Lampiran 20

UTM/RMC/F/0024 (1998)

UNIVERSITI TEKNOLOGI MALAYSIA

BORANG PENGESAHAN
LAPORAN AKHIR PENYELIDIKAN

TAJUK PROJEK : MODELING OF CONVECTIVE RAIN FOR PREDICTING FLASH FLOOD

Saya PM DR ZALINA BT MOHD DAUD
(HURUF BESAR)

Mengaku membenarkan **Laporan Akhir Penyelidikan** ini disimpan di Perpustakaan Universiti Teknologi Malaysia dengan syarat-syarat kegunaan seperti berikut :

1. Laporan Akhir Penyelidikan ini adalah hakmilik Universiti Teknologi Malaysia.
2. Perpustakaan Universiti Teknologi Malaysia dibenarkan membuat salinan untuk tujuan rujukan sahaja.
3. Perpustakaan dibenarkan membuat penjualan salinan Laporan Akhir Penyelidikan ini bagi kategori TIDAK TERHAD.
4. * Sila tandakan (/)

[ ]	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau Kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972).
[ ]	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh Organisasi/badan di mana penyelidikan dijalankan).
[√]	TIDAK TERHAD	

TANDATANGAN KETUA PENYELIDIK

DR. ZALINA MOHD DAUD
Assoc Professor
College of SC & Tech
UTM City Campus Jln Semarak
54100 Kuala Lumpur

Nama & Cop Ketua Penyelidik

Tarikh : 12 JUN 2008

CATATAN : * *Jika Laporan Akhir Penyelidikan ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh laporan ini perlu dikelaskan sebagai SULIT dan TERHAD.*

VOT 74280

MODELING OF CONVECTIVE RAIN FOR PREDICTING FLASH FLOOD

(PERMODELAN HUJAN PEROLAKAN UNTUK MERAMALKAN BANJIR KILAT)

ZALINA BINTI MOHD DAUD
FADHILAH BINTI YUSOF
ZULKIFLI BIN YUSOP
MAIZAH HURA BINTI AHMAD
WAN AZLI BIN WAN HASAN
MOHD NOR BIN MOHD DESA
ROBIAH BINTI ADNAN
NORDILA BINTI AHMAD
NURUL HUDA BINTI MD ADNAN
MOHD AFTAR BIN ABU BAKAR

PUSAT PENGURUSAN PENYELIDIKAN
UNIVERSITI TEKNOLOGI MALAYSIA

2008

VOT 74280

MODELING OF CONVECTIVE RAIN FOR PREDICTING FLASH FLOOD

(PERMODELAN HUJAN PEROLAKAN UNTUK MERAMALKAN BANJIR KILAT)

ZALINA BINTI MOHD DAUD
FADHILAH BINTI YUSOF
ZULKIFLI BIN YUSOP
MAIZAH HURA BINTI AHMAD
WAN AZLI BIN WAN HASAN
MOHD NOR BIN MOHD DESA
ROBIAH BINTI ADNAN
NORDILA BINTI AHMAD
NURUL HUDA BINTI MD ADNAN
MOHD AFTAR BIN ABU BAKAR

RESEARCH VOTE NO:
74280

Jabatan Sains
Kolej Sains & Teknologi
Universiti Teknologi Malaysia

2008

ACKNOWLEDGEMENT

This research was undertaken through the funding by the Ministry of Science, Technology and Innovation under the administration of the Research Management Centre, Universiti Teknologi Malaysia. Invaluable radar data were provided by the Malaysian Meteorology Department and rainfall data were provided by the Department of Irrigation and Drainage Malaysia. We are indebted to the staffs of MMD KLIA for their assistance and inputs on the acquisition of the Doppler Radar data. Our sincere appreciation also to the staffs of the Data Bank in DID Ampang for entertaining our frequent requests of rainfall data. Our heartfelt thanks to Ir. Mohd Zaki Mat Amin of DID Ampang for his invaluable assistance in the development of the IDF curves, Prof Ahris Yaakub of Faculty of Built Environment UTM and his Masters student for teaching the use of as well as sharing the ArcGIS 9.1 software. This research has initiated a number of Masters and a Doctoral thesis which we hope and believe will further propagate more research endeavours towards enriching the body of knowledge and providing more practical solutions to the relevant field of applications.

PERMODELAN HUJAN PEROLAKAN UNTUK MERAMALKAN BANJIR KILAT

(Kata kunci: hujan perolakan, banjir kilat, Neyman-Scott, MCME, MARIMA)

Hujan perolakan lebat biasanya adalah penyebab bagi peristiwa hidrometeorologi ekstrim. Ini termasuk kebanyakkan peristiwa banjir kilat yang merupakan salah satu fenomena paling merosakkan berhubung iklim terutama di kawasan bandar di Malaysia. Untuk kajian ini data hujan diperoleh daripada radar dan juga tolok hujan bagi kawasan Lembah Klang. Hujan perolakan diklasifikasikan kepada perolakan kecil, sederhana dan kuat berdasarkan nilai β. Lengkung '*Areal Reduction Factor*' (ARF) yang diperoleh dari kajian ini adalah setara dengan nilai ARF yang diperoleh dari pengkaji terdahulu. Lengkung '*Intensity Duration Frequency*' (IDF) yang diplot berdasarkan hujan perolakan sahaja mempunyai keamatan yang lebih tinggi jika dibandingkan dengan lengkung IDF yang sedia ada dan mungkin lebih sesuai bagi menentukan *design storms* bagi kawasan bandar yang mengalami banyak hujan perolakan. Siri data sintetik pula dijana bagi mengatasi masalah kekurangan data bertempoh pendek. Dua model stokastik yang terkemuka iaitu model berasaskan proses Neyman-Scott Rectangular Pulses (NSRP) dan Markov Chain Mixed Exponential (MCME) digunakan. Hasil penilaianan model menggunakan data hujan setiap jam selama 10 tahun bagi stesen 3217001 di Wilayah Persekutuan menunjukkan model NSRP berupaya mengekal beberapa ciri statistik dan fizikal pada tempoh masa yang berbeza (1, 6, dan 24-jam). Penilaian kualitatif dan berangka di antara model NSRP dan MCME menunjukkan kedua-dua model adalah setanding dalam keupayaan mengekal ciri skala sejam, walaupun keupayaan diskriptif mengatasi keupayaan menelah.. Bagaimana pun kedua-dua model berupaya mengekal trend bermusim seperti ciri data tercerap. Untuk penelahan siri data sejam model Multivariat Autoregresi Terkamir Purata Bergerak (MARIMA) digunakan. Perbandingan dengan model Autoregresif Purata Bergerak (ARMA) menunjukkan hasil yang setara dan ini memungkinkan MARIMA berpotensi sebagai model penelahan.

Penyelidik
Prof. Madya Dr. Zalina Mohd Daud (Ketua)
Prof. Dr. Zulkifli Yusof
Prof. Madya Dr. Maizah Hura Ahmad
Prof. Madya Dr. Robiah Adnan
Dr. Wan Azli Wan Hassan
Ir. Dr. Mohd Nor Mohd Desa
Dr. Fadhilah Yusof
Cik Nordila Ahmad
Cik Nurul Huda Md Adnan
Encik Mohd Aftar Abu Bakar

e-mail : zalina@citycampus.utm.my
Tel: 03-26154662
No. Vote:74280

MODELING OF CONVECTIVE RAIN FOR PREDICTING FLASH FLOOD

(Keyword: convective rain, flash flood, Neyman-Scott, MCME, MARIMA)

Intense convective rain cells are often responsible for extreme hydrometeorological events including the majority of flash flood episodes, which is one of the most common and destructive weather-related phenomena especially in urban areas of Malaysia. Both ground and radar data from the Klang Valley were the inputs of this study on the spatial and temporal characteristics of convective rains. A classification based on the β value was used to differentiate the slightly, moderately and strongly convective rains. The areal reduction factor (ARF) obtained from this study is comparable with ARF values obtained earlier by other researchers. An intensity duration frequency (IDF) curve plotted based only on convective storms generally result in higher storm intensity compared to the existing IDF curve and is potentially more appropriate for determining design storms for urban areas with high occurrence of convective events. Synthetic rainfall data series was generated to overcome lack of short duration data series. Two predominant stochastic rainfall model namely a point-process model based on the Neyman-Scott Rectangular Pulses (NSRP) stochastic process and the Markov Chain Mixed Exponential (MCME) was employed. Results of the model evaluation using a 10-year hourly rainfall record at station 3217001 in the Wilayah Persekutuan indicated that NSRP models describe adequately various statistical and physical properties at different timescales (1, 6, and 24-hour). Qualitative and numerical evaluation between the NSRP and MCME models indicated both models have comparable abilities in preserving the properties at the hourly scales, even though the models' descriptive ability fared better than their predictive ability. However, they were able to preserve the seasonal trend of the observed properties. For forecasting hourly rainfall series, the Multivariate Autoregressive Integrated Moving Average (MARIMA) model was employed. A comparison with an autoregressive moving average model (ARMA) showed comparable results which highlights the potential of the MARIMA model as a forecasting method.

Key Researchers
Assoc. Prof Dr. Zalina Mohd Daud (Head)
Prof. Dr. Zulkifli Yusof
Assoc. Prof Dr. Maizah Hura Ahmad
Assoc. Prof Dr. Robiah Adnan
Dr. Wan Azli Wan Hassan
Ir. Dr. Mohd Nor Mohd Desa
Dr. Fadhilah Yusof
Ms. Nordila Ahmad
Ms. Nurul Huda Md Adnan
Mr. Mohd Aftar Abu Bakar

e-mail : zalina@citycampus.utm.my
Tel: 03-26154662
Vote No:74280

ABSTRACT

Intense convective rain cells are often responsible for extreme hydrometeorological events including the majority of flash flood episodes, which is one of the most common and destructive weather-related phenomena especially in urban areas of Malaysia. Both ground and radar data from the Klang Valley were the inputs of this study on the spatial and temporal characteristics of convective rains. A classification based on the β value was used to differentiate the slightly, moderately and strongly convective rains. The areal reduction factor (ARF) obtained from this study is comparable with ARF values obtained earlier by other researchers. An intensity duration frequency (IDF) curve plotted based only on convective storms generally result in higher storm intensity compared to the existing IDF curve and is potentially more appropriate for determining design storms for urban areas with high occurrence of convective events. Synthetic rainfall data series was generated to overcome lack of short duration data series. Two predominant stochastic rainfall model namely a point-process model based on the Neyman-Scott Rectangular Pulses (NSRP) stochastic process and the Markov Chain Mixed Exponential (MCME) was employed. Results of the model evaluation using a 10-year hourly rainfall record at station 3217001 in the Wilayah Persekutuan indicated that NSRP models describe adequately various statistical and physical properties at different timescales (1, 6, and 24-hour). Qualitative and numerical evaluation between the NSRP and MCME models indicated both models have comparable abilities in preserving the properties at the hourly scales, even though the models' descriptive ability fared better than their predictive ability. However, they were able to preserve the seasonal trend of the observed properties. For forecasting hourly rainfall series, the Multivariate Autoregressive Integrated Moving Average (MARIMA) model was employed. A comparison with an autoregressive moving average model (ARMA) showed comparable results which highlights the potential of the MARIMA model as a forecasting method.

ABSTRAK

Hujan perolakan lebat biasanya adalah penyebab bagi peristiwa hidrometeorologi ekstrim. Ini termasuk kebanyakkan peristiwa banjir kilat yang merupakan salah satu fenomena paling merosakkan berhubung iklim terutama di kawasan bandar di Malaysia. Untuk kajian ini data hujan diperoleh daripada radar dan juga tolok hujan bagi kawasan Lembah Klang. Hujan perolakan diklasifikasikan kepada perolakan kecil, sederhana dan kuat berdasarkan nilai β. Lengkung '*Areal Reduction Factor*' (ARF) yang diperoleh dari kajian ini adalah setara dengan nilai ARF yang diperoleh dari pengkaji terdahulu. Lengkung '*Intensity Duration Frequency*' (IDF) yang diplot berdasarkan hujan perolakan sahaja mempunyai keamatan yang lebih tinggi jika dibandingkan dengan lengkung IDF yang sedia ada dan mungkin lebih sesuai bagi menentukan *design storms* bagi kawasan bandar yang mengalami banyak hujan perolakan. Siri data sintetik pula dijana bagi mengatasi masalah kekurangan data bertempoh pendek. Dua model stokastik yang terkemuka iaitu model berasaskan proses Neyman-Scott Rectangular Pulses (NSRP) dan Markov Chain Mixed Exponential (MCME) digunakan. Hasil penilaianan model menggunakan data hujan setiap jam selama 10 tahun bagi stesen 3217001 di Wilayah Persekutuan menunjukkan model NSRP berupaya mengekal beberapa ciri statistik dan fizikal pada tempoh masa yang berbeza (1, 6, dan 24-jam). Penilaian kualitatif dan berangka di antara model NSRP dan MCME menunjukkan kedua-dua model adalah setanding dalam keupayaan mengekal ciri skala sejam, walaupun keupayaan diskriptif mengatasi keupayaan menelah.. Bagaimana pun kedua-dua model berupaya mengekal trend bermusim seperti ciri data tercerap. Untuk penelahan siri data sejam model Multivariat Autoregresi Terkamir Purata Bergerak (MARIMA) digunakan. Perbandingan dengan model Autoregresif Purata Bergerak (ARMA) menunujukkan hasil yang setara dan ini memungkinkan MARIMA berpotensi sebagai model penelahan.

MODELING OF CONVECTIVE RAIN FOR PREDICTING FLASH FLOOD

(PERMODELAN HUJAN PEROLAKAN UNTUK MERAMALKAN BANJIR KILAT)

ZALINA BINTI MOHD DAUD
FADHILAH BINTI YUSOF
ZULKIFLI BIN YUSOP
MAIZAH HURA BINTI AHMAD
WAN AZLI BIN WAN HASAN
MOHD NOR BIN MOHD DESA
ROBIAH BINTI ADNAN
NORDILA BINTI AHMAD
NURUL HUDA BINTI MD ADNAN
MOHD AFTAR BIN ABU BAKAR

RESEARCH VOTE NO:
74280

Jabatan Sains
Kolej Sains & Teknologi
Universiti Teknologi Malaysia

2008

ACKNOWLEDGEMENT

This research was undertaken through the funding by the Ministry of Science, Technology and Innovation under the administration of the Research Management Centre, Universiti Teknologi Malaysia. Invaluable radar data were provided by the Malaysian Meteorology Department and rainfall data were provided by the Department of Irrigation and Drainage Malaysia. We are indebted to the staffs of MMD KLIA for their assistance and inputs on the acquisition of the Doppler Radar data. Our sincere appreciation also to the staffs of the Data Bank in DID Ampang for entertaining our frequent requests of rainfall data. Our heartfelt thanks to Ir. Mohd Zaki Mat Amin of DID Ampang for his invaluable assistance in the development of the IDF curves, Prof Ahris Yaakub of Faculty of Built Environment UTM and his Masters student for teaching the use of as well as sharing the ArcGIS 9.1 software. This research has initiated a number of Masters and a Doctoral thesis which we hope and believe will further propagate more research endeavours towards enriching the body of knowledge and providing more practical solutions to the relevant field of applications.

ABSTRACT

Intense convective rain cells are often responsible for extreme hydrometeorological events including the majority of flash flood episodes, which is one of the most common and destructive weather-related phenomena especially in urban areas of Malaysia. Both ground and radar data from the Klang Valley were the inputs of this study on the spatial and temporal characteristics of convective rains. A classification based on the β value was used to differentiate the slightly, moderately and strongly convective rains. The areal reduction factor (ARF) obtained from this study is comparable with ARF values obtained earlier by other researchers. An intensity duration frequency (IDF) curve plotted based only on convective storms generally result in higher storm intensity compared to the existing IDF curve and is potentially more appropriate for determining design storms for urban areas with high occurrence of convective events. Synthetic rainfall data series was generated to overcome lack of short duration data series. Two predominant stochastic rainfall model namely a point-process model based on the Neyman-Scott Rectangular Pulses (NSRP) stochastic process and the Markov Chain Mixed Exponential (MCME) was employed. Results of the model evaluation using a 10-year hourly rainfall record at station 3217001 in the Wilayah Persekutuan indicated that NSRP models describe adequately various statistical and physical properties at different timescales (1, 6, and 24-hour). Qualitative and numerical evaluation between the NSRP and MCME models indicated both models have comparable abilities in preserving the properties at the hourly scales, even though the models' descriptive ability fared better than their predictive ability. However, they were able to preserve the seasonal trend of the observed properties. For forecasting hourly rainfall series, the Multivariate Autoregressive Integrated Moving Average (MARIMA) model was employed. A comparison with an autoregressive moving average model (ARMA) showed comparable results which highlights the potential of the MARIMA model as a forecasting method.

ABSTRAK

Hujan perolakan lebat biasanya adalah penyebab bagi peristiwa hidrometeorologi ekstrim. Ini termasuk kebanyakkan peristiwa banjir kilat yang merupakan salah satu fenomena paling merosakkan berhubung iklim terutama di kawasan bandar di Malaysia. Untuk kajian ini data hujan diperoleh daripada radar dan juga tolok hujan bagi kawasan Lembah Klang. Hujan perolakan diklasifikasikan kepada perolakan kecil, sederhana dan kuat berdasarkan nilai β. Lengkung '*Areal Reduction Factor'* (ARF) yang diperoleh dari kajian ini adalah setara dengan nilai ARF yang diperoleh dari pengkaji terdahulu. Lengkung '*Intensity Duration Frequency*' (IDF) yang diplot berdasarkan hujan perolakan sahaja mempunyai keamatan yang lebih tinggi jika dibandingkan dengan lengkung IDF yang sedia ada dan mungkin lebih sesuai bagi menentukan *design storms* bagi kawasan bandar yang mengalami banyak hujan perolakan. Siri data sintetik pula dijana bagi mengatasi masalah kekurangan data bertempoh pendek. Dua model stokastik yang terkemuka iaitu model berasaskan proses Neyman-Scott Rectangular Pulses (NSRP) dan Markov Chain Mixed Exponential (MCME) digunakan. Hasil penilaianan model menggunakan data hujan setiap jam selama 10 tahun bagi stesen 3217001 di Wilayah Persekutuan menunjukkan model NSRP berupaya mengekal beberapa ciri statistik dan fizikal pada tempoh masa yang berbeza (1, 6, dan 24-jam). Penilaian kualitatif dan berangka di antara model NSRP dan MCME menunjukkan kedua-dua model adalah setanding dalam keupayaan mengekal ciri skala sejam, walaupun keupayaan diskriptif mengatasi keupayaan menelah.. Bagaimana pun kedua-dua model berupaya mengekal trend bermusim seperti ciri data tercerap. Untuk penelahan siri data sejam model Multivariat Autoregresi Terkamir Purata Bergerak (MARIMA) digunakan. Perbandingan dengan model Autoregresif Purata Bergerak (ARMA) menunujukkan hasil yang setara dan ini memungkinkan MARIMA berpotensi sebagai model penelahan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOL/ABBREVIATIONS/NOTATIONS

β	-	Beta parameter for classifying convective rain
ΔT	-	Time interval of accumulation of the precipitation
L	-	Intensity threshold
N	-	total number of ΔT
dBZ	-	decibels of z
z	-	Reflectivity factor
ARF	-	Areal Reduction Factor
IDF	-	Intensity Duration frequency
POT	-	Peak Over Threshold
X_T	-	Quantile value
IDF	-	Intensity Frequency Duration
t	-	time
z_t	-	a time series at time t
ψ_{t-1}	-	information set available at time $t-1$
a_t / ε_t	-	one-step-ahead forecast error of z_t at time origin $t-1$ / white noise process
ρ	-	autocorrelation function
μ	-	mean
σ^2	-	variance
$\bar{z}$	-	sample mean of time series
s_z^2	-	sample variance of time series
γ	-	autocovariance function
r_k	-	estimate of the kth lag autocorrelation ρ_k

c_k	-	estimate of the kth lag autocovariance γ_k,
B	-	backward shift operators
∇	-	backward difference operators
$\phi_p(\mathrm{B})$	-	autoregressive operator of order p
$\theta_q(\mathrm{B})$	-	moving average operator of order q
$\boldsymbol{R}(s)$	-	covariance matrix of lag s
$\boldsymbol{\rho}(s)$	-	correlation matrix of lag s
$\boldsymbol{X}_t$	-	a n-variate time series at time t
$\boldsymbol{a}_i$	-	$n \times n$ matrix of the ith parameter for the autoregressive model
$\boldsymbol{b}_i$	-	$n \times n$ matrix of the ith parameter for the moving average model
$\boldsymbol{\alpha}(B)$	-	autoregressive operator of order p in a form of $n \times n$ matrices
$\boldsymbol{\beta}(B)$	-	moving average operator of order q in a form of $n \times n$ matrices
$\boldsymbol{I}$	-	identity matrix
$\boldsymbol{B}$	-	backward shift operators in a form of $n \times n$ matrices
$\boldsymbol{M}_0$	-	covariance matrix
$\boldsymbol{M}_1$	-	lag 1 covariance matrix
μ_{ε_t}	-	average value of the residuals (error)

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

1.1 Introduction

The globally accepted phenomena of climate fluctuations have received wide attention from all walks of life. As climate and weather dictates our life to a certain extent, even the uncertainty of a dry and wet season serves to disrupt designed activities. Studies on climate change are plentiful in the literature as every country serves to address the issue. On a more local note, the impact of climate change has had some impact. Flash floods resulting from extremely heavy thunderstorms are becoming more frequent. So too are the occurrence of tornado-like activities.

Intense convective rain cells are often responsible for extreme hydrometeorological events with serious and relevant consequences from a social and economic standpoint. Therefore, the analysis of the spatio-temporal properties of these structures is relevant both theoretically and operationally. It is widely acknowledged that storms of convective origins are responsible for the majority of flash flood events, which is one of the most common and destructive weather-related phenomena in the country. In Selangor, these rain events occur mainly during the inter monsoon period as discussed in Mohd Nor and Zalina (1999).

A deeper understanding of the properties and the dynamics of convective rain cells is, therefore, necessary from a physical and operational point of view. Studies on the origin and physics of convective storms have been reported worldwide (Llasat, 2001; Dong and Hyung 2000; Doswell *et. al*., 1996; Pascual, Callado and Berenguer, 2004). Rapid urbanization, which modified the hydrological processes of a catchment is responsible for many water related problems in urban areas, especially in the tropical regions. Urban drainage systems, often cannot cope with intense convective rainfall events. It is also difficult to forecasts convective rain in terms of timing and spatial distribution as it develops over a short period and can happen any time day or night.

In the management of urban and rural water systems, important hydrological processes such as runoff, infiltration and erosion are usually determined using watershed simulation models that require rainfall data as input. Analysis of pollutant migration through water flow system also require rainfall data as input. However, existing historical records of rainfall are often insufficient in length or in adequate in their completeness and spatial coverage to provide a reliable simulation results. Hence, simulations of rainfall data have been widely used through rainfall modeling. These models were used to generate many sequences of synthetic rainfall series that could describe accurately the physical and statistical properties of the observed rainfall process at a given location.

In many situations, stochastic approach is always preferred in the rainfall modeling as compared to the physically-based model due to the complexity in describing the dynamical and randomness properties of the rainfall. Stochastic rainfall modeling involves using the historical rainfall data to estimate the model parameters of an appropriate model, which may then be used to simulate the desired length of rainfall series. The models are also appropriate for the analysis of data collected on a short time scale, e.g. hourly and the synthetic rainfall series produced are said to resemble the observations statistically. This is particularly useful when the observed rainfall data is inadequate in terms of length and completeness for hydrological applications. These models are also known to have the potentials to estimate the frequency of occurrence of critical events generated by rainfall such as flood.

In view of convective rains which are shorter in duration and higher in intensity, the modeling is based on hourly rainfall series. There are two approaches commonly used in describing the rainfall process. The first approach combines both the rainfall occurrence and rainfall amount and parameter estimation is performed from the hourly and the integrated rainfall data. In this approach certain physical processes of rainfall structure, for example, rain cells, storm and cell clusters are described with a stochastic approach (Kavvas and Delleur, 1981; Waymire and Gupta, 1981a). The second described the rainfall occurrence and the rainfall amount separately and then both are superimposed to form the overall rainfall model (Woolhiser et.al,1982, Roldan et.al, 1982). A poisson-cluster process, namely the Neyman-Scott Rectangular Pulse model is used for the first approach and the Markov-Chain Mixed Exponential model is used in the latter.

Forecasting of convective initiation poses a challenge as orographic and diurnal cycles which triggers a convective activity need to be correctly identified and assessed. Rainfall forecasts can help to determine the magnitudes and patterns of the rainfall expected. It helps prevent hazards caused by flash flood such as damages on building structures and casualties. Forecasting rainfalls also allow an efficient real-time control (RTC) of combined sewer systems (CSS), by proper operation of gates and pumping stations. These control actions enable tanks and channels of the sewer system to be kept at low levels, in order to allow the storage of water volumes of the approaching storm, and to limit the overflow, thus reducing damage, costs and pollution. It is also the means to prepare for drought where the water can be stored if there is no rain for a very long term.

Forecasting rain is one of the most difficult tasks in weather prediction due to the scarce knowledge on how to characterize the mechanisms taking part in its formation. Many different techniques have been proposed to forecast rainfalls. Among these, a physically based approach which makes use of meteorological models might be appealing. However there is some limitations using this approach such as the hydro meteorological variables are not available. In these cases, rainfall forecasting based on stochastic models represent a useful tool where one may be able to forecast rainfall based on current and past rainfall measurements even though such forecasts may not be as accurate as those based on meteorological considerations. In the literature, several attempts to forecast rainfall based on mathematical models can be found such as using the Box-Jenkins models, the neural network models and the numerical weather prediction (NWP) models.

Hydrological data such as rainfall and humidity are often collected in roughly equally spaced time intervals such as, hour, week, month, or year. Such time series data may be available on several related variables of interest. In other words, more than one series is involved in such a model. For example, the rainfall data, where the series is the current and past rainfall occurrences observed at several points in the basin, including the point itself. The operational use of multivariate autoregressive integrated moving average (MARIMA) model or also known as multiple time series ARIMA was suggested by Montanari et al. (1994), who highlighted how a multivariate scheme could remarkably improve the forecasts. In view of the limitations regarding the physics of convective rain initiations, the study undertook the stochastic approach of forecasting using MARIMA.

1.2 Objective of the Study

The objectives of the research are:

(i) To define and identify convective rain based on predetermined variables

(ii) To build/identify a model for convective processes for predicting flash floods.

1.3 Scope of Study

The study encompasses a detailed investigation into the spatio-temporal behavior of convective rain. Distinguishing convective from non-convective events, tracking the movement of storms using radar data and building intensity duration frequency curves based on convective rains are included in the research.

For the modeling and generation of synthetic hourly data, two approaches were investigated namely the stochastic Poisson-cluster process and the Markov-chain process. Hourly rainfall data were the main input for building the models and seasonal effect was taken into account.

The third and last part of the study involves forecasting of hourly rainfall based on a multivariate autoregressive integrated moving average model. Physically based model were not considered due to the limitations in hydro meteorological data and difficulty in assessing the physics of such events.

CHAPTER 2

LITERATURE REVIEW

2.1 Identification of Convective Rainfall

Forecasting of convective initiation is one of the main current challenges in operational nowcasting tasks today. Knowledge of areas where convection develops most frequently is very important. It has been widely known that storms of convective origins are responsible for the majority of flash flood events that causing significant loss of life, property damage, soil erosion and other socio-economic problems. Unfortunately, forecasting skill for heavy convective rain still lacking at present. The characteristic of convective rain such as intensity, rainfall duration, spatial distribution and storm movements are still not enough. No specific guideline is giving a better understanding of this rain which is plays an important part of flooding area.

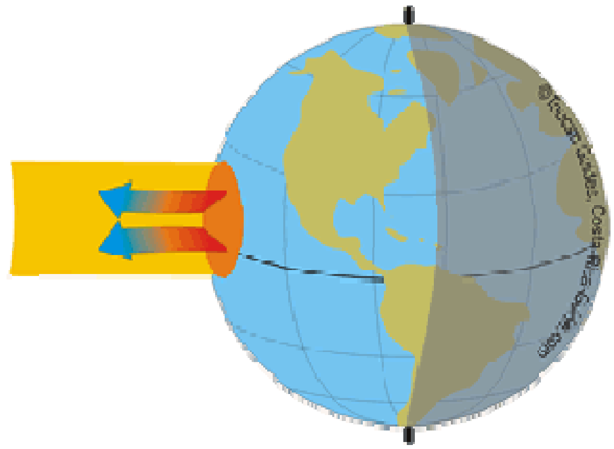
In Malaysia, these events have contributed to substantial damages and losses especially in areas that are prone to flash flood such as Klang Valley. This problem has not been eased even though million of ringgit has been spent or allocated to overcome the drainage problem. Therefore, in this study an effort is made to examine the characteristics of convective rain from Klang Valley's surface rainfall data and radar data.

Although convective rain has long been recognized as important in the area, its contribution has to our knowledge, never been quantified properly. The nearly reason is that convective rain not usually recorded and therefore it is hardly identifiable in meteorological records. Nevertheless, in addition to its meteorological interest, for some applications such as in civil engineering, microwave radiolinks (Burgueno *et al*., 1987, 1988; Vilar *et al.,* 1988), design management of drainage systems and water resources management (Cheng-Lung Chen, 1983; Vazquez *et al.,* 1987; Nix, 1994), it is needed to know the type of rain.

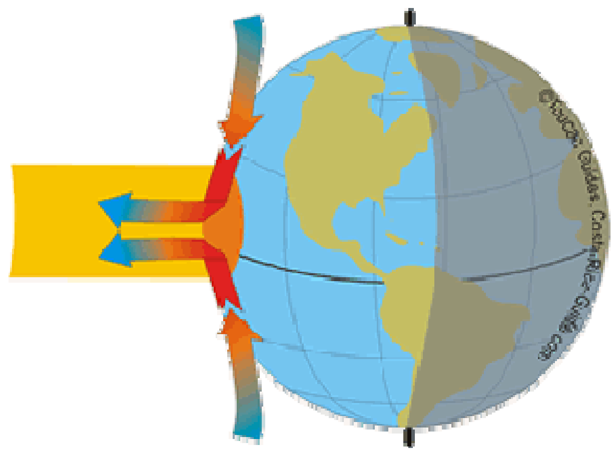
In this chapter, the types of rain, the measurement of rainfall and the previous research on convective rain were described. The probability of the occurrence of flash flood due to convective storm also discussed. Then the methods of spatial interpolation and comparing spatial distribution between all of that method were presented at the end of this chapter.

2.1.1 Convective Rainfall

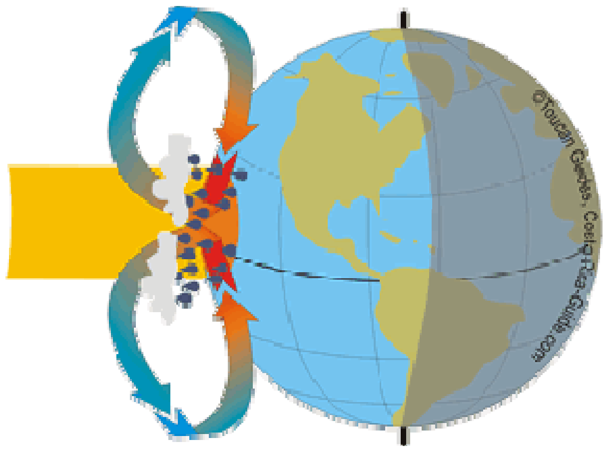
Unlike stratiform precipitation, which is formed in a stable atmosphere, convective precipitation is formed in an unstable atmosphere. Convective rain is a sudden short outburst of rain that brings heavy rainfall in a short period of time. Usually, this short outburst of rain is heavier than normal rainfall. This precipitation occurs from convective clouds e.g., cumulonimbus or cumulus congestus. It falls as showers or a sudden downpour, with rapidly changing intensity. Beside that, the downpour is within one area at a time, as a convective cloud has limited horizontal extent (WikiAnswers, 2007). Convective precipitation usually occurs in the tropics especially in midlatitudes. This phenomenon is due to convection process. Convection is the vertical transport of heat and moisture in the atmosphere, especially by updrafts and downdrafts in an unstable atmosphere. The atmosphere is classified as unstable when the temperature of displaced surface air is warmer than that of the environment surrounding it. This difference in temperature causes the displaced air to rise up into the atmosphere until it gets to a point where it is colder than its surrounding air. At this time, the air begins to fall back towards its original location. This happens because warm air is less dense than cold air at equal pressure (PennState, 2001). Figure 2.1 clearly shows the process of convective rainfall.



(a) warm air rises



(b) Air from surrounding regions move in to replace the warm air as it moves up. The air that moves in to take the place of the rising air has to come from the north or the south because the air to east and west is also extra hot and rising.



(b) As the warm air rises it expands and cools. Since cool air cannot hold as much moisture, this often results in rainfall. The cooled air is then drawn back towards the poles, dropping towards the earth to replace the air moving along the surface near the equator. This cycle of air movement is called convention and causes convective rainfall.

Figure 2.1 : The formation of convective rainfall (After Charles L. Hogue, 2007)

2.1.2 Identifying Convective Rainfall

2.1.2.1 Rainfall Intensity

The intensity of rainfall is dependent on the rate at which storm processes water vapor. In this case, a distinction could be made essentially, between precipitation of convective origin and precipitation stratiform origin. Many researchers used intensity as a method to differentiate among convective and stratiform rainfall. Dutton and Dougherty, 1979; Watson et al., 1982 sets a convective rainfall rate threshold at 50 mm/hr and below as supposedly non-convective. Llasat and Puigcerver (1997) divided their analysis into four kinds of event: (1) non-convective (2) convective with rain rate equal or less than 0.8 mm/min (3) convective in which the rate threshold of 0.8 mm/min was exceeded; and (4) rainfall from thunderstorms. Llasat studied convective rain for a number of years. In 2001, she used 35 mm/hr as a threshold intensity value and was utilized parameter β for characterizing convective rain (Llasat 2001).

Nevertheless, Houze (1993) distinguishes between stratiform or convective precipitation on the basis of vertical air velocity, *w*. If it is less than the terminal fall velocity of ice crystals and snow, it is called as stratiform. But nowadays, radar can also be used to make a distinction between both of these rainfalls. Using 4-D radar imagery, the 'bright band' near the melting level is a signature that helps to distinguish convective mode from stratiform mode (Llasat, 2001). Steiner *et al*. (1995) proposed two methods to distinguish between stratiform and convective precipitation in radar echo patterns. Radar used reflectivity to measure the intensity of rain and usually the reflectivity is expressed in decibels of z (dBZ). Dong and Hyung (2000) used 35 dBZ to determine convective rainfall. Pascual, Callado and Berenguer (2004) used four reflectivity thresholds: 30 dBZ, 35 dBZ, 40 dBZ and 45 dBZ during identify convective cells origin. On the other hand, Rigo and Llasat (2002) used 43 dBZ to analyse convective event which is derived from meteorological radar.

2.1.2.2 Rainfall Duration

As previously mentioned, convective rain is a sudden short outburst of rain that brings heavy rainfall in a short period of time. These are usually inversely related, because high intensity storms are likely to be of short duration and low intensity storms can have a long duration. Brooks *et al.*, (1992) noting that convective cell typically has a lifetime of about 20 min. It follows, then, that any convective storm lasting more than about 20 min is made up of more than one cell. A convection cell is a phenomenon of fluid dynamics which occurs in situations where there are temperature differences within a body of liquid or gas (Wikipedia, 2007).

Ronal and Andrew (1981) studied about duration of convective events related to visible cloud, convergence, radar and rain gage parameters in South Florida. The highly variable response could be understood better by taking into account the duration of the cloud where it is defined as the time from first surface convergence until it is complete dissipation. From their observation, the average of storm duration for nine clouds was

25 min from first convergence to first organization of the cloud area. Another 35 min passed, on the average, until the clouds began a rapid upward growth stage.

2.1.2.3 Analyses on Convective Rain

The poor quality of heavy rain forecast might seem surprising in view of the great improvements over recent years in general weather forecasts. Predicting where such storms will break out or start abruptly is one of the major challenges facing meteorologists today. Furthermore, convective storms always cause downpours and flash flood. This situation motivates many researchers to study about convective rain.

Llasat and Puigcerver, (1997) studied convective rainfall with an objective to obtain the percentage of convective rainfall from the total rainfall amount in Catalonia. Convective events were identified on charts of a rain-rate recorder from 1960-1979. Events were classified into four categories: non-convective, convective with low rainfall rates, convective with moderate to high rates and thunderstorm events. From the result, the ratio of convective to total rainfall amounts ranges from 70 to 80 percent in summer months to less than 30 percent in winter. Next, in year 2001, Llasat characterized convective rain in new event classes and apply it in modelling intensity-duration-frequency (IDF) curves and design hyetographs (Llasat, 2001). A parameter related to the greater or lesser convective character of the precipitation, designated as β is defined. Intensity value of 35 mm/hr is taken as threshold intensity and β parameter was classified into four categories; non-convective, slightly convective, moderately convective and strongly convective. Llasat and Rigo (2002) used radar in their analysis. They studied convective structures with made a comparison between meteorological radar data and surface rainfall data. In year 2007, Llasat and Barnolas studied flood geodatabase and its application in meteorology of climates. In their study, convective rain was divided in three types; (1) very convective rainfall events: episodes of very short duration (less than 6 h) but very high rainfall intensity, (2) very convective and moderate rainfall events: episodes of short duration (between 6 and 72 h) with heavy rain sustained for several hours, and (3) episodes of long duration (approximately 1 week) with weak raingauge intensity values. Geographical Information System (GIS) is used to display all of the information in geodatabase. From the analysis, fall season floods are mainly identified with convective episodes with heavy rain sustained for several hours. The inland region is mainly affected by episodes of types 2 and 3. While episodes of type 1 mainly affected in regions with a high population density.

Nowadays, there are many researchers used meteorological radar to detect convective area and done various analyses. By using radar, two algorithms have been applied to analyze convective structures. First, Johnson *et al.,* (1998) identified convective cells as a region of maximum reflectivity in 3D. Second algorithms were proposed by Steiner *et al*., (1995), where they identify convective structures at the lowest level 2D. These algorithms classify pixels from radar image as rainfall or non-rainfall. Then they choose which rainfalls satisfy certain requirements to consider them as 'convective'and 'stratiform'. Both of these algorithms also have been applied by Rigo and Llasat (2002) where they used radar data and surface data to improve the tracking and nowcasting of convective structures in Catalonia, Spain. For surface rainfall data, they used 35mm/hr as a rain rate threshold of convective events whilst 43dBZ as a reflectivity threshold to do a first identification of convective rainfall. The β parameter (Llasat, 2001) is used to identify the degree of convection of every rainfall event for raingauge data. The comparison of the daily β parameter for raingauges and radar charts allows identifying the areas most prone to convective precipitation, especially for different seasons.

Another study of convective rain using meteorological radar is Pascual *et al.,* (2002) and Callado *et al.,* (2002). They analyzed the origin of convection identified in radar data with low levels convergence zones. After that, Pascual *et al.,* (2004) studied about convective activity during the summer of 2003 and relate it with convergence areas associated to terrain characteristics and to the interaction between different flows at low levels. The 15 C-band Doppler radars are used in this study. The results were presented in term of relative frequency maps. From the observation, higher relative frequencies for all thresholds (30 dBZ, 35 dBZ, 40 dBZ and 45 dBZ) appear in mountainous terrain and most of the frequencies happen between 12:00 and 18:00 hours.

As mentioned before, convective activities are more frequent in the Tropics. The diurnal cycles of convective activity are different and it is depend on the location and weather. If the location is near to the sea, the convective activity may due to wind and water vapour from the sea. The duration also can be different with other location. Hara, Yoshikane and Kimura, (2006) conducted a cloud-resolved simulation using regional climate model to clarify the mechanism of diurnal cycle of convective activity around Borneo Island. The convective activities on top of mountain decay in evening. The diurnal cycle of convective activity in Borneo Island is maintained by sea breeze and upslope wind and is dependent on the distance from the coast to the centre of the mountain. The convective activities continue until the next morning.

Dong and Hyung (2000) studied heavy rainfall with Mesoscale Convective Systems (MCSs) over the Korean Peninsular. A Mesoscale Convective Systems (MCSs) is a complex group of thunderstorms which becomes organized on a scale larger than the individual thunderstorms, and normally persists for several hours or more. It can be round or linear in shape, and include systems such as tropical cyclones, and squall lines (Wikipedia, 2007). The study focused on mesoscale convective systems (MCSs) which were most responsible for flash floods over the central Korean Peninsular for 6 hours. The evolution and movement of convective storms resulting in heavy rainfall were investigated. They used WSR-88D radar data to conduct the study. From their observation, the heavy rainfall was caused from well-organized multi-cell type convective storms in MCSs. The storm abruptly started near the sea and land, and then merged into large convective storm within less than 2 hours. To investigate movement of the convective storms, they tracked the edges of convective storms. It is found that the boundaries changed into a very complex shape with time and the storm movement was very limited.

2.1.2.4 Probability of Flash Flood due to Convective Storm

Convective storms are always related with the flash floods. It is produced by strong convection in a short time. Charles (1993) considered a precipitation rate of about 25 mm/hr as quite heavy, and flash floods often result from rainfall intensities much greater than that value (25 mm/hr). For this time, it is difficult to sort this rainfall rate from non-convective processes. This is because they simply don't process water mass fast enough. Charles also identified the precipitation efficiency which is indicated from water vapour. Precipitation efficiency is defined as the ratio of the water vapour absorbed into the storm to the water dropped as rainfall. This ratio is not meaningfully evaluated in an instantaneous value. At the start of convective storm, no rain is falling, so the ratio is zero, but at the end of the storm, rainfall can continue to fall after the updraft has dissipated. Figure 2.8 shows a schematic diagram of precipitation efficiency. Therefore, this quantity only makes sense as a time essential over the lifetime of convective system (Fankhauser, 1988). Simple basic consideration suggest that of the water vapour passes through a convective storm, what doesn't fall out as precipitation must evaporate.

Barnolas and Llasat, (2007) studied a flood geodatabase in Catalonia. They classified flash flood into three types based on the convective character of rainfall event. Type 1: Very convective rainfall events: episodes of very short duration (less than 6 h) but very high rainfall intensity. They produce flash flood and local damage. Their associated floods are usually ordinary or extraordinary, following the classification shown in Llasat *et al*., (2005). Type 2: Very convective and moderate rainfall events: episodes of short duration (between 6 and 72 h) with heavy rain sustained for several hours (200-500 mm). In the light of their duration and size of catchments, they can produce catastrophic flash floods. Type 3: Episodes of long duration (approximately 1 week) with weak raingauge intensity values, with possible peaks of high intensity. Accumulated rainfall can be over 200 mm and usually ordinary or extraordinary floods occur. From their study, episodes of type 1 mostly occurred in the area which has a high population density. While episodes of type 2 and 3 occurred in the inland region. It seems that rainfall duration, amount of precipitation and areas of rainfall are main factors in identifying flash flood into several classes.

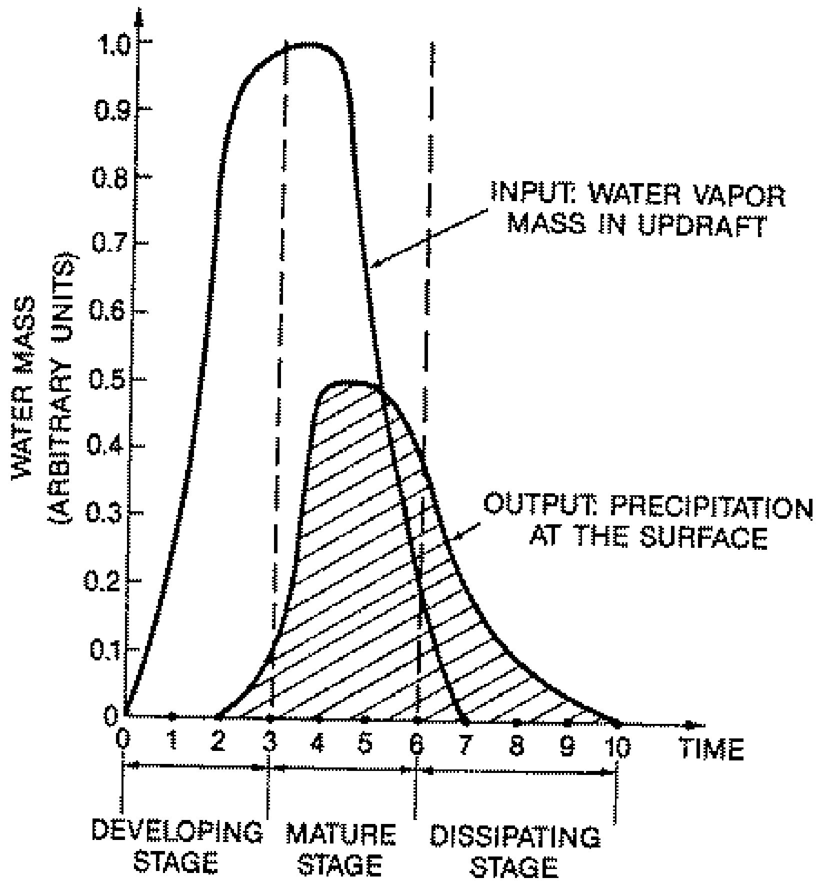


Figure 2.2 : Schematic diagram showing the time history (in arbitrary time units) of water vapor input and precipitation output (hatching) for a convective storm system. The ratio of the areas under the two curves is the precipitation efficiency (after Charles, 1993)

The heavy rainfalls that produce flash floods are the result of high rainfall rates that remain. The high rainfall rates are caused by high water vapour mass flow through convection, coupled with high precipitation efficiency. The previous study also show that convective events and the occurrence of flash flood in a particular area always related to each other. All of these findings are very important to give much more information about convective especially in the areas most prone to convective rainfall.

2.1.4 **Spatial Interpolation**

A very basic problem in spatial analysis is interpolating a spatially continuous variable from point samples. In hydrology, rainfall is always measured only at raingauges. Nevertheless, engineers are interested to estimate the total rainfall in a watershed. Nowadays, the question is how to calculate the individual rain measurements to obtain the best estimate of rainfall at an unmeasured location. Figure 2.3 shows the basic interpolation process in some area.

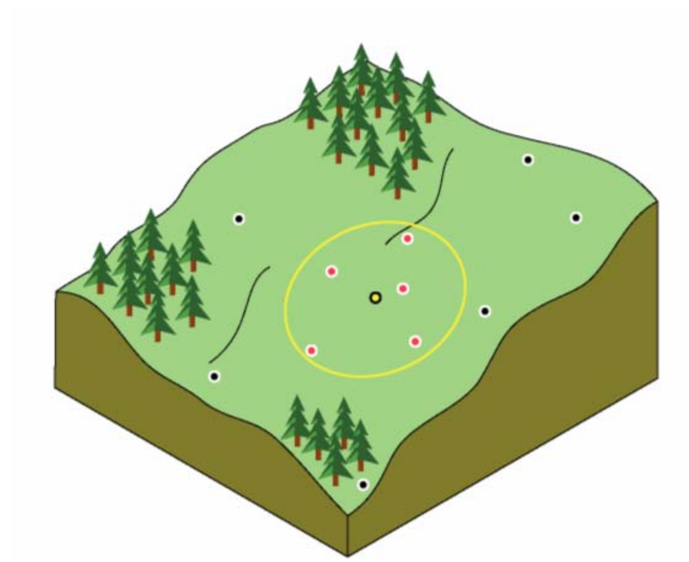


Figure 2.3 : The interpolated value at the unmeasured yellow point is a function of the neighbouring red points (From ArcGIS Help Menu)

Three interpolation techniques, namely the Inverse Distance Weighted (IDW), Kriging and Spline Method are the most commonly used techniques to estimate grid point values from scattered data (Keckler, 1995). In this section, all interpolation techniques will be discussed and a comparison of the spatial interpolation between some of these methods and also from the previous study will also be presented.

2.1.3.1 Kriging Method

Interpolation by Kriging is a geostatistical method based on statistical models that predict spatial correlation of sampled data points (Dille *et. al.,* 2002). Kriging was developed in 1960s by the French mathematician Georges Matheron. Originally, it is proposed by Krige, a South African mining geologist, who is the first to introduce the use of moving averages to avoid overestimation of reserves. The method has been used by gold mining engineers in South Africa and it is used to estimate gold in a rock from a few random core samples. Since this method is widely used in geology, Kriging has become similar with the variety of geological statistics (Matheron, 1963). Today, Kriging has found its way in the earth science and other disciplines. In spatial interpolation, it is an improvement from inverse distance weighting because prediction estimates tend to be less bias and predictions are accompanied by prediction standard errors (quantification of the uncertainty in the predicted values) (Jon and David, 2002).

The objective of Kriging is to estimate values of a field (or linear functions of the field) at a point (or points) from a limited set of observed values (Bras and Rodriguez-Iturbe, 1985). Spatial correlation, is a statistical relationship among measured points in one data set. Kriging also can provide some measure of certainty or accuracy of the prediction models based on correlation. Kriging models use semivariogram or covariance to depict the spatial correlation between measured sample points and to make optimum predictions. Semivariogram modeling is the element that must to separate the spatial modeling from simple spatial description. The model assumes that measurements that are geographically close together are more similar than ones that are farther apart (Donald, 1994). Figure 2.4 shows the spatial correlation in kriging.

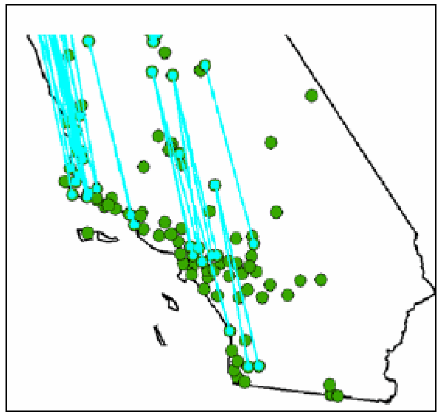


Figure 2.4 : Spatial correlation (a function of distance between pairs of locations)

Semivariograms are described by the parameters of range, sill, and nugget. All of these elements are needed to interpolate data with a Kriging method (Figure 2.5). The range is the distance from a measurement (known sample) point to the point where the semivariance stops increasing with distance from the sample point. Sill is known as the value at which the semivariogram model attains the range. It is mean that the change in semivariance is no longer increasing with increasing distance from the sample point. The nugget is created by measurement errors or spatial sources of variation at distances smaller than the sampling interval. Nugget also recognized as the value of semivariance when the distance from the sample point equals zero (Main *et. al.*, 2004). One more element is partial sill. Partial sill is sill minus the nugget and this value is needed for Kriging interpolation. Figure 2.5 shows one example of semivariogram.

