Multiday Rainfall Simulations for Malaysian Monsoons

Nur S. Muhammad and Pierre Y. Julien

Abstract This study examines the suitability of the discrete autoregressive and moving average [DARMA(1,1)] model to simulate the sequences of daily rainfall data in Malaysia. The daily monsoon rainfall data recorded at Subang Airport are used to test this modeling approach. The autocorrelation function and probability distributions of wet and dry run lengths estimated from the DARMA(1,1) model matched the sample values quite well. Both theoretical and sample autocorrelation functions slowly decay to zero at day 15. The theoretical probability distribution for two consecutive wet days estimated for the DARMA(1,1) is 0.1966, while the observed rainfall data give a probability of 0.2066. Additionally, the sum of squared errors for the DARMA(1,1) model were very small, i.e., 0.0015. Furthermore, two simulations were done, i.e., 100 samples of 9,600 days (simulation A) and a very long sequence of 1,000,000 days (simulation B). This was done to test the capability of DARMA(1,1) to model a long sequence of daily rainfall. The statistics examined in this study include the lag-1 autocorrelation coefficient (lag-1 ACF) and the maximum wet and dry run lengths. Generally, the statistics of generated rainfall for simulation A fall within two standard deviations from the sample. The diversions of these statistics were reasonable, considering the sample size used in this study. The estimated lag-1 ACF for simulation B was slightly lower than the sample. The maximum wet and dry run lengths were much higher than the observed data because of the different sample sizes. It is concluded that the DARMA(1,1) model is able to simulate the long sequences wet and dry days and preserving the statistics within reasonable accuracy.

Keywords Multiday rainfall • Monsoon rainfall precipitation • Stochastic modeling • DARMA(1,1)

N.S. Muhammad (✉)
Department of Civil and Structural Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia
e-mail: shazwani.muhammad@ukm.edu.my

P.Y. Julien
Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523, USA

S.H. Abu Bakar et al. (eds.), *ISFRAM 2014*, DOI 10.1007/978-981-287-365-1_10

1 Introduction

Malaysia is located near the equator and experiences hot and humid climate throughout the year. The country is influenced by two major seasons, namely the North East (NE) and South West (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 inter-monsoons. These monsoons bring lots of moisture, and the country receives a total rainfall of between 2,000 and 4,000 mm with 150–200 rainy days annually [1].

Therefore, multiday rainfall events are common in Malaysia and cause particularly devastating floods, especially on large watersheds [2]. Historical events include the two extreme Kota Tinggi floods in December 2006 and January 2007, which resulted from 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 [3].

Considering the nature of climate in Malaysia and the devastating consequences of multiday rainfall events as discussed above, generating the sequences of daily rainfall using stochastic model should be given serious consideration. Additionally, the planning and designing of water resources projects require the analysis of reliable and long-term hydrological data such as rainfall and streamflow, which can be scarce in developing countries such as Malaysia. Therefore, engineers should consider generating synthetic hydrological data using the parameters estimated from the records available in their analysis.

The low order discrete autoregressive family models, i.e., discrete autoregressive [DAR(1)] and discrete autoregressive and moving average [DARMA(1,1)], are frequently used in simulating the sequence of daily rainfall. DAR(1) is also known as the first-order Markov Chain. Reference [4] introduces the concept of Markov Chain model to simulate the occurrence of daily rainfall at Tel Aviv. 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. References [5–13] are among the studies that were successful in modeling the sequence of rainy and dry days using first-order Markov Chains.

Reference [7] applied the first-order Markov Chain to produce ten synthetic sequences of daily rainfall at Universiti Pertanian Malaysia (UPM), Serdang, Selangor, Malaysia. The authors gathered the daily rainfall data from 1968 to 1978 and divided the data into 11 states according to the amount. The simulations were done for the different monsoon seasons in Malaysia: the Northeast (from November to March), two transitional periods (April and October), and the Southwest (from May to September). They found that the first-order Markov Chain was able to reproduce the daily rainfall of any length in the area. However, the synthetic daily rainfall was generated for a period of 1 year only. Therefore, this research did not indicate if the first-order Markov Chain is able to simulate long daily rainfall sequences.

