DISSERTATION

PROBABILITY STRUCTURE AND RETURN PERIOD CALCULATIONS FOR MULTI-DAY MONSOON RAINFALL EVENTS AT SUBANG, MALAYSIA

Submitted by

Nur Shazwani Muhammad

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2013

Doctoral Committee:

Advisor: Pierre Y. Julien

Larry A. Roesner
Jose D. Salas
Mazdak Arabi
Ellen E. Wohl

ABSTRACT

PROBABILITY STRUCTURE AND RETURN PERIOD CALCULATIONS FOR MULTI-DAY MONSOON RAINFALL EVENTS AT SUBANG, MALAYSIA

Flooding is the most common natural disaster in Malaysia, as a result of heavy rainfall. Malaysia is located in the equatorial zone and experiences a tropical climate with two seasons classified as the Northeast (November to May) and Southwest (May to September) monsoons. Both monsoons bring moisture, and multi-day rainfall events that cause particularly devastating floods on large watersheds.

The objectives of this study are the following: (1) examine the probability structure of multi-day rainfall events; (2) determine the most suitable distribution function to represent the multi-day rainfall amounts; (3) select the most appropriate model to simulate the sequence of daily rainfall using the discrete autoregressive family models; and (4) develop and test an approach to calculate the return period of multi-day rainfall events with respect to the duration and amount. Daily monsoon rainfall data recorded at Subang Airport are gathered from the Malaysian Meteorological Department. Subang Airport is located near Kuala Lumpur (the capital city of Malaysia) and has a long and reliable daily rainfall record, with 18,993 daily measurements from 1960 to 2011.

The majority of wet and dry events at Subang Airport from 1960 to 2011 are multi-days, with the fraction of 57% and 51%, respectively. The analysis of conditional probabilities for t-consecutive wet and dry days shows that the probability of

occurrence for multi-day wet and dry days is increasing as the event duration increases. For example, the probability of rain on any random day is 0.53; and the conditional probability of rain the second day increases to 0.63. Also, the probability of dry on any random day is 0.47; and the probability of the second dry day increases to 0.58. The probability of rain and dry days increases gradually with rainfall duration. This finding shows that the occurrence of rain and dry is time-dependent.

The autocorrelation coefficient for the daily rainfall amounts is very low at 0.0283. It is concluded that this parameter is independent from one day to another.

The two parameter gamma function is most suitable to fit the daily rainfall precipitation data and the cumulative rainfall from t-consecutive rainy days up to 6 days. A graphical method, i.e. the 1:1 plot confirms the goodness-of-fit of the gamma function.

Two discrete autoregressive models are tested in this study, i.e., the low order Discrete Auto Regressive [DAR(1)] and the low order Discrete Auto Regressive and Moving Average [DARMA(1,1)]. These models require data stationarity, therefore the analysis is done separately for the Northeast and Southwest monsoons. The model selection is based on the four-step process suggested by Salas and Pielke (2003). The comparisons between the observed and calculated autocorrelation coefficient and the low sum of squared errors for the probability distributions confirm that DARMA(1,1) is most suitable to simulate daily rainfall sequences at Subang Airport for both monsoons.

The return period for 1-day and multi-day rainfall events is defined as a function of wet run length and rainfall amount. A test of return period calculations up to 20 years based on the mean wet and dry run lengths shows good agreement between calculation and observations of multi-day rainfall amounts up to 150 mm. A very long sequence of daily rainfall (1,000,000 days) is generated to extend the analysis of multi-day events with cumulative rainfall up to 350 mm, which gives an estimated return period of more than 2,000 years. The mean, standard deviation, maximum daily rainfall, lag-1 ACF coefficient and maximum wet and dry run lengths of the generated daily rainfall sequence using DARMA(1,1) are also comparable with the observed data.

The December 2006 rainstorm event at Kota Tinggi, Johor is used as an example of the application of the algorithms developed in this study. This multi-day rainstorm totaling 350 mm caused devastating floods in the area. The December 2006 rainstorm is extremely rare because the cumulative rainfall amount from the multi-day event gives an estimated return period of greater than 2,000 years. The method proposed in this study is helpful for the design of levees on large watersheds (size of more than 1,000 km^2) because multi-day rainstorms are the main cause of flooding to the area. For example, the return period to overtop the current levee at Kota Tinggi is 220 years when considering a 1-day rainstorm, but this period of return decreases to 24 years when considering 4-day rainstorms.

ACKNOWLEDGEMENTS

This dissertation would have never been completed without the guidance of my advisor and committee members, help from friends, and support from my family, husband and the Government of Malaysia.

I would like to express my deepest gratitude to my advisor, Dr. Pierre Julien, for his excellent guidance, encouragement, and providing me with an excellent atmosphere for doing research. I would like to thank Dr. Mazdak Arabi, Dr. Larry Roesner, Dr. Jose Salas and Dr. Ellen Wohl for their services as my committee members and helping me to develop the scope of my research.

I am indebted to my friends and the members of Dr. Julien's Dream Team. This research would not have been possible without their help. Additionally, the comments and suggestions from Dr. Julien and the Dream Team during Friday afternoon seminars have been very helpful in developing my presentation skills.

I would like to thank my parents, siblings, nephews and niece for their prayers and best wishes. Special thanks to the Delap family, from whom I learned a lot about the American family and introducing us to a whole new culture. This dissertation is made possible with the undivided support and encouragement from my husband, Jazuri Abdullah. He is always there cheering me up and stood by me through the good times and bad. We did it!

A special thanks also goes to my employer, the National University of Malaysia (UKM) for supporting my ambition to complete the doctoral degree in Colorado State

University. Additionally, I would like to acknowledge the continuous support from the Deputy Vice Chancellor (Student and Alumni Affairs) of UKM, Prof. Dato' Ir. Dr. Othman A. Karim, the Dean of the Faculty of Engineering and Built Environment UKM, Prof. Ir. Dr. Mohd. Marzuki Mustafa, the Head of the Department of Civil and Structural Engineering UKM, Prof. Dr. Mohd Raihan Taha, the Head of Human Resource Development Division, Puan Normah Adam and my colleagues in UKM.

Last but not least, I would like to acknowledge the financial support from the Government of Malaysia, thru the Ministry of Higher Education.

Nur Shazwani Muhammad

Fort Collins, CO

July 26, 2013

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Malaysia is located in the equatorial zone and experiences a tropical climate with two major seasons classified as the North East (NE) and South West (SW) monsoons. Both monsoons bring lots of moisture and, as a result, Malaysia receives between 2000 to 4000 mm of rainfall with 150 to 200 rainy days annually (Suhaila and Jemain 2007). Multi-day rainfall events are common in the area and cause particularly devastating floods on large watersheds.

This study focuses mainly on the analysis on multi-day rainfalls, particularly on the probability structure and also the amount of rainfall resulting from such events. Understanding the probability structure of multi-day events leads to the selection of the best suited model to simulate the sequences of daily rainfall. Additionally, the method to estimate the return periods of multi-day rainfall events will also be discussed.

This chapter discusses the general information on Malaysian weather, which includes the descriptions of the mechanism of NE and SW monsoons. The motivation of study, objectives and chapter outlines are also given in the following sections.

1.1 GENERAL INFORMATION ON MALAYSIAN WEATHER

Malaysia is exposed to two monsoon seasons, which occur for about 10 months every year. The Malaysian Meteorological Department (2010) classifies the North East

(NE) monsoon between November to March, while the South West (SW) monsoon occurs from May to September. The transition between the NE and SW monsoon (and vice versa) in the months of April and October is referred to as the intermonsoon season, which occurs for about four to seven weeks (Morgan and Valencia 1983; Saadon et al. 1999). Figure 1.1 gives a graphical reference of monsoon seasons in Peninsular Malaysia. The mechanisms of NE and SW monsoons are also given in this section.

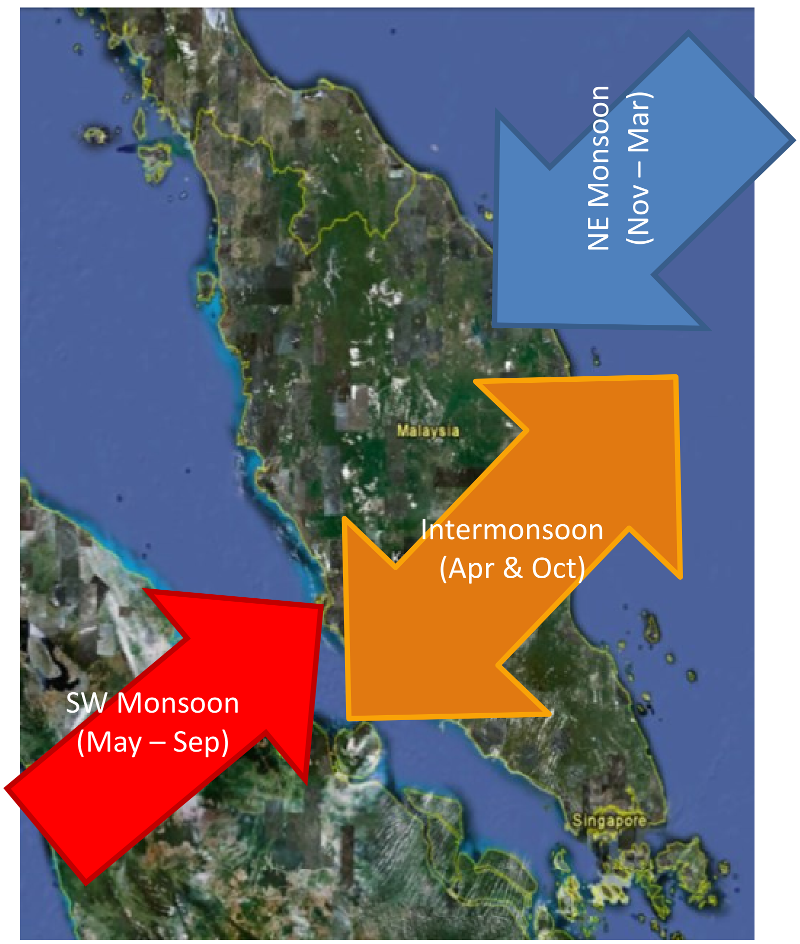

Figure 1.1 Monsoon seasons in Peninsular Malaysia

Figure 1.2 shows that Earth orbits the sun in a counter clockwise direction. From May to June (northern hemisphere summer months), the land mass in the region warms rapidly as compared to the water body (ocean). Higher temperature on the land mass causes warm air to rise, resulting in a low pressure system on the land mass. On the other hand, the water body (ocean) is relatively cool, therefore the cool air falls and causes a high pressure system on the water body. This creates a difference in pressure between the land mass and the water body, which in turn dictates the wind direction. Therefore, during this season, the prevailing winds blows from the SW direction, as shown by the red arrows in Figure 1.2 (Saadon et al. 1999; NAHRIM 2008; Lau 1997).

During the northern hemisphere winter months, i.e., November to March, the monsoon changes direction due to the difference in temperature between the land mass and water body. The land mass becomes relatively colder than the water body. Low temperature on the land mass causes the high pressure. The water body (ocean) is relatively warmer than the land mass, resulting in low pressure on the water body system. Figure 1.3 shows the direction of wind during the NE monsoon season. The cold surges result in prevailing winds in the NE direction (Ngai 1995; Lau 1997).

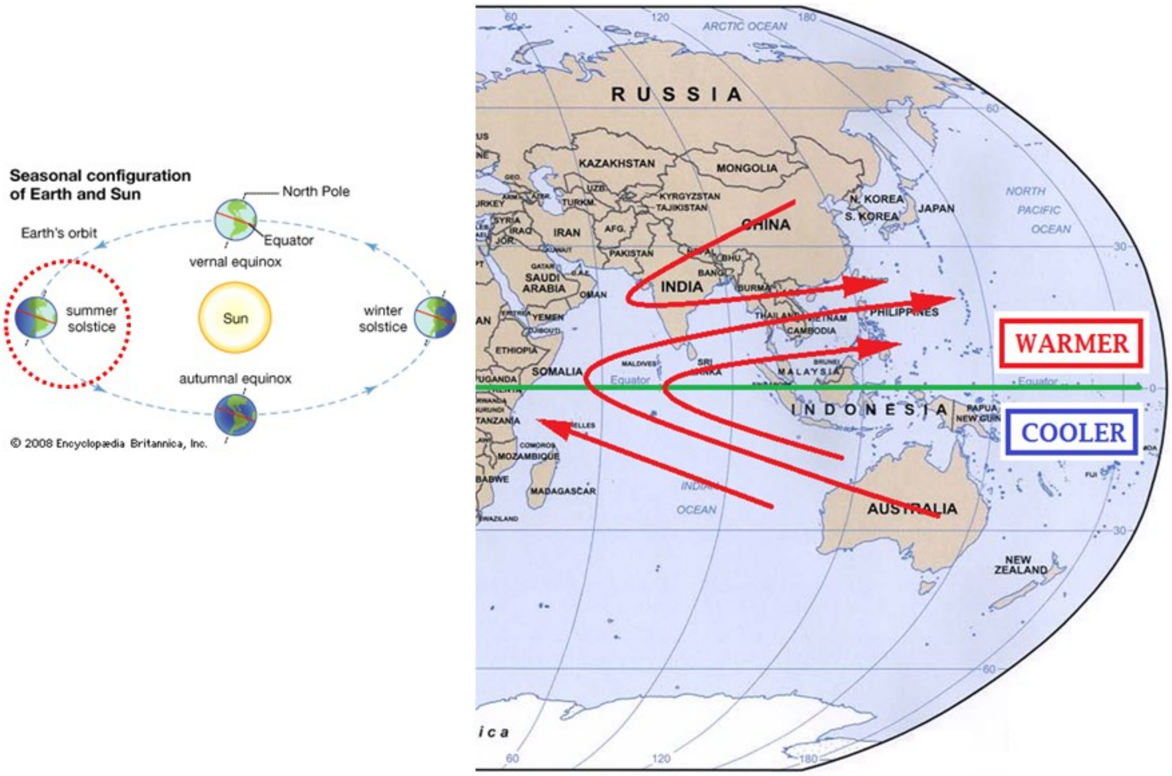

Figure 1.2 Mechanism of SW Monsoon (modified from Wang 2006)

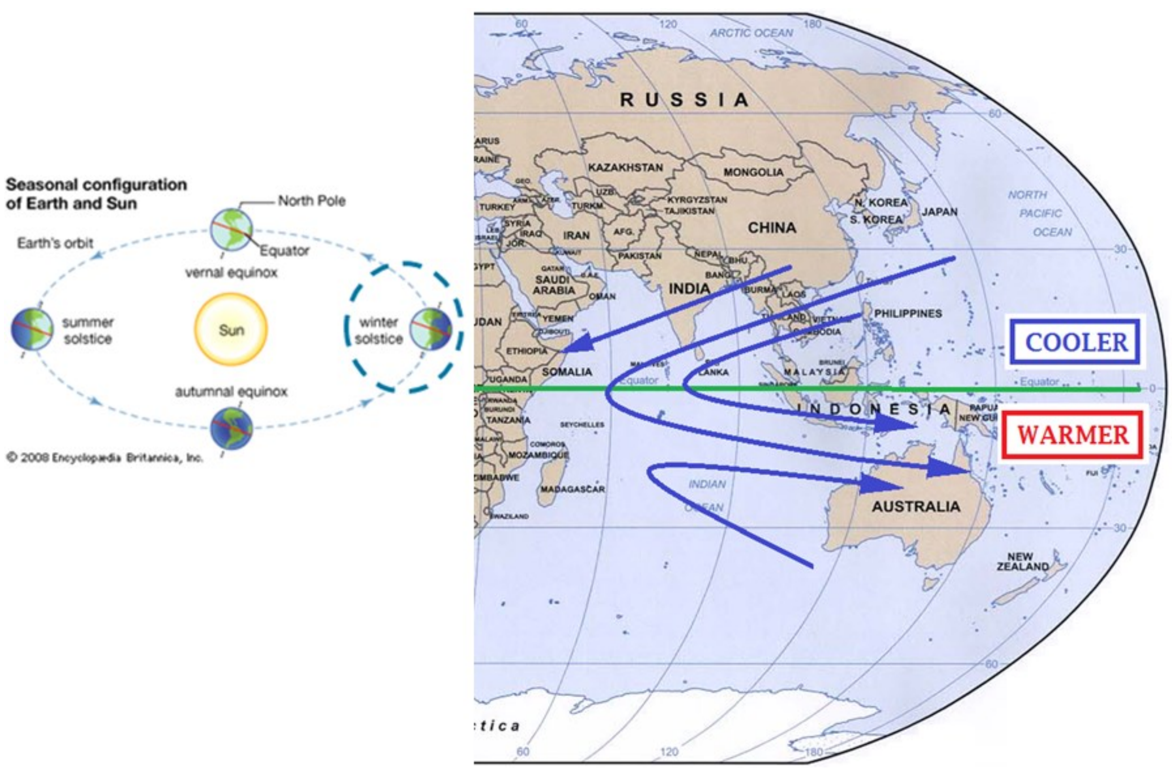

Figure 1.3 Mechanism of NE Monsoon (modified from Wang 2006)

1.2 MOTIVATION

In Malaysia, multi-day rainfall events, especially common during monsoon seasons, are the main causes of flooding (Ngai 1995). There are more examples of the occurrence multi-day rainfall events in other parts of Malaysia (Table 1.1).

Table 1.1 Examples of multi-day rainfall events in Peninsular Malaysia (NAHRIM 2008)

Rainfall Station	Total maximum amount recorded during the multi-day rainfall (mm)			
	2-day	3-day	5-day	7-day
Jasin, Melaka	263.0	276.7	283.0	298.1
Rubber Research Institute of Malaysia, Selangor	225.9	252.0	291.2	293.4
Gua Musang, Kelantan	325.5	373.0	416.5	419.5
Bayan Lepas, Penang	316.4	339.2	375.0	404.6
Kuala Tahan National Park, Pahang	243.8	282.6	309.7	337.2
Kota Tinggi, Johor	922.0	1113.0	1511.0	1722.0

Understanding the probability structure of multi-day rainfall events is extremely important in order to select appropriate rainfall precipitation models. The multi-day rainfalls are time-dependent events, and thus require that the analyses of this stochastic process be done using autoregressive models. This study utilizes the discrete autoregressive models in order to generate the sequence of daily rainfall.

The Discrete Auto Regressive of order 1 [DAR(1)] model is often used to generate the sequence of daily rainfall, under the assumption that the events are time dependent. This model is also known as the first order Markov Chain and assumes that the probability of rain depends only on the current state (rain or dry) and will not be influenced by its past behavior. The model is easy to use, but it lacks long-term persistence. Therefore, it may not be adequate to simulate the long sequence of daily rainfall in tropical and monsoon-affected areas.

Buishand (1978) recommended the low order Discrete Auto Regressive and Moving Average model, which is also known as DARMA(1,1), be used in simulating daily rainfall sequences in tropical and monsoon areas. He found that the DARMA(1,1) model has a long-term persistence, thus it can overcome the problem represented by the first order Markov Chain.

Return periods are usually used in hydrology to measure the severity of an event. This study takes into account the duration, as well as the amount of rainfall to be expected from multi-day events. The joint probability of rainfall amount and duration is utilized to quantify the return period.

This study is intended to enhance the current knowledge of the probability structure and occurrence of multi-day rainfall events caused by tropical monsoons and also the return periods related to them. The findings from this study are important in order to improve the predictability of multi-day rainfall events. There have been a few attempts to use the DARMA(1,1) model in India and Indonesia (Buishand 1978), but the model has not been tested in Malaysia. The determination of return periods of such events may help authorities and engineers quantify the severity of such events.

Additionally, the model proposed in this study may also assist in future planning, including flood warning and evacuation. The methods and results from this study may also help researchers in other monsoon-affected countries, such as India and Pakistan, in managing multi-day rainfall events.

1.3 OBJECTIVES

This study examines the probability structure, generating the sequence of daily rainfall using the discrete autoregressive model and also evaluating the severity of the multi-day events using the concept of return period. The main objectives of this study are to:

1. Examine the probability structure of multi-day rainfall events for tropical monsoons. The daily rainfall data at Subang Airport from 1960 to 2011 are used to calculate the conditional probability of the multi-day rainfall events.
2. Find the most suitable distribution function and give an analytical expression of the rainfall amounts to represent the daily record at Subang Airport.
3. Select the most suitable model to simulate the sequence of daily rainfall using the discrete autoregressive models, i.e., the DAR(1) and DARMA(1,1). The statistics of the generated daily rainfall sequence are compared with the original data in order to evaluate the capability of the model to replicate the observed values.
4. Develop and test an approach to calculate the return period of multi-day rainfall events with respect to rainfall duration and amount. The approach suggested by Shiau and Shen (2001), Salas et al. (2005) and Cancelliere and Salas (2010) is examined in order to calculate the return period for a specific wet run length and rainfall amount, using the conditional probability of both properties.

1.4 CHAPTERS OUTLINE

Various topics which are directly related to the objectives of this study are discussed in the remaining chapters of this report. Chapter 2 gives the details on the related topics pertaining to the determination of threshold to define a wet day, autoregressive models, rainfall amount and return periods.

Chapter 3 discusses the analysis of the definition of wet and dry days and the daily rainfall statistics from Subang Airport. This study uses a long and reliable rainfall record, i.e., from 1960 – 2010 provided by the Department of Meteorology, Malaysia. This chapter gives details pertaining to the annual, monthly and daily statistics of the study area. Additionally, the probability structure of the study area, as well as the distribution function that is suitable to represent the daily rainfall pattern, are also discussed in this chapter.

Chapter 4 details the methods to select the best suited model to generate the sequences of daily rainfall at Subang Airport. Two discrete auto-regressive model are selected, i.e., the DAR(1) and DARMA(1,1). The four-step model selection procedure, i.e., model identification, model estimation, model selection and model verification suggested by Salas and Pielke (2003) is used in this study.

The procedures to simulate the occurrence of daily rainfall as a sequence of binary time series are given in Chapter 5. This step leads to the generation of rainfall amount, the comparison of relevant statistics between observed and simulated data and finally the calculation of return periods.

Chapter 6 provides the details of the application of return period calculations. The analysis concentrates on the most recent rainstorms in the state of Johor, i.e., the Kota Tinggi flood event in December 2006. The estimation of return periods are based on the flood thresholds determined using hydrological modeling by Abdullah (2013). Chapter 7 summarizes the major findings and conclusions of this study.

CHAPTER 2

LITERATURE REVIEW

This chapter discusses the concept and theories that are related to achieve the objectives of this study. The topics included in the section are (1) the method to determine the definition of a rainy day; (2) autoregressive models; (3) distribution functions to represent the observed rainfall amount for a study area; and (4) return period.

2.1 THRESHOLD OF RAINFALL

The threshold (δ in mm) of rainfall is important in determining the occurrence of daily rainfall. A dry state is defined as a day which receives rainfall below a certain threshold value, δ (mm). Buishand (1977) stated that an overestimation of δ gives a bad approximation of the real rainfall process. On the other hand, if δ is underestimated, the daily rainfall sequence may not be homogeneous.

Buishand (1977 and 1978) used the Von Neumann ratio to measure the homogeneity of rainfall data at various locations in the Netherlands, India, Indonesia and Surinam. The analysis was done based on the total annual rainfall and total annual wet days.

Von Neumann (1941) measured homogeneity of a time series based on the ratio of the mean square successive (year to year) difference to the variance. The formulation of Von Neumann ratio (N) is given in Eq. 2.1.

$$N = \frac{\sum_{t=1}^{n-1}(Y_t - Y_{t+1})^2}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}; \quad t = 1, \dots, n \qquad \text{(Eq. 2.1)}$$

Where Y_t , Y_{t+1} = the annual series to be tested;

$\bar{Y}$ = the mean of annual series

The value of N is expected to be 2 if the time series is homogeneous. When N is smaller than 2, it indicates that the sample contains a break. On the other hand, N larger than 2 shows that there are rapid variations in the sample (Bingham and Nelson 1981). The critical values of N can be found in Owen (1962) for N ≤ 50 and Buishand (1981) for N = 70 and N = 100. Table 2.1 summarizes the critical values of N.

Table 2.1 Critical Values of N for the Von Neumann ratio test at 1% and 5%

n	**20**	**30**	**40**	**50**	**70**	**100**
1%	1.04	1.20	1.29	1.36	1.45	1.54
5%	1.30	1.42	1.49	1.54	1.61	1.67

2.2 MARKOV CHAIN

Gabriel and Neumann (1962) successfully developed a Markov Chain model with stationary transitional probabilities for the occurrence of daily rainfall at Tel Aviv for the mid-winter season. The Markov Chain is intended to be a simple model by requiring only two parameters and fit various aspects of the rainfall occurrence pattern. The assumptions in this model are that the probability of a rainy (or dry) day depends only on whether it has rained (or not) the previous day; and the probability of rain (or dry) is assumed to be independent of the preceding days. These probabilities are also known as transitional probabilities, denoted by p_{11} and p_{00} for the sequence of two-rainy days and two-consecutive dry days, respectively. The estimation of p_{11} and p_{00} is by direct counting methods from the available rainfall record. The formula is given in Eq. 2.2 and 2.3.

$$p_{11} = P(X_t = 1|X_{t-1} = 1) = \frac{P(X_{t-1} = 1,\ X_t = 1)}{P(X_{t-1} = 1)} \qquad \text{(Eq. 2.2)}$$

$$p_{00} = P(X_t = 0|X_{t-1} = 0) = \frac{P(X_{t-1} = 0, X_t = 0\ )}{P(X_{t-1} = 0)} \qquad \text{(Eq. 2.3)}$$

Where t = time in days, i.e. 1, 2, …

Gabriel and Neumann (1962) found that if the Markov Chain model is correct, then the geometric distribution represents the probability of occurrence of t-consecutive rainy or dry days, as shown in Eq. 2.4 and 2.5, respectively.

$$P(t - \text{consecutive rainy days}) = p_{11}^{(t-1)}(1 - p_{11}) \qquad \text{(Eq. 2.4)}$$

$$P(t - \text{consecutive dry days}) = p_{00}^{(t-1)}(1 - p_{00}) \qquad \text{(Eq. 2.5)}$$

Where t = time in days ($t \geq 2$).

Since then, the Markov Chain has been widely used in hydrology and meteorological sciences. Richardson and Wright (1984), Hess et al. (1989), Katz (1996) and Baigorria and Jones (2010) reported the use of Markov Chain model to generate weather data.

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 rainfall data from 1951 to 1996 from 25 stations in New York State, USA to simulate the occurrence of daily rainfall using first order Markov Chains. Several statistical tests that were done indicate that the simulated rainfall data match the rainfall data really well. Among the statistical properties are the joint probabilities for both rainy and dry days, mean monthly rainfall and standard deviations of monthly rainfall. Therefore, it was concluded that the model was successful in preserving the dependence nature of daily rainfall at these stations.

Bardaie and Abdul Salam (1981) 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 eleven 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 re-produce the daily rainfall of any length in the area. However, the synthetic daily rainfall was generated for a period of one year only. This research did not indicate if the first order Markov Chain is able to simulate long daily rainfall sequences.

First order Markov Chains are simple and do not require a lot of computational effort. However, research articles in the literature concluded that first order Markov Chains are inadequate to model the sequence of daily rainfall. Feyerherm and Bark (1965) found that first order Markov Chains are unable to describe the random behavior of daily rainfall sequence at six weather stations in the north-central region of the United States of America (USA), but can be used to provide a good approximation. They suggested a higher order of Markov Chain be tested. Farmer and Homeyer (1974) gathered the summer rainfall record from the Wasatch Mountain Range in Utah to compare the probability of occurrence between the measured data and estimation using a simple Markov Chain model. Their study limits the number of consecutive dry days to less than or equal to 30. Their analyses found that the Markov chain model underestimates the probability of occurrence, especially during a long dry-day sequence. They concluded that this result was observed as a result of a strong dry day persistence and that first order Markov Chains are unable to model this phenomenon. The same conclusions were found by Wallis and Griffiths (1995) and Semenov et al. (1998). Another study by Wan et al. (2005) concluded that a modified first-order Markov Chain is more suitable to simulate the Canadian rainfall sequence.

The order of a Markov Chain may be influenced by seasonal change and location (Chin 1977; Cazacioc and Cipu 2004; Deni et al. 2009a). 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 all over the USA. He found that high order Markov Chains are suitable to model the sequence of daily rainfall during winter at most stations, and that first order Markov Chains are appropriate for summer. He also argued that physical environmental causes and geography can influence the order of Markov Chains. The same findings were reported by Cazacioc and Cipu (2004) for the simulation of rainfall sequence at several stations in Romania.

A different approach was used in the analysis of Malaysian daily rainfall data using the Markov Chain model by Deni et al. (2009a). The objective of their study was to find 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. The Akaike's (AIC) and Bayesian information criteria were used to determine the appropriate order. This study uses the available data from 18 rainfall stations located in various parts of Peninsular Malaysia. They also concluded that the optimum order of a Markov Chain varies with the location, monsoon seasons and the level of threshold. For example, the occurrence of rainfall (threshold level 10.0 mm) for 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 model. Additionally, Markov Chain models of higher order 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 maybe used to overcome the problems presented by the first order, more parameters have to be used, which increases the model uncertainty (Jacobs and Lewis 1983) and also makes the calculations more complex.

2.3 AUTOREGRESSIVE MODELS

Autoregressive models (AR) have been used in hydrological research to describe the dependency of hydrological phenomena such as streamflow and rainfall. There are four main categories of AR models in time series modeling, namely Autoregressive model (AR), Autoregressive Moving Average (ARMA), Discrete Autoregressive (DAR) and Discrete Autoregressive Moving Average (DARMA). The model selection depends on the time scale and the persistence of hydrological data, i.e., long or short memory required to preserve the statistics of observed data. ARMA family models, which includes the AR model, are suitable for modeling continuous hydrological processes, rainfall-runoff relationships (Spolia and Chander 1974; Weeks and Boughton 1987; Hsu et al. 1995) and streamflow at various scales, such as annual (Salas and Obeysekera 1982; Vogel et al. 1998; Stedinger 1985; Kendall and Dracup 1991), monthly (Hirsh 1979; Fernando and Jayawardena 1994; Mujumdar and Kumar 1990; Wu et al. 2009) and daily (Kuo and Sun 1996; Yurekli and Ozturk 2003; Wang et al. 2005; Greco 2012). Other applications of the AR family models can be found in Salas et al. (1980) and Marivoet (1983).

DAR(p) and DARMA(p,q) models are suitable for generating binary sequences of time series such as daily rainfall. Among the challenges in modeling daily rainfall is the sequence of dry and wet days, which include zeros for dry days and ones for wet days (Tan and Sia 1997; Detzel and Mine 2011). Previous work and the basic properties of DAR and DARMA models are discussed in the following sections.

2.3.1 DAR (p)

Jacobs and Lewis (1978a) discuss the model that has the correlation structure of an autoregressive process of order p. The mechanism to model the binary sequence of daily rainfall using first order Markov Chain is also referred to as the DAR(1) model (Buishand 1978; Chang et al. 1984a; Delleur et al. 1989).

Evora and Rousselle (2000) have successfully use the DAR(1) model to simulate the sequences of daily rainfall at Bakel Station located near the Senegal River. The simulated daily rainfall was then used as an input to generate the daily flows simulation using the hybrid of DAR-AR and DAR-GAR (Gamma Autoregressive) stochastic models.

The equation for DAR(1) model is given in Eq. 2.6 below ;

$$A_t = V_t A_{t-1} + (1 - V_t) Y_t \qquad \text{(Eq. 2.6)}$$

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

Where: V_t = the independent sequences of independent random variables taking values of 0 (dry day) and 1 (wet day) only and has the probability of λ, i.e.,

$$P(V_t = 1) = \lambda = 1 - P(V_t = 0) \qquad \text{(Eq. 2.7)}$$

Where: Y_t = a sequence of identically and independent distributed (i.i.d.) random variable, with a common probability of $\pi_k = P(Y_t = k), \quad k = 0,1.$

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 theoretical autocorrelation function of DAR(1) model is (Jacobs and Lewis 1978a)

$$\text{corr}(A_t, A_{t-k}) = r_k(A) = \lambda^k, \quad k \geq 1 \qquad \text{(Eq. 2.8)}$$

Where: r_k = the autocorrelation function of lag k-th day

The empirical autocorrelation function for the daily rainfall dataset is calculated based on the sequences of dry and rainy days, i.e., 0s and 1s, and not the rainfall amounts (Delleur et al. 1989). The formula is given in Eq. 2.9.

$$r_k = \left[\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})\right]\left[\sum_{t=1}^{N}(x_t - \bar{x})^2\right]^{-1} \quad \text{(Eq. 2.9)}$$

$$\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t \quad \text{(Eq. 2.10)}$$

Where: x_t = either dry or rainy day (0 or 1, respectively)

N = the number of samples

There are two parameters for DAR(1) model, i.e., π_0 or π_1 and λ. The lag-1 autocorrelation coefficient is the estimator for λ, as given in Eq. 2.8. The parameter π_0 or π_1 are based on the run length property calculated from the observed daily rainfall dataset, and the formulas are given in Eq. 2.11 and 2.12.

$$\pi_0 = \frac{\overline{T_0}}{\overline{T_0} + \overline{T_1}} \quad \text{(Eq. 2.11)}$$

$$\pi_1 = 1 - \pi_0 \quad \text{(Eq. 2.12)}$$

Where: $\overline{T_0}$ = the mean run length for dry days (state 0)

$\overline{T_1}$ = the mean run length for wet days (state 1)

One-step transitional probability, $p(i,j) = P(A_{t+1} = j|A_t = i)$ is given by (Jacobs and Lewis 1978a).

$$p(i,j) = \begin{cases} \lambda + (1-\lambda)\pi_j, & \text{if } i = j \\ (1-\lambda)\pi_j, & \text{if } i \neq j \end{cases} \quad \text{(Eq. 2.13)}$$

Eq. 2.13 can also be represented in terms of transitional probability matrix, as shown in Eq. 2.14.

$$P = \begin{bmatrix} \lambda + (1-\lambda)\pi_0 & (1-\lambda)\pi_1 \\ (1-\lambda)\pi_0 & \lambda + (1-\lambda)\pi_1 \end{bmatrix} \qquad \text{(Eq. 2.14)}$$

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. Eq. 2.15 and 2.16 show the general mathematical representation of wet and dry run lengths.

$$(T_1 = n) = (X_0 = 0, X_1 = 1, \dots, X_n = 1, X_{n+1} = 0 \;\text{ given that } X_0 = 0, X_1 = 1) \qquad \text{(Eq. 2.15)}$$

$$(T_0 = n) = (X_0 = 1, X_1 = 0, \dots, X_n = 0, X_{n+1} = 1 \;\text{ given that } X_0 = 1, X_1 = 0) \qquad \text{(Eq. 2.16)}$$

For DAR(1) model, the probability distribution of wet and dry run lengths can be obtained from Eq. 2.17 and 2.18 (Chang et al. 1984a).

$$P(T_1 = n) = p^{n-1}(1,1)[1 - p(1,1)] \qquad \text{(Eq. 2.17)}$$

$$P(T_0 = n) = p^{n-1}(0,0)[1 - p(0,0)] \qquad \text{(Eq. 2.18)}$$

2.3.2 DARMA (p,q)

DARMA is the discrete process of ARMA. The structure of the DARMA model makes it suitable to be used in modeling the sequence of wet and dry days. Jacobs and Lewis (1977, 1978b, 1978c and 1983) introduced the concept of the DARMA model, which is intended to be a simple tool to model stationary sequences of dependent discrete random variables with specified marginal distribution and correlation structure.

Several studies reported in literature have used the DARMA family models to generate the sequence of daily rainfall (Buishand 1977 and 1978; Chang et al. 1982, 1984a and 1984b; Delleur et al. 1989; Cindrić 2006).

Buishand (1977 and 1978) modeled the sequence of daily rainfall using DARMA(1,1) at several stations in the Netherlands, Suriname, India and Indonesia. DARMA(1,1) is a stationary model, therefore the data in each station were divided into their respective seasons in order to consider the seasonal variations. It is assumed that during a specific season the data may be taken as stationary (Kedem 1980). The results have shown that DARMA(1,1) was successful in simulating the daily rainfall in tropical and monsoon areas, where the long-term persistence model is needed.

Other applications of DARMA include the analysis of drought using annual streamflow (Chung 1999; Chung and Salas 2000; Cancelliere and Salas 2010). Chung and Salas (2000) concluded that the DARMA(1,1) model is suitable to generate the drought occurrence of the Niger River in Africa. This study defined a drought when the annual streamflow of a river is less than the mean annual streamflow of a river. From the results it was shown that there were long periods of drought and high flows. This finding was consistent with the behavior of DARMA(1,1), i.e., it is suitable to model long events and has a longer memory as compared to DAR(1).

Chang et al. (1987) used the DARMA(1,1) to generate the sequence of daily rainfall and then extended the model to estimate the daily streamflow using linear transfer function at an Indiana watershed. They name the extended model as Transfer Discrete Autoregressive Moving Average [T-DARMA(p,q,m,n,c,d)]. The model produces satisfactory results in terms of model building procedures, which can be observed from the ability of the model to preserve the auto correlation function. Additionally, the means of observed and generated daily streamflow are almost the same, and this result shows that water balance is also well preserved. The authors also concluded that T-DARMA is a stationary model, therefore the data should be divided into their respective seasons and the analyses done separately for each season.

The basic properties of the DARMA(1,1) model are given in the following paragraphs. Eq. 2.19 represents the general formula for DARMA(1,1), as shown below:

$$X_t = U_t Y_t + (1 - U_t) A_{t-1} \qquad \text{(Eq. 2.19)}$$

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

And the autoregressive component is given by

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

Where: U_t = the independent sequences of independent random variables take values of 0 (dry day) or 1 (wet day) only and have the probability of β, as given by Eq. 2.20:

$$P(U_t = 1) = \beta = 1 - P(U_t = 0) \qquad \text{(Eq. 2.20)}$$

Where: Y_t = a sequence of identically and independent distributed (i.i.d.) random variable having a common probability of

$\pi_k = P(Y_t = k)$,

k = 0 and 1

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 theoretical autocorrelation function of DARMA(1,1) model is

$$\text{corr}(X_t, X_{t-k}) = r_k(X) = c\lambda^{k-1}, \quad k \geq 1 \qquad \text{(Eq. 2.21)}$$

Where : c $= (1-\beta)(\beta + \lambda - 2\lambda\beta)$ (Eq. 2.22)

The coefficient c can be estimated as the lag-1 autocorrelation function coefficient of the DARMA(1,1) model. It should be noted that the DARMA model has the same correlation structure as ARMA (Buishand 1978; Jacobs and Lewis, 1983).

Three parameters in DARMA(1,1) need to be estimated, namely π_0 or π_1 , λ, and β. These parameters are always positive and less than 1, and π_0 or π_1 is estimated using Eq. 2.11 and 2.12. Buishand (1978) uses the ratio of second to the first autocorrelation function as an estimator for λ, as shown in Eq. 2.23.

$$\hat{\lambda} = \frac{r_2}{r_1} \qquad \text{(Eq. 2.23)}$$

Where: r_2 = the second autocorrelation coefficient

r_1 = the first autocorrelation coefficient

The estimate of λ is calculated by minimizing Eq. 2.24 using the Newton-Raphson iteration techniques and Eq. 2.23 is used as the initial estimate.

$$\phi(\lambda) = \sum_{k=1}^{M} \left[r_k - c\lambda^{k-1}\right]^2; \quad k \geq 1 \qquad \text{(Eq. 2.24)}$$

Where: c = can be estimated as the lag-1 autocorrelation function coefficient of the DARMA(1,1) model, as shown in Eq. 2.21.

After solving equation (2.24) for $\hat{\lambda}$, β is estimated using Eq. 2.25

$$\hat{\beta} = \frac{(3\hat{\lambda}-1) \pm \sqrt{(3\hat{\lambda}-1)^2 - 4(2\hat{\lambda}-1)(\hat{\lambda}-\hat{c})}}{2(2\hat{\lambda}-1)} \qquad \text{(Eq. 2.25)}$$

Alternatively, Eq. 2.24 can also be minimized using the Marquardt (1963) method for nonlinear equations (Chang et al. 1984a; Delleur et al. 1989). The estimator for λ is suggested as

$$\hat{\lambda} = \frac{r_2 + r_3 + r_4}{r_1 + r_2 + r_3} \qquad \text{(Eq. 2.26)}$$

Where: r_1 = the first autocorrelation coefficient

r_2 = the second autocorrelation coefficient

r_3 = the third autocorrelation coefficient

r_4 = the forth autocorrelation coefficient

As mentioned earlier, (X_t, A_t) forms a first order bivariate Markov Chain, therefore the one step transitional probabilities $H_t(u, v)$ is

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

(Eq. 2.27)

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

The transitional probability matrices are given in Eq. 2.28 and 2.29.

$$H_0 = \begin{bmatrix} \lambda(1-\beta) + [1-\lambda(1-\beta)]\pi_0 & (1-\beta)(1-\lambda)\pi_1 \\ \beta(1-\lambda)\pi_0 & \beta\lambda\pi_0 \end{bmatrix} \qquad \text{(Eq. 2.28)}$$

$$H_1 = \begin{bmatrix} \beta\lambda\pi_1 & \beta(1-\lambda)\pi_1 \\ (1-\beta)(1-\lambda)\pi_0 & \lambda(1-\beta) + [1-\lambda(1-\beta)]\pi_1 \end{bmatrix} \qquad \text{(Eq. 2.29)}$$

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 2.2 summarizes the W_t values.

Table 2.2 Four state Markov Chain, W_t

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

The value of 0 and 1 in 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 in W_t.

The transitional probabilities are given as

$$p_W(0,1) = P(X_{t+1} = 0, A_{t+1} = 1 | X_t = 0, A_t = 0) = P(X_{t+1} = 0, A_{t+1} = 1 | A_t = 0)$$
$$= H_0(0,1) \qquad \text{(Eq. 2.30)}$$

$$p_W(0,2) = P(X_{t+1} = 1, A_{t+1} = 0 | X_t = 0, A_t = 0) = P(X_{t+1} = 1, A_{t+1} = 0 | A_t = 0)$$
$$= H_1(0,0) \qquad \text{(Eq. 2.31)}$$

$$p_W(0,3) = P(X_{t+1} = 1, A_{t+1} = 1 | X_t = 0, A_t = 0) = P(X_{t+1} = 1, A_{t+1} = 1 | A_t = 0)$$
$$= H_1(0,1) \qquad \text{(Eq. 2.32)}$$

$$p_W(1,0) = P(X_{t+1} = 0, A_{t+1} = 0 | X_t = 0, A_t = 1) = P(X_{t+1} = 0, A_{t+1} = 0 | A_t = 1)$$
$$= H_0(1,0) \qquad \text{(Eq. 2.33)}$$

Transitional probability matrix Q of the univariate Markov Chain W_t is

$$Q = \begin{matrix} & 0 & 1 & 2 & 3 \\ 0 & H_0(0,0) & H_0(0,1) & H_1(0,0) & H_1(0,1) \\ 1 & H_0(1,0) & H_0(1,1) & H_1(1,0) & H_1(1,1) \\ 2 & H_0(0,0) & H_0(0,1) & H_1(0,0) & H_1(0,1) \\ 3 & H_0(1,0) & H_0(1,1) & H_1(1,0) & H_1(1,1) \end{matrix} \qquad \text{(Eq. 2.34)}$$

and its marginal distribution is

$$
\begin{aligned}
P[W_i = 0] &= P(X_t = 0, A_t = 0) \\
&= P(X_t = 0, A_t = 0 | A_{t-1} = 0)P(A_{t-1} = 0) \\
&\quad + P(X_t = 0, A_t = 0 | A_{t-1} = 0)P(A_{t-1} = 1) \\
&= H_0(0,0)\pi_0 + H_0(1,0)\pi_1 \qquad \text{(Eq. 2.35)}
\end{aligned}
$$

$$
\begin{aligned}
P[W_i = 1] &= P(X_t = 0, A_t = 1) \\
&= P(X_t = 0, A_t = 1 | A_{t-1} = 0)P(A_{t-1} = 0) \\
&\quad + P(X_t = 0, A_t = 1 | A_{t-1} = 1)P(A_{t-1} = 1) \\
&= H_0(0,1)\pi_0 + H_0(1,1)\pi_1 \qquad \text{(Eq. 2.36)}
\end{aligned}
$$

$$
\begin{aligned}
P[W_i = 2] &= P(X_t = 1, A_t = 0) \\
&= P(X_t = 1, A_t = 0 | A_{t-1} = 0)P(A_{t-1} = 0) \\
&\quad + P(X_t = 1, A_t = 0 | A_{t-1} = 1)P(A_{t-1} = 1) \\
&= H_1(0,0)\pi_0 + H_1(1,0)\pi_1 \qquad \text{(Eq. 2.37)}
\end{aligned}
$$

$$
\begin{aligned}
P[W_i = 3] &= P(X_t = 1, A_t = 1) \\
&= P(X_t = 1, A_t = 1 | A_{t-1} = 0)P(A_{t-1} = 0) \\
&\quad + P(X_t = 1, A_t = 1 | A_{t-1} = 1)P(A_{t-1} = 1) \\
&= H_1(0,1)\pi_0 + H_1(1,1)\pi_1 \qquad \text{(Eq. 2.38)}
\end{aligned}
$$

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

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

$$= \frac{P(X_0 = 0, X_1 = 1, \dots, X_t = 1, X_{t+1} = 0)}{P(X_0 = 0, X_1 = 1)} \qquad \text{(Eq. 2.39)}$$

Note that

$$P(X_0 = 0, X_1 = 1, \dots, X_t = 1, X_{t+1} = 0)$$

$$= \left(P[W_0 = 0]\left[H_1^{(n)}(0) - H_1^{(n+1)}(0)\right] + P[W_0 = 1]\left[H_1^{(n)}(1) - H_1^{(n+1)}(1)\right]\right) \qquad \text{(Eq. 2.40)}$$

Where: $$H_1^{(n)}(j) = H_1^{(n)}(j,0) + H_1^{(n)}(j,1); \quad j = 0,1 \qquad \text{(Eq. 2.41)}$$

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

$$P(X_0 = 0, X_1 = 1) = \sum_{k=0}^{1}\sum_{j=0}^{1} H_1(j,k)\left[\pi_j - \sum_{l=0}^{1} H_1(l,j)\pi_l\right] \qquad \text{(Eq. 2.42)}$$

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

$$P(T_0 = t) = P(X_0 = 1, X_1 = 0, \dots, X_t = 0, X_{t+1} = 1 | X_0 = 1, X_1 = 0\,); \quad t = 1,2, \dots$$

$$= \frac{P(X_0 = 1, X_1 = 0, \dots, X_t = 0, X_{t+1} = 1\,)}{P(X_0 = 1, X_1 = 0)} \qquad \text{(Eq. 2.43)}$$

$$P(X_0 = 1, X_1 = 0, \dots, X_t = 0, X_{t+1} = 1\,)$$

$$= \Big(P[W_0 = 2]\Big[H_0^{(n)}(0) - H_0^{(n+1)}(0)\Big]$$

$$+ P[W_0 = 3]\Big[H_0^{(n)}(1) - H_0^{(n+1)}(1)\Big]\Big) \qquad \text{(Eq. 2.44)}$$

Where: $$H_0^{(n)}(j) = H_0^{(n)}(j,0) + H_0^{(n)}(j,1), \;\; j = 0,1 \qquad \text{(Eq. 2.45)}$$

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

$$P(X_0 = 1, X_1 = 0) = \sum_{k=0}^{1}\sum_{j=0}^{1} H_0(j,k)\left[\pi_j - \sum_{l=0}^{1} H_0(l,j)\pi_l\right] \qquad \text{(Eq. 2.46)}$$

Where: $H_0(j,k), H_0(l,j)$ = elements of the n-step transitional probability matrix

2.4 RAINFALL AMOUNT

The rainfall amount is often associated with the occurrence of daily rainfall. In water resources planning and management, the analyses of rainfall amount and daily rainfall occurrence help engineers to better understand the hydrological behavior of a particular study area.