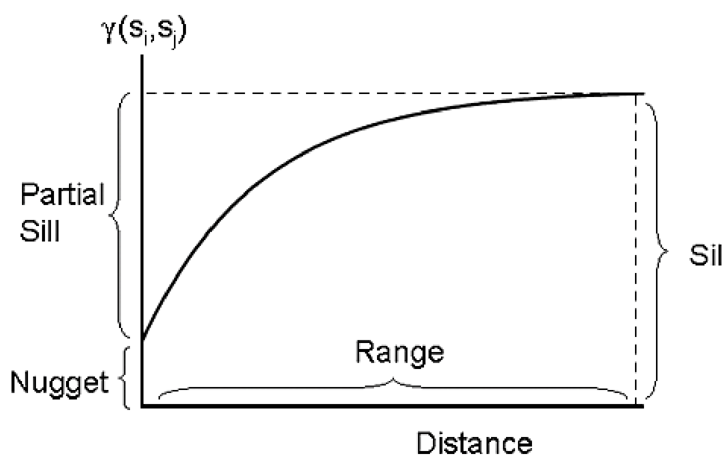


Figure 2.5 : Example semivariogram depicting range, sill, and nugget (after Main *et al.*, 2004).

As already noted, Kriging models use semivariogram or covariance to depict the spatial correlation. Estimation of covariance is similar to the estimation of semivariogram, but covariance requires mean data. However, the data mean usually not known, but estimated and this causes bias. This situation resulted that most geostatistical software use semivariogram as default function tool to characterize spatial data structure (Konstantin, 2006). The equation of semivariogram and covariance can be described as:

Empirical Semivariogram (Equation 2.3)

Semivariogram (distance h) = ½ average [(value at location i – value at location j)^2]

Empirical Covariance (Equation 2.4)

Covariance (distance h) = average [(value at location i – mean)*(value at location j – mean)]

where, for all pairs of locations i and j separated by distance h

Kriging is considered the best predictor of non-sampled locations, because mean residual error is minimized by its calculation (Isaaks and Srivastava, 1989). Actually, Kriging interpolation is similar to IDW where it uses surrounding data points to predict an unknown value for an unmeasured location. The difference with Kriging can be mentioned in three ways:

(a) the predicted point depends on a fitted model to the measured points;
(b) the distance from the unknown point to measured points; and
(c) the spatial relationship among the measured points around the predicted point.

In this study, Kriging Method is chosen to show the spatial distribution of rainfall derived from surface rainfall data.

2.2 Model Building Using Stochastic Rainfall Modelling : Neyman-Scott Rectangular Pulse Model (NSRP) and Markov Chain Mixed Exponential Model (MCME).

The modeling of rainfall has been progressing significantly in the recent decades. It has a long history in literature with significant advances being made over years in the statistical methods and techniques used and the subsequent accuracies achieved. Reviews of previous works on rainfall modeling have been discussed exhaustively in Waymire and Gupta (1981), Foufoula-Georgiu and Krajewski (1995) and Onof C. *et al.,* (2000). Two models based on stochastic rainfalls modeling are adopted in this study. The first is on the cluster point process and the second is on empirically derived models with "fitted" parameters. The developments of both models are presented as follows.

2.2.2 Neyman-Scott Rectangular Pulse (NSRP) Model

Two of the most recognized cluster-based models used in stochastic modeling of rainfall are the Neyman-Scott Rectangular Pulses (NSRP) model and the Bartlett-Lewis Rectangular Pulses (BLRP) model (Rodriguez-Iturbe *et.al*. (1987a). These models represent rainfall sequences in time and rainfall fields in space. Both the occurrence and the depth processes are combined and parameter estimation is performed from the hourly and the integrated rainfall data. To understand properly the models, we begin with the reviews of the theoretical basis of stochastic point processes. These reviews will focus on the study of special processes that have importance application related to rainfall modeling . The developments of techniques for analyzing data generated from such processes can be found for example in Waymire and Gupta (1981c) or Rodriguez-Iturbe *et.al*. (1987a). In stochastic processes, the realizations consist of point events in time or space. This is used in many fields of applications and is discussed exhaustively in literature from several points of views (Cox and Isham, 1980; Waymire and Gupta, 1981a).

2.2.1.1 Development of the Neyman-Scott Rectangular Pulse Model (NSRP)

The Neyman-Scott cluster point process, originally developed in 1958 to describe the distribution of galaxies in space (Neyman and Scott,1958) has become an important representation for a broad range of phenomena in the physical, biological, and social sciences. Vere- Jones (1970) applied Neyman-Scott (N-S) cluster process in the time dimension to model the earthquake occurrences in which he utilized the probability generating functional (pgfl) of N-S process in modeling the occurrences but did not give a derivation of this functional. Lawrence (1972) modeled the earthquake occurrences by deriving the probability generating functional of the Neyman-Scott (N-S) process by counting the cluster centers which represented the main shocks and the aftershocks form the secondary process. In the hydrologic literature Kavvas and Delleur(1975,1981), Kavvas (1982a,b), Gupta and Waymire (1979), and Waymire and Gupta (1981a,b,c) have popularized the use of cluster models.

There are number of variations of the N-S model for representing rainfall events. All variations are essentially the same in the way they model the occurrence of the rainfall events, i.e the occurrence of the rain cells. The variety of N-S models results from how the depth of rain associated with each rain cell is distributed over a time interval. The simplest N-S model, known as the N-S white noise model takes the rain cells as instantaneous bursts and associates some distribution with the depth of rain due to the cell. This model was first introduced for representing rainfall events by Kavvas and Delleur (1975), by deriving expressions for the counting properties of rain cell occurrence. Waymire and Gupta (1981 a,b,c) demonstrated how the probability generating functional taken from the general theory of point processes can be an effective tool for capturing the joint distributional properties of the counting process of rainfall occurrences. Smith and Karr (1985) showed that a N-S process, in which the distribution of cluster sizes is Poisson and the distribution of the distance of the cluster members from the cluster center is exponential, could be represented as a Cox process. Using this representation they derived the maximum likelihood estimates for parameters of the N-S model. Rodriguez-Iturbe et al. (1984) found the second-order moments of the aggregated N-S white noise model. These properties are particularly desirable as rainfall records tend to be available in aggregated form, usually as daily totals. N-S white noise model perform better than other rainfall model over range of time scales (Cowpertwait, 1991). Valdes et al. (1985) also found that the N-S white noise model performs better over time scales from 1 to 24 hours, when compared with some other rainfall models. However, they found that the N-S white noise model appears not to be able to preserve the statistics of extreme rainfall events. Foufoula-Georgiou and Guttorp (1987) also found inadequacies in the N-S white noise model in particular, difficulties were found when estimating the model's parameters, when using both the method of moments and maximum likelihood estimation.

Motivated by the inadequacies of the N-S white noise model, Rodriguez-Iturbe *et al*. (1987a) introduced the N-S and Bartlett-Lewis Rectangular Pulse models for representing rainfall. These models give each rain cell a random duration, and a random intensity which is constant throughout the cell duration. In their paper the second-order moments of the aggregated process for the N-S model are found, under the assumption that the duration and the waiting time for the rain cells after the beginning of the storm are exponentially distributed. An analysis of empirical data using the N-S and Bartlett-L ewis rectangular pulse models has been carried out by Rodriguez-Iturbe *et al*.(1987b), the conclusion being that the rectangular pulse models are able to preserve rainfall statistics, including extreme values, over time scales from 1 hour upward, with the exception of the proportion of dry days.

To solve the problem of overestimation of the probability of observed dry periods, Rodrigue-Iturbe *et al*.(1988) suggested the use of a modified Bartlett-Lewis model with an additional parameter, and Entekhabi *et al.* (1989) introduced a similar modification in the Neyman-Scott model. In the modified NSRP, the rain cell duration η was random and was allowed to change from storm to storm. The probability density function for η is assumed to be a two-gamma distribution with shape parameter α. Burlando and Rosso (1991) questioned the ability of the modified Bartlett-Lewis models to reproduce the historical characteristics of the rainfall series, stating that the original N-S model fits better than the original and the modified Bartlett-Lewis models. They (1996) also pointed out some features limiting the use of stochastic point processes in modeling storms, such as the inability to reproduce variability displayed in the extreme storms, nontrivial mathematical complexities are involved in the construction and implementation of the models, and the presence of subjectivity in the parameter estimation.

Velghe *et al.* (1994) argued that even though the modified models gave a better zero depth probability, owing to the higher complexity of the parameter estimation they did not preserve the second order properties (especially lag-2 and lag-3 auto-correlations) of the rainfall process. The authors also found that the Barlett-Lewis model, especially in the modified version, is very sensitive to the choice of moments used in the parameter estimation. The original and modified versions of the geometric Neyman-Scott model were found to be amenable to practical use in hydrological studies than the Poisson Neyman-Scott model. However, the findings were not conclusive and open to many more research to be undertaken.

Cowpertwait (1994) further developed the model at a single site by allowing each generated cell to be of *n* types. The model developed is called generalized Neyman-Scott Rectangular Pulses [GNSRP(n)]model. The case for two cell types was considered, categorized as either “heavy” or “light” where heavy cells have shorter expected lifetime than the light cells, which agreed with the observational studies on precipitation fields. Cowpertwait (1997) fitted the model and harmonic parameter estimates were regressed on sites variables. The residual errors analysis showed that the regression equations could be used with reasonable confidence for urban sites.

Stochastic spatial-temporal models of rainfall have been formulated based on physical processes observed in precipitation fields which usually incorporate bands of rain, regions of high intensity rain, and rain cells. These fields can be used to simulate fine resolution data over a large geographical region, and are potentially useful in real-time forecasting. Cox and Isham (1988) developed a simple model where storm centers arrived in a two-dimensional space and time Poisson process but empirical analysis of data has shown that rainfall events tend to arrive in clusters. Cowpertwait (1995) then formulated spatial-temporal model where the arrival times of rain cells follow a Neyman-Scott process. The cells are randomly classified from 1 to *n* with different parameters for different cell types, so that the random variables of an arbitrary cell, e.g. radius and intensity, are correlated. The model has a flexible structure, via the generalization, so that a reasonable fit to multi-site extreme values could probably be achieved. However, the range of applications to which the model could be applied may be limited because rain cells are taken to have zero velocity. An extension of the generalized spatial-temporal model was done by introducing the third moment function for the single site model by Cowpertwait (1998). A good fit to the observed extreme values over a range of time scales was found. Lack of fit was evident when the third moment was excluded from the fitting procedure. However, the cell parameter estimates had large standard errors and were related, partly due to the difficulty in identifying cells in physical process. Statistical properties for the spatial-temporal model (Cowpertwait, 1995,1998) were combined into the fitting procedure, which used moments up to third order and cross correlation function (Cowpertwait, 2002), but for a single type of cell. The results indicate that the model is able to preserve regional extremes and support the use of the model in hydrological applications.

The GNSRP(n) that was developed by Cowpertwait (1994) did not include the third moment function. Hence, Cowpertwait (2004) developed a mixed model by using superposed independent NSRP processes to make use of the existing NSRP functions that have been derived and cited in Cowpertwait (2002). The use of superposed processes makes an allowance for different possible storm types, e.g. those with predominantly convective cells or stratiform cells. This model gives further flexibility in the parameterizations thus providing a methodology for obtaining good fits to a wider range of data.

Kim et al.,(2006) developed a new stochastic point rainfall model which considers the correlation structure between rain cell intensity and duration. The model is able to reproduce well the statistical characteristics of the historical rainfall series and the model generated data are robust with different parameter sets when the correlation parameter is appropriately taken.

Previous studies assumed that rain cell intensity follows an exponential distribution due to its small number of parameter (e.g. Rodriguez-Iturbe et al., 1987a, 1988; Cowpertwait, 1996, etc). However, the choice of distribution for the cell intensity in the NSRP model is arbitrary. Cowpertwait (1998) had used gamma to represent the rain cell distribution because past studies have reported lack-of-fit to extreme values under the exponential distribution. Cowpertwait, (1996, 2002, 2004) had also attempted a heavier-tailed distribution such as Weibull to improve the fit in the extremes. Hence, there are still many other distributions that are open to be explored.

2.2.2.2 Parameter Estimation

The fitting of the parameters of the model and the assessment of the adequacy of its fit raise many statistical questions. Calenda and Napolitano (1999) described exhaustively the different methods for parameters estimation of the NSRPM model. The usual procedure as described by them is based on the method of moments (Rodriguez-Iturbe *et al.,* 1987a,b; Entekhabi *et al*., 1989; Cowpertwait, 1991). The maximum likelihood estimates, besides involving heavy mathematical complexity (Smith and Karr, 1985 a,b) are not available and are not computable because the distribution function of the rainfall average intensity in each disjoint time interval for a given scale of aggregation is not known. Even if a likelihood function could be calculated, it would not be a proper basis for fitting the model because the idealization involved leads to sample path with some (short-term) deterministic features (Favre *et al.*, 2004). Kirk (1997) fits the model using importance sampling in order to obtain a product-of-spacings function, but the estimator obtained is biased.

The original NSRP model depends on five parameters, $\lambda, \beta, \eta, \nu, \xi$, so that following the method of moments, five statistical properties of the observed time-series must be equated to their theoretical expression, and the resulting equations solved for the parameters estimates. The most frequently used procedure that used the method of moments, adopted first by Rodriguez-Iturbe *et al.* (1987,a,b) and then by many others (Burlando, 1989; Entekhabi *et al.,*1989; Islam *et al.*, 1990; Cowpertwait, 1991; Velghe *et al.*, 1994). The historical series is aggregated at two different temporal scales using the expressions of the mean at the first level of aggregation, the variances and the lag-1 autocorrelation at both levels with the mean being a linear function of the scale. The equation system obtained is not linear, and it is solved by minimization of

$$Z(x) = \sum_{k,h} \left[1 - \frac{\Theta_k(x,h)}{\Theta_k^*(h)} \right]^2 \qquad (2.1)$$

The use of the ratio function Z(x) ensures that large numerical values do not dominate the fitting procedure. Cowpertwait *et al.* (1996 a,b) suggested the use of a larger set of sample moments. They used mean at 1 hour scale, variance at 1, 6 and 24 hour aggregation, autocorrelations at 1, 6 and 24 hour and, probability of dry time intervals, assigning weights to the different statistics. The used of autocorrelations were found to affect the match on the proportion of dry days because autocorrelations tend to have large sampling errors because of the large number of zero depths. Thus, the autocorrelations at all aggregations were excluded and the transition probabilities were used instead while the other moments remained the same. They also used sample moments and transition probabilities at 3 and 12 hour aggregations besides 1, 6 and 24 that were applied earlier. The results on the proportion of dry periods improved. The couple of scales generally considered in the estimation procedure are combinations of 1, 3, 6, 12 and 24h aggregation scales (Rodriguez-Iturbe *et al.*, 1987 a,b; Entekhabi *et al.,* 1989; Islam et al., 1990; Cowpertwait, 1991, 1992; Velghe *et al*., 1994; Cowpertwait et al., 1996 a,b); but sometimes also scales of 30 min (Burlando, 1991), and 20 min (Sirangelo, 1992). It is generally held that the parameter estimates are not biased by the selection of the aggregation scales of the sample data set; but preliminary results (Calenda and Napolitano,1997) showed a significant variability of the estimates with the scales, that could be ascribed to two different causes: the characteristics of the objective function Z(x) change substantially with the scales and the results of the optimization algorithm vary when the starting point of the search is changed, especially if the selected scales are close together. They then suggested an alternative estimation procedure based on the scale of fluctuation of the observed process. The estimates obtained with the proposed procedure are as good or better than those obtained with the usual procedure for all aggregation scales, with the exception of very long (24h) and very short aggregation times (5 and 10 min), both in term of reproduction of the second order statistics and extreme values for different aggregation scales.

Following Calenda and Napolitano (1999), Favre *et al.* (2002) proposed a modified method of moments using two temporal scales of aggregation, hourly (1h) and daily (24h). The two scales were selected because the estimates of the parameters of the continuous process always depend on the aggregation scales selected for the formulation of the solution system. However, if the scales are more widely spaced the estimation stabilizes.

The choice of the minimization of the objective function of concern, whereby methods like quadratic convergence of Powell have been proposed (Velghe *et al*., 1994; Calenda and Napolitano,1999) to solve the nonlinear optimization problem. The main difficulty relates, however, to the choice of initial parameters values on which the convergence of the algorithm is strongly dependent. The minimization is carried out in a space of five or more dimensions and local minima are difficult to avoid. To avoid these limitations and the related bias an alternative approach is proposed by Favre *et al*., (2004) by reducing the number of parameters to be obtained by minimization. Using the Nelder-Mead simplex the minimization procedure is said to be stable with regard to the starting point of the algorithm and always converges. Nelder and Mead simplex uses direct search complex algorithm that is dependent on the comparison of function values at the (n+1) vertices of a general simplex, followed by the replacement of the vertex with the highest value by another point (Nelder and Mead, 1963). This method is said to be effective and computationally compact.

Duan (1992) developed Shuffle Complex Evolution-University of Arizona (SCE-UA) method that is a general purpose global optimization program. SCE-UA was both effective and efficient, compared with the existing global methods such as adaptive random research (ARS) method and multi-start Simplex method. He also showed that SCE-UA was an effective and efficient optimization technique for calibrating watershed models and these are basically influenced by the choices of algorithmic parameters. Han (2001) used SCE-UA method to optimize the objective function of NSRP and compared that with the Nelder-Mead Simplex method. It was found that the SCE-UA performed better than Nelder-Mead simplex method.

2.2.3 The Markov Chain Mixed Exponential Model (MCME)

A rainfall model based on daily precipitation is attractive because relatively long and reliable records are readily available and such a model is frequently efficient for many practical problems. Stochastic models of daily rainfall are usually divided into two parts, a model of rainfall occurrence which provides a sequence of dry and wet days, and a model of rainfall amounts, which simulates the amount of rainfall occurring on each wet day and then both are superimposed to form the overall rainfall model. (Woolhiser *et.al*,1982, Roldan *et.al,*1982,). One of the popular stochastic modeling of daily rainfall is the Markov Chain-Mixed Exponential (MCME). The first-order two-state Markov Chain model is used to describe the daily rainfall occurrence process and the Mixed Exponential distribution is used to describe the daily amount distribution. Many studies have used the combination of Markov Chain and Mixed Exponential (MCME) to model daily rainfall series and the combined model had proven to be the best in describing rainfall processes (Woolhiser and Pegram. 1979, Woolhiser *et.al,*1982, Han, 2001). Models of this kind are capable of simulating daily rainfall records of any length, based on simulating occurrences and rainfall amounts separately. Parameter estimates are needed for transitional probabilities for occurrences and parameters are fitted through a frequency distribution for rainfall amounts. The research work presented in this thesis on modeling the hourly rainfall series is based on this approach.

2.2.2.1 Modeling of Rainfall Occurrences

The Markov chain model for the daily occurrence of precipitation has achieved widespread use with Gabriel and Neumann (1962) was probably the first mentioned in literature that had described the daily occurrence using a two-state simple Markov Chain. Their work was then adopted by Haan *et al.* (1976) that proposed a stochastic model based on a first-order Markov Chain to simulate daily rainfall series at a point. He was able to justify the capability of the model to simulate a daily rainfall record of any length, based on the estimated transitional probabilities and frequency distributions of rainfall amounts.

According to Chin (1977), the common practice of assuming that the Markov order is always one is unjustified. He used a decision criterion based on a loss function that is composed of a log-likelihood ratio term and a degree-of-freedom term and the order that minimized the loss function is selected. The results showed that the order of conditional dependence of daily precipitation occurrences is dependent upon the season and the geographical locations. Gregory *et al.*(1992) found that the lumping together some of the states of a many-state first-order Markov Chain does not in general give a first-order Markov chain with a smaller number of states. They even suggested that a many-state process, possibly of only first order would actually be a better choice than a two-state process.

Most of the point process models are continuous in time and not directly applicable to discretely sampled data such as the occurrence of rainfall. But Smith (1987) proposed a new family of discrete point process models for daily rainfall occurrences termed as a Markov Bernoulli process that contained Markov chain and Bernoulli trial models. The process in which a discrete time analog of Neyman-Scott models was constructed. Likelihood-based inference procedures for discrete point process models of wet-dry sequences were developed that not only evaluates quantitatively but also qualitatively the significance of the parameter estimates. Foufoula (1987) found an alternative discrete-time point process model termed as Markov renewal model. This model exhibits clustering relative to the independent Bernoulli process.

Another alternative approach is through the use of spell-length models, where observed relative frequencies of dry or wet day spells are fitted to a probability distribution. This process is called the 'alternating renewal process (Buishand, 1977; Roldan and Woolhiser, 1982; Raseko *et al*.,1991) allows for a new spell of opposite type of random length to be generated once a spell of consecutive dry or wet days have ended.

2.2.2.2 Modeling of Rainfall Amounts

Methods of modeling precipitation amounts on wet days have been discussed extensively in the literature. The most common approach is to assume that precipitation amounts on successive days are independent and to fit some theoretical distribution to the precipitation amounts (Todorovic and Woolhiser,1975). A second approach is to assume that precipitation amounts are independent but the distribution function depends on whether the previous day was wet or dry, i.e a chain-dependent process (Katz,1977). Theoretical distributions used include the exponential (Todorovic and Woolhiser, 1975), the Gamma (Katz, 1977, Buishand,1977), and the Weibull (Han,2001). The mixed exponential distribution has been used previously by Foufola-Georgiou and Lettenmaier (1987), Woolhiser and Pegram (1979), and Han (2001).

The statistical distribution of rainfall amounts for different length periods was discussed exhaustively in literature especially in monthly and yearly scales, where good fits using gamma, Gaussian, logarithmic normal and normal distributions were found (Delleur anf Kavvas, 1978; Srikanthan and McMohon, 1982). Distributions on the daily scales or lower, on the other hand has higher variability and that limits the number of applicable distributions (Nguyen and Rouselle, 1981; Woolhiser and Roldan, 1982).

There was generally no single distribution accepted for describing rainfall amount over a wide range of regions and time scales. Richardson (1981) used the one parameter exponential model due to its simplicity, as a first approximation of daily rainfall distribution. However, to improve the fit to the observed the two-parameter gamma was used (Ison *et al.,*1977; Katz,1977; Buishand, 1977). The three-parameter Kappa distribution performed comparably with gamma (Mielke, 1973). A gamma-family distribution such as a two-parameter Weibull has also been used. A three-parameter mixed exponential was found to be the best fit distribution for daily rainfall series for a number of stations in U.S (Woolhiser and Roldan, 1982; Smith and Shreiber, 1974) and also in Quebec, Canada (Nguyen and Mayabi, 1990). The mixed exponential distribution has also given a better representation of precipitation extremes (Wilks, 1999a) than gamma improves the spatial coherency of precipitation simulated at a network of locations (Wilks, 1998).

The method of maximum likelihood (ML) or the method of moments has always been used in the estimation of parameters. An iterative method for the approximations of the ML estimators for gamma was presented by Greenwood and Durand (1960) while Rider (1961) initiated the initial parameter solutions for the mixed exponential function through the method of moments. A faster convergence to the optimal parameter set was done by solving seven likelihood functions with incremental initial guesses for 2 of the parameters within reasonable bound was suggested by Nguyen *et al.* (1990). However, all iterative convergence methods of ML estimates were found to be computationally exhaustive and often provided local optimum solutions. The method of moments on the other hand, has often given an inefficient parameter estimates for asymmetric distributions. It should be noted that robust global optimization methods such as the Shuffled Complex Evolution (SCE) method (Duan *et al.* 1992) and the Direct Search Complex (DSC) algorithm (Nelder and Mead, 1963) have not yet been commonly applied to parameter estimation of probability distribution using the ML method. With the recent advance of computing ability, these global optimization methods could provide more robust and reliable parameter estimates.

2.2.2.4 Modeling the Seasonal Variations

The seasonal variations of parameters of the probabilistic models are usually been accounted for by estimating the parameters in various methods. It can be handled by estimating parameters for discrete periods such as a monthly period or 3 monthly period. To be parsimonious with respect to the number of parameters needed to describe rainfall at a particular location during a climatologically year, many researchers have used Fourier series to describe the periodic seasonal fluctuations of parameters. Fayerherm and Bark (1965) used Fourier series to account for parameter variation in first-order Markov Chain models of precipitation occurrence. Ison *et al.* (1971) used least-squares estimates of Fourier coefficients to examine seasonal variability of gamma distribution parameters for the amount of precipitation for the i day wet period (i=1,2,….,i). Woolhiser and Pegram (1979,1986) studied seasonal and regional variability of parameters for stochastic daily precipitation models. They further used maximum likelihood estimates of the Fourier coefficients to describe the seasonal variability in parameters from a two-state Markov Chain model for occurrence and from a mixed exponential distribution for rainfall amounts.

2.2.2.4 Hourly Series Models

The stochastic models discussed above basically used time series of daily total precipitation but less effort has been devoted to data on shorter time scales (e.g. hourly), with the most prevalent approach being based on so-called conceptual (or physically based) models, which involve chance mechanisms(e.g. clustering) by which storms arrive (e.g. Neyman-Scott model). Katz and Parlange (1995) fitted the hourly precipitation amounts series into an extension of a form of chain-dependent process model that commonly fit to daily amounts. The extensions involve allowing hourly intensities to be auto-correlated and allowing the model parameters to possess diurnal cycles. The results are competitive, if not superior to the so-called conceptual models of the precipitation process.

2.4 Further Advances in Rainfall Modeling

The introduction of several new concepts and ideas in rainfall modeling had been witnessed in the past decade. The spectral theory of rainfall intensity based upon three components of stochastic point processes were used by Waymire *et al* (1984) and similar spectral structure were applied to stochastic modeling of rainfall by Yoo (1996) where the derivation was based on the autoregressive process that considered advection and diffusion. Elsner *et al*. (1993) examined the possibility of using the concept of entropy for the problem of assessing complexity and predictability of precipitation records. Yeboah et al. (1997) used a hybrid point rainfall model for the modeling of rainfall. The recent developments focus more on the refinement of the existing models towards applications to practical problems.

2.4 Weather Forecasting

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a future time at a given location. For millennia, people have tried to forecast the weather. In 650 BC, the Babylonians predicted weather from cloud patterns. In about 340 BC, Aristotle described weather patterns in Meteorological. Chinese weather prediction lore extends at least as far back as 300 BC. Ancient weather forecasting methods usually relied on observed patterns of events. For example, it might be observed that if the sunset was particularly red, the following day often brought fair weather. This experience accumulated over the generations to produce weather lore. However, not all of these predictions proved reliable and many of them have since been found not to stand up to rigorous statistical testing.

It was not until the invention of the telegraph in 1837 that the modern age of weather forecasting began. Before this time, it had not been possible to transport information about the current state of the weather any faster than a steam train. The telegraph allowed reports of weather conditions from a wide area to be received almost instantaneously by the late 1840's. This allowed forecasts to be made by knowing what the weather conditions were like further upwind. The two men most credited with the birth of forecasting as a science were Francis Beaufort, remembered chiefly for the Beaufort scale, and his protégé Robert Fitzroy, the developer of the Fitzroy barometer. Both were influential men in British Naval and Governmental circles, and though ridiculed in the press at the time, their work gained scientific credence, was accepted by the British Navy and formed the basis for all of today's weather forecasting knowledge.

As practiced by the professionally trained meteorologist, weather forecasting today is a highly developed skill that is grounded in scientific principle and the method makes use of advanced technological tools. The notable improvement in forecast accuracy that has been achieved since the 1950s is a direct outgrowth of technological developments, basic and applied research, and the application of new knowledge and methods by weather forecasters. High-speed computers, meteorological satellites, and weather radars are tools that have played major roles in improving weather forecasts.

A policy statement of the American Meteorological Society as adopted by the Council on 13^{th} January 1991 stated that the most impressive gain in forecast accuracy in recent years has been in the prediction for the 1 to 5 day range. A number of factors have contributed to the increase in accuracy. Foremost among these has been the further development of numerical prediction models, based on the laws of physics that are able to forecast the formation and movement of the large high and low pressure systems that govern day-to-day weather changes in middle and high latitudes.

Several other factors have also contributed significantly in increasing the forecasting accuracy. One is the development of statistical methods for enhancing the scope and accuracy of model predictions. Statistical methods allow a wider variety of meteorological elements to be predicted than do the models alone, and they tailor the geographically less precise model forecasts to specific locations.

A number of different statistical and machine learning techniques have emerged in the last decades. These techniques extract the information contained in meteorological databases of historical observations to train specific forecast models such as the regression model, hidden Markov models and neural networks. The resulting models predict future outcomes of a given variable based on the past evidence collected in the database.

There have also been some attempts for combining both database information and the numerical prediction models. This is done by combining the model's predicted patterns with the information available in the databases such as rainfalls, and predictions, such as gridded atmospheric patterns. Employing downscaling methods, sub-grid detail in the prediction is gained by post-processing the outputs from the numerical prediction models using knowledge extracted from the databases (Murphy, 1999). One of the most popular downscaling techniques is the method of analogs, which assumes that similar atmospheric patterns may lead to similar future outcomes. Thus, predictions based on an atmospheric pattern can be derived from an “analog ensemble” extracted from the database.

Another factor that increases forecasting accuracy is the improved observational capability afforded by meteorological satellites (Matthew et al., 2003). The continued improvement of the initial conditions prepared for the forecast models also contributes to the increase in accuracy. Satellites now provide the capability for nearly continuous viewing and remote sensing of the atmosphere on a global scale. The improvement in initial conditions is the result of an increased number of observations and better use of the observations in computational techniques.

2.6 Rainfalls Forecasting Techniques

Forecasting rains is one of the most difficult tasks in weather prediction due to the scarce knowledge on how to characterize the mechanisms taking part in its formation. Short term forecasting of rainfall fields is one of the major tasks to achieve efficient forecasts of flood events. Regardless of the model adopted to predict rainfall, it has been demonstrated that it allows extending of the lead time of flood forecasts, as well as improving the estimate of flood for a given forecast lead time (Brath et al., 1988).

Many different techniques have been proposed to forecast rainfalls. Among these, a physically based approach which makes use of meteorological models might be appealing. One example is the numerical weather prediction (NWP) model. Early in the 20th century, advances in the understanding of atmospheric physics led to the foundation of modern numerical weather prediction. In 1922, Lewis Fry Richardson published "Weather prediction by numerical process," which described how small terms in the fluid dynamics equations governing atmospheric flow could be neglected to allow numerical solutions to be found. They took the analysis as the starting point and evolved the state of the atmosphere forward in time using understanding of physics and fluid dynamics. However, the sheer number of calculations required was too large to be completed without the use of computers. Nowadays, the complicated equations which govern how the state of a fluid changes with time can be solved by supercomputers. The output from the model provides the basis of the weather forecasts. Unfortunately, a major limitation stems from the spatial and temporal resolution of the hydro meteorological variables required for the initialization of deterministic models where wind speed, relative humidity, temperature and pressure profile cannot be provided by most of the operational monitoring networks (Burlando et al., 1996).

In the 1960s, the chaotic nature of the atmosphere was first observed and understood by Edward Lorenz, the founder of the field of chaos theory. These advances have led to the current use of ensemble forecasting in most major forecasting centers and to taking into account uncertainty arising from the chaotic nature of the atmosphere. It is the second limitation of physically based approaches that could also be viewed in the chaotic structure of the thermodynamic equations to be solved (Ghil et al., 1985; Tsonis and Elsner, 1989). This can be detected as an intrinsic limit to predictability of rainfall (Rodriguez Iturbe et al., 1989; Ghilardi and Rosso, 1990).

Since 1986, the neural network technique has drawn considerable attention to many researchers as it can handle the complex and nonlinear problems better than the conventional statistical techniques where it has the ability to predict future values of the time series. Elsner and Tsonis (1992) have shown that the neural network can be successfully used to predict the chaotic series. It is useful for stochastic and deterministic forecast processes where in deterministic forecast process, rainfall time series are treated as deterministic and even chaotic.

Nevertheless, some improvements can be expected as related to further developments of mixed stochastic-deterministic models where they include both deterministic and stochastic aspects in the model such as the so-called limited area models, or simplified meteorological models that act at the basin scale (Georgakakos and Krajewski, 1991).

Rainfall forecasting based on stochastic models may still represents a useful tool. In the literature, several attempts to forecast rainfall based on mathematical models can be found. Most of them are statistical black-box models where the functional relationships between system inputs and system outputs are studied. The main advantage of this model is that they are not as data demanding as the physical models. This model develops the concept of storm tracking, based on cross-correlation between rainfall either observed at various rain gages, or tracked by radar signals (Nguyen et al., 1978; Phanartzis, 1979; Johnson and Bras, 1980).

2.6 Time Series and Forecasting

A time series is a sequence of observations taken sequentially in time. There are many sets of time series data such as a weekly series of the number of customer in a supermarket, a yearly series for the prices of gold and hourly observations made on the yield of a chemical process. Time series are found in many fields such as economics, business, engineering, natural sciences and social sciences.

Time series analysis comprises methods that attempt to understand such time series, often either to understand the underlying context of the data points such as where they came from or what generated them. The term time series analysis is used to distinguish a problem, firstly from more ordinary data analysis problems where there is no natural ordering of the context of individual observations and secondly from spatial data analysis where there is a context that observations often relate to geographical locations. There are additional possibilities in the form of space-time models which are often called spatial-temporal analysis. A time series model will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values in a series for a given time will be expressed as deriving in some way from past values, rather than from future values.

There are many applications of time series. One is where the time series are used to develop models where predictions can be made. This is called time series forecasting. Time series forecasting is the use of a model to forecast future events based on known past events to forecast future data points before they are measured.

Forecasting is the process of estimation in unknown situations. Prediction is a similar, but more general term, and usually refers to estimation of time series, cross-sectional or longitudinal data. Risk and uncertainty are central to forecasting and prediction. In more recent years, forecasting has evolved into the practice of demand planning in every day business forecasting for manufacturing companies. The discipline of demand planning, also sometimes referred to as supply chain forecasting, embraces both statistical forecasting and consensus process.

2.6.1 Statistical Time Series and Forecasting

Statistical analysis of time series data started a long time ago (Tsay, 2000), and forecasting has an even longer history. The objectives of the two studies may differ in some situations but forecasting is often the goal of a time series analysis. Applications played a key role in the development of time series methodology. The following are uses of time series analyses in business and economics:

(i) To study the dynamic structure of a process.
(ii) To investigate the dynamic relationship between variables.
(iii) To perform seasonal adjustments of economic data such as the gross domestic product and unemployment rate.
(iv) To improve regression analysis when the errors are serially correlated.
(v) To produce point and interval forecasts for both level and volatility series.

To facilitate discussion, we denote a time series at time t by z_t and let ψ_{t-1} be the information set available at time $t-1$. It is often assumed that ψ_{t-1} is the σ- field generated by the past values of z_t. A model for z_t can then be written as

$$z_t = f(\psi_{t-1}) + a_t \tag{2.2}$$

where a_t is a sequence of independent and identically-distributed random variables with mean 0 and finite variance σ_a^2. It is evident from the equation that a_t is the one-step-ahead forecast error of z_t at time origin $t-1$ and hence it is often referred to as the innovation or shock of the series at time t. The history of time series analysis is concerned with the evolution of the function $f(\psi_{t-1})$ and the shock a_t.

The publication of Time Series Analysis: Forecasting and Control by Box and Jenkins in 1970 was an important milestone for time series analysis. It provided a systematic approach that enables practitioners to apply time series methods in forecasting. It popularized the autoregressive integrated moving average (ARIMA) model by using an iterative modeling procedure consisting of identification, estimation, and model checking. The success of ARIMA models generated substantial research in time series analysis. Originally, time series analysis was divided into frequency domain and time domain approaches. The time domain approach uses autocorrelation function, ρ_l of the data and parametric models, such as the ARIMA models, to describe the dynamic dependence of the series (Box, Jenkins, and Reinsel, 1994). The frequency domain approach on the other hand focuses on spectral analysis or power distribution over frequency to study theory and applications of time series analysis. A power spectrum of a stationary z_t is the Fourier transform of the autocorrelation function ρ_l (Brillinger, 1975; Priestley, 1981). Cooley and Tukey made an important advance in frequency-domain analysis by making spectral estimation efficient (Tsay, 2000).

The objective of an analysis and experience of the analyst are the determining factors between which approaches to use. In the context of Bayesian and non-Bayesian time series analyses, there remain some differences, but the issue has been shifted those of practicality rather than philosophy. Durbin and Koopman provided both classical and Bayesian perspectives in time series analysis (Tsay, 2000).

The advances in computing facilities and methods have profound impacts on time series analysis. There are many important developments within the so called "traditional analysis", for example, linear Gaussian processes with parametric models. In model diagnostics, outlier analysis and detecting structural breaks have become an integral part of the model. Chang, Tiao, and Chen (1988) for example, looked at outlier detection while Martin and Yohai studied influential functionals (Tsay, 2000). Outlier analysis in time series are concerned with aberrant observations in z_t and a_t, or in other words the observations straying from the right or normal way, and the changes in the mean of z_t and the variance of a_t. Akaike (1974) and Hannan (1980) proposed some model selection criteria to help in the time series model selection. Some important advances in pattern identification methods have also been developed for example, the R- and S-array of Gray, Kelley, and McIntire (1978) and the extended autocorrelation function of Tsay and Tiao (1984). The pattern identification methods are capable of handling both stationary and unit-root nonstationary series. Choi (1992) discussed the many developments in ARMA model identification. The exact likelihood method now becomes the standard method of estimation. The foregoing developments are not in isolation with other developments in the area and their impacts are not limited to linear Gaussian time series models (Tsay, 2000).

Generally speaking, two important technical advances in the recent history of time series analysis have generated much interest on the topic. The first advance is the use of state-space parameterization and Kalman filtering. This happened largely in the 1980s, as evidenced by the explosion in the papers published in statistical journals that have "state-space" or "Kalman filter" in their titles. The original purpose of introducing Kalman filter into time series analysis was mainly to evaluate efficiently the exact

Gaussian likelihood function of a model and to handle missing observations. The usefulness of the technique was extended beyond estimation, where it led to developments of new methods for signal extraction, for smoothing and seasonal adjustment, and for renewal interest in structural models (Tsay, 2000).

The second technical advance in recent time series analysis is the use of Markov Chain Monte Carlo (MCMC) methods, especially Gibbs sampling and the idea of data augmentation. The applicability of MCMC methods to time series analysis is widespread and indeed the technique has also led to various new developments in time series analysis. These include nonnormal and nonlinear state-space modeling and inference and prediction of autoregressive models with random mean and variance shifts, including using explanatory variables to estimate transition probabilities in mean and variance. The MCMC methodology also led to increasing use of simulation methods in time series analysis, especially in tackling complicated problems that were impossible to handle a few years ago (Tsay,2000).

The past several decades also brought many important advances in time series methodology. One of it is for the multivariate process. Methods for analyzing multivariate series have been developed, especially in structural specification of a vector system. The usefulness and need of considering jointly several related time series were recognized a long time ago (Quenouille, 1957). However, multivariate analysis is often confined to vector autoregressive (VAR) models. Two reasons for this lack of progress are:

(i) The generalization of univariate ARMA models to vector ARMA models encounters the problem of identifiability.

(ii) Multivariate models are much harder to estimate and to understand, and there is a propensity to use perceived simpler models.

A related development in multivariate time series analysis is the cointegration of Engle and Granger (1987). Cointegration means that a linear combination of marginally unit-root nonstationary series becomes a stationary series. It has become popular in econometrics because cointegration is often thought of as the existence of some long-term relationship between variables. In the statistical literature, the idea of a linear combination of unit-root nonstationary series becoming stationary was studied by Box and Tiao (1977). Associated with cointegration is the development of various test statistics to test for the number of cointegrations in a linear system. Despite the huge literature on cointegration, its practical importance is yet to be judged. This is due primarily to the fact that cointegration is a long-term concept that overlooks the practical effects of scaling factors of marginal series (Tsay, 2000).

Since the last decade, multivariate forecasting methods have given rise to more research than univariate methods have. This is partly because computational advances have made them more feasible in practice. It seems natural to try to improve forecasts of one variable by including appropriate explanatory variables in the model. Identifying all the relevant variables may not be easy and it is important to study the context, to ask questions and to look for previous empirical regularities. There is always the contrary danger of including unnecessary explanatory variables, which appear to improve the fit but actually lead to poorer out-of-sample forecasts. Although most people expect multivariate forecasts to be better than univariate forecasts, this is not necessarily the case. However, they may still improve our understanding of the interrelationships between variables.

There are many types of multivariate models. One basic question is whether there is a causal relationship between the explanatory variables and the response variable, and also whether the system is of open loop structure or whether changes in the response variable feed back to affect the explanatory variables in a closed loop way. Multiple regression is still the most commonly used method but there can be problems in fitting such models to economic time series data where the variables can be correlated with each other and with time, and where feed-back may be present. Although a good fit can often be obtained, poor forecasts may still result. It is arguable that this is partly because the error structure of regression models is overly simplistic for use with time series data and there has been much work on alternative classes of multivariate time series model notably vector ARMA (VARMA) models. Software has become available but VARMA models are still not easy to fit even with only two or three explanatory variables. Partly because of this, many analysts prefer to restrict attention to vector autoregressive (VAR) models or even further to low order VAR models. Empirical evidence does suggest that restricted VAR models give better out-of-sample forecasts than unrestricted VAR models.

Multivariate methods are worth considering when appropriate expertise is available and when suitable explanatory variables have been identified and measured, especially when one or more of them are leading indicators. Multivariate forecasts are sometimes worth the extra effort that they entail, and multivariate models usually do give a better fit. However, it is important to realize that out-of-sample forecasts from multivariate models are not necessarily more accurate than those from univariate models either in theory or practice, because of the following reasons:

(i) Exogenous variables may have to be forecasted.
(ii) Economic data are generally observational rather than designed data, and so may be unsuitable for fitting multivariate models.
(iii) 'Simple may be best'. It appears that simple univariate methods are often more robust to model misspecifications and to changes in the model than more complicated models are.

Multivariate forecasts are more accurate than univariate extrapolations in many case studies. Despite the research interest in alternatives, such as VAR models, multiple regression is still the most commonly used multivariate model. This is because of its simplicity.

A multivariate autoregressive integrated moving average (MARIMA) model is more likely to be the same as the autoregressive integrated moving average (ARIMA) model. However, instead of analyzing only a series, we observe simultaneously several series. Such time series data may be available on several related variables of interest or in other words, there is more than one series involved in such a model. The reasons for analyzing and modeling such series jointly are to understand the dynamic relationships among them. They may be contemporaneously related, one series may lead the others or there may be feedback relationships. Another reason is to improve the accuracy of forecasts. When information of one series is contained in the historical data of another, better forecasts can be obtained when the series are model jointly (Tiao and Box, 1981). In this view, the operational use of MARIMA model was suggested by Montanari et al. (1994), who highlighted how a multivariate scheme could remarkably improve the forecasts.

CHAPTER 3

METHODOLOGY

3.1 Convective Rainfall

To analyse and characterize convective rain in Klang Valley, the temporal pattern and the spatial distribution between meteorological radar data and surface rainfall (rain gauge) need to be explored. This chapter presents the methodologies used in this research with focus on characterization of rain properties, establishment of criteria for separating convective from non-convective storms and checking discrepancies or similarity between meteorological radar data and observed surface data (rain gauge). The source of data and limitations are also described.

3.1.1 Research Design and Procedure

The research procedure of this study is summarised in Figure 3.1 below:

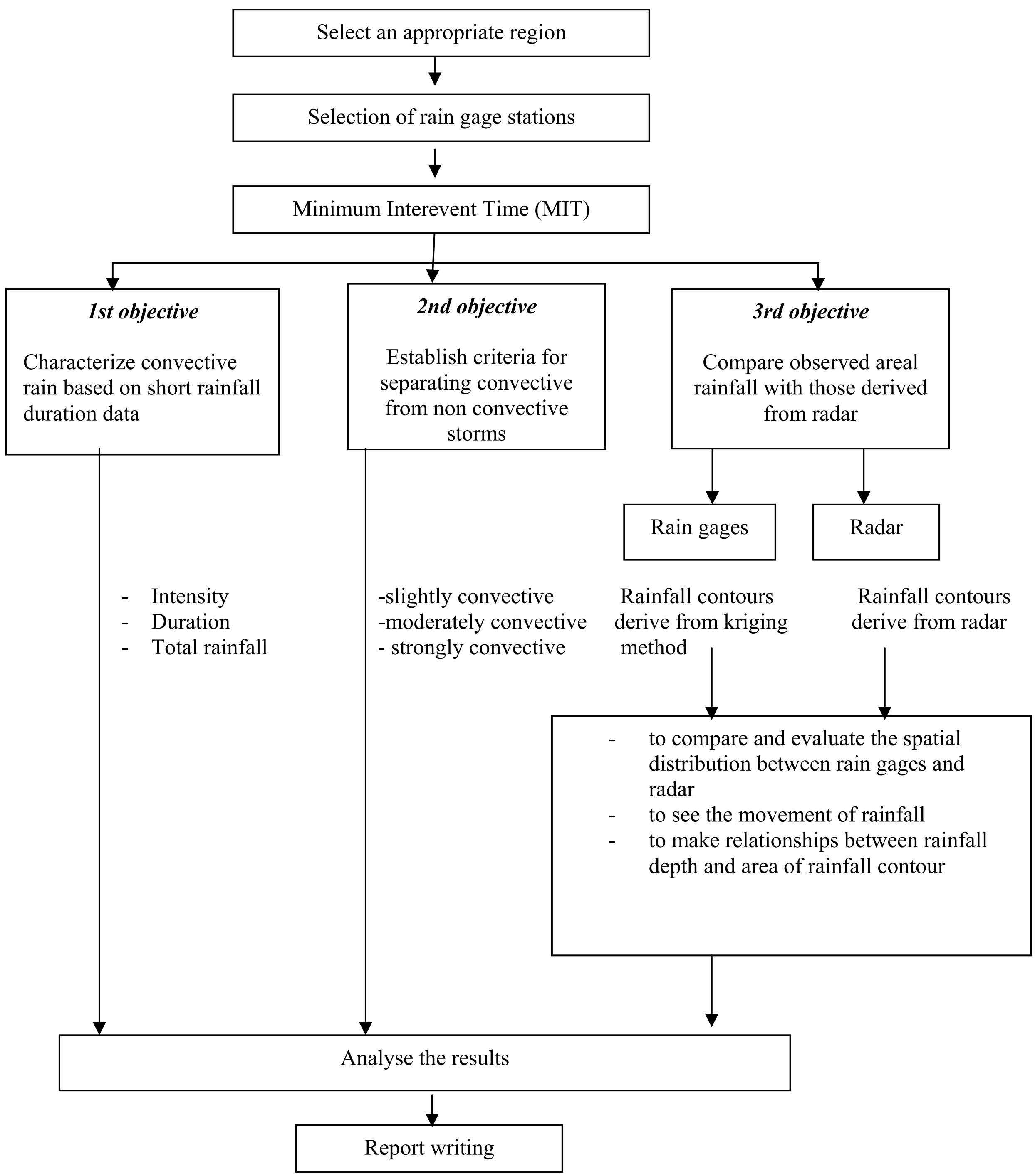


Figure 3.1 : Flow chart of research design and procedure

3.2.3 Study Area

The study area covers the whole Klang Valley, comprises Kuala Lumpur and its surroundings and suburbs. Klang Valley is surrounded by hilly areas especially to the east and northeast and the Port Klang coastline to the west. Based on the most recent census, the population in the Klang Valley has expanded to 26.64 million (Statistics Bulletin, 2006 June), and it has an area of about 3200 sq. Km.(Norhan and Mazian, 1997) The climate of the area is tropical with averages temperature range from 22^{0}C to 33^{0}C throughout the year and the relative humidity as high as 90%. Being located in the equatorial zone, the climate is governed by the northeast and southwest monsoons. The northeast monsoon usually commences in early November and ends in March and the southwest monsoon is usually established in the later half of May or early June and ends in September. These two main monsoon seasons are separated by two relatively short inter-monsoon seasons which usually recorded heavy rainfall. The annual rainfalls vary between 2,000 mm and 2,500 mm and the mean monthly rainfall between 133 mm and 259 mm (Desa *et al*., 2005).

Peninsular Malaysia

Figure 3.2 : The study area in Klang Valley

Figure 3.2 shows the area of Klang Valley (inset) from the map of Peninsular Malaysia and rainfall station 3117070-JPS Ampang which supplies data for the study.

3.1.3 Terminal Doppler Radar

The radar images were derived from the Terminal Doppler Weather Radar (TDR) located at Bukit Tampoi, about 10 km north of Kuala Lumpur International Airport (KLIA). The TDR is primarily used for the detection and warning of wind shear and micro bursts in the vicinity of KLIA. RADAR stands for Radio Detection and Ranging and it's used for detecting the position, velocity and characteristic of target (bearing, range, and altitude). The difference between a conventional weather radar and Doppler weather radar is that the former can only detect the characteristic, size, direction and distance of precipitations while the latter can detect not only the characteristic, size, direction and distance of precipitations but also radial wind speed, wind shear and microburst. Figure 3.3 shows the TDR at KLIA. Table 3.1 summarizes the principle characteristics of this radar.