Reference [14] discusses the optimum order of Markov Chain for daily rainfall during North East (NE) and South West (SW) monsoons using two different thresholds, i.e., 0.1 and 10.0 mm. Eighteen rainfall stations in Peninsular Malaysia were used in this study. They found that the optimum order of a Markov Chain varies with the location, monsoon seasons, and the level of threshold. For stations located in the northwestern and eastern regions of Peninsular Malaysia, the occurrence of rainfall (threshold level 10.0 mm) for both monsoons can be represented using a first-order Markov Chain model. Other than that, Markov Chain models of higher order are suitable to represent rainfall occurrence, especially during the NE monsoon, for both levels of threshold. This study shows that the rainfall events in Peninsular Malaysia requires a longer memory model than first-order Markov Chain to simulate the sequence of daily rainfall. However, high order Markov Chain increased the model uncertainty because more parameters have to be used [15]. It also makes the calculations more complex. These disadvantages can be overcome by DARMA(1,1) model.

The DARMA model is a simple tool to model stationary sequences of dependent discrete random variables with specified marginal distribution and correlation structure [16]. The model is stationary; therefore, the rainfall data should be divided into their respective seasons in order to consider the seasonal variations. References [17] and [18] use this model to simulate the sequence of daily rainfall using data collected from several stations in the Netherlands, Suriname, India, and Indonesia. They concluded that that DARMA(1,1) 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 first-order Markov Chain. Other studies that use this model to simulate the sequence of daily rainfalls include [19–23].

This chapter discusses the simulation of daily rainfall using the DARMA(1,1) model. The daily rainfall measurements from the Subang Airport were used in this study. This station is chosen because it provides a long and reliable record of 52 years, i.e., from 1960 to 2011. Separate analyses are conducted for NE and SW monsoons. However, only the results for NE monsoons will be presented here.

2 DARMA(1,1) Model

The DARMA(1,1) model is represented as [16]

$$X_t = U_t Y_t + (1 - U_t) A_{t-1} \tag{1}$$

with

$$X_t = \begin{cases} Y_t & \text{with probability } \beta \\ A_{t-1} & \text{with probability } (1-\beta) \end{cases}$$

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

$$P(U_t = 1) = \beta = 1 - P(U_t = 0) \tag{2}$$

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

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

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} [18]. This variable has the same probability distribution as Y_t but is independent of Y_t. The X_t is not Markovian, but (X_t, A_t) forms a first-order bivariate Markov Chain.

The theoretical autocorrelation function of the DARMA(1,1) model is [18]

$$\mathrm{corr}(X_t, X_{t-k}) = r_k(X) = c\lambda^{k-1}, \quad k \geq 1 \tag{3}$$

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

$$c = (1-\beta)(\beta + \lambda - 2\lambda\beta) \tag{4}$$

The sample autocorrelation function (r_k) for the time series is estimated based on the sequences of dry and rainy days and not the rainfall amounts [22].

$$r_k = \left[\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})\right] \left[\sum_{t=1}^{N} (x_t - \bar{x})^2\right]^{-1} \tag{5}$$

$$\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t \tag{6}$$

where N is the sample size.

Three parameters of the DARMA(1,1) model are π_0 or π_1, λ, and β. The parameter λ may be estimated from the lag-1 autocorrelation coefficient as given in (3) and (4). The parameters π_0 or π_1 may be estimated from (7) and (8).

$$\pi_0 = \frac{\overline{T_0}}{\overline{T_0} + \overline{T_1}} \tag{7}$$

$$\pi_1 = 1 - \pi_0 \tag{8}$$

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

The estimation of λ may be determined by minimizing (9) using and [18] suggested using the ratio of the second to the first autocorrelation coefficients as an initial estimator for λ, as shown in (10).

$$\phi(\lambda) = \sum_{k=1}^{M} \left[r_k - c\lambda^{k-1}\right]^2; \quad k \geq 1 \tag{9}$$

$$\hat{\lambda} = \frac{r_2}{r_1} \tag{10}$$

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

$$\hat{\beta} = \frac{(3\hat{\lambda} - 1) \pm \sqrt{(3\hat{\lambda} - 1)^2 - 4(2\hat{\lambda} - 1)(\hat{\lambda} - \hat{c})}}{2(2\hat{\lambda} - 1)} \tag{11}$$

The probability distributions of wet and dry run lengths for the DARMA(1,1) model are given in [16].