Delleur et al. (1989) suggested that the daily rainfall amount can be computed using Eq. 2.47.

$$W_t = X_t Z_t \qquad \text{(Eq. 2.47)}$$

Where: X_t = a sequence of daily rainfall generated using DAR(1) or DARMA(1,1) model

Z_t = a sequence of random variables with the suitable distribution function

A number of studies discussing the distribution function of n-consecutive rainy or dry day events were reported in literature. Among these, a few studies reported on Malaysian data (Deni et al. 2008; Deni and Jemain 2009a and 2009b; Deni et al. 2009b; Deni et al. 2010).

Deni et al. (2008) gather the rainfall data from 10 gaging stations which are located all over Peninsular Malaysia. These stations were selected based on the length of record, ranging from 1971 to 2005, and the missing data. Seven probability distributions were tested to determine the best fit function for each station. The suitability of each distribution function was tested using the Chi-squared test. Their research found that, for most stations, the compound geometric distribution and the truncated negative binomial distribution were the best functions to fit the behavior of n-consecutive rainy and dry days, respectively. Deni and Jemain (2009a and 2009b) and Deni et al. (2009b) expanded the research for other stations in Peninsular Malaysia.

These studies also propose the use of mixed probability models, such as two log series distributions, log series Poisson distributions and the combination of log series and geometric distributions. A Chi-squared goodness-of-fit test is used in Deni and Jemain (2009a and 2009b) and Akaike's information criterion was used by Deni et al. (2009b). These researchers found that the best distribution functions varied according to the location of the rainfall gaging stations.

Deni et al. (2010) used a slightly different approach to investigate the most suitable probability models for consecutive dry and rainy days during monsoon seasons. Thirteen distribution functions were tested for 38 rainfall stations located all over Peninsular Malaysia. They concluded that, for most stations, the modified log series and the compound geometric distributions were the best fit functions for n-consecutive dry and rainy days, respectively. The test for goodness-of-fit was done using the Akaike's information criteria and Kolmogorov-Smirnoff test.

Previous studies on fitting the best probability distribution function for maximum rainfall depths in n-consecutive rainy days in a monsoonal climate area were reported in the literature (Upadhyaya and Singh 1998; Ali et al. 2002; Machiwal et al. 2006; Bhakar et al. 2008).

Upadhyaya and Singh (1998) investigated the suitability of eight probability distributions to describe the behavior of maximum rainfall amount for 1- to 6-consecutive rainy days. The daily rainfall data of 42 years from 1950 to 1991 were collected from Orissa University of Agriculture and Technology, Bhubaneswar, India. The authors found that Gringorten's plotting position function is the best fit to represent the maximum rainfall amount in a one-day rainfall event. For 2- to 6-consecutive rainy days, the log extreme value 1 shows the most promising results.

Ali et al. (2002) expanded the research done by Upadhyaya and Singh (1998). The authors used six different stations at the Koraput district, which is a part of the Eastern Ghat High Land Zone of Orissa, India. Their research limits the number of consecutive rainy events to a maximum of 4 days. The two parameter log-normal distribution function was the best fit function for all cases tested in the region and is verified using the standard error method. A regression function was also developed in their study to calculate the amount of maximum rainfall in n-consecutive days.

The probability density functions for one day and 2- to 6-consecutive days of maximum rainfall of Kharagpur, West Bengal in India were examined by Machiwal et al. (2006). A long period of rainfall data was used, i.e., 47 years from 1956 to 2002. Nine different functions were tested. To determine the goodness-of-fit for these functions, they use two different approaches, i.e., the Chi-square and Kolmogorov-Smirnoff methods. Their findings concluded that the best fit function was the Pearson type-V for all cases.

Additionally, a linear regression model was used to generate the equation to calculate the maximum rainfalls for 2- to 6-consecutive rainy days. The relationship was derived from the one-day maximum rainfall.

Bhakar et al. (2008) analyze the daily rainfall collected from the Agricultural Meteorological Observatory, Udaipur, India. A long record was obtained for this site, i.e., 85 years, from 1921 to 2005. The probability distributions used in this study were normal, lognormal and gamma, while the goodness-of-fit test was evaluated using the Chi-square method. They concluded that the gamma distribution is the best fit function for this study area.

2.5 RETURN PERIOD

In hydrology, the term "return period" is used to define the risk of a particular design of a hydraulic structure. There are two different terms of return period reported in literature, i.e., first arrival time and interarrival time or recurrence interval. These terms give different values when the events are dependent in time. On the other hand, for single and independent events, the first arrival time and recurrence interval give the same value (Fernández and Salas 1999a). Most of the completed studies use first order Markov Chain to describe the events as serially dependent. Extensive theories and applications on the return period definitions and serial dependence are discussed in Fernández and Salas (1999a and 1999b).

Woodyer et al. (1972), Kite (1978), Lloyd (1970), Loaiciga and Mariňo (1991) and Şen (1999) define recurrence interval as the average elapsed time between the occurrences of critical events, such as earthquake of high magnitude and extreme floods or drought. The calculation of return period by these authors is for any event that has a value of equal or greater to the critical event.

Lloyd (1970) compared the return periods for river flows using two different conditions, (1) mutually independent, and (2) dependent time series, which was serially correlated using the first order Markov Chain model. His research found that the same return periods were calculated for the independent and dependent time series. Therefore, the method proposed by Lloyd (1970) may not be suitable to estimate the annual streamflow data with strong persistence.

Şen (1999) compares the independent and dependent return periods for short and long records. For short records, there are significant differences in return periods but as the number of records grew, the dependence theory reduces to independent case. This study shows that the length of record may influence the calculation of return period.

Another definition of return period used by other authors (Vogel 1987; Douglas et al. 2002) is the average number of trials required to the first occurrence of a critical event. This definition may be more useful in determining the return period for a reservoir operation because the interest is knowing the first time that the reservoir is at the risk of failure, rather than the average time between failures (Douglas et al. 2002).

Vogel (1987) analyzed the state of operation of a reservoir, and the expected number of years until the first reservoir failure. He used the first order Markov Chain model to represent the time dependency of reservoir operation for the Pacific Northwest Hydroelectric Power System. This study simulated 1,000 sets of 100-year simulations to find the importance of annual streamflow persistence in determining the return period based on the first reservoir failure. The results show that the estimation using the Markov Chain model was appropriate to model the return periods because they matched the simulated data reasonably well. This result also shows that the failures are time dependent.

Douglas et al. (2002) found that the conventional methods to calculate the return period and risk, which assume that the time series are independent, do not take into account the persistence and statistics of annual streamflow. As a result, the return period, risk and the low flow quantiles are often underestimated by the conventional method. This study suggests the method to calculate return period and risk that considers the persistence of annual streamflow and the evaluation of the risk of drought in the United States.

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 and 2010) reported their studies on the calculation of return period and risk that include both the amount (or severity) and duration of the hydrological events.

Shiau and Shen (2001) suggested that the bivariate probability distribution function of drought duration and severity be used in order to describe the conditional distribution of both properties. The relationship is presented in Eq. 2.48.

$$f_{DL,DS}(dl, ds) = f_{DS|DL}(ds|dl)f_{DL}(dl) \qquad \text{(Eq. 2.48)}$$

Where: DL, dl = drought duration

DS, ds = drought severity

$f_{DL,DS}(dl, ds)$ = bivariate probability distribution function of drought duration and severity

$f_{DS|DL}(ds|dl)$ = conditional distribution of drought severity given a drought duration

$f_{DL}(dl)$ = distribution of drought duration.

$f_{DL}(dl)$ is calculated using a simple Markov chain and Loaiciga and Leipnik (1996) showed that the geometric distribution can represent the probability distribution function of drought duration (refer to Eq. 2.49).

$$f_{DL}(dl) = p_{01}(1 - p_{01})^{dl-1} \qquad \text{(Eq. 2.49)}$$

Where: p_{01} = transitional probability of observing given that there is a deficit at time t and a surplus at time t+1

DL, dl = drought duration

The probability of occurrence, denoted by P(E), which includes both the duration and amount of a hydrological event, can be calculated by integrating the conditional distribution ($f_{DS|DL}(ds|dl)$). Then the return periods for the given conditions can be calculated using Eq. 2.50.

$$\text{Return Period,} \quad T(ds|dl) = \left(\frac{p_{01} + p_{10}}{p_{01}p_{10}}\right)\frac{1}{P(E)} \qquad \text{(Eq. 2.50)}$$

Where: ds = drought severity

dl = drought duration

p_{01} = transitional probability of observing given that there is a deficit at time t and a surplus at time t+1

p_{10} = transitional probability of observing given that there is a surplus at time t and a deficit at time t-1

P(E) = the probability of occurrence for an event

Cancelliere and Salas (2010) adapted the same method as Shiau and Shen (2001), but they used the DARMA(1,1) model to calculate the probability distribution of drought length (refer to Eq. 2.39 to 2.46). The formulas for the expected duration of drought and non-drought events are given in Eq. 2.51 and 2.52, respectively.

$$E(L) = \frac{\pi_1[1 - \beta\lambda + \beta(1 - \lambda - \beta + 2\lambda\beta)(1 - \pi_1)]}{(1 - \pi_1)[1 - \lambda(1 - \beta)(1 - \beta\pi_1) - \beta\pi_1\{1 - \beta(1 - \lambda)\}]} \quad \text{(Eq. 2.51)}$$

$$E(Ln) = \frac{\pi_0[1 - \beta\lambda + \beta(1 - \lambda - \beta + 2\lambda\beta)(1 - \pi_0)]}{(1 - \pi_0)[1 - \lambda(1 - \beta)(1 - \beta\pi_0) - \beta\pi_0\{1 - \beta(1 - \lambda)\}]} \quad \text{(Eq. 2.52)}$$

Where: E(L) = expected duration of drought events;

E(Ln) = expected duration of non-drought events;

π_1, π_0, λ and β = the DARMA(1,1) parameters, calculated using Eq. 2.11, 2.12, 2.24 and 2.25, respectively.

The return period for a critical event can be calculated using Eq. 2.53.

$$T = \frac{E(L) + E(Ln)}{P(E)} \qquad \text{(Eq. 2.53)}$$

González and Valdés (2003) use the alternating renewal process to describe the occurrence of droughts, and assume equal distribution and independence of drought and nondrought events when using Eq. 2.53. Cancelliere and Salas (2010) show that Eq. 2.53 gives an excellent approximation for an autocorrelated process.

2.6 SUMMARY

The extensive discussions and literature review given in this chapter show that the study of multi-day rainfall events in Malaysia has received minimal attention from Malaysian researchers. Current studies focus on determining the most appropriate order of Markov Chain. Furthermore, there are no attempts to model the sequence of daily rainfall using the discrete autoregressive models, i.e., DAR(1) and DARMA(1,1) in Malaysia, even though Ngai (1995) shows that multi-day rainfall events are the main causes of floods.

DAR(1) is also known as the first order Markov Chain and is often used in the analysis of daily rainfall sequence. However, this model has a short memory and is unable to simulate the daily rainfall that requires long daily sequences of wet or dry cycles. To overcome the problem, the DARMA(1,1) model can be used. The autocorrelation function for this model decays slowly, which also means that it has long-term persistence.

Rainfall amounts are an important parameter in the analysis of daily rainfall and this is a random variable. A suitable function to represent the probability distributions of rainfall amounts from the observed data needs to be determined for a particular study area. Goodness-of-fit test can be used to determine the suitability of a distribution function.

A return period describes the average waiting time for another event to occur. The bivariate probability distribution function of duration and amount of rainfall can give a quantitative idea of the possibility for a specific event to occur, which leads to the determination of the return period. The return period estimation for multi-day rainfall events is uncommon in the literature. Therefore, this study is intended to explore and fill in the gaps that have not been explored by other researchers in this particular topic.

CHAPTER 3

MULTI-DAY RAINFALL PROBABILITY STRUCTURE AND DISTRIBUTION FUNCTION

This study is motivated by the nature of monsoon rainfall, in which the rainfall duration exceeds one day. The threshold of a rainy day is determined using the Von Neumann ratio test. This chapter discusses the statistics of annual, monthly and daily rainfall of the study area. The statistical dependence of multi-day rainfall events at Subang Airport is determined using the conditional probability structure. Additionally, the distribution functions and statistical dependence of rainfall amount for daily and multi-day rainstorms are also examined.

3.1 STUDY AREA

Subang Airport is located near Kuala Lumpur, the capital city of Malaysia. The coordinate for this rainfall station is 3^{o} 7′ 1.20″ N, 101^{o} 33′ 0.00″ E. Subang Airport was the main gateway to Malaysia via air from the 1960s to 1998. Since 1998, the airport is used for general aviation and commercial services. Based on its history, this site is chosen because of the length, reliability and quality of daily rainfall data that it can provide, i.e., more than 50 years. The location of Subang Airport is shown in Figure 3.1.

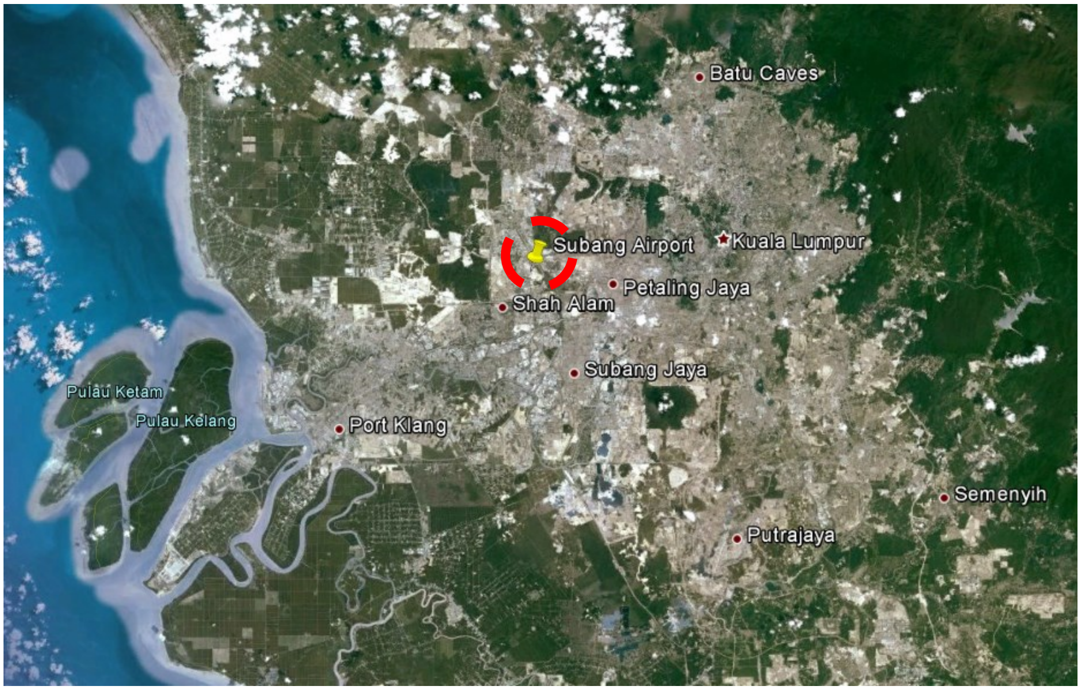

Figure 3.1 Location of Subang Airport and other major cities

3.2 RAINFALL RECORDS AT SUBANG AIRPORT

The daily rainfall data from 1960 to 2011 are provided by the Department of Meteorology, Malaysia. Rain is collected using the tipping bucket method, and the amount is recorded from 8 in the morning until 8 a.m the next day. Constant monitoring of rainfall is done at the Subang Airport because this is a high priority meteorological station. Hence, the dataset is complete, with a total of 18,993 daily measurements.

3.2.1 DETERMINING THE THRESHOLD OF A RAINY DAY

In this study, the definition of rain (or wet) is a day in which the rainfall exceeds a certain threshold, δ (mm). The threshold of rain is determined by the assessment for homogeneity. It is well known that the location of Subang Airport does not change over the years. Therefore, the Von Neumann ratio test is chosen because it is not location specific and does not yield information based on the year of the break, which is unnecessary for this study. Additionally, homogeneity is measured based on the ratio of the mean square successive (year to year) difference from the variance. This method ensures that homogeneity determined from the measurements is statistically consistent.

The Von Neumann ratio test is done based on the total annual rainfall and the number of annual wet days. The formulation for the Von Neumann ratio test is given in Eq. 2.1. Five different values of thresholds are considered, i.e., 0.1, 1.0, 2.5, 5.0 and 10.0 mm of daily rainfall. Table 3.1 shows the statistics of each threshold, i.e., the mean, standard deviation and skewness, as well as the Von Neumann ratio for the total annual rainfall.

Table 3.1 The statistics and Von Neumann ratio based on total annual rainfall

Threshold, δ (mm)	**Mean, (mm)**	**Percentage difference**	**Standard deviation, (mm)**	**Skewness**	**Von Neumann ratio**
0.1	2,531		379	0.11	1.39
1.0	2,515	0.6	380	0.11	1.38
2.5	2,469	2.5	378	0.08	1.38
5.0	2,378	6.0	381	0.09	1.38
10.0	2,160	14	381	0.11	1.28

The Von Neumann ratios for the total annual rainfall show that all thresholds except 10.0 mm do not exceed the 1% significance level (refer to Table 2.1). Figure 3.2 shows the total annual rainfall from 1960 to 2011 for thresholds of 0.1, 1.0, 2.5 and 5.0

mm. The highest total annual rainfalls are shown for threshold 0.1 mm, and smaller values are shown for other thresholds. There are no significant differences in the total annual rainfall between thresholds of 0.1 and 1.0 mm. Additionally, the decreasing mean with the increasing amount of threshold is expected because the smaller rainfall amounts are being neglected. The changes in mean are quantified in terms of percentage based on the 0.1 mm threshold. Small differences are shown between the thresholds of 1.0, 2.5 and 5.0 mm. Other statistics, such as standard deviation and skewness, also show insignificant differences.

The second homogeneity test is done based on the total annual wet days. Table 3.2 summarized the statistics, percentage difference of the mean of total annual wet days based on threshold 0.1 mm, and also the Von Neumann ratio.

Table 3.2 The statistics and Von Neumann ratio based on total annual wet days

Threshold, δ (mm)	**Mean, (days)**	**Percentage difference**	**Standard deviation, (days)**	**Skewness**	**Von Neumann ratio**
0.1	194		14	-0.21	1.91
1.0	156	20	13	-0.01	2.06
2.5	130	33	12	-0.08	1.89
5.0	105	46	11	-0.02	1.94
10.0	75	61	10	0.13	1.41

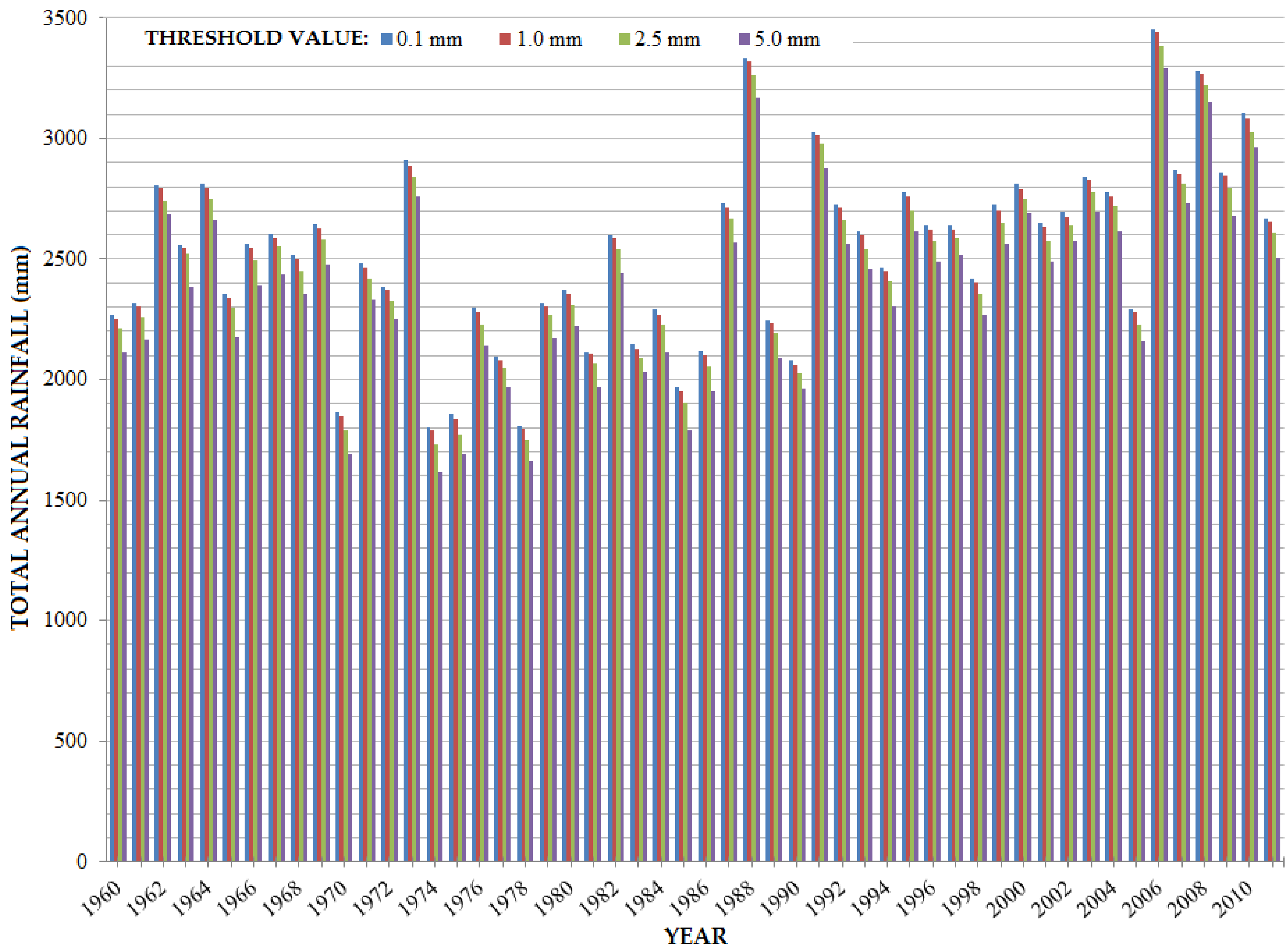

Figure 3.2 Total annual rainfall from 1960 to 2011 for thresholds 0.1, 1.0, 2.5 and 5.0 mm

The Von Neumann ratios for thresholds of 0.1, 1.0, 2.5 and 5.0 mm show excellent results; these values did not exceed the 5% critical values. Figure 3.3 illustrates the total annual wet days for thresholds of 0.1, 1.0, 2.5 and 5.0 mm.

Substantial differences in the total annual wet days are shown for all thresholds. For example, the difference in total annual wet days between thresholds of 0.1 and 1.0 mm is 20% and it increased significantly to 33% when the threshold is 2.5 mm.

The Von Neumann ratios give an indication that all thresholds, except for 10.0 mm, are suitable to be used. However, the statistics of the test show that there is a significant difference in the mean of total annual number of wet days for each of the thresholds. Selecting a larger value of threshold may result in the underestimation of rainfall occurrence. Considering this factor, therefore, 0.1 mm is selected as the threshold for this study.

The occurrence of rainfall event in this study is treated as discrete and the time frame for a day is from 8 a.m. to 8 a.m the next day, based on the daily data provided by the Department of Meteorology, Malaysia. 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 classification can be summarized as follows;

$$X_t = \begin{cases} 1 \text{ if rainfall} > 0.1 \text{ mm} \\ 0 \text{ if rainfall} \leq 0.1 \text{ mm} \end{cases}$$

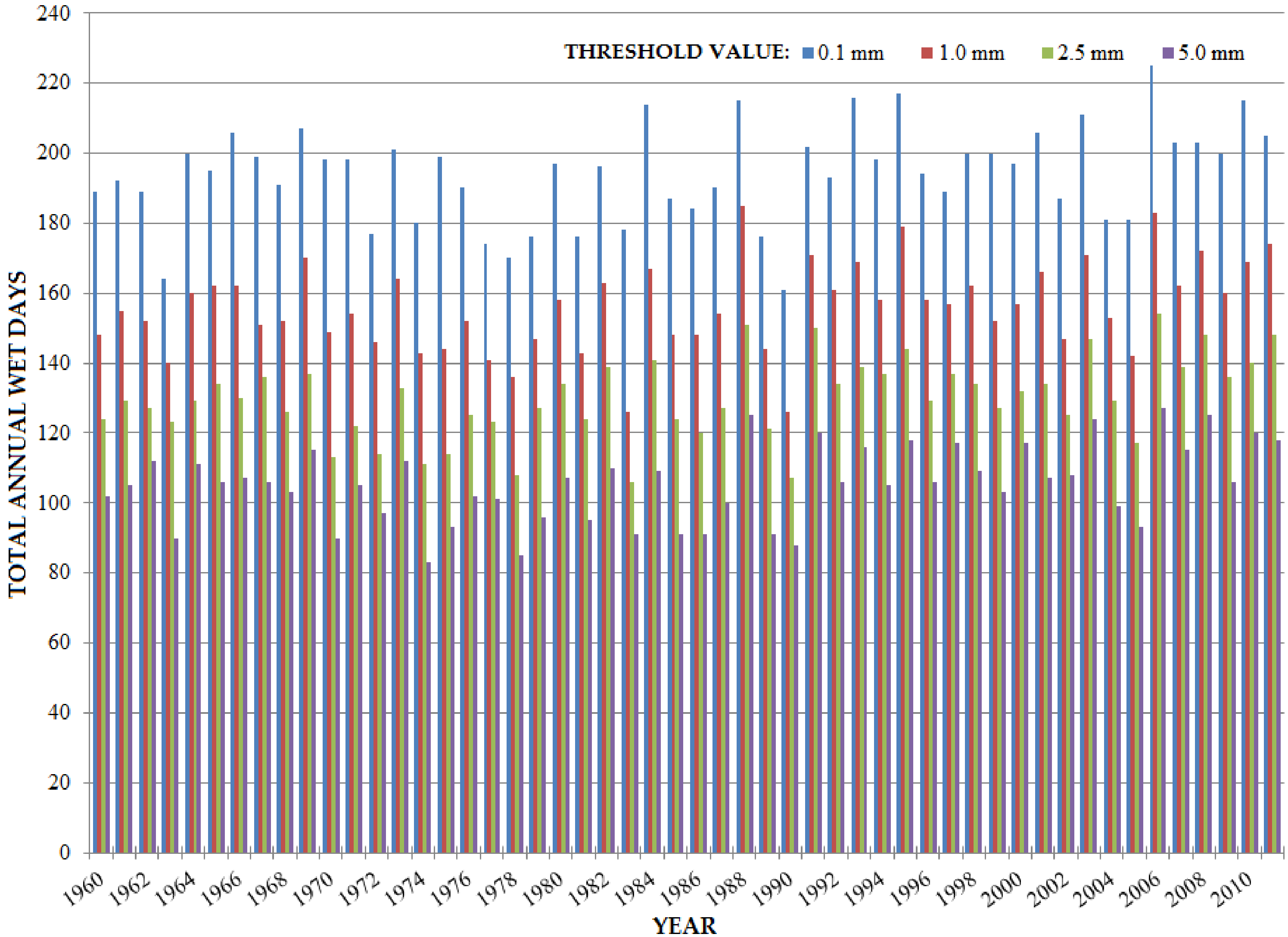

Figure 3.3 Total annual wet days from 1960 to 2011 for thresholds 0.1, 1.0, 2.5 and 5.0 mm

3.2.2 ANNUAL RAINFALL STATISTICS

The total annual rainfall statistics from 1960 to 2011 at Subang Airport shows that the study area receives an average annual rainfall of 2,531 mm, with a standard deviation of 379 mm. Figure 3.4 shows the plot of total annual rainfall from 1960 to 2011.

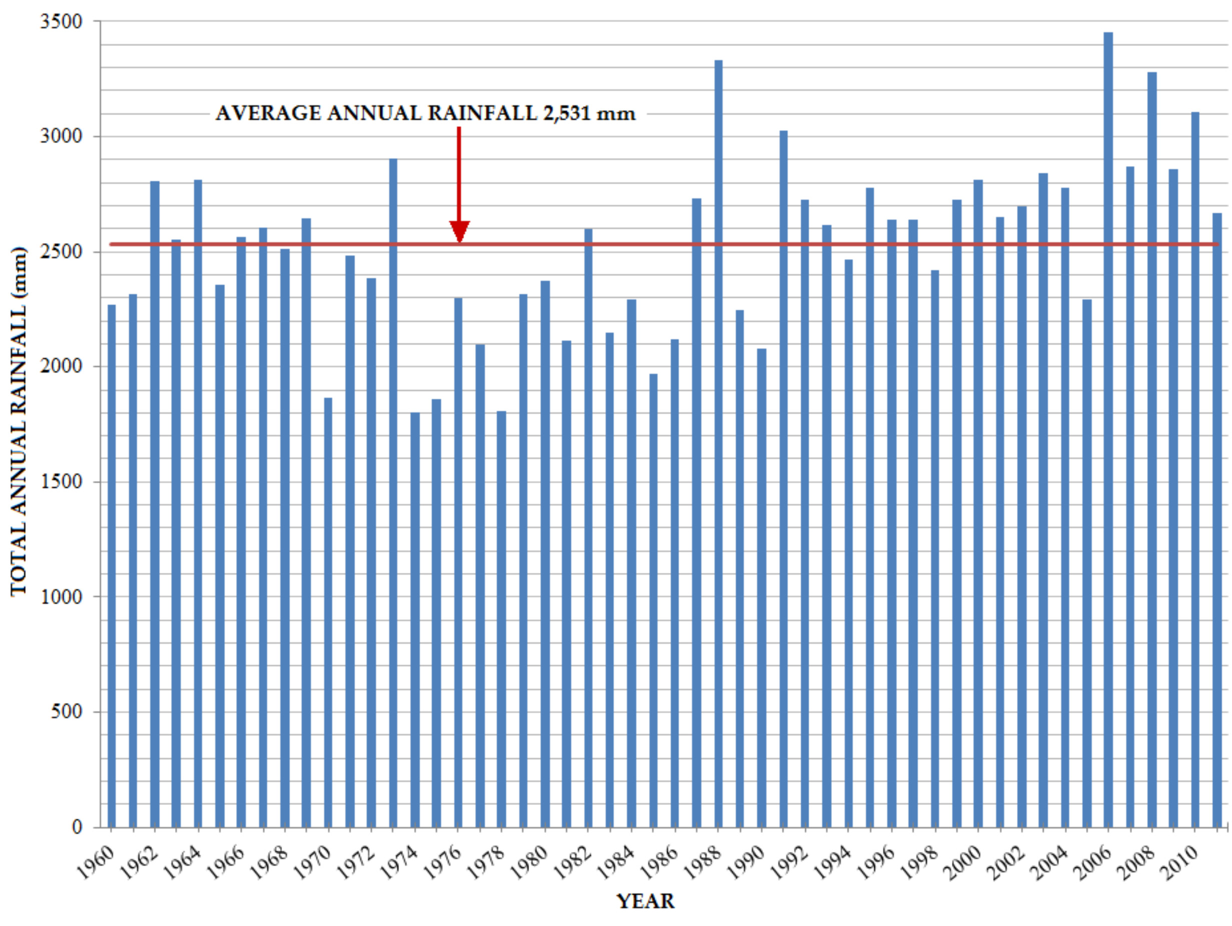

Figure 3.4 Total annual rainfall at Subang Airport from 1960 to 2011

The discussions of total annual rainfall are concentrated on five years and these include:

1) 1968 - the total annual rainfall is very close to the average annual rainfall;
2) 1971 – the maximum rain in a day occurs in this particular year;
3) 1974 – the minimum total annual rainfall;
4) 2003 – the longest wet run;
5) 2006 – the maximum total annual rainfall.

Figure 3.5 shows the plot of daily rainfall recorded in 1968. The year 1968 recorded the amount of rainfall closest to the average rainfall, with a total of 2,515 mm. For an average year, there are a few occurrences of rainfall where Subang Airport receives considerably more than 50 mm. The highest rainfall in a day was recorded in August 23rd, which is more than 80 mm. The measurement was taken during the occurrence of the SW monsoon. Another important observation is that multi-day rainfalls are common and these events can be observed in Figure 3.5.

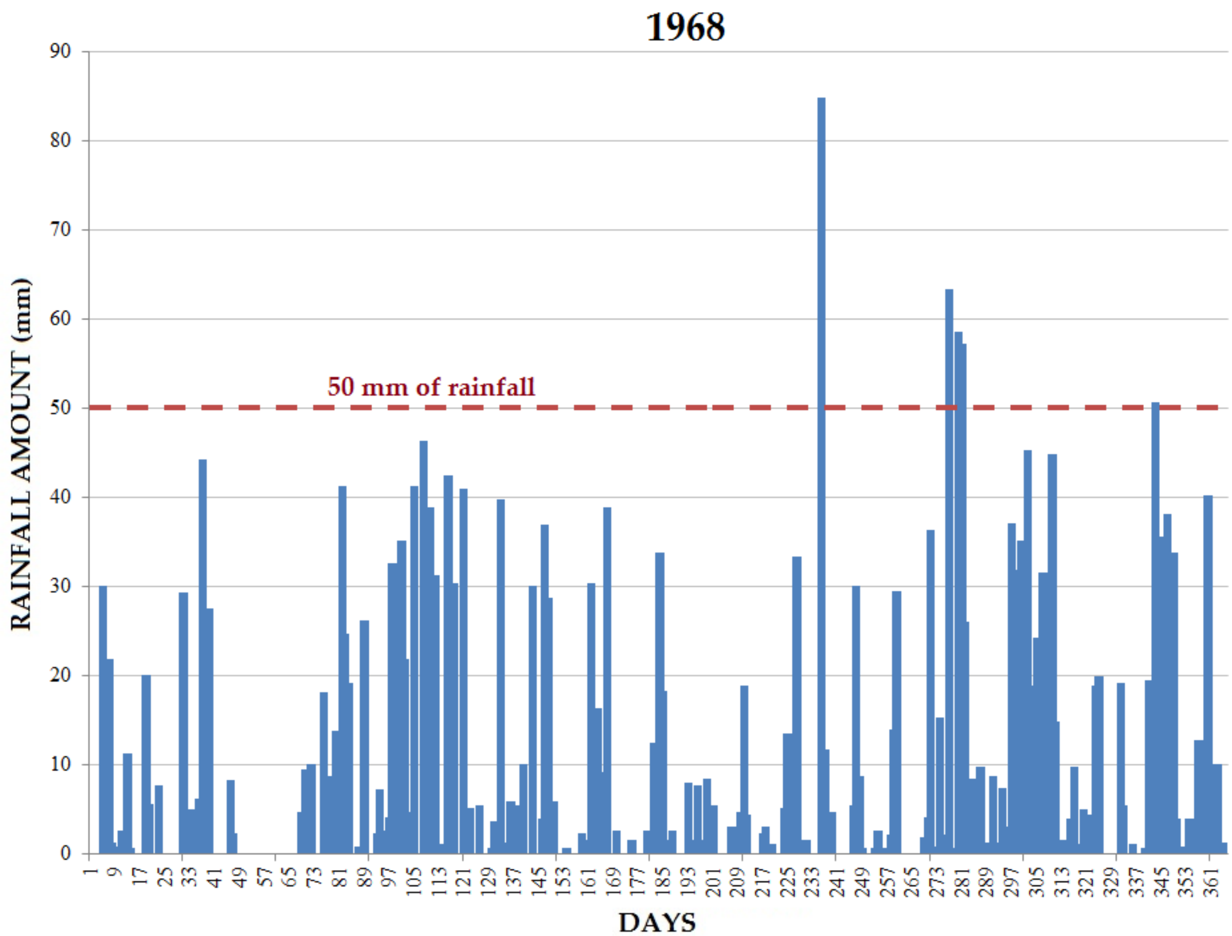

Figure 3.5 Daily rainfall recorded in 1968

Figure 3.6 shows the daily rainfall recorded in 1971. The maximum rainfall in a day was recorded on Jan 4th, 1971 with the measurement of 171.5 mm. This event occurred as a part of an 11-consecutive rainy days event starting on December 26th, 1970, which is classified in the NE monsoon. After this event, the occurrences of other multi-day events are commonly observed, but the magnitudes of rainfall are much less than the January 4th rainfall. However, multi-day rainfall events with the magnitude of more than 50 mm are observed mainly towards the end of 1971, which is during the NE monsoon season.

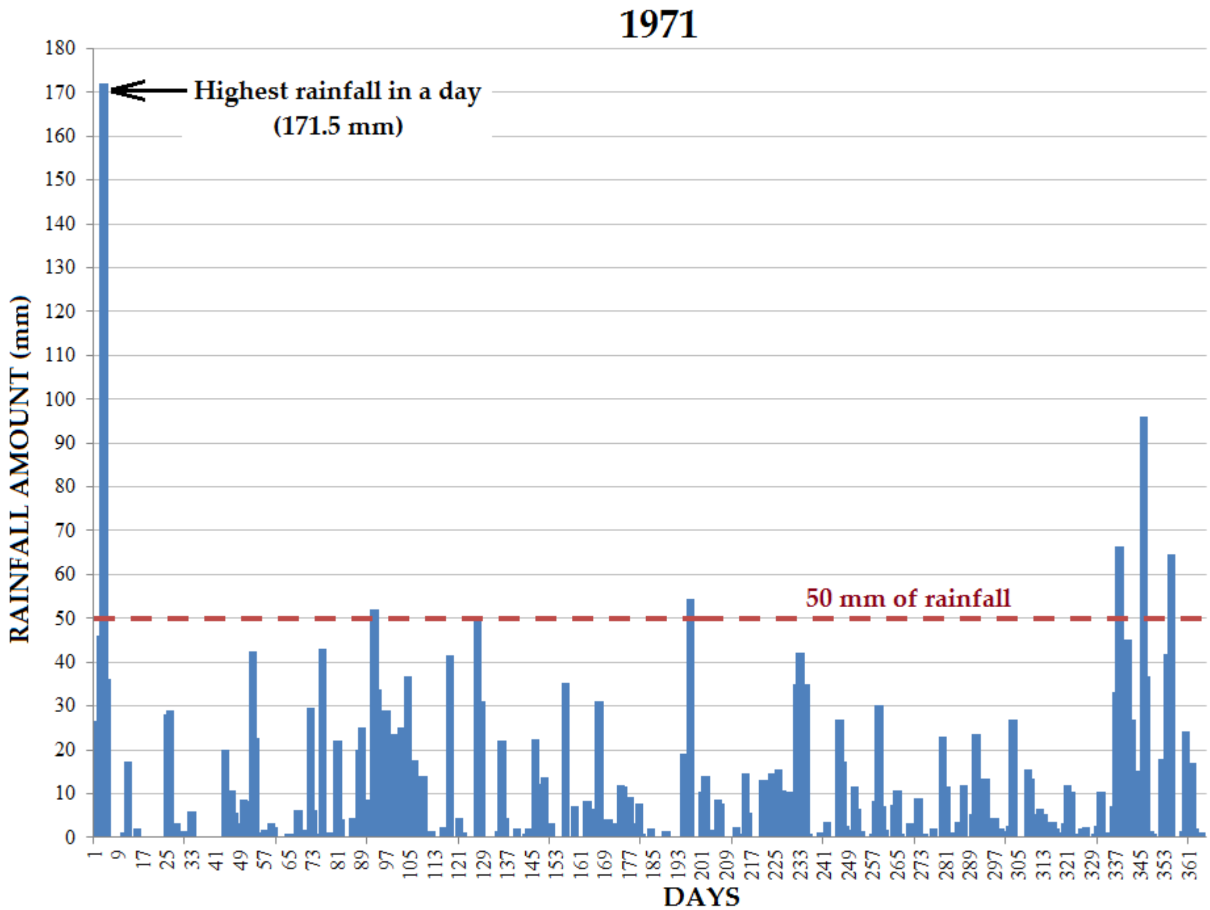

Figure 3.6 Daily rainfall recorded in 1971

Figure 3.7 shows the daily rainfall recorded in 1974, in which the Subang Airport received the least amount, with the measurement of 1,802 mm. Although this particular year recorded minimum rainfall, there are a few occurrences of events with magnitude of more than 50 mm. Another important observation is that multi-day rainfall events are common in 1974, as shown in Figure 3.7.

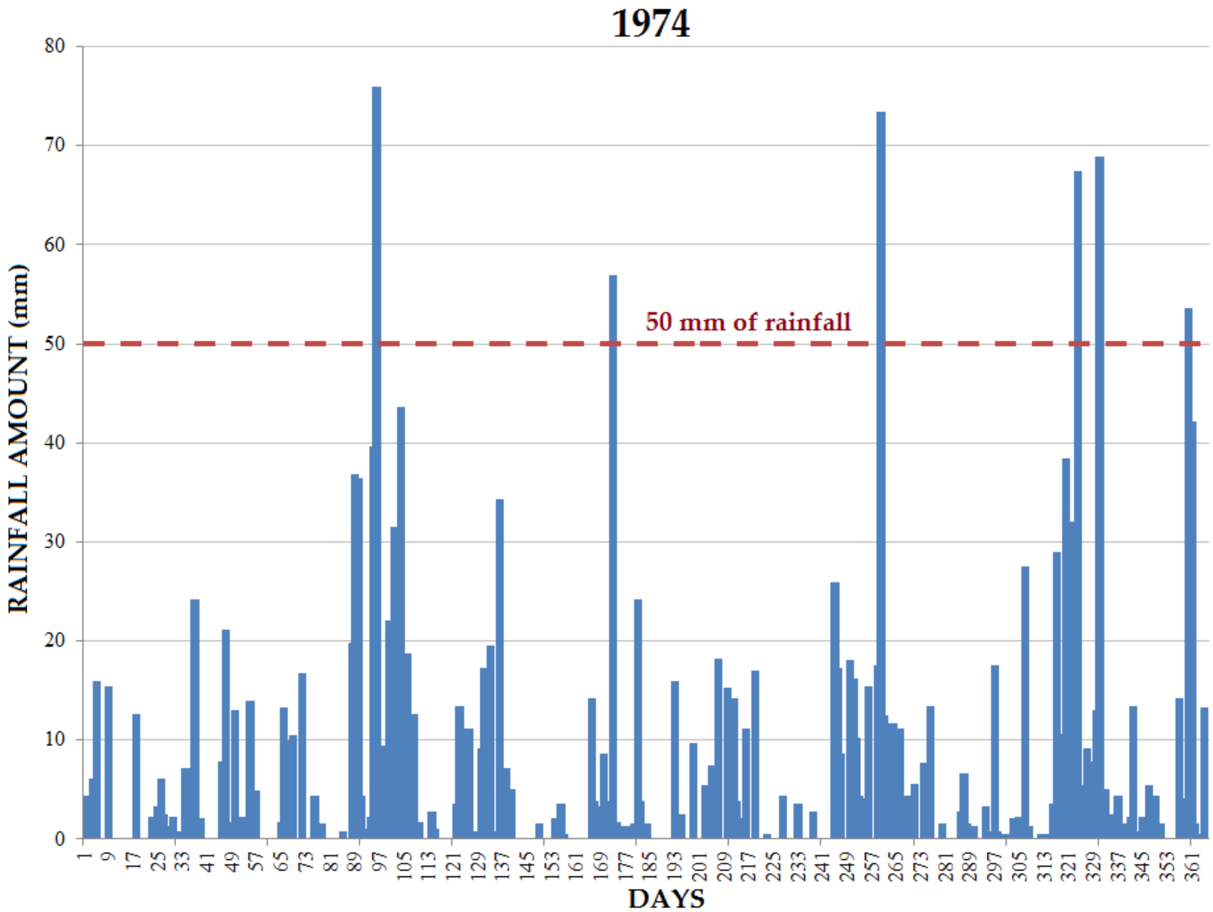

Figure 3.7 Daily rainfall recorded in 1974

Figure 3.8 shows the daily rainfall recorded in 2003. This particular year recorded the longest multi-day event. The event was recorded from 27th October to 26th November, 2003, i.e., 31 consecutive wet days, which resulted in a total rainfall of 624 mm. This event happens during the NE monsoon season. Additionally, Figure 3.8 also shows that the multi-day rainfall events are common. Some of the events at Subang Airport recorded significant rainfall amounts of more than 50 mm.

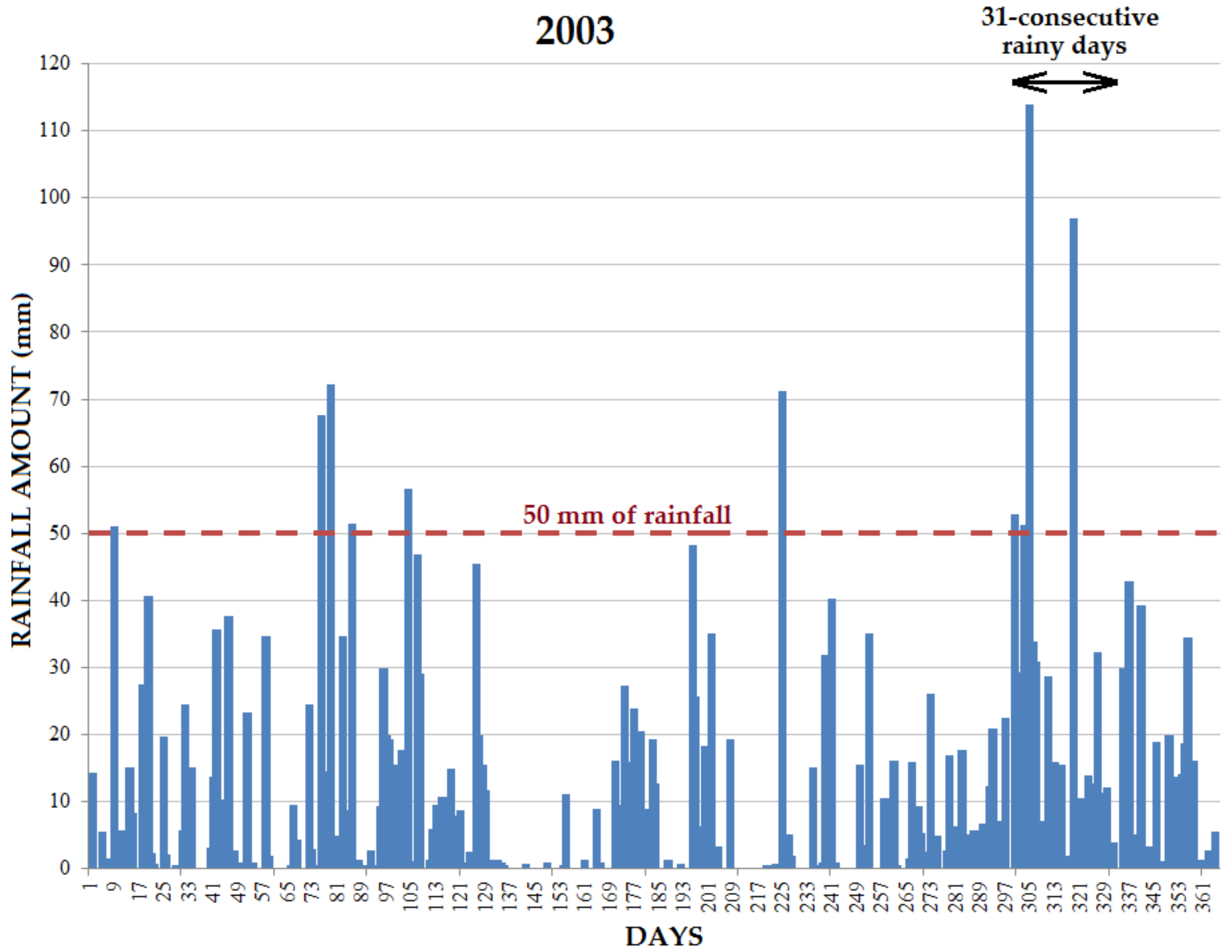

Figure 3.8 Daily rainfall recorded in 2003

Figure 3.9 shows the daily rainfall recorded in 2006 at Subang Airport. This particular year was chosen because it records the highest total amount of annual rainfall, i.e., 3,455 mm. Substantial amounts of rainfall were recorded at Subang Airport. Seventeen rainy days recorded amounts of more than 50 mm. The highest rainfall magnitude in a day was measured on October 20th, 2006 with 163.0 mm. Multi-day rainfall events are also commonly observed in Figure 3.9.