Figure 3.3 : Terminal Doppler Radar at KLIA

Table 3.1 : Main characteristics of KLIA Terminal Doppler radar used in this study

Radome	- 12 m. diameter
Parabolic Reflector	- 8.5 m. diameter
Wavelength	- 10 cm
Frequency	- 2874.5 MHz
Peak power	- 750 KW
Pulse Width	- 1.0 µs /3.0 µs
Pulse Repetition	- 1000Hz (1.0 µs pulse width)
Frequency	- 300 Hz (3.0 µs pulse width)
Azimuth Resolution	- 0.7°
Range Resolution	- 125m
Doppler Velocity	- 1.0m/s

The colours on radar images represent the values of energy reflected toward the radar. The reflected intensities or echoes are measured in dBZ (decibles of z). The scale of dBZ values is also related to the intensity of rainfall. Typically, light rains have dBZ value of less than 20. The higher the dBZ, the stronger the rain intensity. The Doppler radar does not determine where rain is located, only areas of returned energy (National Weather Service, 2006). The "dB" in the dBZ is logarithmic and has no numerical value, but is used only to express a ratio. The "z" is the ratio of the density of water drops (measured in milimeters, raised to the 6th power) in each cubic meter (mm^6/m^3). Mathematically:

$$dBZ = 10*\log (z/z0) \qquad (3.1)$$

where,

z = reflectivity factor

z0 = 1 mm^6/m^3

When the "z" is large (many drops in a cubic meter), the reflected power is large. A small "z" means little returned energy. In fact, "z" can be less than 1 mm^6/mm^3 and since it logarithmic, dBZ values will become negative, as often in the case when the radar is in clear air mode and indicated by earthtone colours (National Weather Service, 2006). Figure 3.4 shows rainfall image from Doppler radar at KLIA. The intensity was measured in two units. On the left side, the scale is in dBZ and on the right in mm/hr. In this study, rainfall intensity in mm/hr was used to show the rainfall rate in digitized image. The Doppler radar image has too many colours for

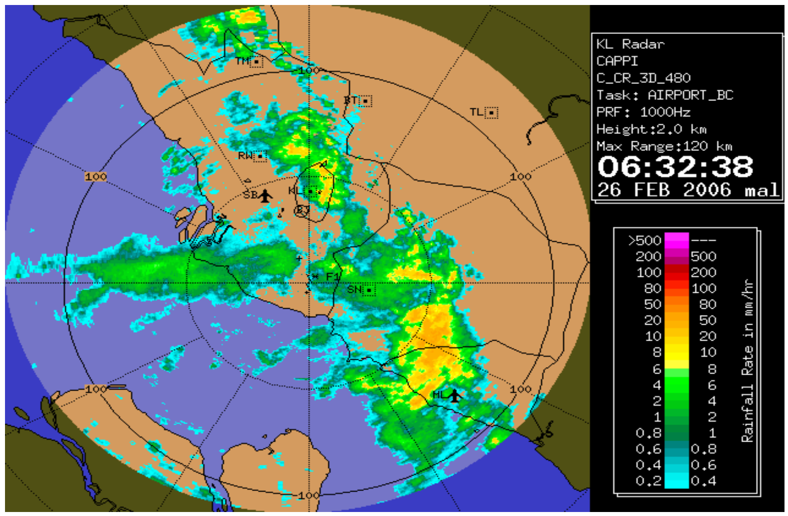


Figure 3.4 : Radar image

(a)

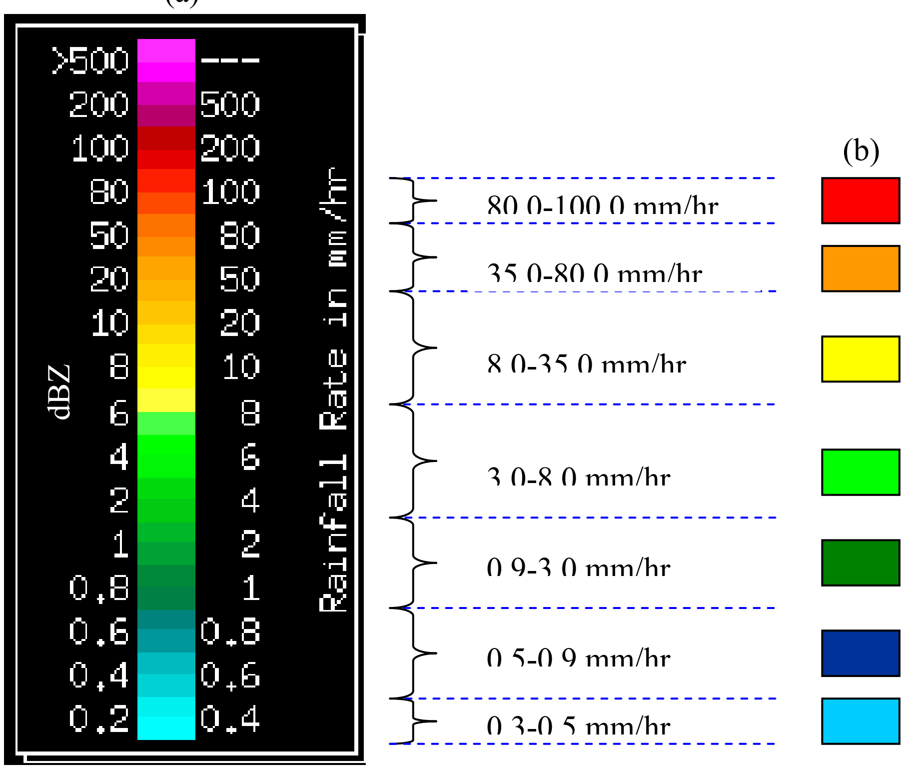


Figure 3.5 : Various level of reflectivity colour derived from radar image (a) and (b) simplified rainfall intensity colour after digitization

various intensity scales. However, it is visually difficult to differentiate these colours. To simplify the data analysis, the colour scales were reduced to seven by redigitizing the radar image. The new intensity scales and the corresponding radar intensity values are shown in Figure 3.5. These scales were used in determining of rainfall contours. These scales were used to construct rainfall contours.

3.1.4 Data Source and Collection

In order to analyse convective rain of the study area, several different data sources are used. In the first stage, a five year (2000-2004) rainfall data recorded from hydrological data bank, Department of Irrigation and Drainage (DID) at station 3117070-JPS Ampang was extracted. All data from this station were used to execute first and second objectives. In the second stage, rainfall data from 20 raingauges (9 raingauges in Wilayah Persekutuan and 11 raingauges in Selangor) were selected to achieve the fourth objective, which is determine the spatial distribution between meteorological radar data and observed surface data (raingauge). Ground data was obtained from DID, while radar data were taken from Malaysian Meteorological Department (MMD), KLIA in Sepang. Heavier rainfalls were selected for this analysis. These events coincided with major flood events. These events occurred on June 10, 2003, Nov 5, 2004, Jan 6, Feb 26, Apr 6, and May 10, 2006. Table 3.2 lists the various data sources of Klang Valley.

Table 3.2 : Data sources

<table>
<tr><th></th><th colspan="2">Data Description</th><th>Year/Date</th><th>Sources</th><th>Method of data collection</th></tr>
<tr><td>1st and 2nd objectives</td><td>Rain gauge</td><td>3117070 – JPS Ampang</td><td>2000-2004</td><td rowspan="2">Department of Irrigation & Drainage (DID), Malaysia</td><td rowspan="2">Hydrological data bank</td></tr>
<tr><td rowspan="2">3rd and 4th objective</td><td>Rain gauges</td><td>WILAYAH PERSEKUTUAN
3116003 – Ibu Pejabat JPS
3116006 – Ldg Edinburgh Site 2
3216001 – Kg. Sg Tua
3217001 – KM 16, Gombak
3217002 – Emp. Genting Klang
3217003 – KM 11, Gombak
3217004 – Kg Kuala Sleh
3317001 – Air Terjun Sg Batu
3317004 – Genting Sempah

SELANGOR
2917001 – JPS Kajang
3014084 – JPS Klang
3014091 – UiTM Shah Alam
3018101 – Emp. Semenyih
3115079 – Pt Penyelidikn Sg Buloh
3117070 – JPS Ampang
3118102 – SK Kg Lui
3119104 – Jln Genting Peres
3216004 – SMJK Kepong
3315037 – Tmn Bukit Rawang
3315038 – Country Home</td><td rowspan="2">10th Jun 2003
05th Nov 2004
06th Jan 2006
26th Feb 2006
06th Apr 2006
10th May 2006</td></tr>
<tr><td>Radar</td><td>The whole Klang Valley</td><td>Malaysian Meteorological Department (MMD), KLIA, Sepang</td><td>Radar data</td></tr>
</table>

3.1.5 Data Analysis

3.1.5.1 Separation of Rainfall Events

Rainfall events must be isolated before they can be analysed. The period without rainfall or interevent time definition is a typical criterion used to isolate an individual rainfall event from continuous rainfall. The criterion is also well known as minimum interevent time (MIT) (Figure 3.6). Many researchers used MIT values between 0 and 50 hours to separate rainfall events (e.g. Hydroscience, 1979; Bedient and Huber, 2002) while Adams et. al., (1986) suggested MIT values between 1 and 6 hours for urban applications. In this study, a rainfall event is defined from Minimum Interevent Time (MIT) method. The annual numbers of rainfall events were plotted against different MIT values and an appropriate MIT value is selected from the graph at point after which increases in the MIT do not result in significant changes in the number of event.

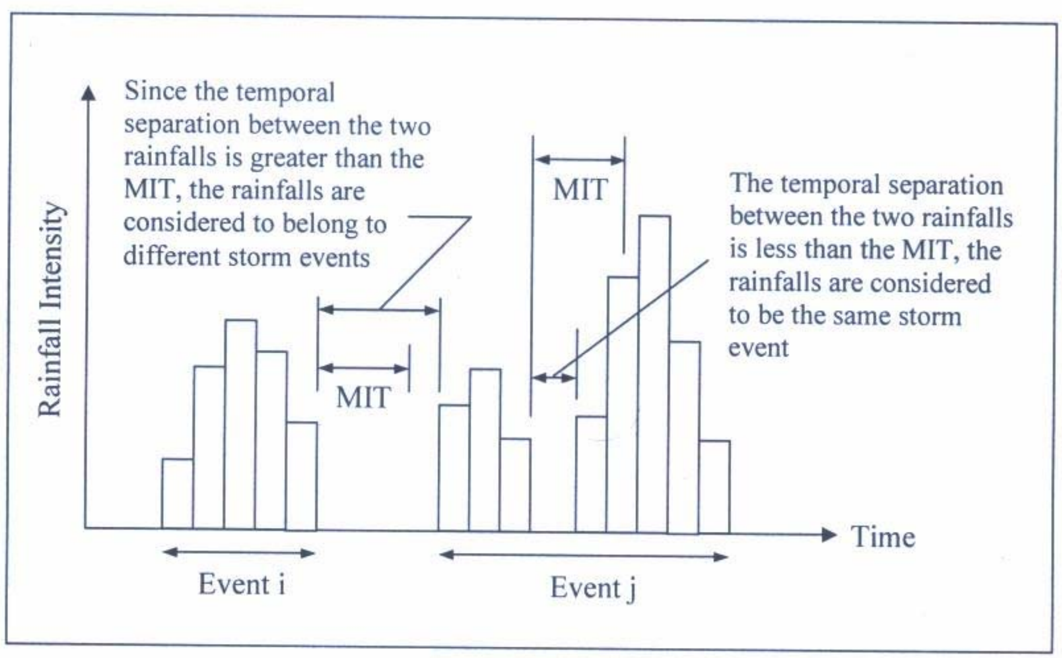


Figure 3.6 : Separation of rainfall events based on minimum interevent time (MIT)

3.1.6 Analysis of Convective Rain

3.1.6.1 Temporal

The aim of this study is to characterize convective rain in Klang Valley. Initially rainfall data is analysed in terms of intensity, rainfall duration and total rainfall. Short interval rainfall data recorded between years 2000 and 2004 were used. In year 2000, DID has installed automatic raingauges that can record short intervals of 1-minute or 5-minutes rather than 15-minutes intervals as previously recorded. Shorter rainfall aggregation can give more accurate information about the duration of a storm and thus short intervals data is needed for analyse convective rain. This is because convective storms usually lasted over a short period of time.

A five year rainfall data recorded at JPS Ampang (3117070) was analysed. In the beginning, the diurnal and monthly rainfall patterns at Ampang station were studied. The separation between non-convective and convective event were carried out based on a 35mm/hr threshold for each 5 minute interval. This threshold is very often used in precipitation models for engineering applications to set apart non-convective from convective precipitation (Llasat, 2001). Five minute intensity is used because rainfall data are already collected in 5 minutes interval. The convective characteristics were clearly shown in storm shape where 10 storms were selected to show the rainfall pattern. Next, convective event was divided into four classes based on the β parameter. This classification is according to their greater or lesser convective character (Llasat, 2001). The β parameter is determined using equation (3.2):

$$\beta_{L,\Delta T} = \frac{\left[\sum_{i=1}^{N} I(ti, ti + \Delta T) > L\right]}{\sum_{i-1}^{N} I(ti, ti + \Delta T)} \qquad (3.2)$$

where,

ΔT	=	time interval of accumulation of the precipitation
$I(ti,ti+\Delta T)$	=	precipitation measured between ti and ti+ΔT

L = is set at 35 mm/hr

N = total number of ΔT integration steps into which the episode is divided

Llasat further divided the storms into four categories based on the β values as follows:

$\beta = 0$ non-convective

$0 < \beta \leq 0.3$ = slightly convective

$0.3 < \beta \leq 0.8$ = moderately convective

$0.8 < \beta \leq 1.0$ = strongly convective

3.1.6.2 Spatial Distribution

The spatial distribution of rainfall derived from meteorological radar data was compared with surface rainfall data (rain gauge) using Geographical Information System (GIS). There are a number of softwares available in GIS, for example ArcView, ArcInfo and ArcGIS. All of these softwares are developed by ESRI, which is one of the most analytically developed GIS products. In this study, ArcGIS 9.1 is used to digitize radar data and displaying the image in rainfall contour. Radar data need to be digitized first because the image which is taken from KLIA Meteorological Station is in JPEG format. This format is the end product of Interactive Radar Information System (IRIS), the radar software used at KLIA and IRIS cannot give rainfall image in GIS format. Figure 3.7 shows radar image taken from KLIA Meteorological Station.

The digitized images using ArcGIS can give the area of every colour code and the corresponding rainfall intensity. On the other hand, the isohyetal line for surface rainfall was constructed using TIDEDA database. TIDEDA is a computer program for processing time-dependent data, particularly hydrological data. Comparison was made based on a 5-minutes rainfall. For similar event and time four heavier rainfalls were selected for this analysis. These events coincided with major flood events. These events occurred on 10th Jun 2003, 05th Nov 2004, 06th Jan, 26th Feb, 06th Apr, and 10th May 2006. For every event, several images at different time were selected and digitized. By matching the same occurrence time, line rainfall contour from surface data (Kriging) were compared with rainfall contour radar image (digitized image). Finally, a relationship between areas of rainfall contour (derived from Kriging) with rainfall depth was examined. Table 3.3 shows the time of images, which are selected for spatial comparison and correlation.

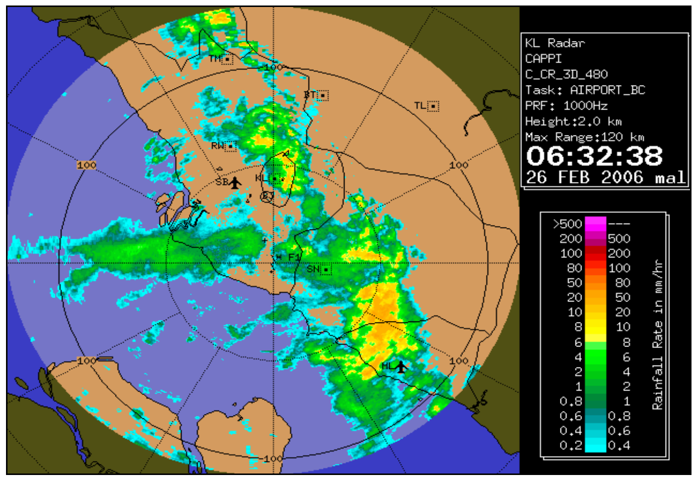


Figure 3.7 : Radar image in JPEG format

During these events, twenty rain gauge stations in Klang Valley exhibited relatively good continuity of rainfall data. All of the rain gauges are selected to compare spatial distribution between radar data and surface rainfall data. Figure 3.8 shows the locations rainfall stations used in this study.

Table 3.3 : Times during which the digitized images were captured by TDR

	Date of events			
	Jan 6, 2006	**Feb 26, 2006**	**Apr 6, 2006**	**May 10, 2006**
Capturing Time (hh:mm)	18:19	03:23	15:08	15:01
	18:25	04:55	15:13	15:12
	18:30	06:21	15:19	15:28
	18:36	06:32	15:29	15:33
		06:38	15:35	15:39
		06:43	15:41	

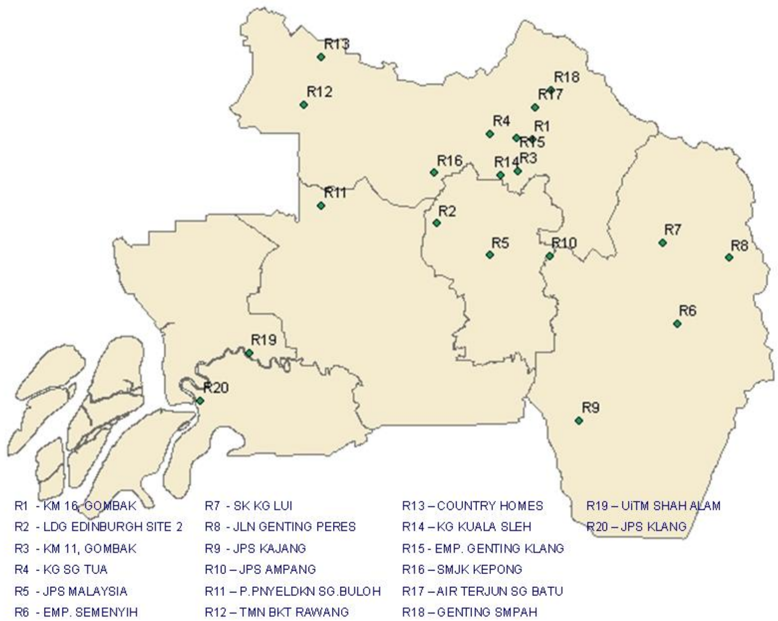


Figure 3.8 : Locations of twenty rain gauge stations selected in this study

3.1.6.3 Procedure To Derive Rainfall Contour from Radar and Raingauge Data Using GIS

As already noted in section 3.6.2.2, radar images which is taken from KLIA Meteorological Station is in JPEG format. All images need to be digitized before rainfall contours is created. Radar images need to be digitized with layer by layer according to the colour of intensity in that image. Due to the number of intensities, it is visually to differentiate those colours. To simplify the data analysis, the colour scales were reduced to seven by redigitizing the radar image (see Figure 3.5). The new intensity scales and the corresponding radar intensity values were used in radar's contour. Figure 3.9 shows the flow chart to produce rainfall contour derived from radar. The process of digitizing radar image is shown in Appendix A.

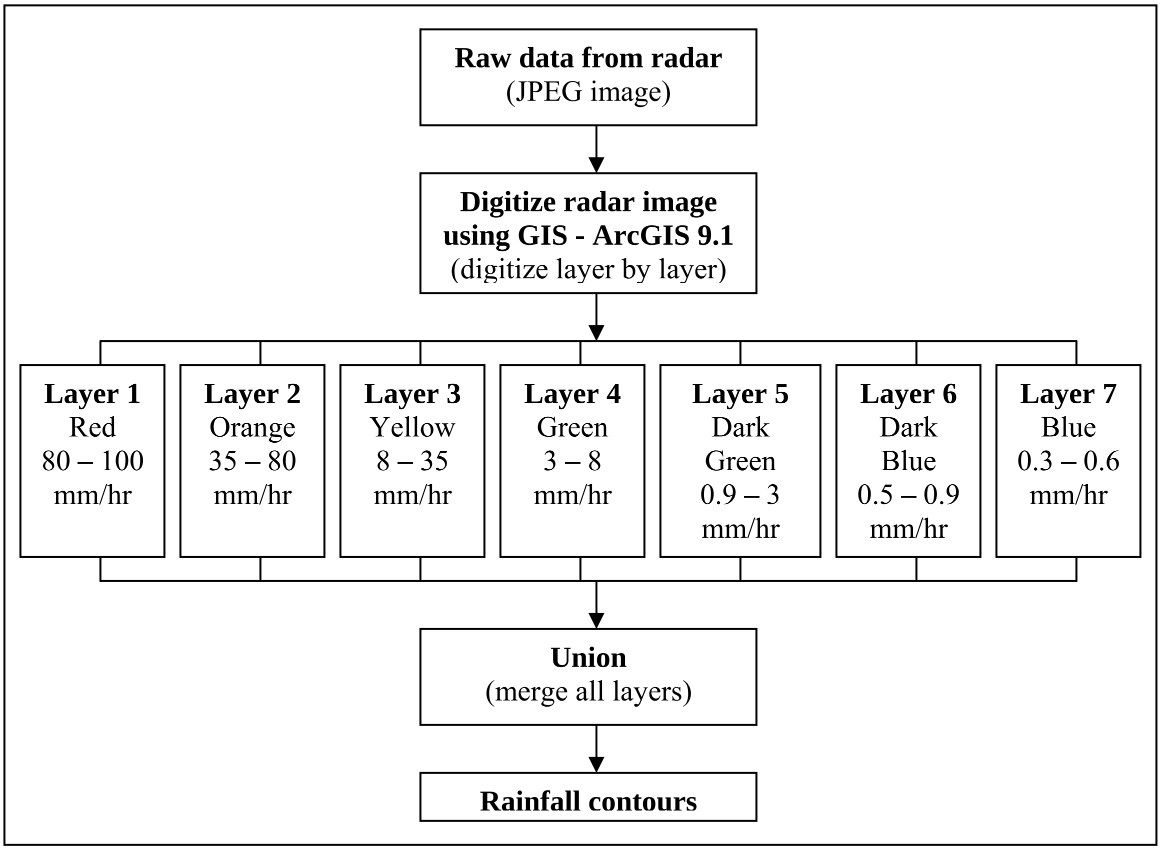


Figure 3.9 : Flow chart of making rainfall contours derived from radar

Rainfall contours from surface rainfall were derived by GIS also where Kriging Method was used in ArcGIS 9.1. As noted in Chapter 2, Kriging produces an estimate of the underlying (usually assumed to be smooth) surface by a weighted average of the data, with weights declining with distance between the point at which the surface is being estimated and the locations of the data points. Since raingauge station is selected, the location of rainfall station in Klang Valley was shown in point features in GIS. All intensities for every raingauge station were key-in in GIS. Using ArcGIS, Kriging Method can be implemented in two ways either Spatial Analyst or Geostatistical Analyst. In this study, Geostatistical Analyst is chosen because the Matern model (now it is recognized as K-Bessel) tends to produce surfaces that are smoother locally (on a very fine scale) than some other models (such as the exponential or spherical). Beside that, among the advantages of the implementation of kriging in Geostatistical Analyst relative to that in Spatial Analyst are the ability to handle directionality in the data and the ability to make plots of prediction errors as a way of assessing uncertainty. There have four steps to execute kriging in Geostatistical Analyst. Figure 3.10 shows the flow chart of producing rainfall contours by ground data. The four steps during interpolate the rainfall contour in ArcGIS can be seen in Appendix B. After both of rainfall contours were created, the spatial distributions of rainfall were compared in term of intensity and area. The area of rainfall contours also determined by GIS.

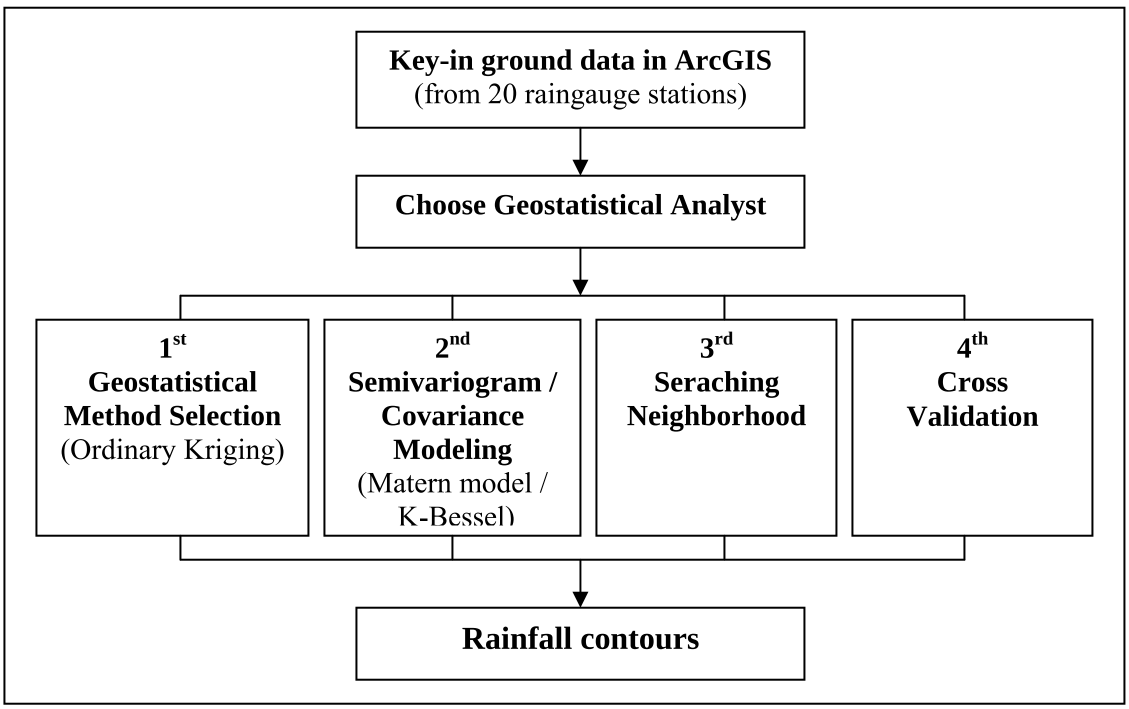


Figure 3.10 : Flow chart of making rainfall contours derived from ground data

3.1.6.4 Storm Movements and Depth Area Relationship

The movement of rainfall pattern also observed. In this study, four flash flood events that had occurred in the Klang Valley were chosen. The storms bringing rains leading to the flash floods had exhibited convective characters. These events also are a good example of unusually strong convective events responsible for heavy rainfall. To identify convective rainfall in radar images, a value of 35 dBZ is taken as reflectivity threshold. This technique was developed by Dong and Hyung (2000), where they were used this value in study of heavy rainfall with mesoscale convective systems over the Korean Peninsular. Beside that, this value also is already noted in radar's rate, so it is easy to read the reflectivity according to radar's colour code. The highest reflectivity, which is greater than 35 dBZ is chosen as centre of the storm for convective events. The centre of the storm is used as reference to show the movement of the storms. The coordinates of every movements of centre of the storms were plotted in RSO (Rectified Skew Ortomorphic) Malaysia, which is one of coordinate system and it is interpreted from GIS (ArcGIS 9.1).

Next, in order to get the relationship between area and rainfall depth, surface rainfall data from eleven raingauge stations were used. The rainfall depth pattern and the area for every color code of rainfall contours in small catchment were presented in six selected storms. The area of catchment is about 241.34 km^2. The areas between all pairs of neighbouring isohyets of the six selected storms were computed by ArcGIS 9.1. These rainfall contours also derived by Kriging Method as stated in section 3.6.2.3. After all of the area of every colour code were calculated, mean area precipitation (MAP) were computed. MAP is the mean areas between all pairs of neighbouring isohyets. Then, the percentage reduction of storm depth is determined and lastly, areal reduction curves for all storms were plotted. All calculations to produce areal reduction curves were shown in Appendix C.

3.1.7 Limitations in Analysing Convective Rainfalls

The above sections have described the research methodologies for analyzing convective rains. The data used in this analysis has some limitation as follows.

(a) Some rainfall stations in Klang Valley are no longer in operation and some stations have missing data. This limit the numbers of rainfall stations used in this study.

(b) Although a number of flash flood events occurred between year 2001 and 2006, complete sets of rainfall data for both surface rainfall and radar rainfall are not always available.

(c) Due to the small numbers of rainfall stations, rainfall contours derived by Kriging Method cannot give a smooth contour. This is because Kriging works best with large input data and prediction errors are larger in areas with small number of samples.

3.3 Stochastic Modeling of Rainfall Series using Neyman-Scott Rectangular Pulses Model (NSRP)

This study emphasizes on the single-site rainfall modeling for data collected on short time scale, that is hourly, both for describing adequately the high variability of these rainfall processes and for providing a basis for simulating rainfall processes for longer time scales. The progress made in this area is crucial to the generalization of the approach to temporal rainfall modeling. The description of the model mathematical structure will be presented in this section. The presentation will consider many important features of temporal rainfall processes such as the structure of the rainfall depth, duration, intensity and occurrence. Some improvements that are proposed in the present study will be derived in this chapter.

3.2.1 Determining the Best-fit Distribution for the Hourly Rainfall Series

According to WMO, a wet day is defined as a day with a rainfall amount above a fixed threshold of 0.1 mm. This threshold will be chosen in this study with amount of greater than or equal to 0.1mm to be identified as wet hours. The sequence of rainfall amounts on wet hours is also considered as the intensity process (Katz and Parlange, 1995).

3.3.2.1 Types of Distribution

In this study, the distribution of hourly rainfall amounts is described by four functions the Exponential, Gamma, Weibull, and Mixed Exponential distribution.

The probability density functions along with the log likelihood functions are as follows:

a. The Exponential distribution with parameter λ represents mean while x represents the hourly rainfall amounts.

$$f(x) = \frac{1}{\lambda} e^{\frac{-x_i}{\lambda}}, \qquad x, \lambda > 0 \tag{3.3}$$

b. The Weibull distribution with two-parameters, namely α and β to represent shape and scale parameters respectively while x represents the hourly rainfall amounts.

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x_i}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x_i}{\beta}\right)^{\alpha}}, \qquad \alpha > 0, \beta > 0, x > 0 \tag{3.4}$$

c. The Gamma distribution with two-parameters, namely α and β to represent shape and scale parameters respectively while x represents the hourly rainfall amounts.

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x_i^{\alpha-1} e^{-\frac{x_i}{\beta}}, \qquad \alpha > 0, \beta > 0, x > 0 \qquad (3.5)$$

d. The Mixed-Exponential distribution is a weighted average of two one-parameter exponential distributions. The mixture distribution has three parameters, with α representing the mixing probability and β_1 and β_2 representing the scale parameters, while x representing the hourly rainfall amounts.

$$f(x) = \left(\frac{\alpha}{\xi}\right) e^{\frac{-x}{\xi}} + \left(\frac{1-\alpha}{\theta}\right) e^{\frac{-x}{\theta}} \qquad (3.6)$$
$$x > 0, 0 \leq \alpha \leq 1, 0 < \xi < \theta$$

3.3.2.2 Parameter Estimation Methods

The maximum likelihood method that is claimed to being a minimum variance unbiased estimator is used in estimating the parameters of the distributions. However, the method of moments is still being used to set up the initial points of the maximum likelihood method. In the maximum likelihood estimation, it is assumed that X_i's are independent and identically distributed where i=1, 2, …, n. The function $f(X_i|\theta_1, \ldots, \theta_m)$ is the conditional density function of the observations X_i given the parameters $\theta_1, \ldots, \theta_m$. When the X_i's are independent, the joint density function of X_i is the product of the marginal densities.

The parameters $\theta_1, \ldots, \theta_m$ are estimated by maximizing the following likelihood function:

$$L(\theta_1, \ldots, \theta_m) = \prod_{i=1}^{n} f(X_i \mid \theta_1, \ldots, \theta_m) \qquad (3.7)$$

The parameters are determined by taking the partial derivative of $L(\theta_1,\ldots,\theta_m)$ with respect to each parameter setting the resulting equations to zero. These m partial derivative equations are solved for the m unknown parameters. In order to get the unknown parameters, it is easier to maximize the natural logarithm of the likelihood function because most of probability distributions involve the exponential function.

a. Exponential :

The first-order moment about the origin is

$$M_1 = E(X) = \lambda \tag{3.8}$$

The corresponding sample moment is

$$\text{Mean } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.9}$$

The estimate of the parameter $\hat{\lambda}$ is $\bar{x}$.

The log-likelihood is

$$\log L = \sum_{i=1}^{n} \log\left[\frac{1}{\lambda} e^{\frac{-x_i}{\lambda}}\right] \tag{3.10}$$

b. Gamma:

The first two moments about the origin is

$$M_1 = E(X) = \beta / \lambda \tag{3.11}$$

$$M_2 = E(X^2) = \frac{\beta(\beta+1)}{\lambda^2} \tag{3.12}$$

The corresponding sample moments are

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.13}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2 \tag{3.14}$$

Equating the population and sample moments, the parameters are

$$\hat{\lambda} = \frac{M_1}{M_2 - M_1^{\,2}} = \frac{\overline{X}}{s^2} \tag{3.15}$$

$$\hat{\beta} = \hat{\lambda} M_1 = \frac{M_1^{\,2}}{M_2 - M_1^{\,2}} = \frac{\overline{X}^2}{s^2} \tag{3.16}$$

The log-likelihood is:

$$\log L = \sum_{i=1}^{n} \log\left[\frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}\right] \tag{3.17}$$

c. Weibull

The first two moments about the origin is

$$M_1 = E(X) = \alpha\Gamma\left(\frac{1}{\gamma} + 1\right) \tag{3.18}$$

$$M_2 = E(X^2) = \alpha^2\Gamma\left(\frac{2}{\gamma} + 1\right) \tag{3.19}$$

The above equations are nonlinear and cannot be solved directly. The coefficient of variation (COV) and the shape parameter γ is used to estimate the parameters (Cohen, 1965):

$$COV = \frac{(Variance)^{\frac{1}{2}}}{Mean} = \frac{(M_2 - M_1^{\,2})^{1/2}}{M_1} = \frac{[\Gamma(\frac{2}{\gamma}+1) - \Gamma^2(\frac{1}{\gamma}+1)]^{1/2}}{\Gamma(\frac{1}{\gamma}+1)} \tag{3.20}$$

The log-likelihood is

$$\log L = \sum_{i}^{n} \log\left[\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}\right] \tag{3.21}$$

d. Mixed Exponential

The first three moments about the origin are

$$M_1 = E(X) = \alpha\xi + (1-\alpha)\xi \tag{3.22}$$

$$M_2 = E(X^2) = 2\alpha\xi^2 + 2(1-\alpha)\theta^2 \tag{3.23}$$

$$M_3 = E(X^3) = 6\alpha\xi^3 + 6(1-\alpha)\theta^3 \tag{3.24}$$

The corresponding sample moments are

$$K_1 = \hat{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3.25}$$

$$K_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^{\,2} \tag{3.26}$$

$$K_3 = \frac{1}{n}\sum_{i=1}^{n} x_i^{\,3} \tag{3.27}$$

The log-likelihood is

$$\log L = \sum_{i=1}^{n} \log\left[\left(\frac{\alpha}{\xi}\right)e^{\frac{-x_i}{\xi}} + \left(\frac{1-\alpha}{\theta}\right)e^{\frac{-x_i}{\theta}}\right] \tag{3.28}$$

3.2.1.3 Goodness of fit tests

In determining the best-fit distributions five quantitative methods are used in this study.

a. The mean and median absolute difference between the hypothesized distribution $F(x)$ and the empirical distribution, $F_n(x)$.

$$\text{Mean} = \frac{\sum_{i=1}^{n} \left| F_n(x_i) - F(x_i, \hat{\theta}) \right|}{n} \tag{3.29}$$

$$\text{Median}\left| F_n(x_i) - F(x_i, \hat{\theta}) \right|$$

b. Kalmogorov-Smirnov (KS) test calculates the maximum difference between the hypothesized distribution and empirical distribution.

$$D^{+} = \max_i \{1/n - Z_i\}, \qquad D^{-} = \max\{Z_i - (i-1)/n\}$$
$$KS = \max(D^{+}, D^{-}). \tag{3.30}$$

c. Cramer-Von-Mises (CVM) calculates the squared difference between $F(x)$ and $F_n(x)$.

$$W^2 = \sum_{i=1}^{n} \{Z_i - (2i-1)/2n\}^2 + \frac{1}{12n} \tag{3.31}$$

d. Anderson-Darling (AD) test calculates the squared difference between $F(x)$ and $F_n(x)$, and divided them by the weight function $\left[F(x)\left(1-F(x)\right)\right]^{-1}$.

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}\left[(2i-1)\log Z_i + (2n+1-2i)\log(1-Z_i)\right] \tag{3.32}$$

e. Akaike Information Criterion (AIC) is derived by minimizing the Kullback Leibler distance between the proposed model and true one. The best model is the one having the smallest AIC. The AIC is given by

$$\begin{aligned} \text{AIC} &= -2 \log (\text{maximum likelihood}) + 2\,k \\ &= -2\text{MLL} + 2k \end{aligned} \tag{3.33}$$

in which k denotes the number of parameters.

3.2.2.4 Exceedance Probability

The exceedance probability is defined as the probability of a rainfall amount occurring greater than that of a given amount. Example, the probability of rainfall exceeding a low amount (< 1 mm) would be high, while the probability of rainfall exceeding above 100 mm is a more unlikely event. This probability is plotted on a semi-log scale and it is a qualitative tool to assess the performance of distributions considered. The horizontal axis represents the wet hours amount and the vertical axis represents the

[1- $F(x)$] and [1- $F_n(x)$] where $F(x)$ represents the hypothesized distribution and the $F_n(x)$ represents the empirical distribution. This plot will display every wet hours data distinctly.

3.2.3 The Neyman-Scott Rectangular Pulses model (NSRP)

The theoretical basis of stochastic point processes is needed in order to understand the Neyman-Scott model properly. This will focus on special processes of potential importance in applications related to rainfall.

3.2.2.3 Theory of Point Processes

The use of point process theory have received widespread attention by scientists for the development of realistic rainfall models (e.g.,Rodriguez-Iturbe et al.1987a,b; Entekhabi et al. 1989; Islam et al.1990; Cowpertwait, 1991; Onof and Wheater,1993; Onof et al. 1994; Velghe et al. 1994; Cowpertwait et al. 1996a,b; Khaliq and Cunnane, 1996; Cowpertwait, 1998,2002,2004) since the pioneering work of Kavvas and Delleur (1981); Waymire and Gupta (1981a,b,c); Rodriguez and Iturbe et al. (1984); Waymire et al. (1984); Smith and Karr (1985b,a); Valdes et al. (1985). A point process is a model of points randomly distributed in some space E. The points may represent times of events, locations of objects or paths followed by a stochastic system. One example of a point process event is the emission of radioactive from a source that occurs in an irregular sequence in time. Each emission defines a time distant. When a point process is defined, it is often of interest to count the numbers of points in subsets of the space E. Let assume a realization T of a random point process on E is a denumerable point set of E. This mean that T can be enumerated as $T = \{t_1, t_2, ...\}$ where each t_i denotes the coordinate of a point in E. Let A be a subset of E .Then

$$N(A,T) = \sum_i I_R(t_i), \tag{3.34}$$

is the number of points that lie in A and each t_i is the coordinate of a point in T. $I_A(t)$ denotes the set characteristic function of A and is defined as follows:

$$I_R(q) = \begin{cases} 1 & \text{if } q \in R \\ 0 & \text{if } q \notin R \end{cases} \tag{3.35}$$

$N(A,T)$ defines a non-negative, integer-valued random process on E. This process is called a counting process.

3.2.2.4 The Poisson Process

Consider the process as defined over the whole time axis $(-\infty, \infty)$. Let H_t denotes the history of the process at time t, i.e. a specification of the positions of all points in $(-\infty, t]$. For $u < v$, let $N(u,v)$ be a random variable giving the number of points in $(u,v]$. Then for a given constant ρ with dimensions [time]$^{-1}$, the Poisson process of rate ρ is defined by the requirements that for all t, as $\delta \to 0+$,

$$P\{N(t, t+\delta) | \mathrm{H}_t\} = \lambda\delta + o(\delta), \tag{3.36}$$

$$P\{N(t, t+\delta) | \mathrm{H}_t\} = o(\delta) \tag{3.37}$$

so that

$$P\{N(t,t+\delta)=0|\mathrm{H}_t\}=1-\lambda\delta+o(\delta). \qquad (3.38)$$

From Eqs.(3.36) to (3.38), the probabilities concerned do not depend on H_t. It follows that the probability of finding a point in $(t,t+\delta]$ does not depend on the number of points occurring just before t. In fact, the expression of Eq.(3.37) excludes the possibility of multiple simultaneous occurrences. However, there are two important results that can be deduced from the above specifications of the process.

i. Consider the points $0<T_1<T_2<T_3<......$ building a Poisson process of a constant rate λ. The random variables $X_1=T_1,.......T_n=T_n-T_{n-1}(n\geq 2)$ are independent and each has probability density function of $f_x(.)=\lambda\exp(-\lambda x)$. This property provides the interval specification of the process.
ii. Consider the number of events $N(a_i,b_i)$ of the process that falls in $a_i<b_i\leq a_{i+1}$. The Poisson process on the line is completely defined by the following equation.

$$P\{N(a_i,b_i)=n_i,i=1,......,k\}=\prod_{i=1}^{k}\frac{(\lambda(b_i-a_i))^{ni}}{ni!}\exp(-\lambda(b_i-a_i)) \qquad (3.39)$$

This counting specification includes three important features: the number of points in each finite interval $[a_i,b_i]$ has a Poisson distribution ; the number of points in disjoint intervals are independent random variables; the distributions are stationary and are dependent upon the respective lengths b_i-a_i of the intervals.

Cox and Isham (1980) defined the above Poisson process with three mutually equivalent specifications: the intensity specification Eqs.(3.36-3.37), the interval specification, and the counting specification. The interplay between the three specifications of the Poisson process is a recurring theme in the study of point processes. Note that the intensity specification can be used for building a realization of a Poisson process while the interval specification gives an efficient basis for such a construction.

3.2.2.3 Some basic definitions

The complete intensity function is an important characteristic of point processes. It is defined as

$$\rho(t;\mathrm{H}_t) = \lim_{\delta\to 0^+} \delta^{-1} P\{N(t,t+\delta) > 0 \,|\, \mathrm{H}_t\}. \qquad (3.40)$$

where H_t specifies the point process up to and including t. For the Poisson process, the complete intensity function is equal to $\rho(t;H_t)=\lambda$. The probability of a point in $[t,t+\delta]$ given the fact that there is a point at the origin is specified by the conditional intensity function

$$h(t) = \lim_{\delta_1,\delta_2\to 0^+} \delta_2^{-1} P\{N(t,t+\delta_2) > 0 \,|\, N(-\delta_1,0) > 0\} \qquad (3.41)$$

The conditional intensity function will be used to derive the covariance of the counting process later.

Stationarity and orderliness are another two important properties in point processes. The intuitive notion of stationarity means that the distribution of the number of points lying in an interval depends on its length but not on its location; that is

$$P\{N(t,t+x)=k\} \qquad (x>0, k=0,1,2,.....) \qquad (3.42)$$

depends on the length x but not on location t.

The following definitions explain the characteristics in stationarity.

Definition 1

A point process is stationary when for every r = 1,2,….., and all bounded Borel subsets $A_1, A_2,\ldots\ldots A_r$ of the real line of the joint distribution of

$$\{N(A_1+t),......., N(A_r+t)\} \qquad (3.43)$$

does not depend on t $(-\infty < t < \infty)$.

Definition 2

A point process is interval stationary when for every r = 1,2,…, and all integers $i_1, \ldots, i_r$ the joint distribution of $\{\tau_{i1} + k, \ldots\ldots.\tau_{ir} + k\}$ does not depend on k, (k = 0, ±1,….).

The non-existence of a multiple simultaneous occurrences in a process is called orderliness, that is

$$P\{N(t) > 1 \quad \text{for some } t \in \mathbb{R}\} = 0, \tag{3.44}$$

It can be shown that (3.36) implies (3.43) and for most point processes they are in fact equivalent.

The probability generating functional (pgf) is a generalization of the probability generating functions that provides a complete description of the random variable and is useful in the calculation of moments. $G_N(.)$ is defined by (see Cox and Isham (1980, eq.(2.42))

$$G[\xi] = E\left[\exp\left(\int_{\mathbb{R}} \log \xi(t) dN(t)\right)\right] = E\left[\prod_{i=1}^{n} \xi(t_i)\right] \tag{3.45}$$

where $\{t_i\}$ are the random co-ordinates of the points. The two forms of $G[\xi]$ are equivalent because N is a step function. In Eq.(3.43) the product is unity if n = 0, and zero if n > 0 and $\xi(t_i) = 0$ for any i. In order for the expectation to exist, $0 \le \xi(t) \le 1$ is required to be imposed.

A more intuitive approach for the probability generating functional (pgf) is obtained by taking $A_1, A_2, \ldots\ldots, A_r$ to be a measurable partition of E and setting:

$$\xi(x) = \sum_{i=1}^{r} z_i I_{A_i}(x), \tag{3.46}$$

where $I_A(x)$ is the indicator function of the set A and $|z_i| \le 1$ for i = 1,……,r. Substitution in (3.43) leads to

$$G\left[\sum_{i=1}^{r} z_i I_{A_i}(.)\right] = E\left[\prod_{i=1}^{r} z_i^{N(A_i)}\right] \tag{3.47}$$

that is the multivariate probability generating function of the number of points in the sets of the given partition.

An example is given by Cox and Isham (1980) with the probability generating functions for non-homogenous Poisson process with rate function $\lambda(t)$,

$$G[\xi] = \exp\left(-\int_{i} (1-\xi(t))\lambda(t)dt\right), \tag{3.48}$$

which is equal to the probability generating function for a Poisson variable with parameter λ, given by $G(z) = \exp(-\lambda(1-z))$.

Superposition of process is concerned with two or more independent processes that are superposed in term of summation. Let say there are two independent processes, namely N_1 and N_2. and $N(A) = N_1(A) + N_2(A)$ for all sets A. The resulting generating functional satisfy the relation $G_N[\xi] = G_{N_1}[\xi]G_{N_2}[\xi]$. This is in fact a useful property of probability generating function and is shared by the probability generating functional as well..

3.2.2.5 Moments

In this section the theory related to the moments of the counting process which will be used to derive the cross-covariance of the rainfall process is presented.
Consider the first two moments of the counting process in the arbitrary sets A and B:

$$E[N(A)], V[N(A)], Cov[N(A), N(B)] \tag{3.49}$$

For stationary orderly processes of finite and fixed rate λ,

$$E[N(A)] = \lambda|A|, \tag{3.50}$$

where $|A|$ is the Lebesgue measure of the set A. Considering the covariance for the counting process for two disjoint sets A and B, we have

$$2Cov\left[N(A),N(B)\right]=Var\left[N(A\cup B)\right]-Var\left[N(A)\right]-Var\left[N(B)\right] \tag{3.51}$$

This is the simplest case of a point process on a line. Lets consider A has the interval of $[0,t]$.

$$N(t)=\int_0^t dN(z). \tag{3.52}$$

Applying the above formula,

$$Var\left[N(t)\right]=\int_0^t Var\left[dN(z)\right]+2\iint_{\substack{0<z<t\\0<u\leq t-z}} Cov\left[dN(z),dN(z+u)\right], \tag{3.53}$$

where the integral is to be considered as the limit of a sum.

The definition of orderliness implies that $N(z,z+\delta)$ can take only the values of zero and one. Hence, for an orderly manner, we get (Cox and Isham, 1980)

$$\begin{aligned}Var\left[N(z,z+\delta)\right]&=E\left[N(z,z+\delta)^2\right]-\left(E\left[N(z,z+\delta)\right]\right)^2\\&=P\left\{N(z,z+\delta)=1\right\}-\left(P\left\{N(z,z+\delta)=1\right\}\right)^2+o(\delta)\\&=\lambda\delta+o(\delta),\end{aligned} \tag{3.54}$$

$$\begin{aligned}&u>0,\\&Cov\left[N(z,z+\delta_1),N(z+u,z+u+\delta_2)\right]\\&=E\left[\ \right]E\left[N(z+u,z+u+\delta_2)\middle|N(z,z+\delta_1)\right]-\\&E\left[N(z,z+\delta_1)\right]E\left[N(z+u,z+u+\delta_2)\right]\\&=P\left\{N(z,z+\delta_1)=1\right\}P\left\{N(z+u,z+u+\delta_2)=1\middle|N(z,z+\delta_1)=1\right\}\\&-P\left\{N(z,z+\delta_1)=1\right\}P\left\{N(z+u,z+u+\delta_2)=1\right\}+o(\delta_1\delta_2)\\&=\lambda h(u)\delta_1\delta_2-\lambda_2\delta_1\delta_2+\delta(\delta_1\delta_2),\end{aligned} \tag{3.55}$$

where $h(.)$ is the conditional intensity function. Merging the above in the limit as δ_1 and δ_2 move simultaneously to zero, we can now evaluate

$$\begin{aligned} Var\left[N(t)\right] &= \int_0^t \lambda dz + 2\int_0^t dz \int_0^{t-z} (\lambda h(u) - \lambda^2) du \\ &= \lambda t + 2\lambda \int_0^t (t-u) h(u) du - \lambda^2 t^2. \end{aligned} \qquad (3.56)$$

Hence, the variance can be written as follows

$$Var\left[N(t)\right] = \int_0^t dz \int_0^t c(u-z) du \qquad (3.57)$$

where

$c(u) = \lambda\delta(u) + \lambda h(u) - \lambda^2, \qquad u \geq 0,$

with $\delta(.)$ being the Dirac delta function.

3.2.2.5 Cluster Processes

Kavvas and Delleur (1975,1981), Kavvas (1982a,b), Gupta and Waymire (1979), and Waymire and Gupta (1981a,b,c) have popularized the use of cluster models. Amorocho and Wu (1977) and Burlando (1989) suggested that cluster models are able to simulate the cellular structure of actual precipitation fields and able to preserve theoretically at least the relevant statistics on a wide range of temporal aggregation scales. Shaw (1983) found that cluster models were more appealing for rainfall time series simulation as they are able to preserve rainfall statistics over a range of time scales, and they have built into their structure the capability of representing rain cells, which are known to exist in actual rainfall events. Hence, the discussion of what is cluster process is discussed as follows.

The general structure of cluster processes involves the existence of a point process of cluster centers. Each cluster center is associated with a random number of points forming a subsidiary process or cluster. These subsidiary points are being distributed about the cluster center in some specified ways. The cluster process then consists of the superposition of all the separate clusters, points belonging to the same cluster are not being identified as such.

Let say N_c denotes the counts connected with the process of cluster centers. $N_{(t)}(A)$ is the number of subsidiary points in A arising from a cluster given to have center at t. The total number of points $N(A)$ in A is

$$E\{N(A)\} = \int E\{N_{(t)}(A)\}E\{dN_c(t)\} \tag{3.58}$$

Suppose that the cluster centers have occurred at points t_i. The independence of the separate clusters implies that the conditional probability generating functional of the point process is

$$\prod_{i=-\infty}^{\infty} E\left[\exp\left(\int \log \xi(t) dN_{(t_i)}(t)\right)\right] = \prod_{i=-\infty}^{\infty} G_s[\xi; t_i], \tag{3.59}$$

where $G_s[\xi; t_i]$ is the probability generating functional for a cluster centered at t_i . An immediate consequence obtained by taking the expectation of Eq.(3.57) is

$$G[\xi] = G_c\left[G_s[\xi;.]\right] \tag{3.60}$$

where $G_c[\xi] = E\left[\prod_i \xi(T_i)\right]$ refers to the process of cluster centers.

The cluster process based on the Poisson process is the most frequently used and the simplest is obtained by treating the Poisson point as sites and locating at each site a random number of points. Neyman-Scott process is one of an example of the cluster-based Poisson process and it is sometimes called the center-satellite process (Neyman and Scott, 1958).This process uses Poisson points as centers or parents. At each center, independent of other centers, a random number of satellites are generated. The number of satellites per center is given by independent, identically distributed non-negative random variables. Each satellite is displaced from the center according to some dispersal distribution. Hence, in the Neyman-Scott process the points in a cluster are independently and identically distributed around the cluster. Besides Neyman-Scott process there is another process called the Bartlet-Lewis process. In this process, at each center point a renewal process generates satellites. In relation to that, in the Bartlet-Lewis process, the intervals between successive points in a cluster are independently and identically distributed (idd).

The second-order counting properties of the Neyman-Scott and the Bartlett-Lewis can be derived with the conditional intensity function h(.). Cox and Isham (1980) derived this function for the Neyman-Scott process in considering two cases depending on the position of two points (they either belong to the same cluster or not):

$$h(u) = E[C]\lambda_c + \frac{E[C(C-1)]}{E[C]}\int_{-\infty}^{\infty} f(x)f(x+u)dx \tag{3.61}$$

where λ_C is the rate of the Poisson process of the cluster centers.