3 Results and Discussion

3.1 Simulating the Sequences of Daily Rainfall Using DARMA(1,1)

The occurrence of rainfall event in this study is treated as a discrete variable. Threshold value is determined using the Von Neumann ratio [24]. The definition of wet is any day with rainfall of more than 0.1 mm, and a dry day received less than or equal to the said amount. This value is chosen because it ensures homogeneity of the time series.

The first step in simulating the sequences of daily rainfall using DARMA(1,1) is to estimate the model parameters, i.e., π_0 or π, λ, and β. The average wet and dry run lengths are calculated from the observed daily rainfall dataset, and the values are $\overline{T_1} = 3.00$ days and $\overline{T_0} = 2.19$ days. Following this, the estimated probabilities of a wet and dry day are $\hat{\pi}_1 = 0.58$ and $\hat{\pi}_0 = 0.42$, respectively. The parameter λ is calculated using (9), based on the Newton–Raphson iteration techniques and initial

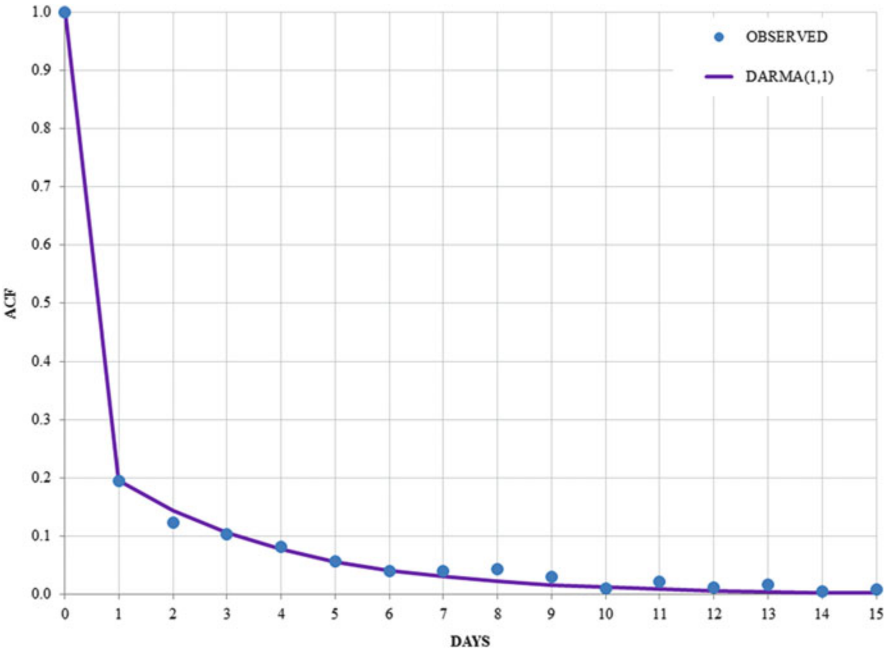

Fig. 1 Sample and theoretical ACF

estimation using (10). Then (11) is applied to estimate the parameter β. These give estimated model parameters as $\hat{\lambda} = 0.7339$ and $\hat{\beta} = 0.5775$.

After all parameters were determined, the theoretical and sample autocorrelation functions (ACFs) are estimated using (3) and (5), respectively. These values are compared using graphical method, as shown in Fig. 1. Excellent agreements between the sample and theoretical ACFs for the estimated DARMA(1,1) model are shown. The ACFs matched even though the number of lag (day) increased. Both sample and theoretical ACFs decay slowly and eventually reach to nearly zero at day 15.

Other than the ACFs, the theoretical and sample probability distributions of wet and dry run lengths were also estimated and compared to further examine the suitability of DARMA(1,1) model to simulate the sequence of daily rainfall.