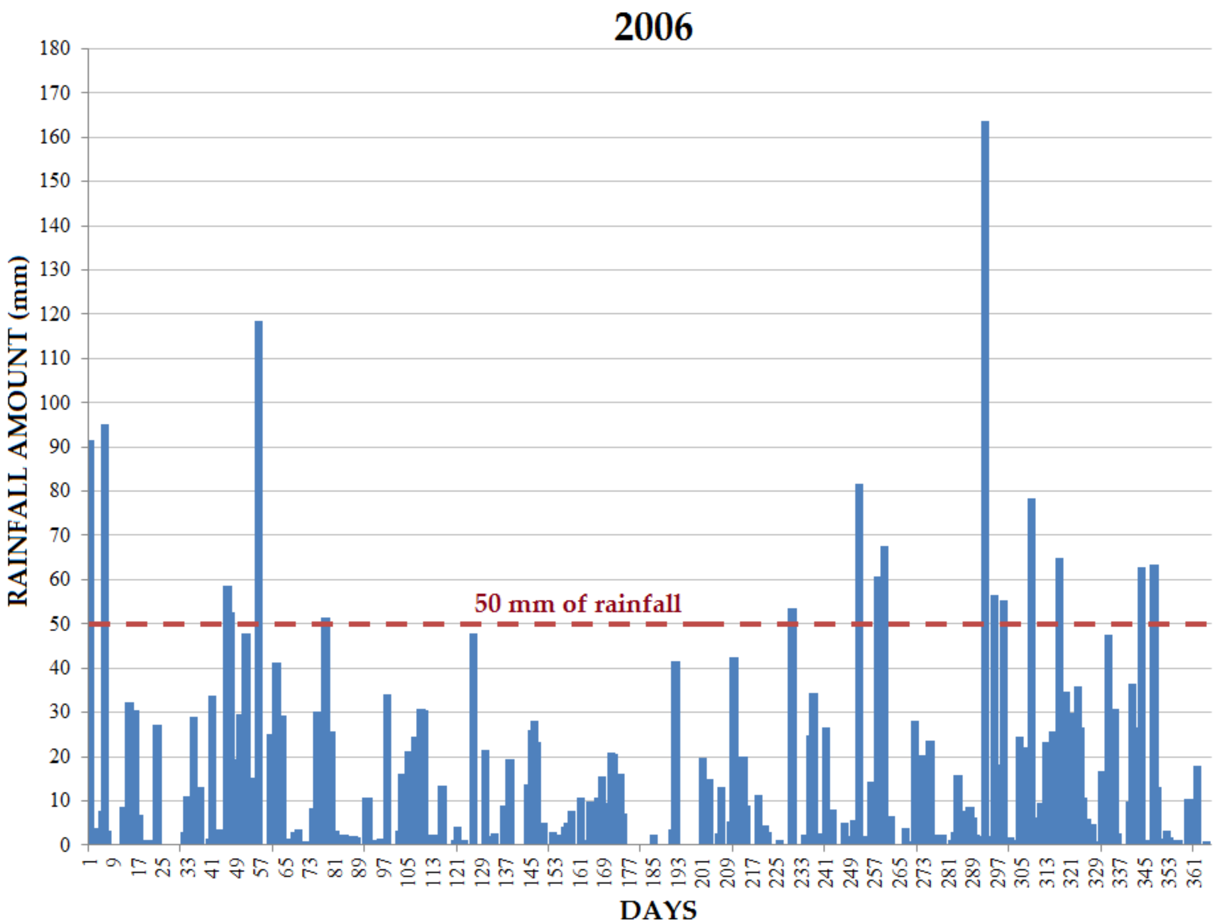

Figure 3.9 Daily rainfall recorded in 2006

Figures 3.6 to 3.9 show the various conditions of annual rainfall at Subang Airport. The common observations from these figures are the many occurrences of multi-day events and that the study area received considerable rainfall amounts each year.

3.2.3 MONTHLY RAINFALL STATISTICS

The plot of average monthly rainfall is given in Figure 3.10. The monthly rainfall amounts are affected by the respective monsoon seasons.

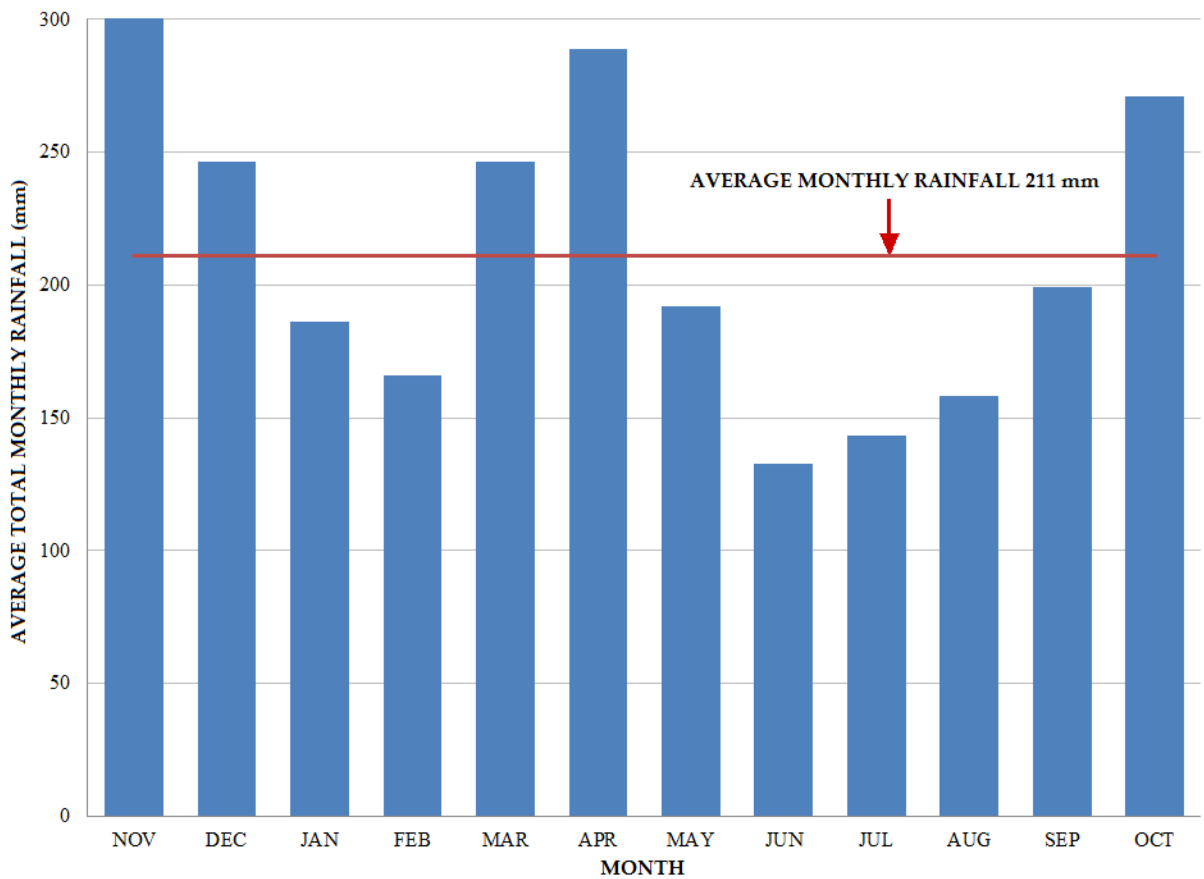

Figure 3.10 Average total monthly rainfall at Subang Airport

The average total monthly rainfall at Subang Airport shows that November (the beginning of the NE Monsoon) receives the highest amount. Lower magnitude of monthly rainfall was recorded in December, continues to decrease steadily until February and increases again in March. The total average rainfall amount recorded during the NE monsoon is 1,146 mm.

During the SW monsoon season, the maximum rainfall amount is recorded in May. The rainfall amounts decrease steadily for a few months and increase again in August and September. The months of June and July (SW Monsoon) recorded the lowest monthly rainfall. The total average rainfall amount recorded during the SW monsoon is slightly lower than the NE monsoon, with the measurement of 825 mm.

The inter monsoon months, i.e., April and October, show considerable rainfall amount as a result of convective rain, which usually occurs in the afternoon.

3.2.4 DAILY RAINFALL STATISTICS

The daily rainfall data measured at Subang Airport from 1960 to 2011 have a total of 18,993 days and from that number, 10,092 are rainfall days of more than 0.1 mm. The average daily rainfall is 13 mm and standard deviation of 17 mm.

The estimated numbers of wet run lengths observed at Subang Airport are given in Figure 3.11. There are a total of 3,726 wet run lengths. From that amount, 1,586 runs are one-rainy day, which accounts for 43% of the total wet run lengths. That gives 57% of the wet run lengths as equal to or more than 2-consecutive rainy days, i.e., multi-day events.

The details of the number of dry run lengths observed at Subang Airport are shown in Figure 3.12. The daily rainfall records give an estimated total of 3,727 dry run lengths. The majority of the dry run lengths are equal to or longer than 2-consecutive dry days, with the fraction of 52%, i.e., 1938 dry runs. The percentage shown for the dry run lengths is similar to the wet run lengths, i.e., the occurrence of multi-day events is more than the single day event.

Figures 3.13 and 3.14 give the estimated probability distribution based on the wet and dry run lengths, respectively. The estimated probability distribution for a single

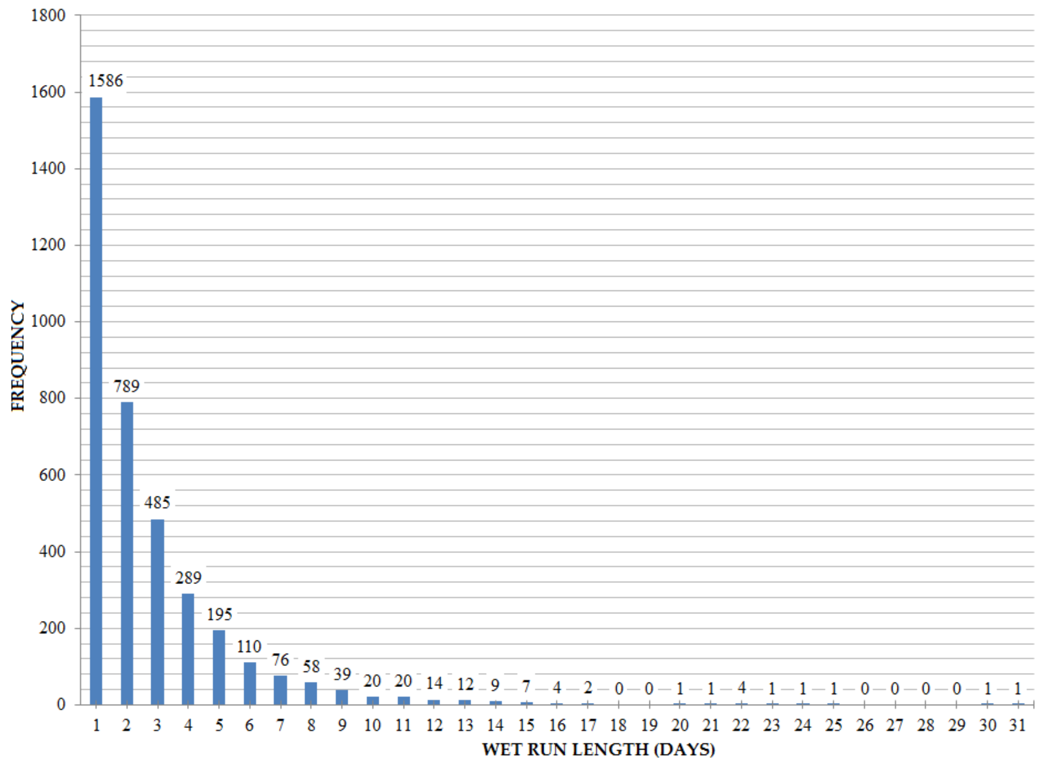

Figure 3.11 Number of wet run lengths

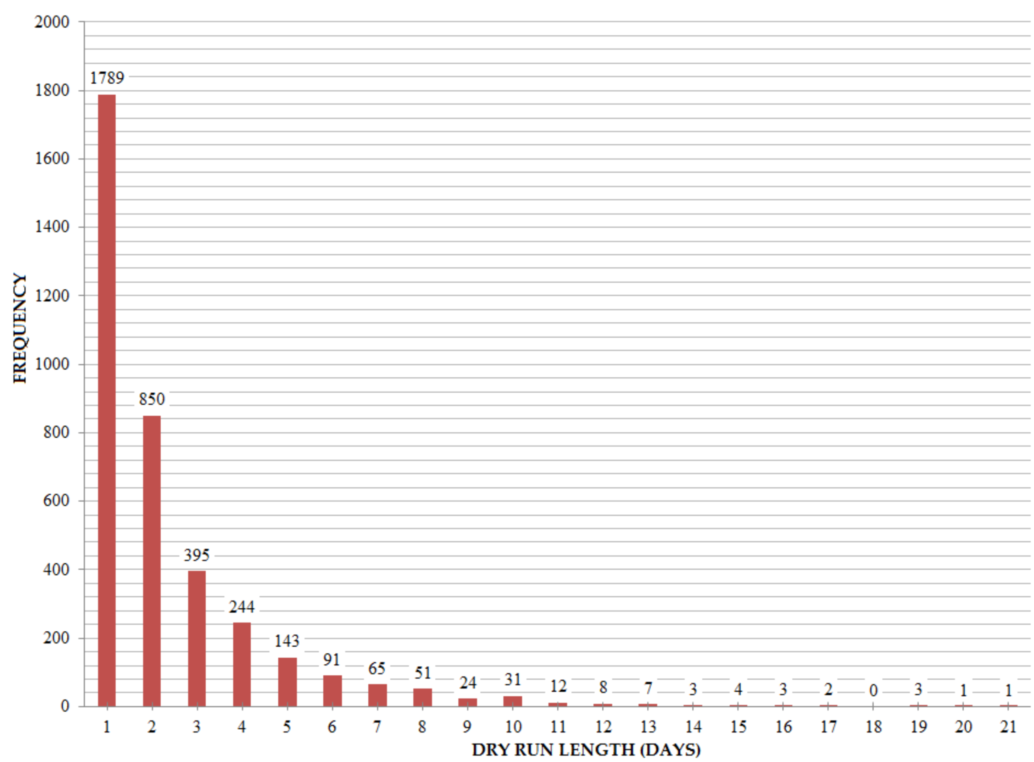

Figure 3.12 Number of dry run lengths

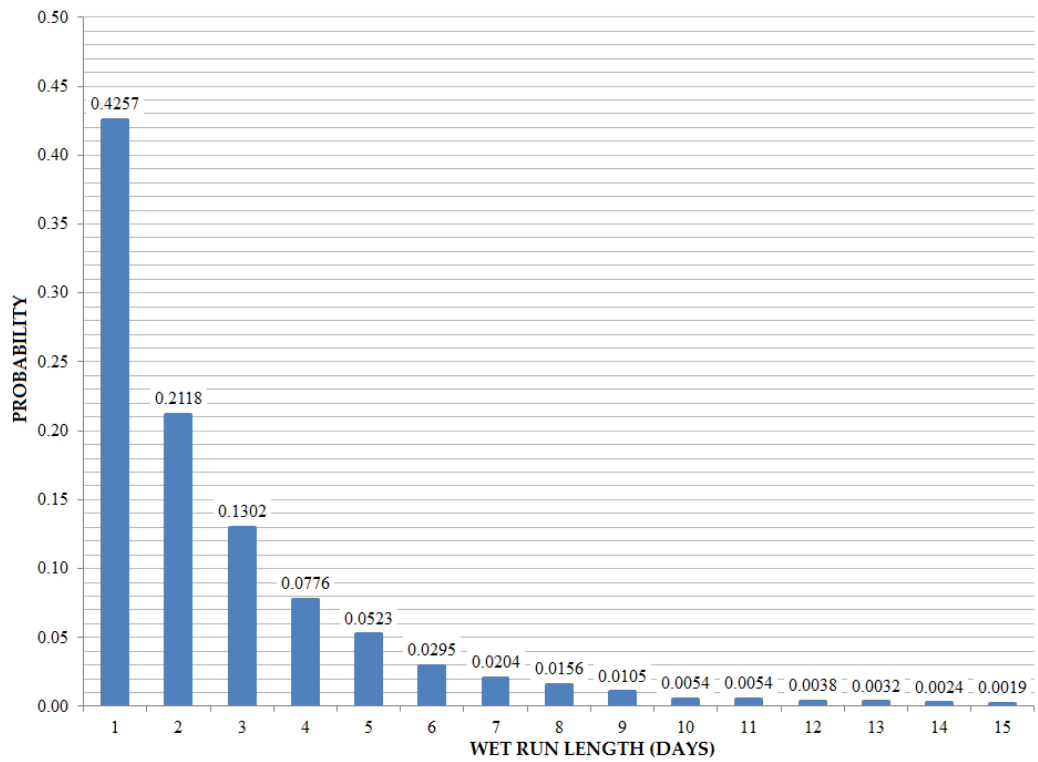

Figure 3.13 Probability distribution of wet run lengths

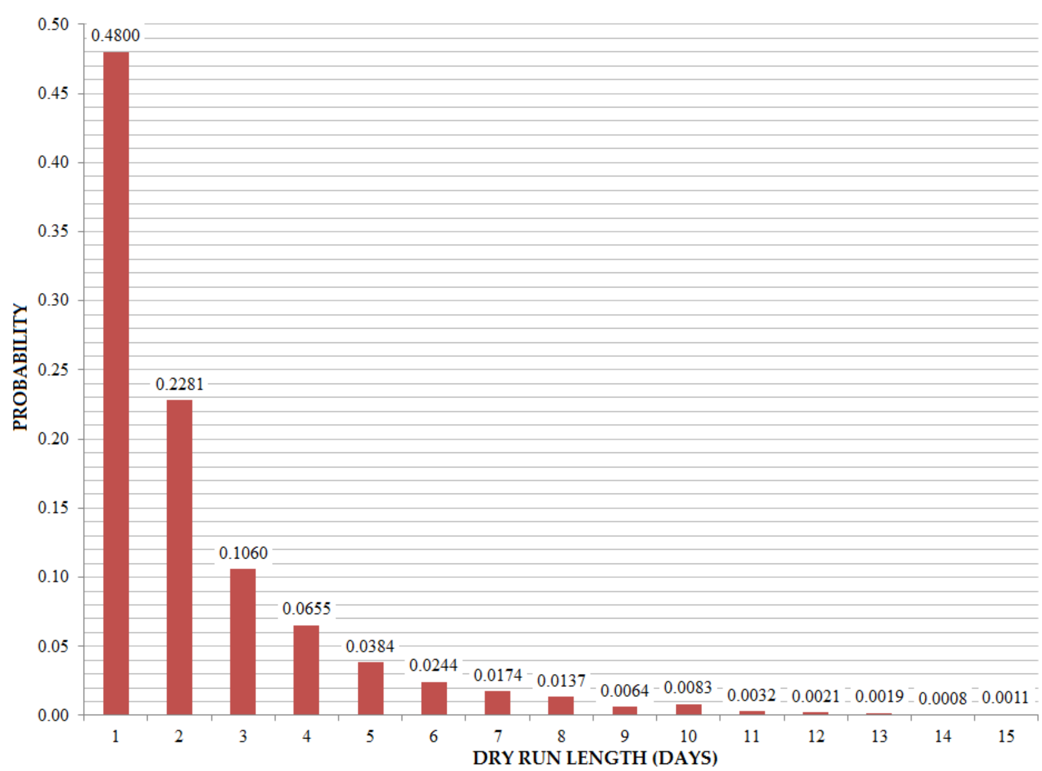

Figure 3.14 Probability distribution of dry run lengths

wet and dry day is 0.43 and 0.48, respectively. The mean wet and dry run lengths for the measured data is 2.7809 and 2.3882 days, respectively. These numbers also show that most of the rainfalls are multi-day events.

3.3 CONDITIONAL PROBABILITY OF MULTI-DAY RAINFALL EVENTS AT SUBANG AIRPORT

The conditional probability theory may determine whether the occurrence of random events is statistically independent or dependent. In this study, the interest is to determine whether future rainfall events are dependent on the given state (rain or dry) the previous day by analyzing the probability structure of multi-day events at Subang Airport.

Let R and D be the events in a probability space. R denoted a rainy day and D is a dry day on any random (t-th) day. Say P(R) is the probability of rain (wet) and P(D) is the probability of dry on any random day. Therefore the conditional probabilities are defined as

$$P(X_t = R|X_{t-1} = D) = \frac{P(R \cap D)}{P(D)} \qquad \text{(Eq. 3.1a)}$$

$$P(X_t = R|X_{t-1} = R) = \frac{P(R \cap R)}{P(R)} \qquad \text{(Eq. 3.1b)}$$

$$P(X_t = D|X_{t-1} = R) = \frac{P(D \cap R)}{P(R)} \qquad \text{(Eq. 3.1c)}$$

$$P(X_t = D|X_{t-1} = D) = \frac{P(D \cap D)}{P(D)} \qquad \text{(Eq. 3.1d)}$$

Where: X_t = any random day

t = time (days) = 1,2,3…

Events R and D are independent if and only if the following conditions are true:

$$P(X_t = R|X_{t-1} = D) = P(R) \quad \text{if } P(D) > 0 \qquad \text{(Eq. 3.2)}$$

$$P(X_t = D|X_{t-1} = R) = P(D) \quad \text{if } P(R) > 0 \qquad \text{(Eq. 3.3)}$$

The analysis started by counting the frequency of these events, and later Eq. 3.1b and 3.1d are used to calculate the conditional probabilities of t-consecutive wet and dry events at Subang Airport.

Table 3.3 gives the details of the frequency and the estimated conditional probability of 1 to 15-consecutive wet and dry days. Figures 3.15 and 3.16 show plots of the estimated conditional probability of n-consecutive wet and dry days, respectively.

It is estimated that more than 50% of the events observed at Subang Airport are rainy days. The estimated probability of rain on any random day is 0.5314. If day to day rainfall events are independent at this particular study area, the probability of rain on any day will remain constant at 0.5314 (shown in blue lines in Figure 3.15). However, Figure 3.15 shows that the probability structure increased significantly when the number of consecutive rainy day increased, i.e., from 0.5314 for a single rainy day to 0.80 for 15-consecutive days. The estimated conditional probability of a fourth rainy

Table 3.3 Frequency and estimated conditional probability 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	0.5314	1	8,901	0.4686
2	6,366	0.6308	2	5,174	0.5813
3	4,226	0.6638	3	3,236	0.6254
4	2,875	0.6803	4	2,148	0.6638
5	2,009	0.6988	5	1,455	0.6774
6	1,432	0.7128	6	1,006	0.6914
7	1,050	0.7332	7	700	0.6958
8	778	0.741	8	485	0.6929
9	582	0.7481	9	335	0.6907
10	444	0.7629	10	236	0.7045
11	345	0.7770	11	161	0.6822
12	266	0.7710	12	117	0.7267
13	207	0.7782	13	85	0.7265
14	162	0.7826	14	61	0.7176
15	129	0.7963	15	44	0.7213

day, given that it had rained for 3-consecutive days, is 0.6908. This probability is far greater than the estimated probability of the first day of rain, i.e., 0.5314. The examples given above show that the events are dependent; therefore, the estimated probability of rain on a given day is not constant. The occurrence of rain on a given day affects the probability of rain the following day.

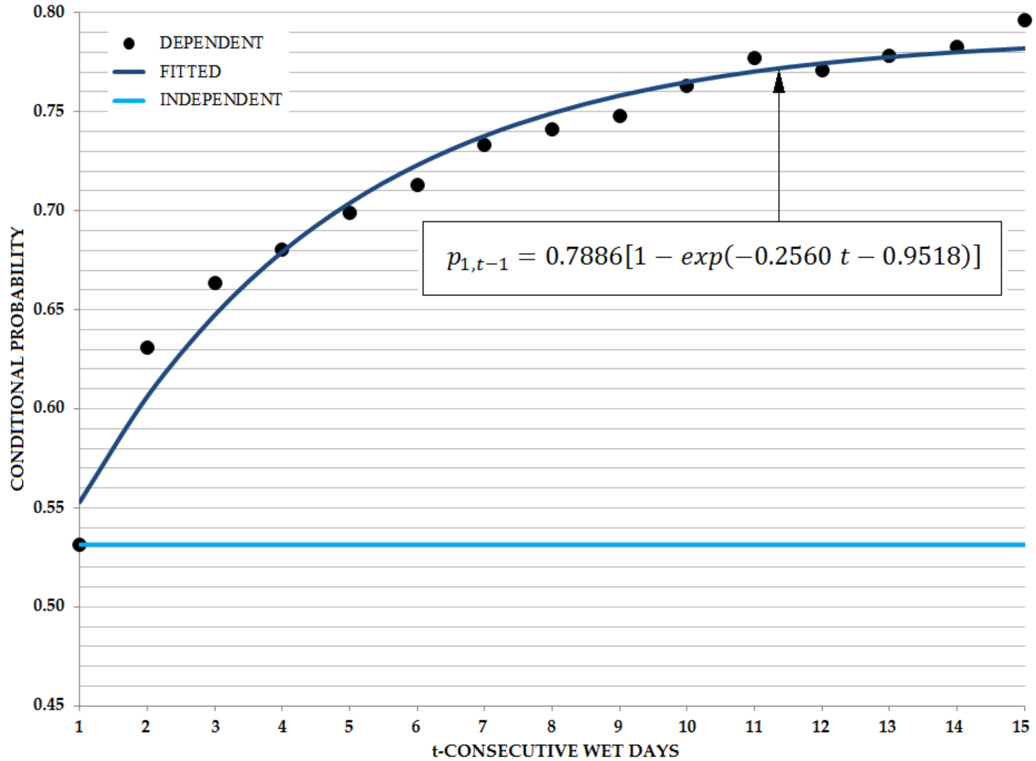

Figure 3.15 Plot of conditional probability of t-consecutive wet days

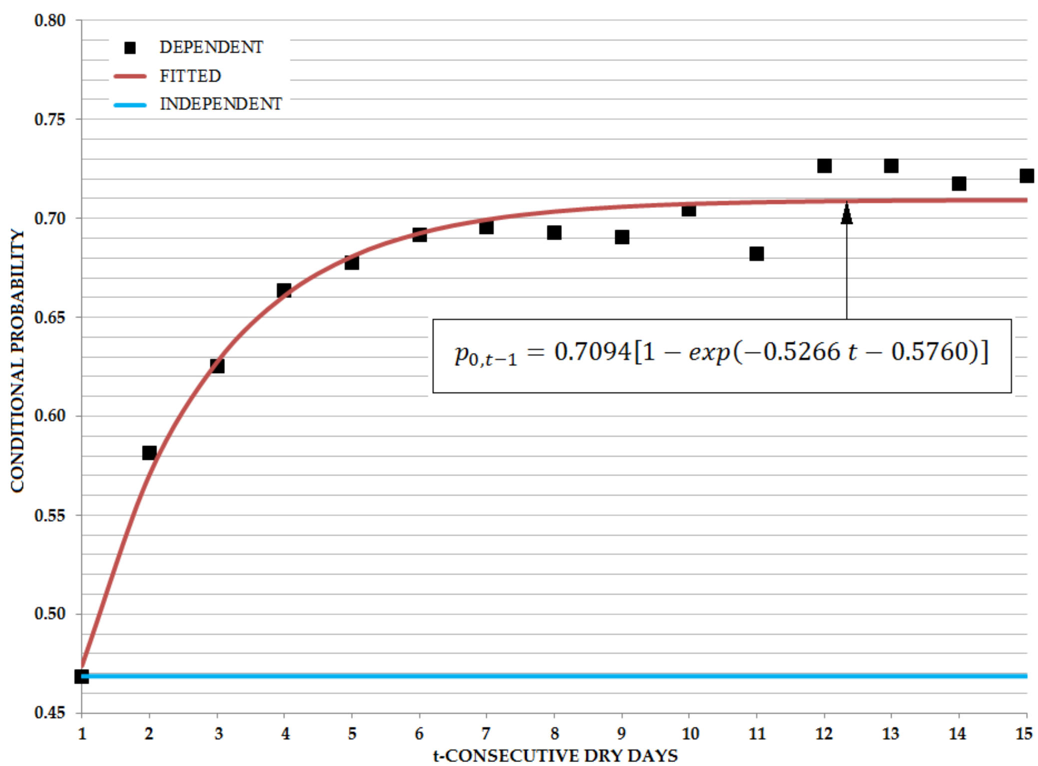

Figure 3.16 Plot of conditional probability of t-consecutive dry days

An exponential function is fitted to the conditional probability structure for t-consecutive wet days, and the equation is

$$p_{1,t-1} = 0.7886[1 - \exp(-0.2560\,t - 0.9518)] \quad \text{(Eq. 3.4)}$$

Where: $p_{1,t-1}$ = conditional probability of t-consecutive wet days

t = time (days) = 1,2,3,…

According to Eq. 3.4, the estimated probability of rain after a long multi-day rainfall gradually increases from 0.5314 to 0.7886.

Figure 3.16 shows the probability structure of n-consecutive dry days at Subang Airport. The estimated probability that a randomly chosen day will be dry is 0.4686, which increases significantly to 0.7213 after 15 consecutive dry days. Another example is that the estimated conditional probability for a 2 consecutive dry day is 0.5813, and the estimated probability for the third dry day increases to 0.6254. Thus, the probability structure of t-consecutive dry days is also dependent. Note that the estimated conditional probabilities of t-consecutive dry days are fitted using the exponential function, as shown in Eq. 3.5.

$$p_{0,t-1} = 0.7094[1 - \exp(-0.5266\,t - 0.5760)] \quad \text{(Eq. 3.5)}$$

Where: $p_{0,t-1}$ = conditional probability of t-consecutive dry days

3.4 DAILY RAINFALL DISTRIBUTION FUNCTION AT SUBANG AIRPORT

The Cumulative Distribution Function (CDF) for the observed data is represented using the plotting position formula known as the Weibull method. The formula for the Weibull method is given in equation 3.6.

$$F(x) = \frac{i}{N+1} \qquad \text{(Eq. 3.6)}$$

Where: x = rainfall amount (mm)

i = rank (ordered sample from the smallest to the largest)

N = sample size

The two parameter gamma function is used to fit the daily rainfall data, and the density function is given in Eq. 3.7.

$$f(x) = \frac{1}{|\alpha|\Gamma(\beta)}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(-\frac{x}{\mu}\right) \qquad \text{(Eq. 3.7)}$$

Where: x = rainfall amount (mm)

μ = average daily rainfall (mm)

α = scale parameter

β = shape parameter

The shape and scale parameters are estimated using the method of moments, and are given in Eq. 3.8 and 3.9, respectively.

$$\hat{\beta} = \left(\frac{\hat{\mu}}{\hat{\sigma}}\right)^2 \qquad \text{(Eq. 3.8)}$$

$$\hat{\alpha} = \frac{\hat{\mu}}{\hat{\beta}} \qquad \text{(Eq. 3.9)}$$

Where: $\hat{\mu}$ $= \frac{1}{N}\sum_{t=1}^{N} x_t =$ sample mean

$\hat{\sigma}$ $= \sqrt{\frac{1}{N-1}\sum_{t=1}^{N}(x_t - \bar{x})^2}$ = sample standard deviation

N = sample size

The estimated sample mean and standard deviation are $\hat{\mu} = 12.77$ mm and $\hat{\sigma} = 17.24$ mm. These values lead to the estimation of shape (β) and scale (α) parameters, and the values are $\hat{\beta} = 0.55$ and $\hat{\alpha} = 23.29$, respectively. The density function for a rainy day is given in Eq. 3.10.

$$f(x, 1) = \frac{1}{|23.29|\Gamma(0.55)}\left(\frac{x}{23.29}\right)^{-0.45} \exp\left(-\frac{x}{23.29}\right) \qquad \text{(Eq. 3.10)}$$

Figure 3.17 shows the CDF for the daily rainfall amount. For 1-rainy day, about 60% of the amounts of rainfall are less than 10 mm. The CDF also shows that the Subang Airport had received 50 mm or more rainfall in a day, i.e., 5% of the data, which is about 500 days from 1960 to 2011.

3.5 MULTI-DAY DISTRIBUTION FUNCTION AT SUBANG AIRPORT

Multi-day rainfall events are common in this study area, therefore the distribution function to represent t-consecutive rainfall days is examined in this section.

Eq. 3.11 to 3.15 show the empirical representation of 2- to 6- consecutive rainy days, and Eq. 3.16 gives the general form of the probability distribution for the total amount of rainfall resulted from 1- and t-consecutive rainy days.

2-consecutive rainy days

$$f(x, 2) = \frac{1}{|24.33|\Gamma(1.12)}\left(\frac{x}{24.33}\right)^{0.12} \exp\left(-\frac{x}{24.33}\right) \qquad \text{(Eq. 3.11)}$$

3-consecutive rainy days

$$f(x, 3) = \frac{1}{|24.20|\Gamma(1.75)}\left(\frac{x}{24.20}\right)^{0.75} \exp\left(-\frac{x}{24.20}\right) \qquad \text{(Eq. 3.12)}$$

4-consecutive rainy days

$$f(x, 4) = \frac{1}{|24.54|\Gamma(2.36)}\left(\frac{x}{24.54}\right)^{1.36} \exp\left(-\frac{x}{24.54}\right) \qquad \text{(Eq. 3.13)}$$

5-consecutive rainy days

$$f(x, 5) = \frac{1}{|25.14|\Gamma(2.95)}\left(\frac{x}{25.14}\right)^{1.95} \exp\left(-\frac{x}{25.14}\right) \qquad \text{(Eq. 3.14)}$$

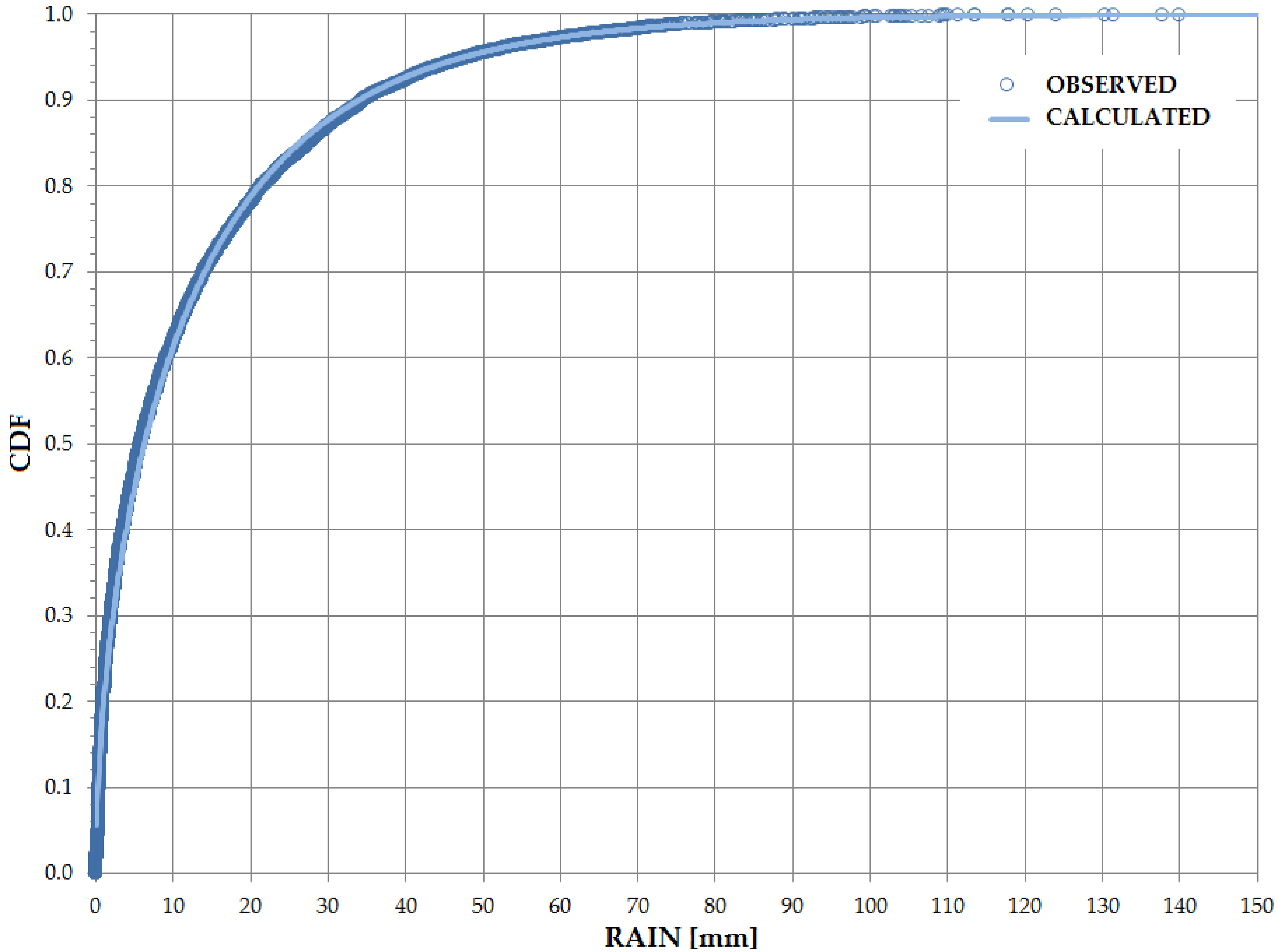

Figure 3.17 CDF for daily rainfall amount at Subang Airport

6-consecutive rainy days

$$f(x,6) = \frac{1}{|25.81|\Gamma(3.57)}\left(\frac{x}{25.81}\right)^{2.57}\exp\left(-\frac{x}{25.81}\right) \qquad \text{(Eq. 3.15)}$$

General equation

$$f(x,t) \cong \frac{1}{|24.0|\Gamma(0.6t)}\left(\frac{x}{24.0}\right)^{0.6t-1}\exp\left(-\frac{x}{24.0}\right) \qquad \text{(Eq. 3.16)}$$

Where: x = total amount of rainfall for t-consecutive rainy days (mm)

t = number of consecutive rainy days

Sections 3.4 and 3.5 show that the two-parameter gamma function is most suitable to represent the one-day and multi-day rainfall events at Subang Airport. This is because the moment generating function for the distribution of sum is also gamma distributed. The derivations for the moment generating function of a gamma function are given in Eq. 3.17 to Eq. 3.24.

The Moment Generating Function for a gamma function is

$$M_X(t) = E(e^{tx}) = \int_0^{\infty} \exp(tx)\frac{1}{\alpha\Gamma(\beta)}\left(\frac{x}{\alpha}\right)^{\beta-1}\exp\left(-\frac{x}{\alpha}\right)$$

$$M_X(t) = \int_0^{\infty} \frac{1}{\alpha\Gamma(\beta)}\left(\frac{x}{\alpha}\right)^{\beta-1}\exp\left(t-\frac{x}{\alpha}\right)dx \qquad \text{(Eq. 3.17)}$$

Where: x = daily rainfall measurements at Subang Airport (mm)

t = time (days)

α = scale parameter

β = shape parameter

After defining $$z = tx - \frac{x}{\alpha} \quad \text{(Eq. 3.18)}$$

Rearranging Eq. 3.18 and applying the $(\beta - 1)$ to both sides gives Eq. 3.19 and 3.20, respectively.

$$\frac{x}{\alpha} = \frac{z}{(1-\alpha t)} \quad \text{(Eq. 3.19)}$$

$$\left(\frac{x}{\alpha}\right)^{\beta-1} = \left(\frac{z}{1-\alpha t}\right)^{\beta-1} \quad \text{(Eq. 3.20)}$$

Differentiating Eq. 3.20 gives

$$\frac{dz}{dx} = \frac{d}{dx}\left[\frac{x}{\alpha}(1-\alpha t)\right] = \frac{1}{\alpha}(1-\alpha t) \quad \text{(Eq. 3.21)}$$

$$dx = \frac{\alpha}{(1-\alpha t)} dz \quad \text{(Eq. 3.22)}$$

Substituting Eq. 3.22 into Eq. 3.17 gives

$$M_X(t) = \int_0^\infty \frac{1}{\alpha\Gamma(\beta)}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(tx - \frac{x}{\alpha}\right) dx = \frac{1}{\alpha\Gamma(\beta)}\int_0^\infty \left(\frac{z}{1-\alpha t}\right)^{\beta-1} \exp(-z)\frac{\alpha}{(1-\alpha t)} dz$$

$$M_X(t) = \frac{1}{\alpha\Gamma(\beta)}\left(\frac{1}{1-\alpha t}\right)^{\beta-1}\left(\frac{\alpha}{1-\alpha t}\right)\int_0^{\infty} z^{\beta-1}\exp(-z)dz \qquad \text{(Eq. 3.23)}$$

$\int_0^{\infty} z^{\beta-1}\exp(-z)dz$ is gamma function, therefore Eq. 3.23 becomes

$$M_X(t) = \frac{1}{\alpha\Gamma(\beta)}\left(\frac{1}{1-\alpha t}\right)^{\beta-1}\left(\frac{\alpha}{1-\alpha t}\right)\Gamma(\beta)$$

$$M_X(t) = \left(\frac{1}{1-\alpha t}\right)^{\beta} = (1-\alpha t)^{-\beta} \qquad \text{(Eq. 3.24)}$$

Eq. 3.24 shows that the moment generating function for one-day is a function of scale (α) and shape (β) parameters.

Let B be the rainfall on any first day and C is the measurement on any second day, then A is the sum of any t-consecutive rainy days; i.e., A = B+C. The moment generating function for two-consecutive rainy days is shown in Eq. 3.25.

$$M_A(t) = M_B(t)M_C(t) = (1-\alpha t)^{-\beta_B}.(1-\alpha t)^{-\beta_C} = (1-\alpha t)^{-\beta_B-\beta_C} \qquad \text{(Eq. 3.25)}$$

Eq. 3.25 shows that the moment generating function for the sum is $(1-\alpha t)^{-\beta_B-\beta_C}$ which has two parameters, that is scale (α) and shape ($\beta_B + \beta_C$). Therefore it is shown that the distribution of a gamma function is also gamma distributed.

Figure 3.18 shows the CDF for 2- to 6- consecutive rainy days at Subang Airport. The CDFs show that the multi-day rainfall events resulted in a significant amount of rainfall to the study area. Historical records show that 2- and 3-consecutive rainy days are capable of producing more than 100 mm of rain. Half of the data for 4-, 5- and 6-consecutive events resulted in more than 55, 65, and 85 mm of rainfall, respectively. The probability of receiving more than 100 mm of rain increases as the number of consecutive rainy days increases. These results show that the duration and magnitude have the same significance in the analysis of multi-day events.

3.6 GOODNESS-OF-FIT TEST

The goodness-of-fit test is used to confirm the selection of a distribution function that has been proposed to represent the observed data. As shown in sections 3.6 and 3.7, the two-parameter gamma distribution best suited the distribution of daily rainfalls at Subang Airport.

Goodness-of-fit test can be performed using graphical or analytical methods. In this study, the graphical method, i.e., the 1:1 plot, is preferable to the analytical methods because it gives an excellent visual representation for the comparison between the observed data and the calculated values. Analytical methods such as the Chi-Square and Kolmogorov-Smirnov are best suited for a small sample size. Therefore, these methods are not suitable to be used in this study, because the sample size is large, i.e., 10,092 rainy days.

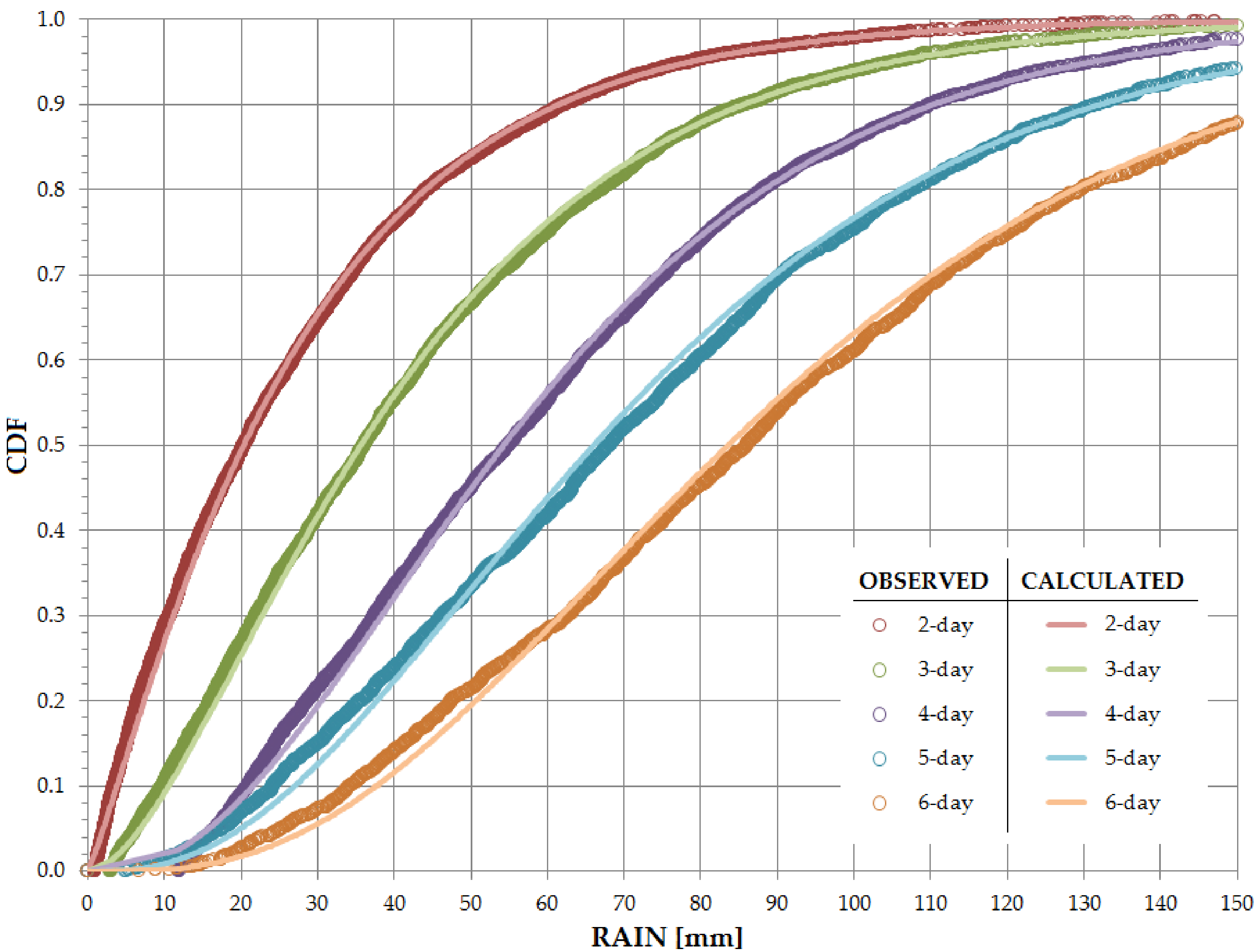

Figure 3.18 CDF of t-consecutive rainy days

Figure 3.19 shows the 1:1 plot for the observed data and calculated values. There are some insignificant differences which can be seen in 1-day of rainfall for probability distribution of less than 0.1. There is more than one occurrence of light rain (less than 1 mm), and therefore the same values of estimated probability are calculated using the two-parameter gamma function. Excellent fit is observed for other probabilities, including the large events (having the exceedence probability of 0.1 or less).