The probability generating functional for both processes can be obtained from Eq. (3.48) and Eq. (3.61) as

$$G[\xi] = \exp\left(-\lambda_C \int_{-\infty}^{\infty} (1 - G_S[\xi;t])dt\right)$$

The generating functional for a cluster with center at *t* may be expressed for the Neyman-Scott process as (see Cox and Isham (1980, eq. (3.58))

$$
\begin{aligned}
G_s[\xi;t] &= E\left[\exp\int_{-\infty}^{\infty}\log\xi(u)dN_{(t)}(u)\right] \\
&= \sum_{m=0}^{\infty} g_m \int_{-\infty}^{\infty} f(u_1)\xi(t+u_1)du_1 \ldots\ldots \int_{-\infty}^{\infty} f(u_m)\xi(t+u_m)du_m \qquad (3.62) \\
&= G_M\left(\int_{-\infty}^{\infty}\xi(t+u)f(u)du\right).
\end{aligned}
$$

3.2.4.6 Description of the Neyman-Scott Rectangular Pulses Model (NSRP)

The first proposed rainfall modeling scheme, referred hereafter as the Neyman-Scott Rectangular Pulse (NSRP) model is a clustered point process model. This model is used in modeling the rainfall event where in any event there exists a generating mechanism called the storm origin. The storm origin may be passing fronts or some other criteria for convection storms from which rain cells arise. The Neyman-Scott models are described by 3 independent elementary stochastic processes: They are

- A process that sets the origin of the events;
- A process that sets the number of rain cells generated by each event;
- A process that sets the origin of the cells.

Storm origins are governed by a Poisson process with parameter λ. At a point on the ground the storm is conceptualized as a random number C of rain cells. Natural candidates for the distribution of the number of cells C are the Poisson distribution and the geometric distribution. The cell origins are independently separated from the storm origin by distances that are exponentially distributed with parameter b. It is assumed that there is no cell origins being located at the storm origin. A rectangular pulse is associated independently with each cell origin with its duration and intensity (depth) being independent. The duration is assumed to be exponentially distributed with parameter h. The intensity is assumed to be exponentially distributed with parameter $\frac{1}{m_X} = 1/x$. In summary, the NSRP model reproduces the characteristics of intermittency, persistency and periodicity of the rainfall series.

3.2.2.7 Mathematical Representation of the NSRP model

The precipitation intensity at time *t, Y*(*t*), is given by the sum of the intensities of the individual active cells at time t:

$$Y(t) = \int_{u=0}^{\infty} X_{t-u}(u)\,dN(t-u) \tag{3.63}$$

where $X_u(k)$ is the random depth of the pulse originating at time u measured a time k later and {*N*(*t*)} counts occurrences in the Poisson process of pulse origins. Note that the intensity of N(*t*) is $\lambda \mu_c$, where μ_c denotes the mean number *E*[*C*] of cells per storm.

The derivative of the counting process is

$$dN(t-u) = \begin{cases} 1 & \text{if there is a cell origin at } t-u \\ 0 & \text{otherwise} \end{cases} \tag{3.64}$$

and for the rectangular pulses, we have

$$X_{t-u}(u) = \begin{cases} X & \text{with probability } R(x) \\ 0 & \text{with probability } 1-R(x) \end{cases} \tag{3.65}$$ 80

where $X_{t-u}(u)$ is the intensity of the rectangular pulse triggered at time u and *N*(*t*) represents the counting stochastic process of the arrivals of the individual cells. *R*(*x*) is the survival function of *X*.

The moments of the counting process *N*(*t*) have been obtained by Waymire and Gupta (1981c) by derivation of the probability generating functional of the Neyman-Scott process defined in Eq.(3.61). The second order properties of *Y*(*t*) can be derived in various ways, most simply through (3.61). This method has been used by Rodriquez-Iturbe et al. (1987a). The mean of the depth process can be represented directly as the product of the rate at which cell origins occur, the mean length of a cell and the mean depth of a cell, that is,

$$E[Y(t)] = \frac{\lambda}{\eta}\mu_c\mu_x \tag{3.66}$$

The variance and auto covariance at lag-τ have been expressed in terms of the conditional intensity function $h(.)$ of the Neyman-Scott process defined in Equations (3.60) which leads to the following expressions (Rodriguez-Iturbe et al., 1987a)

$$Var[Y(t)] = \frac{\lambda}{\eta}\mu_c E[X^2] + \frac{\lambda\beta\mu_x^2 E[C^2 - C]}{2\eta(\beta+\eta)}$$

$$\begin{aligned} c_Y(\tau) &= Cov[Y(t), Y(t+\tau)] \\ &= \frac{\lambda}{\eta}e^{-\eta\tau}\left(\mu_c E[X^2] + \frac{\beta^2\mu_X^2 E[C^2 - C]}{2(\beta^2 - \eta^2)}\right) - \frac{\lambda\beta\mu_X^2 e^{-\beta\tau}E[C^2 - C]}{2(\beta_2 - \eta_2)} \end{aligned} \tag{3.67}$$

Since rainfall data are usually available only as rainfall depths in discrete time intervals (e.g. historical records of hourly or daily totals), the aggregated properties are needed to estimate the parameters of the model. The aggregated process at time scale h (the total depth in a time interval h) is given by:

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} Y(t)dt \tag{3.68}$$

The second-order properties of the aggregated process (Rodriguez-Iturbe et al., 1984) are

81

$$\begin{aligned} &E[Y_i^h] = hE[Y(t)], \\ &Var[Y_i^h] = 2\int_0^h (h-u)c_Y(u)du, \\ &Cov[Y_i^{(h)}, Y_{i+k}^{(h)}] = \int_{-h}^{h} c_Y(kh+v)(h-|v|)dv, \end{aligned} \tag{3.69}$$

Thus, if h is measured in hours, the series $\{Y_i^h : i = 1,2,....\}$ is a rainfall time series at the h-hour level of aggregation, i.e. an h-hourly rainfall time series. The second-order properties of the aggregated process [Rodriguez-Iturbe et al., 1987a] are

Mean:

$$E\{Y_i^{(h)}\} = h\lambda E\{C\}E\{X\}/\eta \tag{3.70}$$

Variance:

$$\begin{aligned} Var\{Y_i^{(h)}\} &= \lambda\eta^{-3}(\eta h - 1 + e^{-\eta h})[2\mu_c E\{X^2\} + E\{C^2 - C\}\mu_x^2\beta^2/(\beta^2-\eta^2)] \\ &- \lambda(\beta h - 1 + e^{-\beta h})E\{C^2 - C\}\mu_x^2\beta^{-1}/(\beta^2-\eta^2) \end{aligned} \tag{3.71}$$

Covariance:

$$\begin{aligned} Cov\{Y_i^{(h)}Y_{i+k}^{(h)}\} &= \lambda\eta^{-3}(1-e^{-\eta h})^2 e^{-\eta(k-1)h} \\ &\times[\mu_c E\{X^2\} + \frac{1}{2}E\{C^2 - C\}\mu_x^2\beta^2/(\beta^2-\eta^2)] \\ &- \lambda(1-e^{-\beta h})^2 e^{-\beta(k-1)h}E\{C^2 - C\}\mu_x^2/[2\beta(\beta^2-\eta^2)] \end{aligned} \tag{3.72}$$

From now on, $E\{Y_i^{(h)}\}$, $Var\{Y_i^{(h)}\}, Cov\{Y_i^{(h)}, Y_{i+k}^{(h)}\}$ will be denoted as $\hat{\mu}(h)$, $\hat{\gamma}(h)$, $\hat{\gamma}(h,k)$ respectively for convenience. The lag k autocorrelation function $\hat{\rho}(h,k)$ is given by $\hat{\gamma}(h,k)/\hat{\gamma}(h)$.

3.2.2.8 The choice of distributions for the rain cells numbers, *C* and the rain cell intensities, *X*.

For the model to be completely defined distributions need to be chosen for C and X. Natural candidates for C are the Poisson distribution and the geometric distribution. Velghe et al. (1994) found that geometric N-S performed better than the Poisson N-S with regards to its ability to reproduce several properties of rainfall, but the result may not be representative since it was only applied to one station.

In this study the Poisson distribution is chosen to represent the distribution for C. Following Velghe et al. (1994), the derivations of $E(C)$ and $E(C^2\text{-}C)$ are as follows:

Assume that C is strictly positive, then C-1 is said to have a Poisson distribution. Let y be a Poisson distribution with probability density function $f(y)$ given by

$$f(y) = \frac{e^{-\beta}\beta^{y}}{y!} \qquad y = 0,1,2,..... \tag{3.73}$$

The mean E(y) and variance V(y) of the above distribution are both β.

Since C is strictly positive, then

$$C = 1 + y \tag{3.74}$$

the expected value for C and C^2 is given as follows:

$$E[C] = \mu_c = 1 + \beta \tag{3.75}$$

$$E[C^2] = E[(1+y)^2] = 1 + 2\beta + E[y^2] \tag{3.76}$$

From the above

$$E[y^2] = Var[y] + E[y]^2 = \beta + \beta^2 \tag{3.77}$$

$$E[C^2] = 1 + 3\beta + \beta^2 \tag{3.78a}$$

$$E[C^2 - C] = \beta^2 + 2\beta = \mu_c^2 - 1 \tag{3.78b}$$

Therefore from (14), if we let $\mu_c = \nu$,then

$$E[C-1] = \nu - 1 \tag{3.79a}$$

$$E[C] = \mu_c = \nu \tag{3.79b}$$

$$E[C^2 - C] = \nu^2 - 1 \tag{3.79c}$$

The rain cell intensity X in the model is following the exponential distribution. The cumulative distribution function is,

$$F(x) = P\{X \le x\} = 1 - P\{X > x\} = \begin{cases} 0 & \text{if } \; x < 0, \\ 1 - e^{-\xi x} & \text{if } \; x \ge 0. \end{cases} \tag{3.80}$$

and the probability distribution function is,

$$f(x) = \begin{cases} 0 & \text{if } \; x < 0, \\ \xi e^{-\xi x} & \text{if } \; x > 0. \end{cases} \tag{3.81}$$

The expected values for X and X^2 are

$E[X] = \frac{1}{\xi}$ and $E[X^2] = \frac{2}{\xi^2}$ respectively.

Therefore in Eqs.(3.68) to (3.70) the followings have to be substituted in order for the mean and the second order properties are to be defined:

$\mu_c \equiv E(C) = \nu; E(C^2 - C) = \nu^2 - 1; \mu_x \equiv E(X) = \xi^{-2}; E(X^2) = 2\xi^{-2}$.

With the above properties, the NSRP model has five parameters

$$\Theta_{\exp} = (\lambda, \nu, \beta, \eta, \xi). \tag{3.82}$$

The parameters λ represents the storm origin, ν represents the number of cells, β represents the position of cells, η represents the duration of cells and ξ represents the intensity of the rain cells. Therefore, the model structure is based on the following assumptions. The diagram in Figure 3.11 explains the following structure in details.

i. The inter-arrival time of the storm origin follows the exponential distribution:

$$P_{L_n}(l_n) = 1 - e^{-\lambda l_n} \tag{3.83}$$

ii. The number of rain cells is described by Poisson distribution:

$$P(C) = \frac{(\nu t)^C e^{-\nu t}}{C!} \tag{3.84}$$

iii. The waiting times from the origin to the rain cells origin is described by exponential distribution:

$$P_{B_m}(b_m) = 1 - e^{-\beta b_m} \tag{3.85}$$

iv. The duration of the rain cells is also described by exponential distribution function:

$$P_T(t) = 1 - e^{-\eta t} \tag{3.86}$$

v. The intensities are described by exponential distribution:

$$P_X(x) = 1 - e^{-\xi x} \tag{3.87}$$

3.2.2.9 The proposed distribution for the rain cell intensities, *X*.

The choice of the distributions to represent the rain cell intensities in the NSRP model is arbitrary. The exponential distribution was selected so that the model would have only a small number of parameter. However, a heavier-tailed distribution could be used to model the cell intensity to improve the fit to the historical extreme values. An obvious alternative to the exponential distribution which could be used to improve the fit to the extremes is the Weibull distribution (Cowpertwait,1996,2002) or Gamma distribution (e.g. Onof and Wheater ,1993,1994). Whether such distributions are needed would depend on the intended application for the model.

In this study the mixed exponential distribution is proposed to represent the rain cell intensities. This distribution is chosen following the results obtained for the fitting of the hourly amount using the goodness of fit test. It was found that the mixed exponential distribution was the best among the other candidate distributions namely exponential, Gamma and Weibull in describing the hourly rainfall amount used in this study (Fadhilah et. al. 2007).

The probability distribution function for the mixed exponential distribution is given as:

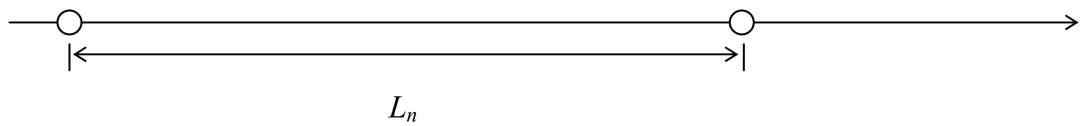
$$f(x) = \frac{\alpha}{\xi} e^{\left(\frac{-x}{\xi}\right)} + \frac{(1-\alpha)}{\theta} e^{\left(\frac{-x}{\theta}\right)} \tag{3.88}$$

$$x > 0; 0 \le \alpha \le 1;), 0 < \xi < \theta$$

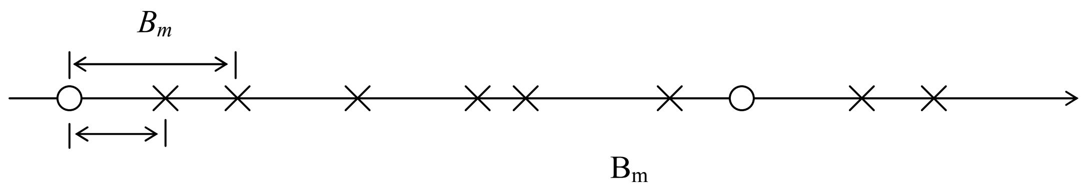
The mixed-exponential distribution is a weighted average of two one-parameter exponential distributions. The mixture distribution has three parameters, with α representing the mixing probability, ξ and θ representing the scale parameters and x representing the hourly rainfall amounts per hour. The distribution function *F(x)* is given as:

$$F(x) = \alpha e^{\frac{-x}{\xi}} + (1-\alpha) e^{\frac{-x}{\theta}} \tag{3.89}$$

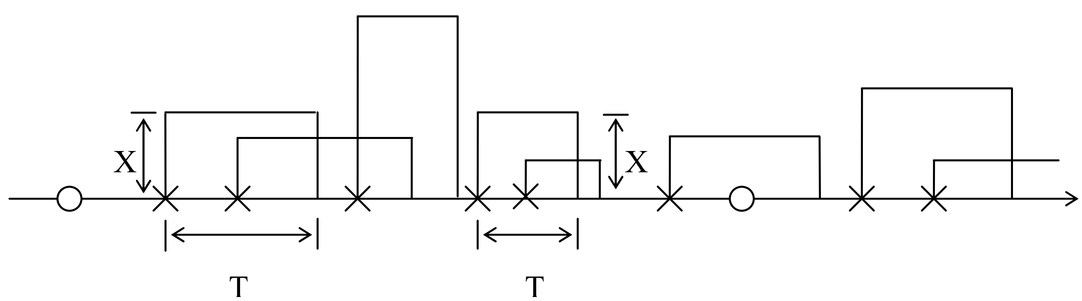
i) Storm origins (○) arrive according to a Poisson Process



ii) Each storm origin generates a random number of rain cells beginning at (×)



iii) The duration and intensity of each rain cell are exponentially distributed



iv) The total intensity at any point in time is the sum of the intensities due to all active rain cells at that point

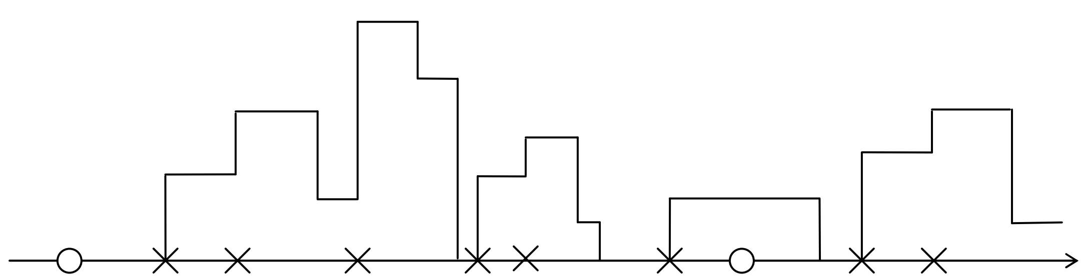


Figure 3.11: A scheme for the Neyman-Scott rectangular pulses model.

The exceedence probability function *R(x)= 1-F(x)* is defined as :

$$R(x) = 1 - \left[\alpha e^{\frac{-x}{\xi}} + (1-\alpha) e^{\frac{-x}{\theta}} \right] \qquad (3.90)$$

The first three moments about the origin are

$$M_1 = E(X) = \int_{-\infty}^{\infty} x f(x) dx = \alpha\xi + (1-\alpha)\theta \qquad (3.91)$$

$$M_2 = E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = 2\alpha\xi^2 + 2(1-\alpha)\theta^2 \qquad (3.92)$$

$$M_3 = E(X^3) = \int_{-\infty}^{\infty} x^3 f(x) dx = 6\alpha\xi^3 + 6(1-\alpha)\theta^3 \qquad (3.93)$$

Hence, the mean rain cell intensity $\mu_x = E(X)$ and $E(X^2)$ for the NSRP model are given by Eqs. (3.68) and (3.69) respectively. In addition, the $E(C)$ and $E(C^2\text{-}C)$ are

$$\mu_c = E(C) = \nu \qquad (3.94a)$$

$$E(C^2 - C) = \mu_c^2 - 1 = \nu^2 - 1 \qquad (3.94b)$$

With the above properties, the NSRP model with mixed exponential distribution has seven parameters, namely λ, ν , β, η , α, ξ, and θ that characterize respectively the origin of storm, the number of cells, the positions of cells relative to the storm origin, the duration of rain cells, the mixing probability and the intensity of the rain cells that is described by the last two parameters. Hence,

$$\Theta = (\lambda, \nu, \beta, \eta, \alpha, \xi, \theta) \qquad (3.95)$$

With the mixed exponential distribution to represent the rain cell intensities, the model structure follows Equations (3.82) to (3.85) but the rain cell intensities is described by:

$$P_X(\alpha,\xi,\theta) = \int f(x)dx = \alpha\left(1-e^{\frac{-x}{\xi}}\right)+\left(1-\alpha\right)\left(1-e^{\frac{-x}{\theta}}\right) \tag{3.96}$$

3.2.2.10 Probability of dry periods

The expression for the probability of an arbitrary interval of any chosen length being dry is a useful property to be derived as it may be used in fitting the model or comparing the model with field data. Cowpertwait (1991) derived this expression in the case where the rain cells are distributed according to Poisson law.

$$\mathrm{P}\left(Y_i^{(h)} = 0\right) = \exp\begin{bmatrix} -\lambda h + \lambda\beta^{-1}(\nu-1)^{-1}\{1-\exp[1-\nu+(\nu-1)e^{-\beta h}]\} \\ -\lambda\int_0^\infty [1-p_h(t)]dt) \end{bmatrix} \tag{3.97}$$

in which

$$p_h(t) = \left\{e^{-\beta(t+h)} + 1 - (\eta e^{-\beta t} - \beta e^{-\eta t}/(\eta-\beta)\right\}\times$$
$$\exp\left\{-(\nu-1)\beta(e^{-\beta t}-e^{-\eta t})/(\eta-\beta) - (\nu-1)e^{-\beta t} + (\nu-1)e^{-\beta(t+h)}\right\}$$

and

$$\int_0^\infty [1-p_h(t)]dt = \frac{1}{\beta}\left\{\gamma + \ln[(\frac{\eta}{\eta-\beta} - e^{-\beta h})\cdot(\nu-1)]\right\}$$

where γ is a Euler's constant = 0.5772.

The transition probabilities, $P(Y_{i+1}^{(h)} > 0 \big| Y_i^{(h)} > 0)$ and $P(Y_{i+1}^{(h)} = 0 \big| Y_i^{(h)} = 0)$, denoted as $\phi_{WW}(h)$ and $\phi_{DD}(h)$, respectively, can be expressed in terms of the probability of dry period $\mathrm{P}\left(Y_i^{(h)} = 0\right) = \phi(h)$ as follows (Cowpertwait, 1996):

$$\phi_{DD}(h) = \phi(2h)/\phi(h) \tag{3.98}$$

$$\phi(h) = \phi_{DD}(h)\phi(h) + \{1-\phi_{WW}(h)\}\{1-\phi(h)\} \tag{3.99}$$

so that

$$\phi_{WW}(h) = \{1 - 2\phi(h) + \phi(2h)\} / \{1 - \phi(h)\} \tag{3.100}$$

3.2.2.11 Parameter Estimation

The fitting of the parameters and the assessment of the adequacy of the fit raise many statistical questions. The different methods of parameter estimation of the NSRP model have been discussed extensively in Chapter 2. However, the method of moments is the most frequently used for estimating the parameters of the NSRP(Rodriguez-Iturbe et. al., 1987a,b; Entekhabi et.al.,1989; Cowpertwait, 1991). Following Cowpertwait et. al.,(1996) the historical hourly series of rainfall data is aggregated at three different temporal scales that is, 1,6 and 24 hours scales using the expressions of the mean at 1-hour (3.69), the variances at 1, 6 and 24-hour (3.70), lag-1 autocorrelation at 1, 6 and 24-hour(3.70) and probability of dry at 24-hour (daily) (3.96). The using of mean of more than one level of aggregation is not possible since $E[Y_i^{(kh)}] = kE[Y_i^{(h)}]$. Therefore, in this study parameter estimation procedure is to be achieved by minimizing the sum of squares, where the squared terms are the differences between the selected expressions of the model and their equivalent historical sampled values. Let $M_i \equiv M_i(\lambda, \nu, \beta, \eta, \xi)$ be a function of the original NSRP model, and let M_i^s be its historical sampled value.

$$S = \sum_{i=1}^{m} w_i \left[1 - \frac{M_i}{M_i^s}\right]^2 \qquad m \geq 5 \quad \lambda, \beta, \eta, \xi > 0, \nu > 1 \text{ and } M_i^s > 0. \tag{3.101}$$

w_i is a weight and it allows greater weight to be given to fitting some sample moments relative to others. The use of a ratio of model function is to ensure that large numerical values do not dominate the fitting procedure. Cowpertwait (1996) applied a weight of 100 to the term relating to the sample mean to ensure that this is matched almost exactly by the model. He also suggested the use of a larger set of sample moments (e.g. mean, variances and autocorrelations at different aggregations, probability of dry days and transition probabilities), assigning weights to different statistics. However, in this study, there are two proposed fitting procedures to be used:

3.2.2.11.1 Model parameter estimation using autocorrelations

As suggested by Rodriguez-Iturbe et al. (1987), Entekhabi et al. (1989) and Cowpertwait (1991, 1996), the sample moments to be used are 1 hour mean [$\hat{\mu}(1)$], variances at 1, 6 and 24 hourly [$\hat{\gamma}(1), \hat{\gamma}(6), \hat{\gamma}(24)$], lag-1 autocorrelations at 1, 6, and 24 hourly [$\hat{\rho}(1,1), \hat{\rho}(6,1), \hat{\rho}(24,1)$], and probability of dry days [$\hat{\phi}(24)$]. The following estimators of $\mu(h), \gamma(h), \gamma(h,1)$ were employed to avoid bias (e.g. see Trenberth, 1984):

Mean: $\hat{\mu}(h) = \sum_{i=1}^{n} \sum_{j=1}^{n_k^{(h)}} Y_{i,j,k}^{(h)} / \left\{ n_k^{(h)} n \right\}$ (3.102)

Variance : $\hat{\gamma}(h) = \sum_{i=1}^{n} \sum_{j=1}^{n_k^{(h)}} \left\{ Y_{i,j,k}^{(h)} - \hat{\mu}_k(h) \right\}^2 / \left\{ n_k^{(h)} n \right\}$ (3.103)

Covariance: $\hat{\gamma}(h,1) = \sum_{i=1}^{n} \sum_{j=1}^{n_k^{(h)}-1} \left\{ Y_{i,j,k}^{(h)} - \hat{\mu}_k(h) \right\} \left\{ Y_{i,j+1,k}^{(h)} - \hat{\mu}_k(h) \right\} / \left\{ \left(n_k^{(h)} - 1 \right) n \right\}$ (3.104)

Where k is a calendar month index (k=1 for January, 2 for February, etc), $Y_{i,j,k}^{(h)}$ is the jth h-hourly total in year I for month k, $n_k^{(h)}$ is the number of h-hourly totals in month k and n is the number of years of record. The autocorrelations of lag-k is $\hat{\rho}(h,k) = \hat{\gamma}(h,k) / \hat{\gamma}(h)$.

The weight of 100 is applied to term relating to sample mean and one to the others (Cowpertwait et. al.,1996). Therefore, in this study, based on procedures proposed by Cowpertwait et.al (1996) the following equation is optimized to estimate the parameters.

$$S = \sum_{i=1}^{m} w_i (1 - M_i / M_i^s)^2$$

$$= 100 \cdot \left(1 - \frac{\mu(1)}{\hat{\mu}(1)}\right)^2 + w_2 \cdot \left(1 - \frac{\gamma(1)}{\hat{\gamma}(1)}\right)^2 + w_3 \cdot \left(1 - \frac{\rho(1,1)}{\hat{\rho}(1,1)}\right)^2 + w_4 \cdot \left(1 - \frac{\gamma(6)}{\hat{\gamma}(6)}\right)^2$$
$$+ w_5 \cdot \left(1 - \frac{\rho(6,1)}{\hat{\rho}(6,1)}\right)^2 + w_6 \cdot \left(1 - \frac{\gamma(24)}{\hat{\gamma}(24)}\right)^2 + w_7 \cdot \left(1 - \frac{\rho(24,1)}{\hat{\rho}(24,1)}\right)^2 + w_8 \cdot \left(1 - \frac{\phi(24)}{\hat{\phi}(24)}\right)^2 \quad (3.105)$$

3.2.2.11.2 Model parameter estimation using transition probabilities

Cowpertwait et.al. (1996) found that NSRP model matched poorly the historical proportion of dry days when autocorrelations were used in the fitting procedures. The possible explanation for this is that the model is unable to match both the autocorrelations and the proportions of dry days. Moreover, autocorrelations tend to have large sampling errors due to large number of zero depths. Hence, the lag-1 autocorrelations are excluded and transition probabilities in Eqs. (3.97) and (3.99) are used. The choice of sample moments in this study was based upon G. Calenda Napolitano (1999) where the choice of aggregation scale must not be too close or else the optimization procedure may fail. Hence the sample moments and the transition probabilities used were one-hour mean [$\hat{\mu}(1)$], variances at one, six and 24 hourly [$\hat{\gamma}(1), \hat{\gamma}(6), \hat{\gamma}(24)$], transition probabilities of *P00* (dry-dry event) and *P11*(wet-wet event) at hourly and daily scales [$\phi_{DD}(h), \phi_{WW}(h)$] and the probability of dry days [$\phi(h)$]. The estimators for the mean and variances are given in Equations (3.100) and (3.101) respectively. The observed transition probabilities were computed using the following formulas:

$$\hat{\phi}_{DD}(h) = \hat{p}_{DD}(h) = \frac{\hat{a}_{DD}(h)}{\hat{a}_{DD}(h) + \hat{a}_{DW}(h)}$$
$$\hat{\phi}_{WW}(h) = \hat{p}_{WW}(h) = \frac{\hat{a}_{WW}(h)}{\hat{a}_{WW}(h) + \hat{a}_{WD}(h)} \quad (3.106)$$

Where $\hat{a}_{ij}(h)$ denotes the number of times in the sample of observations of rainfall occurrences that a transition from state i on the h^{th} hour to state j on the $(h+1\}^{\text{th}}$ hour occurs and

$$\hat{a}_{i.} = \hat{a}_{iD}(h) + \hat{a}_{iW}(h). \tag{3.107}$$

Equal weights are given to all terms. Therefore the following equation is to be optimized:

$$\begin{aligned} S &= \sum_{i=1}^{m} w_i (1 - M_i / M_i^s)^2 \\ &= w_1 \cdot \left(1 - \frac{\mu(1)}{\hat{\mu}(1)}\right)^2 + w_2 \cdot \left(1 - \frac{\gamma(1)}{\hat{\gamma}(1)}\right)^2 + w_3 \cdot \left(1 - \frac{\gamma(6)}{\hat{\gamma}(6)}\right)^2 + w_4 \cdot \left(1 - \frac{\gamma(24)}{\hat{\gamma}(24)}\right)^2 \\ &\quad + w_5 \cdot \left(1 - \frac{\phi_{DD}(1)}{\hat{\phi}_{DD}(1)}\right)^2 + w_6 \cdot \left(1 - \frac{\phi_{DD}(24)}{\hat{\phi}_{DD}(24)}\right)^2 + w_7 \cdot \left(1 - \frac{\phi_{WW}(1)}{\hat{\phi}_{WW}(1)}\right)^2 \\ &\quad + w_8 \cdot \left(1 - \frac{\phi_{WW}(24)}{\hat{\phi}_{WW}(24)}\right)^2 + w_9 \cdot \left(1 - \frac{\phi(24)}{\hat{\phi}(24)}\right)^2 \end{aligned} \tag{3.108}$$

3.2.2.12 Optimization Techniques

The parameters of the model are to be estimated by minimizing Eqs.(3.104) or (3.107). There are many methods discussed in literature on minimizing the objective functions. However in this study the Shuffled Complex Evolution-University of Arizona (SCE-UA) method by Duan et. al.(1992) is used in minimizing the model function. The SCE-UA is a global optimization method that has been shown to be able to provide more accurate and more efficient search for the optimal solution of complex nonlinear objective functions as compared to the local optimization technique such as Nelder and Mead Simplex or Quasi Newton Search (Duan et al.,1992). This algorithm requires the knowledge of the model parameters upper and lower bounds before it can be implemented.

The SCE-UA method starts with a population of points sampled randomly from the feasible space and the population is partitioned into several communities. The communities evolve based on a statistical reproduction process that uses the simplex geometric shape to direct the search in an improvement direction. As the search progresses, the entire population are shuffled and points are reassigned to communities to ensure information sharing. If the initial population is large enough, the entire population tends to converge to the neighborhood of the global optimum.

The SCE-UA method combines the strengths of the simplex procedure with the concepts of controlled random search, competitive evolution and the newly developed concept of complex shuffling. The strategy of the SCE-UA method is as follows (Duan et.al., 1992):

i. Initializing process

   To select $p \geq 1$ and $m \geq n+1$, and to compute the sample size $s = pm$ where p is the number of complexes, m is the number of points in each complex, and n is the dimension of the problem.

ii. Generation of a sample

   To sample s points $x_1, \ldots, x_s$ in the feasible space and to compute the function value f_i at each point x_i using a uniform sampling distribution.

iii. Rank of points

   To sort the s points in order of increasing function value and to store them in an array $D = \{x_i, f_i, i = 1, \ldots, s\}$.

iv. Partition of array D

   To partition D into p complexes $A_1, \ldots, A_p$, each containing m points, such that $A_k = \{x_j^k, f_j^k | x_j^k = x_{k+p(j-1)}, f_j^k = f_{k+p(j-1)}, j=1, \ldots, m\}$.

v. Evolution

   To evolve each complex A^k, $k = 1, \ldots, p$, according to the competitive complex evolution algorithm.

vi. Shuffling the complexes

   To replace $A_1, \ldots, A_p$ into D, such that $D = \{A_k, k = 1, \ldots, p\}$ and to sort D in order of increasing function value.

vii. Convergence

To stop if the convergence criteria are satisfied, or to return to step (iv).

The population is portioned into several communities (complexes), each of which is permitted to evolve independently. After a certain number of generations, the communities are mixed and new communities are formed through a process of shuffling. This procedure enhances survivability by a sharing of the information (about the search space) gained independently by each community (Duan et al., 1992). This strategy uses the information contained in the sub complex to direct the evolution in an improved direction. The processes of competitive evolution and complex shuffling inherent in the SCE-UA algorithm help to ensure that the information contained in the sample is efficiently and thoroughly exploited. They also help to ensure that the information set does not become degenerate. These properties provide the SCE-UA method with good global convergence properties over a broad range of problems.

As mentioned earlier the SCE-UA method requires the knowledge of the upper and lower bounds of the model parameters before the algorithm can be implemented. Based on the results by Cowpertwait et.al. (1996) and Calenda et.al. (1999), Table 3.4 presents the range of parameter values used in this optimization computation. The mixed exponential distribution is represented by parameters ξ and θ with α represents the mixing probabilities. The ξ is always smaller than θ. Table 3.5 presents the SCE-UA method options in optimization program. These options are part of the requirements in the SCE-UA algorithm before the parameters could be estimated. The parameters obtained will be used in the generation of hourly rainfall data. Since seasonal variations are considered in this modeling the parameters are evaluated for each month with twelve sets of parameters for the calibrated NSRP model.

Table 3.4: Parameter ranges in optimization procedure

Parameters	Description	Parameter ranges
λ	Inter-arrival times of storms	0.001 - 0.05
v	Number of rain cells	1 - 20
β	Waiting times from the origin to the rain cell	0.01 - 0.5
η	Rain cell duration	0.1 - 5
ξ	Rain cell intensities (Exponential)	0.01 - 4
ξ	Rain cell intensities (Mixed Exponential)	0.001 - 20
α	Mixing probabilities	0 - 1
θ	Rain cell intensities (Mixed Exponential)	10 - 100

Table 3.5: SCE-UA method options in optimization program

Option	Description	Value
MAXN	Maximum number of trials	10000
KSTOP	Number of shuffling loops	10
PECNTO	Percentage by which the criterion value must change in the specified number of shuffling loops	0.01
NGS	Number of complexes used in optimization search	2
ISEED	Random seed used in optimization search	-1
INIFLG	Flag on whether to include the initial point in the starting population	1

3.2.3 Simulation of the hourly rainfall series

The MATLAB program was designed to simulate the rainfall data based upon the Neyman-Scott Rectangular Pulse (NSRP) model. The NSRP model consists of five processes for describing the following properties:

i. Numbers of storms from inter-arrival times - Storm origins occur as a Poisson process with a mean rate λ/hour.
ii. Number of rain cells - Average number of rain cells per storm is ν.

iii. Waiting times from the origin to the rain cell - Average waiting time from the origin to the rain cell is $1/\beta$ hours.
iv. Rain cell duration - Average rain cell duration is $1/\eta$ hours.
v. Rain cell intensities - Average rain cell intensity is $1/\xi$ if the rain cells intensities are described by the Exponential Distribution. If the rain cells intensities are described by the mixed exponential distribution, then the average rain cell intensity is $1/\xi$ and $1/\theta$ with a mixing probability of α.

The program includes Poisson and exponential random number generation procedure.

The hourly rainfall simulation procedure of the NSRP model is illustrated in Figure 3.3. The following steps are followed in generating the hourly rainfall:

i. Generate the number of storms in which the arrival rate is a Poisson process.
ii. Generate a number of rain cells based upon the Poisson distribution originated from the storm origin.
iii. Generate the time intervals, t, between the rain cells and the storm origin where t is exponentially distributed.
iv. Generate the duration for each rain cell based upon the exponential distribution.
v. Generate the intensities for each rain cell based upon the exponential or the mixed exponential distribution.
vi. Calculate the position of the storms by adding up the waiting time between the storm origins.
vii. Calculate the position of each rain cells by adding up the position of storm origin and the intervals between rain cells and storm origin.
viii. Calculate the duration and the intensities of each storm.
ix. Calculate the total intensities of the storm.
x. Calculate the hourly intensities generated by the storms.

Series of rainfall data will be generated depending on the number of simulation chosen. Data is generated according to months. Sample of MATLAB programs are given in Apendix D.

3.2.4 Models Assessment

In this study, the performance of the traditional NSRP using the exponential distribution for the rain cell intensities will be assessed and compared with the performance of the proposed NSRP using the mixed exponential. For each model, two fitting strategies were adopted, using autocorrelations or using transition probabilities as mentioned previously. More specifically, the following cases of NSRP model calibration are considered.

1. The NSRP model with exponential distribution to describe rain cell intensities.

- i. Using autocorrelations in the fitting procedure and is referred hereafter as the EXP.
- ii. Using transition probabilities in the fitting procedure and is referred hereafter as the EXPTRAN.

2. The proposed NSRP model with mixed exponential distribution to describe rain cell intensities.

- i. Using autocorrelations in the fitting procedure and is referred hereafter as the MEXP.
- ii. Using transition probabilities in the fitting procedure and is referred hereafter as the MEXPTRAN.

The flowchart of the working strategies for the NSRP models is given in Figure 3.14

3.2.4.1 Graphical Method

Graphically, the simulated rainfall properties represented by the box-plots (Figure 3.12) are compared with the observed properties (represented by the dots connected by the dashed lines). If the observed value is comparable to the median value (the middle 50% value) of the boxplots, then the proposed model is said to have an "excellent" or "very well" ability in preserving the properties of the historical data. If the observed value falls on the whiskers and within the range defined by the simulated minimum and maximum, then the proposed model is said to have a "fair" ability in preserving the properties of the historical data. Otherwise, the model either underestimates or overestimates the observed statistical characteristics. Figure 3.12 show the characteristics of a box plot.

3.2.4.2 Root-mean-square error (RMSE).

Quantitatively both models are compared using the root-mean-square errors (RMSE) calculated for each property tested. The root-mean-square error formula is as follows:

$$\text{RMSE} = R_M = \left\{ \frac{\sum_{i=1}^{n} \left(S - \hat{S}_m \right)^2}{n} \right\}^{\frac{1}{2}} \tag{3.109}$$

where S is statistics of the observed, $\hat{S}_m$ is the median of the simulated, *n* is the number of simulated statistics.

3.2.4.3 Statistical Properties

Statistical properties of 30 synthetic hourly time series produced by each model were analyzed graphically using box plots for the monthly comparisons of the 30 simulated series with the observed. The statistical properties examined include:

a. One-hour series

The mean, variance, autocorrelation, coefficient of skewness of the hourly rainfall amount are to be computed from the generated hourly series. These properties will determine the model's suitability and accuracy in preserving the observed at the same scale as the generated series using the generated hourly series.

b. Six-hour series

The generated hourly series are lumped or aggregated to six- hourly rainfall series. The mean, variance, autocorrelation and coefficient of skewness of the six-hour rainfall will be computed. These properties will determine the ability of the model in preserving the six-hour rainfall process.

c. Twenty-four or Daily series

The generated hourly series are then lumped or aggregated to 24-hourly series. This is equivalent to the daily scale. The mean, variance, autocorrelation and the coefficient of skewness of the 24-hour rainfalls will be computed. These are daily rainfall properties and the properties will determine the model ability in describing the daily rainfall process.

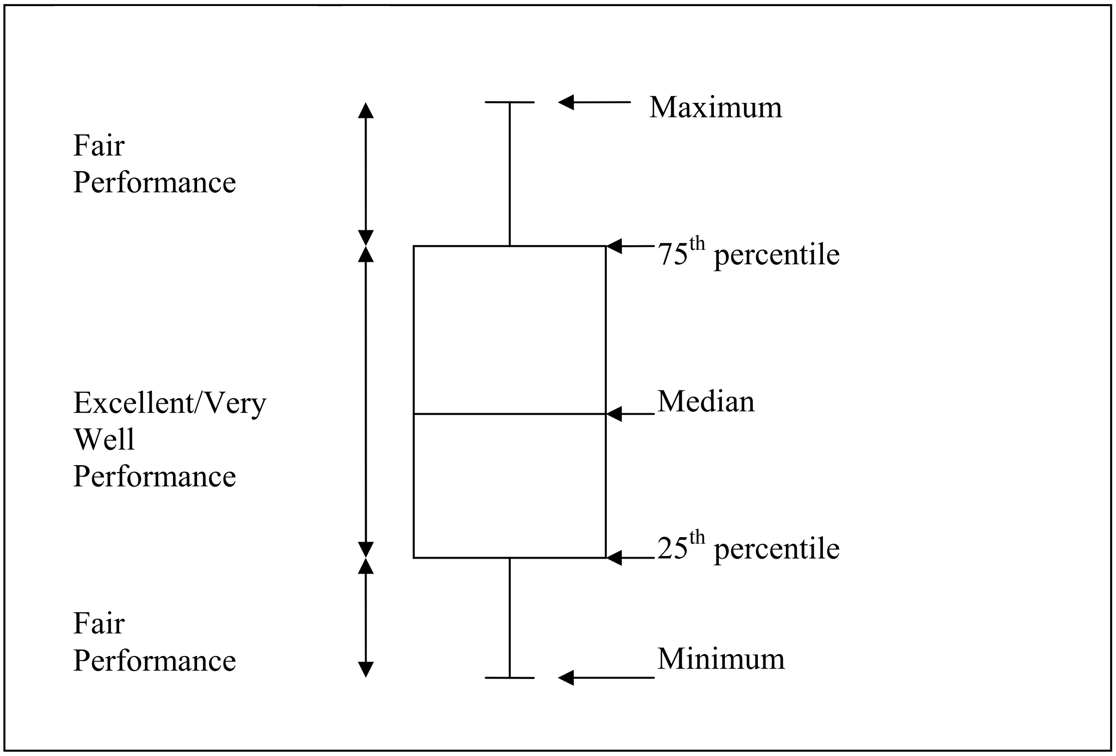


Figure 3.12: Characteristics of a Box plot

d. Monthly series

The generated hourly series are lumped or aggregated to become monthly series. The lumping is done by accumulating from hourly to 24 hourly, then to one-month scale. Only the mean and variance will be computed to represent the statistical characteristics of the monthly scales. These properties will determine the model ability to describe the rainfall process at monthly scale.

3.2.4.4 Physical Properties

The physical properties of the rainfall series will represent the underlying process of rainfall events. The properties identified as physical properties are:

a. One-hour series

The physical properties include the distribution of the maximum rainfall amounts, the probability of dry hours and the hourly transition probabilities of rainfall occurrences *P00* (dry-dry hours) and *P10* (wet-dry hours). These describe the rainfall process physically at hourly scale.

b. Twenty-four-hour or daily series

The physical properties include the distribution of the maximum rainfall amounts, the probability of dry days and the hourly transition probabilities of dry-dry days and wet-dry days. These are important physical characteristics that are required in the daily series and also crucial in the water management planning. This will determine the model ability in describing the physical process of daily rainfall.

c. Monthly series

The physical properties include the distribution of the maximum and minimum rainfall amounts. These properties are important for water management planning.

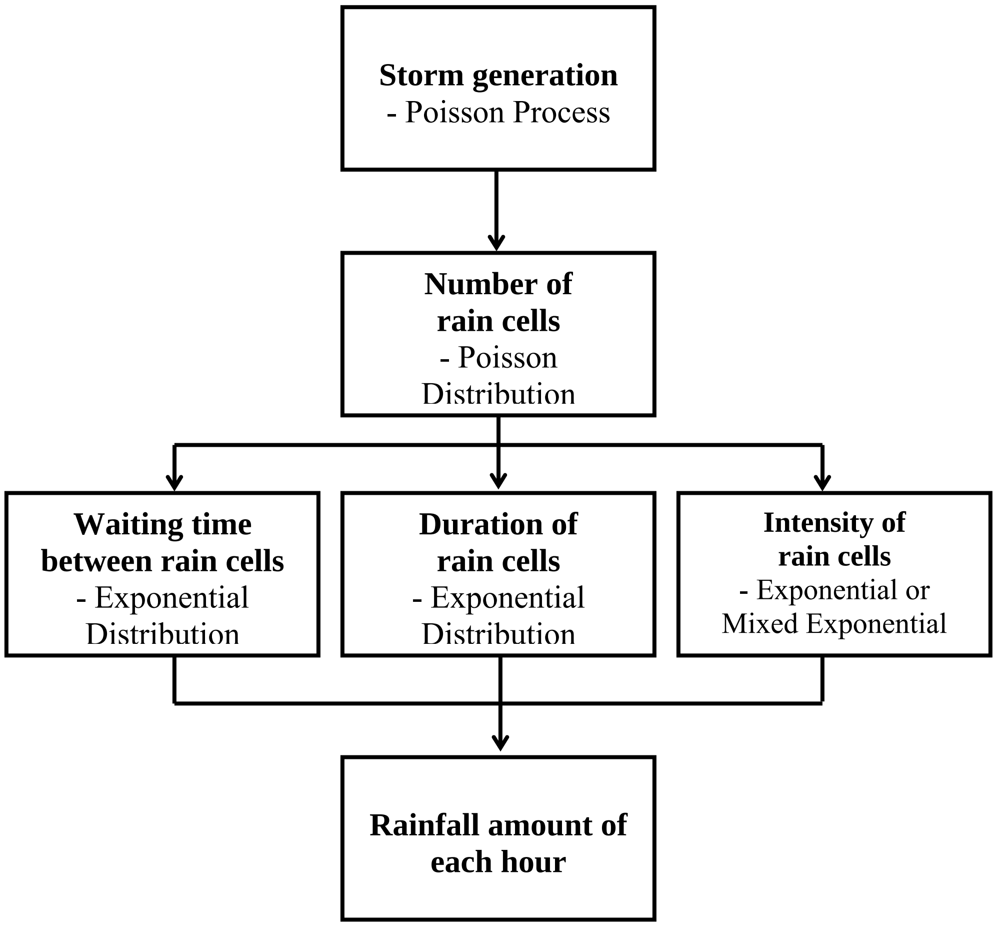


Figure 3.13: Flowchart of simulation procedures of the NSRP model

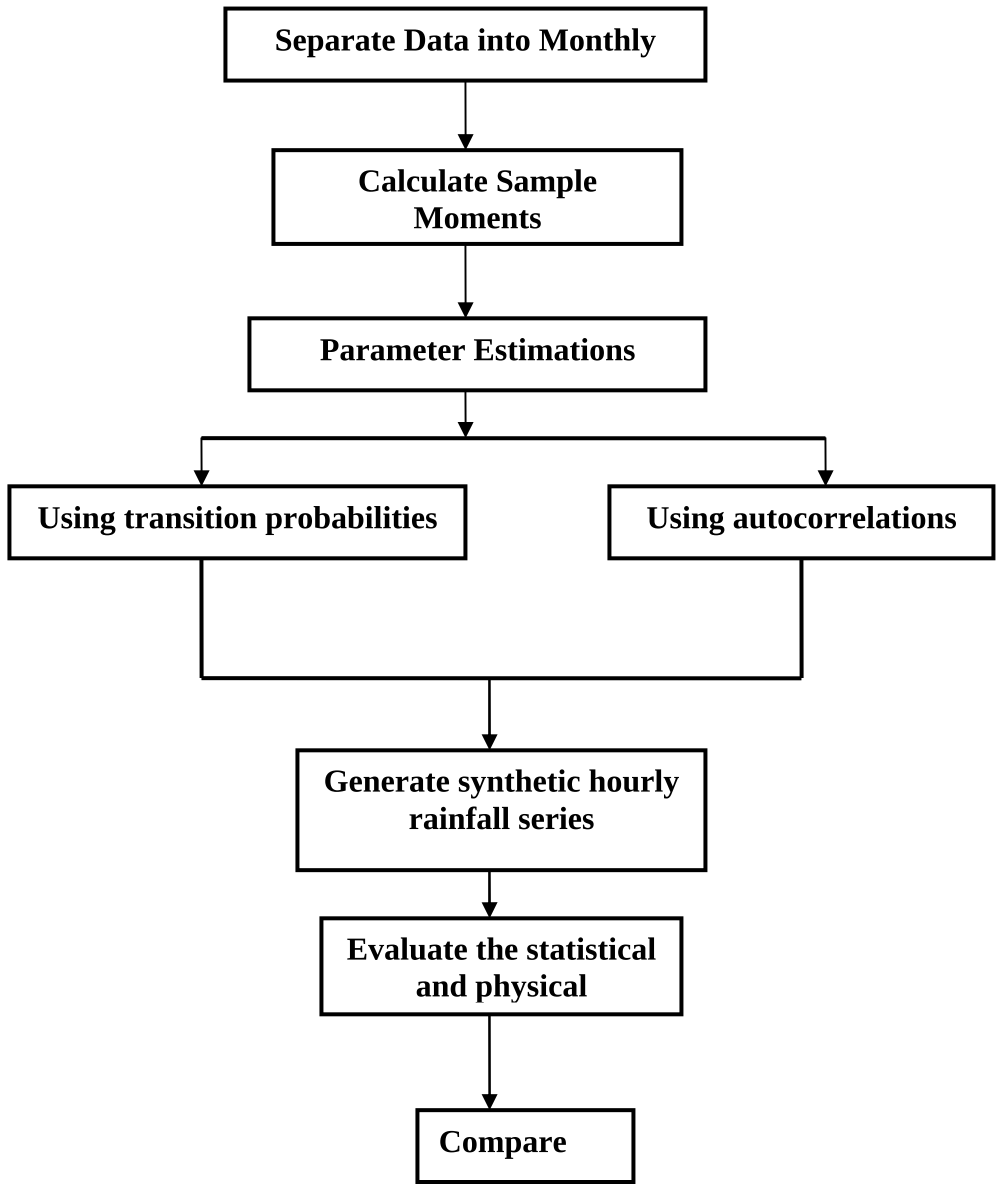


Figure 3.14: Flowchart of the working strategy for the NSRP models

3.3 Stochastic Rainfall Modeling using Markov Chain Mixed Exponential Model (MCME)

3.3.1 Introduction

A rainfall model based on daily precipitation is attractive because relatively long and reliable records are readily available and such a model is frequently efficient for many practical problems. Stochastic models of daily rainfall are usually divided into two parts, a model of rainfall occurrence which provides a sequence of dry and wet days, and a model of rainfall amounts, which simulates the amount of rainfall occurring on each wet day and then both are superimposed to form the overall rainfall model. (Eagleson, 1978; Woolhiser et.al,1982, Roldan et.al,1982,).

One of the popular stochastic modeling of daily rainfall is the Markov Chain-Mixed Exponential (MCME). The first-order two-state Markov Chain model is used to describe the hourly rainfall occurrence process and the Mixed Exponential distribution is used to describe the hourly amount distribution.). Many studies have used the combination of Markov Chain and Mixed Exponential(MCME) to model daily rainfall series and the combined model had proven to be the best in describing rainfall processes (Woolhiser and Pegram. 1979, Woolhiser et.al,1982, Han, 2001).

An effort on modeling the hourly rainfall series using the two parts modeling was done by Katz and Parlange (1995) that fitted stochastic models to time series hourly data by using an extension of chain-dependent process commonly fit to daily rainfall amount, and the amount distribution is described by a power transformation of the normal. The model was said to be competitive to the so-called conceptual model (pulse-based) but failed to reproduce the statistics of 12h and 24 h aggregation. However using MCME on hourly series has never been reported in literature yet. Han et al (1982) pointed out that rainfall for short time intervals, is more difficult to model than long time period because of the sequential persistence between rainfall amounts, and also because the time-series are dominated by zero values (intermittent process). Therefore, this study will explore the possibility of using MCME in modeling the hourly rainfall process.

3.3.2 The Hourly MCME Model Derivation

According to Pattison (1965), a first-order Markov Chain could be used to model hourly rainfall observations during sequences of nonzero rainfall (wet hours). Using the WMO guideline, a wet day is defined as a day with a rainfall amount above a fixed threshold of 0.1 mm. This threshold will be used as well for defining the wet hours with rainfall amount of greater than or equal to 0.1 mm.