Excellent agreements were observed between the theoretical and sample probability distributions of wet and dry run lengths. For example, the theoretical probability distribution for two consecutive wet days estimated for the DARMA (1,1) was 0.1966, while the observed rainfall data give a probability of 0.2066. Additionally, the sum of squared errors for the DARMA(1,1) model was very small, i.e., 0.0015. The probability distributions of wet run lengths are illustrated in Fig. 2. Based on the analyses performed on the ACFs and probability distributions of wet and dry run lengths, the authors concluded that the DARMA(1,1) model is suitable to represent the occurrence of daily rainfall at Subang Airport.

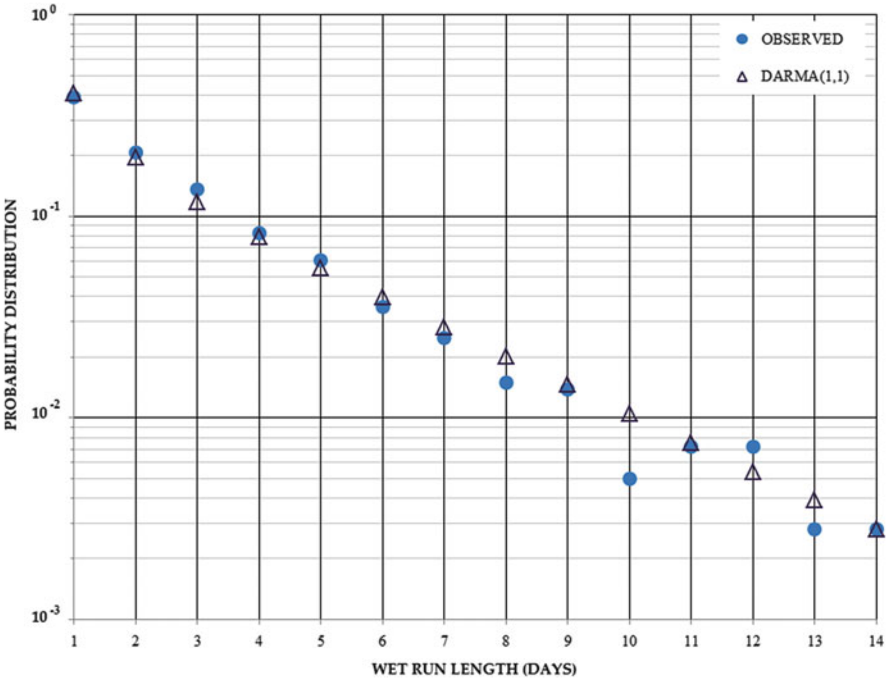

Fig. 2 Probability distribution of wet run lengths

We further analyzed the simulations of daily rainfall sequences using DARMA (1,1) model. A total of two simulations are performed, i.e., simulation A and simulation B. Simulation A was performed to compare it with the statistics of measured rainfall at Subang Airport, in order to ensure that the model applications were persistent and examine the variability of the simulated samples. This simulation consists of 100 samples of 9,600 days. A sample size of 9,600 days is chosen because this is about the same as the sample data for NE monsoon. On the other hand, simulation B consists of a sample with the size of 1,000,000 days. This was done to test the capability of DARMA(1,1) to model a long sequence of daily rainfall.

The statistics examined in this study include the lag-1 autocorrelation coefficient (lag-1 ACF) and the maximum wet and dry run lengths, which are given in Table 1.

Generally, the statistics of generated rainfall for simulation A fall within two standard deviations from the sample. The diversions of these statistics were reasonable, considering the sample size used in this study.

The estimated lag-1 ACF for simulation B was slightly lower than the sample. The maximum wet and dry run lengths were much higher than the observed data because of the difference in sample sizes. These values are significant in estimating the highest possible consecutive wet and dry days over a long period of time. This information is valuable for water resources engineers to plan their strategy. For

Table 1 Statistics for observed and simulated daily rainfall during NE monsoon

Statistics	Sample	Simulation A		Simulation B
		Mean	Standard deviation	
Lag-1 ACF	0.196	0.179	0.012	0.181
Maximum wet run length (days)	31	24	4	34
Maximum dry run length (days)	21	16	3	25