In general, the 1:1 plots show good agreement for 1-day to 6-consecutive rainy days, which confirms that the two-parameter gamma function is suitable to represent the distribution of rainfall for t-consecutive wet days.

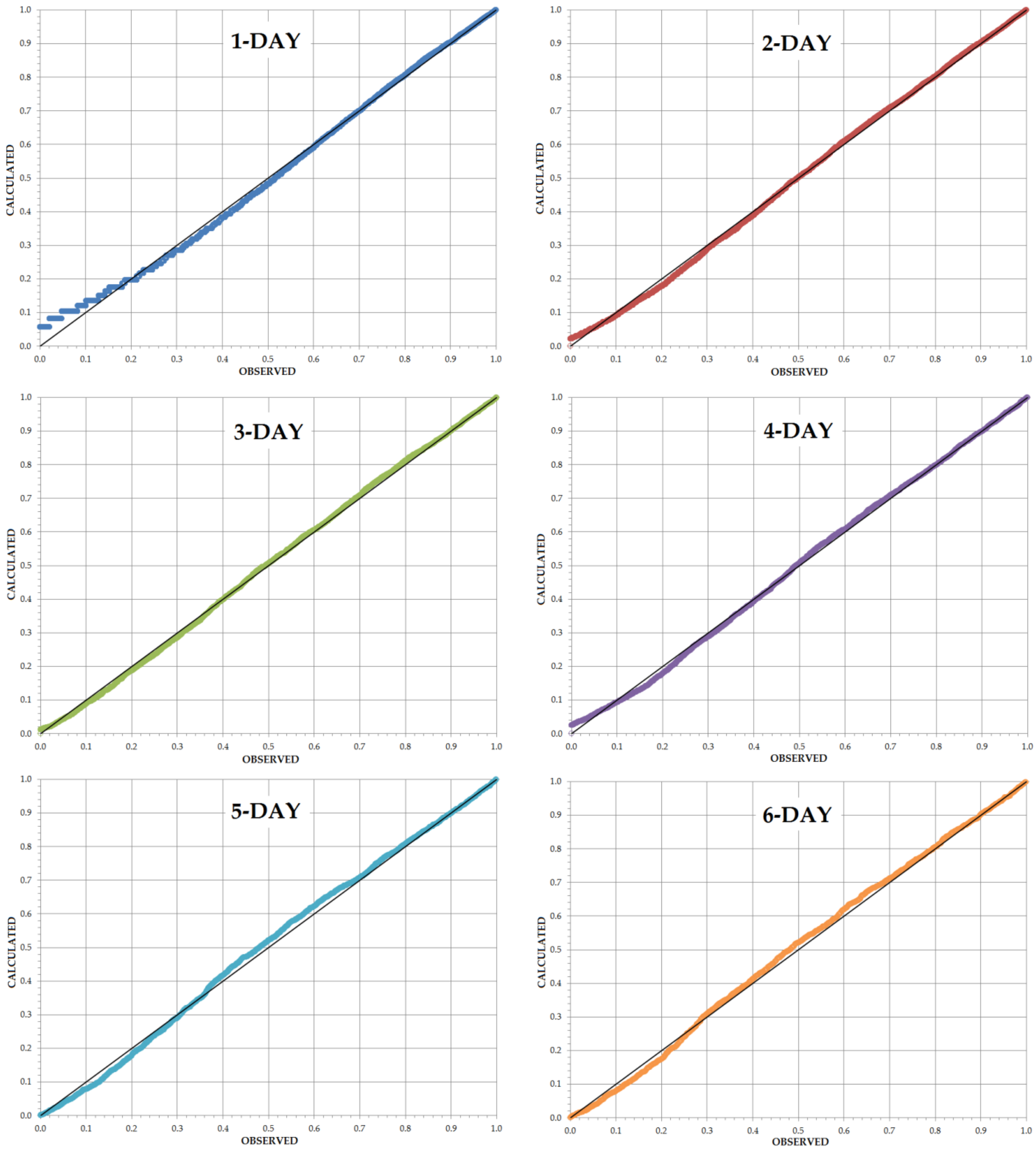

Figure 3.19 Comparison of CDF between calculated and observed using 1:1 plot

3.7 DEPENDENCE OF RAINFALL AMOUNT

The dependency of rainfall amount from one rainy day to the next is tested in this section, using three different scenarios: (1) all consecutive wet days; (2) rainfall on Day 1 and Day 2 (D1 & D2); and (3) rainfall on day 2 and day 3 (D2 & D3). The tests are done using two methods, i.e., determining the Auto Correlation Function (ACF) of the rainfall amount which is based on the rainfall amount (similar to equation 2.8) and by plotting the scatter plot.

For the first method, i.e., the ACFs for all scenarios are very low, which shows that the rainfall amounts are independent of each other. The ACFs are 0.0283, 0.0451, 0.0066 for all consecutive rainy days, D1 & D2 and D2 & D3, respectively. The results are summarized in Table 3.4.

Table 3.4 The ACFs for all consecutive rainy days, D1 & D2 and D2 & D3

Scenario	**Sample Size (Days)**	**ACF**
All consecutive rainy days	6,367	0.0283
D1 & D2	2,140	0.0451
D2 & D3	1,351	0.0066

Figures 3.20 and 3.21 show the scatter plot of the amounts of rainfall for D1 & D2 and D2 & D3. The observations for both graphs are the same, there are no structured appearances at any of the points and the plots are totally random. These plots further prove that there is no dependency between the amounts of rainfall for consecutive rainy days.

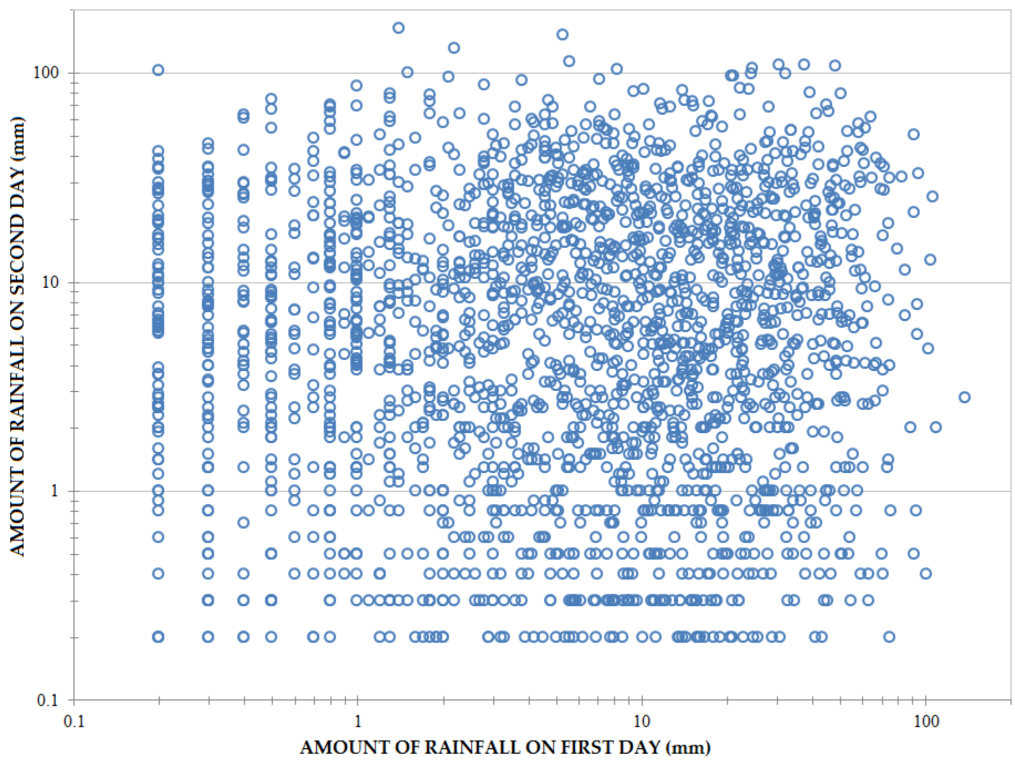

Figure 3.20 Amounts of Rainfall on D1 and D2

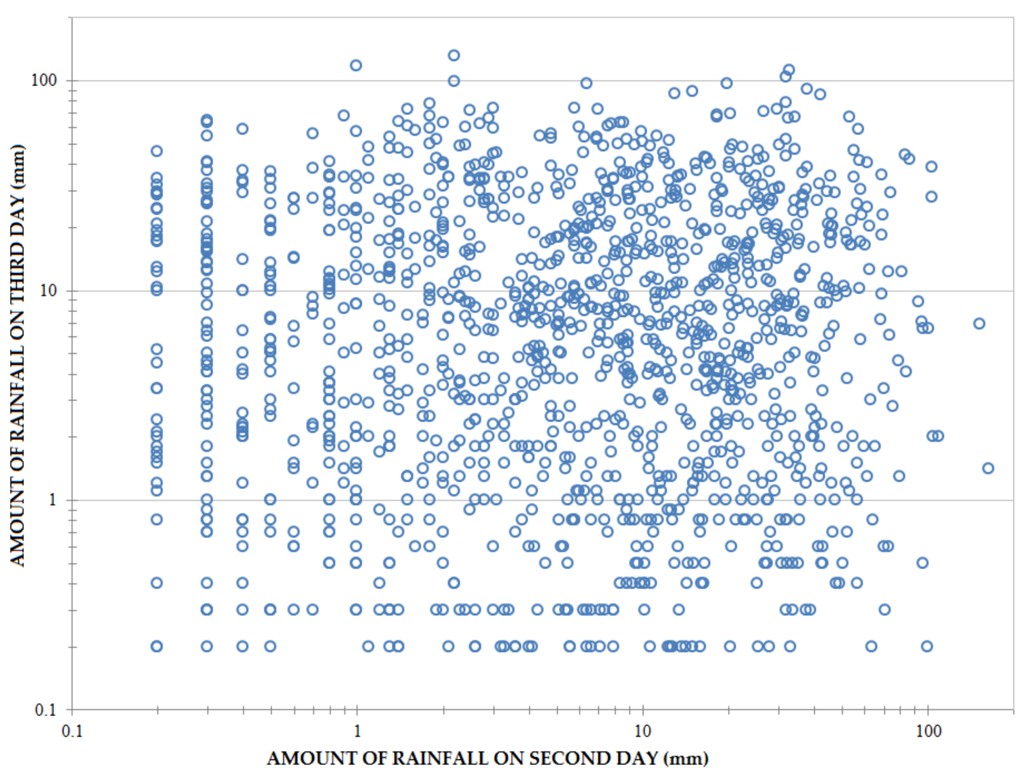

Figure 3.21 Amounts of Rainfall on D2 and D3

3.8 SUMMARY

The daily rainfall statistics show that most of the rainfall events at Subang Airport are multi-days. The conditional probabilities of t-consecutive wet and dry days increase with the event duration. This finding shows that the events are time-dependent. However, the rainfall amounts in this study area are independent from one rainy day to another. The two-parameter gamma distribution function is suitable to represent the daily rainfall data at Subang Airport.

CHAPTER 4

SIMULATION OF WET AND DRY DAY SEQUENCES FOR NORTH EAST AND SOUTH WEST MONSOONS

The rainstorms at Subang Airport are time-dependent and the majority of the events are multi-days rainfall, as shown in Chapter 3. Therefore, the daily rainfall sequence should be simulated using a discrete autoregressive model. This chapter gives detailed procedures of the model selection process. Two models are tested, i.e., the low order Discrete Auto Regressive [DAR(1)] and the low order Discrete Auto Regressive and Moving Average [DARMA(1,1)]. These models are chosen because of the different characteristics that they have. DAR(1) is a short memory model, while DARMA(1,1) has a long-term persistence.

4.1 MODEL FOR THE SIMULATION OF WET AND DRY SEQUENCES

A four-step process, as suggested by Salas and Pielke (2003), is adopted to simulate the sequences of daily rainfall. The four-step process is: model identification, model estimation, model selection and model verification. There are two major monsoon seasons that affect Peninsular Malaysia, i.e., the North East (NE) and South West (SW) monsoon. Separate analyses are conducted for these monsoons. In this study the NE and SW monsoons are classified as the daily rainfall recorded in the months of

October to March and April to September, respectively. The details of each step are given in the following sections.

4.1.1 Step 1: Model Identification

The first step is the model identification, in which the behaviour of the observed data is determined using the empirical Auto Correlation Function (ACF). The estimated empirical ACFs for each monsoon coefficient calculated using equation 2.9, based on the sequence of wet and dry days. The DAR(1) is a short memory model, therefore the ACF plot is expected to decay exponentially and tails off. If the empirical ACF exhibits this behaviour, then a DAR(1) model may be suitable. On the other hand, DARMA(1,1) has a long memory, therefore the ACF for this model should decay gradually. Figure 4.1 shows the observed plot of estimated ACFs for both monsoons.

Figure 4.1 shows a similar trend in the estimated values of ACFs for both monsoons. The estimated lag-1 ACFs for both monsoons are about 0.20. After that, the estimated ACFs for both monsoon decreased to 0.12 on day 2 and continue to gradually drop to 0.10 on day 3. The decreasing trend also continues from day 4 until the estimated ACFs are almost zero at day 15. The plot of estimated ACFs for NE and SW monsoons as shown in Figure 4.1 gives an early indication that the DARMA (1,1) model may be suitable to simulate the sequence of daily rainfall at Subang Airport for the NE and SW monsoons.

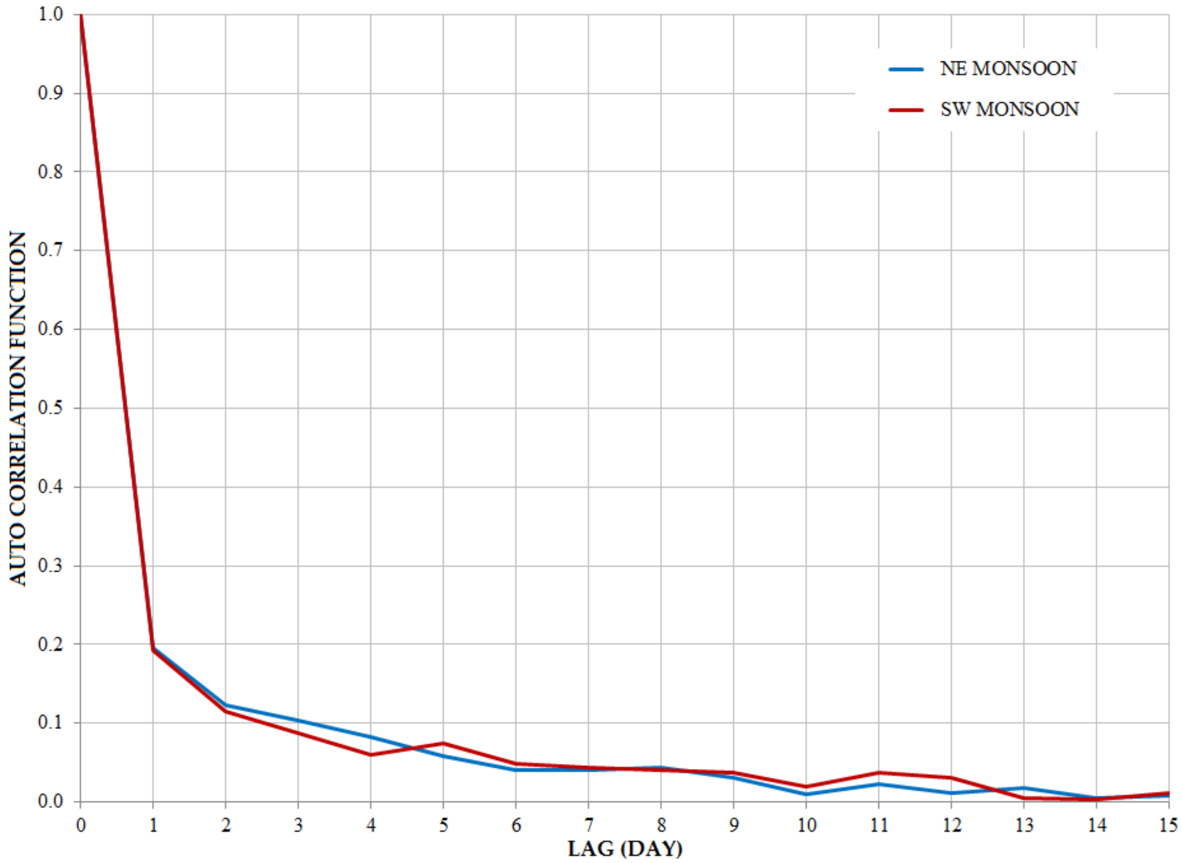

Figure 4.1 Observed ACFs for NE and SW monsoons

The analysis continues to the model estimation; that is, to determine the parameters needed for DAR(1) and DARMA(1,1). From there, the theoretical ACF for each model can be calculated and compared with the observed ACF.

4.1.2 Step 2: Model Estimation

In this step, the related parameters in the DAR(1) and DARMA(1,1) models are estimated using the Method Of Moments (MOM). The DAR(1) model has two parameters, that is π_1 (or π_0) and λ; while the DARMA(1,1) model is made up of three parameters, namely λ, β and π_1 (or π_0).

The sample mean wet ($\overline{T_1}$) and dry ($\overline{T_0}$) run lengths are estimated from the observed daily rainfall dataset, therefore the values are the same for both the DAR(1) and DARMA(1,1) models. During the NE monsoon, $\overline{T_1} = 3.000$ days and $\overline{T_0} = 2.1893$ days, and the values for the SW monsoon differ slightly, i.e., $\overline{T_1} = 2.4300$ days and $\overline{T_0} =$ 2.5798 days.

The probability distributions of dry and wet run lengths are estimated following Eq. 2.11 and 2.12, respectively. The observed data for the NE monsoon give the estimated wet and dry probability distributions of $\widehat{\pi_1} = 0.5781$ and $\widehat{\pi_0} = 0.4219$, respectively. The SW monsoon shows a slightly smaller value of wet probability distribution, i.e., $\widehat{\pi_1} = 0.4851$, which resulted in a bigger value of dry probability distribution, $\widehat{\pi_0} = 0.5149$.

The other parameter in DAR(1), λ, is calculated using Eq. 2.8, which is estimated as the observed lag-1 autocorrelation coefficient. The estimated values for $\hat{\lambda}$ are 0.1960 and 0.1918 for NE and SW monsoons, respectively. There are no significant differences in the estimation of $\hat{\lambda}$ for the two separate monsoon seasons.

The model estimation continues with the estimation of model parameters for the DARMA(1,1) model, namely λ and β. Using Eq. 2.23 as the initial value, parameter λ is estimated using Newton-Raphson iteration techniques (refer to Eq. 2.24). Eq. 2.25 leads to the estimation value of β.

For the NE monsoon, the estimated model parameters for DARMA(1,1) are $\hat{\lambda} = 0.7339$ and $\hat{\beta} = 0.5775$, while the observed daily rainfall series during the SW monsoon gives the estimation of $\hat{\lambda} = 0.7827$ and $\hat{\beta} = 0.5789$. Table 4.1 summarizes the estimated model parameters for DAR(1) and DARMA(1,1) for both monsoon seasons.

The final task in this step is to compare the observed and theoretical ACF of the DAR(1) and DARMA(1,1) models using the graphical method. Theoretical ACFs are determined using Eq. 2.8 and 2.21 for DAR(1) and DARMA(1,1), respectively, with the estimated parameters that have been determined earlier in this section (refer to Table 4.1). Figures 4.2 and 4.3 show the comparison between the observed and theoretical ACFs for the NE and SW monsoons, respectively.

Table 4.1 Model Parameters for DAR(1) and DARMA(1,1)

Monsoon Seasons	Estimated by run lengths		Estimated by ACF		
	DAR(1) and DARMA(1,1)		DAR(1)	DARMA(1,1)	
	$\widehat{\pi_1}$	$\widehat{\pi_0}$	$\hat{\lambda}$	$\hat{\lambda}$	$\hat{\beta}$
NE	0.5781	0.4219	0.1960	0.7339	0.5775
SW	0.4851	0.5149	0.1918	0.7827	0.5789

For the NE monsoon (refer to Figure 4.2), excellent agreements are shown in the observed and theoretical ACFs estimated for DARMA(1,1). The ACFs estimated using the DAR(1) formulation decay abruptly to zero after day 2, which is expected since it is a short persistence model. The theoretical ACFs calculated from the daily rainfall dataset collected for the SW monsoon using the formulation for DARMA(1,1) match the observed ACFs quite well, as shown in Figure 4.3.

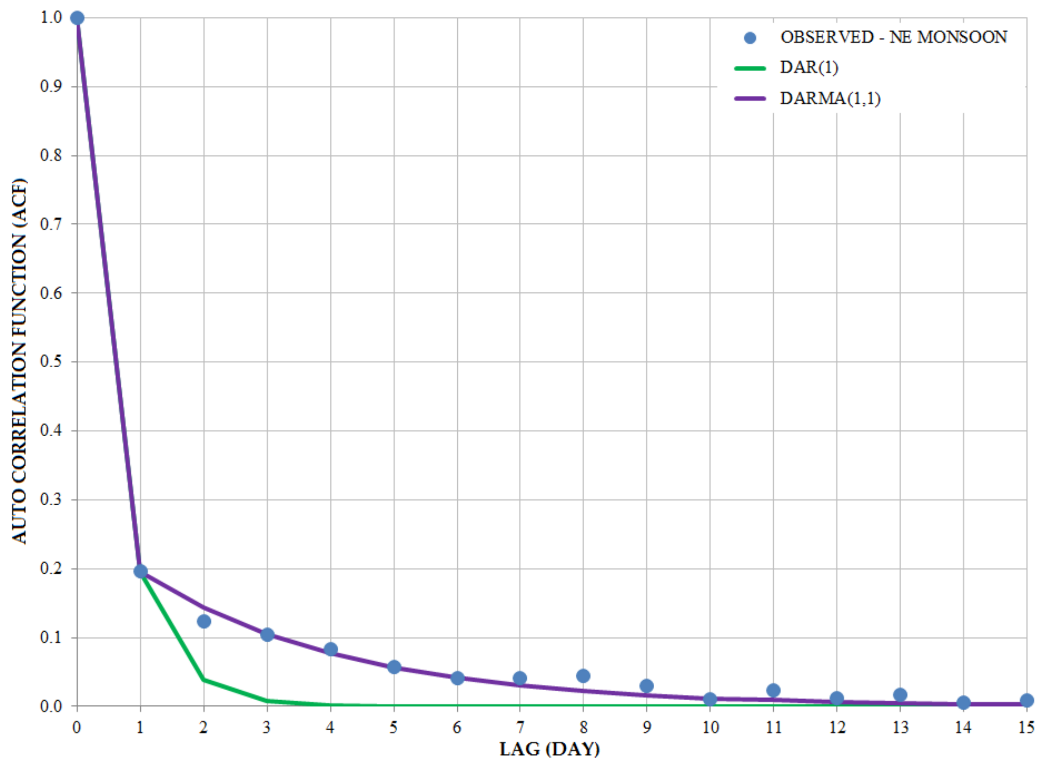

Figure 4.2 Observed and theoretical ACF for NE Monsoon

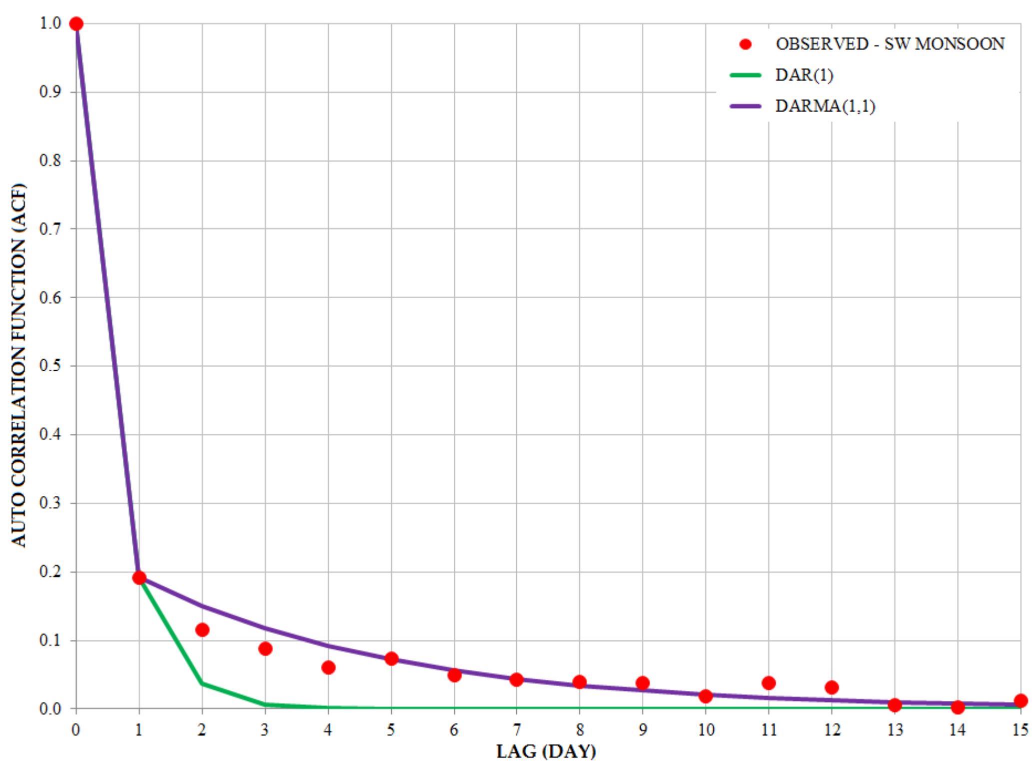

Figure 4.3 Observed and theoretical ACF for SW Monsoon

Overall, the theoretical ACFs for both NE and SW monsoons decay slowly and eventually reach zero at day 15. The behaviour matches the characteristics of a DARMA(1,1) model. This observation confirms the results shown in the previous section, i.e., the DARMA(1,1) model is suitable to represent the sequence of daily rainfall for any season at Subang Airport. The ACF for the DAR(1) model shows that it decays exponentially, as expected.

The parameters determined in this step ease the process of model selection, that is the comparison between the observed and theoretical probability distributions of wet and dry run lengths.

4.1.3 Step 3: Model Selection

Chung et al. (1984a and 1984b) suggested that the model selection process is based on the minimum of sum of squared errors between the observed and theoretical probability distributions of wet and dry run lengths. Wet and dry run lengths are equally important in determining the behavior of the model, therefore both sums of squared errors are added and the model that produces the minimum error will be selected. The model selection is done separately for each season, as shown below.

NE Monsoon

Eq. 2.13 is used to calculate the transitional probabilities for DAR(1), which include:

1) the probability of a dry day followed by another dry day, denoted as p_{00};

2) the probability of a dry day followed by a wet day, denoted as p_{01};

3) the probability of a wet day followed by another wet day, denoted as p_{11};

4) the probability of a wet day followed by a dry day, denoted as p_{10}.

These transitional probabilities are used to calculate the probability distribution function of wet and dry run lengths. The parameters for the DAR(1) model determined in the previous sections lead to the estimated transitional probabilities of $\widehat{p_{00}} = 0.5352$, $\widehat{p_{01}} = 0.4648$, $\widehat{p_{10}} = 0.3392$ and $\widehat{p_{11}} = 0.6608$.

Eq. 2.39 to 2.46 are used to calculate the probability distributions of wet and dry run lengths for the DARMA(1,1) model. The probability distributions of wet and dry run lengths are determined based on the elements in transitional probability matrices, H_0 and H_1. Eq. 2.28 and 2.29 give the details of H_0 and H_1, which are calculated based on the parameters for DARMA(1,1). The transitional matrices H_0 and H_1 for the NE monsoon are given as

$$\mathrm{H}_0 = \begin{bmatrix} 0.6124 & 0.0529 \\ 0.0531 & 0.1911 \end{bmatrix} ; \ \mathrm{H}_1 = \begin{bmatrix} 0.2619 & 0.0727 \\ 0.0386 & 0.7172 \end{bmatrix}$$

Figures 4.4 and 4.5 show the probability distributions of wet and dry lengths from the DAR(1) and DARMA(1,1) models and NE monsoon observations. Both plots show that DARMA(1,1) performs better than DAR(1). The DARMA(1,1) model is able to generate the probabilities with the least amount of error from one to 15 consecutive

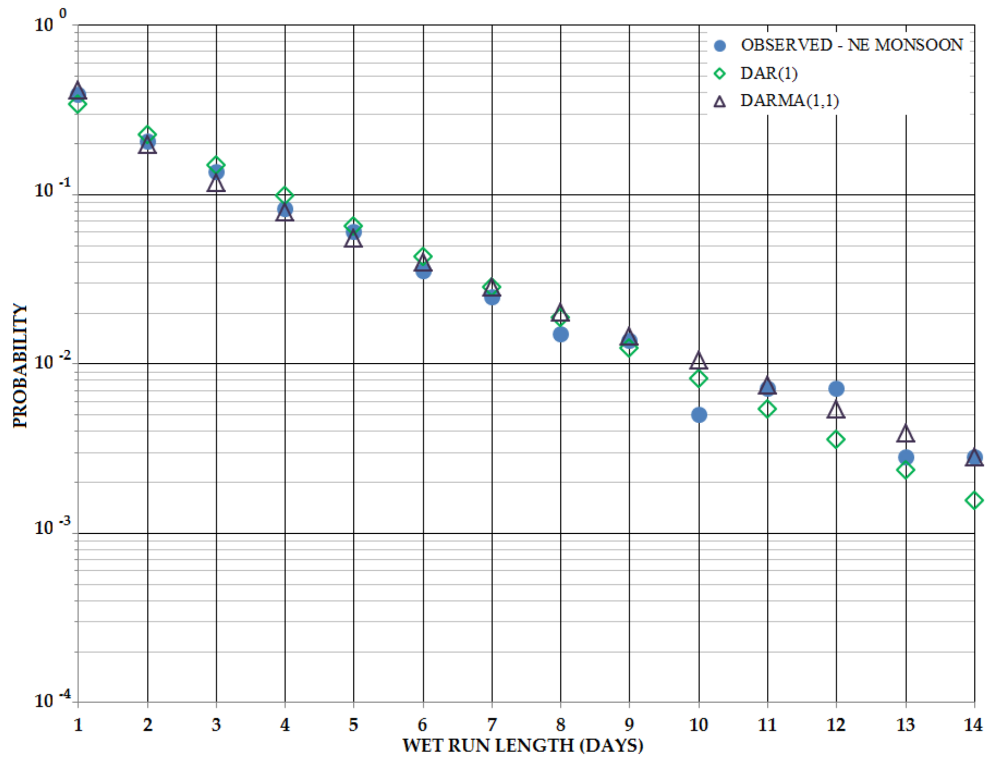

Figure 4.4 Probability distribution of wet run lengths for NE monsoon

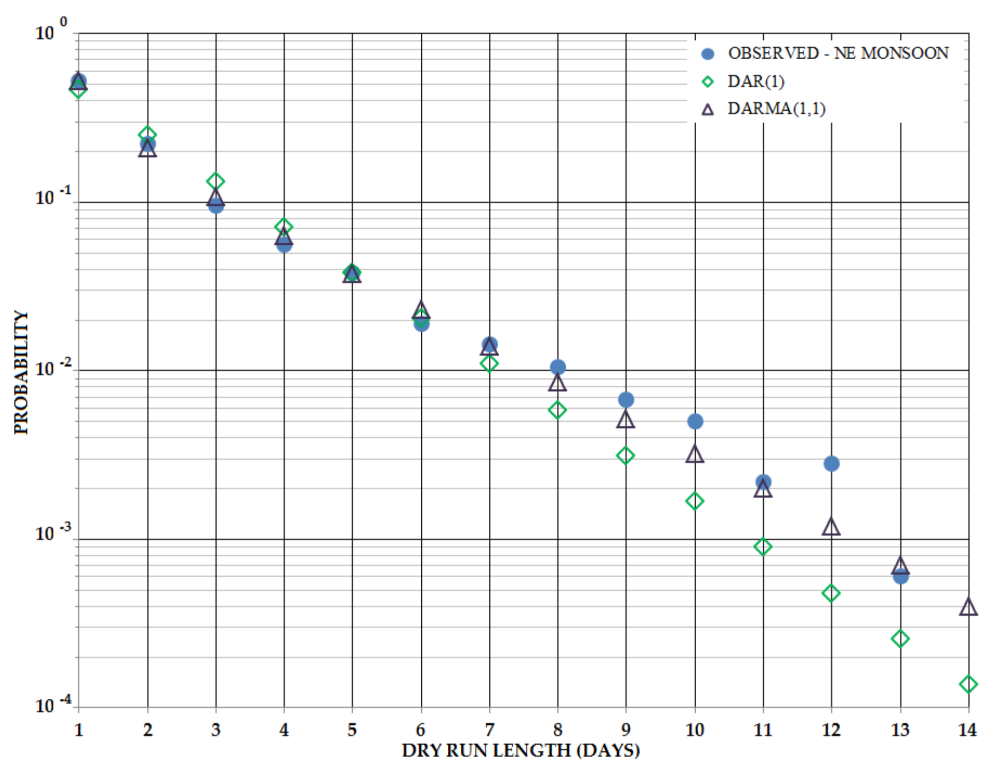

Figure 4.5 Probability distribution of dry run lengths for NE monsoon

rainy days. For example, the probability distribution for 4-consecutive wet days estimated using the theoretical formula for DARMA(1,1) is 0.0792, while the observed dataset gives a probability of 0.0829. Excellent agreement is also observed for the probability distributions of dry run lengths. The theoretical formula for DARMA(1,1) gives a probability of 0.5245 for a single dry day, compared to 0.5236 for the observed dataset.

The DAR(1) model performance is poor, and it can be observed from Figures 4.4 and 4.5 that the errors are bigger as the number of day increased. For instance, the probability of 2-consecutive wet days is 0.2241 using the DAR(1) formula, while the observed value is 0.2066. Additionally, the estimated probability of 3-consecutive dry days using the DAR(1) formula is 0.1331 and the observed value is 0.0961.

The sum of squared errors for wet run lengths given by DARMA(1,1) and DAR(1) is 0.0010 and 0.0033, respectively. The value of error recorded by the DARMA(1,1) model is 3 times smaller as compared to DAR(1). The DARMA(1,1) model also produced smaller error for the probability distributions for dry run lengths when compared to DAR(1). The sum of squared error for dry run lengths estimated using the DARMA(1,1) model is 0.0005, compared to 0.0058 when DAR(1) is used. The details of the sum of squared errors for both wet and dry run lengths are summarized Table 4.2.

Therefore, the DARMA(1,1) model is chosen to simulate the sequences of daily rainfall for the NE monsoon because it gives the least amount of errors for both wet and dry probability distributions. This conclusion also confirms the initial findings reported in model identification and model estimation.

Table 4.2 Sum of squared errors of wet (S_1) and dry run lengths (S_0) for DAR(1) and DARMA(1,1) models during NE monsoon

Model	$\mathbf{S_1}$	$\mathbf{S_0}$	$\mathbf{S_1 + S_0}$	Selection
DAR(1)	0.0033	0.0058	0.0091	DARMA(1,1)
DARMA(1,1)	0.0010	0.0005	0.0015	

SW Monsoon

The same procedures described in the previous section are used in determining the probability distribution of wet and dry run length for the DAR(1) and DARMA(1,1) models during the SW monsoon.

The transitional probabilities for the DAR(1) model during the SW monsoon are $\widehat{p_{00}} = 0.6079$, $\widehat{p_{01}} = 0.3921$, $\widehat{p_{10}} = 0.4161$ and $\widehat{p_{11}} = 0.5839$.

The transitional probability matrices for the DARMA(1,1) model are given below;

$$H_0 = \begin{bmatrix} 0.6748 & 0.0444 \\ 0.0648 & 0.2333 \end{bmatrix} \; ; \; H_1 = \begin{bmatrix} 0.2198 & 0.0610 \\ 0.471 & 0.6548 \end{bmatrix}$$

Figures 4.6 and 4.7 show the probability distribution of wet and dry lengths from the DAR(1), DARMA(1,1) and SW monsoon observations. Both plots show that DARMA(1,1) performs better than DAR(1). As an example, the probability of 3-consecutive rainy days is 0.1257; DARMA(1,1) estimated the value to be 0.1114, while the DAR(1) model gives an estimation of 0.1419. The sum of squared errors for the DARMA(1,1) model is 0.0021, as compared to 0.0079 for DAR(1). From the SW monsoon

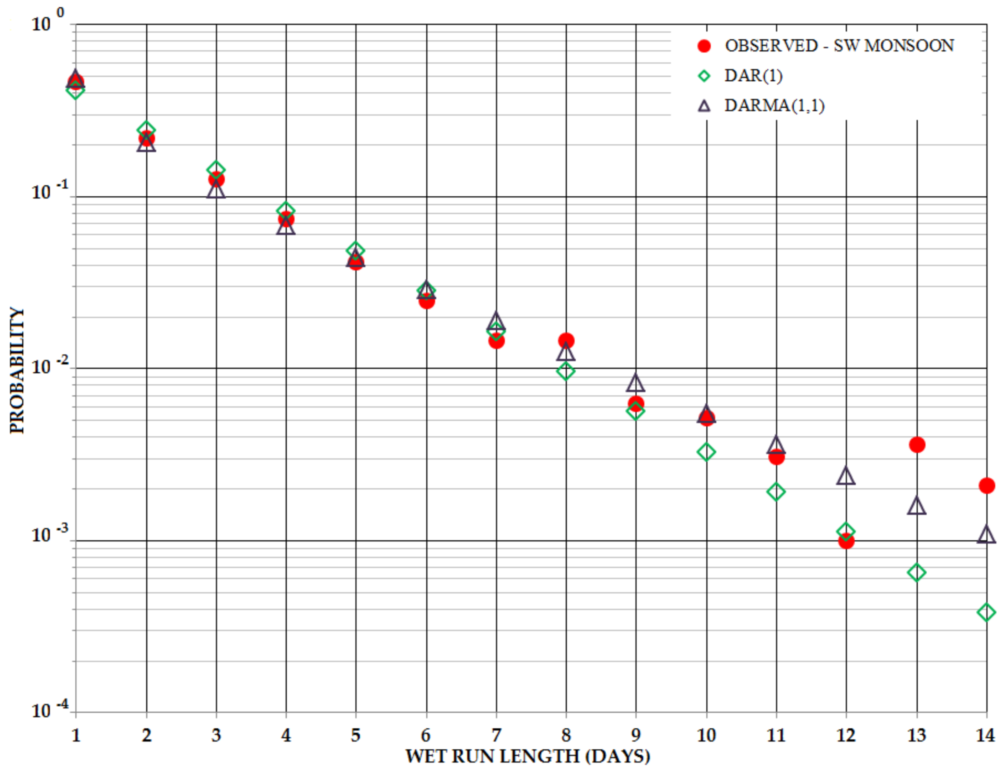

Figure 4.6 Probability distribution of wet run lengths for SW monsoon

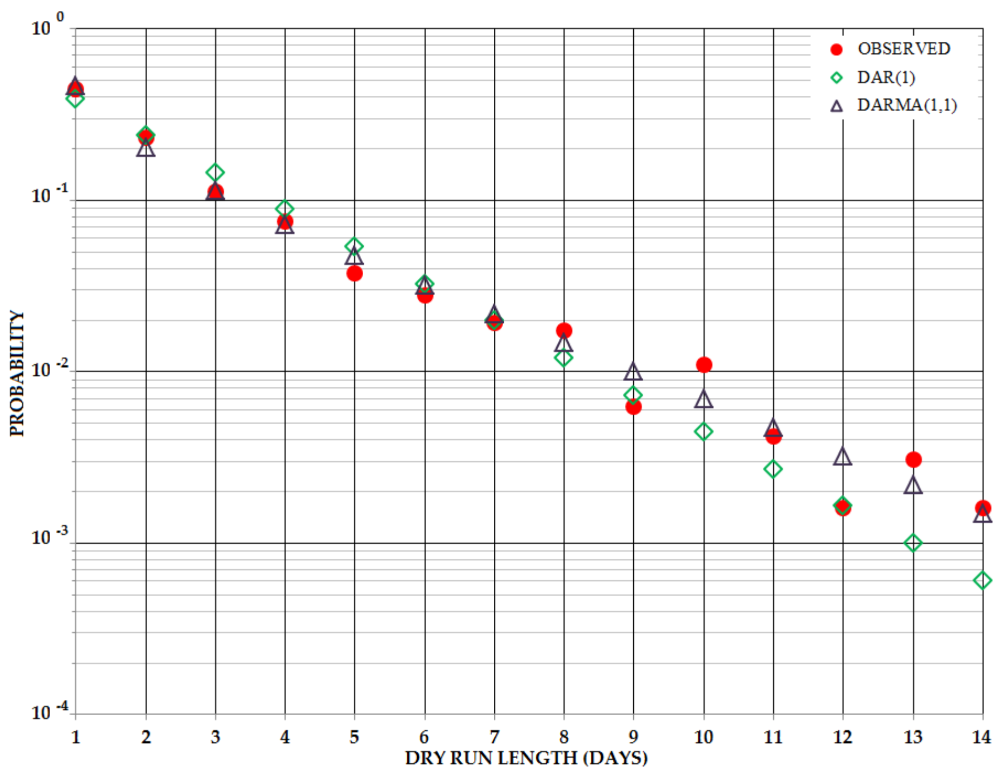

Figure 4.7 Probability distribution of dry run lengths for SW monsoon

observations, the probability distribution of 3-consecutive dry days is 0.1122. The DARMA(1,1) formulation gives a very close estimation to the observed value, i.e., 0.1135. DAR(1) performs poorly, giving the estimated probability of 0.1449 for 3-consecutive dry days.

The sum of squared error for wet run lengths given by DARMA(1,1) and DAR(1) is 0.0009 and 0.0033, respectively. The value of error recorded by the DARMA(1,1) model is almost 4 times smaller as compared to DAR(1). The DARMA(1,1) model also produced smaller error for the probability distributions for dry run lengths when it is compared to DAR(1). The sum of squared error for dry run lengths estimated using the DARMA(1,1) model is 0.0012, compared to 0.0045 when DAR(1) is used. The sum of squared errors for DARMA(1,1) is 0.0021. The DAR(1) model gives a total error almost 4 times larger compared to DARMA(1,1) at 0.0079. The details of the sum of squared errors for both wet and dry run lengths are summarized Table 4.3.

Table 4.3 Sum of squared errors of wet (S_1) and dry run lengths (S_0) for DAR(1) and DARMA(1,1) models during SW monsoon

Model	$\mathbf{S_1}$	$\mathbf{S_0}$	$\mathbf{S_1 + S_0}$	**Selection**
DAR(1)	0.0033	0.0045	0.0079	DARMA(1,1)
DARMA(1,1)	0.0009	0.0012	0.0021	

The findings as discussed in the previous paragraphs clearly indicate that the DARMA(1,1) model is most suitable to simulate the sequences of daily rainfall for the SW monsoon because it recorded the least amount of errors for both wet and dry probability distributions. This conclusion also confirms the initial findings reported in model identification and model estimation.

Conclusion for the model selection process

Significant differences in the sum of squared errors are calculated for both monsoons using the DAR(1) model when compared with DARMA(1,1). Therefore, the DARMA(1,1) model is most suitable to simulate the sequences of daily rainfall for the NE and SW monsoons because it recorded the least amount of errors for both wet and dry probability distributions.

4.1.4 Step 4: Model Verification

A separate verification process is performed for the NE and SW monsoons. The model verification process is done by comparing the probability distributions of wet and dry lengths of the observed and simulated datasets using the Monte Carlo method, with 9,600 days. The wet and dry probability distributions for the generated sequence are estimated using the theoretical formula given by the DARMA(1,1) model.

NE Monsoon

The Monte Carlo method is used to simulate the sequence of daily rainfall during the NE monsoon season, with a sample size of 9,600 days. The DARMA(1,1) parameters estimated from the observed and generated sequence are given in Table 4.4.

Table 4.4 Model Parameters for DARMA(1,1) estimated from observed (NE monsoon) and generated using the Monte Carlo method

Monsoon Seasons	Model Parameters			
	$\hat{\lambda}$	$\hat{\beta}$	$\widehat{\pi_1}$	$\widehat{\pi_0}$
Observed	0.7339	0.5775	0.5781	0.4219
Generated	0.7111	0.6023	0.5725	0.4275

The model parameters estimated from the generated sequence are comparable with the observed data. For instance, the probability of a wet day ($\widehat{\pi_1}$) estimated from the generated sequence is 0.5725, while the observed data give a value of 0.5781. Another example is the $\hat{\lambda}$, where observed data give an estimated value of 0.7330; compared with 0.7111 calculated from the generated sequence.

Figures 4.8 and 4.9 show the verification process for the NE monsoon. Both plots show excellent agreement between the observed (NE monsoon) and the simulated data. This observation concludes that the simulated sequence of daily rainfall is capable of reproducing the parameters and characteristics of the original dataset.

SW Monsoon

Model verification process continues to the SW monsoon dataset. The DARMA(1,1) parameters estimated from the observed and generated sequence are given in Table 4.5.