In this study the MCME is to be applied on the hourly rainfall series with two states: dry or wet hours; modeled as either a 0 or 1 respectively with a first order Markov Chain explaining the dependence between wet and dry hours on successive hours. The rainfall amounts is modeled using the mixed exponential distribution.

3.3.2.1 The Occurrence Process

Let assume the amount of precipitation falling on h^{th} hour and t^{th} day is a random variable

$$Z_t(h) = X_t(h).Y_t(h) \tag{3.110}$$

where X_t (h) represents the occurrence process and Y_t (h) represents the amount of precipitation when X_t (h) is wet.

The hourly occurrence process $\{X_t(h): h = 1, 2, ..., 24; t = 1, 2, ...\}$ is defined as

$$X_t(h) = \begin{cases} 1 & \text{if } h\text{th hour of } t\text{th day is wet} \\ 0 & \text{otherwise} \end{cases} \tag{3.111}$$

where wet hour refers to one on which measurable precipitation occurs. The conventions adopted are as follows:

$$X_t(-1) = X_{t-1}(23), X_t(0) = X_{t-1}(24), X_t(25) = X_{t+1}(1), X_t(26) = X_{t+1}(2),.......$$

It is assumed that the process $X_t(h)$ constitutes a two-state, first order Markov Chain with transition probabilities

$$\begin{aligned}
&P_{ij}(h) = P\left\{X_t(h) = j \middle| X_t(h-1) = i\right\} \qquad i, j = 0,1 \\
&t = 1,2,........365 \\
&h = 1,2,........24 \\
&P_{i1}(h) = 1 - P_{i0}(h) \qquad i = 0,1
\end{aligned} \tag{3.112}$$

Let $Y_t(h)$ be the amount of rainfall that falls on day t and hour h when $X_t(h) = 1$. We assume that $Y_t(h)$ is serially independent and independent of $X_t(h-1) = 1$. This means that there is dependence on rainfall occurrence from hour to hour but that the amount is independent of previous occurrences and amounts. The assumption of independence between the amounts of rainfall on successive days leads to significant simplifications in the model structure and has been used by several previous researchers (Coe and Stern, 1982; Richardson and Wright, 1984).

Woolhiser and Pegram [1979] recommended the maximum likelihood method to estimate the parameters of the Markov Chain using the daily series. Therefore, the same procedures are applied to the hourly series. The log likelihood function:

$$\begin{aligned}
\ln L\left(\{X_t\}\right) &= \sum_{i=0}^{1} \sum_{j=0}^{1} \sum_{h=1}^{24} a_{ij}(h) \ln P_{ij}(h) \\
&= \sum_{h=1}^{24} \begin{bmatrix} a_{00}(h) \ln P_{00}(h) + a_{01}(h) \ln(1 - P_{00}(h)) \\ + a_{10}(h) \ln P_{10}(h) + a_{11}(h) \ln(1 - P_{10}(h)) \end{bmatrix}
\end{aligned} \tag{3.113}$$

where

$$P_{ij}(h) = \frac{a_{ij}(h)}{a_{i.}(h)} \tag{3.114}$$

and $a_{ij}(h)$ denotes the number of times in the sample of observations of precipitation occurrences that a transition from state i on the hth hour of the day to state j on the (h+1}th hour occurs and

$$a_{i.} = a_{i0}(h) + a_{i1}(h). \tag{3.115}$$

In relation to the above definition, the maximum likelihood estimates of Markov Chain parameters that are calculated by computing the observed number of transitions $a_{ij,k}$(h) from state (i=0 or 1) on hour h to state (j=0 or 1) on hour $h+1$ in period k across the entire length of record where 0 represents a dry hour and 1 represents a wet hour are represented as follows by substituting equation(3.102) to equation (3.101). The two parameters to be estimated are P_{00} and P_{10} and the definitions are as follows:

$$\begin{aligned} p_{00,k}(h) &= \frac{a_{00,k}(h)}{a_{00,k}(h) + a_{01,k}(h)} \\ p_{10,k}(h) &= \frac{a_{10,k}(h)}{a_{10,k}(h) + a_{11,k}(h)} \end{aligned} \tag{3.116}$$

The unconditional probability of being wet on day t and hour h can also be approximated by:

$$P\{X_t(h) = 1\} \approx \frac{[1 - p_{00,t}(h)]}{1 + p_{10,t}(h) - p_{00,t}(h)} \tag{3.117}$$

3.3.2.2The Amount Process

In describing the rainfall amounts of the hourly series, the empirical observed frequency distribution were fitted to the theoretical probability density function. The mixed exponential distribution was found to be the most accurate for describing the distribution of hourly rainfall amounts as compared to other popular candidate distributions such as simple exponential, gamma and Weibull (Fadhilah et al. 2007).

Let $Y_t(h)$ denotes the precipitation amount on the h^{th} hour of the t^{th} day. If $X_t(h)$=1, then $Y_t(h) > 0$ and is referred to as intensity. The convention is adopted that

$$Y_t(-1) = Y_{t-1}(23), Y_t(0) = Y_{t-1}(24), .Y_t(25) = Y_{t+1}(1), Y_t(26) = Y_{t+1}(2), ...$$

The distribution of the hourly rainfall amounts is described by the Mixed Exponential function.

$$\begin{aligned} &Y_t(h) = \frac{\alpha_t(h)}{\xi_t(h)} e^{\left(\frac{-y}{\xi_t(h)}\right)} + \frac{(1-\alpha_t(h))}{\theta_t(h)} e^{\left(\frac{-y}{\theta_t(h)}\right)} \\ &y > 0; 0 \le \alpha_t(h) \le 1;), 0 < \xi(h) < \theta(h) \\ &h = 1, ...24. \\ &t = 1, ...365 \end{aligned} \tag{3.118}$$

The mixed exponential distribution can be interpreted as the result of a random sample from two exponential distributions where the smaller mean $\xi(h)$ is sampled with probability $\alpha(h)$ and the distribution with the larger mean $\theta(h)$ is sampled with probability $(1-\alpha(h))$. he maximum likelihood estimates of the parameters of the mixed exponential distribution were obtained by maximizing the log likelihood function :

$$\ln L_k \{Y_t(h)\} = \sum_{j=1}^{N(k)} \left\{ \ln \left[\frac{\alpha_k(h)}{\xi_k(h)} e^{\left(\frac{-y_{kj}(h)}{\xi_k(h)}\right)} + \frac{(1-\alpha_k(h))}{\theta_k(h)} e^{\left(\frac{-y_{kj}(h)}{\theta_k(h)}\right)} \right] \right\} \tag{3.119}$$

where $\alpha_k(h), \xi_k(h),$ and $\theta_k(h)$ are the parameter values for the k^{th} period, y_{kj} is the amount of rainfall for the j^{th} wet hours in period k, and $N(k)$ is the number of wet hours in period k.

3.3.3 Parameter Estimation

The hourly data is pooled according to calendar months. Instead of using a local optimization technique as in most previous studies, the Shuffled Complex Evolution (SCE) global optimization method (Duan et.al.,1992) is employed for finding the optimal solution of the minimization of the likelihood function. This global optimization technique was found to be able to provide more accurate and more robust results than the local optimization procedures (Peyron and Nguyen, 2004). The SCE-UA technique has been discussed at length in section 3.4.6.

There are five parameters in which the two is used to describe the transitional probabilities and three explaining the mixed exponential can be found for 12 sets of monthly data. Each parameter set is then fitted to a finite Fourier series (Woolhiser and Pegram, 1979), where the parameters change periodically through the 12 months of the year. The parameter set for the rainfall process for each month *m* can be written as:

$$\gamma(m) = \{ p_{00}(m), p_{10}(m), \alpha(m), \xi(m), \theta(m) \} \tag{3.120}$$

The same orientation as in the daily MCME model is used in estimating the parameters of the model. The parametric monthly Fourier series representation of the parameters for $m = 1, 2, \ldots, w$ where $w = 12$ can be written as:

$$\gamma_m = \hat{\mu}_m + \sum_{j=1}^{h} \left\{ A_j \cos\left(\frac{2\pi jm}{w} \right) + B_j \sin\left(\frac{2\pi jm}{w} \right) \right\} \tag{3.121}$$

Here, h is the maximum number of harmonics needed to specify the variation of parameter concerned, it is however set to a constant $h = 5$ for the purposes of this research based on the research of Sang-Yoon Han (2001). Thus, a maximum of $2h + 1$ coefficients are needed to describe each parameter , To make a parsimonious estimation, a maximum of $2h+1$ coefficient are needed to describe each parameter γ_m. $\hat{\mu}_m$ is defined as the sample estimate of the unknown population periodic parameter γ_m where

$$\hat{\mu}_m = \frac{1}{w}\sum_{m=1}^{w}\mu_m \tag{3.122}$$

The coefficients of the Fourier series in Equation (3.108) are determined through maximum likelihood estimates as follows, for all j = 1, 2,….h harmonics specified as:

$$A_j = \frac{2}{w}\sum_{m=1}^{w}\mu_m \cos\left(\frac{2\pi jm}{w}\right) \tag{3.123}$$

$$B_j = \frac{2}{w}\sum_{m=1}^{w}\mu_m \sin\left(\frac{2\pi jm}{w}\right) \tag{3.124}$$

An alternate polar form of the Fourier series were also considered but not applied to the final model.

$$\gamma_m = \hat{\mu}_m + \sum_{j=1}^{h}\left[C_j \cos\left(\frac{2\pi jm}{w} + \theta_j\right)\right] \tag{3.125}$$

3.4.4 Simulation of Hourly Rainfall Process

MATLAB functions were developed to create a software package to simulate hourly rainfall using the MCME model for any time series data. The stochastic model was created such that the occurrence and amounts on any given hour would be random. The software package for the hourly series was created based on the daily series software package developed by Hussain (2007). Sample of this hourly MCME simulation programs are found in Appendix D.

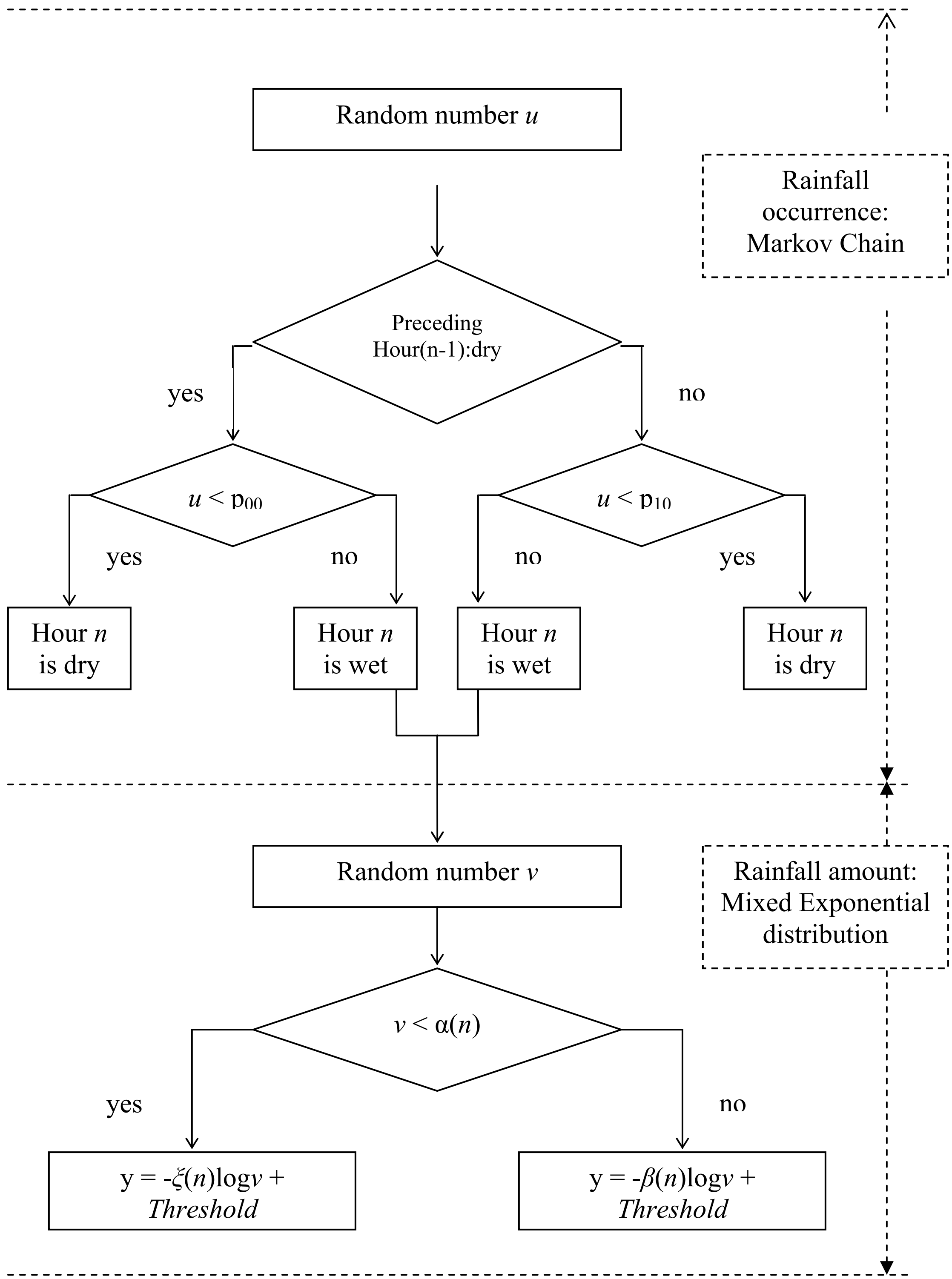


Figure 3.15: Flowchart of the simulation procedures of the MCME model.

3.3.4.1 Hourly Scale

In hourly simulation model, a day contained 24 values for occurrences and amounts. All the hourly data are separated into monthly data sets and twelve sets of monthly parameters are derived for the hourly rainfall MCME model. The data handling and random number generation procedure was much more computationally intensive for generating rainfall series for the hourly scale as compared to the procedure for the daily scale.

The parameters obtained for each month will be used to simulate hourly rainfall series. The simulation program uses the first-order, two-state Markov Chain for hourly rainfall occurrences and the mixed exponential for hourly rainfall amounts. The procedure for generating simulated hourly rainfall series is shown in Figure 3.15. The algorithm for the simulation of the hourly MCME model is as follows:

1. For any given hour, a uniform random number, u between 0 and 1 is generated.

2. The parameter set of the month to which the simulated hour belongs is extracted.
   i. If the preceding hour is dry and $u < p_{00}$, then the current hour is said to be dry and the process restart at step 1. However, if $u > p_{00}$, the hour is said to be wet and a rainfall amount is then required to be generated.
   ii. If the preceding hour is wet and $u < p_{10}$, then the current hour is said to be dry and the process restarts at step 1. However, if $u > p_{10}$, the hour is said to be wet and a rainfall amount is then required to be generated,

3. If step 2 determined a wet hour, another uniform random number, v is generated.

For the mixed exponential distribution, $\xi(n)$ and $\theta(n)$ are means of the smaller and the larger exponential distributions, respectively. If $u(n)$ is the mean of the hourly rainfall amount, it can be described by the following relation:

$$\mu(n) = \alpha(n)\xi(n) + (1 - \alpha(n))\theta(n)\,. \tag{3.126}$$

If $v < \alpha(n)$, the depth, y, is generated from an exponential distribution with smaller means, $\xi(n)$, using the transformation:

$$y = -\xi(n)\log v + threshold \tag{3.127}$$

If $v \geq \alpha(n)$, the depth y is generated from an exponential distribution with larger means, $\theta(n)$

$$y = -\theta(n)\log v + threshold \tag{3.128}$$

where threshold value in this study is equal to zero because a non-zero amount is considered as wet hour.

3.3.4.2 Daily Scale

The daily simulation MATLAB software package has been created by Hussain (2007). The same algorithm as stated above for the hourly process was used to describe successive day states and rainfall amounts. As expected, the data handling and random number generation was less intensive computationally.

3.4.5 Assessment of the MCME Model

The assessment of the MCME model performance is carried out for both hourly and daily rainfall simulations.

3.3.5.1 Assessment of the Hourly MCME Model

The MCME stochastic hourly rainfall generator was calibrated with the hourly data from 1981 to 1990 available at the Gombak. The mixed exponential goodness of fit test was assessed using observed hourly rainfall frequency within each month. Based on this calibration, a set of 50 simulated rainfall series were generated. Both graphical and numerical comparisons were used in the comparisons of simulated and observed statistical and physical properties as described in sections (3.3.14.1) and (3.3.14.2). The following parameters will be considered in this assessment.

a. Hourly series

The statistical properties consist of mean, standard deviation, coefficient of skewness, autocorrelations and hourly correlogram. The physical properties consist of maximum hourly rainfall amount and number of wet and dry hours.

b. Twenty-four hourly or daily series

The hourly series are lumped or aggregated to 24 hourly series rainfall. The statistical properties consist of mean, standard deviation, coefficient of skewness, autocorrelations and daily correlogram. The physical properties consist of maximum daily amount of rainfall and the number of wet and dry days.

c. Monthly series

The hourly series are lumped or aggregated to a one-month series rainfall. Statistical properties consist of mean and standard deviation while the physical properties consist of monthly maximum and minimum rainfall amount.

3.3.5.2 Assessment of the Daily MCME Model

Similarly, the daily model was applied to daily data from the Gombak station for the 1981-1990 period. The mixed exponential goodness of fit was assesses using observed daily rainfall frequency within each month. Based on the calibration, 50 simulations were generated. For each simulation output, a set of statistical and physical properties describe above were used to evaluate the ability of the MCME model in preserving the observed characteristics of rainfall. Therefore, similar assessments as for the hourly model were carried out for the daily model.

To test the accuracy of the hourly model in describing daily rainfall characteristics, the hourly simulations were lumped to form daily simulations and compared to the observed daily series. Similarly, to evaluate the performance of both hourly and daily models in preserving monthly rainfall properties both simulations were lumped to form monthly simulations and compared to the observed monthly rainfall series. The same statistical and physical criteria were used in these assessments.

3.5 NSRP and MCME Model Comparisons

Following assessment of the accuracy of the NSRP and MCME models in preserving observed rainfall characteristics, the performance of these two models can be compared. The comparisons to be made are as follows:

a. To compare the performance of NSRP with MCME hourly model
b. To compare the performance of NSRP with MCME daily model

To compare the model performance, only the properties at one-hour and 24-hour (daily) scales are chosen. While these two series have many applications in the water management process, they are also important in understanding the underlying process of any rainfall events.

a. Hourly series

The statistical properties consist of mean, standard deviation, coefficient of skewness, and autocorrelations. The physical properties consist of maximum hourly amount and probability of dry hours.

b. 24-hourly or daily series

The statistical properties consist of mean, standard deviation, coefficient of skewness and autocorrelations. The physical properties consist of maximum daily amount and probability of dry days. Both graphical and numerical comparisons were used in this evaluation as discussed in sections 3.3.14.1 and 3.3.14.2 to determine the best stochastic model that could describe the rainfall process at the study site.

3.6 NSRP and MCME Model Validation

Much of the work done in this study was done in the calibration period, assessing the descriptive ability of both models. To assess the predictive ability, the validation of the best NSRP hourly model and hourly MCME model can be carried out using available data from the 1991-2000 period at WP station. The monthly descriptive statistics for rainfall data are given in Appendix B. For this validation purposes, each rainfall simulation was generated for 20 years. The first 10-year simulated rainfall series were used for assessing the descriptive (calibration) ability of the models while the second 10-year series were used for evaluating their predictive (validation) ability. Statistical and physical properties of the observed and synthetic hourly and daily time series considered in the model validation are similar to the properties used to compare models. Both graphical and numerical comparisons discussed in sections 3.3.14.1 and 3.3.14.2 were used to evaluate and compare both models in the validation period in order to determine the best stochastic model that could predict the rainfall process at the study site.

3.6 Stationary and Nonstationary Stochastic Models

Stochastic means being or having a random variable. A stochastic model is a tool for estimating probability distributions of potential outcomes by allowing random variation in one or more inputs over time. The random variation is usually based on fluctuations observed in historical data for a selected period. The models for time series are in fact stochastic models.

An important class of stochastic models for describing time series is called stationary models. It is assumed that the process remains in equilibrium about a constant mean level or the process in a particular state of statistical equilibrium. A stochastic process is said to be strictly stationary if its properties are unaffected by a change of time origin, that is, if the joint probability distribution associated with m observations $z_{t_1}, z_{t_2}, ..., z_{t_m}$, made at any set times $t_1, t_2, ..., t_m$, is the same as that associated with m observations $z_{t_1+k}, z_{t_2+k}, ..., z_{t_m+k}$, made at times $t_1 + k, t_2 + k, ..., \ t_m + k$.

When m=1, the stationarity assumption implies that the probability distribution $f(z_t)$ is the same for all times t and may be written as $f(z)$. Hence, the stationary process has a constant mean,

$$\mu = E[Z_t] = \int_{-\infty}^{\infty} zp(z)\, dz \tag{3.129}$$

and a constant variance

$$\sigma_z^2 = E\left[(Z_t - \mu)^2\right] = \int_{-\infty}^{\infty} (z - \mu)^2 f(z)\, dz. \tag{3.130}$$

The mean μ of the stochastic process can be estimated by the sample mean

$$\bar{z} = \frac{1}{N}\sum_{t=1}^{N} z_t \tag{3.131}$$

of the time series, and the variance σ_z^2 of the stochastic process can be estimated by the sample variance

$$s_z^2 = \frac{1}{N}\sum_{t=1}^{N} (z_t - \bar{z})^2 \tag{3.132}$$

of the time series.

The stationarity assumption also implies that the joint probability distribution $p(z_{t_1}, z_{t_2})$ is the same for all times t_1, t_2 which are a constant interval apart. The covariance between z_t and z_{t+k} that is separated by k intervals of time, which under the stationarity assumption must be the same for all t, is called the autocovariance at lag k. It is defined by

$$\gamma_k = \text{cov}[Z_t, Z_{t+k}] = E[(Z_t - \mu)(Z_{t+k} - \mu)] \tag{3.133}$$

Similarly, the autocorrelation at lag k is

$$\rho_k = \frac{E[(Z_t - \mu)(Z_{t+k} - \mu)]}{\sqrt{E[(Z_t - \mu)^2]E[(Z_{t+k} - \mu)^2]}} = \frac{E[(Z_t - \mu)(Z_{t+k} - \mu)]}{\sigma_z^2} \tag{3.134}$$

since for a stationary process, the variance $\sigma_z^2 = \gamma_0$ is the same at time $t+k$ as at time t. Thus the autocorrelation at lag k, that is, the autocorrelation between z_t and z_{t+k}, is

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{3.135}$$

which implies that $\rho_0 = 1$. A number of estimates of the autocorrelation function have been suggested in the literature but the most satisfactory estimate of the kth lag autocorrelation ρ_k is

$$r_k = \frac{c_k}{c_0} \tag{3.136}$$

where

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}(z_t - \bar{z})(z_{t+k} - \bar{z}), \qquad k = 0, 1, 2, ..., K \tag{3.137}$$

is the estimate of the autocovariance γ_k, and $\bar{z}$ is the sample mean of the time series. It should also be noted that K should not be larger that $N/4$. Three conditions that should be considered are

$$
\begin{aligned}
&r_0 = 1 \\
&-1 \le r_k \le 1 \\
&r_k = r_{-k}
\end{aligned}
\tag{3.138}
$$

which implies that ρ_k should be in the range [-1,1].

Autocorrelation function is important in model identification because it can identify whether the model is stationary or nonstationary. Theoretically, the series is stationary if the estimated autocorrelation function quickly reduces to zero with increasing lag k.

Generally, there are three basic models for a Box-Jenkins stationary stochastic model. The models are

(i) Autoregressive model AR(p)

(ii) Moving Average model MA(q)

(iii) Mixed Autoregressive-Moving Average model (p,q).

However, forecasting has been of particular importance in many fields where many time series are often represented as nonstationary and, in particular, as having no natural constant mean level over time. Sometimes there are some trends in time series. Therefore, some simple operators that is the backward difference operators ∇, as follows, can be employed to the time series.

$$y_t = \nabla z_t = z_t - z_{t-1}, \qquad t = 2,3,\ldots,n \tag{3.139}$$

If the series is still nonstationary, then it can be differentiated once again so that it would be stationary.

$$\begin{aligned} y_t &= \nabla^2 z_t = \nabla(z_t - z_{t-1}) \\ &= z_t - 2z_{t-1} + z_{t-2} \qquad t = 2,3,\ldots,n \end{aligned} \tag{3.140}$$

Theoretically, if a time series have been differentiated twice, the series will be stationary. In Box-Jenkins, the model for the nonstationary stochastic model is called the Autoregressive Integrated Moving Average ARIMA(p,d,q) model.

3.7 Univariate Box-Jenkins Model

There are two main models in Box-Jenkins. These are seasonal model and nonseasonal model. Here we will only discuss the nonseasonal model that consists of the stationary model and the nonstationary model for the univariate time series that is the Autoregressive model, AR(p), Moving Average model, MA(q), Mixed Autoregressive-Moving Average model, ARMA(p,q) and the Autoregressive Integrated Moving Average model, ARIMA(p,d,q).

3.7.1 Autoregressive Model, AR(*p*)

The general autoregressive model is given by

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \cdots + \phi_p \tilde{z}_{t-p} + a_t \tag{3.141}$$

which is known as AR(p) model or autoregressive model of order p. It is a transformation from

$$z_t = \mu + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_p z_{t-p} + a_t \tag{3.142}$$

where

$z_t =$ observation at time t

$a_t =$ shock at time t

$\mu =$ mean

$\tilde{z}_t = z_t - \mu$

In (3.14), the variable z is regressed on previous values of itself. If we define an autoregressive operator of order p by

$$\phi_p(\mathrm{B}) = 1 - \phi_1 - \phi_2 \mathrm{B}^2 - \ldots - \phi_p \mathrm{B}^p \tag{3.143}$$

then, equation (3.14) can be written as

$$\begin{aligned} \phi_p(\mathrm{B})\tilde{z}_t &= a_t \\ \tilde{z}_t &= \phi_p^{-1}(\mathrm{B})a_t \end{aligned} \tag{3.144}$$

where $\mathrm{B}^p(\tilde{z}_t) = \tilde{z}_{t-p}$.

The model contains $p+2$ unknown parameters $\mu, \phi_1, \ldots, \phi_p, \sigma_a^2$, which have to be estimated from the data. The parameter σ_a^2 is the variance of the white noise process a_t.

3.7.2 Moving Average Model, MA(q)

Moving average model of order q is given by

$$z_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \tag{3.145}$$

where

$z_t =$ observation at time t

$a_t =$ shock at time t

$\mu =$ mean

Let $\tilde{z}_t = z_t - \mu$. Then, equation (3.18) will be

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \tag{3.146}$$

If we define a moving average operator of order q by

$$\theta_q(\mathrm{B}) = 1 - \theta_1 \mathrm{B} - \theta_2 \mathrm{B}^2 - \ldots - \theta_q \mathrm{B}^q \tag{3.147}$$

the moving average model may be written as

$$z_t = \left(1 - \theta_1 \mathrm{B} - \theta_2 \mathrm{B}^2 - \ldots - \theta_q \mathrm{B}^q\right) a_t \tag{3.148}$$

where $\mathrm{B}(a_t) = a_{t-1}$.

This model contains $q+2$ unknown parameters $\mu, \theta_1, \ldots, \theta_q, \sigma_a^2$ which are estimated from the data. σ_a^2 is the variance of the white noise process a_t.

3.7.3 Mixed Autoregressive-Moving Average Model, ARMA(*p*,*q*)

Mixed autoregressive-moving average model, ARMA(p,q) model consists of both the autoregressive model of order p, AR(p) and the moving average model of order q, MA(q). This model gives some flexibility in the fitting of actual time series by combining both of the models. Thus, the autoregressive-moving average model of order p and q is given by

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \ldots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \qquad (3.149)$$

By using the autoregressive operator of order p and the moving average operator of order q, the ARMA(p,q) model can be written as

$$\phi_p(\mathrm{B})\tilde{z}_t = \theta_q(\mathrm{B})a_t \qquad (3.150)$$

where

$$\phi_p(\mathrm{B}) = 1 - \phi_1 \mathrm{B} - \phi_2 \mathrm{B}^2 - \ldots - \phi_p \mathrm{B}^p$$

$$\theta_q(\mathrm{B}) = 1 - \theta_1 \mathrm{B} - \theta_2 \mathrm{B}^2 - \ldots - \theta_q \mathrm{B}^q$$

For ARMA(p,q) model, there are $p+q+2$ unknown parameters $\mu, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, \sigma_a^2$ that are estimated from the data. In practice, it is frequently true that adequate representation of actually occurring stationary time series can be obtained with autoregressive, moving average or mixed autoregressive-moving average models, in which p and q are not greater than 2.

3.7.4 Autoregressive Integrated Moving Average Model, ARIMA(*p,d,q*)

There are many empirical time series that behave as though they have no fixed mean. However, they exhibit homogeneity in the sense that apart from local level and trend, one part of the series behaves much like any other part.

Models that describe such homogeneous nonstationary behavior can be obtained by supposing some suitable difference of the process to be stationary. We now consider the properties of the important class of models for the dth differences which $d \leq 2$, of a stationary mixed autoregressive-moving average process so that the series will be homogeneous and stationary. These models are called autoregressive integrated moving average (ARIMA) process.

The following table shows the dth difference for $d \leq 2$:

Table 3.6: dth difference for the ARIMA model.

	$d=0$	$d=1$	$d=2$
y_1	y_1		
y_2	y_2	$z_2 = y_2 - y_1$	
y_3	y_3	$z_3 = y_3 - y_2$	$z_3 = y_3 - 2y_2 + y_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_{n-1}	y_{n-1}	$\vdots$	$\vdots$
y_n	y_n	$z_n = y_n - y_{n-1}$	$z_n = z_n - 2z_{n-1} + z_{n-}$

After differentiating the ARMA(p,q) model, the model will be the pth order autoregressive, qth order moving average with dth difference autoregressive integrated moving average, ARIMA(p,d,q). The ARIMA(p,d,q) can be written as

$$\begin{aligned}\varphi(\mathrm{B})z_t &= \phi_p(\mathrm{B})\nabla^d z_t^* \\ &= \theta_q(\mathrm{B})a_t\end{aligned} \tag{3.151}$$

with

$$\nabla^d = (1-\mathrm{B})^d$$

$$\phi_p(\mathrm{B}) = 1 - \phi_1\mathrm{B} - \phi_2\mathrm{B}^2 - \ldots - \phi_p\mathrm{B}^p$$

$$\theta_q(\mathrm{B}) = 1 - \theta_1\mathrm{B} - \theta_2\mathrm{B}^2 - \ldots - \theta_q\mathrm{B}^q$$

where

$\phi_p(\mathrm{B})$ = Autoregressive operator of order p

$\theta_q(\mathrm{B})$ = Moving average operator of order q

∇^d = dth differences

z_t^* = Time series data

a_t = Shock at time t

B = Backward shift operator

3.8 Multivariate Box-Jenkins Model

Multivariate process arise when instead of observing just a single process $X(t)$, we observe simultaneously several processes, $X_1(t), X_2(t), ..., X_n(t)$. For example, in an engineering context we may wish to study the simultaneously variations, over time, of current and voltage, or pressure, temperature and volume, or seismic records taken at a number of different geographical locations. In economics we may be interested in studying inflation rates and money supply, unemployment and interest rates, or the supply and demand of a particular commodity.

Although this would give us some information about each quantity, it could never reveal what might, in fact, be the most important feature of the study, namely, the interrelationships between the various quantities. Just as in probability theory, we cannot examine relationships between random variables knowing only their marginal distributions. Instead we also need to know their joint probability distribution. So, in dealing with multivariate processes we need a framework for describing not only the properties of the individual processes but also the cross-links which may exist between them. This is achieved by introducing the notions of cross-covariance or cross-correlation functions.

3.8.1 Correlation of Multivariate Stationary Processes

To introduce the new ideas involved in the study of multivariate processes we consider first the case of bivariate processes.

Suppose we are given two stochastic processes, $\{X_{1,t}\}, \{X_{2,t}\}, t = 0, \pm 1, \pm 2, \ldots$. We may define the autocovariance functions of $\{X_{1,t}\}, \{X_{2,t}\}$ in the usual way, namely,

$$R_{11}(s) = E\left[\{X_{1,t} - \mu_1\}\{X_{1,t+s} - \mu_1\}\right] \tag{3.152}$$

$$R_{22}(s) = E\left[\{X_{2,t} - \mu_2\}\{X_{2,t+s} - \mu_2\}\right] \tag{3.153}$$

where $\mu_1 = E\lfloor X_{1,t} \rfloor, \mu_2 = E\lfloor X_{2,t} \rfloor$. The cross-covariance function is defined by

$$\begin{aligned} R_{21}(s) &= Cov\{X_{1,t}, X_{2,t+s}\} \\ &= E\left[\{X_{1,t} - \mu_1\}\{X_{2,t+s} - \mu_2\}\right] \end{aligned} \tag{3.154}$$

The autocorrelation functions are then

$$\rho_{11}(s) = R_{11}(s)/R_{11}(0) \tag{3.155}$$

$$\rho_{22}(s) = R_{22}(s)/R_{22}(0) \tag{3.156}$$

and the cross-correlations function is given by

$$\rho_{21}(s) = \frac{R_{21}(s)}{\sqrt{R_{11}(0)}\sqrt{R_{22}(0)}} \tag{3.157}$$

Let $R_{21}(s)$ denotes the cross-covariance function with "$X_{1,t}$ leading $X_{2,t}$". For the sake of symmetry, define the cross-covariance function with "$X_{2,t}$ leading $X_{1,t}$" as

$$R_{12}(s) = E[\{X_{2,t} - \mu_2\}\{X_{1,t+s} - \mu_1\}] \tag{3.158}$$

with $\rho_{12}(s)$ defined analogously to (3.29). Note that the functions $R_{12}(s), R_{21}(s)$ contain equivalent information since, for all s,

$$R_{12}(s) = R_{21}(-s) \tag{3.159}$$

The complete covariance properties of the bivariate process $\{X_{1,t}, X_{2,t}\}$ are then summarized by the sequence of matrices, which is called the covariance matrix of lag s,

$$\boldsymbol{R}(s) = \begin{bmatrix} R_{11}(s) & R_{12}(s) \\ R_{21}(s) & R_{22}(s) \end{bmatrix} \tag{3.160}$$

The correlation matrix of lag s is defined as

$$\boldsymbol{\rho}(s) = \begin{bmatrix} \rho_{11}(s) & \rho_{12}(s) \\ \rho_{21}(s) & \rho_{22}(s) \end{bmatrix} \tag{3.161}$$

If we have n parameter processes, $X_1(t), X_2(t), ..., X_n(t)$, we define the covariance matrix at lag s by

$$\boldsymbol{R}(s) = \begin{bmatrix} R_{11}(s) & R_{12}(s) & \cdots & R_{1n}(s) \\ R_{21}(s) & R_{22}(s) & \cdots & R_{2n}(s) \\ \vdots & \vdots & \ddots & \vdots \\ R_{n1}(s) & R_{n2}(s) & \cdots & R_{nn}(s) \end{bmatrix} \tag{3.162}$$

where

$$R_{ij}(s) = E[\{X_{j,t} - \mu_j\}, \{X_{i,t+s} - \mu_i\}] \tag{3.163}$$

If $\boldsymbol{X}_t$ denotes the column vector,

$$\boldsymbol{X}_t = \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ \vdots \\ X_{n,t} \end{bmatrix}$$

then we may write,

$$\boldsymbol{R}(s) = E[\boldsymbol{X}_{t+s}\boldsymbol{X}_t *] \tag{3.164}$$

where the asterisk denotes both conjugate and transposition, and $\boldsymbol{R}(s)$ clearly has the property

$$\boldsymbol{R}*(s) = \boldsymbol{R}(-s) \tag{3.165}$$

The sample autocovariance function of $X_{i,t}$ and the sample cross covariance is given by

$$\begin{aligned} \hat{R}_{ii}(s) &= \frac{1}{N}\sum_{t=1}^{N-|s|}\left(X_{i,t} - \overline{X}_i\right)\left(X_{i,t+|s|} - \overline{X}_i\right) \\ \hat{R}_{ij}(s) &= \frac{1}{N}\sum_{t=1}^{N-|s|}\left(X_{j,t} - \overline{X}_j\right)\left(X_{i,t+|s|} - \overline{X}_i\right), \qquad s = 0, \pm 1, \ldots, \pm(N-1) \end{aligned} \tag{3.166}$$

with

$$\overline{X}_i = \frac{1}{N}\sum_{t=1}^{N} X_{i,t}$$

3.8.2 Multivariate AR, MA and ARMA Models

In section 3.2 we discussed the three main types of univariate models, namely the autoregressive (AR), moving average (MA) and mixed autoregressive-moving average (ARMA). Each of these models has its corresponding multivariate extension which is obtained by replacing the scalar parameters in the univariate model by matrix parameters.

3.8.2.1 Autoregressive Models

The n-variate AR(p) model is given by

$$\boldsymbol{X}_t + \boldsymbol{a}_1 \boldsymbol{X}_{t-1} + \ldots + \boldsymbol{a}_p \boldsymbol{X}_{t-p} = \boldsymbol{\varepsilon}_t \tag{3.167}$$

where $\boldsymbol{X}_t = \left[X_{1,t}, \ldots, X_{n,t}\right]'$, $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$ are $n \times n$ matrices, and $\boldsymbol{\varepsilon}_t = \left[\varepsilon_{1,t}, \ldots, \varepsilon_{n,t}\right]'$ is a multivariate shock.

For example, for a bivariate AR(2) model, the $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ parameters can be defined as

$$\boldsymbol{a}_1 = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix}, \qquad \boldsymbol{a}_2 = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$$

3.8.2.2 Moving Average Models

The n-variate MA(q) model is

$$\boldsymbol{X}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{b}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{b}_2 \boldsymbol{\varepsilon}_{t-2} + \ldots + \boldsymbol{b}_q \boldsymbol{\varepsilon}_{t-q} \tag{3.168}$$

where $\boldsymbol{b}_1, ..., \boldsymbol{b}_q$ are $n \times n$ matrices. For example, for a bivariate MA(2) model, the $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ parameters can be defined by

$$\boldsymbol{b}_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}, \qquad \boldsymbol{b}_2 = \begin{bmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{bmatrix}$$

3.8.2.3 Mixed Autoregressive-Moving Average Models

The n-variate ARMA(p,q) model is written as

$$\boldsymbol{X}_t + \boldsymbol{a}_1 \boldsymbol{X}_{t-1} + ... + \boldsymbol{a}_p \boldsymbol{X}_{t-p} = \boldsymbol{\varepsilon}_t + \boldsymbol{b}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{b}_2 \boldsymbol{\varepsilon}_{t-2} + ... + \boldsymbol{b}_q \boldsymbol{\varepsilon}_{t-q} \qquad (3.169)$$

or in operator form

$$\boldsymbol{\alpha}(B)\boldsymbol{X}_t = \boldsymbol{\beta}(B)\boldsymbol{\varepsilon}_t \qquad (3.170)$$

where the matrix polynomials $\boldsymbol{\alpha}(B), \boldsymbol{\beta}(B)$ are defined as

$$\boldsymbol{\alpha}(B) = \sum_{u=0}^{p} \boldsymbol{a}_u B^u, \qquad (\boldsymbol{a}_0 = \boldsymbol{I})$$

$$\boldsymbol{\beta}(B) = \sum_{u=0}^{q} \boldsymbol{b}_u B^u, \qquad (\boldsymbol{b}_0 = \boldsymbol{I})$$

with $\boldsymbol{I}$, the identity matrix.

3.8.3 Multivariate Autoregressive Integrated Moving Average Models

For several processes in which after dth differences, $\Delta^d X_{i,t}$, $X_1(t), X_2(t), ..., X_n(t)$ that are nonstationary will be a stationary process. It can then be modeled by the Multivariate Autoregressive Integrated Moving Average, MARIMA model. Thus, writing $Y_{i,t} = \Delta^d X_{i,t}$, where $\Delta = (1 - B)$ denotes the difference operator, we may write

$$\boldsymbol{\alpha}(B)\boldsymbol{Y}_t = \boldsymbol{\beta}(B)\boldsymbol{\varepsilon}_t \tag{3.171}$$

where

$$\boldsymbol{Y}_t = \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ \vdots \\ Y_{n,t} \end{bmatrix}$$

The corresponding model for $\boldsymbol{X}_t$ is

$$(\boldsymbol{I} - \boldsymbol{B})^d \boldsymbol{\alpha}(B)\boldsymbol{X}_t = \boldsymbol{\beta}(B)\boldsymbol{\varepsilon}_t \tag{3.172}$$

where

$$\boldsymbol{B} = \begin{bmatrix} B & 0 & \cdots & 0 \\ 0 & B & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B \end{bmatrix}$$

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

This chapter will discuss in the results obtained from this study. It begins by discussing the results on the identification of the convective rainfalls, followed by the generation of the hourly rainfall series using the stochastic rainfall modeling and proceed with the method of forecasting the short-term rainfall.

4.2 Identification of Convective Rainfall

4.2.1 Diurnal and Monthly Distribution

In order to characterize the convective storms, historical rainfall of 5-min intervals was extracted from the hydrological data bank of the Department of Irrigation and Drainage Malaysia. Station 3117070 – JPS Ampang is chosen because the data sets have relatively good continuity. Only about 0.66 percent of data was missing. The rainfall station is located at North 3° 9’ 20” and 101° 45’ 00” East.

Knowledge of the diurnal cycle of rainfall is important for evaluating convective activity. Previous studies by Ohsawa *et.al.* (2001) on the diurnal variations of convective activity and rainfall in tropical Asia suggests a strong possibility that late night-early morning maxima of convective activity and rainfall have a great effect on energy and water cycles. Figure 4.1 shows the diurnal and monthly distributions of rainfall (greater than 5 mm) in 2004 at Ampang station (3117070). About, 79% of the total rainfall occurred during the daytime (07:00 – 19:00h). The bulk of the rainfall, 75% occurred between 13:00 and 19:00 and 12.5% fall between 19:00 and 22:00. It means that most of rainfall occurred in the afternoon. Convective storms are caused by differential solar heating of the ground and lower air layers, which typically occur during afternoons when warm moist air covers an area (Hewlett, 1969). Consequently, most afternoon rainstorms in this area can be classified as convectional storms.

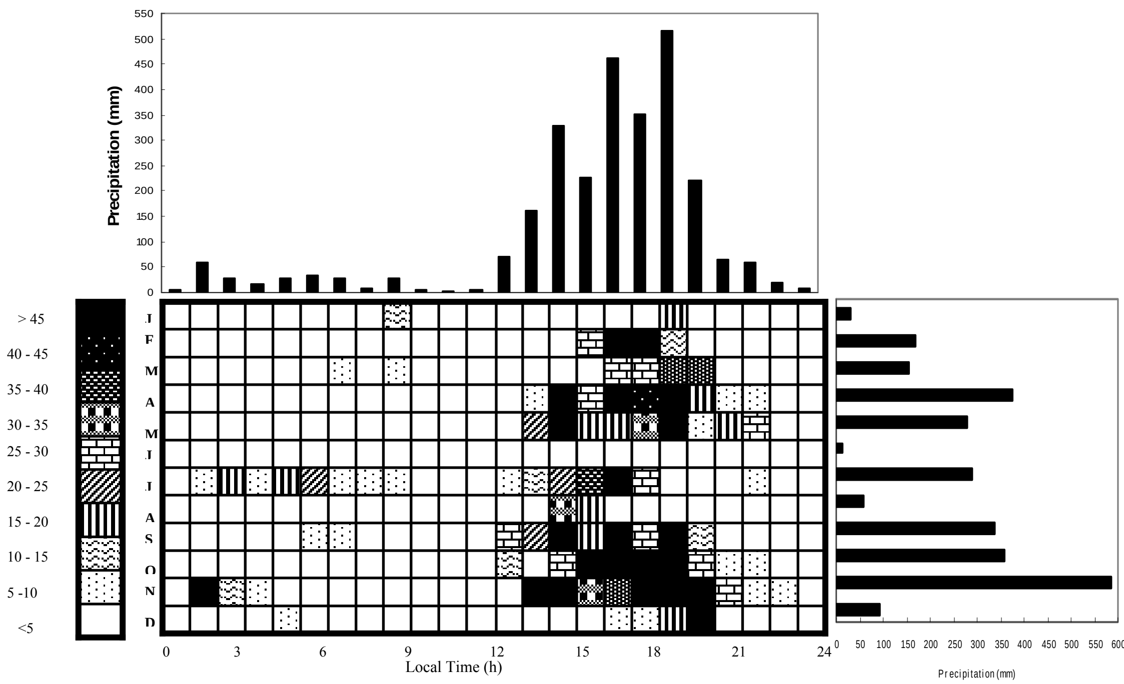


Figure 4.1 : Diurnal and monthly distributions of rainfall (greater than 5 mm) in 2004 at JPS Ampang station

4.2.2 Minimum Interevent Time (MIT)

In this analysis, a rainfall event is defined based on the Minimum Interevent Time (MIT) method. One year rainfall data is used to define this analysis. The annual number of rainfall events were plotted against different MIT values and an appropriate MIT value is selected from the graph at a point after which increases in the MIT do not result in significant changes in the number of event. An MIT value of three hours is chosen. As can be seen from Figure 4.2 after an MIT value of 3, changes in the numbers of events with respect to MIT values has become insignificant. Therefore, rainfall events used in the analysis must were have a minimum separation time of 3 hours. This value can be accepted because Adams *et. al.,* (1986) suggested MIT values between 1 and 6 hours for urban applications**.**

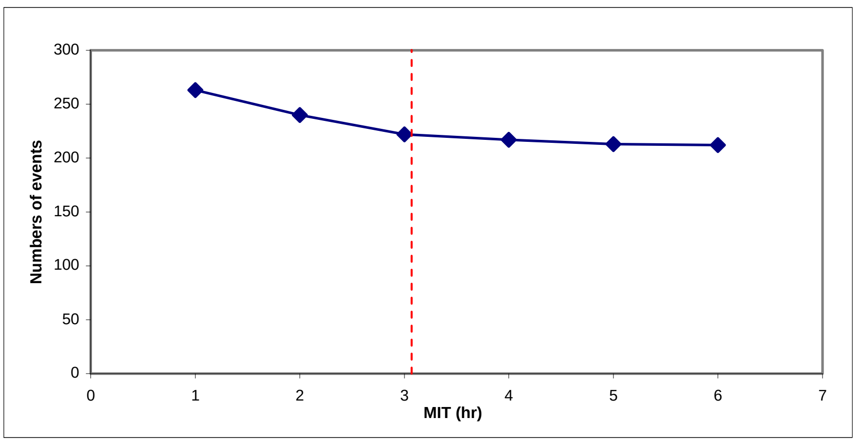


Figure 4.2 : Annual number of rainfall events as a function of MIT

4.2.3 Characterization of Convective Rain Based on Short Rainfall Duration Data

4.2.3.1 Preliminary Analysis

In this stage, the preliminary results on the characteristics of convective and non-convective storms are presented in terms of total rainfall, intensity and duration. Table 4.1 presents the statistical summary of the event rainfalls between year 2000 and 2004. The separation between convective and non-convective storms is based on the 35 mm/hr threshold intensity as described by Llasat (2001). Convective rain occurred most frequently in November (45 times). Of the total 297 convective storm events which exceeded 35 mm/hr, 130 storms or 44% occurred during inter-monsoon months (Oct – Nov and Apr – May). The southwest and northeast monsoons recorded 27% and 30% of the events respectively. This is maybe influenced by inter-monsoon process where during the inter-monsoon period the weather in Malaysia will be typically fair in the morning with strong convective clouds developing in the late morning and early afternoon. Beside that, the wind direction during this period is often variable and the wind speeds seldom exceed 10 knots. The frequency of storms event in different monsoon period is shown in Table 4.2.

Table 4.1 : Summary statistics of monthly convective and non-convective rainfalls between 2000 and 2004 at Ampang station

Precipitation class and totals	*Month*	Northwest				Inter		Southwest				Inter	
		Dec	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**
Nonconvective precipitation (mm) with rate < 35 mm/hr	Total rainfall amounts	696.2	405.6	529.5	686.8	902.1	360.6	409.3	419.1	450.9	834.6	824.1	1312.6
	Mean	139.2	81.1	105.9	137.4	180.4	72.1	81.9	83.8	90.2	166.9	164.8	262.5
	Median	99.9	66.4	66.2	114.8	164.4	80.5	86.0	61.8	50.6	146.6	163.0	263.8
	Standard Deviation	121.2	71.1	123.1	88.0	68.3	31.0	60.0	66.4	90.9	92.3	51.3	97.4
	Coefficient of variation	0.9	0.9	1.2	0.6	0.4	0.4	0.7	0.8	1.0	0.6	0.3	0.4
	Number of event	65	44	39	60	79	38	34	47	46	74	63	99
	Precipitation event-1	10.7	9.2	13.6	11.4	11.4	9.5	12.0	8.9	9.8	11.3	13.1	13.3
Convective precipitation (mm) with rate > 35 mm/hr	Total rainfall amounts	483.1	200.1	331.2	809.0	716.5	396.9	454.6	317.6	309.0	632.1	917.9	883.5
	Mean	96.6	40.0	66.2	161.8	143.3	79.4	90.9	63.5	61.8	126.4	183.6	176.7
	Median	92.7	27.4	61.3	118.8	139.1	52.8	15.7	79.1	27.2	95.9	192.2	239.2
	Standard Deviation	51.0	36.1	24.8	150.9	50.2	60.6	129.3	61.3	70.7	69.9	102.4	121.1
	Coefficient of variation	0.5	0.9	0.4	0.9	0.4	0.8	1.4	1.0	1.1	0.6	0.6	0.7
	Number of event	18	15	22	33	33	16	18	14	17	30	36	45
	Precipitation event-1	26.8	13.3	15.1	24.5	21.7	24.8	25.3	22.7	18.2	21.1	25.5	19.6
Bulk all kinds (mm)	Total rainfall amounts	1179.3	605.7	860.7	1495.8	1618.6	757.5	863.9	764.5	759.9	1466.7	1742.0	2196.1
	Mean	235.9	121.1	172.1	299.2	323.7	151.5	172.8	152.9	152.0	293.3	348.4	439.2
	Median	183.9	93.8	129.0	295.0	351.7	143.9	190.1	140.9	172.1	324.5	359.9	470.5
	Standard Deviation	143.4	102.5	119.7	176.9	90.4	81.3	148.5	111.6	97.7	107.7	77.1	136.3
	Coefficient of variation	0.6	0.8	0.7	0.6	0.3	0.5	0.9	0.7	0.6	0.4	0.2	0.3
	Number of event	83	59	61	93	112	54	52	64	63	104	99	144
	Precipitation event-1	13.4	9.0	16.2	15.9	14.6	13.7	14.3	11.6	11.6	14.0	18.1	15.7

Table 4.2 : Frequency of convective storms events during monsoon and inter-monsoon periods

Monsoon	Frequency	%Frequency
Southwest	79	27
Northeast	88	30
Intermonsoon	130	44

4.2.4.2 Classification of Convective Events

In order to classify convective events, it is useful to have a parameter for each one of them. As noted in Chapter III, an intensity of 35 mm/hr is taken as the 5 minute mean intensity threshold (Llasat, 2001). This threshold is useful in order to derive convective storm properties. Table 4.3 shows the number of non-convective and convective events between 2000 and 2004. In this analysis, it is found that convective events were contribute 30.1% from all of rainfall events whereas non-convective events represent 69.9%. The highest number of convective event was fall in inter-monsoon months where 9 convective events were recorded in November.

