (1998). “Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates.” *Clim. Res.*, 10, 95–107.

Şen, Z. (1999). “Simple risk calculations in dependent hydrological series.” *Hydrol. Sci. J.*, 44(6), 871–878.

Sharma, T. C. (1996). “Simulation of the Kenyan longest dry and wet spells and the largest rain-sums using a Markov model.” *J. Hydrol.*, 178(1–4), 55–67.

Shiau, J. T., and Shen, H. W. (2001). “Recurrence analysis of hydrologic droughts of differing severity.” *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2001)127:1(30), 30–40.

Small, M. J., and Morgan, D. J. (1986). “The relationship between a continuous-time renewal model and a discrete Markov chain model of precipitation occurrence.” *Water Resour. Res.*, 22(10), 1422–1430.

Suhaila, J., and Jemain, A. A. (2007). “Fitting daily rainfall amount in Malaysia using the normal transform distribution.” *J. Appl. Sci.*, 7(14), 1880–1886.

Tan, S. K., and Sia, S. Y. (1997). “Synthetic generation of tropical rainfall time series using an event-based method.” *J. Hydrol. Eng.*, 10.1061/(ASCE)1084-0699(1997)2:2(83), 83–89.

Todorovic, P. (1968). “A mathematical study of precipitation phenomena.” *Rep. CER 67-68 PT65*, Engineering Research Center, Colorado State Univ., Fort Collins, CO.

Todorovic, P., and Yevjevich, V. (1969). “Stochastic processes of precipitation.” Colorado State Univ., Fort Collins, CO, 61.

Verhoest, N. E. C., Vandenberghe, S., Cabus, P., Onof, C., Meca-Figueras, T., and Jameleddine, S. (2010). “Are stochastic point rainfall models able to preserve extreme flood statistics?” *Hydrol. Processes*, 24(23), 3439–3445.

Vogel, R. M. (1987). “Reliability indices for water supply systems.” *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1987)113:4(563), 563–579.

Von Neumann, J. (1941). “Distribution of the ratio of the mean square successive difference to the variance.” *Ann. Math. Stat.*, 12(4), 367–395.

Wallis, T. W. R., and Griffiths, J. F. (1995). “An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator (EPIC).” *Agric. Forest Meteorol.*, 73(1–2), 115–133.

Wilks, D. S. (1998). “Multisite generalization of a daily stochastic precipitation generation model.” *J. Hydrol.*, 210(1–4), 178–191.

Woodyer, K. D., McGilchrist, C. A., and Chapman, T. G. (1972). “Recurrence intervals between exceedances of selected river levels: 4. Seasonal streams.” *Water Resour. Res.*, 8(2), 435–443.

Yevjevich, V. (1984). *Probability and statistics in hydrology*, Water Resources, Littleton, CO.