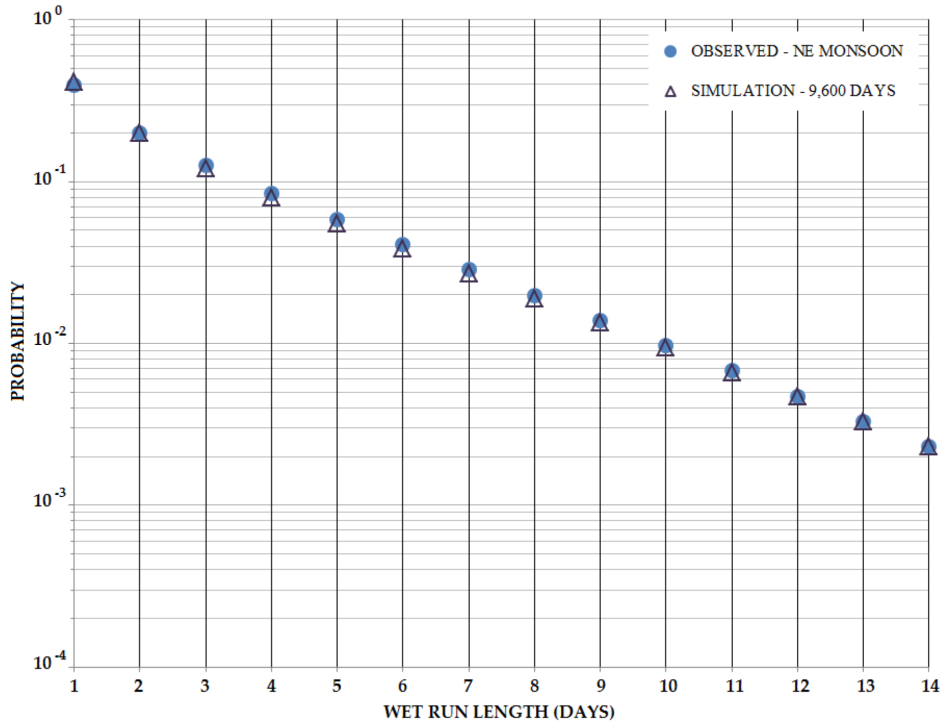

Figure 4.8 Model verification for NE monsoon using the probability distributions of wet run length

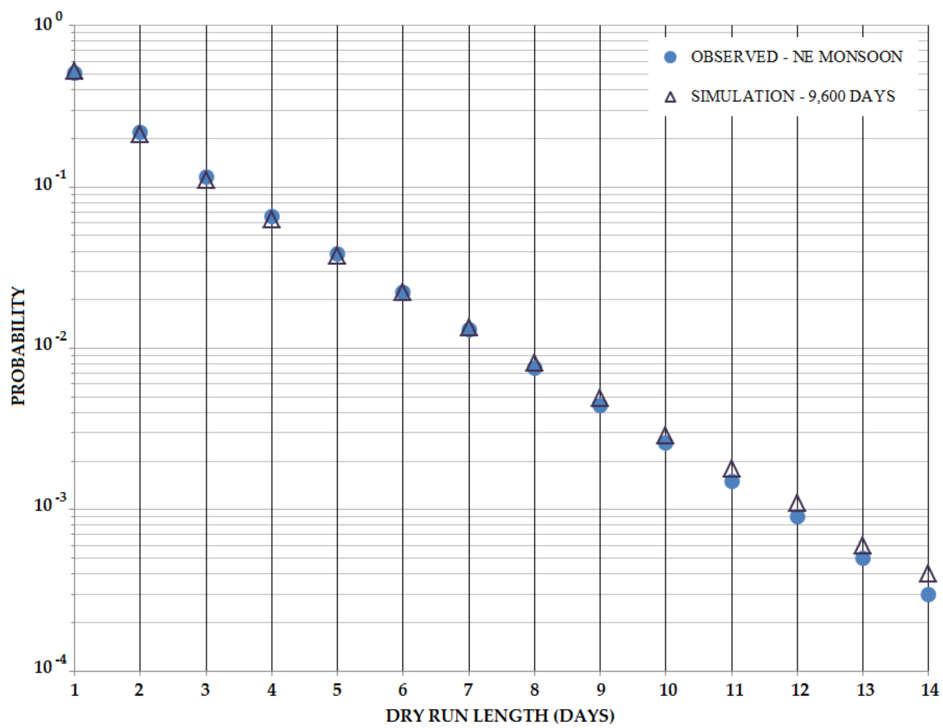

Figure 4.9 Model verification for NE monsoon using the probability distributions of dry run length

Table 4.5 Model Parameters for DARMA(1,1) estimated from observed (SW monsoon) and generated using the Monte Carlo method

Monsoon Seasons	**Model Parameters**			
	$\hat{\lambda}$	$\hat{\beta}$	$\widehat{\pi_1}$	$\widehat{\pi_0}$
Observed	0.7827	0.5789	0.4851	0.5149
Generated	0.7398	0.5765	0.4897	0.5103

The model parameters estimated from the generated sequence are comparable with the observed data. For instance, the probability of a dry day ($\widehat{\pi_0}$) estimated from the generated sequence is 0.5103, while the observed data give a value of 0.5149. Another example is the $\hat{\lambda}$, where observed data give an estimated value of 0.7827, compared with 0.7398 calculated from the generated sequence.

The plots of wet and dry run lengths probabilities are shown in Figures 4.10 and 4.11, respectively. Generally, both plots demonstrate good agreement between the observed and simulated data. There are insignificant errors shown in the longer duration of wet and dry events. This finding confirms that the DARMA(1,1) model is suitable to be used in generating the sequence of daily rainfall during the SW monsoon season.

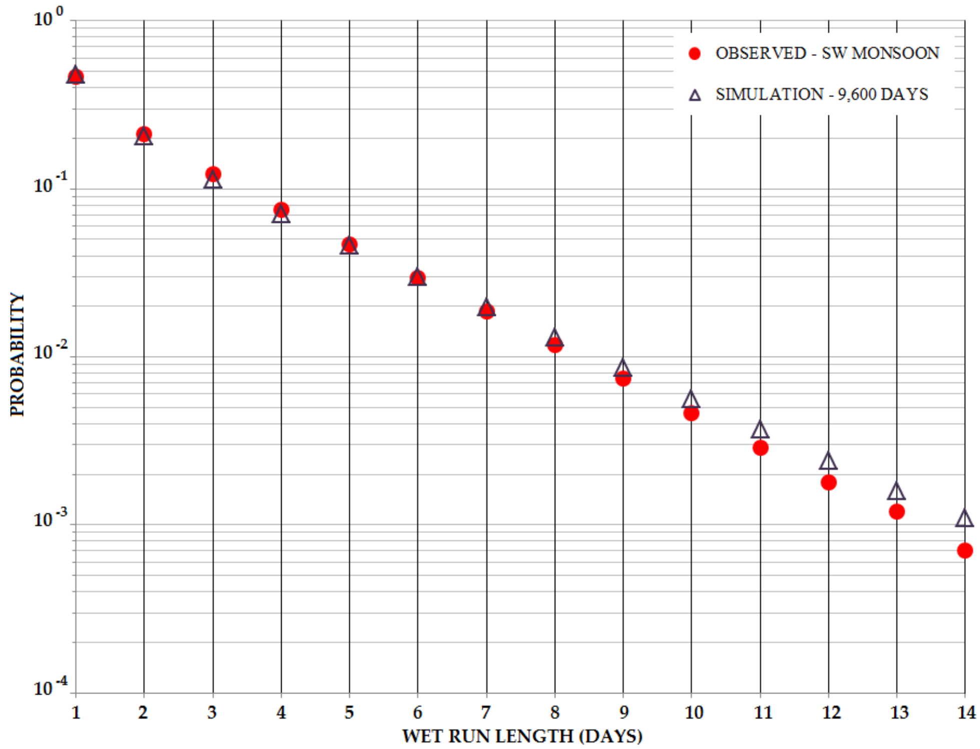

Figure 4.10 Model verification for SW monsoon using the probability distributions of wet run length

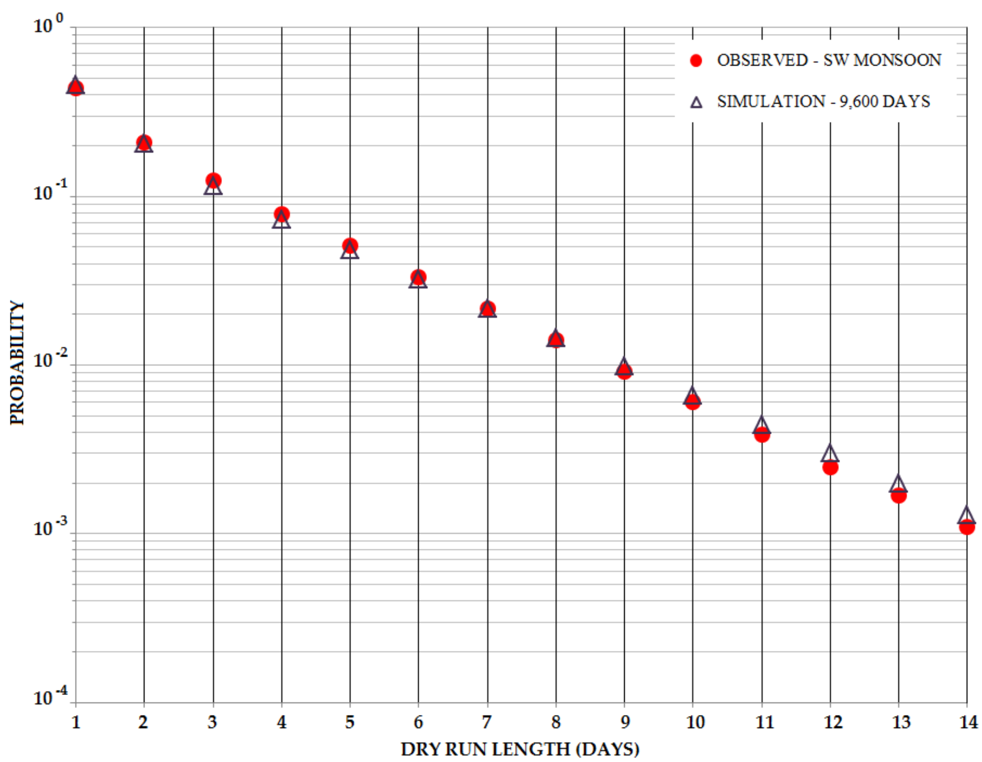

Figure 4.11 Model verification for SW monsoon using the probability distributions of dry run length

4.2 SUMMARY

The four-step model selection procedure leads to the conclusion that DARMA(1,1) is the most suitable model to simulate the sequence of daily rainfall for both scenarios, i.e., NE and SW monsoons at Subang Airport. The occurrences of multi-day rainfall events are common at the study area, therefore a long-term persistence model is required to replicate of the characteristics the observed data.

CHAPTER 5

SEQUENCE OF DAILY RAINFALL AND RETURN PERIOD CALCULATIONS

This chapter discusses the simulation of daily rainfall using a discrete binary time series model, i.e., low order Discrete Auto Regressive and Moving Average [DARMA(1,1)]. Additionally, the rainfall amounts are generated randomly using the two-parameter gamma distribution function. Long sequences of data are simulated for both North East (NE) and South West (SW) monsoons. Return period curves are produced from the generated sequences and compared with the observed data.

5.1 MODELING THE SEQUENCE OF DAILY RAINFALL USING DARMA(1,1)

As shown in equation 2.19, there are a few different components in simulating the sequence of daily rainfall using the DARMA(1,1) model. The first step is generating a sequence of an identical and independent distributed random variable (Y_t), with the discrete probability distribution of π_1 for a wet day (denoted as 1) and π_0 for a dry day (denoted as 0).

The second step is to randomly select the value of U_t, either a 0 or 1. The parameter β is the probability of moving average component, denoted as 1, and the autoregressive component is selected with the probability of (1-β), i.e., when 0 is chosen.

If a moving average component is chosen, ($U_t = 1$), then X_t equals Y_t (as described earlier). On the other hand, $U_t = 0$ indicates that the autoregressive (A_t)

component is selected. For A_t, the sequence is generated with λ and (1- λ) probabilities of 1 and 0, respectively. The regression part is selected when $A_t = 1$, therefore $A_t = A_{t-1}$. A_t of 0 means that there is no regression, and the autoregressive part is $A_t = Y_t$. After these steps are carefully followed, the sequence of daily rainfall generated using the DARMA(1,1) model is simulated. The autoregressive and moving average parts of order 1 are chosen with the probability of (1-β) and β, respectively.

The final step is to randomly generate the rainfall amounts using the two-parameter gamma distribution function (as shown in equation 3.16). The whole process described in this section is done using the Matlab software. This software is chosen because of its availability, ease of use and ability to randomly generate numbers for the sequences of binary time series using the specified probability and also the amount of rain.

In this study, the sequences of daily rainfall are generated separately for the NE and SW monsoons. For each monsoon season, two simulations are done; simulation A and simulation B. Simulation A consists of 100 samples with the size of 9,600 days; while Simulation B is done by generating a sample of 1,000,000 days, which is equivalent to 2,740 years. The summary of both simulations is given in Table 5.1.

Table 5.1 Simulations of daily rainfall for NE and SW monsoons

Simulation	**No. of samples**	**No. of rainfall days**	**Monsoon Seasons**	
			NE	**SW**
A	100	9,600	X	X
B	1	1,000,000	X	X

Three parameters are needed to simulate the daily rainfall sequences using the DARMA(1,1) model, namely λ, β and π_1 (or π_0). The model parameters estimated for the NE and SW monsoons used in simulations A and B are shown in Table 5.2.

Table 5.2 Model Parameters for DARMA(1,1)

Monsoon Seasons	Model Parameters			
	$\hat{\lambda}$	$\hat{\beta}$	$\widehat{\pi_1}$	$\widehat{\pi_0}$
NE	0.7339	0.5775	0.5781	0.4219
SW	0.7827	0.5789	0.4851	0.5149

The main purpose of simulation A is to make sure that the DARMA(1,1) model is capable of reproducing the statistics of the observed data. Therefore, 9,600 days are chosen as the sample size because this is about the same as the observed data for each monsoon. Additionally, 100 samples are produced for simulation A to check the consistency of the DARMA(1,1) model in simulating the daily rainfall sequences at Subang Airport.

Simulation B is done to test the ability of DARMA(1,1) to model a long sequence of daily rainfall and also to produce comparable statistics with the observed data. Therefore, a million days are simulated to represent the long sequence of daily rainfall. Only one sample is generated because the consistency of the model has been tested in simulation A. The statistics of the daily rainfall sequence for simulation are also compared to the observed data. A detailed discussion is given in the next section.

5.2 STATISTICS OF THE OBSERVED AND GENERATED DAILY RAINFALL SEQUENCE AT SUBANG AIRPORT

The relevant statistics of the simulated sequence of daily rainfall using the DARMA(1,1) model are examined in this section. This analysis is important in order to ensure that the simulated sequences are able to reproduce the same statistics as the observed data. The statistics chosen in this study are mean and standard deviation of the amount of rainfall, maximum rainfall in a day, lag-1 Auto Correlation Function (lag-1 ACF) and the maximum wet and dry run lengths. The mean, standard deviation and the maximum daily rainfall are chosen to observe the statistics of the generated rainfall amounts, while the lag-1 ACF and maximum wet and dry run lengths are used to evaluate the statistics of the simulated sequences of daily rainfall. Additionally, further verification process is done comparing the probability of wet and dry run lengths from observed data, Simulation A and Simulation B.

5.2.1 NE Monsoon

Table 5.3 summarizes the statistics of the observed and simulated daily rainfall events at Subang Airport during the NE monsoon. Generally, for simulation A (100 samples of 9,600 days), the statistics for the rainfall amounts generated show excellent results. The mean and standard deviation of daily rainfall are comparable with the observed data. Even though the maximum rainfall in a day is slightly higher than the observed data, it is still acceptable, with a difference of about 4%. Simulation A gives

Table 5.3 Statistics for observed and simulated daily rainfall during NE monsoon

Statistics	NE Monsoon (Observed Data)	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	12.9	0.3	12.9
Standard deviation (mm)	17.6	17.2	0.4	17.3
Maximum rainfall in a day (mm)	171.5	178.9	24.6	292.2
Lag-1 ACF	0.1960	0.1790	0.0116	0.1805
Maximum wet run length (days)	31	24	4	34
Maximum dry run length (days)	21	16	3	25

a reasonable value for the standard deviation of maximum daily rainfall, that is 24.6 mm, which gives the lower and upper bound of 154.3 mm and 203.5 mm, respectively.

Similarly, for simulation B, the long sequence of daily rainfall (a sample of 1,000,000 days) is also capable of reproducing the statistics of observed data. The maximum rainfall in a day is much higher than the observed data, but this is needed in order to perform future predictions for the study area.

The statistics for daily rainfall sequences are also examined in this section. The lag-1 ACFs estimated for the generated sequence are comparable but slightly lower than the observed data. DARMA(1,1) is a long persistence model, therefore the model is capable of producing long sequences of wet and dry days, as shown in Table 5.3.

Figure 5.1 presents the wet run lengths probability distributions from the observed data, simulation A and simulation B. The generated daily sequences (simulations A and B) are capable of reproducing the same values of wet run lengths probability distributions when they are compared with the observed data.

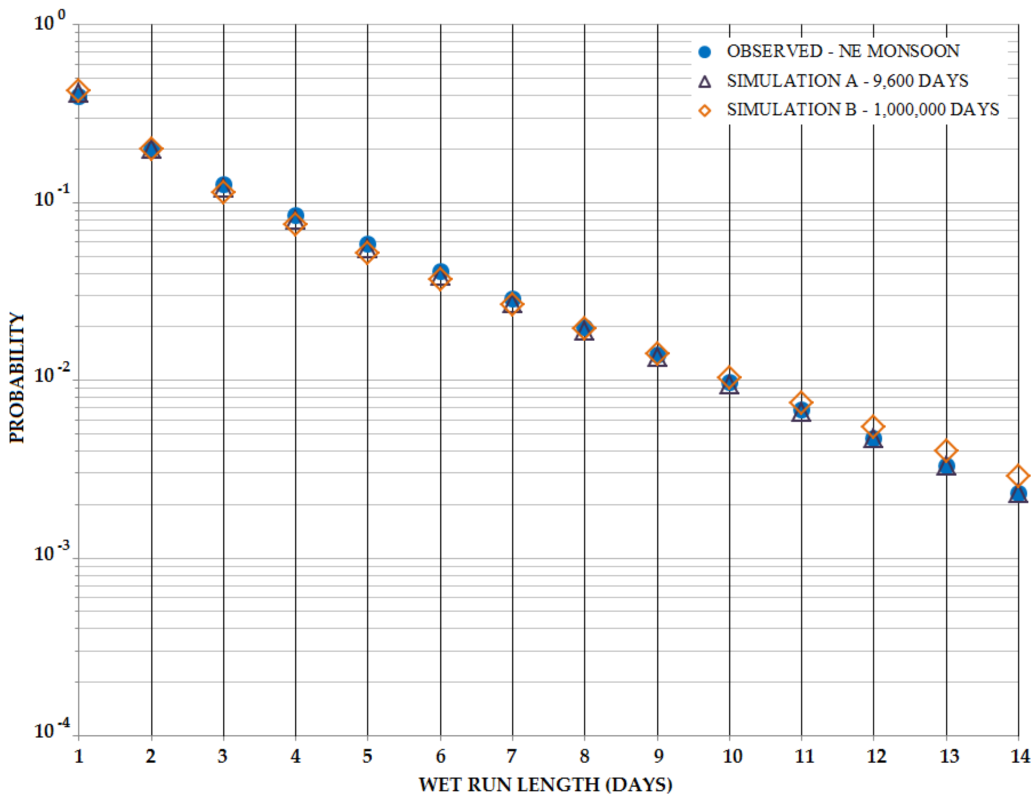

Figure 5.1 Probability distributions of wet run lengths for NE monsoon generated from simulations A and B

The probability distributions of dry run lengths from the observed data, simulation A and simulation B are shown in Figure 5.2. Excellent results are shown, where the generated daily sequences (simulations A and B) are capable of reproducing the same values of dry run lengths probability distributions when they are compared with the observed data.

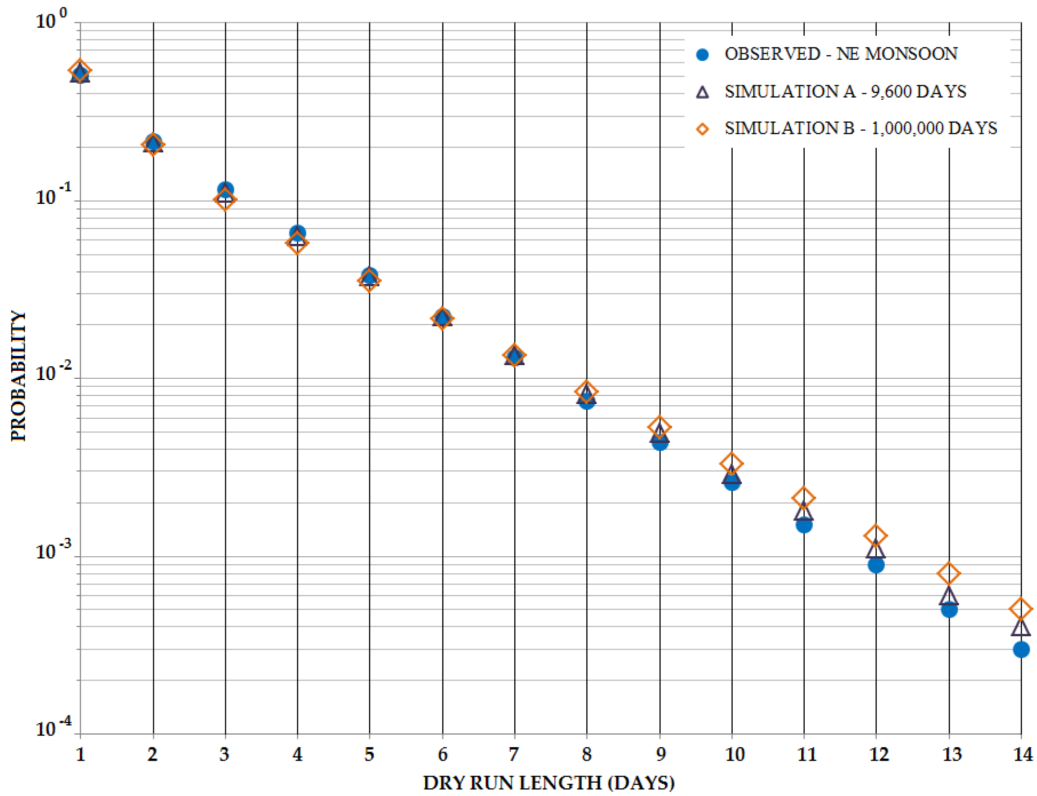

Figure 5.2 Probability distributions of dry run lengths for NE monsoon generated from simulations A and B

5.2.2 SW Monsoon

The statistics for the observed and simulated sequences of daily rainfall during the SW monsoon are given in Table 5.4.

Table 5.4 Statistics for observed and simulated daily rainfall during SW monsoon

Statistics	SW Monsoon (Observed Data)	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)	12.0	12.9	0.3	12.9
Standard deviation (mm)	16.8	17.2	0.4	17.2
Maximum rainfall in a day (mm)	158.3	173.6	26.4	325.1
Lag-1 ACF	0.1918	0.1813	0.0114	0.1798
Maximum wet run length (days)	17	20	3	27
Maximum dry run length (days)	20	20	3	28

The mean and standard deviation for the generated sequences are comparable with the observed data. Maximum daily rainfalls in a day for all simulated sequences are expected to be higher than the observed value. This property is useful for the calculations of return period, which will be discussed later in this chapter.

The maximum wet and dry run lengths given by the simulated sequences are comparable with the observed data. The sequence of 1,000,000 days of generated rainfall shows that the DARMA(1,1) is capable of modeling long wet and dry run lengths. The lag-1 ACF for the generated sequences are slightly lower than the observed data.

The probability distributions of wet run lengths from the observed data, simulation A and simulation B are shown in Figure 5.3. Excellent agreements are shown, which further proves that Simulations A and B are capable of reproducing the

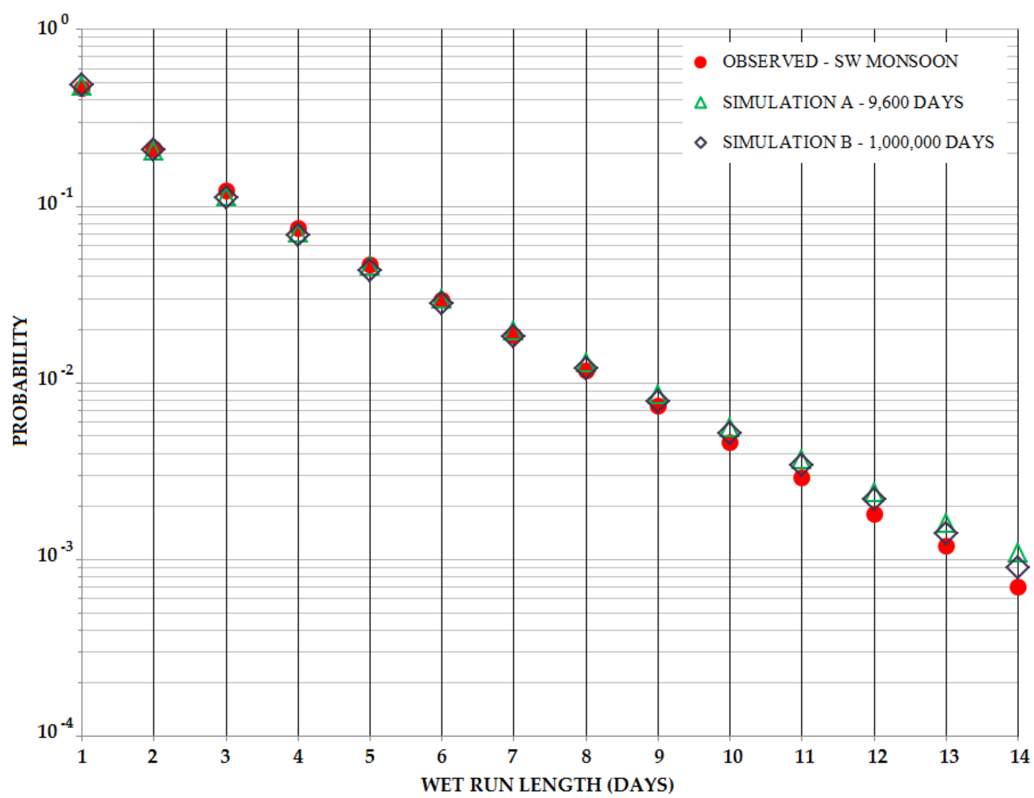

Figure 5.3 Probability distributions of wet run lengths for SW monsoon generated from simulations A and B

same values of dry run lengths probability distributions when they are compared with the observed data.

Figure 5.4 presents the dry run lengths probability distributions from the observed data, simulation A and simulation B. The generated daily sequences (simulations A and B) perform well in terms of reproducing the comparable values of observed dry run lengths probability distributions.

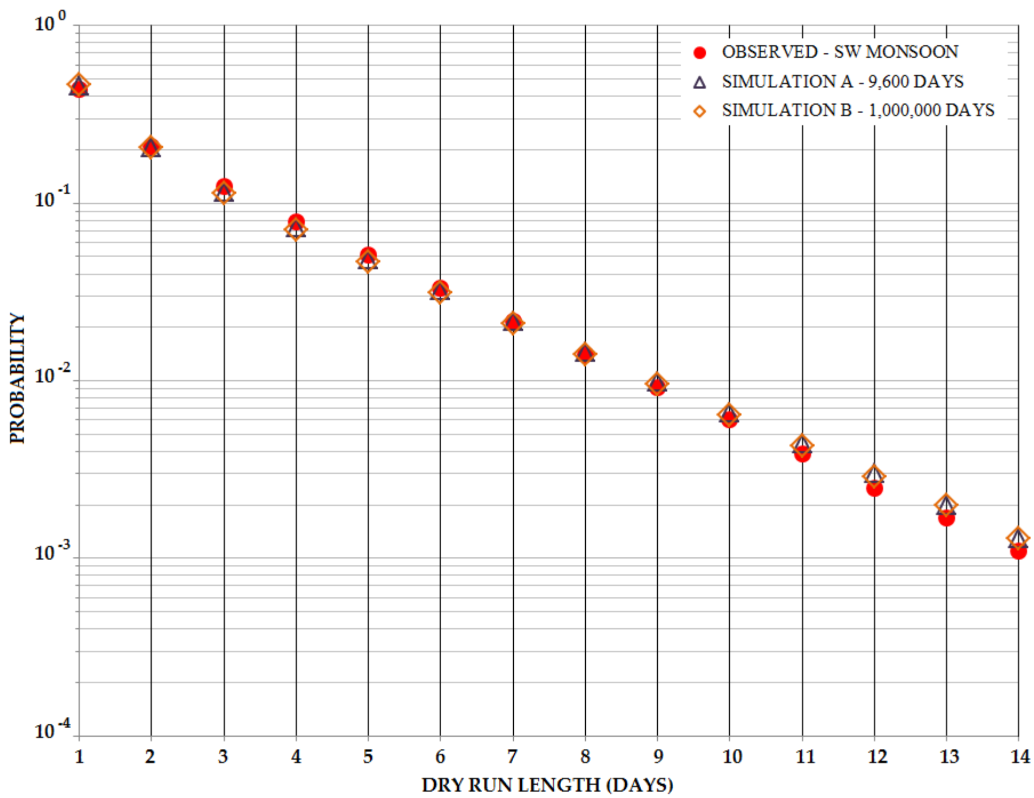

Figure 5.4 Probability distributions of dry run lengths for SW monsoon generated from simulations A and B

5.3 RETURN PERIOD CALCULATIONS

The methods suggested by Shiau and Shen (2001) and Salas et al. (2005) shown in equation 2.50 concentrate on the estimation of return periods for annual drought. This study modified the approach presented in equation 2.50 in order to calculate the return periods of daily rainfall, with the emphasis on multi-day events. Detailed procedures for the estimation of return periods for multi-day events are presented in the following paragraphs.

The bivariate probability distribution functions of rainfall amount and duration are used in order to describe the conditional distribution of both properties. The relationship is presented in Eq. 5.1.

$$f(x,t) = f(x|t)f(t) \qquad \text{(Eq. 5.1)}$$

Where:
- t = number of consecutive rainy days;
- x = total amount of rainfall (mm);
- $f(x,t)$ = bivariate probability distribution function of rainfall amount and duration;
- $f(x|t)$ = conditional distribution of the amount of rainfall given a rainfall duration;
- $f(t)$ = distribution of rainfall duration.

The bivariate probability distribution function of rainfall amount and duration, $f(x|t)$, has been derived in Chapter 3, i.e., the general equation for t-consecutive rainy days. The two-parameter gamma equation is given below;

$$f(x|t) = \frac{1}{24.0\Gamma(0.6t)}\left(\frac{x}{24.0}\right)^{0.6t-1} \exp\left(-\frac{x}{24.0}\right) \qquad \text{(Eq. 5.2)}$$

The distribution of rainfall duration, f(t), has been discussed in detail in Chapter 2 (refer to Eq. 2.39 to Eq. 2.42). To recall, the probability distribution function of wet run lengths is estimated using the equation given below;

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

$$= \frac{P(X_0 = 0, X_1 = 1, \dots, X_t = 1, X_{t+1} = 0)}{P(X_0 = 0, X_1 = 1)} \qquad \text{(Eq. 5.3)}$$

The probability of an event occurring, P(E) for $x \geq x_0$ and t= 1,2,3, ... can be calculated by integrating Eq. 5.4 as shown below;

$$P(E|t) = \int_{x_0}^{\infty} \frac{1}{24.0\Gamma(0.6t)}\left(\frac{x}{24.0}\right)^{0.6t-1} \exp\left(-\frac{x}{24.0}\right). f(t) \qquad \text{(Eq. 5.4)}$$

Thus, the return period is calculated using equation 5.5.

$$T = \frac{\overline{T_1} + \overline{T_0}}{P(E)} \qquad \text{(Eq. 5.5)}$$

Where: $\overline{T_1}$ = mean run length for wet days;

$\overline{T_0}$ = mean run length for dry days;

This study modified the approach used by Cancelliere and Salas (2010) (Eq. 2.53) in order to calculate the return period for 1-day and multi-day rainfall events. Eq. 5.5 gives the best theoretical estimation of return periods for 1-day and multi-day rainfall events, which are shown in the following section.

5.4 RETURN PERIOD CURVES

The return period curves are developed separately for each of the monsoon seasons, i.e., NE and SW. Eq. 5.1 to Eq. 5.5 are used to calculate the theoretical return periods, which are then compared with the observed data. The return periods for the observed data (daily rainfall measurements at Subang Aiport from 1960 to 2011) are estimated by counting. Next, two sequences of daily rainfall, 9,600 and 1,000,000 days long, are generated using the DARMA(1,1) model.

The first sequence is done to make sure that the estimated return periods from a generated sample (which has the same size as the observed data) are comparable with the observed data. The amounts of rainfall (in mm) selected for this analysis are 1, 13, 30, 60, 90, 120 and 150. 1 mm is selected to represent the majority of rainfall events and 13 mm is the average daily rainfall. The remaining amounts are selected because these values are considered as significant rainfall, especially during multi-day events.

The second generated sample is done to represent a long sequence of daily rainfall, i.e., 1,000,000 days (2,740 years). The return period estimations are performed for significant rainfall amounts (in mm), i.e., 50, 100, 150, 200, 250, 300 and 350. These values (more than 150 mm) are chosen to represent rare events.

The details for each analysis are given in the following subsections.

5.4.1 NE Monsoon

Figure 5.5 shows the comparison between the observed and theoretical return periods (estimated using Eq. 5.1 to Eq. 5.5). In general, the theoretical values show good agreement with the observed data. The estimated return periods for multi-day rainfall events for any amounts are higher as compared to the 1-day event. The 1-day events occur more often as compared with 2-consecutive days or more. For higher rainfall amounts (in mm), i.e., 13, 30, 60, 90, 120 and 150, the return periods decreased for several t-consecutive rainy days and increased steadily after that. This trend is observed because more amounts are collected during multi-day events, as compared to a 1-day rainfall. Excellent agreements are shown in the calculation of return periods for multi-day events. For example, the observed and calculated return periods of 4-consecutive rainy days, with the total amount of 60 mm is 211 days.

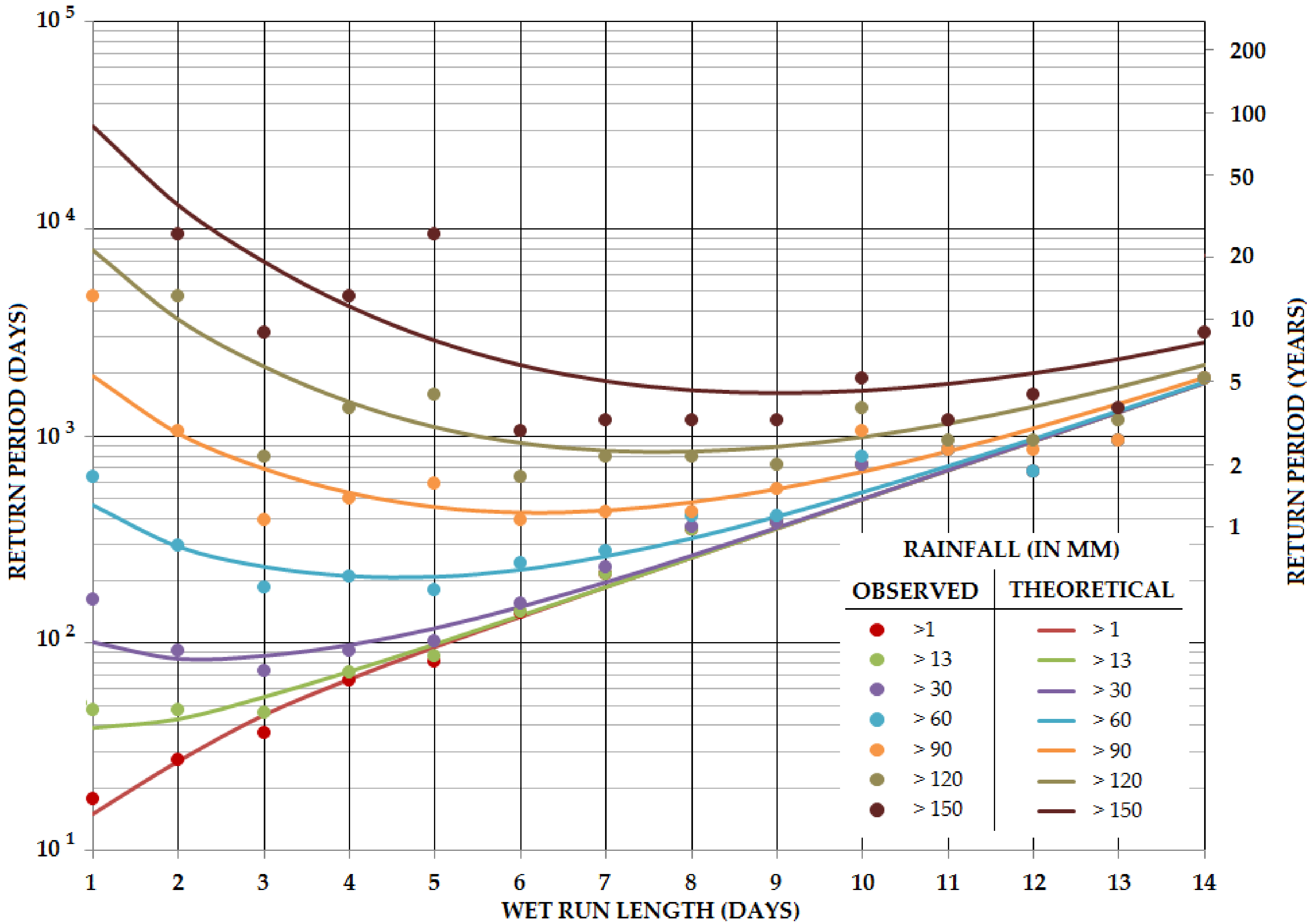

Figure 5.5 Observed and theoretical return periods for NE monsoon

A generated sample of 9,600 days is simulated using the DARMA(1,1) model. The return periods from this generated sample are compared with the observed data, as shown in Figure 5.6. Generally, the estimated return periods from the generated time series are comparable with the observed data. Same trends are observed, i.e., the estimated return periods for multi-day rainfall events (rainfall amounts more than 1 mm) are higher as compared to 1-day event. Other return period curves, i.e., rainfall amounts (in mm) of 13, 30, 60, 90, 120 and 150, show that the estimated return periods decreased for several rainy days and increased steadily after that. The calculation of return periods for multi-day events shows excellent results. For instance, the observed return period of 6-consecutive rainy days, with the total amount of 60 mm, is 243 days (1.37 years), while the calculated value is 233 days (1.34 years). The findings from this analysis show that the generated time-series has the same characteristics as the observed data, and hence is able to represent the return periods very well. Additionally, it also proves that Eq. 5.1 to Eq. 5.5 can be used to estimate the return periods for a generated daily rainfall sequence using DARMA(1,1).

Daily rainfall measurements collected from Subang Airport have limited sample size. Therefore, DARMA(1,1) is used to generate a long sequence of daily rainfall. Furthermore, the occurrences of rare events are also being simulated in this sequence. This section discusses the capability of DARMA(1,1) to give reliable return periods for a long sequence of daily rainfall, i.e., 1,000,000 days.

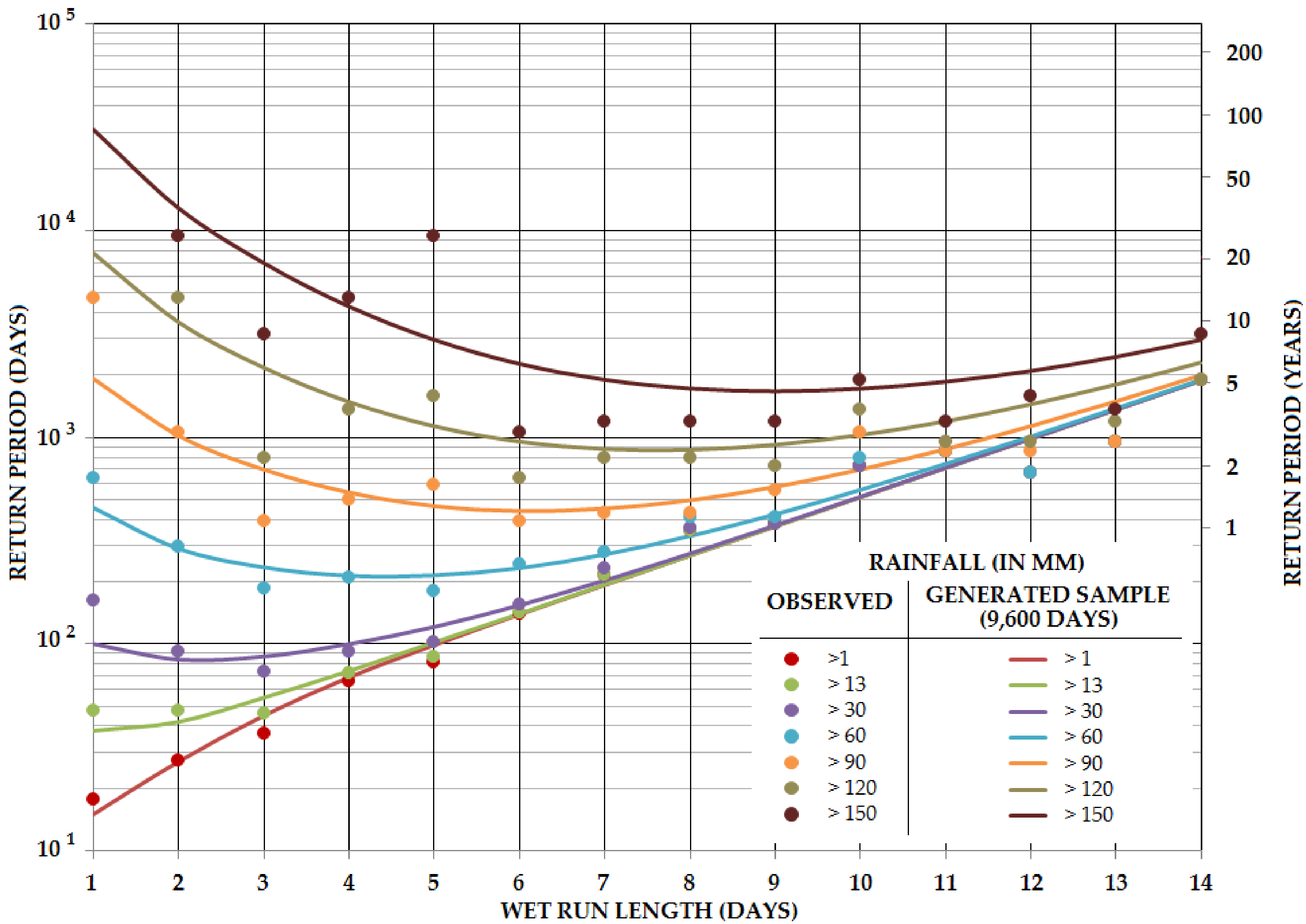

Figure 5.6 Observed and theoretical return periods from generated daily rainfall sequence (9,600 days) for NE monsoon

Figure 5.7 shows the comparison between calculated (by counting) and theoretical (calculated using Eq. 5.1 to Eq. 5.5) return periods. The return period estimations are performed for significant rainfall amounts (in mm), i.e., 50, 100, 150, 200, 250, 300 and 350. The return period curves for all classifications of rainfall amount show excellent agreement, which further verifies that Eq. 5.1 to Eq. 5.5 are reliable to estimate the return periods for multi-day events. For instance, the calculated return period of 5-consecutive rainy days, with the total amount of 200 mm, is 14,280 days (about 39 years), while the theoretical value is 16,100 days (about 44 years).

5.4.2 SW Monsoon

The return periods for various rainfall amounts during the SW monsoon are estimated using the proposed method, as shown in Eq. 5.1 to Eq. 5.5. Figure 5.8 shows that the estimated return periods using the proposed method (theoretical values) are comparable with the observed data. The theoretical return periods for any rainfall event totaling more than 1 mm give excellent agreement with the observed data. Good agreement is shown in the theoretical return periods for multi-day rainfall events. For example, the observed return period of 4-consecutive rainy days, with the total amount of 60 mm is 221 days, and the calculated value is 229 days. That gives 3.6% difference between the observed and theoretical values.

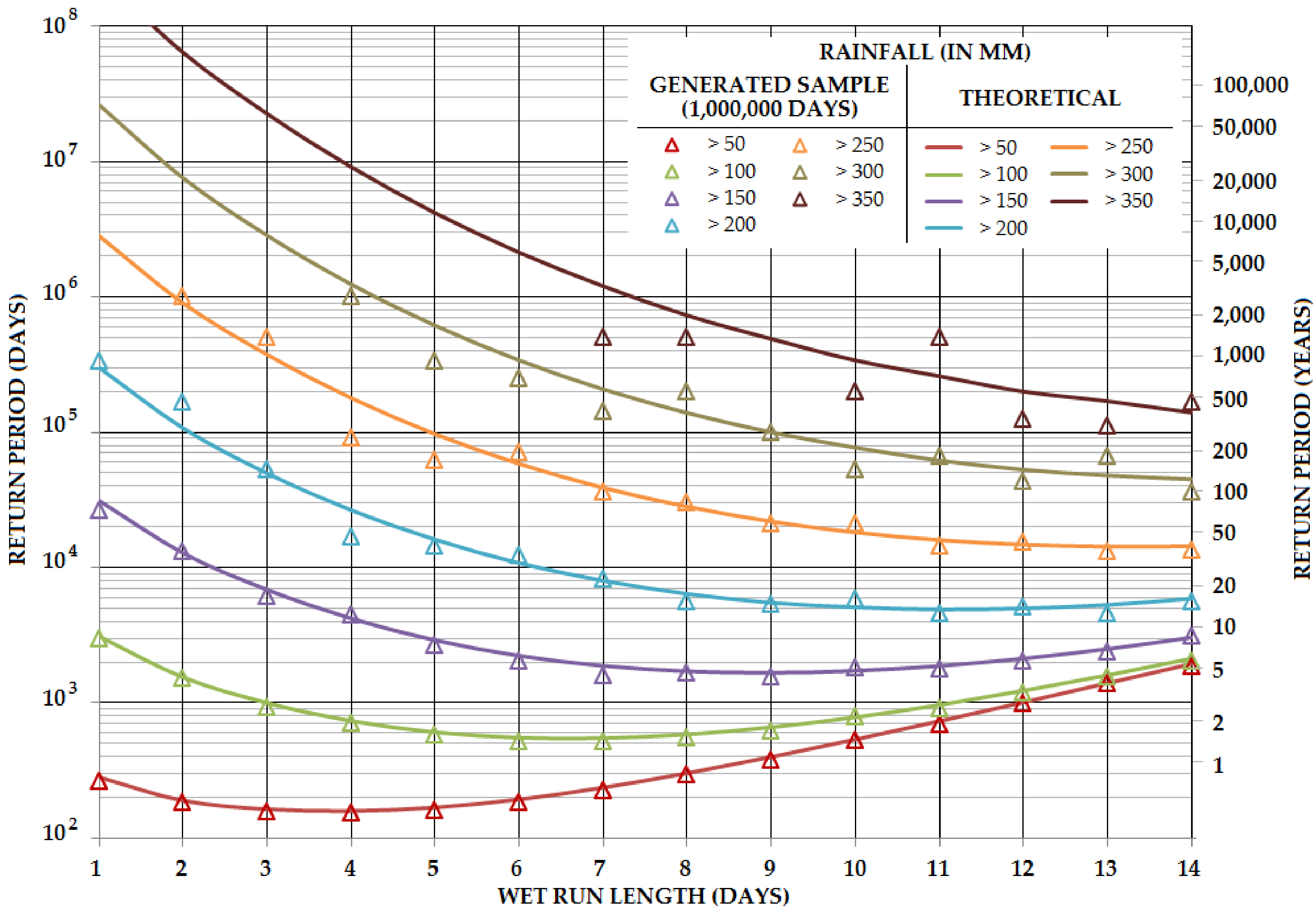

Figure 5.7 Calculated (by counting) from generated daily rainfall sequence (1,000,000 days) and theoretical return periods for NE monsoon

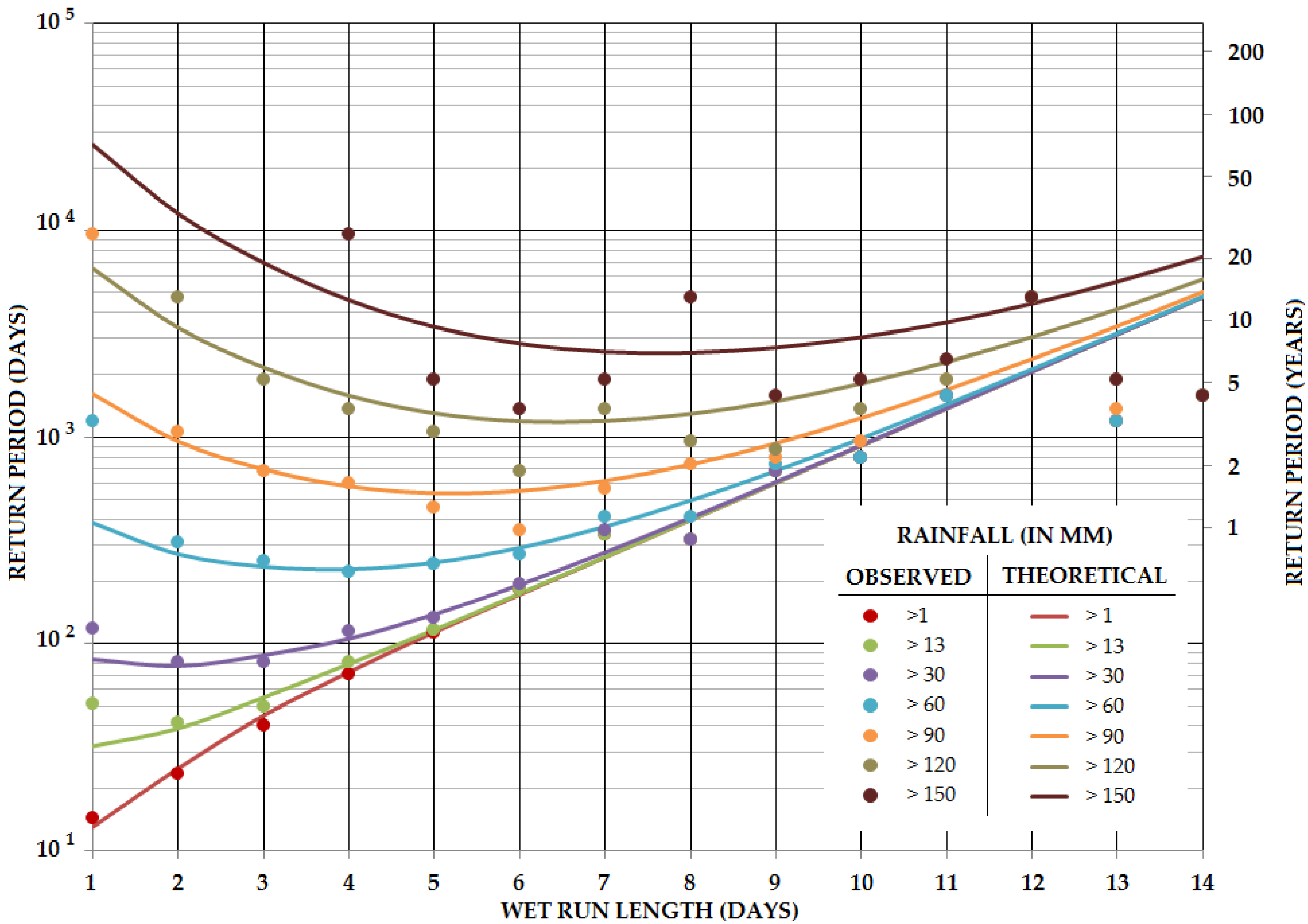

Figure 5.8 Observed and theoretical return periods for SW monsoon

Figure 5.9 shows that the return periods calculated from a generated sequence of daily rainfall are comparable with the observed values. The DARMA(1,1) model is used to generate the sequence of 9,600 days of daily rainfall. The generated return period curves show good agreement with both observed and theoretical values. For instance, the observed return period of 7-consecutive rainy days, with the total amount of 90 mm, is 560 days, while the theoretical value is 549 days.