Table 4.3 : Number of convective and non convective events

Season	**Northwest**				**Inter-monsoon**		**Southwest**				**Inter-monsoon**	
Month	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Non-convective events (< 35 mm/hr)	13	9	8	12	16	8	7	9	9	15	13	20
Convective events (> 35 mm/hr)	4	3	4	7	7	3	4	3	3	6	7	9

A classification of episodes based on the β parameter is shown in Figures 4.3 and 4.4. This classification is according to their greater or lesser convective character (Llasat, 2001). The number of event which falls under moderately convective class is the highest in all months (Figure 4.3). On a yearly basis the percentage of events trend fall under moderately convective storm range from 51.5 % to 69.3 % (Figure 4.4). All percentages from Figure 4.4 were not include non convective events. Only for event which have intensity greater than 35 mm/hr.

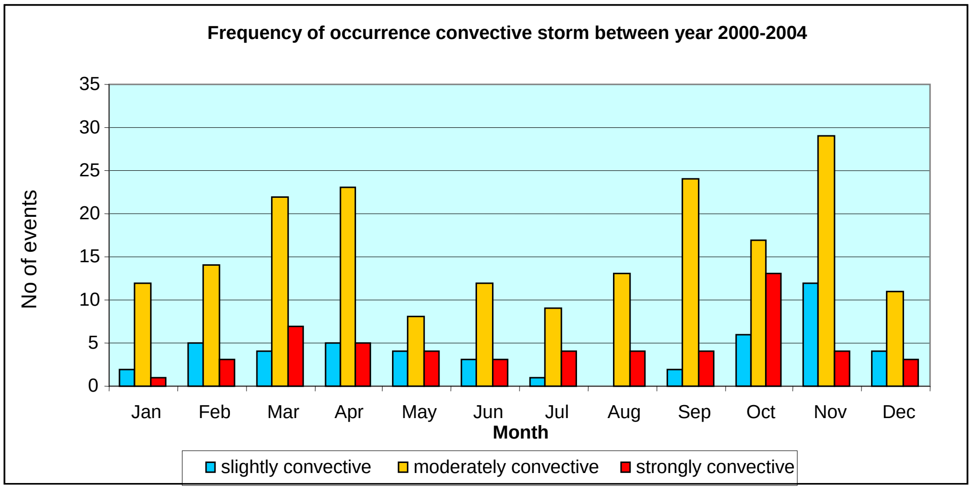


Figure 4.3 : Monthly number of event for each class of convective storm

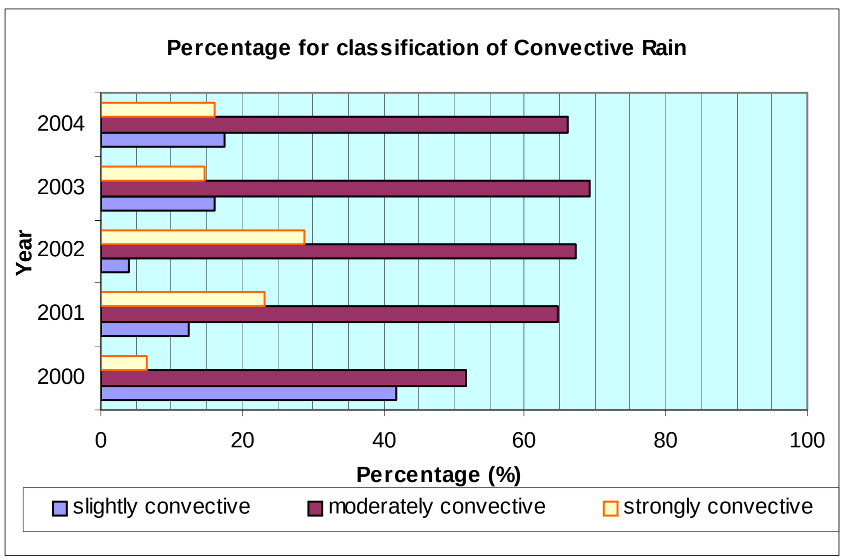


Figure 4.4 : Yearly percentage of occurrence of convective storm

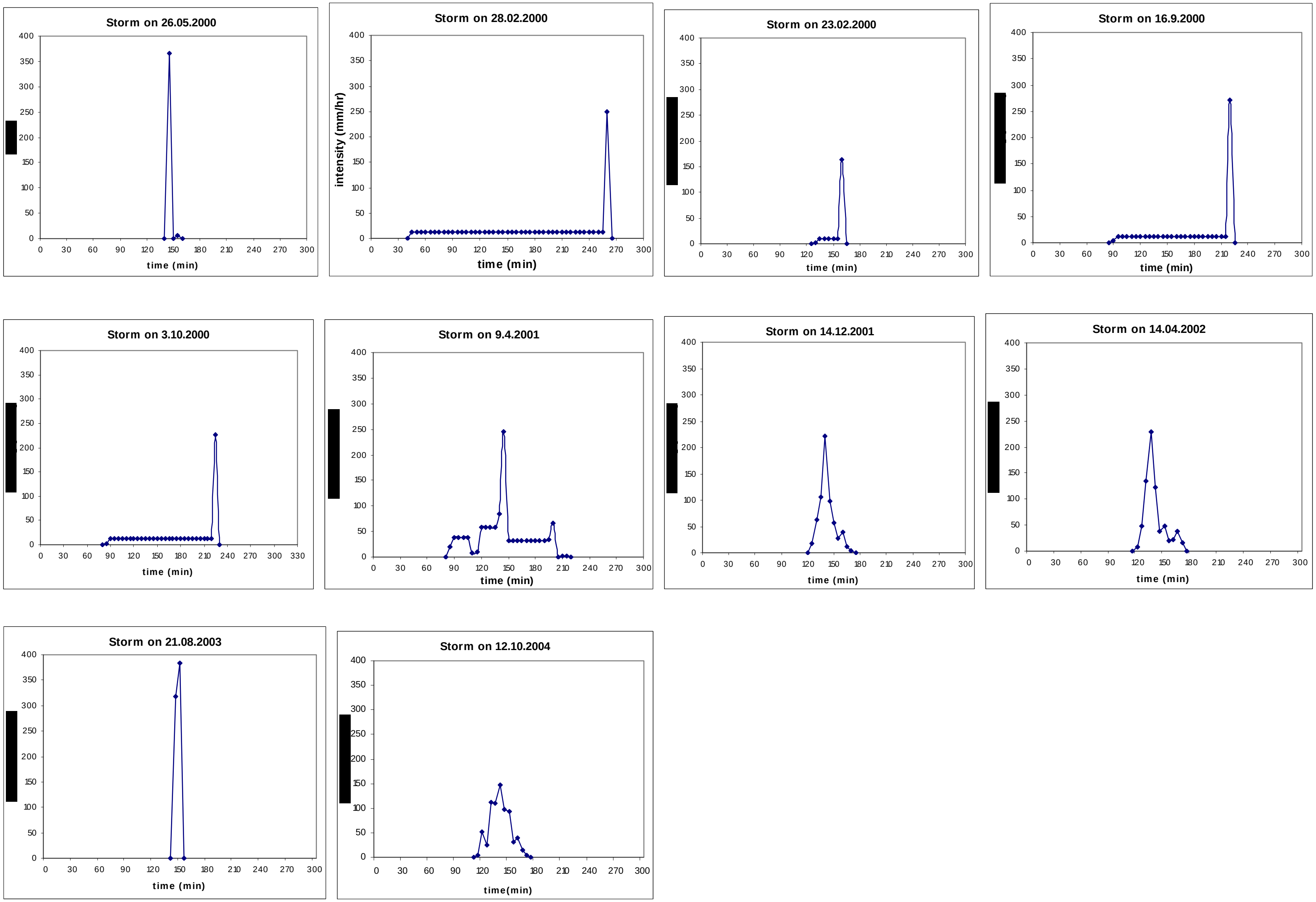


Figure 4.5 : Convective storms with the highest 5 –minutes intensity for each year

4.2.4 Spatial Distribution of Convective Rainfall between Meteorological Radar Data and Surface Data

In this analysis, the comparison of spatial distribution between meteorological radar data and surface rainfall were presented in terms of intensity and the area between isohyetal line. In addition, the movement of storm centre for selected convective events were observed. Finally, the depth-area relationship was plotted for six single events.

4.3.4.1 Digitizing Radar Image

In order to analyse storm areal coverage, the radar images were finest digitized to get a layer of isohyetal contour in GIS format. The real images (JPEG image) from KLIA Meteorological Station were rectified with Klang Valley Map. Then, the colours of rainfall image are digitized one by one until a rainfall contour is produced. Figure 4.6 shows the image of rainfall contour after being digitized using GIS (ArcGIS 9.1).

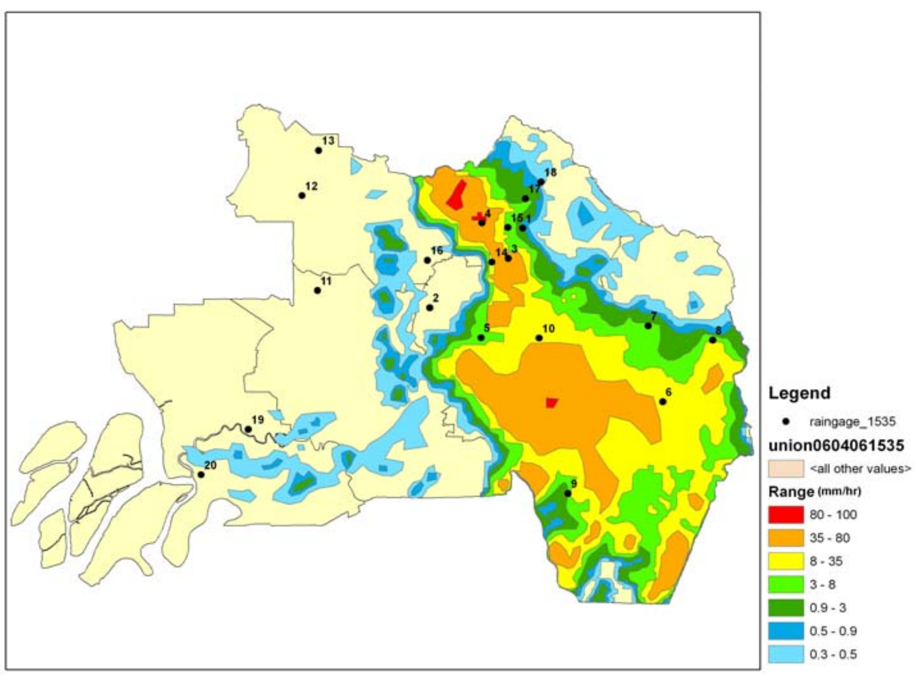


Figure 4.6 : Digitized image using ArcGIS 9.1

4.2.4.2 Comparison on Intensity

A temporal comparison on intensity values between surface rainfall data and meteorological radar data was carried out for selected events. Tables 4.4, 4.5, 4.6 and 4.7 show the rainfall intensity between radar data and surface rainfall of the events. From Table 4.4, four times (18:19, 18:25, 18:30 and 18:36) was selected on January 6, 2006 to compare the rainfall intensity. All of these times were chosen during heavy rainfall. There have four raingauge stations (R4, R5, R12 and R13) were got missing data. Between this comparison, there was no similarity in intensity values from all selected times. Table 4.5 shows event on February 26, 2006 and six times was selected (06:21, 06:32, 06:38, 06:43, 04:55 and 03:23). There was no missing data observed but the case is same with Table 4.6 where differences in intensity value between raingauge and radar are too large. Some intensity values from raingauge are bigger than radar data and vice versa. Two more events April 6 and May 10, 2006 (Tables 4.6 and 4.7) also had shown a bad comparison between raingauge and radar data. Even though many times were chosen to make this comparison, the results were still the same.

Overall, it is observed that both data produced remarkable different in intensity. For a given storm, the radar data can both overestimate or underestimate the surface rainfall. The differences in intensity value between raingauge and radar are too large. The main challenge in getting close approximately between radar rainfall and surface rainfall is the difficulty in establishing the relationship between decibel of, Z-R in unit mm^6/m^3 and rainfall, R in unit mm/hr (Ray et., al 1988). Another factor leading to error is evaporation of precipitation before reaching the ground, which could happen frequently in the tropics. Also, winds may carry precipitation away from beneath the producing cloud. Beside that, the discontinuities in the vertical distribution of precipitation in the cloud affect radar reflectivity and thus are also sources of error (Ray et al., 1988).

In this analysis, the spatial distributions of the rainfall were derived by Kriging Method using intensity data for every raingauge. However, out of four storms, only one event or January 6, 2006 produced smooth circular isohyetal lines. The rainfall contour patterns for this event exhibits very similar patterns with radar data. This storm started at 06.10 pm lasted about two hours. Figure 4.7 comprises the spatial distribution between Kriging and the observed radar data for event on January 6, 2006 at different times. This storm also shows increasing intensity as the storm centre moves from the northeast to the southwest. However, the other three storms, fail to show good agreement between radar and raingauge data (February 26, April 6 and May 10).

Table 4.4 : Surface and radar rainfall intensity on January 6th 2006

	Raingages	Time		18:19		18:25		18:30		18:36	
		Latitude	Longitude	RG	RDR	RG	RDR	RG	RDR	RG	RDR
R1	3217001-KM16 Gombak	3.2680	101.7291	0	0.6	0	0.5	0	0.8	6	2.0
R2	3116006-Ldg Edinburgh Site 2	3.1833	101.6333	0	no rain	5	1.5	0	10	0	20.0
R3	3217003-KM11 Gombak	3.2361	101.7139	0	0.5	0	0.7	0	1	6	0.8
R4	3216001-Kg Sg Tua	3.2722	101.6861	?	0.5	?	no rain	?	0.6	?	0.7
R5	3116003-JPS Msia	3.1514	101.6847	?	no rain	?	no rain	?	2	?	8.0
R6	3018101-Emp. Semenyih	3.0856	101.8892	0	4	0	0.8	0	1.5	0	1.5
R7	3118102-SK Kg Lui	3.1736	101.8722	21	0.5	21	0.9	4	3	1	2.0
R8	311104-Jln Genting Peres	3.1403	101.9297	4.8	1	4.8	2	8.4	4	3.6	9.0
R9	2917001-JPS Kajang	2.9917	101.7972	0	0.9	0	0.3	0	no rain	0	0.7
R10	3117070-JPS Ampang	3.1556	101.7500	0	no rain	7.2	0.3	7.2	3	8.4	7.0
R11	3115079-Pusat Penyldkn Sg Buloh	3.1583	101.5597	22.8	20	22.8	20	52.8	35	50.4	25.0
R12	3315037-Tmn Bkt Rawang	3.3014	101.5008	?	35	?	20	?	20	?	5.0
R13	3315038-Country Homes	3.0167	101.5022	?	0.9	?	no rain	?	no rain	?	0.7
R14	3217004-Kg Kuala Sleh	3.2583	101.7903	6	1	6	0.3	0	0.7	0	0.8
R15	3217002-Emp. Genting Klang	3.2361	101.7528	0	no rain	0	0.6	6	0.5	0	2.0
R16	3216004-SMJK Kepong	3.2319	101.6361	0	15	0	20	0	10	0	0.8
R17	3317001-Air Terjun Sg Batu	3.3347	101.7042	6	3	0	3	0	1.5	0	2.0
R18	3317004-Genting Sempah	3.3681	101.7708	0	2	0	2	0	3	0	0.7
R19	3014091-UiTM Shah Alam	3.0022	101.4019	15.6	2	10.8	1	8.4	1.5	79.2	6.0
R20	3014084-JPS Klang	3.0389	101.4444	0	no rain	0	no rain	1.2	0.4	1.2	0.3

?

= missing data

RG = rain gauge

RDR = radar

Table 4.5 : Surface and radar rainfall intensity on February 26th 2006

		Time		6:21		6:32		6:38		6:43		4:55		3:23	
	Raingages	**Latitude**	**Longitude**	**RG**	**RDR**	**RG**	**RDR**	**RG**	**RDR**	**RG**	**RDR**	**RG**	**RDR**	**RG**	**RDR**
R1	3217001-KM16 Gombak	3.2680	101.7291	12	9	18	6	18	4	18	6	48	0.4	0	0.3
R2	3116006-Ldg Edinburgh Site 2	3.1833	101.6333	0	no rain	0	no rain	0	0.3	0	0.4	20	6	5	20
R3	3217003-KM11 Gombak	3.2361	101.7139	0	9	0	2	0	1.5	6	3	12	0.6	0	3
R4	3216001-Kg Sg Tua	3.2722	101.6861	6	6	24	6	24	15	12	15	0	0.6	48	65
R5	3116003-JPS Msia	3.1514	101.6847	0	2	6	0.8	0	0.3	0	no rain	6	1.5	6	0.9
R6	3018101-Emp. Semenyih	3.0856	101.8892	0	no rain	0	no rain	0	0.8	0	4	0	4	0	no rain
R7	3118102-SK Kg Lui	3.1736	101.8722	0	no rain	33	0.9	28	10	28	8	0	no rain	0	no rain
R8	311104-Jln Genting Peres	3.1403	101.9297	0	no rain	0	0.5	0	no rain	21.6	no rain	0	no rain	0	no rain
R9	2917001-JPS Kajang	2.9917	101.7972	0	no rain	0	no rain	0	no rain	0	no rain	0	5	0	no rain
R10	3117070-JPS Ampang	3.1556	101.7500	50.4	20	19.2	5	6	3	3.6	4	1.2	0.5	1.2	no rain
R11	3115079-Pt Penyldkn Sg Buloh	3.1583	101.5597	0	no rain	0	no rain	0	no rain	0	no rain	0	4	18	2
R12	3315037-Tmn Bkt Rawang	3.3014	101.5008	4	0.8	0	0.5	0	0.3	0	0.3	0	no rain	25	50
R13	3315038-Country Homes	3.0167	101.5022	1	0.9	0	0.8	0	0.6	0	0.5	0	0.3	6	1.5
R14	3217004-Kg Kuala Sleh	3.2583	101.7903	30	4	6	1.5	6	0.7	12	1.5	0	0.6	0	7
R15	3217002-Emp. Genting Klang	3.2361	101.7528	30	6	18	1.5	6	9	6	20	0	1.5	0	0.6
R16	3216004-SMJK Kepong	3.2319	101.6361	6	0.4	6	1	6	1	6	0.4	6	15	6	50
R17	3317001-Air Terjun Sg Batu	3.3347	101.7042	18	2	6	6	48	5	42	5	0	0.4	0	no rain
R18	3317004-Genting Sempah	3.3681	101.7708	12	0.8	6	0.8	6	1.5	6	1	6	no rain	0	no rain
R19	3014091-UiTM Shah Alam	3.0022	101.4019	0	no rain	0	no rain	0	no rain	0	no rain	7.2	0.8	16.8	2
R20	3014084-JPS Klang	3.0389	101.4444	0	0.7	0	no rain	0	no rain	0	no rain	0	0.5	0	no rain

? = missing data

RG = rain gauge

RDR = radar

Table 4.6 : Surface and radar rainfall intensity on April 6th 2006

	Raingages	Time		15:08		15:13		15:19		15:29		15:35		15:41	
		Latitude	Longitude	RG	RDR	RG	RDR	RG	RDR	RG	RDR	RG	RDR	RG	RDR
R1	3217001-KM16 Gombak	3.2680	101.7291	72	65	42	7	12	6	6	0.3	6	0.3	0	0.3
R2	3116006-Ldg Edinburgh Site 2	3.1833	101.6333	5	0.5	15	0.8	5	no rain	5	no rain	5	no rain	0	no rain
R3	3217003-KM11 Gombak	3.2361	101.7139	0	0.4	0	0.3	0	0.4	12	3	90	15	48	9
R4	3216001-Kg Sg Tua	3.2722	101.6861	108	50	54	65	30	35	24	50	18	50	12	10
R5	3116003-JPS Msia	3.1514	101.6847	0	0.9	6	6	6	15	24	35	24	6	12	9
R6	3018101-Emp. Semenyih	3.0856	101.8892	0	0.5	0	2	0	20	0	15	0	15	0	7
R7	3118102-SK Kg Lui	3.1736	101.8722	0	0.6	11	no rain	1	no rain	1	0.7	0	0.4	0	0.6
R8	311104-Jln Genting Peres	3.1403	101.9297	1.2	1	1.2	1	4.8	0.7	25.2	1.5	20.4	2	6	0.9
R9	2917001-JPS Kajang	2.9917	101.7972	?	no rain	?	no rain	?	4	?	50	?	7	?	1.5
R10	3117070-JPS Ampang	3.1556	101.7500	2.4	15	3.6	35	7.2	20	10.8	35	8.4	35	14.4	35
R11	3115079-Pt Penyldkn Sg Buloh	3.1583	101.5597	0	0.4	0	0.3	0	no rain	0	no rain	0	no rain	0	no rain
R12	3315037-Tmn Bkt Rawang	3.3014	101.5008	0	no rain	0	no rain	0	no rain	0	no rain	0	no rain	5	no rain
R13	3315038-Country Homes	3.0167	101.5022	0	no rain	0	no rain	0	no rain	0	no rain	0	no rain	0	0.3
R14	3217004-Kg Kuala Sleh	3.2583	101.7903	0	no rain	0	no rain	0	0.6	0	50	0	35	0	65
R15	3217002-Emp. Genting Klang	3.2361	101.7528	?	65	?	50	?	20	?	1	?	1	?	0.9
R16	3216004-SMJK Kepong	3.2319	101.6361	0	no rain	0	0.4	0	no rain	0	0.4	0	no rain	0	no rain
R17	3317001-Air Terjun Sg Batu	3.3347	101.7042	12	6	6	3	0	1.5	0	0.5	0	0.4	0	0.5
R18	3317004-Genting Sempah	3.3681	101.7708	0	0.5	6	no rain	0	no rain	0	0.6	0	no rain	0	no rain
R19	3014091-UiTM Shah Alam	3.0022	101.4019	0	no rain	0	no rain	0	no rain	0	no rain	0	no rain	0	no rain
R20	3014084-JPS Klang	3.0389	101.4444	0	no rain	0	no rain	0	no rain	0	no rain	0	no rain	0	no rain

? = missing data

RG = rain gauge

RDR = radar

Table 4.7 : Surface and radar rainfall intensity on May 10th 2006

	Raingages	Time		15:01		15:12		15:28		15:33		15:39	
		Latitude	Longitude	RG	RDR	RG	RDR	RG	RDR	RG	RDR	RG	RDR
R1	3217001-KM16 Gombak	3.2680	101.7291	?	0.8	?	80	?	35	?	65	?	50
R2	3116006-Ldg Edinburgh Site 2	3.1833	101.6333	45	20	20	35	0	50	0	35	0	50
R3	3217003-KM11 Gombak	3.2361	101.7139	?	no rain	?	6	?	35	?	15	?	7
R4	3216001-Kg Sg Tua	3.2722	101.6861	102	6	66	9	6	2	0	4	6	7
R5	3116003-JPS Msia	3.1514	101.6847	90	65	20	65	10	50	10	15	10	7
R6	3018101-Emp. Semenyih	3.0856	101.8892	?	50		20	?	2	?	2	?	0.9
R7	3118102-SK Kg Lui	3.1736	101.8722	0	no rain	0	0.5	0	35	0	5	0	5
R8	311104-Jln Genting Peres	3.1403	101.9297	0	1	0	35	0	6	0	0.8	0	3
R9	2917001-JPS Kajang	2.9917	101.7972	15	15	10	9	0	1.5	0	0.7	0	no rain
R10	3117070-JPS Ampang	3.1556	101.7500	21.6	15	42	0.4	0	0.3	0	no rain	0	0.3
R11	3115079-Pusat Penyldkn Sg Buloh	3.1583	101.5597	0	no rain	0	0.3	23	0.3	5	1	11	2
R12	3315037-Tmn Bkt Rawang	3.3014	101.5008	25	no rain	5	no rain	5	no rain	5	no rain	7	no rain
R13	3315038-Country Homes	3.0167	101.5022	0	15	0	4	1	no rain	3	no rain	2	no rain
R14	3217004-Kg Kuala Sleh	3.2583	101.7903	0	80	0	20	0	65	0	10	0	7
R15	3217002-Emp. Genting Klang	3.2361	101.7528	0	20	0	80	0	35	6	35	24	50
R16	3216004-SMJK Kepong	3.2319	101.6361	?	20	?	7	?	0.3	?	0.3	?	0.3
R17	3317001-Air Terjun Sg Batu	3.3347	101.7042	0	0.3	0	20	12	65	18	50	36	50
R18	3317004-Genting Sempah	3.3681	101.7708	0	no rain	0	0.3	0	20	6	7	12	9
R19	3014091-UiTM Shah Alam	3.0022	101.4019	0	no rain	0	no rain	0	no rain	0	no rain	0	no rain
R20	3014084-JPS Klang	3.0389	101.4444	0	no rain	0	no rain	0	no rain	0	no rain	0	no rain

? = missing data

RG = rain gauge

RDR = radar

Most of the isohyetal lines derived from raingauge data are not smooth as those derived from digitized images. Moreover, there was no similarity in the spatial distributions between the radar and surface rainfall. This might be due to the small number of raingauge station employed in the study and further complicated by the occurrence of missing data for some of the events. Kriging methods require a large number of rainfall stations to produce smooth curves. Prediction errors tend to be larger in areas with small number of station. Beside the small number of rainfall station, the discrepancies arise from rainfall data but it also the way Doppler radar estimate rainfall intensity. Doppler radar does not determine actual rainfall intensity, but only areas of returned energy. It means the energy that is reflected back toward the radar (National Weather Service, 2006). The more intense the precipitation, the greater the reflectivity (Ray et al., 1988). Figures 4.8, 4.9 and 4.10 show the spatial distribution of rainfall on February 26, April 6, and May 10, 2006.

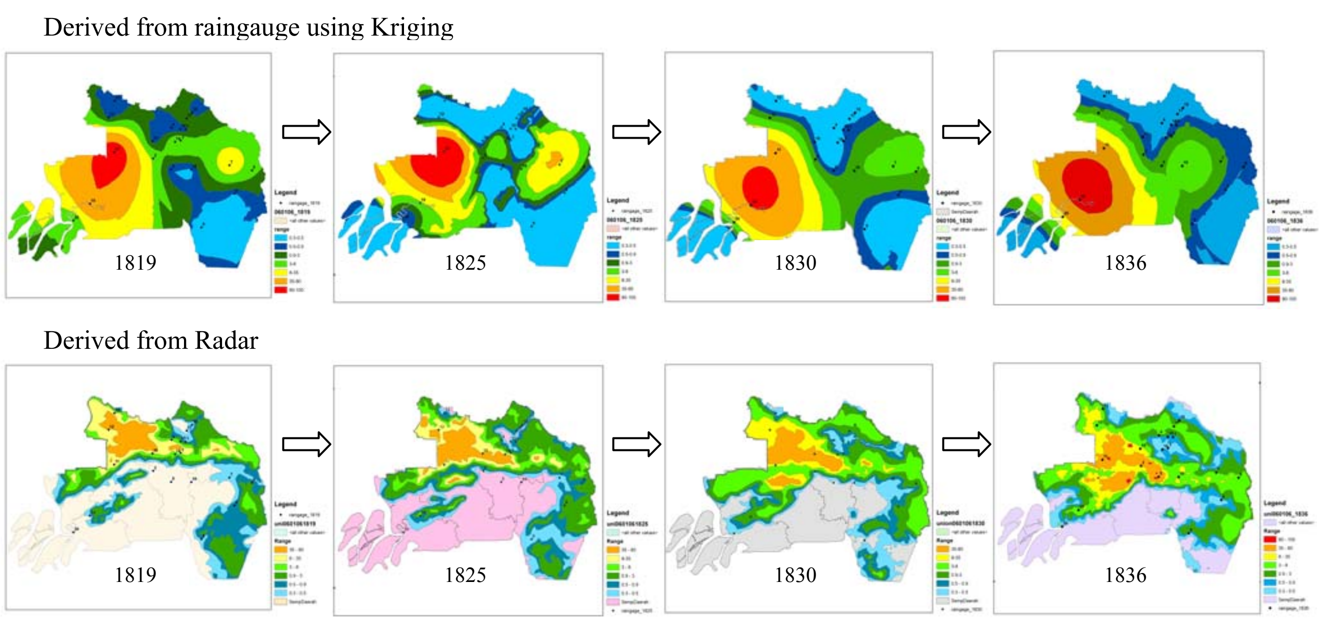


Figure 4.7 : Comparison of rainfall distribution derived from raingauge and radar for event on January 6,2006

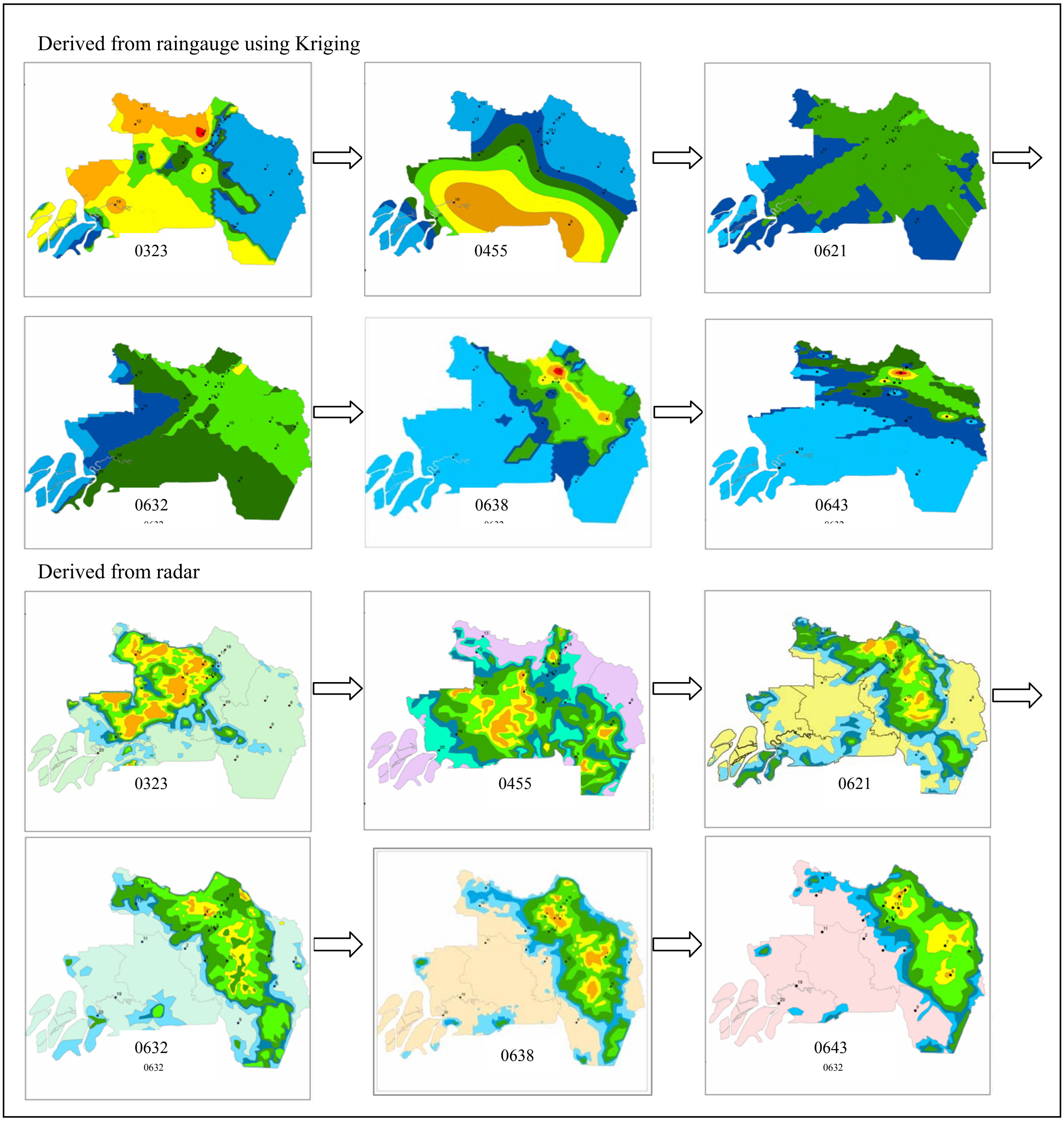


Figure 4.8 : Comparison of rainfall distribution derived from raingauge and radar for event on February 26,2006

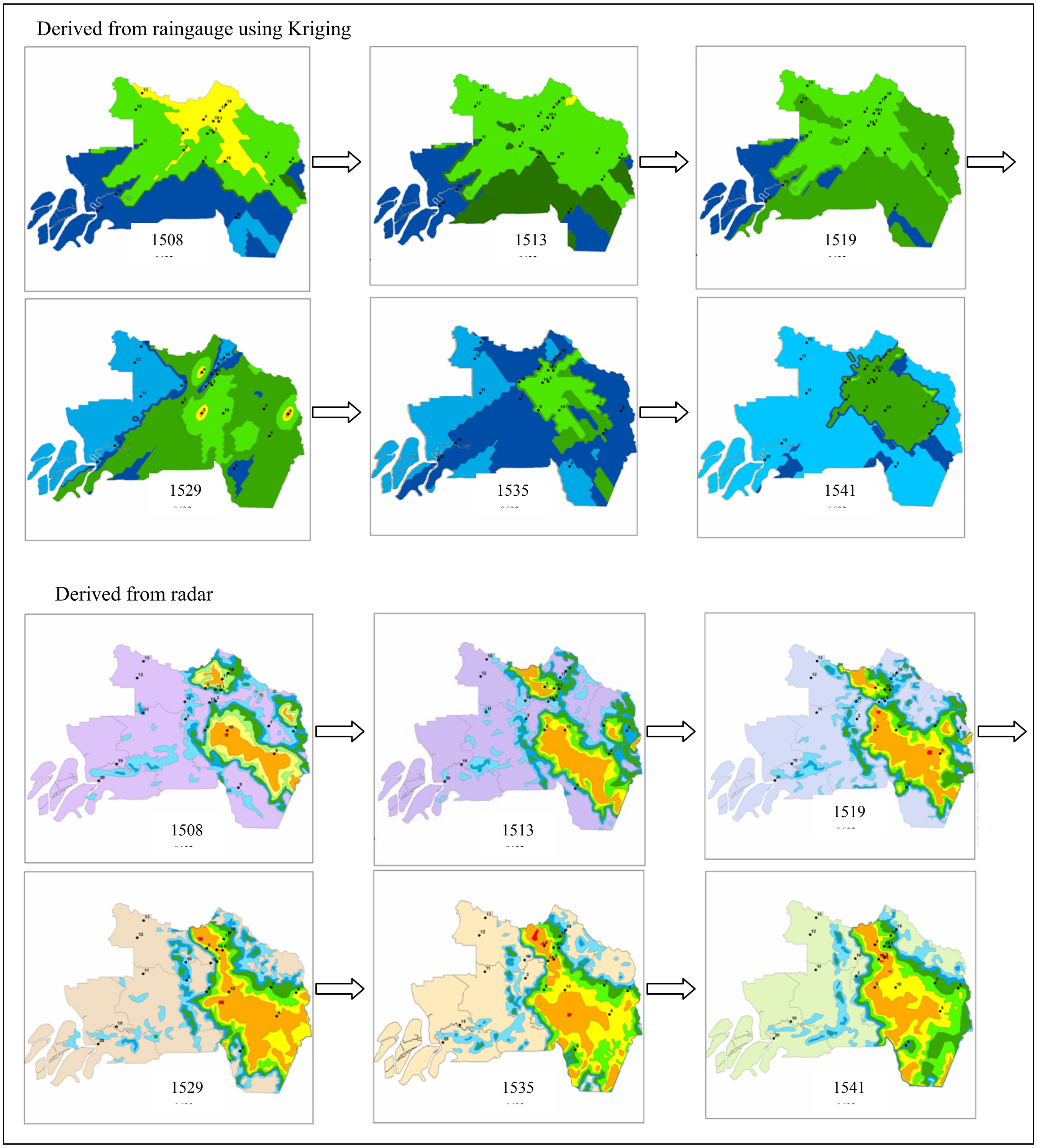


Figure 4.9 : Comparison of rainfall distribution derived from raingauge and radar for event on April 6,2006

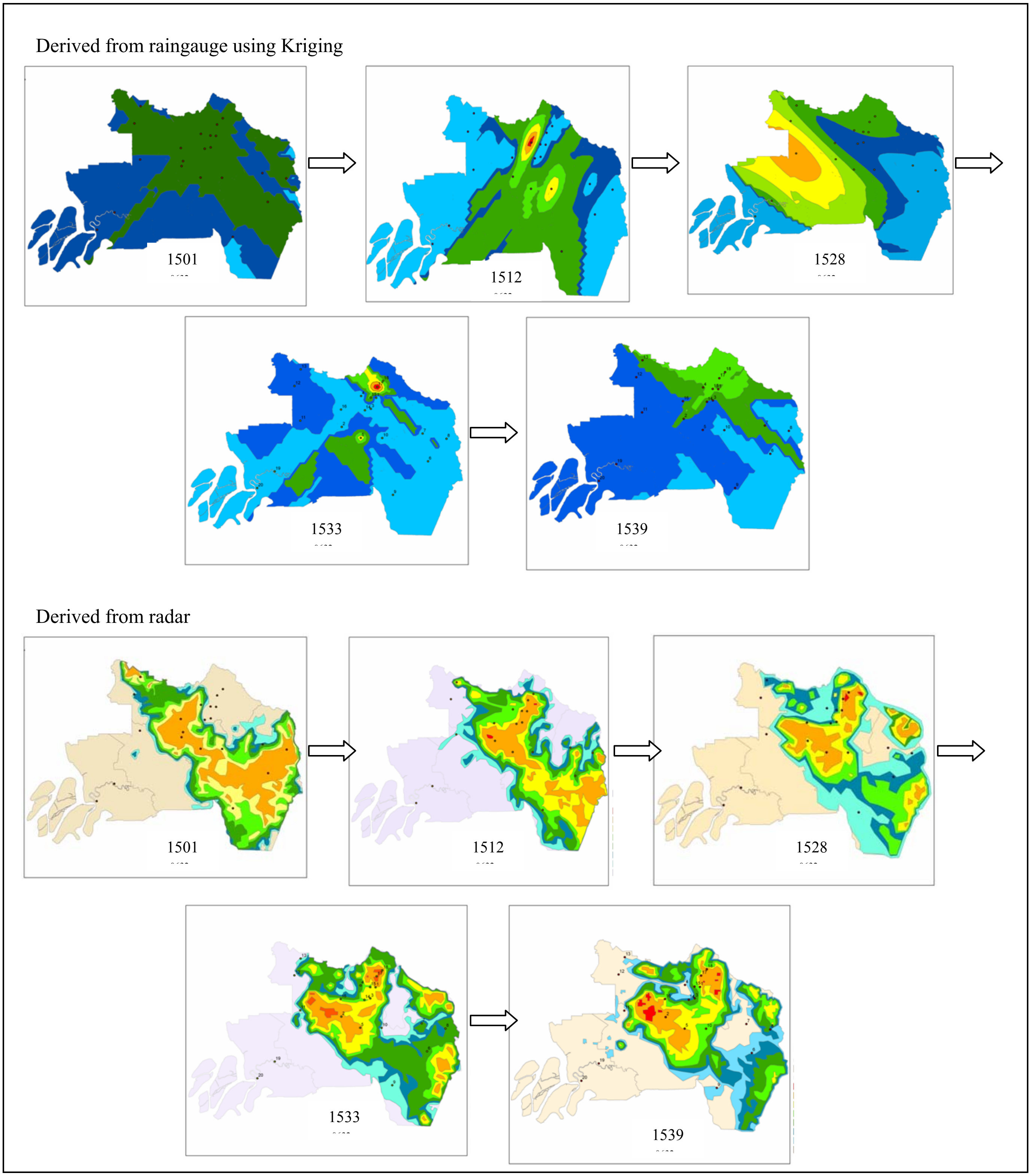


Figure 4.10 : Comparison of rainfall distribution derived from raingauge and radar for event on May 10, 2006

4.3.5 Comparison of Area Rainfall between Radar and Surface Rainfall

A comparison on the areal rainfall derived from radar and surface rainfall is computed using GIS software (ArcGIS 9.1). The colour represents the intensity level. The analysis used four selected storms. Three of the storms analysed occurred in the afternoon. Table 4.8 compares the areal coverage of rainfall intensity derived from radar against those from raingauge. For event on January 6, 2006, the heaviest rainfall was detected at 18:36 pm. Both of centre of the storms occurred in the western part of Klang Valley. The area distribution between radar and surface rainfall is different. The area of centre of the storm for rainfall contour derived from raingauge is bigger than those derived from radar (red colour). This might be due to the number of raingauge station is small and rainfall data which is recorded the highest rainfall amount is less. From twenty rainfall data which is recorded from twenty raingauge station, only one raingauge (R19) shows the highest intensity compared to the others with value of 79.2 mm/hr (red colour). This situation was made the interpolation process in Kriging did not produce smooth rainfall contours as those derived from digitized images (radar). This is also caused the centre of the storm was not captured accurately by ground data.

Comparison of area distribution for event on February 26, 2006 was taken at 04:55 am. It is indicated that the highest intensity was within 35 – 80 mm/hr. The area distribution still differs between ground data and radar data. Most area of rainfall contour from ground data was bigger than those derived from radar. This situation might be same with event on January 6, 2006 where the number of raingauge station is small and rainfall data which is recorded the highest rainfall amount is less. This is caused the centre of the storm was not captured accurately by ground data. The highest intensity at this time is only 48 mm/hr and that is why both of rainfall contours were show that orange colour in each image as the highest rainfall amount in that area. The total rainfall at this moment is 8.9 mm. From surface rainfall data, only 8 raingauge stations were recorded rainfall amount.

For event on April 6, 2006, there have two centres of the storm (red colour) with intensity within 80 – 100 mm/hr at 15:29 pm. From Table 4.8, it is indicated that low intensities were give a bigger area compared to high intensities from surface rainfall data. This is might be influenced by ground data where no high intensity value is recorded at this moment. From rainfall contour which is derived from surface rainfall (Figure 4.11), there have three centres of the storm in that area. All of centres of the storm were occurred at raingauges numbered R4, R5 and R8 with intensity values of 24, 24 and 25.2 mm/hr respectively. Both of these spatial distributions were give a different result. It seems that radar shows more accurate than surface rainfall. This might be due to the effectiveness of radar detecting rainfall area. The colours of radar represent the values of energy reflected toward the radar. The higher the dBZ, the stronger the rain intensity. Beside that, only eight raingauges was recorded rainfall intensity. This is also might be one of factor that why rainfall contour from ground cannot capture accurately. This is because contour from ground needs more raingauge stations to interpolate in Kriging. Wind also could be one factor. Wind can bring rain far from the location where it is start to fall.

Event on May 10, 2006 is quiet similar with event on April 6, 2006. There have only one centre of the storm in ground contour but in radar contour shows two centres of the storm at 15:12 pm. Low intensities were giving a bigger area than high intensities. The location of centre of the storm between both of contours is also different. Rainfall contour derived from radar shows more accurate compared to those derived from surface rainfall. Beside that, only six raingauges was recorded rainfall amount. This is caused the interpolation process cannot give a smooth rainfall contour because the more data used for interpolation, the better contour can be produced. Figure 4.11 comprises the area distribution between radar and surface rainfall for four selected storms.

Overall, it is evident that the two analyses produced remarkably different results. Such discrepancies could be due to interpolation process in Kriging Method where the procedure of spatial interpolation require an estimate of unknown values of a variable at unsampled points by using measured values from other points (Weise, 2001). Moreover, a few raingauges had missing data. This has worsen the interpolation process in Kriging compared to the digitized images (radar). Beside that, it is not only due to the raingauge data but other factor also impacted to this result. Another factor leading to error is evaporation of precipitation before reaching the ground, which happen frequently in tropics. Also winds may carry precipitation away from beneath the producing cloud. All of these are sources of error.

Table 4.8 : Areal distribution of storm intensity obtained from radar and raingauge

Date	6-Jan-06		26-Feb-06		6-Apr-06		10-May-06	
Time	18:36		4:55		15:29		15:12	
Intensity (mm/hr)	Area (km^2)		Area (km^2)		Area (km^2)		Area (km^2)	
	Radar	Raingauge	Radar	Raingauge	Radar	Raingauge	Radar	Raingauge
0.3-0.5	309.86	767.68	463.11	893.28	303.83	765.27	213.81	1270.45
0.5-0.9	277.37	560.18	408.87	331.33	159.15	223.4	189.88	375.71
0.9-3.0	457.4	425.49	539.74	306.71	167.55	1423.71	237.34	999.32
3.0-8.0	555.11	206.00	370.48	411.08	128.86	408.68	239.36	151.87
8.0-35	234.24	285.05	202.63	500.26	240.51	29.07	303.98	44.42
35-80	186.24	549.19	94.90	413.16	362.60	5.42	284.56	11.11
80-100	5.76	62.24	0.00	0.00	3.03	0.28	2.38	2.95

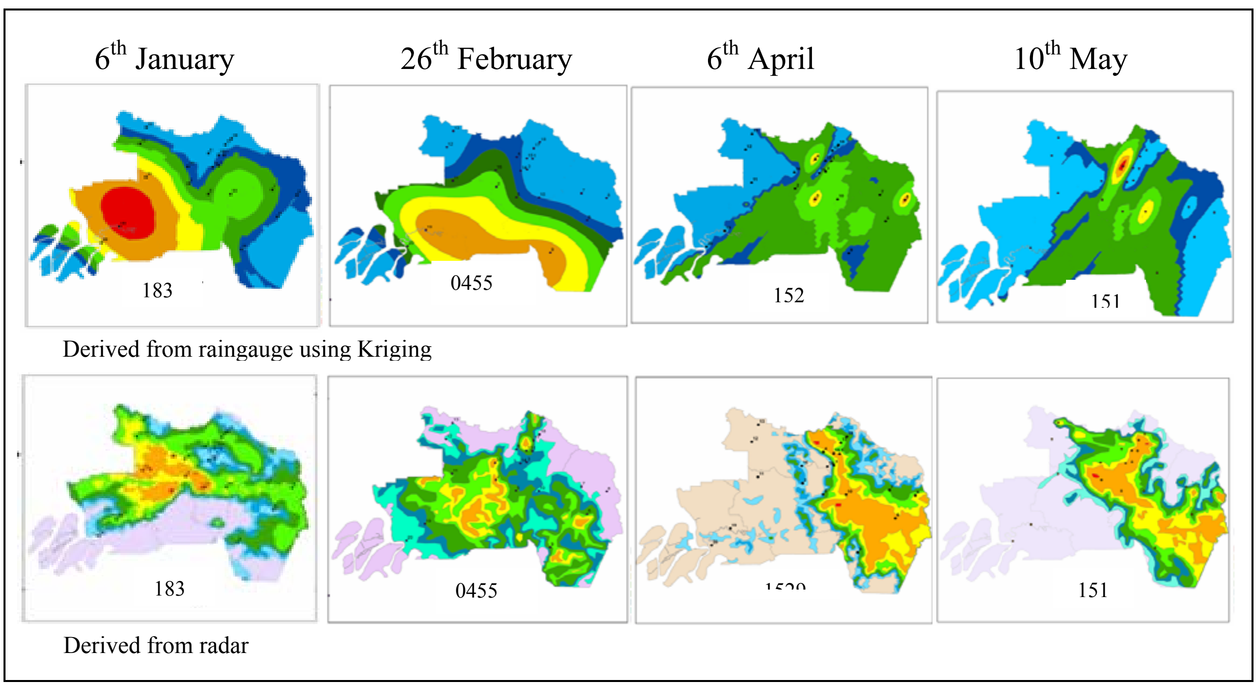


Figure 4.11 : Comparison of areal distribution of intensity between surface rainfall and radar

4.3.6 Storm Movement

In is interesting to investigate the movement pattern of convective storms by tracking the centre of the storm. It is known that an area situated in the tropics experiences predominantly convective precipitation, which is an active component of the tropical weather system (Hastenrath, 1991). Two features of storms which receive attention from researchers are the velocity and direction storm cells movement. It was found that the storm velocities and directions change seasonally (Niemczynowicz and Dahlblom 1984; Chaudry et. al., 1994). The movement and intensity of convective storm are important to predict the magnitude and location of flash flood (Doswell *et. al*., 1996). This section is to investigate what are indicators and predictors were in the evolution and movement of convective storms resulting in heavy rainfall, and the reliability of radar retrieved rainfall data to improve very short-range forecasts. In this analysis, four flash flood events that had occurred in the Klang Valley were chosen. The storms bringing rains leading to the flash floods had exhibited convective characters. These events also are a good example of unusually strong convective events responsible for heavy rainfall. Radar images were used to perform this analysis. Figures 4.12, 4.13, 4.14 and 4.15 illustrate the storm movement for the events.

Pascual et al., (2004) used 30 to 45 dBZ to differentiate convective and stratiform precipitation. On the other hand, Rigo and Llasat (2002) used 43 dBZ to analyse convective event derived from meteorological radar. Whilst Dong and Hyung (2000) used 35 dBZ in study of heavy rainfall with mesoscale convective systems over the Korean Peninsular. In this study a value of 35 dBZ is taken as reflectivity threshold to identify convective rainfall from radar images. This value also corresponds with the radar's rate, thus ease the reading the reflectivity according to radar's colour code. The highest reflectivity, (> 35 dBZ) is chosen as centre of the storm. The centre of the storm is used track the movement of the storms (Figures 4.12, 4.13, 4.14 and 4.15). The coordinates of storm movement were then plotted in Malaysia's RSO (Rectified Skew Ortomorphic), which is a coordinate system in GIS (ArcGIS 9.1). Tables 4.9 and 4.10 present the coordinates of the storm centre and the corresponding reflectivity values. For storm on January 6, 2006, the storm centre developed at 18:03 hr with reflectivity of 65 dBZ or 90 mm/hr. This storm exhibited decreasing reflectivity as it move from northeast to the southwest (Figure 4.12). The duration of this movement was 1 hour and 5 minutes. The storm on February 26, 2006 moved from northwest to southeast and the storm centre at 03:39 hr (Figure 4.13). The storm duration was 1 hour and 16 minutes until the centre of the storm disappeared. Initially, the reflectivity was 65 dBZ or 90 mm/hr and decreased to 35 dBZ until the storm ceased.