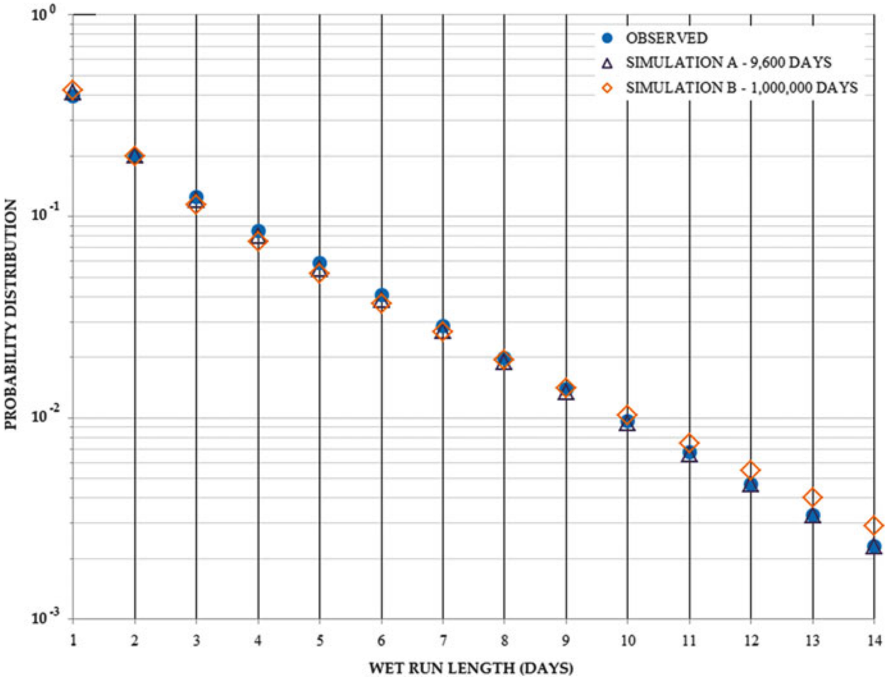

Fig. 3 Probability distributions of wet run lengths generated from simulations A and B

example, what is the best way to ensure constant water supply if it does not rain for 25 days or longer.

In terms of probability distribution functions of wet and dry run lengths, both simulations A and B show good agreement with the sample. As an example, for wet run lengths of 10 days, the estimated probability distribution of the sample and simulations A and B were 0.0097, 0.0094, and 0.0103, respectively. Similarly, the estimated probability distributions for dry run lengths of 7 days were 0.0130, 0.0134, and 0.0135 for the sample and simulations A and B, respectively. Figure 3 detailed the wet run lengths of the sample and simulations A and B. These results indicate that the DARMA(1,1) model is able to simulate the long sequences wet and dry days. These characteristics are important because Malaysia is affected by extreme floods that occur as a result of multiday rainfall events.

4 Conclusions

The DARMA(1,1) model was applied to the daily rainfall data at Subang Airport from 1960 to 2011. Model parameters were estimated from the sample. The autocorrelation functions of the sample are similar to the theoretical values. Additionally, the probability distributions of wet and dry run lengths also show good agreements between the sample and theoretical estimations. These results indicate that DARMA(1,1) was suitable to model the sequences of daily rainfall at Subang station.

Two simulations were performed, including a very long sequence of daily rainfall (1,000,000 days). The statistics estimated for this exercise include lag-1 autocorrelation functions and maximum wet and dry run lengths. The generated sequences' statistics fall within two standard deviations of sample. Additionally, the probability distributions of wet and dry run lengths estimated for the simulated rainfall sequences using DARMA(1,1) are also comparable with the sample. It is concluded that the DARMA(1,1) model is able to simulate the long sequences wet and dry days and preserving the statistics within reasonable accuracy.

Acknowledgment This study was supported by the Universiti Kebangsaan Malaysia (grant number GGPM-2014-046). Also, we would like to thank the Department of Meteorology, Malaysia, for providing the rainfall data. Their assistance is highly appreciated. The authors also appreciate the comments from three anonymous reviewers that led to the improvement of this article.