The DARMA(1,1) model is used to generate a long sequence of daily rainfall during the SW monsoon, i.e., 1,000,000 days. The objective of simulating this sequence is to estimate the return periods for rare rainstorm events.

Figure 5.10 shows the comparison between calculated (by counting) and theoretical (calculated using Eq. 5.1 to Eq. 5.5) return periods. The return period estimations are performed for significant rainfall amounts (in mm), i.e., 50, 100, 150, 200, 250, 300 and 350. The return period curves for all classifications of rainfall amount show excellent agreement, which further verifies that Eq. 5.1 to Eq. 5.5 are reliable to estimate the return periods for multi-day events. For instance, the calculated return period of 4-consecutive rainy days, with the total amount of 200 mm, is 26,310 days (about 72 years), while the theoretical value is 29,100 days (about 80 years).

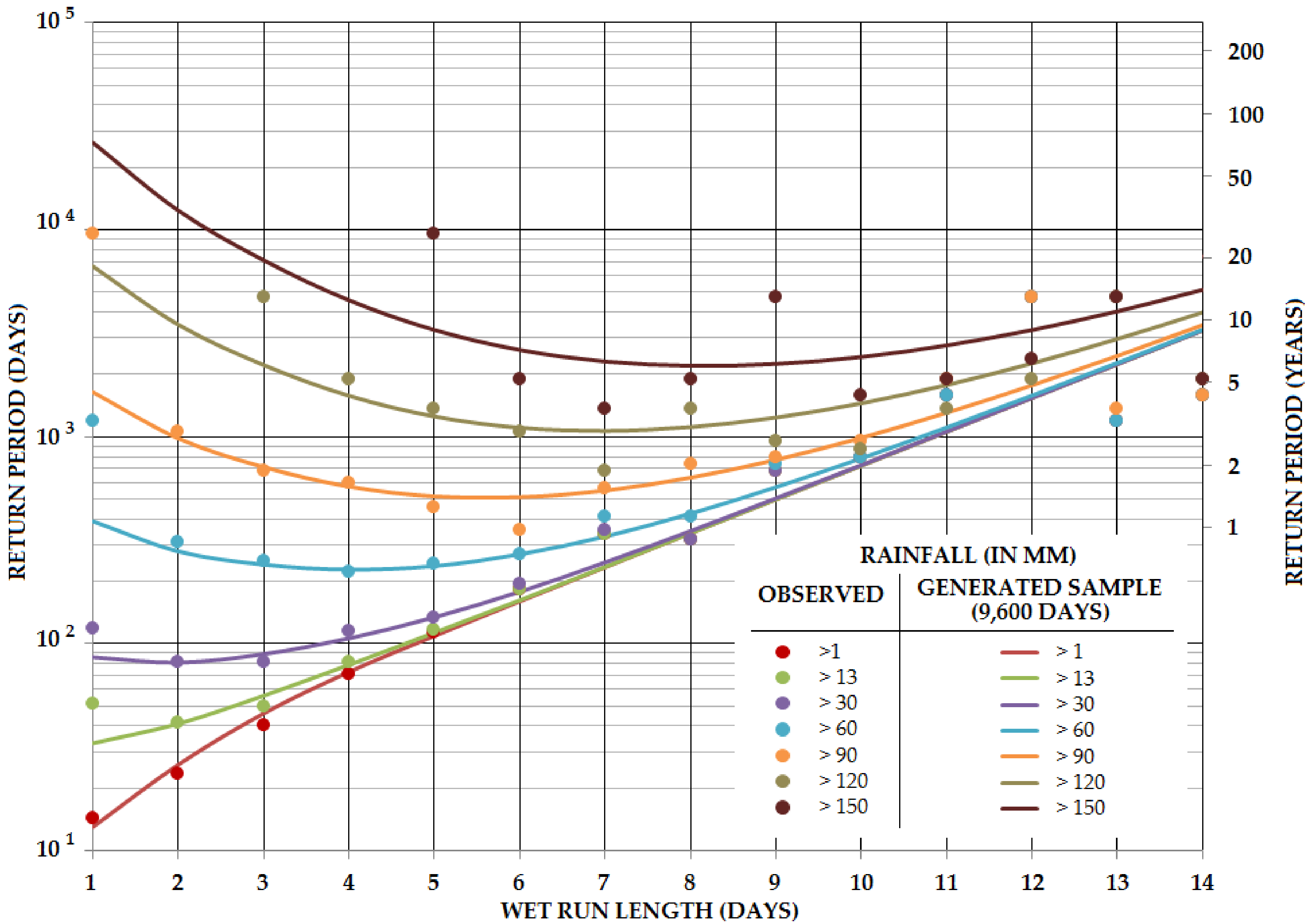

Figure 5.9 Observed and theoretical return periods from generated daily rainfall sequence (9,600 days) for SW monsoon

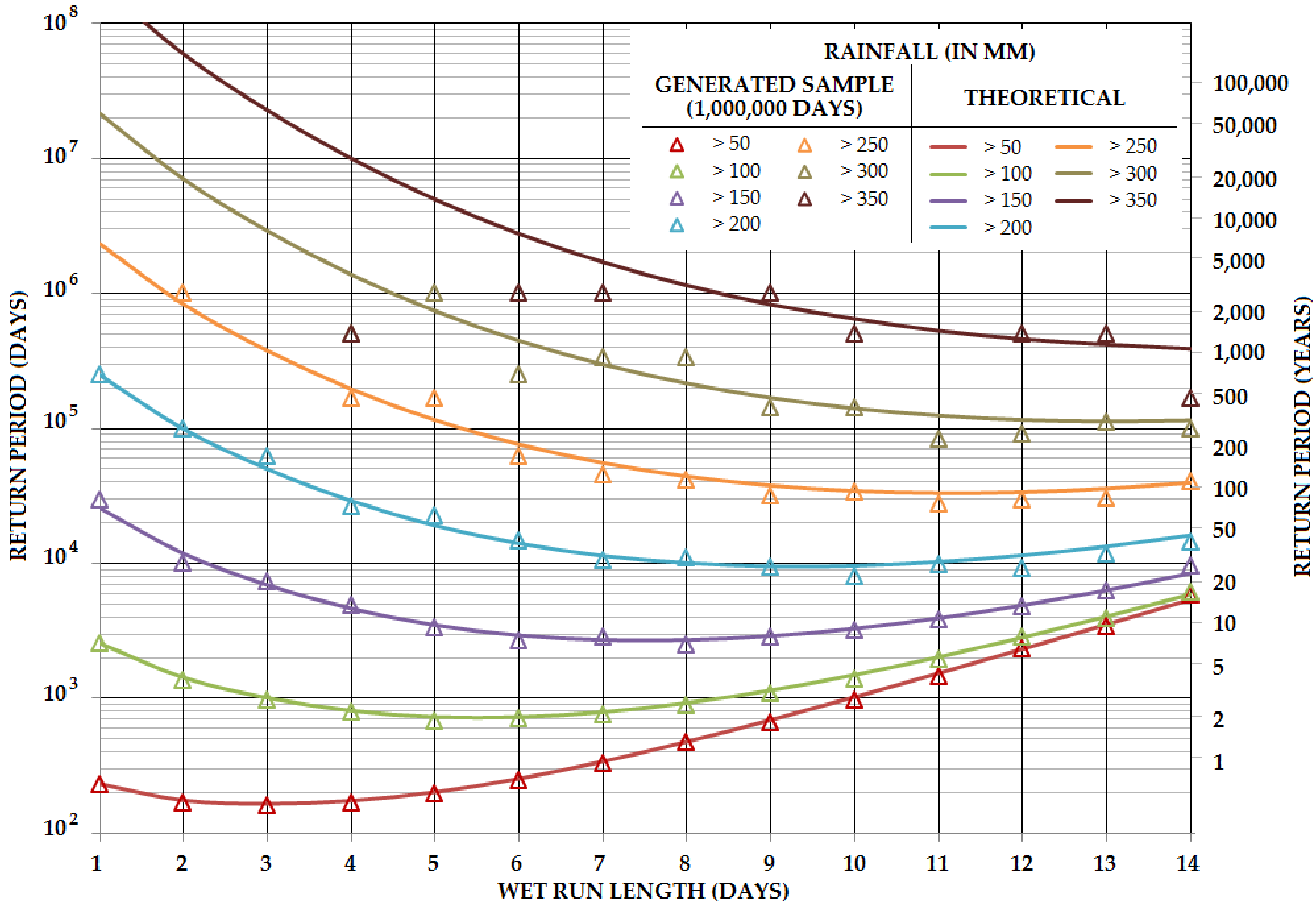

Figure 5.10 Calculated (by counting) from generated daily rainfall sequence (1,000,000 days) and theoretical return periods for SW monsoon

5.5 SUMMARY

The statistical properties of the generated sequences of daily rainfall shows the DARMA(1,1) model capable of reproducing the statistics of the observed data for both NE and SW monsoons at Subang Airport. Additionally, the DARMA(1,1) model is also able to generate a long sequence of daily rainfall, i.e., 1,000,000 days.

The return periods are calculated using the proposed method shown in Eq. 5.1 to Eq. 5.5. Good agreements in the estimation of return periods are shown between the observed, theoretical (calculated) and generated daily rainfall sequence. Return period curves for rare rainstorm events (rainfall amount of more than 150 mm) are also produced using a long sequence of daily rainfall.

CHAPTER 6

MODEL APPLICATION: MULTI-DAY MONSOON RAINFALL EVENTS AT KOTA TINGGI WATERSHED

Multi-day rainfalls are common in Malaysia and the occurrences of these events can be simulated using the DARMA(1,1) model. This section discusses the return period estimation for multi-day rainstorms using the methods and algorithms that have been developed in this study (as shown in Chapter 5). The most recent multi-day rainstorms in the city of Kota Tinggi, Johor are used as an example. These events occurred in December 2006 and January 2007 resulting in more than 350 and 450 mm of cumulative rainfall.

6.1 KOTA TINGGI RAINSTORMS

Kota Tinggi is located in the central part of the state of Johor. The Kota Tinggi watershed has an area of 1,639 km^2 and numerous rivers and tributaries with total channel length of 122.7 km. The location of this study area and the rivers are shown in Figure 6.1.

Kota Tinggi receives significant amounts of rainfall, and the total annual average is 2,470 mm. There were historical floods recorded in 1926, 1967, 1968 and 1971 (Badrul Hisham et. al. 2010). However, the worst floods were reported recently in December 2006 and January 2007, which occurred 3 weeks apart. An economic loss of

Figure 6.1 The location of Kota Tinggi and water bodies surrounding it (after Shafie 2009)

RM1.5 billion (equivalent to about half billion U.S dollars) occurred, and more than 100,000 local residents have to be evacuated during both events (Abu Bakar 2007).

The severe floods in December 2006 and January 2007 are the results of 5 and 4 consecutive rainy days, respectively. Table 6.1 gives the total amount of daily rainfall at several gaging stations for these events. For the December 2006 event, most of the stations recorded an accumulated amount of close to 100 mm for 2-consecutive days. A significant amount of rain was recorded on the 3rd day, December 19, 2010. Figure 6.2 shows the rainfall gage stations around Kota Tinggi and the amount of rainfall measured in a day on December 19, 2006. The highest rainfall was recorded at Bukit Besar station, with 200 mm, and this measured value is the same as the average monthly rainfall. The Ulu Sebol station, which is located in the northeastern part of the Kota Tinggi watershed, recorded 189 mm of rainfall on December 19, 2006. Other

Table 6.1 Total amount of daily rainfall recorded at several gaging stations around Kota Tinggi during December 2006 and January 2007 floods (after Shafie 2009)

Date	Layang-layang	Ulu Sebol	Bukit Besar	Kota Tinggi
		December 2006		
Dec-17	66 mm	33 mm	29 mm	48 mm
Dec-18	52 mm	23 mm	47 mm	43 mm
Dec-19	156 mm	189 mm	200 mm	161 mm
Dec-20	73 mm	78 mm	69 mm	39 mm
4 days total	367 mm	353 mm	345 mm	287 mm
		January 2007		
Jan-11	145 mm	124 mm	147 mm	167 mm
Jan-12	135 mm	290 mm	234 mm	122 mm
Jan-13	84 mm	76 mm	42 mm	49 mm
Jan-14	20 mm	44 mm	35 mm	-
4 days total	384 mm	534 mm	458 mm	338 mm

stations also recorded significant amount of rainfalls and these values are given in Figure 6.2.

The January 2007 flood was more severe than the December 2006 event. Figure 6.3 shows the satellite images of a band of clouds from 11^{th} to 14^{th} January, 2007. The Kota Tinggi watershed received a significant amount of rainfall for 4 consecutive days from these clouds. The maximum magnitude of rainfall was recorded for the first two days, i.e., January 12 – 13, 2006. For example, the accumulated rainfall for two days in Ulu Sebol station was 366 mm, which is almost double the average monthly rainfall. This station also recorded the highest total rainfall for the 4-consecutive rainy days, with 534 mm. In general, the gaging stations in Kota Tinggi recorded an average total rainfall of more than 400 mm.

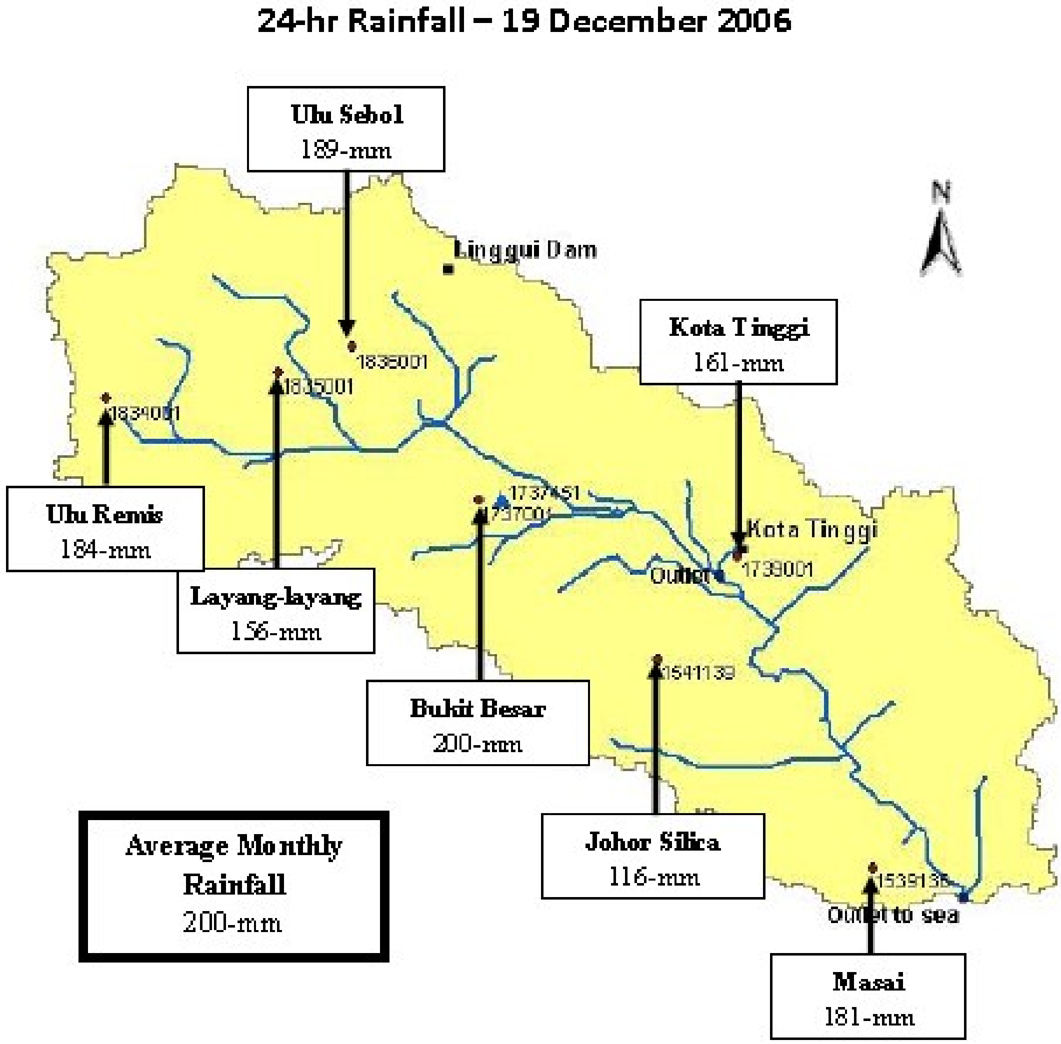

Figure 6.2 Rainfall gage stations around Kota Tinggi and the amount of daily rainfall on December 19, 2006 (after Shafie 2009)

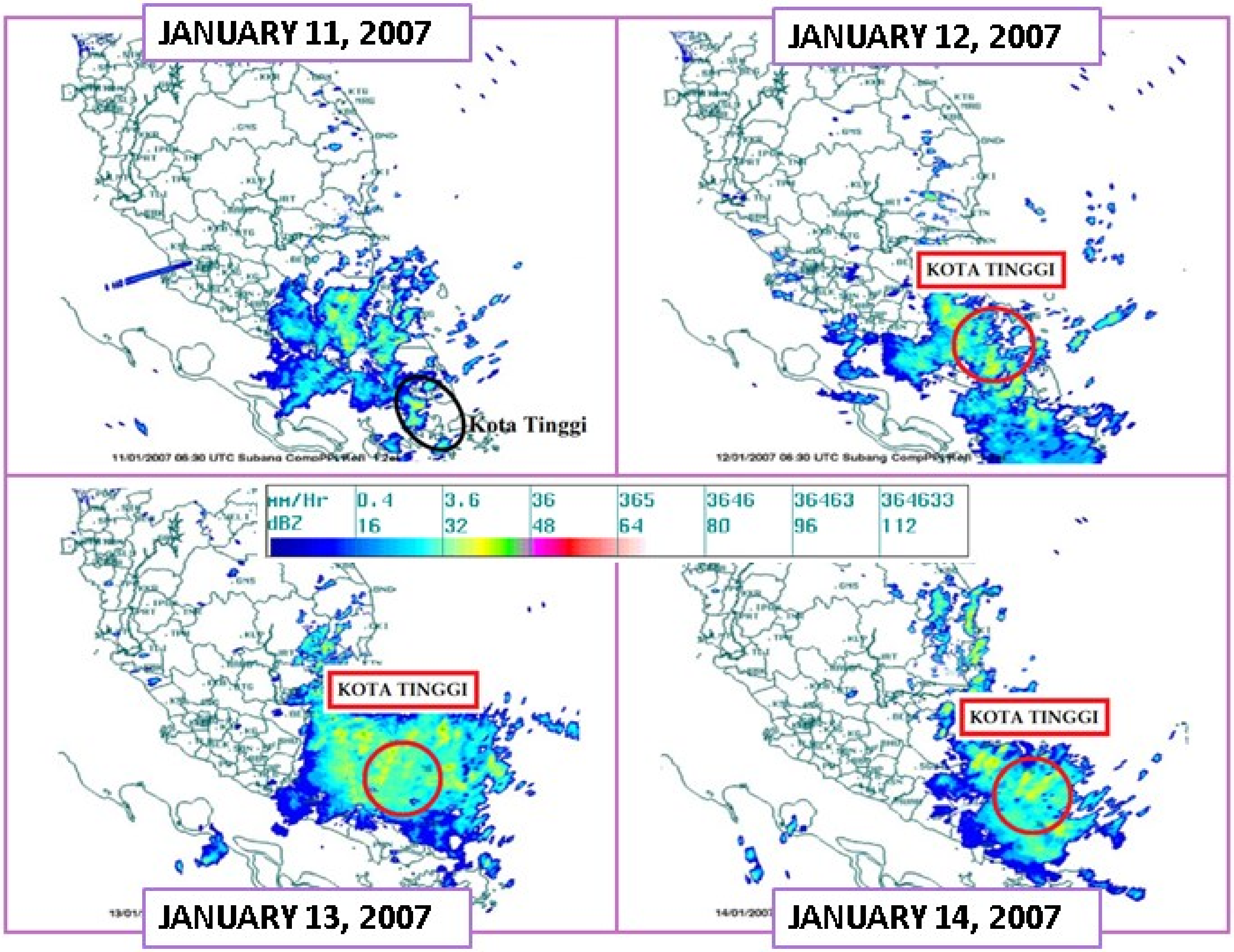

Figure 6.3 Satellite images rainfall distribution (modified from Shafie 2009)

6.2 ESTIMATION OF RETURN PERIODS FOR KOTA TINGGI RAINSTORMS

The procedures and algorithms presented in Chapter 5 are used to estimate the return periods of the December 2006 and January 2007 rainstorm events. The detailed discussions of the return periods pertaining to the December 2006 rainstorm event are given in the following section. The estimation of return periods for the January 2007 storm is given in Appendix A.

6.2.1 RETURN PERIODS FOR THE DECEMBER 2006 RAINSTORM EVENT

Table 6.2 summarizes the estimation of return periods for the December 2006 rainstorm event using Eq. 5.1 to Eq. 5.5. Rainfall measurements from four rainfall gaging stations are used: Layang-Layang, Ulu Sebol, Bukit Besar and Kota Tinggi.

The highest rainfall on the first day was measured at the Layang-Layang station, with 66 mm. This rainfall amount corresponds to the return period of 2 years. Other stations recorded small rainfall amounts, and the return periods estimated for these measurements are less than 1 year.

The amounts measured for the second (Dec 18), third (Dec 19) and fourth (Dec 20) of the multi-day rainstorm events are much more significant as compared to the first day. Layang-Layang recorded the most rainfall with the estimated return period of 8 years, followed by Kota Tinggi (3 years), Bukit Besar (1.5 years) and Ulu Sebol (0.7 years). These values continue to increase on the third and fourth day. Most of the stations recorded the rainfall amount with return period of more than 1,000 years. Bukit Besar station received 276 mm on the third day, which corresponds to 2,750 years

Table 6.2 Estimation of return periods for the December 2006 rainstorm event

Date	December 2006 Dec-17	Dec-18	Dec-19	Dec-20
Station : Layang-layang				
Cumulative Rainfall (mm)	66	118	274	347
Return period (years)	2	8	2,534	20,575
Station : Ulu Sebol				
Cumulative Rainfall (mm)	33	56	245	323
Return period (years)	0.3	0.7	778	7,910
Station : Bukit Besar				
Cumulative Rainfall (mm)	29	76	276	345
Return period (years)	0.3	1.5	2,750	19,013
Station : Kota Tinggi				
Cumulative Rainfall (mm)	48	91	252	291
Return period (years)	0.7	3	1,036	2,247

of return period. On 20th December, 2006, the Kota Tinggi watershed received between 291 to 347 mm of rainfall. The return periods measured from these stations are greater than 2,000 years.

6.3 KOTA TINGGI FLOODS

The water levels at Sungai Johor gaging station during the December 2006 rainstorm event are illustrated in Figure 6.4. Figure 6.4A shows the water depth one day before the event, which is at the normal level. The water level increases significantly to an alert level on December 19, 2006, as shown in Figure 6.4B. Figures 6.4C and 6.4D show the flooding as a result of the multi-day rainfall events. The stage recorded reached the danger level of 2.75 m, making it the highest level ever recorded since 1950, resulting in a declared emergency curfew (Badrul Hisham et. al. 2010).

Figure 6.5 shows the flood levels observed during December 2006 and January 2007 at the same location. On December 21, 2006, one day after the multi-day rainstorms stopped, the stage was at the same level as the flood in 1948 (refer to Figure 6.5a). Figures 6.5b to 6.5d show flood level for the 4 consecutive days of rainstorms in January 2007. On January 12th, 2007, i.e., day 2 of the multi-day rainstorms, the flood level exceeds the December 2006 rainstorm and also the historical event in 1948 (refer to Figure 6.5b). The water level continues to rise the third day, as shown in Figure 6.5c. Figure 6.5d shows that the flood has subsided 5 days after the multi-day rainstorm occurred.

Figure 6.4 Water level indicators [A] On Dec. 18, 2006 – 14:56, [B] On Dec. 19, 2001 – 08:01 [C] On Dec. 20, 2006 – 08:01 and [D] On Dec. 21, 2006 – 08:16 (after Shafie 2009)

a) 21 December 2006 b) 12 January 2007

c) 13 January 2007 d) 19 January 2007

Figure 6.5 Water level indicators a) On Dec. 21, 2006 b) On Jan. 12, 2007 c) On Jan 13, 2007 and d) On Jan 19, 2007 (after Shafie 2009)

6.4 HYDROLOGICAL MODELING AT KOTA TINGGI

Abdullah (2013) simulated the flood events at Kota Tinggi using the Two-dimensional Runoff, Erosion and Export (TREX) model. The 1,635 km^2 watershed area was discretized using a grid size of 230 m.

Figure 6.6 shows detailed water depths at Kota Tinggi watershed using 3-dimensional representation when the water level reached the alert level. The stage continued to increase and easily passed the alert and danger level as a result of the continuous rainfall. Figure 6.7 gives the 3-dimensional representation of the flooding areas at Kota Tinggi watershed on December 21, 2006. The maximum stage was reached on December 22, 2006, 2 days after the rainfall stopped.

The TREX model was able to simulate the hydrological conditions of the study area with reasonable accuracy, as shown in Figure 6.8. The calibration process was done using the historical storm event that occured from November 23 to December 4, 2010. The observed daily discharge and stage are provided by the Department of Irrigation and Drainage (DID).

The validation process was performed using the stage data from December 14, 2006 to January 25, 2007. The comparison between observed and simulated stage for these events is presented in Figure 6.9. The validated model shows that the multi-day rainfall event in December 2006 passed the normal level after 2 days.

The stage increased more rapidly during the second event in January 2007. The stage increased to the alert and danger level after one day of rainfall. This condition is driven by the high intensity of rainfall for 2 consecutive days. The maximum stage was

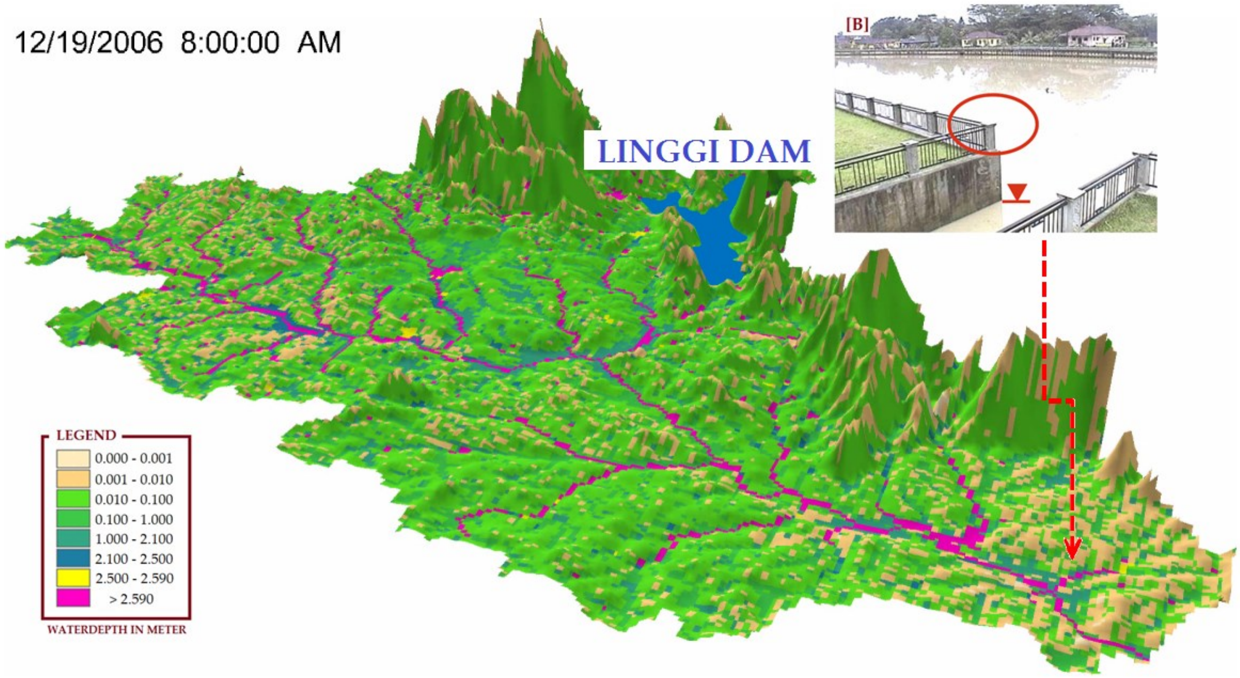

Figure 6.6 3-dimensional representation of the water depths at Kota Tinggi watershed on December 19, 2006 (adapted from Abdullah 2013)

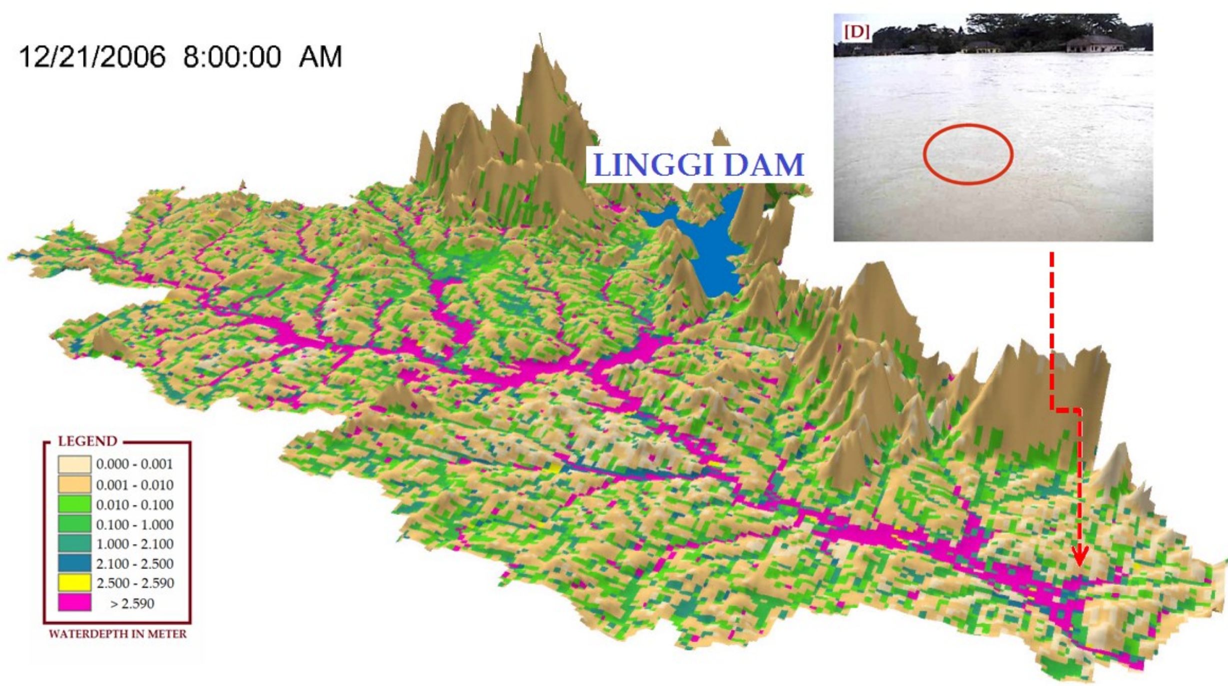

Figure 6.7 3-dimensional representation of the water depths at Kota Tinggi watershed on December 21, 2006 (adapted from Abdullah 2013)

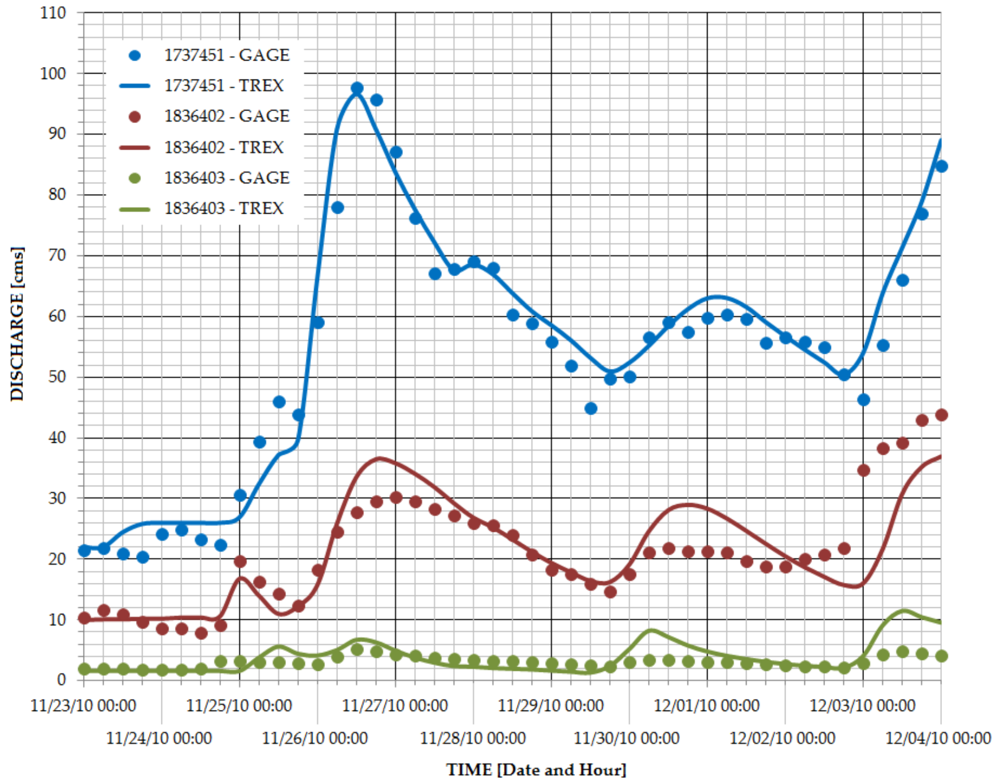

Figure 6.8 Hydrologic calibrations for large watershed (adapted from Abdullah 2013)

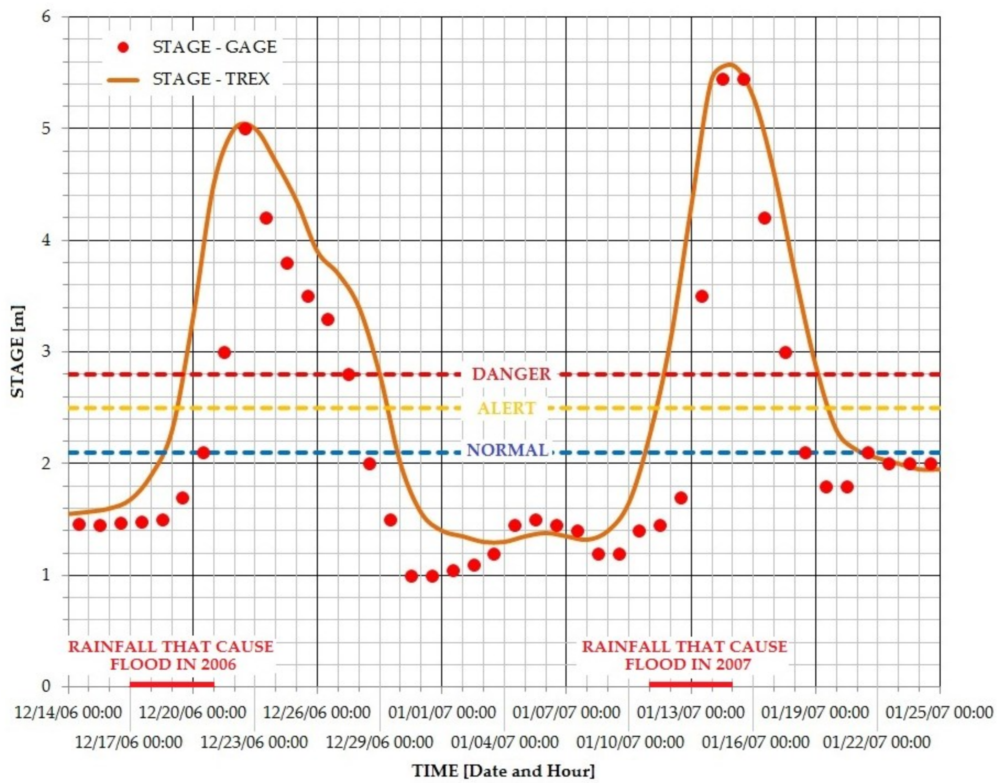

Figure 6.9 Hydrologic validations for large watershed using stage (adapted from Abdullah 2013)

reached on the 4th day of the multi-day rainfall event. It took 6 days for the stage to return to the normal level.

The hydrological modeling performed by Abdullah (2013) gives a physical representation of the flooding in Kota Tinggi. The results further prove that the multi-day rainfall events are the main cause of severe flooding in the area.

Using the calibrated and validated TREX model of the Kota Tinggi watershed (from Abdullah, 2013), the thresholds of flood were determined for rainfall durations of 1- to 4-consecutive days. An average value of rainfall intensity (in mm/hr) was used to model the rainstorm for each duration. The thresholds of flooding for each rainfall duration were determined by the total rainfalls that reach the danger level of 2.8 m. Figure 6.10 shows the stage hydrograph for single and multi-day rainfall events. The simulations give a range of flood threshold of between 140 and 170 mm for rainfall durations of 1 to 4-consecutive days. The range of flood threshold values take into account the uncertainty of model parameters in the TREX model such as Manning's n and hydraulic conductivity.

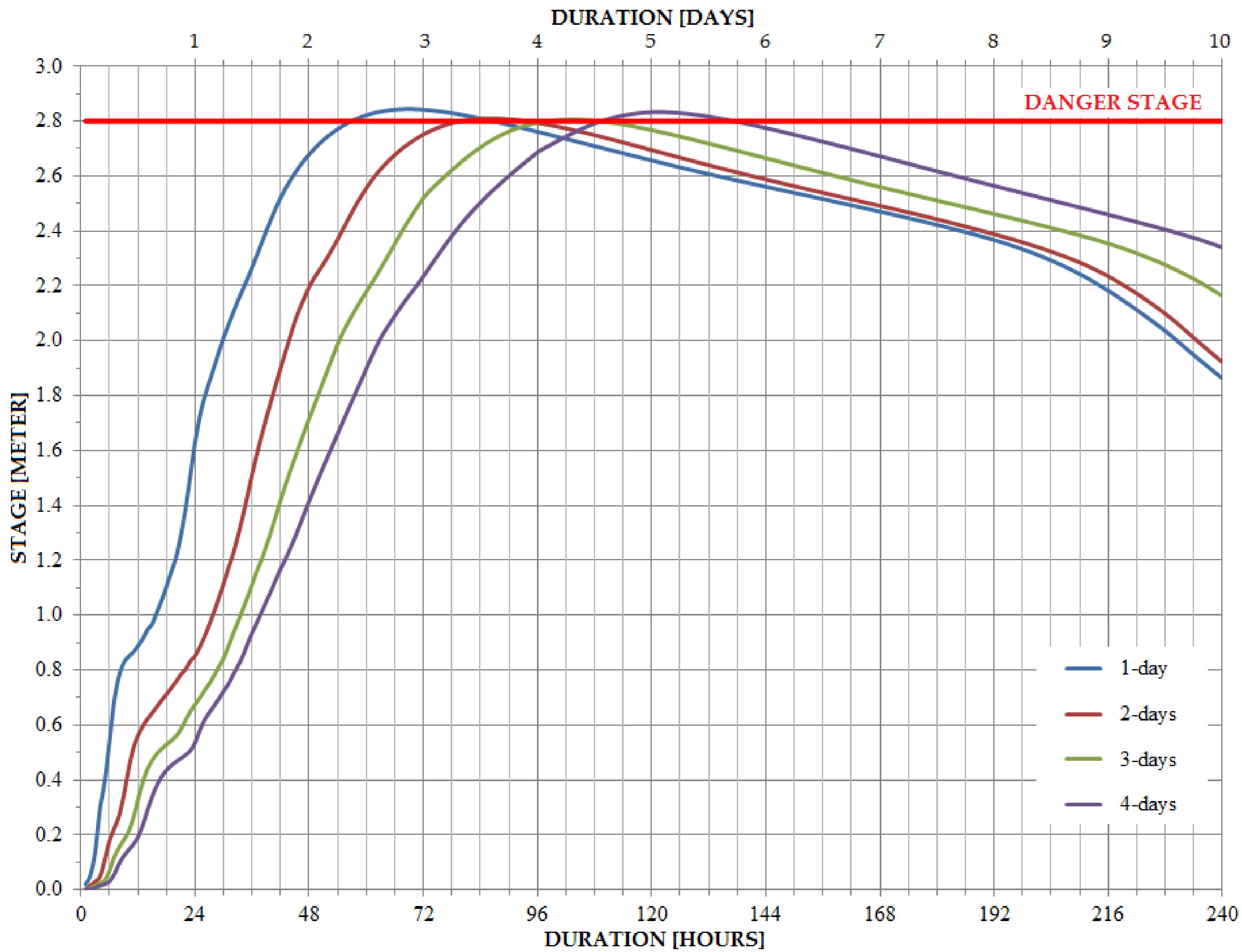

Figure 6.10 Stage hydrograph for 1-day and multi-day rainfall event

6.5 RETURN PERIODS FOR FLOOD THRESHOLD

The multi-day rainstorm caused flooding at Kota Tinggi watershed. Hydrological modeling using TREX was performed by Abdullah (2013) to determine the flood thresholds for 1-day and multi-day rainstorm events. Eq. 5.1 to Eq. 5.5 are used to estimate the return periods corresponding to the total rainfall that can cause flooding in Kota Tinggi. Table 6.3 provides the summary of the return periods for each rainfall duration.

Table 6.3 Rainfall duration, flood threshold and the respective return period

Rainfall Duration (t-consecutive days)	**Flood Threshold (mm)**	**Return Period (years)**	
		Upper values	**Lower values**
1	Between 140 to 170	220	54
2		83	23
3		42	13
4		24	7

A return period of 220 years (upper value) is the flood threshold for 1 day of rainfall. The return period decreased significantly to 83 years for 2 consecutive days of rainfall. The return period for 2 consecutive days is significantly lower than the 1-day event because the probability to receive 170 mm of rainfall in 2 days is higher than a single day. For the same reason, it can be observed from Table 6.3 that the return periods for 3- and 4-consecutive rainy days are lower than the 2-day event at 42 and 24 years, respectively. Overall, the return period estimated for the multi-day rainfall is significantly lower than a single day event. For example, the return period to reach the flood threshold in a day is 220 years, while the return period for 4 consecutive rainy days is 24 years.

These results are useful in determining the design rainfall for a flood mitigation structure at Kota Tinggi watershed. The recommended design rainfall at Kota Tinggi for this historical storm event is 220 years. The structure is estimated to be exceeded once (on average) every 220 years. Additionally, it is estimated that the flood mitigation structure will contain a 2-day event on average of about 3 times in 220 years. The 220-year design is adequate to contain the 3- and 4-consecutive day rains. On average, the structure will be used 5 and 9 times in the period of 220 years for 3- and 4-consecutive day rains, respectively.

Figure 6.11 shows the rainfall durations for the gaging stations in Kota Tinggi, its corresponding return periods and also the flood threshold for the December 2006 rainstorm. The plot shows that the cumulative rainfall at all gaging stations crossed the flood threshold level after day 2 of the multi-day rainstorm event.

6.6 SUMMARY

The algorithms developed in this study are used to estimate the return periods for flood thresholds, and also the December 2006 and January 2007 rainstorm events. The return period estimated for the multi-day rainfall is significantly lower than a single day event. Multi-day rainstorms in December 2006 crossed the flood threshold value after 2 days of continuous rainfall.