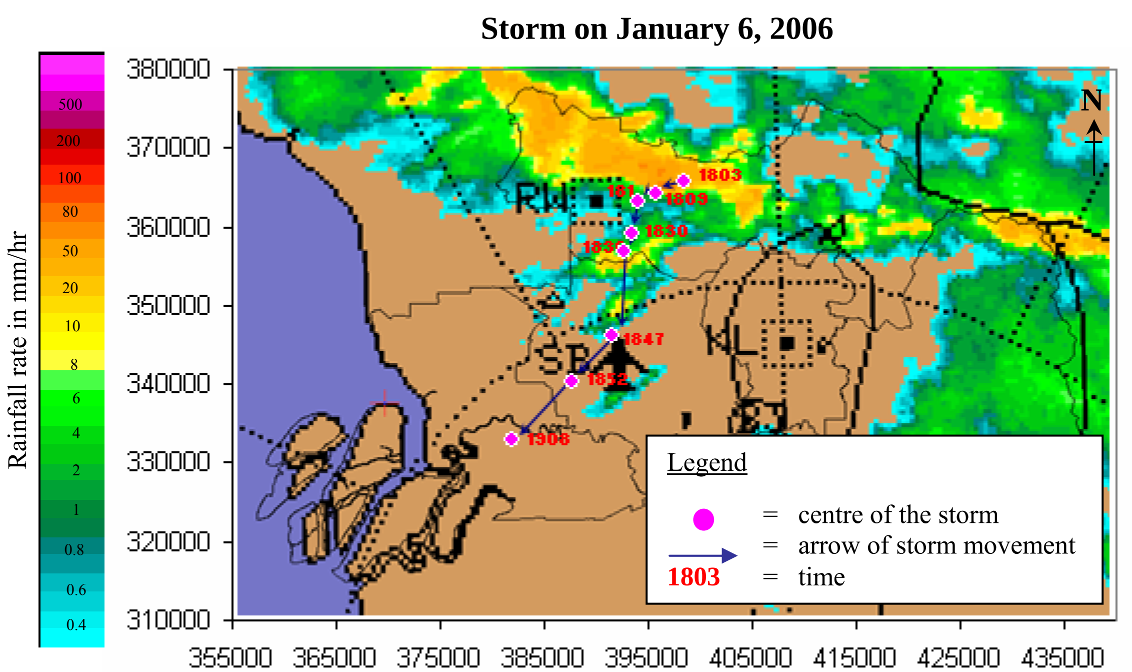


Figure 4.12 : Storm movement on January 6, 2006

Storm on February 26, 2006

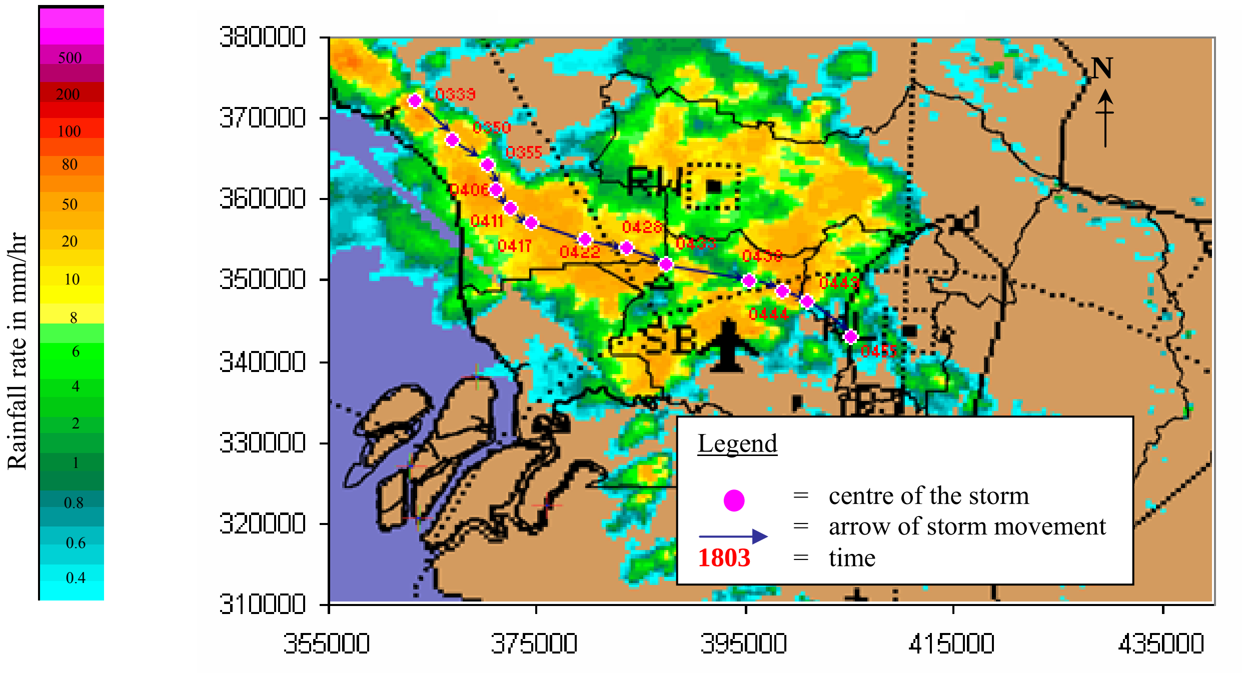


Figure 4.13 : Storm movement on February 26, 2006

Table 4.9 : The coordinates and intensity of storm centres on 6.01.2006 and 26.02.2006

No	6-Jan-06					26-Feb-06				
	Time	Coordinate x	Coordinate y	dBZ	mm/hr	Time	Coordinate x	Coordinate y	dBZ	mm/hr
1	18:03	403611.86	366344.86	65	90	3:39	363432.7	371967.6	65	90
2	18:09	395780.33	364193.73	65	90	3:50	366902.1	367303.8	65	90
3	18:14	394085.73	363303.39	50	80	3:55	370233.0	364128.4	65	90
4	18:30	393554.94	359183.38	50	80	4:06	371106.8	360999.8	65	90
5	18:36	392603.98	356918.98	35	65	4:11	372464.8	358668.4	35	65
6	18:47	391620.26	346201.21	35	65	4:17	374450.8	357020.6	35	65
7	18:52	387676.04	340387.28	35	65	4:22	379764.0	355024.8	35	65
8	19:08	381964.49	332887.73	35	65	4:28	383585.9	353876.2	35	65
9						4:33	387431.4	351956.7	35	65
10						4:38	395388.3	349947.7	35	65
11						4:44	398607.4	348651.6	35	65
12						4:49	400997.3	347367.0	35	65
13						4:55	405145.4	343049.8	35	65

For the other two storms, their durations were very short, only 20-30 minutes and over short paths. As such it is difficult to determine the centre of these storms. Beside that, an observation from radar images was shown that during convective storm developed in some part of Klang Valley, the convective lines (the movement of convective storms) were broken abruptly and another strong convective storms were generated at different location and then pre-existing convective storms began at a new time (not shown). The boundaries of convective storm developed into a very complex shape with time. Figures 4.14 and 4.15 show the storm movement on April 6 and May 10, 2006. These figures show the movement of very strong convective storms during those events. Table 4.10 presents the storm centres coordinates and their reflectivity values.

From overall analysis, it is showed that an area situated in the tropics experiences predominantly convective precipitation. Heavy rainfall was resulted from strong convective events. The movement could be one line and varied. The duration of this movement was taken about 20 minutes to 1 hour until the centres of the storms were shrunk. Sometime, the evolution of centre of the storm is difficult to predict especially for short duration movement. This is because the centre of the storm abruptly initiated and broken rapidly then new strong convective storms were produced and begans at a new time. Beside that, it is indicated that the storm movement for short duration was very limited. The highest intensity of centre of the storm from all events analysed is 80 dBZ or 100 mm/hr in events on April 6, and May 10, 2006.

Table 4.10 : The coordinates and intensity of storm centres on 6.04.2006 and 10.05.2006

No	6-Apr-06					10-May-06				
	Time	Coordinate x	Coordinate y	dBZ	mm/hr	Time	Coordinate x	Coordinate y	dBZ	mm/hr
1	15:46	408014.4	354555.7	80	100	14:39	407001.2	357150.4	80	100
2	15:51	403815.7	351078.0	65	90	14:45	406613.6	357002.2	80	100
3	15:57	403583.0	350619.2	65	90	14:50	406296.3	349207.7	65	90
4	16:02	405663.6	349146.6	35	65	15:01	403297.0	349818.1	65	90
5	16:08	409915.8	345370.8	50	80					
6	16:13	409608.4	343723.8	35	65					

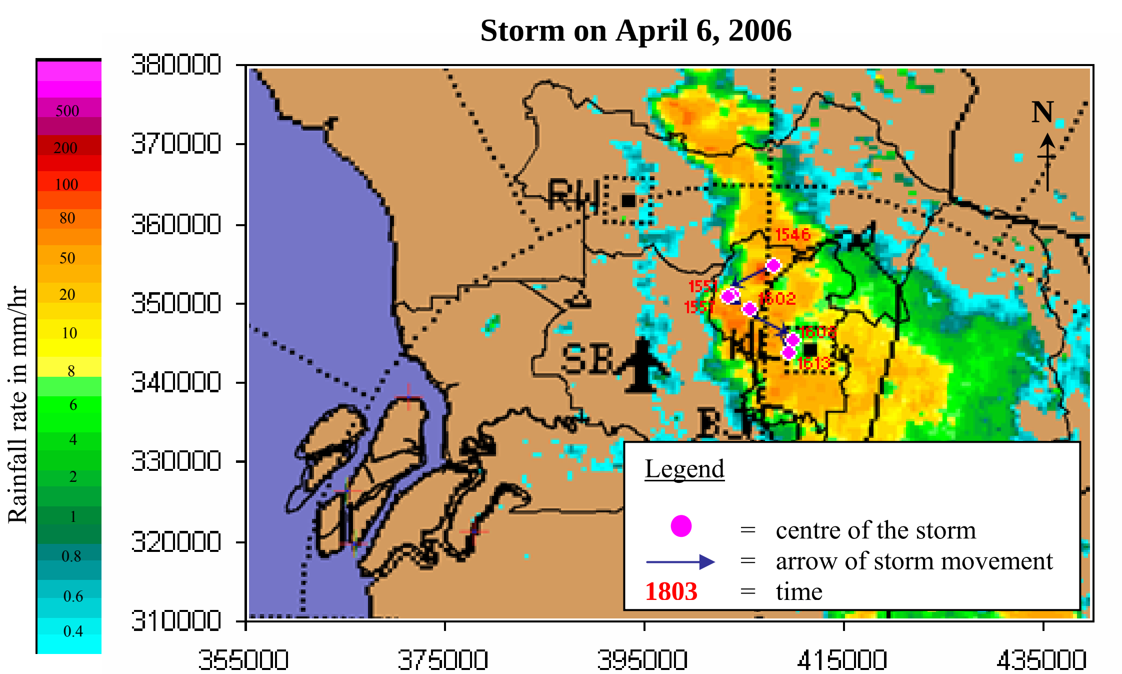


Figure 4.14 : Storm movement on April 6, 2006

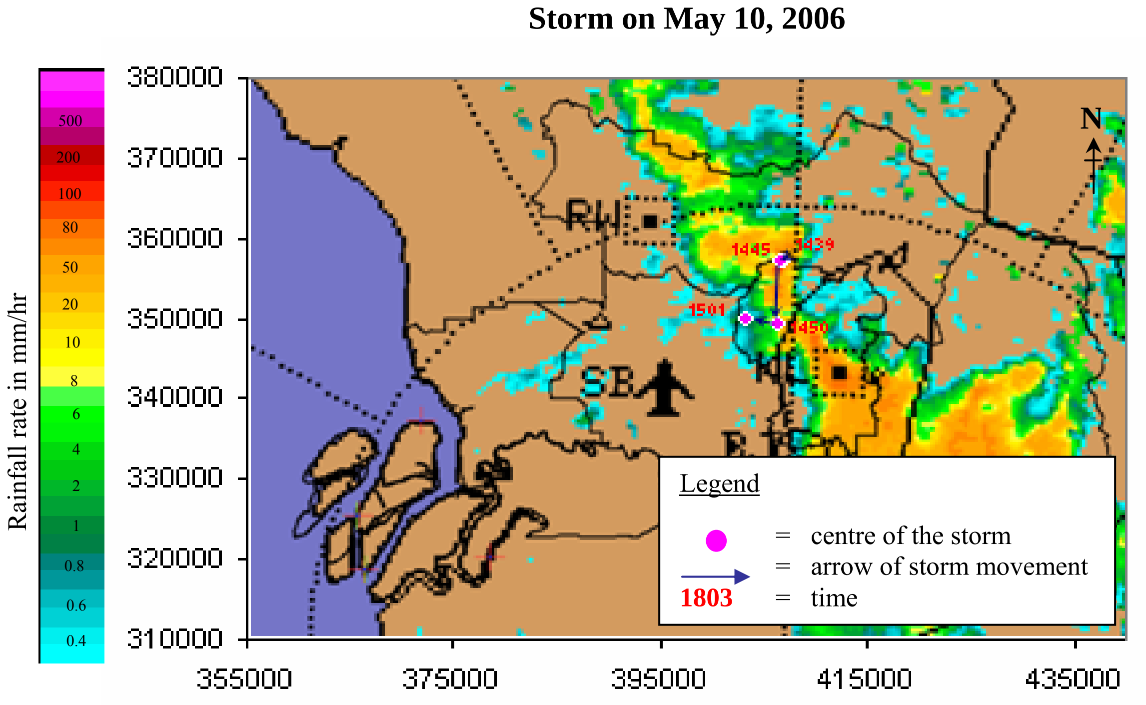


Figure 4.15 : Storm movement on May 10, 2006

4.3.7 Depth-Area Relationship

In order to obtain information on the size of rainfall cells and on the areal volume distribution during a single event depth-area relationships were derived. This analysis focused on a smaller area using eleven raingauges which cover 241.34 km^2. The areas between all pairs of neighbouring isohyets of the six selected storms computed by ArcGIS 9.1 are shown in Figure 4.16. As shown, four of the storms (on January 6, 2006, February 26, 2006, May 10, 2006 and November 5, 2004) have the highest rainfall depth at the southwest and decrease as the storm move to the northestern part of the catchment. However, the storms on April 6, 2006 and June 10, 2003 exhibited no direction of rainfall depth. It is observed that six raingauges had missing data in these events and might be one of factor that made the interpolation process in Kriging did not produce smooth rainfall contours. The percentages reduction of rainfall depth is plotted against the cumulative area from the storm centre (Figure 4.17). The shapes of the areal reduction curves were different for all the storms analysed. An average curve for all six storms was also drawn. Despite the large differences in the depth area curve patterns, the graph generally show that total rainfall depth decrease as the area increase. This finding is consistent with the property of convective events in section 4.5.1 where the highest intensity covers a small fraction of area.

From all curves plotted, it seems that the Areal Reduction Factors (ARF) values are consistent among each curve. Next, the ARF curve was then determined and compared with the ARFs from other areas. Figure 4.18 shows other curves derived by Desa (1997), Niemczynowicz (1984) for Lund in Sweden and by Shaw (1989) in the United Kingdom (1986). Desa (1997) was plotted ARF curve in small urban area (23 km^2) in Kuala Lumpur region. In his study, it is shown that the average ARF curve is lower than average ARF curve in this study but the curve almost similar with study by Niemczynowicz (1984). This is might be due to similar time and space resolution, similar size of area and of raingauge density and both catchments are situated in urban areas (Desa, 1997). From the graph also, it can be noticed that the area reduction curve derived in this study is quiet similar with previous curve derived for Malaysia by Yan and Lin (1986) for 1 hour. Nevertheless, the difference between this study is their curves were derived from data with poorer temporal and spatial resolution: 0.5 mm per tipping bucket with a weekly paper chart recorder and 23 raingauges covering an area of 200km^2. This study used 0.2 per tipping bucket with 20 raingauges covering an area of 241.34 km^2. Therefore, this graph possibly more accurate than graph by Yan and Lin (1986). However, further studies need to be done because the used of 20 raingauges might be not sufficient for an area of more than 200 km^2.

The results indicate that the shapes of such curves can only be compared between other locations if the temporal and spatial resolutions of the measurements are similar. This conclusion must be verified by more detailed analyses of areal and dynamic properties of single rainfall cells.

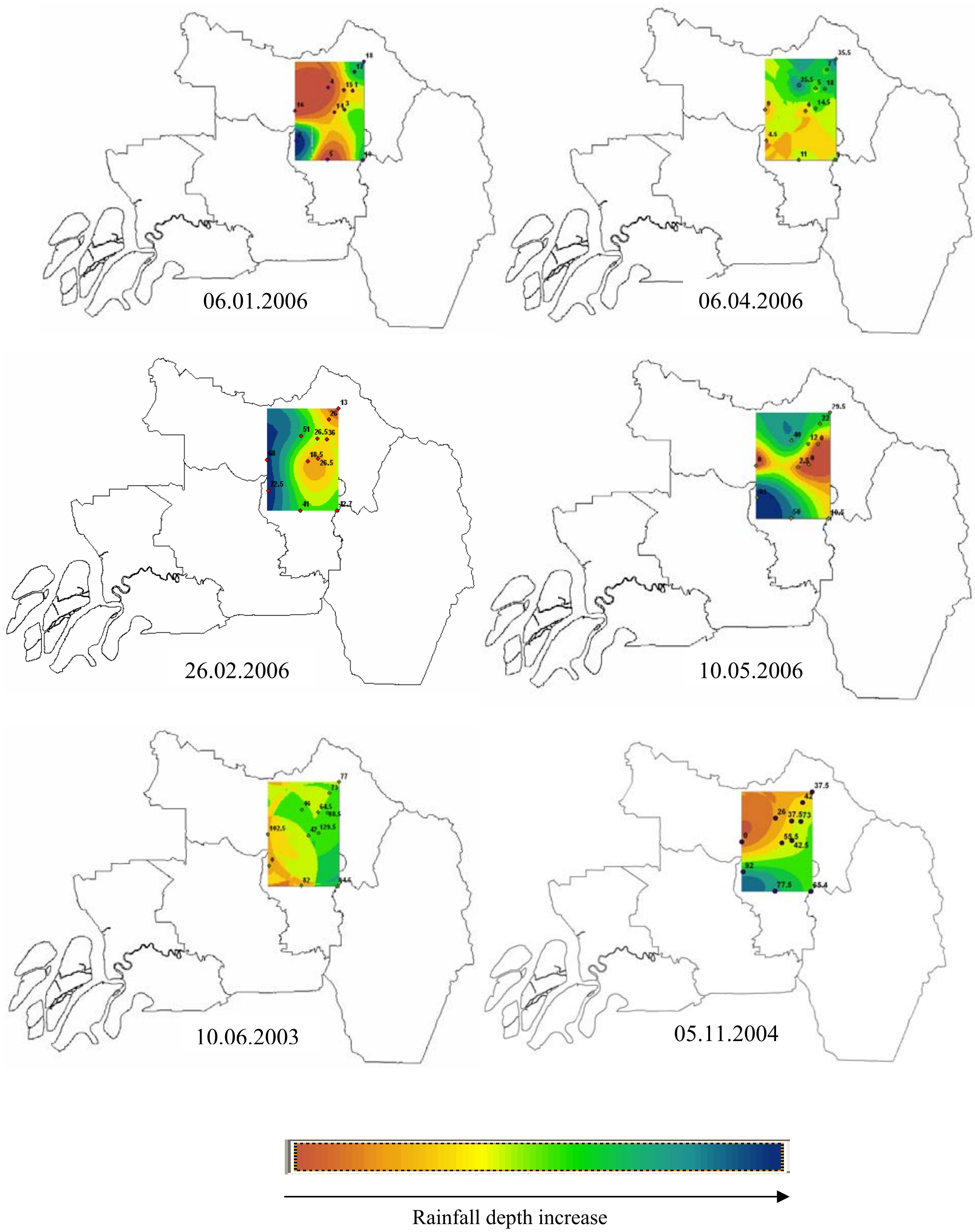


Figure 4.16 : Spatial variation of rainfall depth (mm) of six selected storms

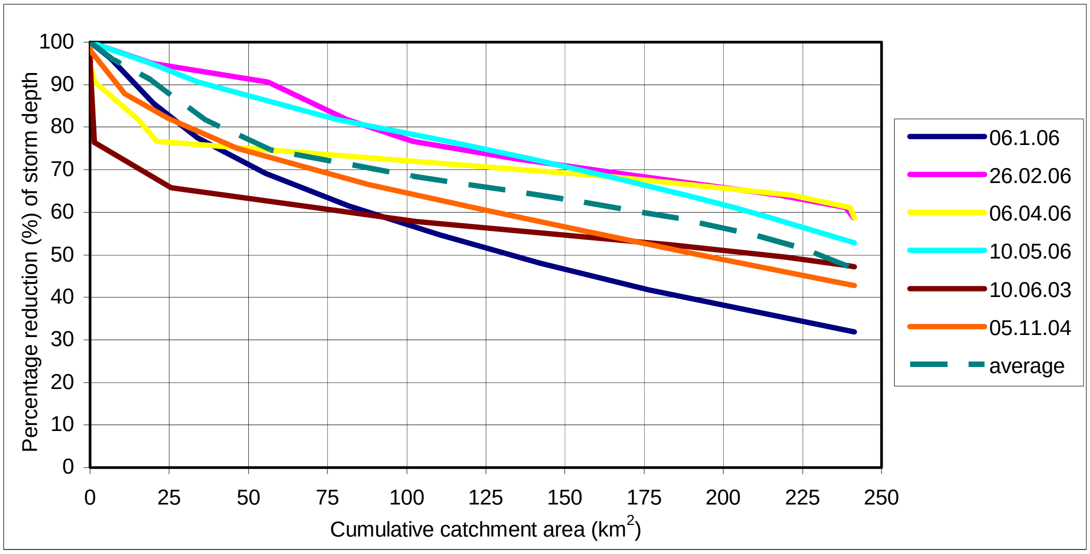


Figure 4.17 : Depth-area relationships for six selected storms.

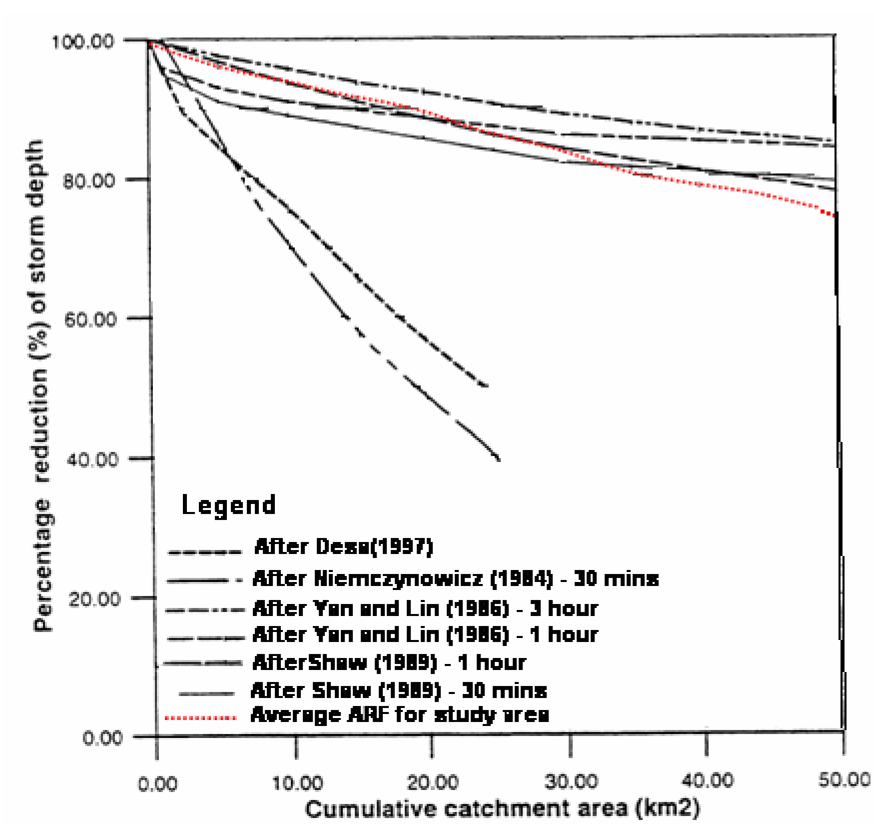


Figure 4.18 : Comparison of depth-area curves obtained in this study and at other locations (After Desa, 1997)

4.4 Stochastic Modeling of Hourly Rainfall Series

The modeling of hourly rainfall begins by finding the best-fit distribution for the hourly rainfall series.

4.3.1 Fitting the Best-fit Distribution for the Hourly Rainfall Amounts

Several methods have been proposed in literature for modeling rainfall amounts at the daily scale. The most common approach is to assume that rainfall amounts on successive days are independent and fit some theoretical distribution to the rainfall amounts (Todorovic and Woolhiser, 1975; Woolhiser and Roldan, 1982). However, there is no attempt so far to extend the method to the hourly rainfall amounts. Hence, this study will explore the methods proposed by Todorovic and Woolhiser (1975) andWoolhiser and Roldan (1982) for the hourly rainfall amounts in the Wilayah Persekutuan area. The best fitting distribution for the hourly rainfall amounts based upon several criteria of goodness-of-fit tests is to be determined. Four theoretical distributions considered include the Exponential, the Weibull, the Gamma and the Mixed-Exponential.

4.3.2 Fitting Distributions

There are 13 rainfall stations located in the vicinity. Historical rainfall data of every 15 minutes and daily amount are supplied by Department of Irrigation and Drainage (DID) Selangor for this study. The 15 minutes data are then aggregated to become hourly data. For this study, twelve stations were chosen based upon the completeness of the data. The study period ranges from 1981-1991 with most stations having a ten-year period hourly data (see Figure 4.1 and Table 4.1 for further details regarding these data).

The summary of the descriptive statistics for the stations used in this study is shown in Table 4.1. The means and variances are ranged from 3 mm/h to 4.3mm/h and from 5.5mm/h to 7.3mm/h respectively. As a result, the coefficients of variations are rather consistent throughout the state ranging from 1.533 to 1.794. This shows that the hourly rainfall variability over the whole state is quite homogenous.

In the ten-year periods, station 3217003 (KM 11 Gombak) shows the highest hourly maximum amount followed by station 3217001(KM 16 Gombak) having the lowest hourly maximum. However, station 3217001 experienced the highest number of wet days. All stations are positively skewed with the values of the coefficients are consistent throughout the stations.

Four theoretical distributions namely the Exponential, the Gamma, the Weibull and the Mixed Exponential are used in determining the best-fit distribution to describe the hourly rainfall amounts in Wilayah Persekutuan. Using the goodness-of-fit tests that has been discussed in Chapter 3, the best-fit distribution is chosen based upon the minimum error. The distributions are ranked according to these criteria. Table 4.12 shows the results of the tests.

Among the four distributions tested, the Mixed-Exponential was found to be the best fitting distributions for all stations where almost all the criteria of goodness-of-fit tests resulted in a minimum error to the Mixed-Exponential. This is followed by the Weibull, the Gamma, and finally the Exponential distributions.

The above results can be verified further by presenting the graphical representations through the plot of the exceedance probability. From the graphs given in Figures 4.19a to 4.19d, the Mixed-Exponential plot has the nearest plot to the observed. Hence, the Mixed-Exponential distribution was found to be the best in describing the hourly rainfall amounts in the Wilayah Persekutuan.

The estimated parameters for the Mixed-Exponential distribution are shown in Table 4.3. The mixing probability that indicated the percentage of variation of the hourly rainfall amounts in the Wilayah Persekutuan has shown an approximate value of between 0.6 to 0.7. The weighted average of two exponential distributions in the mixed-exponential distributions may refer to the two types of rainfall, namely “light” or “heavy”. Hence, it can be interpreted that between 60% and 70% of the hourly rainfall series in the Wilayah Persekutuan is contributed by the light rain. Hence, the remainder is being contributed by the heavy rain. This is true due to the higher frequency of light rain for the hourly data.

However, the total estimated mean shows that about 80% is attributed to heavy rainfall. This implies that most of the rainfall amounts recorded in the study area is received from heavy rains even though there is a higher occurrence of light rainfall. The hourly duration used indicates short duration heavy rainfall has a large impact on the rainfall amount received and potentially is the main contribution to flash flood events.

Table 4.11: Descriptive statistics of the rainfall amounts for the Wilayah Persekutuan **.**

Station no.	Station names	Duration	Hourly Mean	Std. Dev.	CV	Skewness	Kurtosis	No.of wet hours	Max. amount rainfall (mm)
3015001	Puchong Drop	1982-1990	3.997	7.17	1.794	3.712	18.814	4057	82.10
3116005	Sek.Ren. Taman Maluri	1981-1990	3.663	6.337	1.73	3.768	20.946	6466	92.50
3116006	Ladang Edinburgh Site 2	1981-1990	3.68	6.249	1.698	3.808	20.234	5598	72.70
3216001	Kampung Sg. Tua	1981-1990	3.98	6.102	1.533	3.28	14.504	6074	69.60
3216004	SMJK Kepong	1982-1991	4.3	7.346	1.708	3.736	19.277	4328	75.50
3217001	KM 16 Gombak	1981-1990	3.359	5.682	1.692	3.815	20.313	7102	58.20
3217002	Empangan Genting Kelang	1981-1990	3.145	5.495	1.747	3.901	20.313	6819	57.70
3217003	KM11 Gombak	1981-1990	3.779	6.318	1.672	3.789	21.830	5551	92.90
3217004	Kpg. Kuala Saleh	1981-1990	4.16	7.046	1.694	3.682	18.399	4549	72.30
3217005	Gombak Damsite	1982-1991	3.768	6.753	1.792	3.771	18.954	3447	70.10
3317001	Air Terjun Sg.Batu	1985-1994	4.042	6.732	1.666	3.524	16.986	5279	69.70
3317004	Genting Sempah	1981-1990	3.018	5.272	1.747	4.2	27.805	7484	83.00

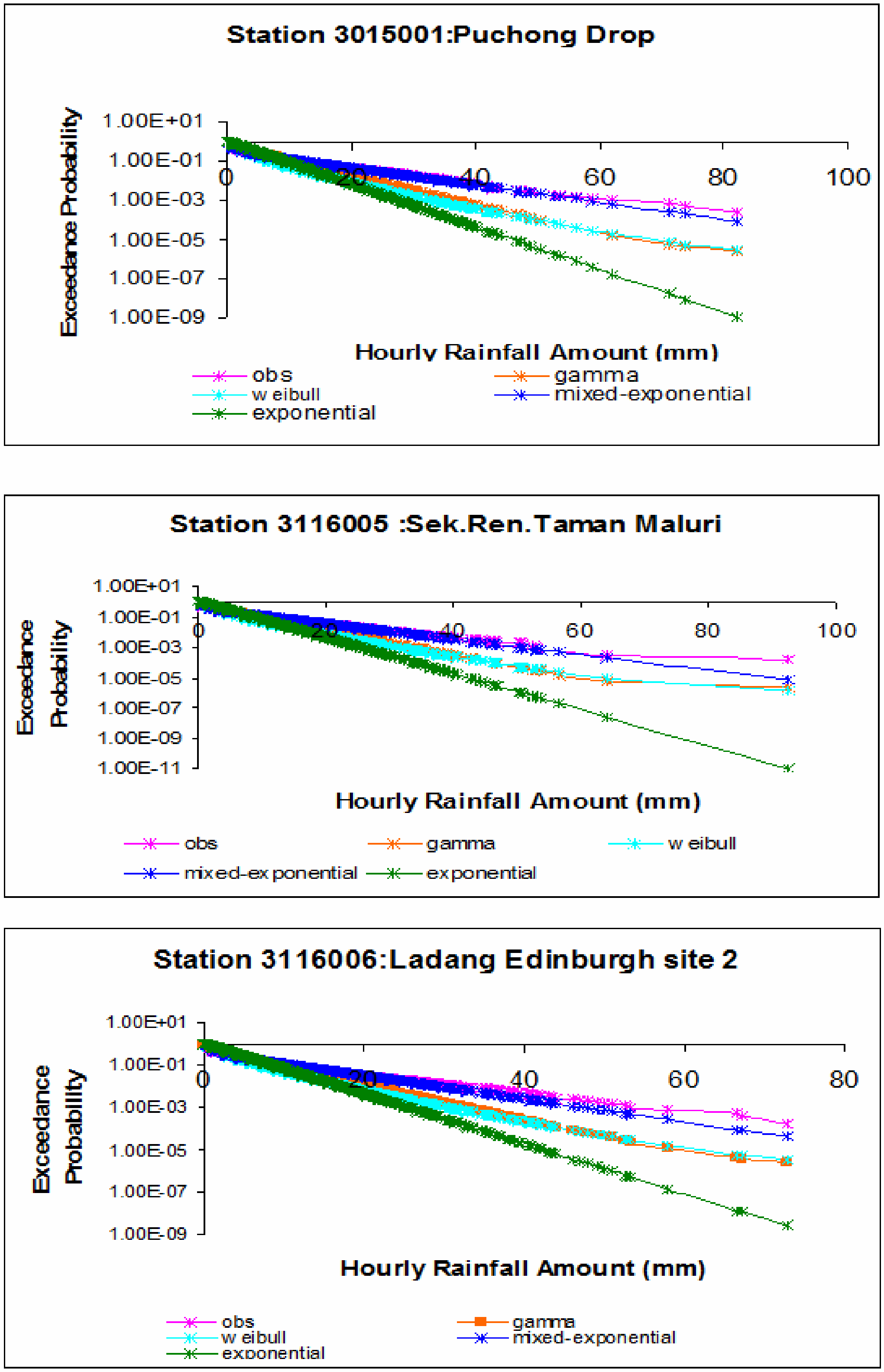
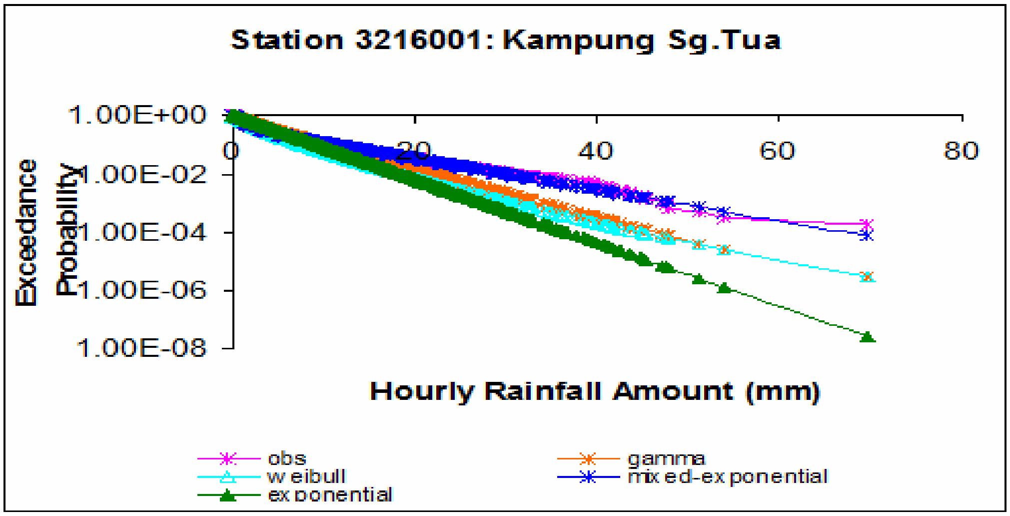
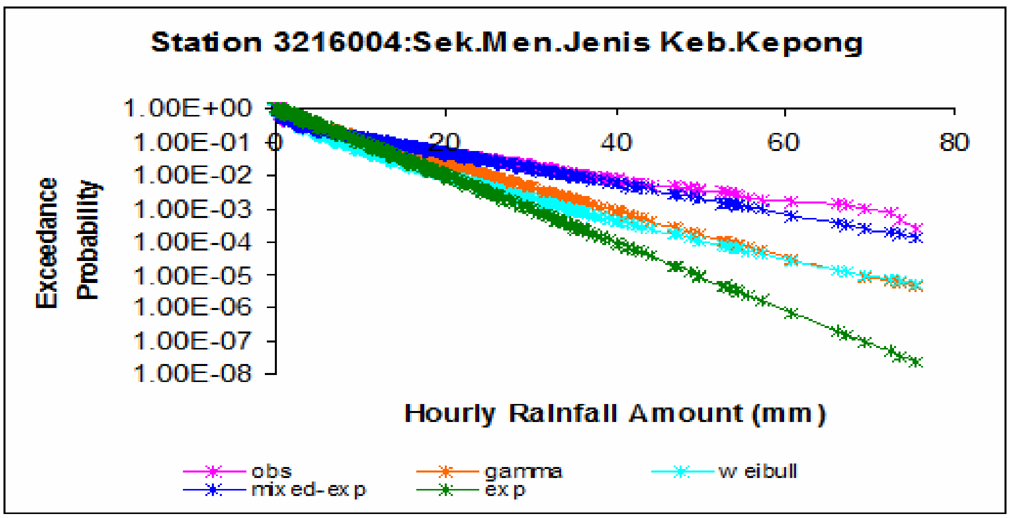


Figure 4.19a: Exceedance Probabilities for the Hourly Rainfall Amount





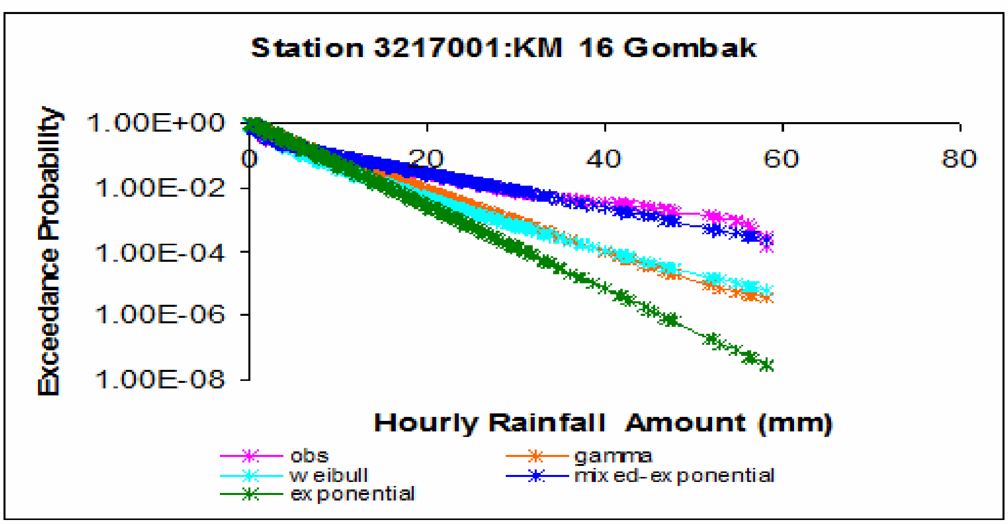
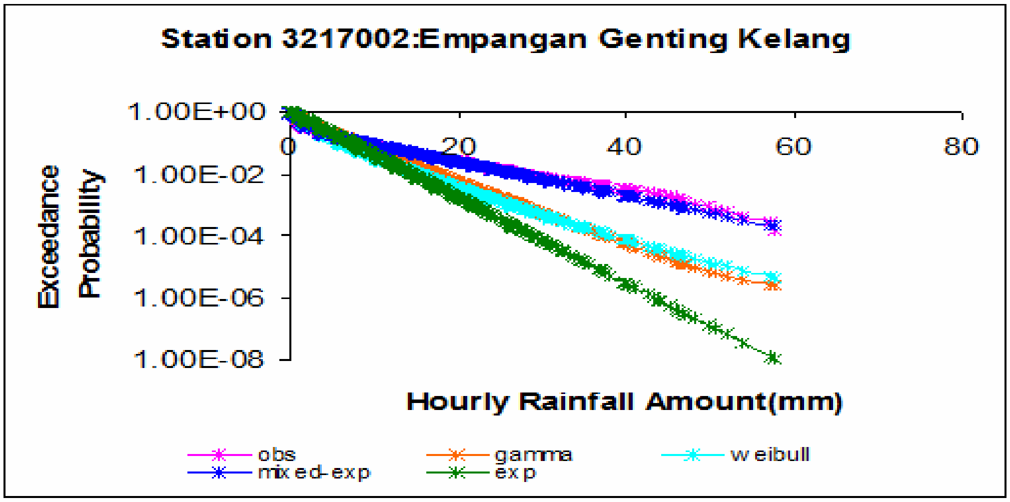
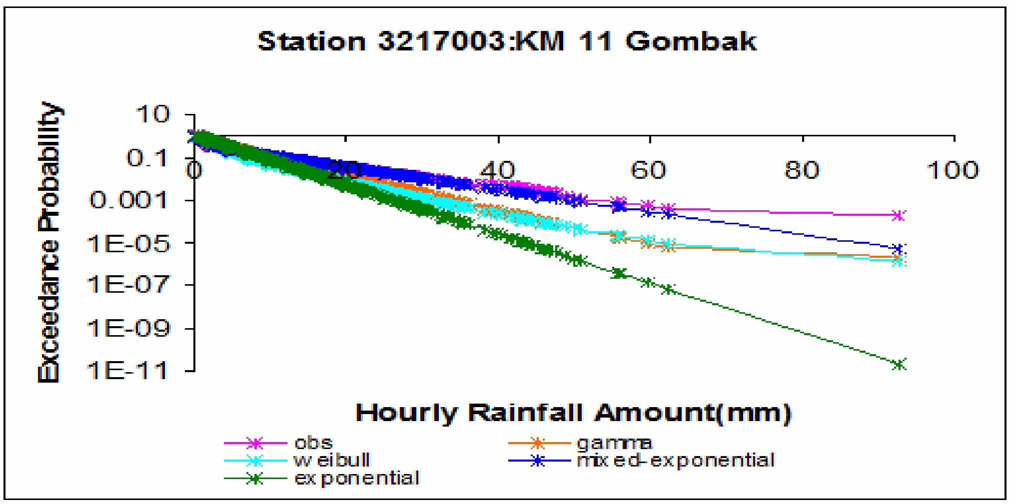


Figure 4.19b: Exceedance Probabilities for the Hourly Rainfall Amount





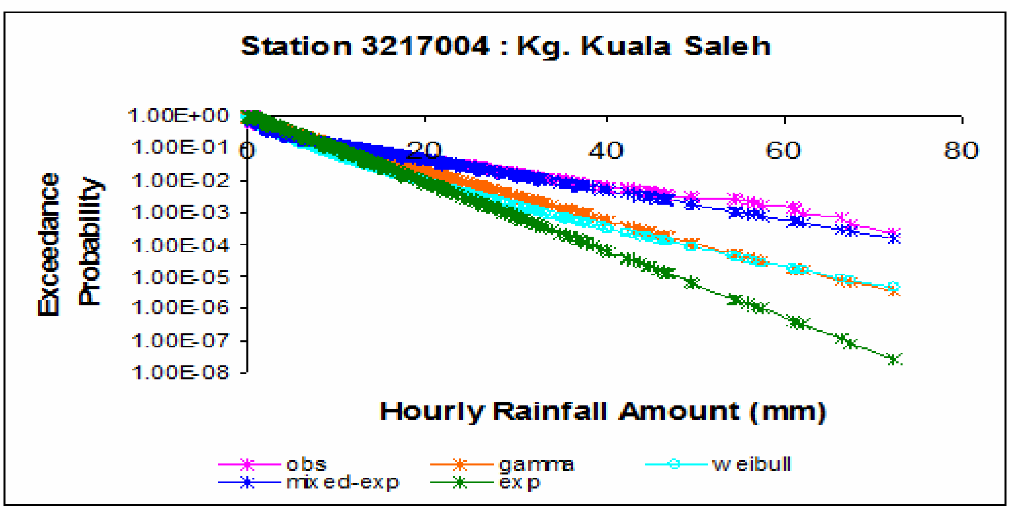


Figure 4.19c: Exceedance Probabilities for the Hourly Rainfall Amount

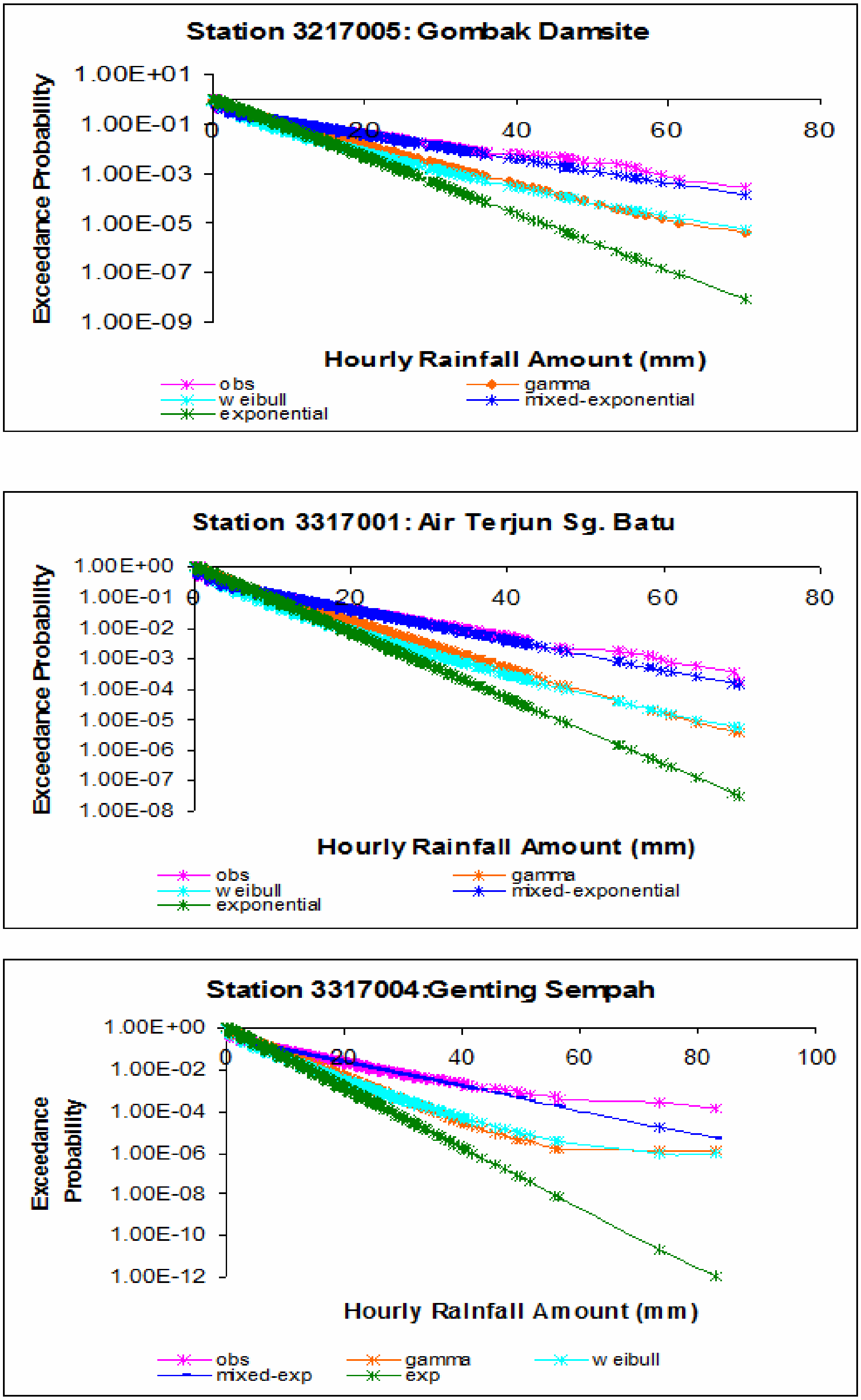


Figure 4.19d: Exceedance Probabilities for the Hourly Rainfall Amount

Table 4.12: The ranking of distributions using AIC and goodness-of-fit tests

No.	Stations	AIC	KS	CVM	AD	Means	Median
1	3015001	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.GM 3.WE 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP
2	3116005	1.MEX 2.WE 3.GM 4.EXP	1.GM 2.MEX 3.WE EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP
3.	3116006	1.MEX 2.WE 3.GM 4.EXP	1.MEXP 2.GM 3.WE 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP
4.	3216001	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.GM 3.WE 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 3.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP
5.	3216004	1.MEX 2.WE 3.GM 4.EXP	1.GM 2.MEX 3.WE 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP
6.	3217001	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.GM 3.WE 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP
7.	3217002	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.GM 3.WE 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP
8.	3217003	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.GM 3.WE 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP
9	3217004	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.GM 3.WE 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP
10.	3217005	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.GM 3.WE 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP
11.	3317001	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.GM 3.WE 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP
12.	3317004	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.GM 3.WE 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP	1.MEX 2.WE 3.GM 4.EXP

MEX=MIXED-EXPONENTIAL;WE=WEIBULL;GM=GAMMA;EXP=EXPONENTIAL;AIC=AIKAKE INFORMATION CRITERION;KS=KOLMOGOROV-SMIRNOV;CVM=CRAMER-VON-MISES;AD=ANDERSON-DARLING

Table 4.13: The estimated parameters for the Mixed Exponential distribution

Station no.	Station names	Mixing probability (α)	Scale 1 (β_1)	Scale 2 (β_2)	Estimated mean
3015001	Puchong Drop	0.6772	1.137 [19%]	9.996 [81%]	3.997
3116005	Sek.Ren. Taman Maluri	0.6504	1.077 [19%]	8.474 [81%]	3.663
3116006	Ladang Edinburgh Site 2	0.6261	1.108 [19%]	7.985 [81%]	3.68
3216001	Kampung Sg. Tua	0.6218	1.44 [23%]	8.154 [77%]	3.977
3216004	SMJK Kepong	0.6302	1.253 [18%]	9.48 [82%]	4.295
3217001	KM 16 Gombak	0.687	1.193 [24%]	8.114 [76%]	3.359
3217002	Empangan Genting Kelang	0.702	1.114 [25%]	7.93 [75%]	3.145
3217003	KM11 Gombak	0.6433	1.211 [21%]	8.409 [79%]	3.778
3217004	Kpg. Kuala Saleh	0.6482	1.313 [20%]	9.4 [80%]	4.158
3217005	Gombak Damsite	0.6477	1.002 [17%]	8.853 [83%]	3.768
3317001	Air Terjun Sg.Batu	0.6245	1.178 [18%]	8.804 [82%]	4.042
3317004	Genting Sempah	0.6998	1.066 [25%]	7.57 [75%]	3.019

The percentages in the brackets refer to the estimated means of the hourly rainfall amounts contributed by both scales.

4.3.5 Summary

The distribution of the hourly rainfall amounts in the Wilayah Persekutuan is best described by the Mixed-Exponential distribution. The Weibull and the Gamma distribution are ranked second and third respectively, and the last in the ranking is the Exponential distribution. These are based on the goodness-of-fit tests performed on the studied station, as discussed in section 3.3.

From the estimated parameters of the Mixed-Exponential distribution obtained, it could be interpreted that between 60% and 70% of the wet hourly series in the Wilayah Persekutuan is contributed by the light rainfall and the remainder by the heavy rainfall. However the total estimated mean shows that about 80% is attributed to heavy rainfall. This implies that most of the rainfall amounts recorded in the study area are received from heavy rains even though there is a higher occurrence of light rainfalls. The hourly duration used indicates short duration heavy rainfalls have a large impact on the rain amounts received and potentially could be the main contribution to flash flood events. These would indeed provide grounds for further studies on convective rainfall and flash floods.