References

1. J. Suhaila, A.A. Jemain, Fitting daily rainfall amount in Malaysia using the normal transform distribution. J. Appl. Sci. **7**(14), 1880–1886 (2007)
2. J. Abdullah, Distributed runoff simulation of extreme monsoon rainstorms in Malaysia using TREX, Ph.D. thesis, Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO, 2013
3. S. Abu Bakar, M.F. Yusuf, W.S. Amly, Johor hampir lumpuh (. . .), Utusan Melayu, 15 Jan 2007, http://www.utusan.com.my/utusan/info.asp?y=2007&dt=0115&pub=Utusan_Malaysia&sec=Muka_Hadapan&pg=mh_01.htm. Accessed 11 Feb 2013
4. K.R. Gabriel, J. Neuman, A Markov chain model for daily rainfall occurrence at Tel Aviv. Q. J. Roy. Meteorol. Soc. **88**(375), 90–95 (1962)
5. C.T. Haan, D.M. Allen, J.O. Street, A Markov chain model of daily rainfall. Water Resour. Res. **12**(3), 443–449 (1976)
6. R.W. Katz, Precipitation as a chain-dependent process. J. Appl. Meteorol. **16**(7), 671–676 (1977)
7. M.Z. Bardaie, A.C.A. Salam, A stochastic model of daily rainfall for Universiti Pertanian Malaysia, Serdang. Pertanika **4**(1), 1–9 (1981)
8. J. Roldan, D.A. Woolhiser, Stochastic daily precipitation models 1: A comparison of occurrence process. Water Resour. Res. **18**(5), 1451–1459 (1982)

9. M.J. Small, D.J. Morgan, The relationship between a continuous-time renewal model and a discrete Markov chain model of precipitation occurrence. Water Resour. Res. **22**(10), 1422–1430 (1986)
10. O.D. Jimoh, P. Webster, The optimum order of a Markov chain model for a daily rainfall in Nigeria. J. Hydrol. **185**, 45–69 (1996)
11. T.C. Sharma, Simulation of the Kenyan longest dry and wet spells and the largest rain-sum using a Markov model. J. Hydrol. **178**, 55–67 (1996)
12. S.K. Tan, S.Y. Sia, Synthetic generation of tropical rainfall time series using an event-based method. J. Hydraul. Eng. **2**(2), 83–89 (1997)
13. D.S. Wilks, Multisite generalization of a daily stochastic precipitation generation model. J. Hydrol. **210**, 178–191 (1998)
14. S.M. Deni, A.A. Jemain, K. Ibrahim, Fitting optimum order Markov chain models for daily rainfall occurrences in Peninsular Malaysia. Theor. Appl. Meteorol. **97**, 109–121 (2009)
15. P.A. Jacobs, P.A.W. Lewis, Stationary discrete autoregressive-moving average time series generated by mixtures. J. Time Ser. Anal. **4**(1), 19–36 (1983)
16. P.A. Jacobs, P.A.W. Lewis, Discrete time series generated by mixtures. I: Correlational and runs properties. J. Roy. Stat. Soc. B (Methodol.) **40**(1), 94–105 (1978)
17. T.A. Buishand (ed.), *Stochastic Modelling of Daily Rainfall Sequences* (Veenman & Zonen, Wageningen, 1977)
18. T.A. Buishand, The binary DARMA (1,1) process as a model for wet-dry sequences. Technical Note, Agriculture University, Wageningen, 1978
19. T.J. Chang, M.L. Kavvas, J.W. Delleur, Stochastic daily precipitation modeling and daily streamflow transfer processes. Technical Report No. 146, Purdue University Water Resources Research Center, 1982
20. T.J. Chang, M.L. Kavvas, J.W. Delleur, Modeling of sequences of wet and dry days by binary discrete autoregressive moving average processes. Am. Meteorol. Soc. **23**, 1367–1378 (1984)
21. T.J. Chang, M.L. Kavvas, J.W. Delleur, Daily precipitation modeling by discrete autoregressive moving average processes. Water Resour. Res. **20**(5), 565–580 (1984)
22. J.W. Delleur, T.J. Chang, M.L. Kavvas, Simulation models of sequences of dry and wet days. J. Irrigat. Drain. Eng. **115**(3), 344–357 (1989)
23. K. Cindrić, The statistical analysis of wet and dry spells by binary DARMA (1,1) model in Split, Croatia, in *BALWOIS Conference 2006*, Ohrid, 2006
24. J. Von Neumann, Distribution of the ratio of the mean square successive difference to the variance. Ann. Math. Stat. **12**(4), 367–395 (1941)