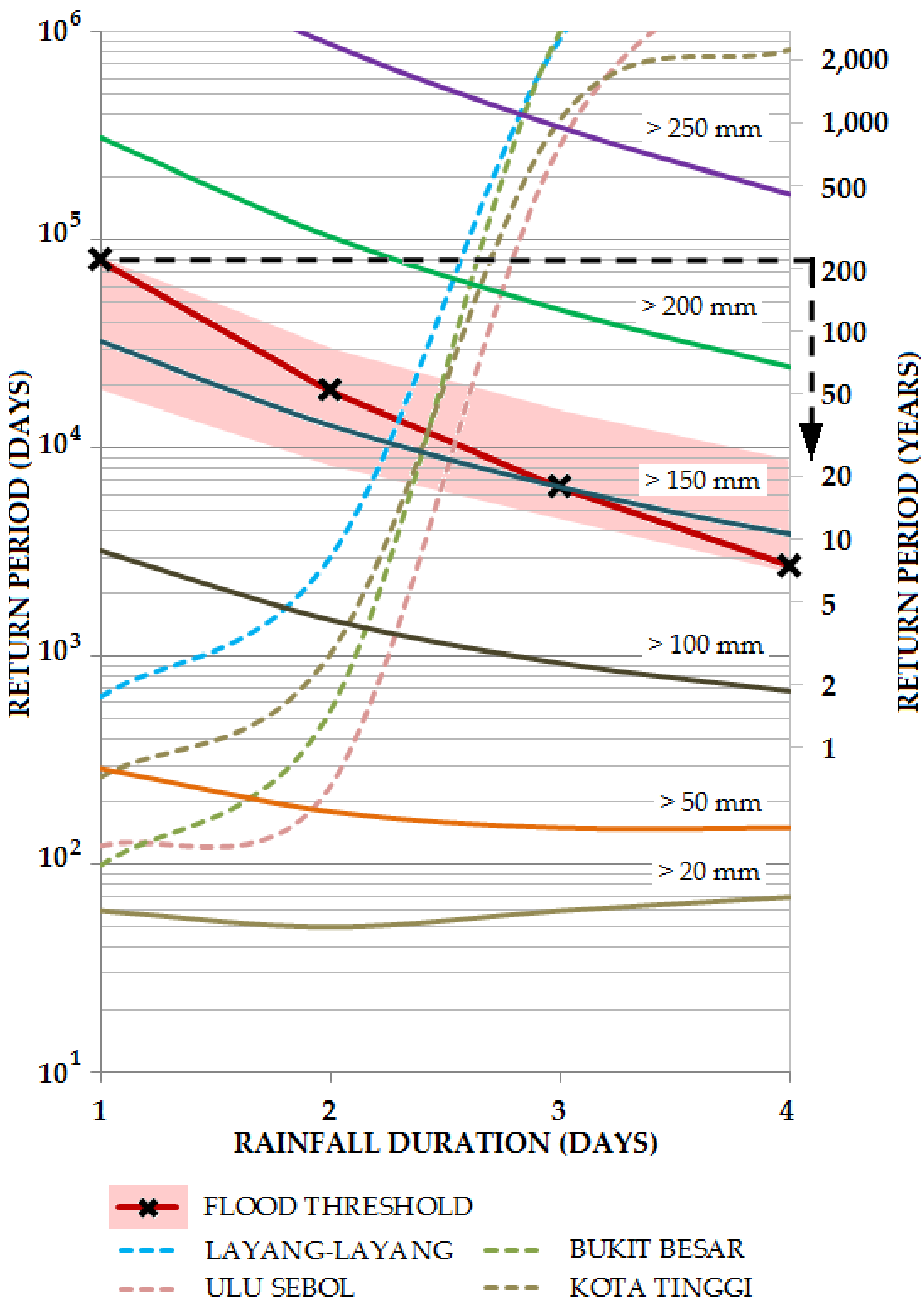

Figure 6.11 Rainfall durations versus return periods for the December 2006 rainstorms

The estimation of return periods using Eq. 5.1 to Eq. 5.5 is suitable to be used at a large watershed (size of more than 1,000 km^2) because multi-day rainstorms are the main cause of flooding in the area.

CHAPTER 7

CONCLUSIONS

Peninsular Malaysia is exposed to two major monsoon seasons, the North East (NE) and South West (SW). The NE and SW monsoons occur between the months of November to March and May to September, respectively. These monsoons result in significant amounts of rainfall, the majority of it resulting from multi-day events. Multi-day rainstorms are also very important because large rainstorms cause major floods on large watersheds (more than 1,000 km^2). The December 2006 and January 2007 multi-day rainstorm events at Kota Tinggi, Johor are the example of such circumstances.

This study examines various aspects pertaining to the characteristics of the monsoon-affected rainfall, with the emphasis on the multi-day events. The specific objectives and conclusions are given in the following sections:

Objective 1: Examine the probability structure of multi-day rainfall events.

A day is classified as wet when the recorded rainfall amount exceeds 0.1 mm. This value is selected based on the Von Neumann ratio test. The daily rainfall data at Subang Airport from 1960 to 2011 show that the majority of wet and dry events are multi-days, with the fraction of 57% and 51%, respectively. Conditional probabilities of t-consecutive wet and dry days are used to prove the dependency of the events. The probability of occurrence for both wet and dry days increases from day to day. For instance, the probability of rain on any random day is 0.53, and the conditional probability of rain the second day increases to 0.63. The probability of rain after 9-consecutive rainy days exceeds 0.75. Similarly, the probability of dry on any raindom day is 0.4686, and the conditional probability of another dry day increases to 0.58. The significant increments in the conditional probabilities for both t-consecutive wet and dry days show that the events in the study area are time dependent.

A dependency test is performed on the amount of rainfall. The estimated autocorrelation coefficients are very low, which proves that there is no significant correlation between the amounts of rainfall from one day to another.

Objective 2: Find the most suitable distribution function and derive an analytical expression of the daily rainfall amount to represent the daily rainfall record at Subang Airport.

The mean and standard deviation of daily rainfall at Subang Airport are 12.77 mm and 17.24 mm, respectively. The two parameter gamma function is suitable to represent the distribution of one and t-consecutive wet days at the study area. The 1:1 plot shows excellent agreement between the observed data and calculated values for multi-day events up to 6 consecutive days.

Objective 3: Select and simulate the sequence of daily rainfall using the discrete autoregressive family models, i.e., the DAR(1) and DARMA(1,1).

The DAR(1) and DARMA(1,1) models are applied separately to the NE and SW monsoons. The best model to generate the sequences of daily rainfall at Subang Airport is selected based on the four-step process suggested by Salas and Pielke (2003). The four-step process includes model identification, model estimation, model selection and model verification. The autocorrelation coefficients of the observed data do not decay rapidly to zero, and this characteristic shows that a long-persistence model such as DARMA(1,1) is more suitable than DAR(1). Additionally, the low sum of squared errors for the probability distributions confirm that DARMA(1,1) is most suitable to simulate daily rainfall sequences at Subang Airport for both monsoons.

Objective 4: Develop and test an approach to calculate the return period of multi-day rainfall events with respect to the duration and amount.

The return period for 1-day and multi-day rainfall events is estimated from the properties of wet run length and rainfall amount. The proposed method shows good agreement between calculated and observed values for multi-day rainfall amounts up to 150 mm and return period of 20 years. A very long sequence of daily rainfall (1,000,000 days) is generated to extend the analysis of multi-day events with cumulative rainfall up to 350 mm, which gives an estimated return period as high as 2,000 years. The mean, standard deviation, maximum daily rainfall, lag-1 ACF coefficient and maximum wet and dry run lengths of the generated daily rainfall sequence using DARMA(1,1) are also comparable with the observed data.

The algorithms developed in this study are applied to the December 2006 rainstorm event at Kota Tinggi, Johor. This rainstorm is extremely rare because the multi-day rainstorm resulted in 350 mm of cumulative rainfall and the estimated return period is greater than 2,000 years. The method proposed in this study is helpful for the design of levees on large watersheds (size of more than 1,000 km^2) because multi-day rainstorms are the main cause of flooding to the area. The return period to overtop the current levee at Kota Tinggi is 220 years for a 1-day rainstorm, but this period of return decreases to 24 years when considering 4-day rainstorms.

REFERENCES

Abdullah, J. (2013). "Distributed runoff simulation of extreme monsoon rainstorms in Malaysia using TREX." Ph.D thesis, Department of Civil and Environmental Engineering, Colorado State University, CO

Abu Bakar, S., Yusuf, M. F., and Amly, W. S. (2007). "Johor hampir lumpuh (…)", Utusan Melayu, January 15, 2007<http://www.utusan.com.my/utusan/ info.asp?y=2007&dt=0115&pub=Utusan_Malaysia&sec=Muka_Hadapan&pg=mh_01.htm> [Accessed on February 11, 2013]

Ali, S., Patnaik, U. S., and Prakasah, C. (2002). "Frequency analysis of one day and consecutive days maximum rainfall of eastern Ghat high land zone of Orissa for Marshy Watershed management planning." *Journal of the Institution of Engineers (India)*, 83, 33-36

Badrul Hisham, A.S., M.I. Marzukhi, & A.R. Daud, 2010. The worst flood in 100 years: Johore Experience. Community Health Journal, 15: 1-14. < http://www.communityhealthjournal. org /pdf/Vol15(K)-Badrul1.pdf> [Accessed on March 27, 2013]

Baigorria, G. A., and Jones, J. W. (2010). "GiST: A stochastic model for generating spatially and temporally correlated daily rainfall data." *Journal of Climate*, 23, 5990-6008

Bardaie, M. Z., and Salam A. C. A. (1981). "A stochastic model of daily rainfall for Universiti Pertanian Malaysia, Serdang." *Pertanika*, 4(1), 1-9

Bhakar, D. S., Kansal, A. K., and Chhajed, N. (2006). "Frequency analysis of consecutive days of maximum rainfall at Udaipur." *Journal of the Institute of Engineers (India)*, 89, 14-16

Bingham, C., and Nelson, L. S. (1981). "An approximation for the distribution of the von Neumann Ratio." *Technometrics*, 23(3), 285-288

Buishand, T. A. (1982). "Some methods for testing the homogeneity of rainfall records." *Journal of Hydrology*, Elsevier Scientific Publishing Company, Amsterdam, 58, 11-27

Buishand, T. A. (1982). "Some methods for testing the homogeneity of rainfall records." *Journal of Hydrology*, 58, 11-27

Buishand, T. A. (1978). *The binary DARMA (1,1) process as a model for wet-dry sequences*, Technical Note, Wageningen: Agriculture University

Buishand, T. A. (1981). "The analysis of homogeneity of long-term rainfall records in the Netherlands." *KNMI Scientific Report WR*, De Bilt, Netherlands

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

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

Cancelliere, A., and Salas, J. D. (2010). "Drought probabilities and return period for annual streamflow series." *Journal of Hydrology*, 391, 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*, Proceedings of the 3rd International Colloquium, 7-9 October 2004, Bucharest, Romania, 82-92

Chang, T. J., Kavvas, M. L., and Delleur, J. W. (1984). "Modeling of sequences of wet and dry days by binary discrete autoregressive moving average processes." *American Meteorological Society*, 23, 1367-1378

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

Chang, T. J., Delleur, J. W., and Kavvas, M. L. (1987). "Application of discrete autoregressive moving average models for estimation of daily runoff." *Journal of Hydrology*, 91, 119-135

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

Chin, E. H. (1977). "Modeling daily precipitation occurrence process with Markov chain." *Water Resources Research*, 13(6), 949-956

Chung, C. H., and Salas, J. D. (1999). "Drought occurrence probabilities and risks of dependent hydrologic processes." *Journal of Hydrologic Engineering*, 5(3), 259-268

Chung, C. H. (1999). "Probability distribution, risk, and return period of dependent hydrologic events." PhD Thesis, Department of Civil and Environmental Engineering, Colorado State University, CO

Cindrić, K. (2006). "The statistical analysis of wet and dry spells by binary DARMA (1,1) model in Split, Croatia." *BALWOIS Conference 2006*, Ohrid

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." *Theoretical and Applied Climatology*, 49, 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 Journal of Radio and Space Physics*, 39, 39-44

Delleur, J. W., Chang, T. J., and Kavvas, M. L. (1989). "Simulation models of sequences of dry and wet days." *Journal of Irrigation and Drainage Engineering*, 115(3), 344-357

Deni, S. M., Jemain, A. A., and Ibrahim, K. (2009a). "Fitting optimum order Markov chain models for daily rainfall occurrences in Peninsular Malaysia." *Theoretical and Applied Meteorology*, 97, 109-121

Deni, S. M., Jemain, A. A., and Ibrahim, K. (2009b). "Mixed probability models for dry and wet spells." *Statistical Methodology*, 6, 290-303

Deni, S. M., and Jemain, A. A. (2009a). "Fitting the distribution of dry and wet spells with alternative probability models." *Meteorological Atmospheric Physics*, 105, 13-27

Deni, S. M., and Jemain, A. A. (2009b). "Mixed log series geometric distribution for sequences of dry days." *Atmospheric Research*, 92, 236-243

Deni, S. M., Jemain, A. A., and Ibrahim, K. (2008). "The spatial distribution of wet and dry spells over Peninsular Malaysia." *Theretical and Application Climatology*, 94, 163-173

Deni, S. M., Jemain, A. A., and Ibrahim, K. (2010). "The best probability models for dry and wet spells in Peninsular Malaysia during monsoon seasons." *International Journal of Climatology*, 30, 1194-1205

Detzel, D. H. M., and Mine, M. R. M. (2011). "Generation of daily synthetic precipitation series: Analyses and application in La Plata River Basin." *The Open Hydrology Journal*, 5, 69-77

Douglas, E. M., Vogel, R. M., and Kroll, C. N. (2002). "Impact of streamflow persistence on hydrologic design." *Journal of Hydrologic Engineering*, 7(3), 220-227

Evora, N. D., and Rousselle, J. (2000). "Hybrid stochastic model for daily flows simulation in semiarid climates." *Journal of Hydrologic Engineering*, 5(1), 33-42

Farmer, E. E., and Homeyer, J. W. (1974). "The probability of consecutive rainless days." *Water Resources Bulletin, American Water Resources Association*, 10(5), 914-924

Fernando, D. A. K., and Jayawardena, A. W. (1994). "Generation and forecasting of monsoon rainfall data." *Affordable Water Supply and Sanitation*, 20th WEDC Conference, Colombia, Sri Lanka, 1994

Fernández, B., nd Salas, J. D. (1999). "Return period and risk of hydrologic events. I: Mathematical formulation." *Journal of Hydrologic Engineering*, 4(4), 297-307

Fernández, B., nd Salas, J. D. (1999). "Return period and risk of hydrologic events. II: Applications." *Journal of Hydrologic Engineering*, 4(4), 308-316

Feyerherm, A. M., and Bark, L. D. (1964). "Statistical methods for persistent precipitation patterns." *Journal of Applied Meteorology*, 4, 320-328

Gabriel, K. R., and Neuman, J. (1962). "A Markov chain model for daily rainfall occurrence at Tel Aviv." *Quaterly Journal of the Royal Meteorological Society*, 88(375), 90-95

Goel, N. K., Seth, S. M., and Chanra, S. (1998). "Multivariate modeling of flood flows." *Journal of Hydraulic Engineering*, 124(2), 146-155

González, J., and Valdés, J. B. (2003). "Bivariate drought recurrence analysis using tree ring reconstruction." *Journal of Hydrologic Engineering*, 8(5), 247-258

Greco, R. (2012). "A fuzzy-autoregressive model of daily river flows." *Computer and Geosciences*, 43, 17-23

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

Hess, G. D., Leslie, L. M., Guymer, A. E., and Fraedrich, K. (1989). "Application of a Markov technique to the operational, short-term forecasting of rainfall." *Australian Meteorological Magazine*, 2 June 1989, 83-91

Hirsch, R. M. (1979). "Synthetic hydrology and water supply reliability." *Water Resources Research*, 15(6), 1603-1615

Hsu, K. L., Gupta, H. V., Sorooshian, S. (1995). "Artificial neural network modeling of the rainfall-runoff process." *Water Resources Research*, 31(10), 2517-2530

Jacobs, P. A., and Lewis, P. A. W. (1983). "Stationary discrete autoregressive-moving average time series generated by mixtures." *Journal of Time Series Analysis*, 4(1), 19-36

Jacobs, P. A., and Lewis, P. A. W. (1977). "A mixed autoregressive-moving average exponential sequence and point process (EARMA 1,1)." *Advance in Applied Probability*, 9(1), 87-104

Jacobs, P. A., and Lewis, P. A. W. (1978a). *Discrete time series generated by mixtures III: Autoregressive process*, Technical note, Naval Postgraduate School, Monterey, California

Jacobs, P. A., and Lewis, P. A. W. (1978b). "Discrete time series generated by mixtures. I: Correlational and runs properties." *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(1), 94-105

Jacobs, P. A., and Lewis, P. A. W. (1978c). "Discrete time series generated by mixtures. II: Asymptotic properties." *Journal of the Royal Statistical Society. Series B (Methodological)*, 40(2), 222-228

Jimoh, O. D., and Webster, P. (1996). "The optimum order of a Markov chain model for a daily rainfall in Nigeria." *Journal of Hydrology*, 185, 45-69

Katz, R. W., and Parlange, M. B. (1998). "Overdispersion phenomenon in stochastic modeling of precipitation." *American Meteorological Society*, 591-601

Katz, R. W. (1996). "Use of conditional stochastic models to generate climate change scenarios." *Climatic Change*, 32, 237-255

Katz, R. W. (1977). "Precipitation as a chain-dependent process." *Journal of Applied Meterology*, 16(7), 671-676

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

Kendall, D. R., and Dracup, J. A. (1991). "A comparison of index-sequential and AR(1) generated hydrologic sequences." *Journal of Hydrology*, 122, 335-352

Kim, T. W., Valdés, J. B., and Yoo, C. (2003). "Nonparametric approach for estimating return periods of droughts in arid regions." *Journal of Hydrologic Engineering*, 8(5), 237-246

Kite, G. W. (ed.) (1978). *Frequency and risk analyses in hydrology*, 2nd Edition, Water Resources Publication, Colorado

Kuo, J. T., Sun, Y. H. (1996). "An ARMA-type section model for average ten-day streamflow synthesis." *Water Resources Management*, 10, 333-354

Lau, K. M. (1997). "Climatology and interannual variability of the Southeast Asia Summer Monsoon." *Advances in Atmospheric Sciences*, 14(2), 141-162

Llyod, E. H., and Saleem, S. D. (1979). "Waiting time to first achievement of specified levels in reservoirs subject to seasonal Markovian inflows." *Input for risk analysis in water systems*, McBean, E. A., K. W. Hipel, and T. E. Unay (eds.), Water Resources Publications, Littleton, Colorado

Lloyd, E. H. (1970). "Return periods in the presence of persistence." *Journal of Hydrology*, 10, 291-298

Loaiciga, H. A., and Marino, M. A. (1991). "Recurrence interval of geophysical events." *Journal of Water Resources Planning and Management*, 117(3), 367-382

Loaiciga, H. A., and Leipnik, R. B. (1996). "Stochastic renewal model of low-flow streamflow sequences." *Stochastic Hydrology and Hydraulics*, 10, 65-85

Machiwal, D., Jha, M. K., and Mal, B. C. (2006). "Forecasting of salient consecutive days' maximum rainfalls of Kharagpur, India using probabilistic approach." *International Agricultural Engineering Journal*, 15(2-3), 65-77

Marivoet, J. L. (1983). "Real time water quality forecasting models based on the water quantity/quality relationship." *Dissolved Loads of Rivers and Surface Water Quantity/Quality Relationship*, Proceedings of the Hamburg Symposium, August 1983, Hamburg, Germany

Mimikou, M. (1983). "Daily precipitation occurrences modeling with Markov chain of seasonal order." *Journal of Hydrological Sciences*, 28(2), 221-232

Marquardt, D. W. (1963). "An algorithm for least-squares estimation of nonlinear parameters." *Journal of the Society for Industrial an Applied Mathematics*, 11(2), 431-441

Morgan, J. R. and Valencia, M. J. (1983). "The natural environment setting." *Atlas for marine policy in Southeast Asian Seas*, Morgan, J. R. and Valencia, M. J. (eds.), Univ. of California Press

Mujumdar, P. P., and Kumar, D. N. (1990). "Stochastic models of streamflow: some case studies." *Hydrological Sciences*, 34(4), 395-410

NAHRIM (2008). "Technical guideline for estimating probable maximum precipitation for design floods in Malaysia." *NAHRIM Technical Research Publication No. 1 (TRP 1)*

Neumann, J. V. (1941). "Distribution of the ratio of the mean square successive difference to the variance." *The Annals of Mathematical Statistics*, 12(4), 367-395

Ngai, W. C. (1995). "Flood disaster management in Malaysia: an evaluation of the effectiveness of government resettlement schemes." *Disaster Prevention and management, MCB Uni. Press*, 4(4), 22-29

Owen, D. B. (ed.) (1962). *Handbook of statistical tables*, Addison-Wesly, Reading, Massachusetts, USA

Roldan, J., and Woolhiser, D. A. (1982). "Stochastic daily precipitation models 1: A comparison of occurrence process." *Water Resources Research*, 18(5), 1451-1459

Richardson, C. W., and Wright, D. A. (1984). "WGEN: A model for generating daily weather variables." USDA ARS Bulletin

Saadon, M. N., Anawat, P. R. and Snidvongs, A. (1999). "Physical characteristics of watermass in the South China Sea, Area I: Gulf of Thailand and East Coast of Peninsular Malaysia." *Proceedings of the first technical seminar on Marine Fishery Resources Survey in the South China Sea Area I, Gulf of Thailand and East Coast of Peninsular Malaysia, (SEAFDEC),* 1-5

Salas, J. D., and Obeysekera, J. T. B. (1982). "ARMA model identification of hydrologic time series." *Water Resources Research*, 18(4), 1011-1021

Salas, J. D., and Pielke, S. R. A. (2003). "CHAPTER 32: Stochastic characteristics and modeling of hydroclimatic processes." *Handbook of Weather, Climate, and Water*, Potter, T. D., and R. C. Bradley, (eds.),John Wiley & Sons, Inc., 587-605

Salas, J. D., Fu, C., Cancelliere, A., Dustin, D., Bode, D., Pineda, A., and Vincent, E. (2005). "Characterizing the severity and risk of drought in the Poudre River, Colorado." *Journal of Water Planning and Management*, 131(5), 383-393

Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L. (Eds.) (1988). *Applied modeling of hydrologic time series*, Water Resources Punblications, 3rd Edition

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

Şen, Z. (1999). "Simple risk calculations in dependent hydrological series." *Hydrological Science Journal*, 44(6), 871-878

Sharma, T. C. (1996). "Simulation of the Kenyan longest dry and wet spells and the largest rain-sum using a Markov model." *Journal of Hydrology*, 178, 55-67

Shiau, J. T., and Shen, H. W. (2001). "Recurrence analysis of hydrologic droughts of differing severity." *Journal of Water Resources Planning and Management*, 127(1), 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 Resources Research*, 22(10), 1422-1430

Spolia, S. K., and Chander, S. (1974). "Modelling of surface runoff systems by an ARMA model." *Journal of Hydrology*, 22, 317-332

Stedinger, J. R., Lettenmaier, D. P., and Vogel, R. M. (1985). "Multisite ARMA (1,1) and disaggregation models for annual streamflow generation." *Water Resources Research*, 21(4), 497-509

Suhaila, J. and Jemain, A. A. (2007). "Fitting Daily Rainfall Amount in Malaysia Using the Normal Transform Distribution." *Journal of Applied Sciences*, 7(14), 1880-1886

Tan, S. K., Sia, S. Y. (1997). "Synthetic generation of tropical rainfall time series using an event-based method." *Journal of Hydraulic Engineering*, 2(2), 83-89

Upadhyaya, A., and Singh, S. R. (1998). "Estimation of consecutive days maximum rainfall by various methods and their comparison." *Indian Journal of Soil Conservation*, 26(3), 193-201

Vogel, R. M., Tsai, Y., and Limbrunner, J. F. (1998). "The regional persistence and variability of annual streamflow in the United States." *Water Resources Research*, 34(12), 3445-3459

Vogel, R. M. (1987). "Reliability indices for water supply systems." *Journal of Water Resources Planning and Management*, 113(4), 563-579

Wan, H., Zhang, X., and Barrow, E. M. (2005). "Stochastic modeling of daily precipitation for Canada." *Atmospheric-Ocean*, 43(1), 23-32

Wallis, T. W. R., and Griffiths, J. F. (1995). "An assessment of the weather generator (WXGEN) used in the erosion/productivity impact calculator (EPIC)." *Agricultural and Forest Meteorology*, 73, 115-133

Wang, B. (Ed.) (2006). *The Asian Monsoon*. Springer, Berlin, Germany

Wang, W., Gelder, P. V., and Vrijling, J. K. (2005). "Improving daily stream flow forecasts by combining ARMA and ANN models." *International Conference on Innovation Advances and Implementation of Flood Forecasting Technology*, 17-19 October 2005, Tromsø, Norway

Wilks, D. S. (1998). "Multisite generalization of a daily stochastic precipitation generation model." *Journal of Hydrology*, 210, 178-191

Week, W. D., and Boughton, W. C. (1987). "Tests of ARMA model forms for rainfall-runoff modeling." *Journal of Hydrology*, 91, 29-47

Woodyer, K. D., McGilchrist, C. A., and Chapman, T. G. (1972). "Recurrence intervals between exceedence of selected river levels 4. Seasonal streams." *Water Resources Research*, 8(2), 435-443

Wu, C. L., Chau, K. W., and Li, Y. S. (2009). "Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques." *Water Resources Research*, 45, 1-23

Yurekli, K., and Oztruk, F. (2003). "Stochastic modeling of annual maximum and minimum streamflow of Kelkit Stream." *Water International*, 28(4), 433-441

APPENDIX A

RETURN PERIODS FOR THE JANUARY 2007 RAINSTORM EVENT

Table A1 summarized the estimation of return periods for the January 2007 rainstorm event using Eq. 5.1 to Eq. 5.5. The cumulative rainfalls measured from Layang-Layang, Ulu Sebol, Bukit Besar and Kota Tinggi gaging stations are used in this section.

Table A1 Estimation of return periods for the January 2007 rainstorm event

Date	January 2007 Jan-11	Jan-12	Jan-13	Jan-14
Station : Layang-layang				
Cumulative Rainfall (mm)	145	280	364	384
Return period (years)	71	8,493	102,466	89,863
Station : Ulu Sebol				
Cumulative Rainfall (mm)	124	414	490	534
Return period (years)	27	2,630,137	19,150,685	38,849,315
Station : Bukit Besar				
Cumulative Rainfall (mm)	147	381	423	458
Return period (years)	77	641,370	1,178,082	1,753,424
Station : Kota Tinggi				
Cumulative Rainfall (mm)	167	289	338	-
Return period (years)	192	12,603	35,069	-

The rainfall magnitudes for January 2007 are much higher than December 2006 rainstorm. The highest rainfall on the first day was measured at the Kota Tinggi station, with 167 mm. This rainfall amount corresponds to the return period of almost 200 years. Other stations also recorded significant rainfall amounts, ranging from 124 to 147 mm.

The amounts measured for the second (Jan 12), third (Jan 13) and fourth (Jan 14) of the multi-day rainstorm events were much more significant as compared to the first day. Ulu Sebol recorded the most rainfall with the estimated return period of more than 2,000,000 years, followed by Bukit Besar (641,370 years), Kota Tinggi (12,603 years) and Layang-Layang (8,493 years). These values continue to increase on the third and fourth day. The gaging stations recorded the rainfall amount with return period greater than 35,000 years on the third day (Jan 13). The high estimated return periods on Jan 13 and 14 were reasonable since the cumulative rainfall amount measured for 3- and 4-consecutive days exceeds the average monthly rainfall of 200 mm. Ulu Sebol station received more than twice the average monthly rainfall for 4-consecutive days, i.e., 534 mm.

Figure A1 shows the rainfall durations for the gaging stations in Kota Tinggi, its corresponding return periods and also the flood threshold for the January 2007 rainstorm. Significant rainfalls on day-1 resulted in flooding almost immediately. Figure A1 also shows that the cumulative rainfall at all gaging stations crossed the flood threshold level after day 1.

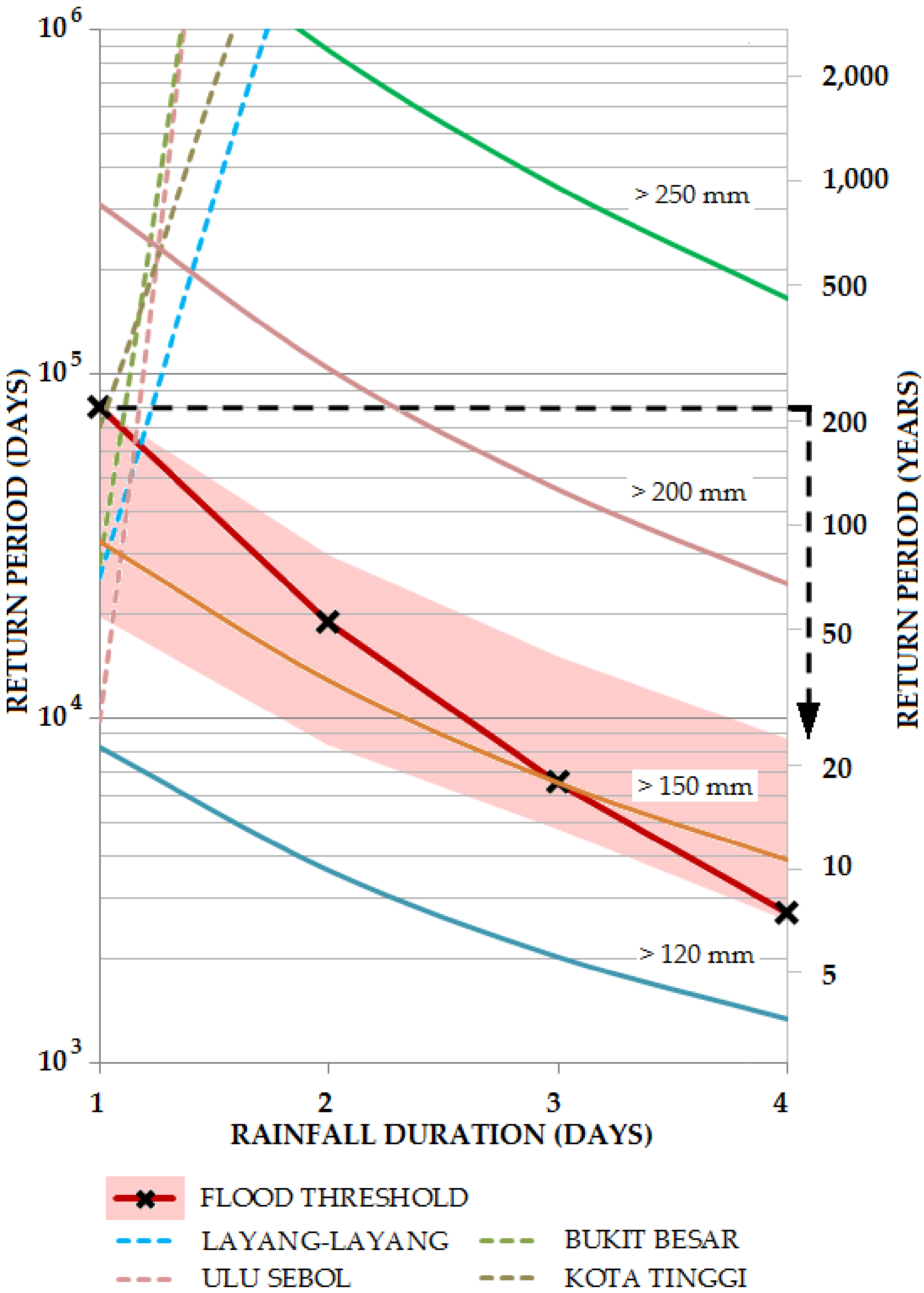

Figure A1 Rainfall durations versus return periods for the January 2007 rainstorms

APPENDIX B

THE STATISTICS OF WET AND DRY YEARS

Figure B1 shows the total annual rainfall at Subang Airport from 1960 to 2011. The total annual rainfalls at Subang Airport show that there are distinctive dry and wet periods, from 1970 to 1986 and 1987 to 2011, respectively. In this section, the rainfalls from 1970 to 1986 are referred to as the dry years, while the wet years are for the rainfall between 1987 and 2011.

The daily rainfall statistics, such as the mean, standard deviation and wet and dry run lengths for these periods are given in the following sections. These statistics are compared with the whole time series (rainfall from 1960 to 2011) in order to give the difference between the overall statistics, wet and dry years.

DAILY RAINFALL STATISTICS

The daily rainfall data measured at Subang Airport from 1960 to 2011 have an average daily rainfall of 12.77 mm. The average daily rainfalls during the period of dry and wet years are 11.72 mm and 13.84 mm, respectively. The difference between the dry and wet years and the whole time series is about $\mp$8%.

The standard deviation of the daily rainfall data for the whole time series is 17.24 mm. Higher standard deviation is estimated for the wet years, i.e., 18.23 mm, which gives the difference of 5.7% when it is compared with the whole time series. The dry

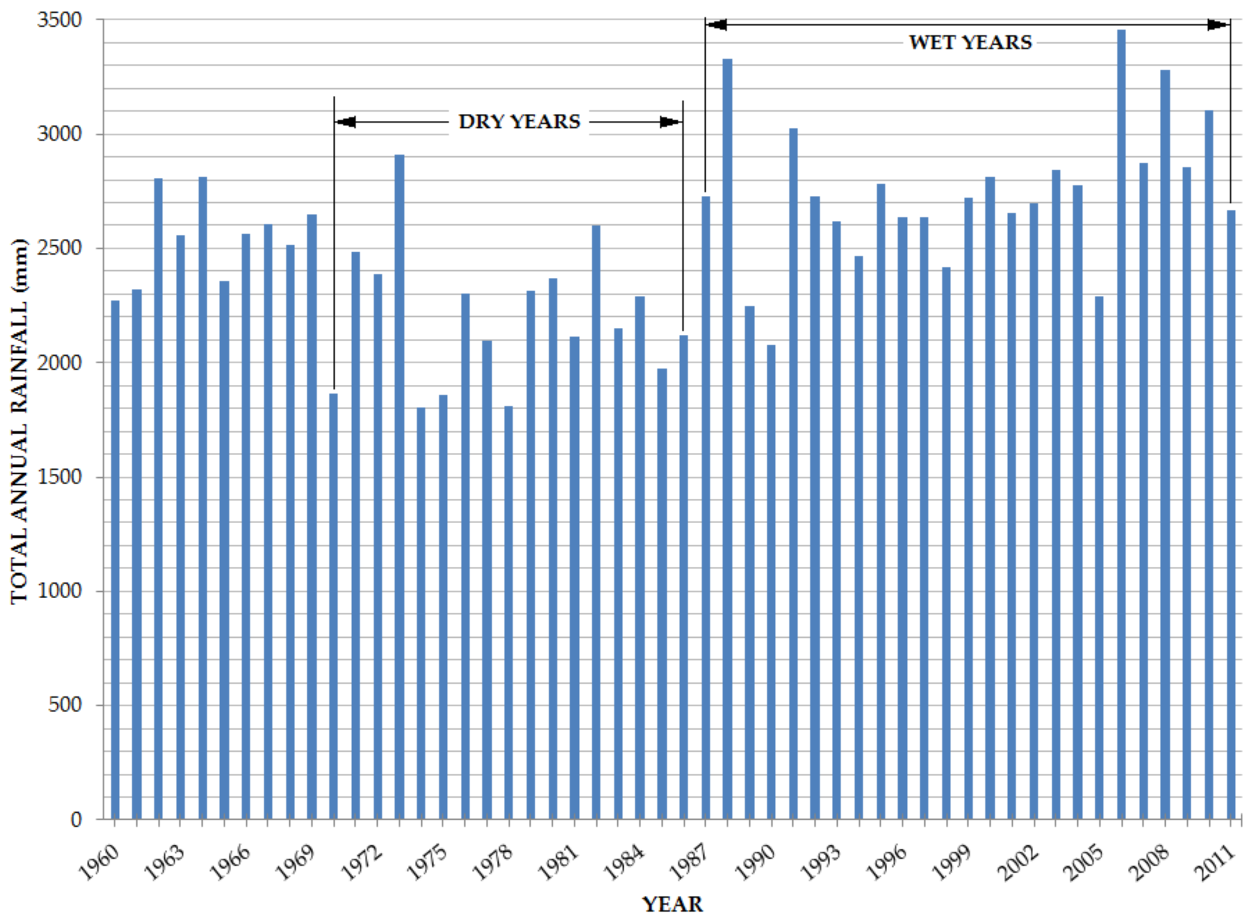

Figure B1 Total Annual Rainfall from 1960 to 2011 at Subang Airport

years gives a standard deviation of 15.79 mm, that is 8.4% lower that the whole time series.

The estimated numbers of wet run lengths observed at Subang Airport for the wet and dry years and also the whole time series are given in Figure B2. There are a total of 1,787 and 1,206 wet run lengths for wet and dry years respectively. From that amount, 42% and 45% are 1-day events for wet and dry years respectively. That gives more than 50% of the wet run lengths as equal to or more than 2-consecutive rainy days, i.e., multi-day events for both wet and dry years. These numbers are almost similar to the whole time series, which show that 43% of the total wet run lengths are one-day events, while the remaining 57% are multi-day rainfall. The estimated mean wet run lengths for the whole time series is 2.71 days. The estimation for the period of wet years is slightly higher at 2.78 days, while the calculated value for dry years is 2.65 days. These values are consistent with the percentage of wet run lengths, i.e., most of the rainfall events at Subang Airport are multi-day. The longest wet run lengths for the wet and dry years are 31 and 30 days, respectively.

The comparison of the estimated number of dry run lengths from wet, dry and whole time series at Subang Airport are shown in Figure B3. The daily rainfall records give an estimated total of 1,788, 1,205 and 3,727 dry run lengths for wet, dry and whole time series respectively. The majority of the dry run lengths for all cases are equal to or longer than 2-consecutive dry days, with the fraction of more than 50%. The percentage shown for the dry run lengths is similar to the wet run lengths, i.e., the occurrence of multi-day events is more than the single day event. The estimated mean dry run lengths for the whole time series is 2.39 days. The estimation for the period of dry years is slightly higher at 2.50 days, while the calculated value for wet years is 2.33 days. These values further verify that most of the events at Subang Airport are multi-days. The longest dry run lengths for the wet and dry years are 19 and 20 days, respectively. Table B1 summarizes the daily rainfall statistics of the whole time series, dry years and wet years.

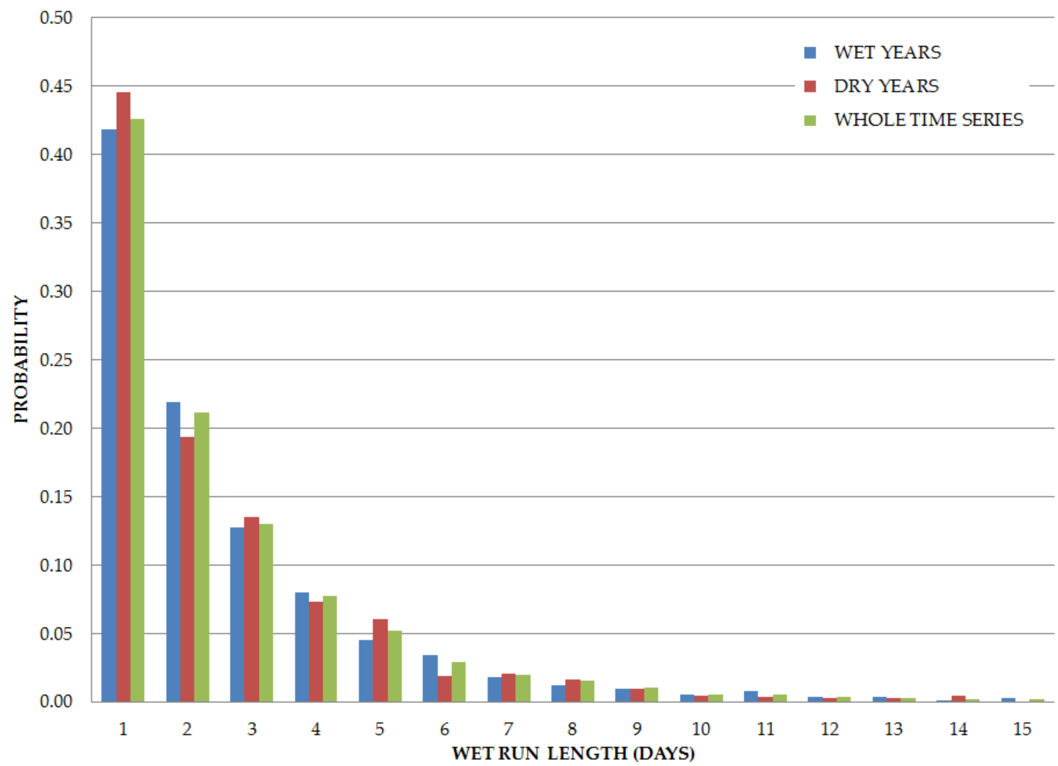

Figure B2 Probability distributions of wet run lengths for whole time series, wet years and dry years

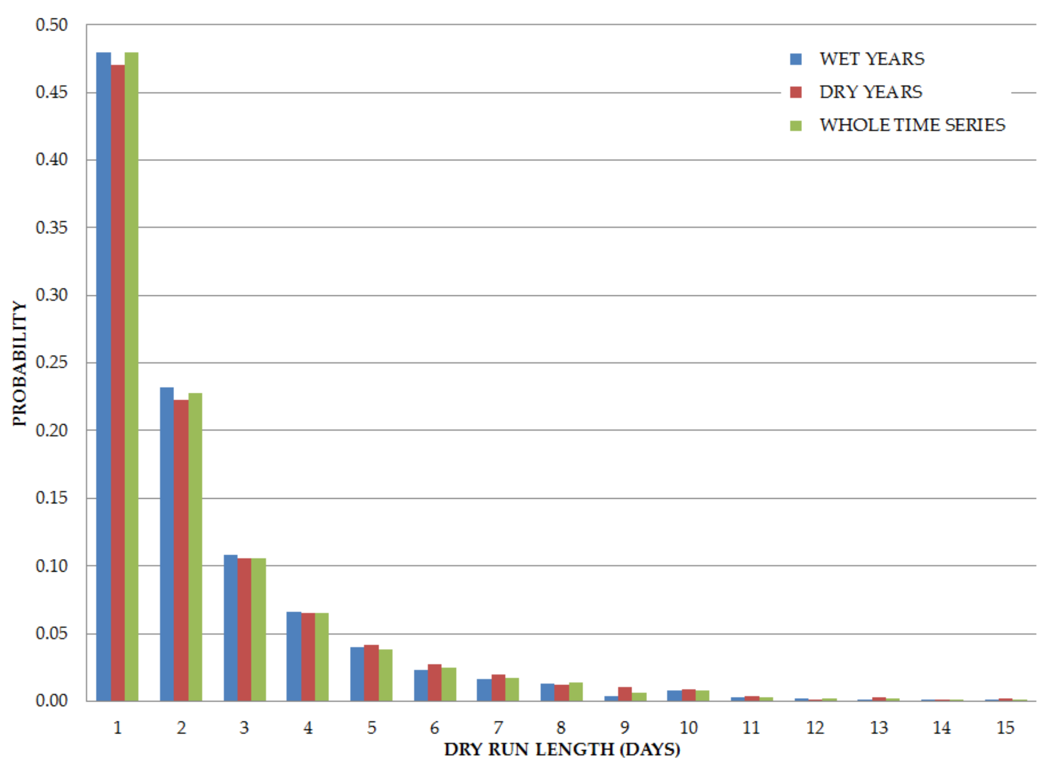

Figure B3 Probability distributions of dry run lengths for whole time series, wet years and dry years

Table B1 Daily rainfall statistics for the whole time series, dry years and wet years

Parameters	Whole time series	Dry Years	Wet Years
Mean, $\hat{\mu}$ (mm)	12.77	11.72	13.84
Standard deviation, $\hat{\sigma}$ (mm)	17.24	15.79	18.23
Mean wet run length, $\hat{\bar{T}}_1$ (days)	2.71	2.65	2.78
Mean dry run length, $\hat{\bar{T}}_0$ (days)	2.39	2.50	2.33

CONDITIONAL PROBABILITIES OF T-CONSECUTIVE WET AND DRY DAYS FOR THE WHOLE TIME SERIES, WET YEARS AND DRY YEARS

Figure B4 shows the plot of conditional probabilities for t-consecutive wet days, considering the whole time series, wet years and dry years. The highest probability for a wet day is estimated for the wet years' time series, i.e., 0.54, followed by the whole time series at 0.53 and dry years give a calculated value of 0.51. These estimations indicate that there are some differences in the probability wet of any random day for the three different scenarios. In general, the average difference between the conditional probabilities of t-consecutive wet days for wet and dry years is 5.2%. The highest difference can be seen at 9-consecutive wet days, where the wet years give an estimated conditional probability of 0.79, while the calculated value for dry years is 0.73. This gives a difference of more than 7%. Smaller differences are shown for the comparison between the whole time series with the wet and dry years. For instance, the average difference between the whole time series and wet years is 3.2%. An average of 2.2%

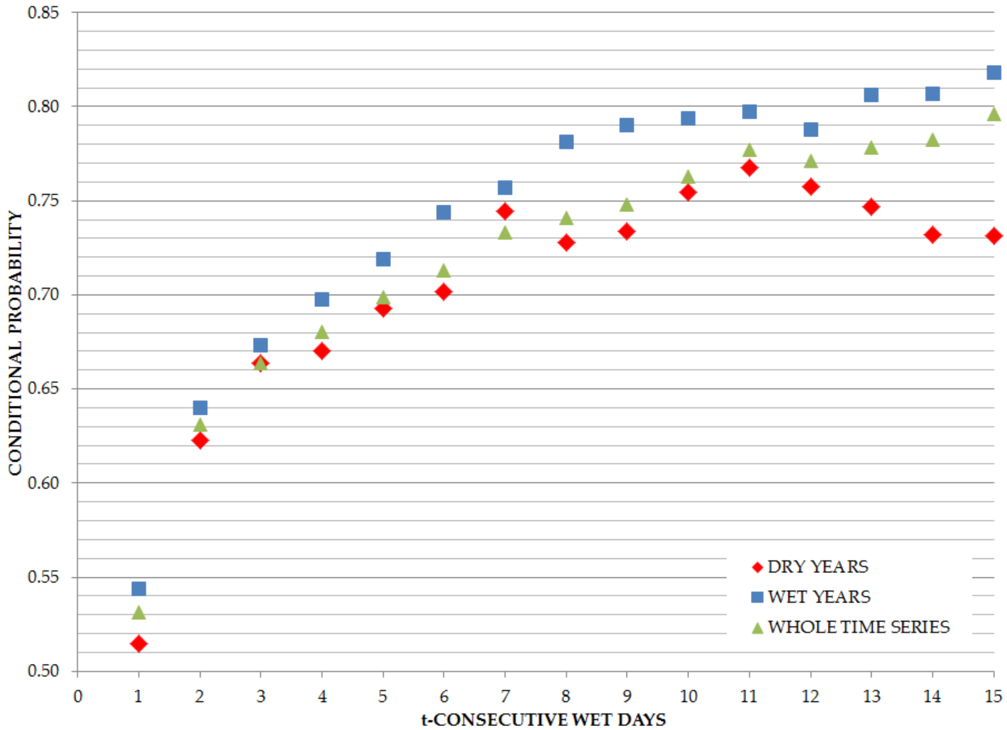

Figure B4 Conditional Probability of t-consecutive wet days

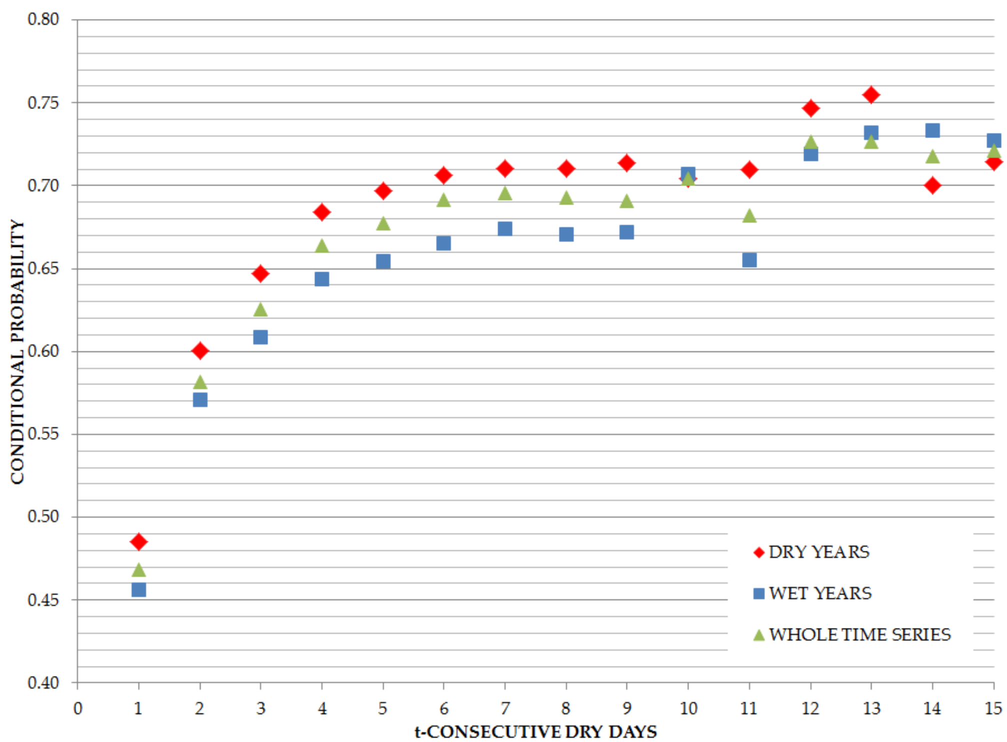

Figure B5 Conditional Probability of t-consecutive dry days difference is estimated for the conditional probabilities of t-consecutive wet days for whole time series and dry years.

The probability structure increased significantly when the number of consecutive rainy day increased for all the scenarios tested in this section. For example, the estimated probability for the whole time series from one rainy day to 15-consecutive days increased significantly, i.e., from 0.53 to 0.80. The significant increments are also seen for the wet and dry years (refer to Figure B4). With reference to the wet years, the conditional probability of a fourth rainy day, given that it has rained for 3-consecutive days, is 0.70. This probability is far greater than the probability of the first days of rain, i.e., 0.54. The examples given above show that the events are dependent; therefore, the probability of rain in a day is not constant. The occurrence of rain in one day affects the probability of rain the following day.

Figure B5 shows the plot of conditional probabilities for t-consecutive dry days, considering the whole time series, wet years and dry years. The highest probability for a dry day is estimated for the dry years' time series, i.e., 0.49, followed by the whole time series at 0.47 and wet years give an estimated value of 0.46. These estimations indicate that there are some differences in the probability of a dry day for any random day for the three different scenarios. The average difference between the conditional probabilities of t-consecutive dry days for wet and dry years is 4.2%. The highest difference can be seen at 11-consecutive dry days, where the dry years give an estimated conditional probability of 0.71, while the calculated value for wet years is 0.66. This gives a difference of 8.4%. Smaller differences are shown for the comparison between the whole time series with the wet and dry years. For instance, the average difference between the whole time series and wet years is 1.8%. An average of 2.2% difference is estimated for the conditional probabilities of t-consecutive wet days for whole time series and dry years.

DEPENDENCE OF RAINFALL AMOUNT

The dependency of rainfall amount from one rainy day to the next is tested in this section, using three different scenarios: (1) all consecutive wet days; (2) rainfall on Day 1 and Day 2 (D1 & D2); and (3) rainfall on day 2 and day 3 (D2 & D3). The tests are done using two methods, i.e., determining the Auto Correlation Function (ACF) of the rainfall amount which is based on the rainfall amount and by plotting the scatter plot.

Wet Years

For the first method, i.e., the ACFs for all scenarios are very low, which shows that the rainfall amounts are independent of each other. The ACFs are 0.0115, 0.0231, -0.0126 for all consecutive rainy days, D1 & D2 and D2 & D3, respectively. The results are summarized in Table B2.

Table B2 ACFs for all consecutive rainy days, D1 & D2 and D2 & D3

Scenario	**Sample Size (Days)**	**ACF**
All consecutive rainy days	3,178	0.0115
D1 & D2	1,039	0.0231
D2 & D3	647	-0.0126

Figures B6 and B7 show the scatter plot of the amounts of rainfall for D1 & D2 and D2 & D3. The observations for both graphs are the same, there are no structured appearances at any of the points and the plots are totally random. These plots further prove that there is no dependency between the amounts of rainfall for consecutive rainy days.

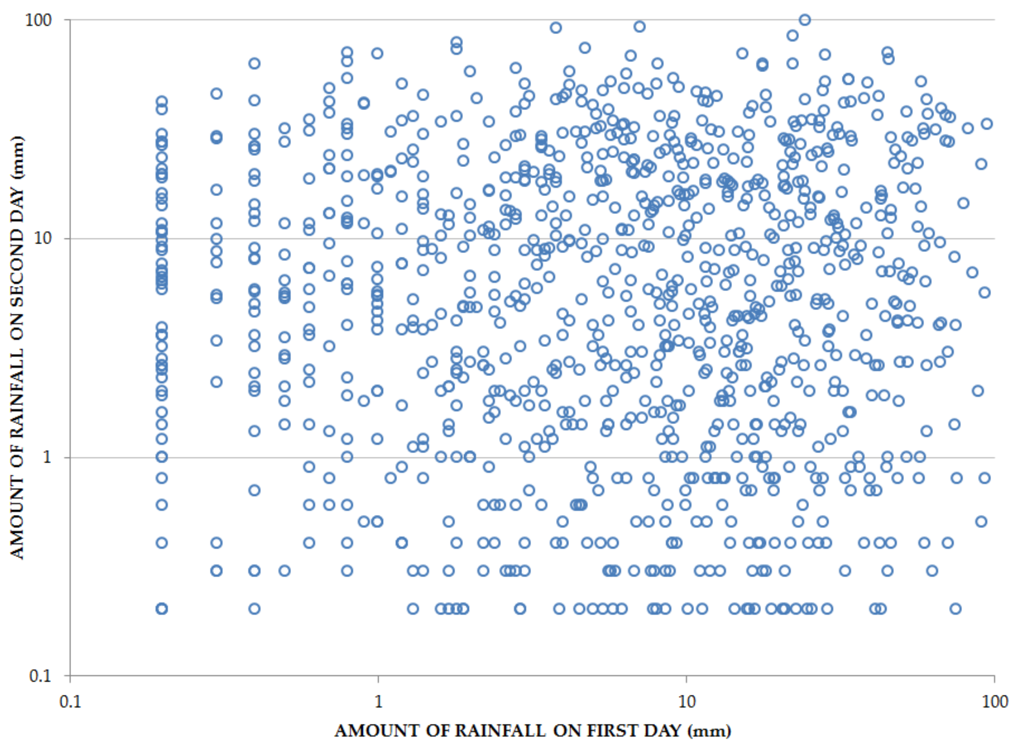

Figure B6 Amounts of rainfall on D1 and D2 for wet years

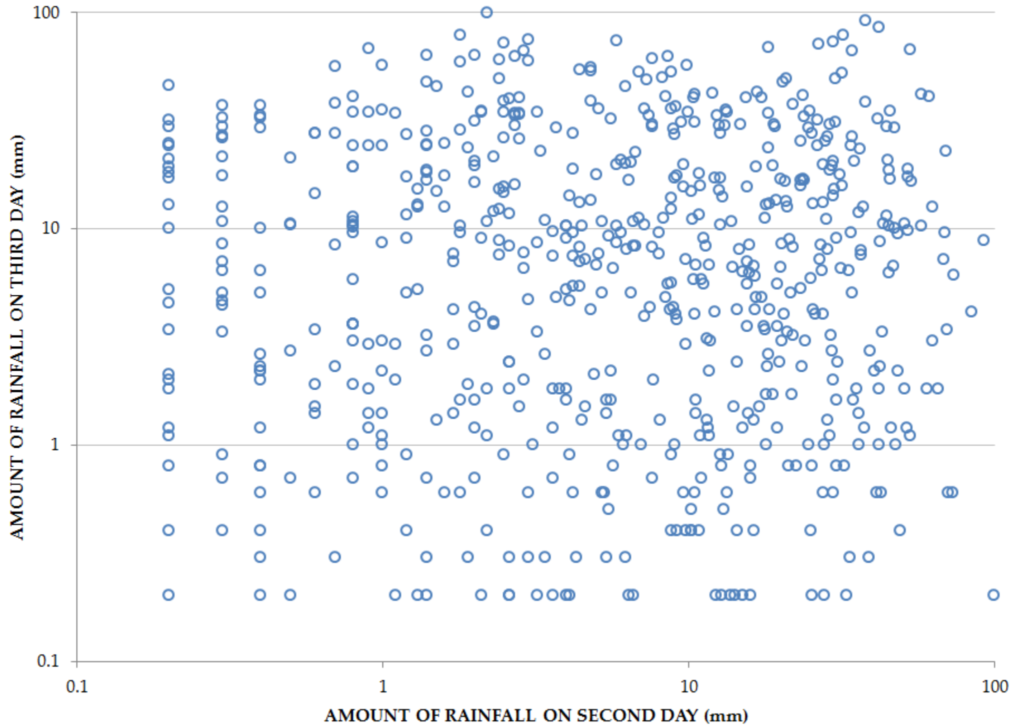

Figure B7 Amounts of rainfall on D2 and D3 for wet years

Dry Years

Similar results are shown for the dry years. The ACFs for all scenarios are very low, which shows that the rainfall amounts are independent of each other. The ACFs are 0.0551, 0.1139, 0.0126 for all consecutive rainy days, D1 & D2 and D2 & D3, respectively. The results are summarized in Table B3.

Table B3 ACFs for all consecutive rainy days, D1 & D2 and D2 & D3

Scenario	**Sample Size (Days)**	**ACF**
All consecutive rainy days	1,989	0.0551
D1 & D2	669	0.1139
D2 & D3	435	0.0126

Figures B8 and B9 show the scatter plot of the amounts of rainfall for D1 & D2 and D2 & D3. There are no structured appearances at any of the points and the plots are totally random. These plots further prove that there is no dependency between the amounts of rainfall for consecutive rainy days.

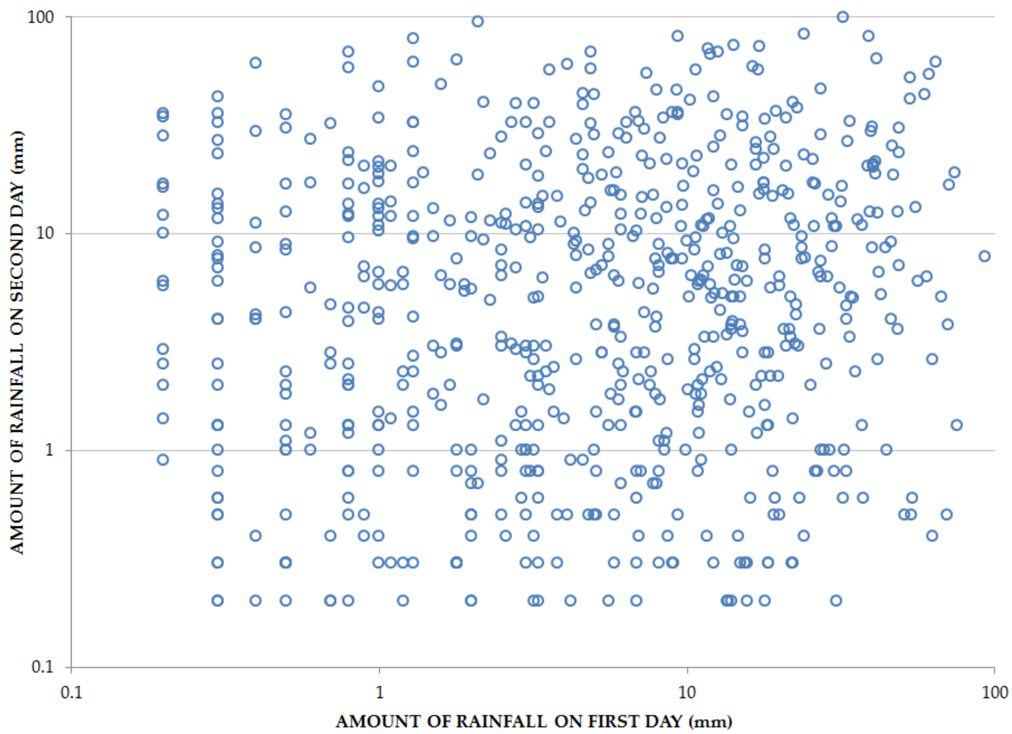

Figure B8 Amounts of rainfall on D1 and D2 for dry years

Figure B9 Amounts of rainfall on D2 and D3 for dry years

RETURN PERIODS

The estimated return periods for the whole time series, wet years and dry years are given in Figure B10. This analysis is done for several amounts of rainfall (in mm), that is 1, 13, 30, 60, 90, 120 and 150. 1 mm is selected to represent the majority of rainfall events and 13 mm is the average daily rainfall. The remaining amounts are selected because these values are considered as significant rainfall, especially during multi-day events.

For rainfall amount of more than 1 mm, the estimated return periods for whole time series, wet years and dry years are the same from one to 7-consecutive days. As the number of t-consecutive rainy days increased, the difference in estimated return periods also increased. Larger differences are seen in high rainfall amounts, i.e., 60 mm or more. Other rainfall amounts, i.e., from 13 to 150 mm show similar trend, that is the lowest return periods are seen for the wet years, followed by the whole time series and dry years. There are minimal differences between the estimated return periods for the whole time series and wet years. For example, the estimated return period for rainfall amount of more than 60 mm and rainfall duration of 5 consecutive days is 222 days for whole time series, compared to 199 days for wet days. The difference between these values is about 10%. The estimated return period of dry years (with the same conditions) is 270 days which gives a difference of 27%. Higher differences can be observed for rainfall amounts of 90, 120 and 150 mm.

This analysis shows that there is a difference in the estimated return periods between the whole time series, wet years and dry years. The dry years give larger return periods for all rainfall amounts, while the wet periods show the smallest return period.

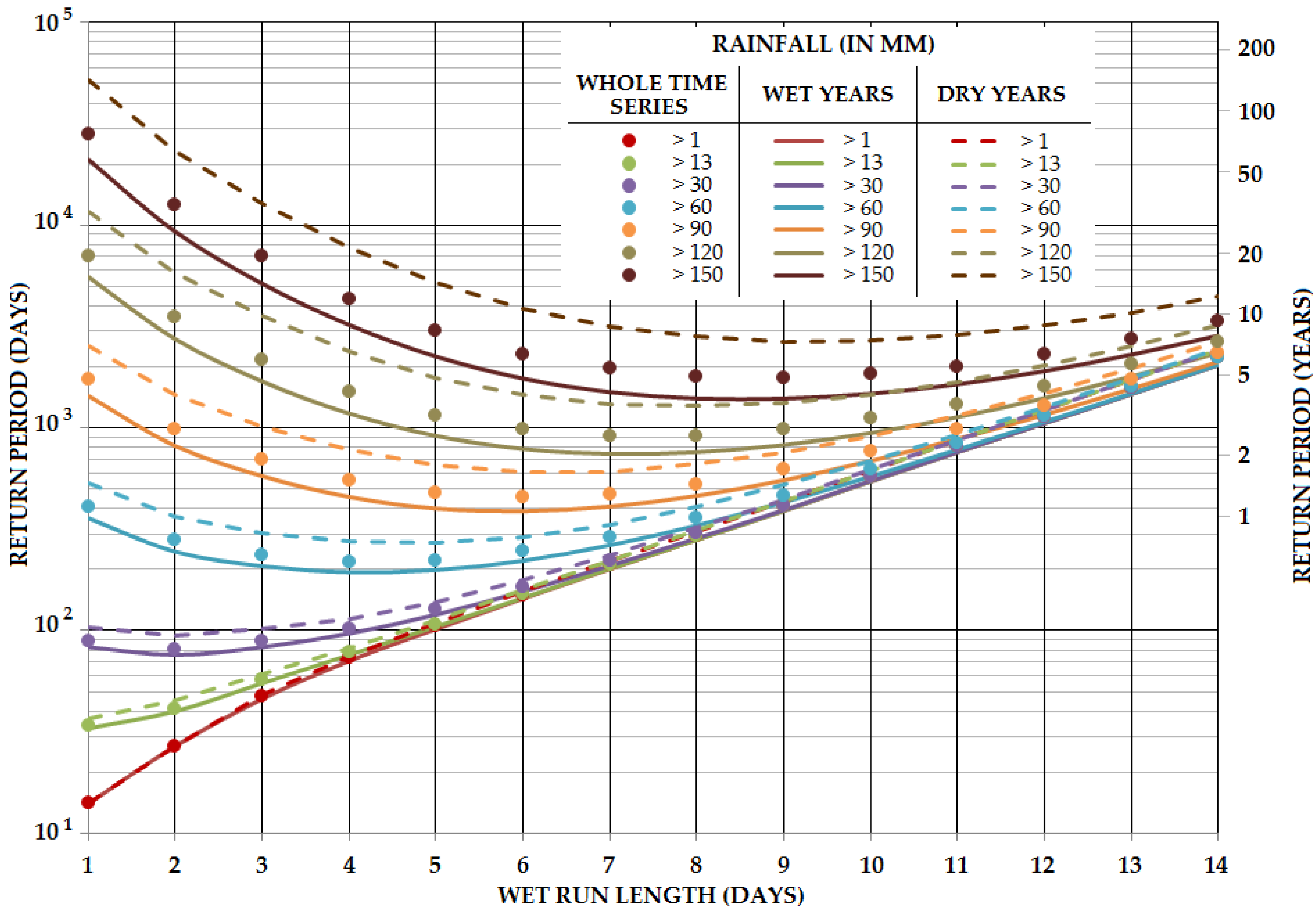

Figure B10 Return period curves for the whole time series, wet years and dry years

CONCLUSIONS

The statistics of wet and dry years, such as the mean, standard deviation, probability distributions and mean wet and dry run lengths, conditional probabilities and return periods are compared in this section. There are a few differences in the statistics of wet and dry years.

The estimated average daily rainfall during the period wet years is higher at 11.72 mm, as compared to 13.84 mm for dry years. Similarly, the standard deviation of the daily rainfall estimated for the wet years is 18.23 mm and for the dry years is 15.79 mm, which gives a difference of 13%.

The mean wet and dry run lengths for wet and dry years do not change significantly. However, the conditional probabilities for t-consecutive wet (dry) days are higher during wet (dry) years as compared to the dry (wet) years. The rainfall amounts for both scenarios are dependent from one day to another.

Significant differences are shown in the estimation of return periods for wet and dry years. Shorter return periods are estimated for wet years for all rainfall amounts that are considered in this analysis. The bigger differences are shown for the large amount of rainfalls, i.e., 60, 90, 120 and 150 mm.

APPENDIX C

FREQUENCY ANALYSIS FOR THE ANNUAL MAXIMUM DAILY RAINFALL AT SUBANG AIRPORT

The Cumulative Distribution Function (CDF) for the observed annual maximum daily rainfalls from 1960 to 2011 at Subang Airport is represented using the plotting position formula known as the Weibull method. The formula for the Weibull method is given in Eq. C1.

$$F(x) = \frac{i}{N+1} \qquad \text{(Eq. C1)}$$

Where: x = annual maximum daily rainfall (mm)

i = rank (ordered sample from the smallest to the largest)

N = sample size

Log-Pearson Type III distribution (LPIII) is used to fit the annual maximum daily rainfalls at Subang Airport. The probability density function of LPIII is given in Eq. C2.

$$f(x) = \frac{0.4343}{|\alpha|\Gamma(\beta)x}\left[\frac{\log(x) - y_o}{\alpha}\right]^{\beta-1} \exp\left[-\frac{\log(x) - y_o}{\alpha}\right] \qquad \text{(Eq. C2)}$$

Where: x = annual maximum daily rainfall (mm)

β = shape parameter

α = scale parameter

y_o = location parameter

LPIII has three parameters, namely the shape (β), scale (α) and location (y_o). These parameters are estimated based on the log transformation of the annual maximum daily rainfall i.e., $Y = \log x$. The indirect method of moments is used to estimate these parameters and the formulations are given in Eq. C3 to Eq. C5.

$$\hat{\beta} = \left(\frac{2}{\hat{\gamma}}\right)^2 \quad \text{(Eq. C3)}$$

$$\hat{\alpha} = \frac{\hat{\sigma}\hat{\gamma}}{2} \quad \text{(Eq. C4)}$$

$$\widehat{y_o} = \hat{\mu} - \hat{\alpha}\hat{\beta} \quad \text{(Eq. C5)}$$

Where: $\hat{\mu}$ = sample mean

$\hat{\sigma}$ = sample standard deviation

$\hat{\gamma}$ = sample skewness coefficient

The CDF for the annual maximum daily rainfall at Subang Airport is shown in Figure C1. The Kolmogorov-Smirnoff (KS) method is used to test the goodness of fit for the fitted CDF at quantile point of 0.95. The maximum difference between the empirical and fitted CDF is 0.084, which is well below the KS test statistical value of 0.189. Therefore, it is concluded that the LPIII is suitable to represent the distribution of annual maximum daily rainfall at Subang Airport.

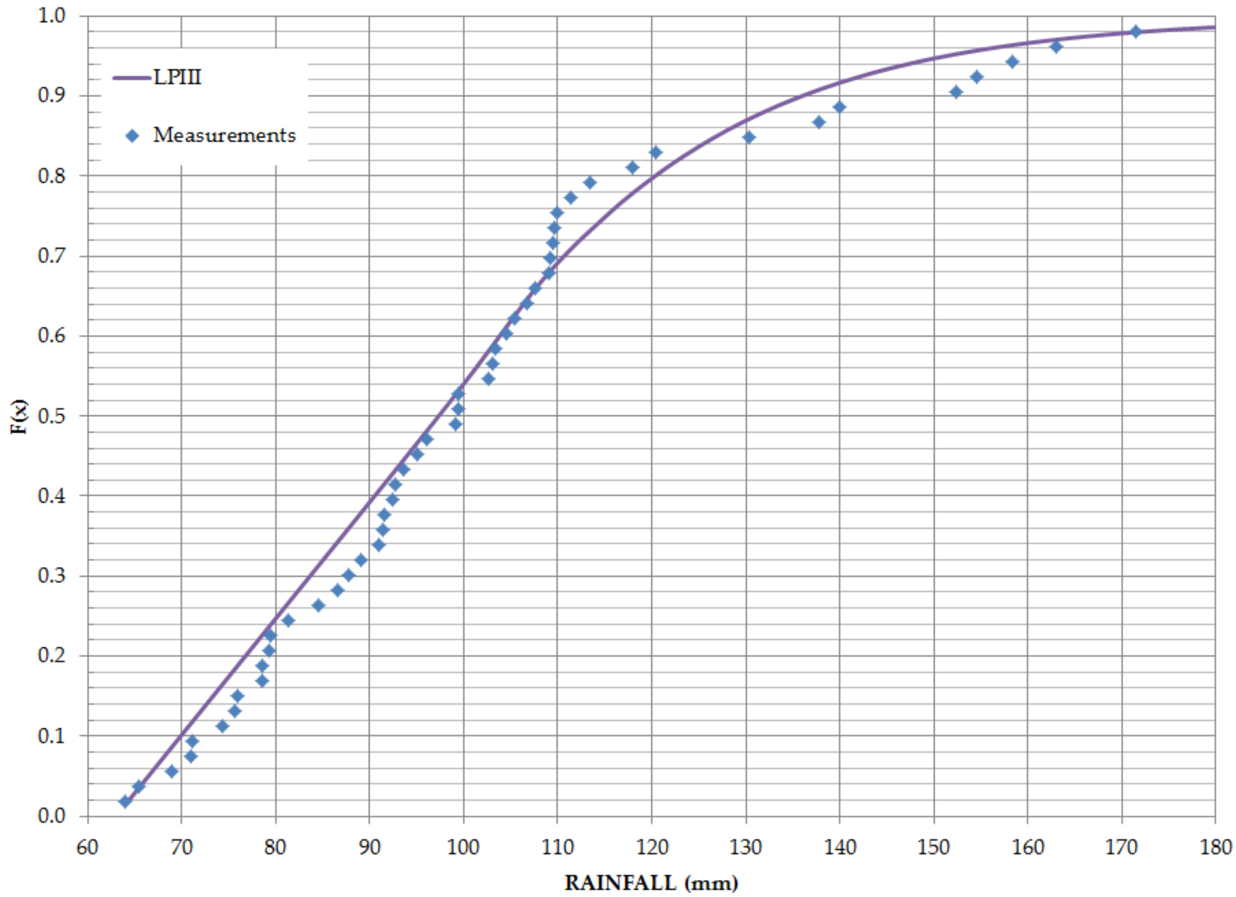

Figure C1 Empirical and fitted CDF using LPIII for the annual maximum daily rainfall at Subang Airport

CONFIDENCE LIMITS ON QUANTILES OF THE LOG-PEARSON TYPE III DISTRIBUTION

This section summarized the estimation of 95% Confidence Limits (CL) on the quantiles of return periods of 10, 25, 50, 100 and 500 years for the annual maximum rainfall at Subang Airport. The formulations for the calculations are given in Eq. C6 to Eq. C13.

$$CL_{1-\alpha}(Y_q) = \widehat{Y_q} \pm z_{1-\alpha/2}\widehat{S_q} \qquad \text{(Eq. C6)}$$

Where: $\widehat{Y_q}$ = quantile estimator corresponding to the non-exceedence probability q

$z_{1-\alpha/2}$ = 1-α quantile of the standard normal deviation

α = significance level

$\widehat{S_q}$ = standard error

The Y_q and S_q are estimated using the formulations given in Eq. C7 and C11.

$$\widehat{Y_q} = \hat{\mu} + \widehat{K_q}\hat{\sigma} \qquad \text{(Eq. C7)}$$

Where: $\hat{\mu}$ = sample mean

$\hat{\sigma}$ = sample standard deviation

$\widehat{K_q}$ = frequency factor, which is calculated using the Eq. C8 to Eq. C9

$$\widehat{K_q} = z_q + (z_q^2 - 1)\left(\frac{\widehat{\gamma_1}}{6}\right) + \left(\frac{1}{3}\right)(z_q^3 - 6z_q)\left(\frac{\widehat{\gamma_1}}{6}\right)^2 - (z_q^2 - 1)\left(\frac{\widehat{\gamma_1}}{6}\right)^3 + z_q\left(\frac{\widehat{\gamma_1}}{6}\right)^4 + \left(\frac{1}{3}\right)\left(\frac{\widehat{\gamma_1}}{6}\right)^5$$

(Eq. C8)

$$\widehat{\gamma_1} = \frac{2\hat{\alpha}}{|\hat{\alpha}|\sqrt{\hat{\beta}}} \qquad \text{(Eq. C9)}$$

$$\widehat{S_q^2} = \frac{\widehat{\sigma^2}}{N}\left[1 + \widehat{K_q}\widehat{\gamma_1} + \widehat{K_q^2}\left(\frac{1}{2} + \frac{3}{8}\widehat{\gamma_1^2}\right) + \widehat{K_q}\left(3\widehat{\gamma_1} + \frac{3}{4}\widehat{\gamma_1^3}\right)\left(\frac{\delta K_q}{\delta\gamma_1}\right) + \left(6 + 9\widehat{\gamma_1^2} + \frac{15}{8}\widehat{\gamma_1^4}\right)\left(\frac{\delta K_q}{\delta\gamma_1}\right)^2\right]$$

(Eq. C10)

$$\frac{\delta K_q}{\delta \gamma_1} = \frac{(z_q^2 - 1)}{6} + \frac{z_q(z_q^2 - 6)}{54}\widehat{\gamma_1} - \frac{(z_q^2 - 1)}{72}\widehat{\gamma_1}^2 + \frac{z_q}{324}\widehat{\gamma_1}^3 + \frac{5}{23328}\widehat{\gamma_1}^4 \qquad (Eq.\ C11)$$

Finally, the confidence limits and quantile estimator must be transformed from the log form, i.e.,

$$CL_{1-\alpha}(X_q) = \exp[CL_{1-\alpha}(Y_q)] \qquad \text{(Eq. C12)}$$

$$X_q = \exp(Y_q) \qquad \text{(Eq. C13)}$$

Where: $CL_{1-\alpha}(Y_q)$ = estimated confidence limits of the corresponding LPIII

Y_q = estimated for the quantile of the corresponding LPIII

X_q = estimated for the quantile, transformed from LPIII

Table C1 summarizes the 95% confidence limits and quantile values for return periods of 10, 25, 50, 100 and 500 years. Figure C2 show the plot of 95% confidence limits and quantile values estimated using LPIII.

In general, the observed values are well within the estimated upper and lower limits of LPIII distribution. With reference to Figure C2, the observed annual maximum daily rainfalls for return periods of 1 to 3 years are close to the estimated quantile value. After that, the observed values for return periods of 3 to 6 years are close to the estimated lower limits. The estimated annual maximum daily rainfall for the 10-years return period is between 124 to 149 mm, while the observed value is slightly higher at 152 mm.

Attention should also be given to the upper and lower limits calculated for the return periods of 25 and 50 years. The observed values for these return periods are well within the confidence limits estimated using the LPIII distribution. Based on these findings, it is concluded that the LPIII distribution is able to give reasonable estimates of annual maximum daily rainfalls of rare events, such as the return periods of 100 or more. For example, the annual maximum daily rainfall for return period of 100 years is estimated to be between 154 to 229 mm. The annual maximum daily rainfall with the return period of 500 years is expected to be in the range of 169 to 303 mm.

Table C1 Confidence Limits and Quantile for the annual maximum daily rainfall at Subang Airport estimated using the LPIII distribution

Return Period, T	Non-exceedence Probability, q	Lower Limit (mm)	Quantile (mm)	Upper Limit (mm)
10	0.90	124	136	149
25	0.96	137	156	177
50	0.98	146	172	202
100	0.99	154	187	229
500	0.998	169	226	303

GENERATED ANNUAL MAXIMUM DAILY RAINFALL AT SUBANG AIRPORT

The sequence of daily rainfall at Subang Airport from 1960 to 2011 is simulated using the DARMA(1,1) model. This model is chosen because the generated sequence of daily rainfall has similar statistical properties as the measured daily rainfall at

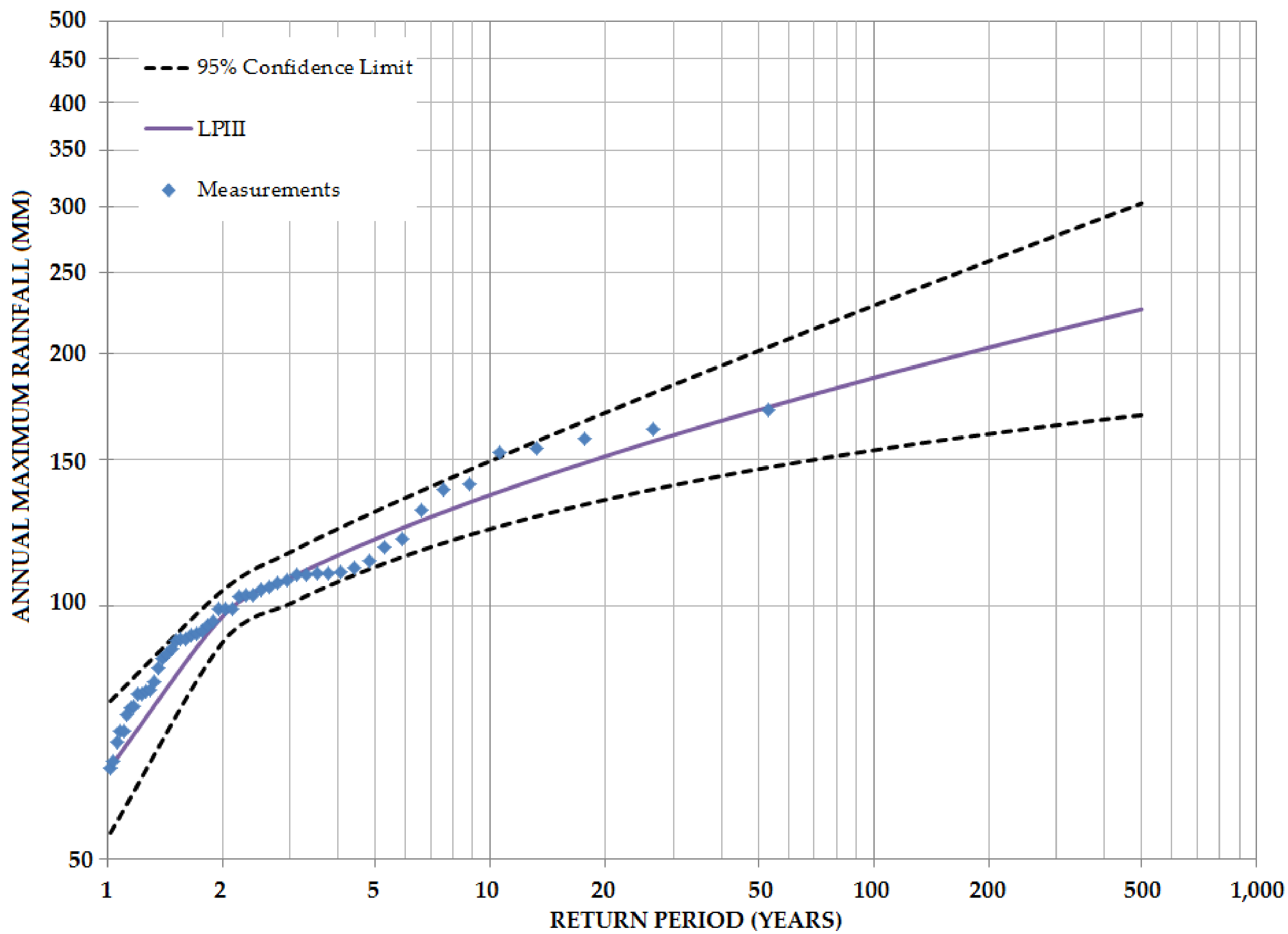

Figure C2 Empirical frequency distribution, fitted CDF and 95% confidence limits on quantiles for the LPIII distribution for the annual maximum daily rainfalls of the Subang Airport

Subang Airport. The two-parameter gamma function has been shown to represent the amount of rain at this particular station.

Two simulations are done, i.e., Simulation A using the parameters derived from the 52 years of observed data (from 1960 to 2011) and Simulation B where parameters are estimated based on the statistical properties of the last 25 years of measured data (from 1987 to 2011). The return periods examined in this section are 10, 25, 50, 100 and 500 years. For each simulation and return period, 1,000 samples are generated in order to give a range of annual maximum rainfall values. The generated sequences of daily rainfalls are then divided into individual groups of 365 days. Then the highest values for each group are recorded as the annual maximum daily rainfall.

There are three parameters in DARMA(1,1) need to be estimated, namely λ, β and π_1(or π_0). The DARMA(1,1) model parameters for Simulations A and B are given in Table C2. There are no significant differences in the values of DARMA(1,1) parameters between Simulations A and B. The estimated values of λ and β for Simulation A are 0.8445 and 0.5446, respectively while Simulation B gives the estimation of $\hat{\lambda} = 0.8284$ and $\hat{\beta} = 0.5490$.

Table C2 DARMA(1,1) model parameters for Simulations A and B

Simulation	Model Parameters			
	$\hat{\lambda}$	$\hat{\beta}$	$\widehat{\pi_1}$	$\widehat{\pi_0}$
A	0.8445	0.5466	0.5314	0.4686
B	0.8284	0.5490	0.5439	0.4561

For Simulation A, the estimated wet and dry probability distributions are $\widehat{\pi_1} = 0.5314$ and $\widehat{\pi_0} = 0.4686$, respectively. Simulation B shows a slightly higher value of wet probability distribution, i.e., $\widehat{\pi_1} = 0.5439$, which resulted in a smaller value of dry probability distribution, $\widehat{\pi_0} = 0.4561$. Table C2 summarizes the DARMA(1,1) model parameters for simulations A and B.

Table C3 summarizes the range of annual maximum daily rainfall estimated using LPIII, Simulation A and Simulation B. In general, Simulations A and B are capable to produce reasonable annual maximum daily rainfall and wider range of values when it is compared with the LPIII. For example, the estimated annual maximum daily rainfall using the LPIII method for return period of 25 years is between 137 to 177 mm; while Simulation A gives the estimated value of between 135 to 243 mm. For the same return period, Simulation B provides an estimation of 143 to 254 mm of annual maximum daily rainfall.

Table C3 Annual maximum daily rainfall estimated using LPIII, Simulations A and B

Return Periods (years)	LPIII			Simulation A – 1000 Generated Samples (95% CL)			Simulation B - 1000 Generated Samples (95% CL)		
	Lower Limit (mm)	Quantile Estimate (mm)	Upper Limit (mm)	Lower Limit (mm)	Mean (mm)	Upper Limit (mm)	Lower Limit (mm)	Mean (mm)	Upper Limit (mm)
10	124	136	149	115	156	226	122	164	233
25	137	156	177	135	177	243	143	186	254
50	146	172	202	151	192	264	157	200	272
100	154	187	229	166	208	284	175	218	290
500	169	226	303	202	241	307	212	255	331

The range of annual maximum daily rainfall for each return periods, i.e., 10, 25, 50, 100 and 500 years for Simulations A and B are shown graphically in Figures C3 and C4, respectively.

With reference to Figure C3, Simulation A gives a smaller minimum value for return periods of 10, 25 and 50 years as compared with the lower limit estimated using the LPIII. Simulation A also shows that the estimated mean for return periods of 10, 25 and 50 years are close to the upper limits calculated using the LPIII method. The range of values for these return periods are also within the measured value.

For higher return periods, i.e., 100 and 500 years, the minimum values are closer to lower limits calculated with LPIII. The mean annual maximum daily rainfalls in Simulation A for these return periods are close to the quantile estimated given by LPIII. Additionally, the range of annual maximum daily rainfall in Simulation A for return periods of 100 and 500 years are reasonable.

Figure C4 shows that the annual maximum daily rainfalls estimated from Simulation B have a wider range than the values determined using the LPIII method. For example, Simulation B estimated that the annual maximum daily rainfall for return period of 10 years to be between 104 to 298 mm, while the LPIII method estimated the value to be in the range of 124 to 149 mm. The minimum values for return periods of 10 and 25 years are smaller than the lower limits estimated using the LPIII method. It should also be noted that the means estimated from Simulation B are close to the upper limits determined by LPIII method.

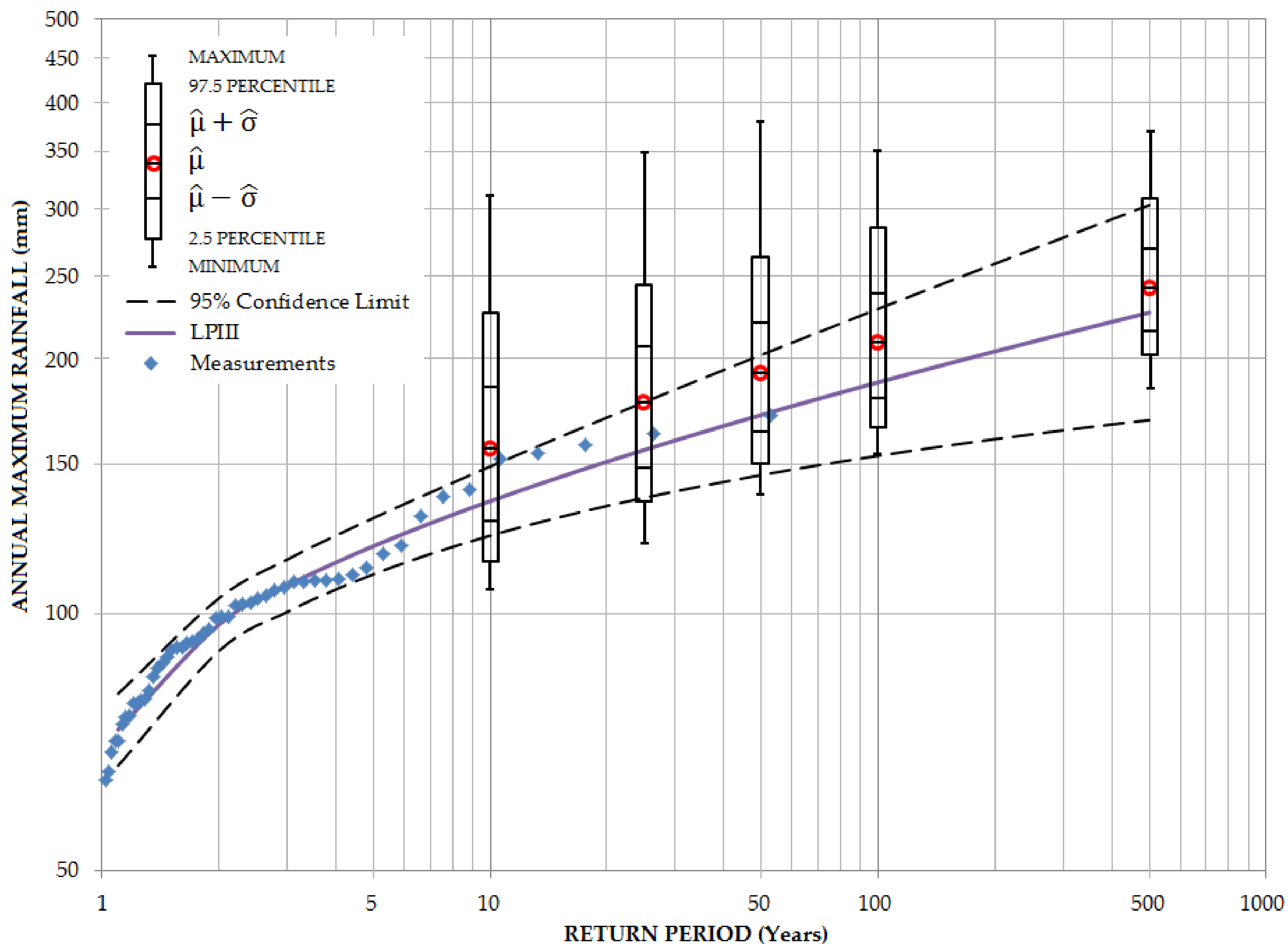

Figure C3 Range of simulated annual maximum rainfall for return periods 10, 25, 50, 100 and 500 years for whole time series

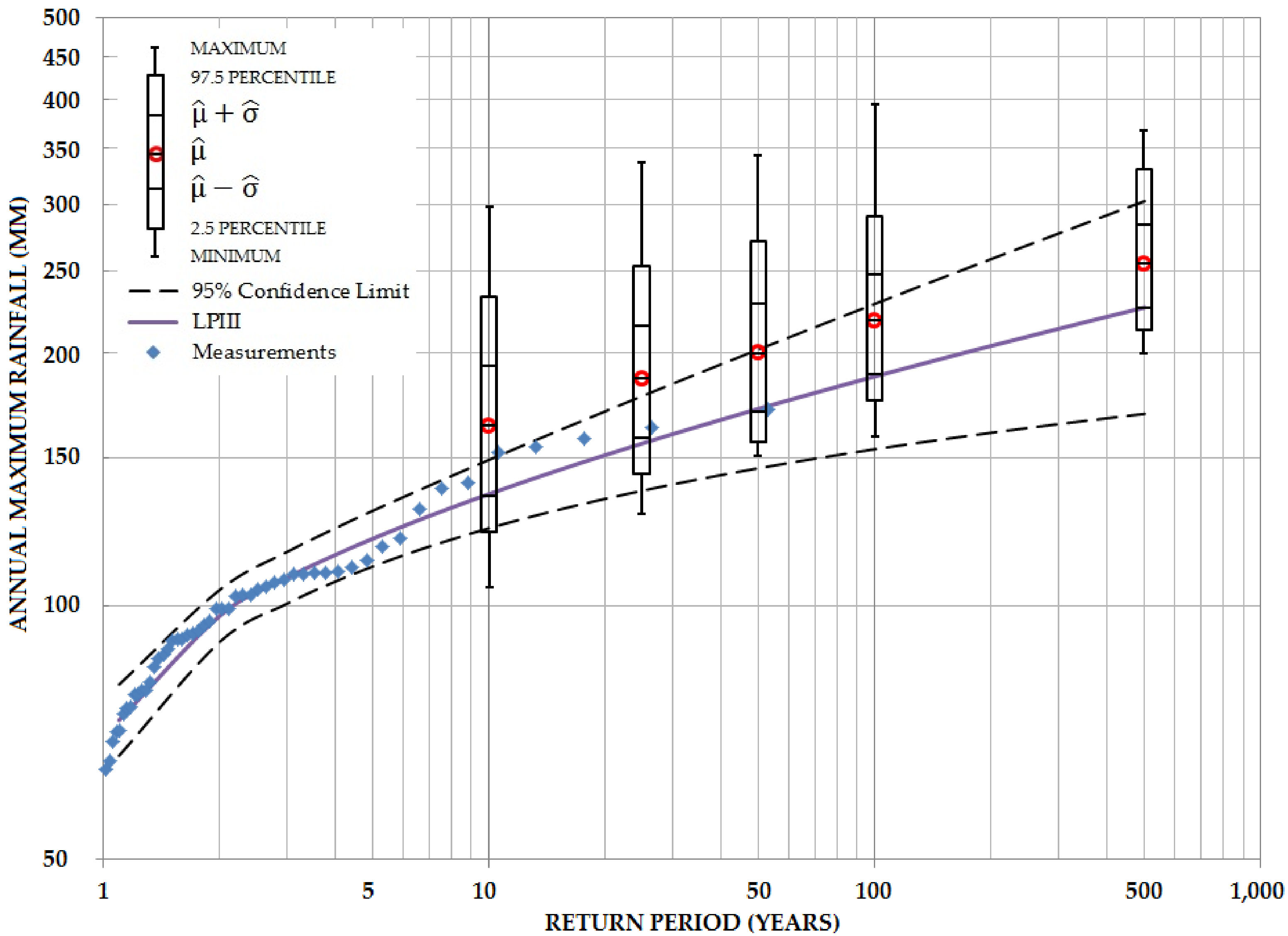

Figure C4 Range of simulated annual maximum rainfall for return periods 10, 25, 50, 100 and 500 years for wet years

Simulation A gives a wider range of annual maximum daily rainfall for all return periods when it is compared with Simulation B. However, considering that the parameters from Simulation B are estimated from a shorter period of observed data, the results are comparable to the values from Simulation A. Table C4 shows the difference (in percentage) between the annual maximum daily rainfall estimated from Simulations A and B. In general the differences for all return periods are very small, with an average of 4.8%. The highest difference is shown at the upper limit for return period of 500 years, with 7.8%.

Table C4 Percentage difference between the estimated values from Simulations A and B

Return Periods (years)	Difference (%)		
	Lower Limit (mm)	Mean (mm)	Upper Limit (mm)
10	6.1	5.1	3.1
25	5.9	5.1	4.5
50	4.0	4.2	3.0
100	5.4	4.8	2.1
500	5.0	5.8	7.8

CONCLUSION

The estimated annual maximum daily rainfalls using the parameters calculated from 52 years of sample gives more reasonable values when it is compared with the measured data. It should be noted that the annual maximum daily rainfalls estimated from simulation B (using the parameters from a shorter measurement) show encouraging results even though bigger differences are calculated when they are compared to simulation A.