4.3.4 NSRP model with mixed exponential distribution

The model is referred as the MEXPTRAN in this study. Figure 4.20 shows the comparison between the observed and the simulated statistical properties of rainfalls for the one-hour scale. The model simulation accurately preserved the observed values of the one-hour mean and variance. The one-hour rainfall coefficients of skewness and autocorrelations were matched very well the observed values for some of the months.

Figure 4.21 shows the comparison between the observed and the simulated physical properties of rainfalls for the one-hour scale. The model matched fairly well the one-hour maximum rainfall for the whole year. The transition probabilities of rainfall occurrence *P10* (wet-dry hour) and *P00* (dry-dry hour) were matched poorly by the model. Similarly, the model underestimated the probability of dry hours of rainfall.

Figure 4.22 shows the comparison between the observed and the simulated of rainfalls for six-hour scale. The mean, variance and coefficients of skewness of six-hour rainfalls were preserved accurately by the model simulation. However, autocorrelations of six-hour observed rainfalls were overestimated.

Figure 4.23 shows the comparison between the observed and the simulated statistical properties of rainfalls for the 24-hour scale. The mean, variance and the coefficients of skewness of the 24-hour rainfalls of the observed were accurately reproduced by the model. The autocorrelations of 24-hour rainfalls were adequately preserved.

Figure 4.24 shows the comparison between the observed and the simulated physical properties of rainfalls for the 24-hour scale. The 24-hour maximum rainfalls were preserved fairly well by the model. However, the probability of dry days of the observed rainfalls were preserved very well by the model. Similarly, the daily transition probabilities of rainfall occurrences *P10* (wet-dry day) and *P00*(dry-dry day) of the observed could be preserved accurately by the model simulation.

Figure 4.25 shows the comparison between the observed and the simulated at the properties of rainfalls for the monthly scale. The observed monthly mean had close agreement with the medians of the box plots for the whole year. However, the standard deviations, maximum and minimum monthly rainfalls of the observed were fairly matched by the model.

In general, the MEXPTRAN performed very well in preserving the observed means and variances of rainfalls at various time scales. The model has also managed to describe accurately the probability of dry days and the transition probabilities of rainfall occurrences for the the whole year. However, the autocorrelations and the coefficients of skewness of rainfalls at various timescales were only fairly preserved, but within the range of the simulated properties considered. Nevertheless, the MEXPTRAN simulation preserved the seasonal trend of the observed properties very well.

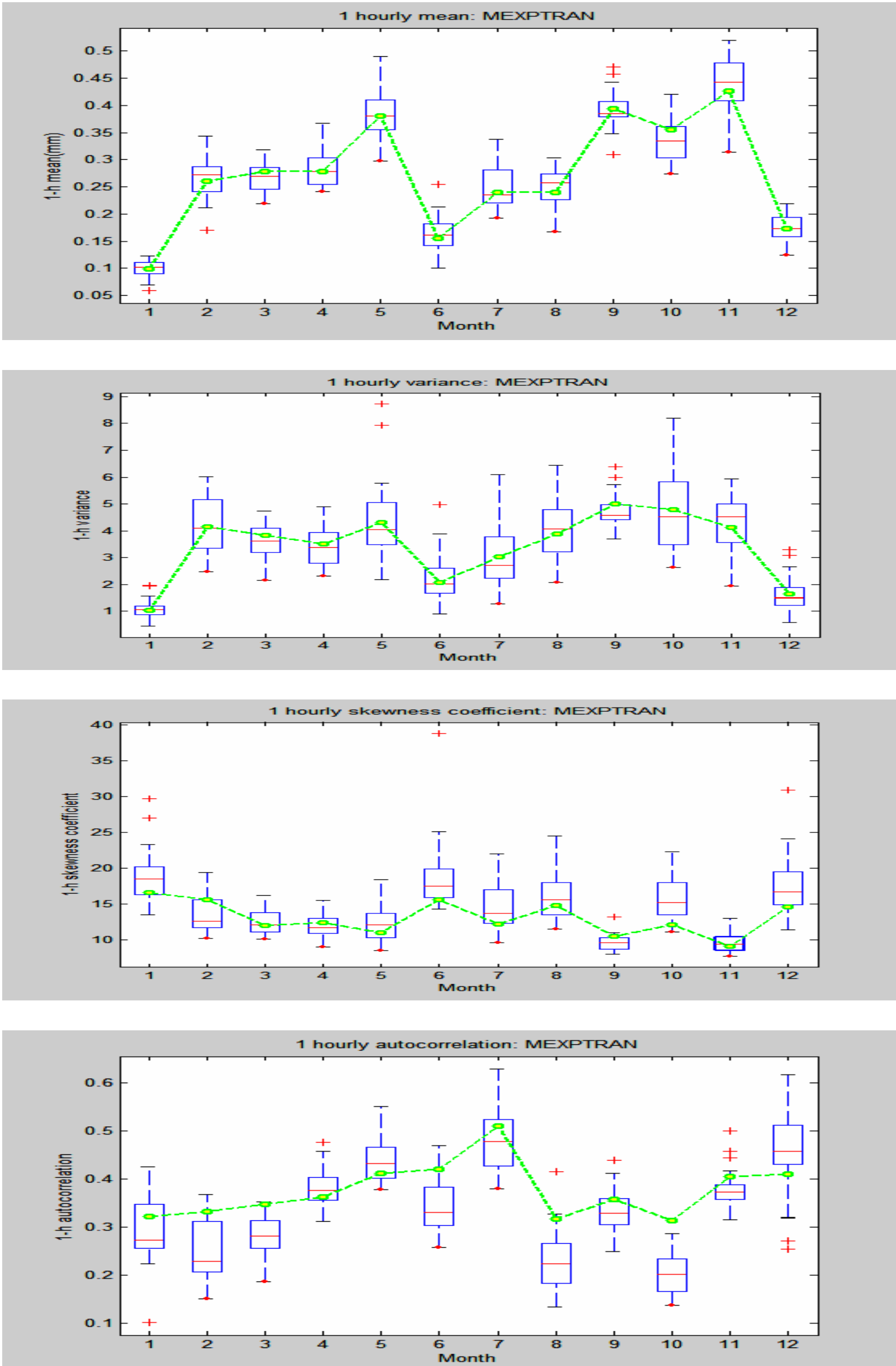


Figure 4.20: Monthly Statistical Properties of 1-Hour Rainfall (in mm) of simulated MEXPTRAN

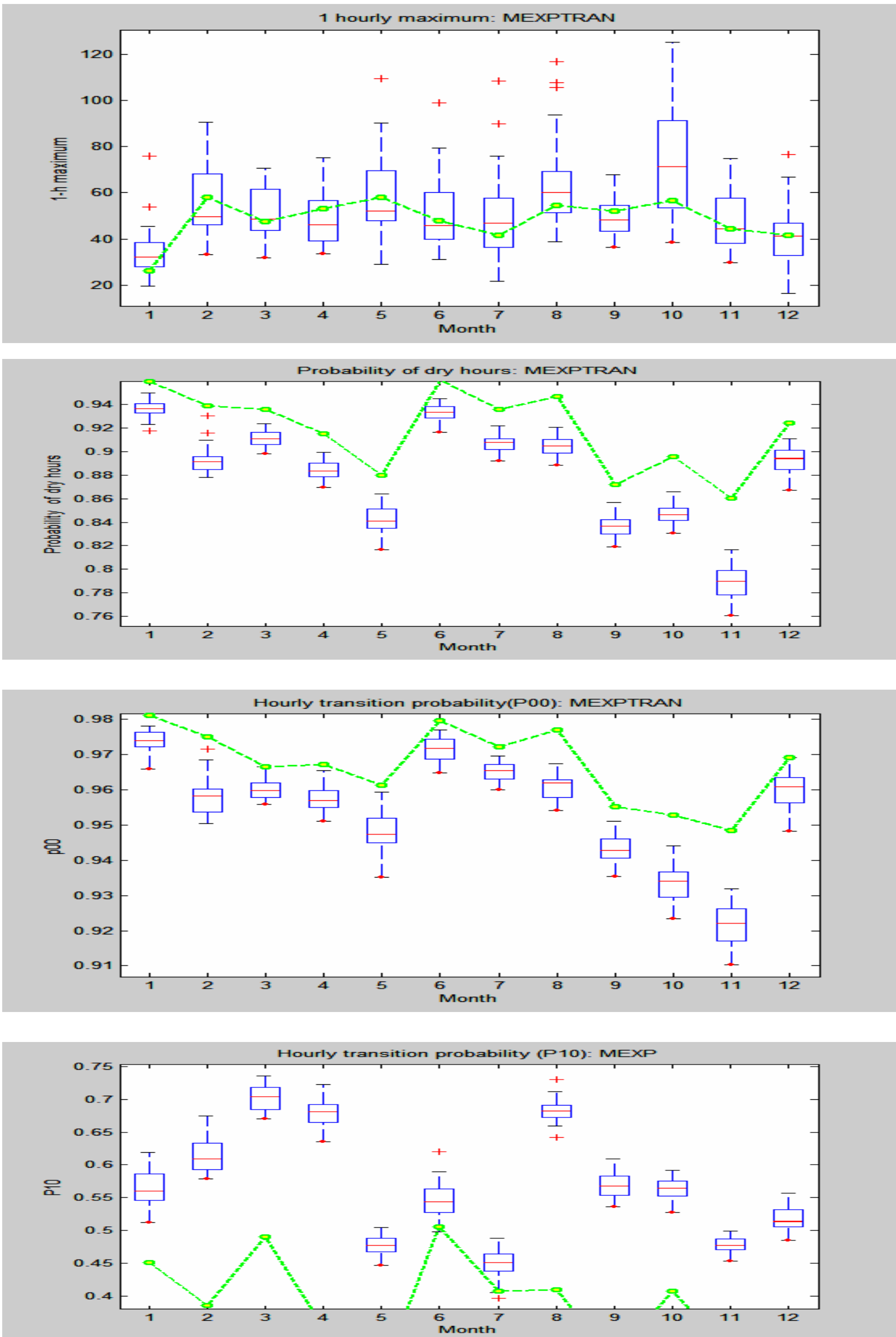


Figure 4.21: Monthly Statistical Properties of 1-Hour Rainfall (in mm) of simulated MEXP

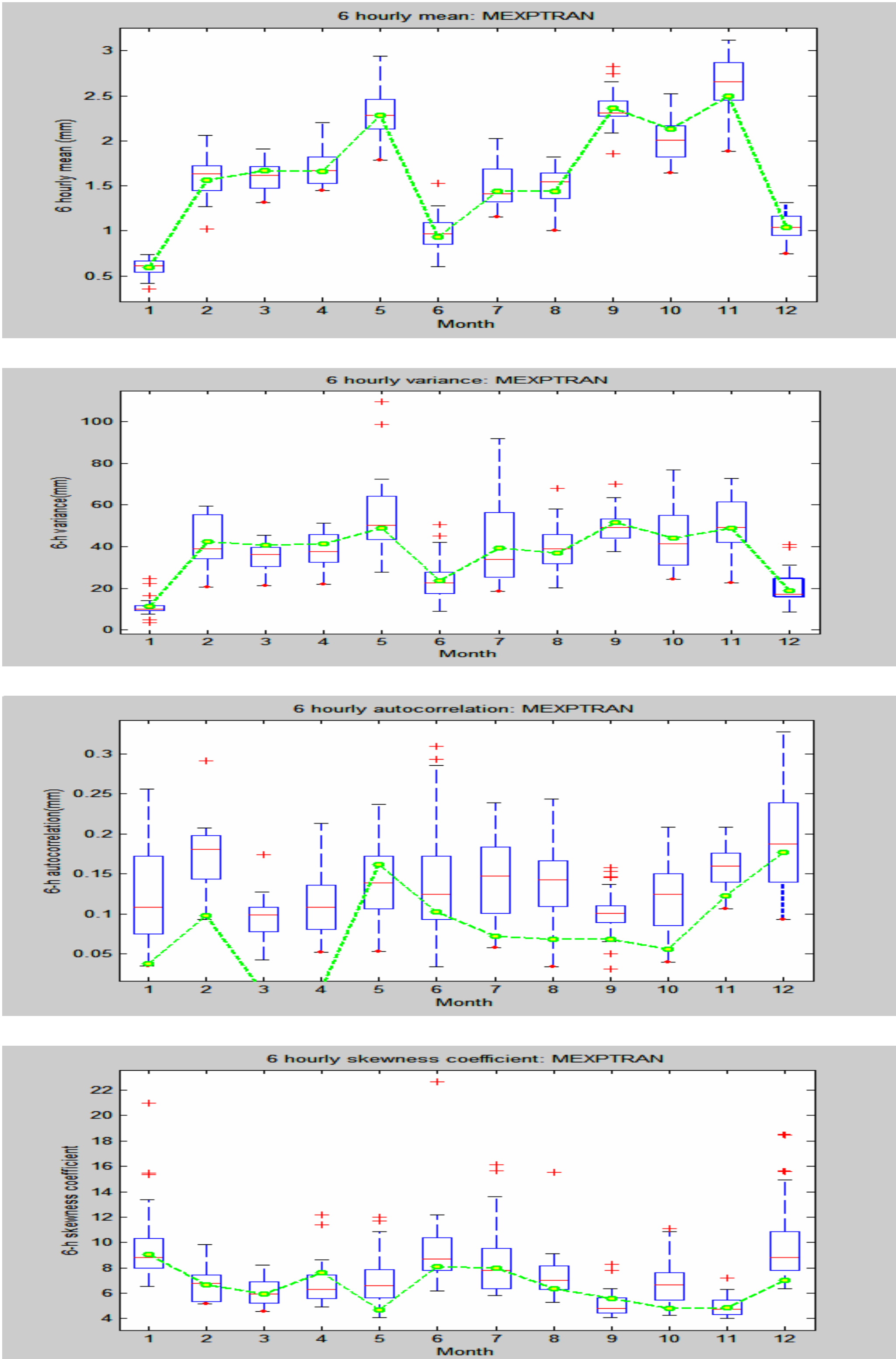


Figure 4.22: Monthly Properties of 6-Hour Rainfall (in mm) of simulated MEXPTRAN

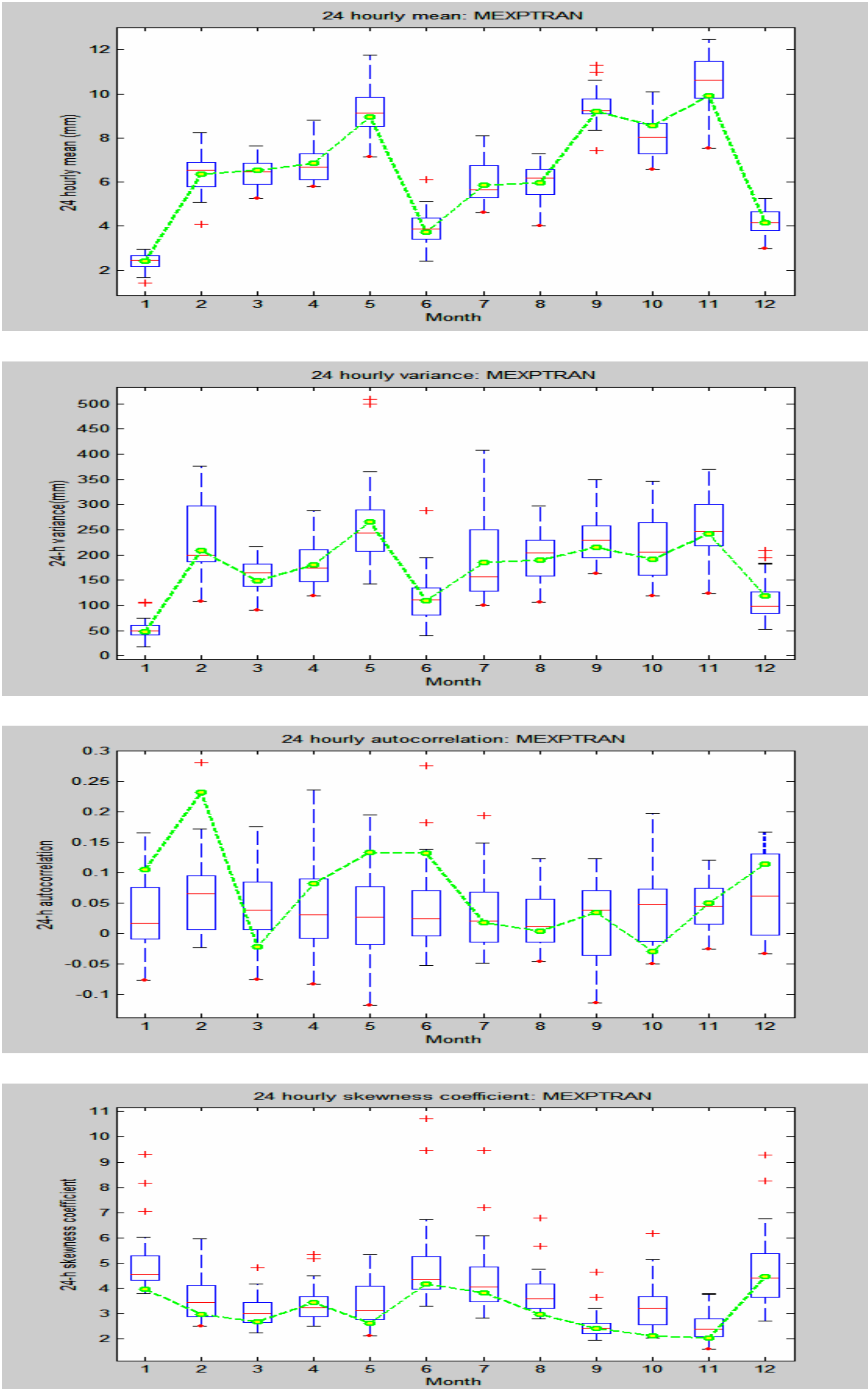


Figure 4.23: Monthly Statistical Properties of 24-Hour Rainfall (in mm) of simulated MEXPTRAN

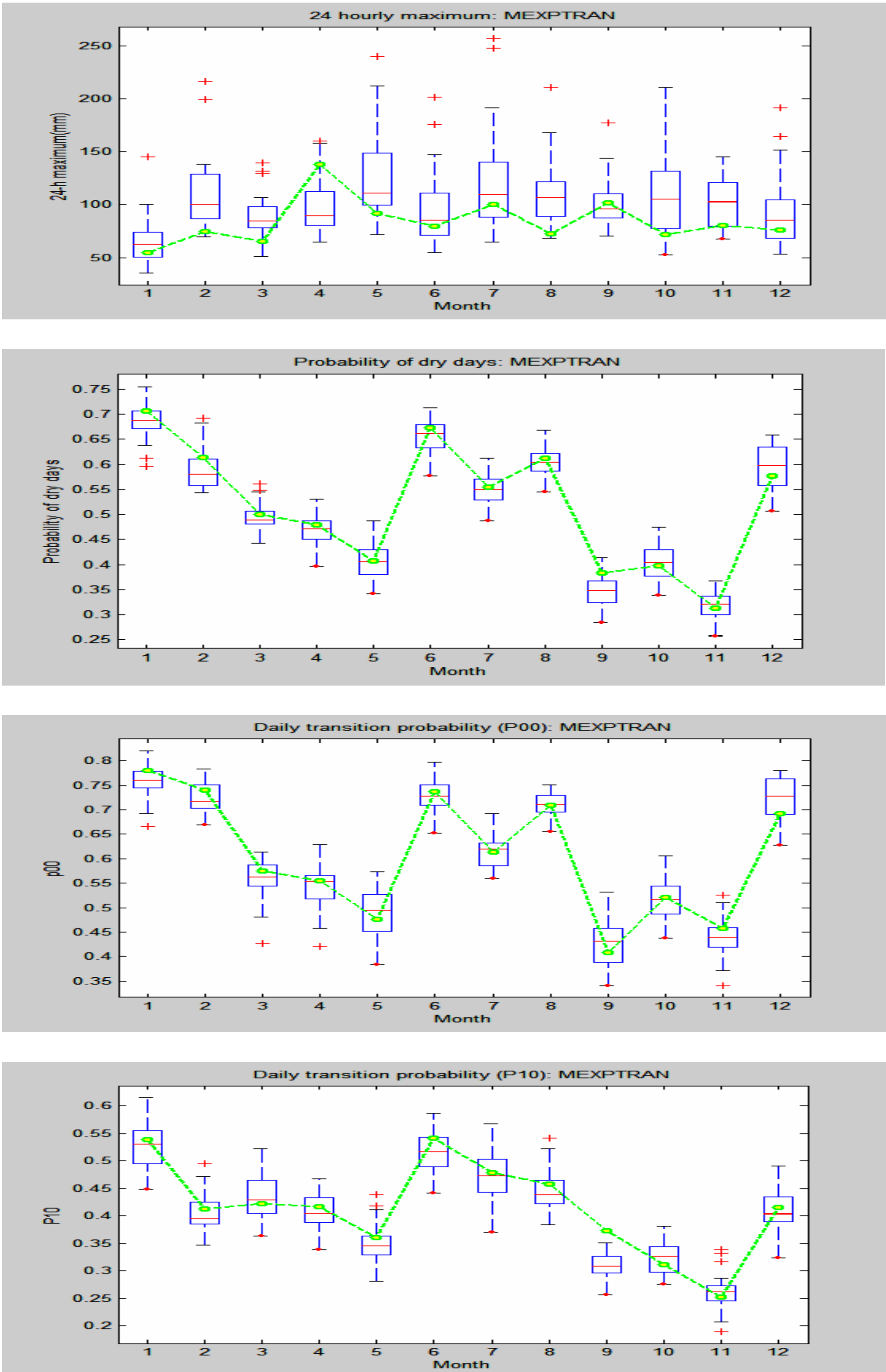


Figure 4.24: Monthly Physical Properties of 24-Hour Rainfall (in mm) of simulated MEXPTRAN

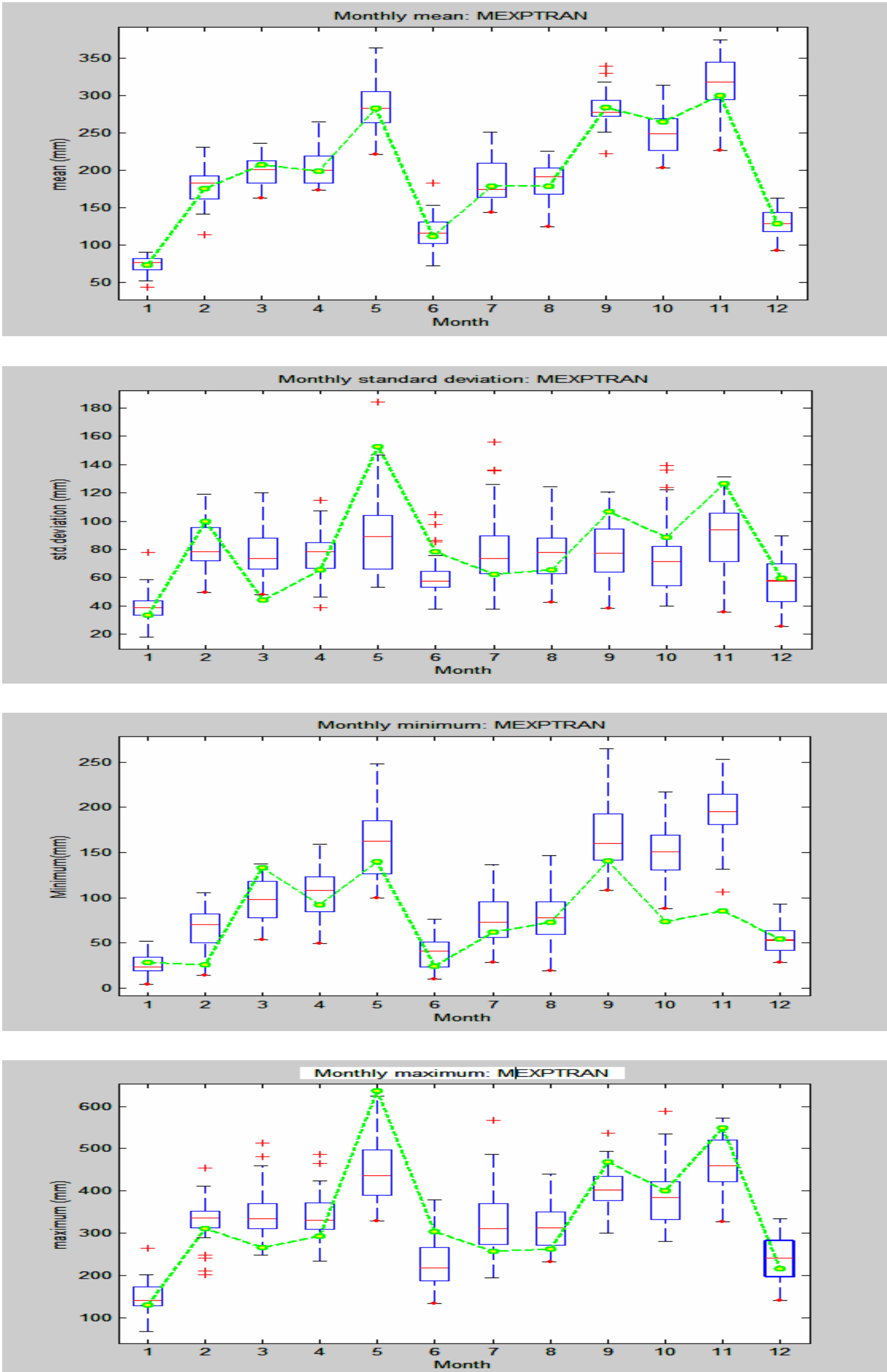


Figure 4.25: Monthly Properties of 1-Month Rainfall (in mm) of simulated MEXPTRAN

4.3.5 MCME model

The Markov Chain Mixed Exponential (MCME) model is a daily rainfall model. The performance of the model on the daily rainfall series was extensively acknowledged in literature (e.g. Woolhiser et al. 1984, Eagleson, 1978; Woolhiser et.al,1982, Roldan et.al,1982, Richardson, 1981). However, in this study the model was modified and applied to the hourly rainfall series. This chapter basically discussed the suitability and applicability of the modified hourly MCME model. The performance of the model on the daily series was also evaluated and compared with the performance of the modified hourly MCME model..

4.3.5.3 Performance of Hourly MCME model

The Markov Chain process for hourly series was applied from 1981-1990 obtained from Station 3217001 at KM 16 Gombak. The Mixed-Exponential represents the hourly rainfall amounts from the same station. Monthly parameters for the rainfall distribution and occurrences were estimated using the SCE method. Following Fourier series fitting of each variable for seasonal variability throughout the year, simulations for the synthetic time series were conducted for 10-year period using the parameter sets obtained from the hourly series.

4.3.5.2 Fitting of the Mixed Exponential Distribution to Observed data

In Section 4.3.3, the study on finding the best distribution for the hourly rainfall amounts in the Wilayah Persekutuan has found that the mixed exponential distribution was the best distribution. However, to assess the descriptive ability of the mixed exponential distribution the exceedance probability curves were used for each month. The exceedence probability of monthly rainfall plotted on a semi-log scale provides a qualitative tool to assess the performance of the mixed exponential distribution. The semi-log scale helps to determine the mixed exponential nature of the data if it exists. The rainfall distribution of a particular month would follow an exponential function if the observed probability follows a straight line. In Figures 4.26a and 4.26b, the dots represent the observed probabilities while the dashed line represents the theoretical values. However, for all months the exceedance probability curves contain at least two slopes, which indicate a mixed-exponential distribution. The break in slopes points to the physical evidence concerning the presence of at least two different types of storm rainfall (convective and non-convective) and this further supports the use of mixed exponential distribution (Hussain, 2007). The use of the mixed exponential is the most appropriate because of its flexibility in capturing the mixture of storm types, as well as a single exponential pattern.

4.3.5.3 Fourier Series Fit to Parameter Sets.

Two Markov Chain transitional probabilities (*P00* and *P10*) and three mixed exponential parameters (α, ξ and θ) were generated for each month. Thus, a total of sixty parameters were needed to describe the rainfall process. However, the number can be reduced by using a truncated polar Fourier series. The seasonal variability of each parameter through the twelve months of the year was represented by using maximum likelihood estimates of the periodic parameters using five harmonics (Han, 2001). The number of harmonics may be reduced or increased to create a more parsimonious model. If the number of harmonic increases the total parameters to be estimated would also increase. The use of five harmonics for all five parameters would lower the total number of parameters from sixty to fifty-five.

The Fourier series fit is compared to the non-fitted transition probabilities and mixed exponential parameters as shown to determine whether the parameters are well represented by the Fourier series. As shown in Figure 4.27 the Fourier fits for all transition probabilities are in very close agreement. The dots represent the MCME parameters and the dashed lines represent the Fouries series fit. However, the parameter

θ that represents the higher (larger) mean is not as well presented as the parameter ξ that represent the lower (smaller) mean. This perhaps implies that the larger mean is not be well predicted by Fourier series fit. Nevertheless, as whole the seasonal variability of the rainfall process is well described by the Fourier series fit.

4.3.5.4 Simulation Verification

Following the calibration of the rainfall amounts using the Fourier series fits of the MCME, 50 simulations of hourly rainfall series of the same length were generated. The statistical properties and the physical properties of the generated series were compared with the observed series.

4.3.5.5 Simulated Transitional Probabilities

The box plot in Figure 4.28 represents the transition probabilities calculated for 50 sets of monthly data from 50 simulations as compared to the empirical transition probabilities (represented by dots connected by the dashed line). The simulated transition probabilities are well preserved and comparable to the empirical values. The median of the box plots is excellently matched with the empirical value and the spread around the median represent the variability that exists in 50 simulations. The seasonal variability can also be seen in general trend of the monthly box plots for the whole year. Therefore, it can be said that the simulation of the hourly rainfall occurrences is comparable to that of the observed pattern.

The probability of a dry hour following a dry hour is as expected, very high. A further investigation of the transition probabilities shows that an hour is more likely to be dry if the previous hour is wet in the months of June and March.

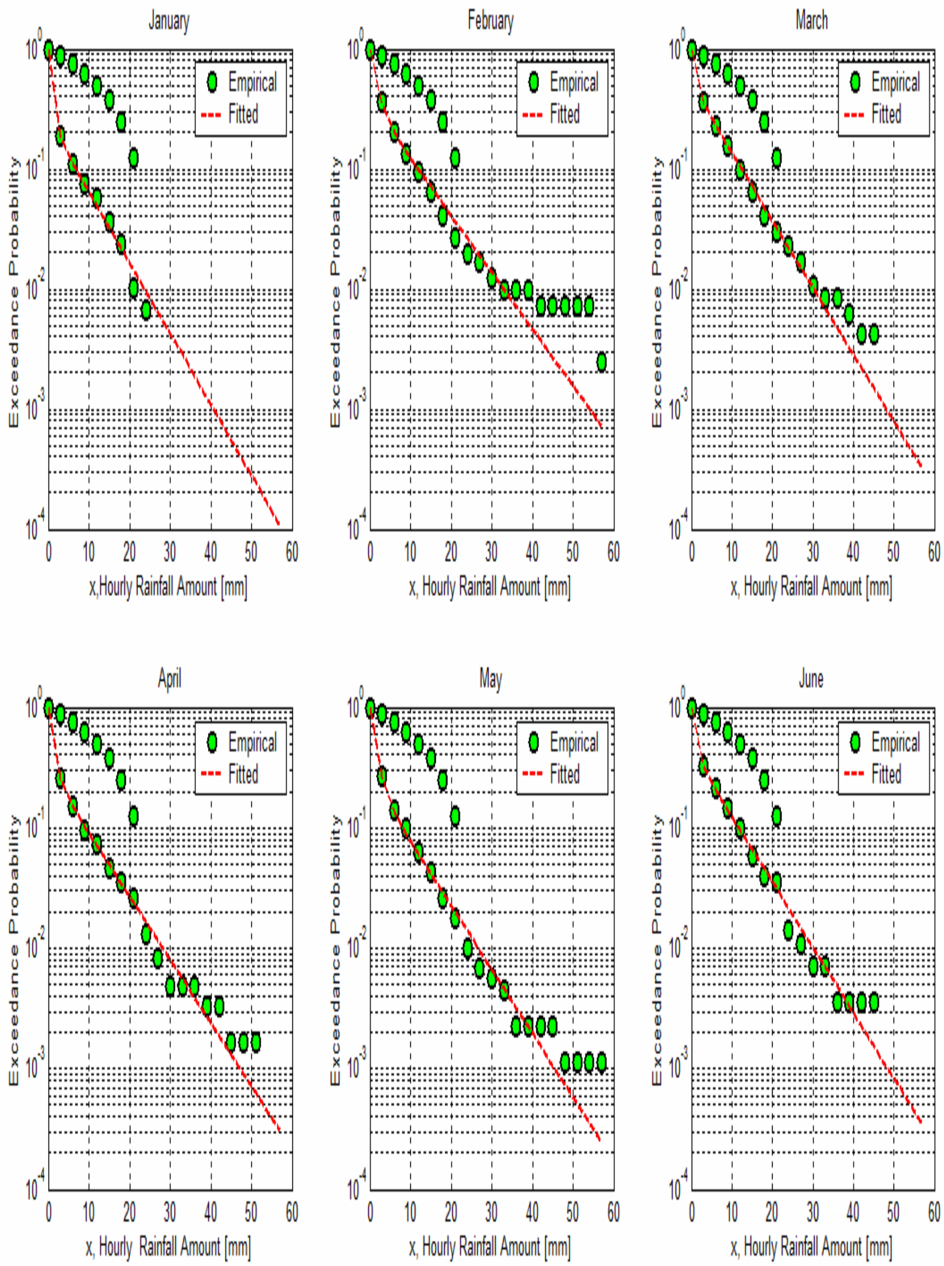


Figure 4.26a: Exceedance Probabilities for Hourly Rainfall from January to June.

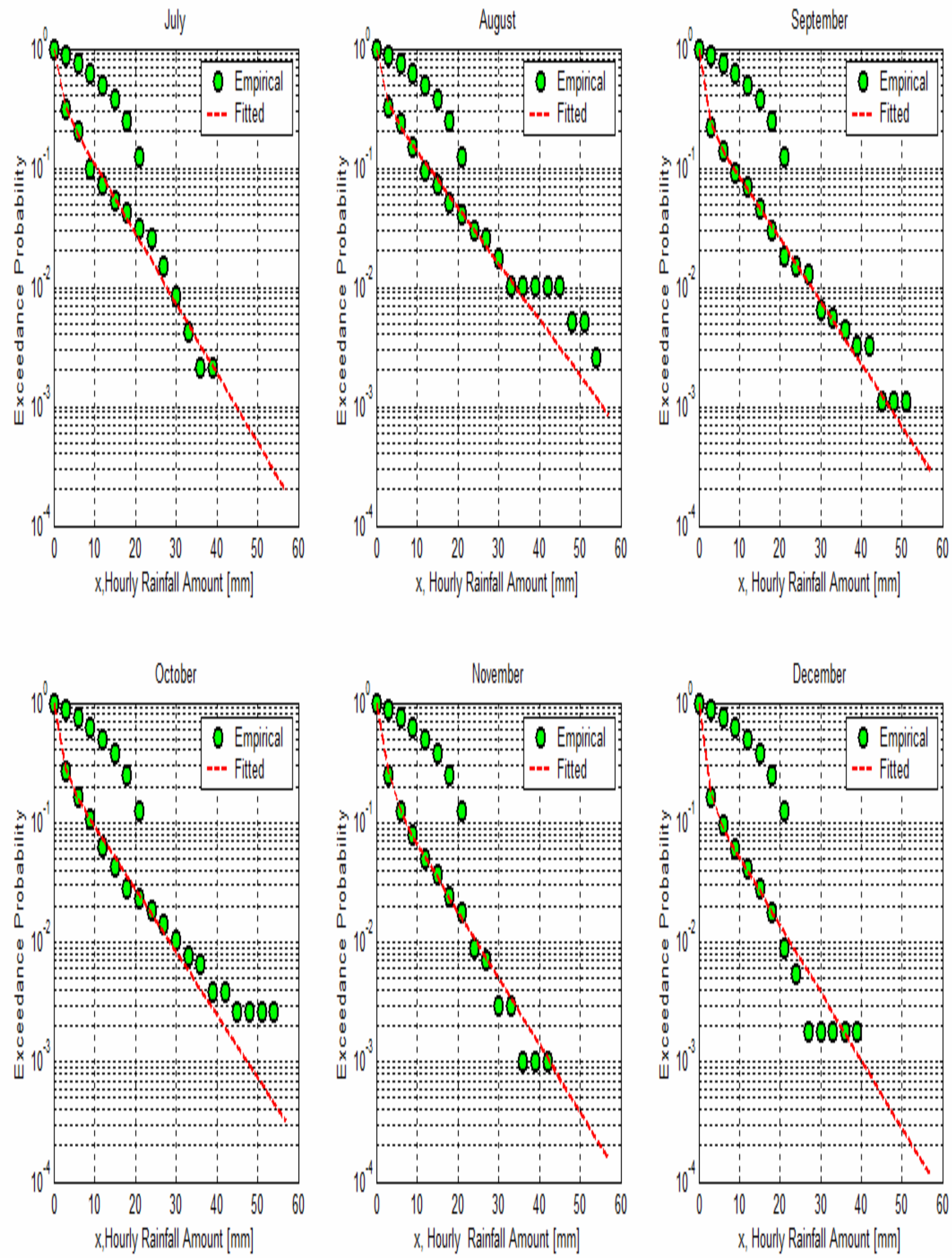


Figure 4.26b: Exceedance Probabilities for Hourly Rainfall from July to December

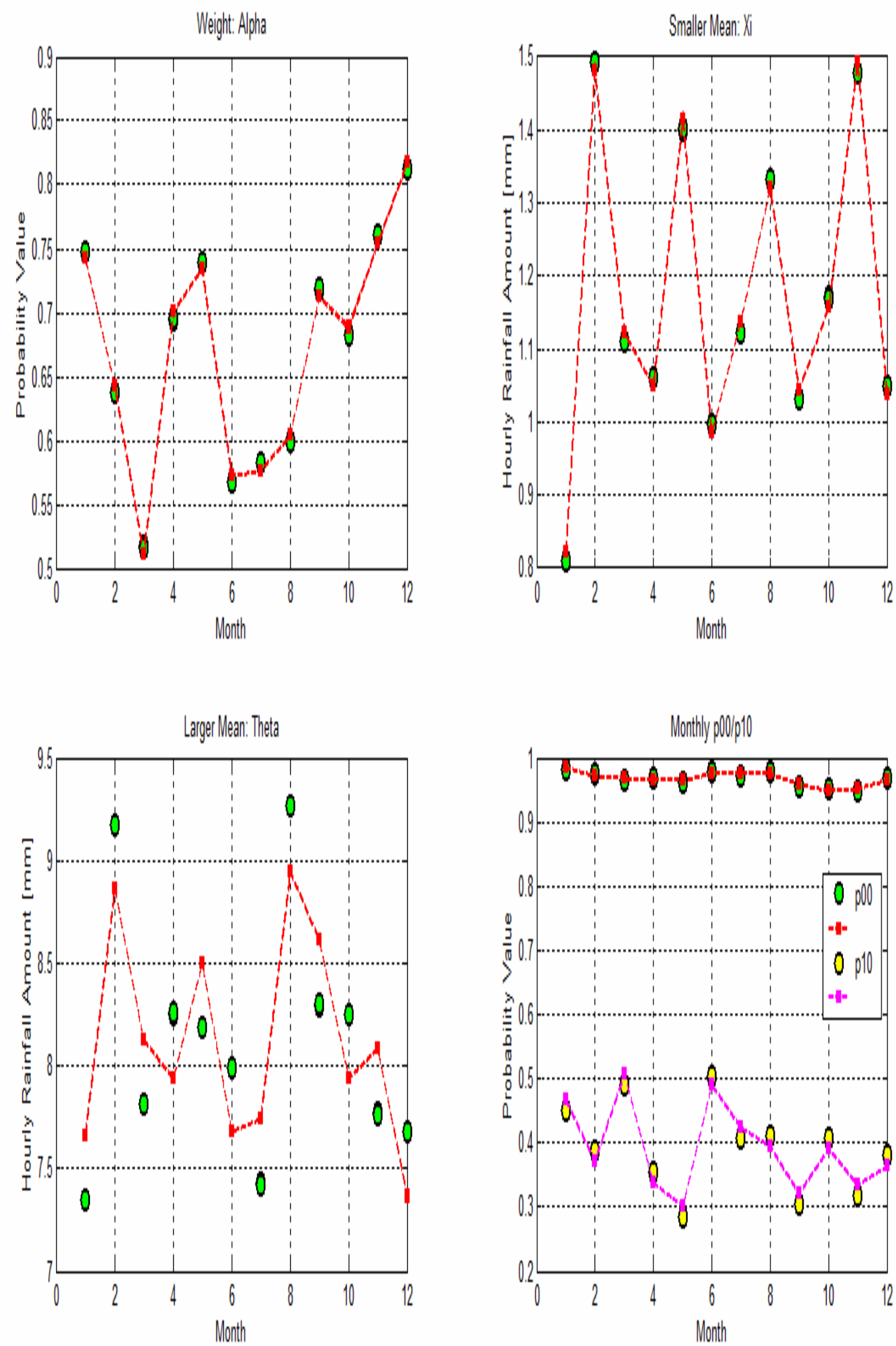


Figure 4.27: Fourier Series Fits (dashed line) and MCME parameters (dots).

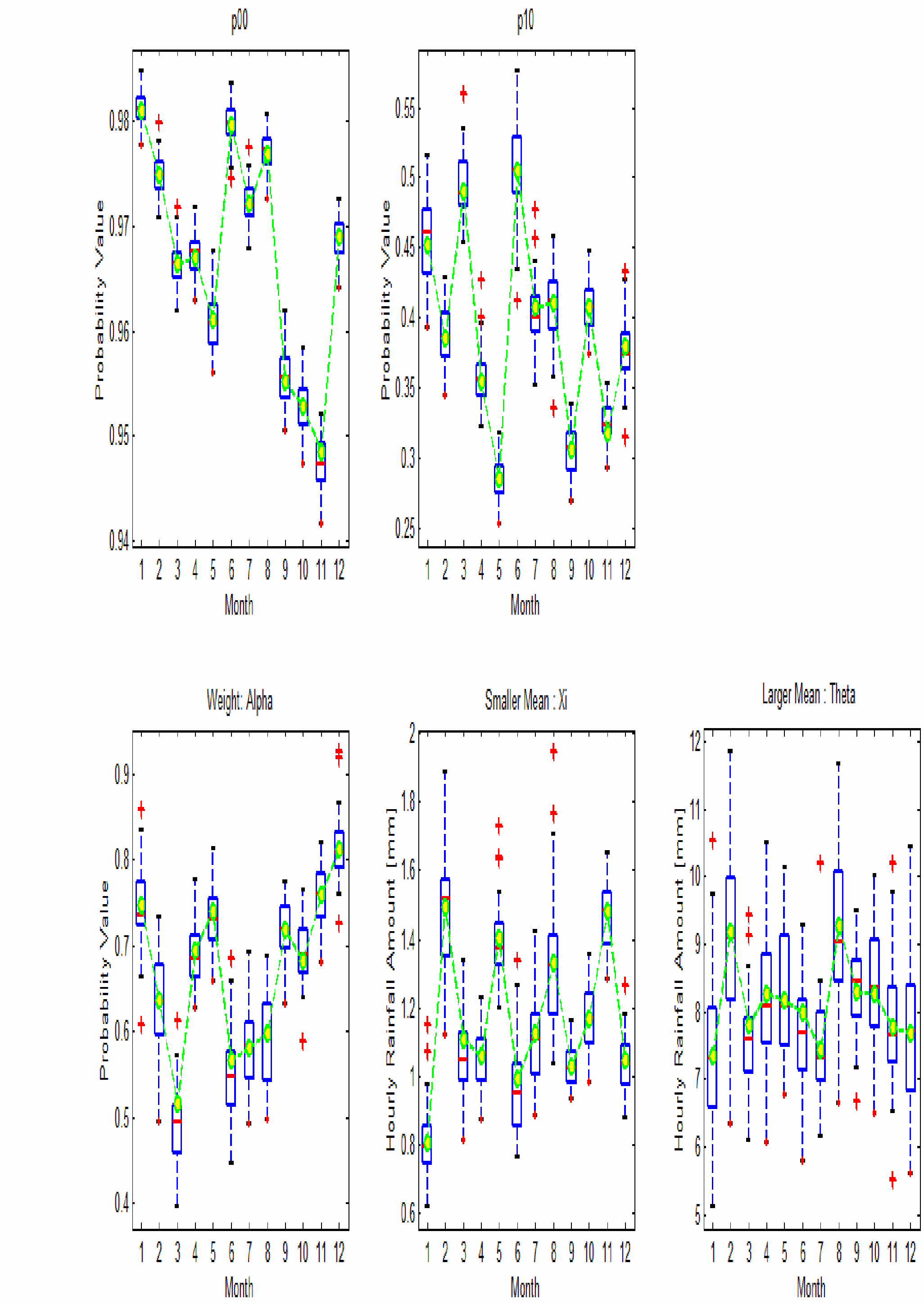


Figure 4.28: Comparison between simulated (box plots) and Empirical (dots connected by dashed lines) MCME parameters of the hourly rainfall series.

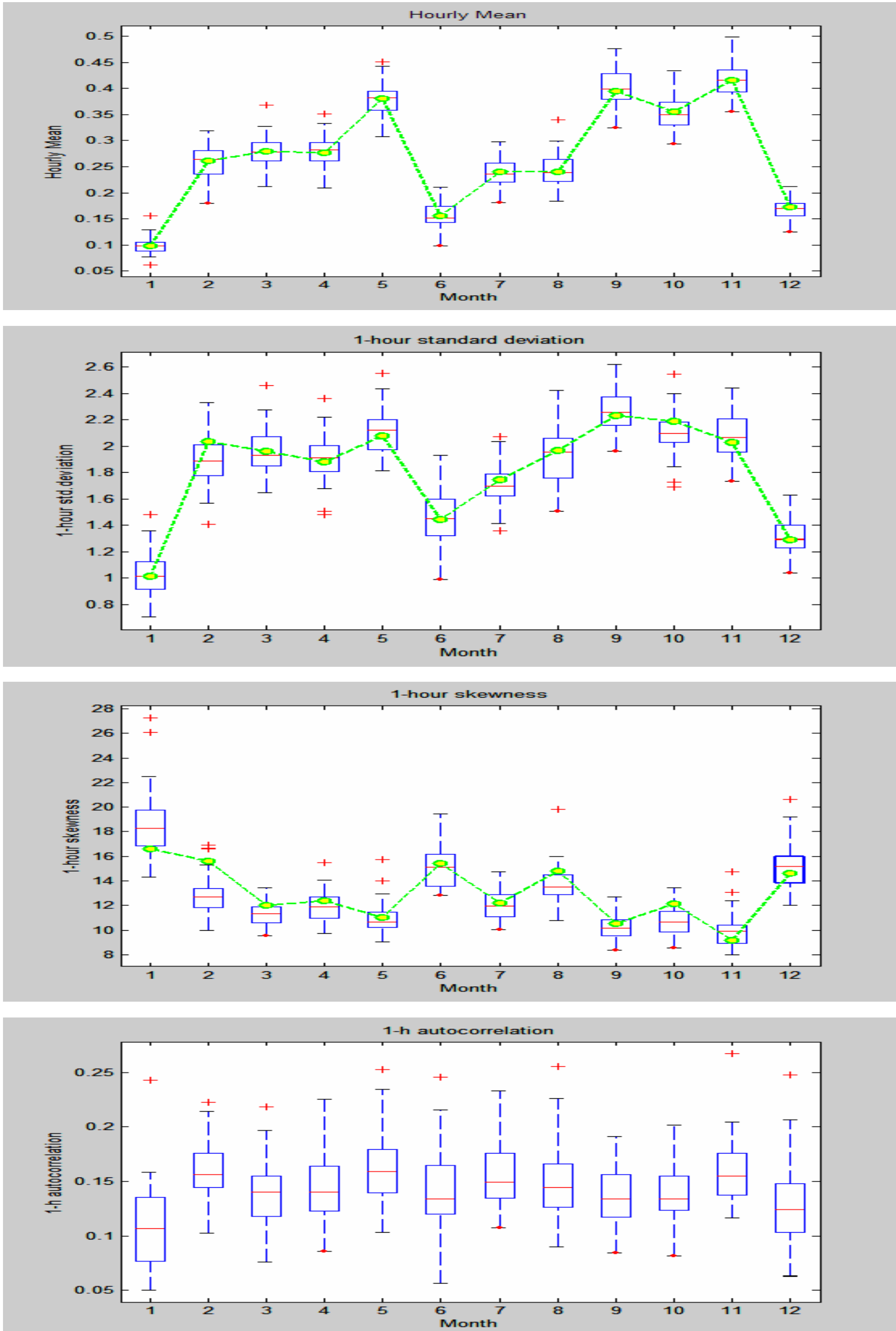


Figure 4.29: Monthly Statistical Properties of 1-Hour Rainfall (in mm) using hourly MCME model

Similarly, there is a higher probability of rain on a given hour if the previous hour was also rainy in September and May. The probability for a wet hour occurring following a dry hour is very low, but more likely in November and September. Therefore, the hourly rainfall occurrence characteristics were well described by the MCME hourly model.

4.3.5.6 Simulated Mixed Exponential Parameters

Figure 4.28 also shows the comparison between the simulated and the empirical MCME parameters. When comparing simulated mixed exponential parameters to the empirical (observed values), it can be seen that the median of the simulated box plots and the empirical parameter values show close agreement in value as well as the trend. There is also a noticeable seasonal periodic variation in all parameters. The higher mean has large range of simulated values. However, the empirical values are still in the middle 50% of the simulated values in the box plots. This is consistent with the Fourier series fit results where the higher mean is not well represented by the Fourier series.

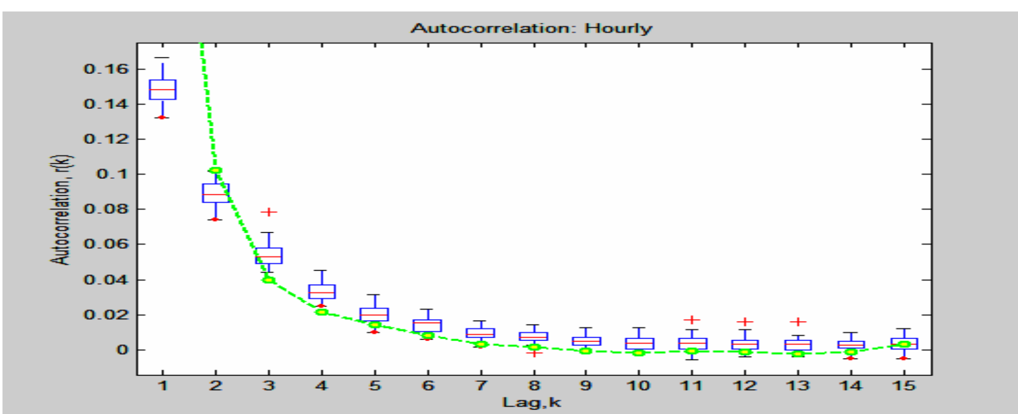


Figure 4.30: Comparison of observed and simulated correlograms of hourly rainfall series using hourly MCME model

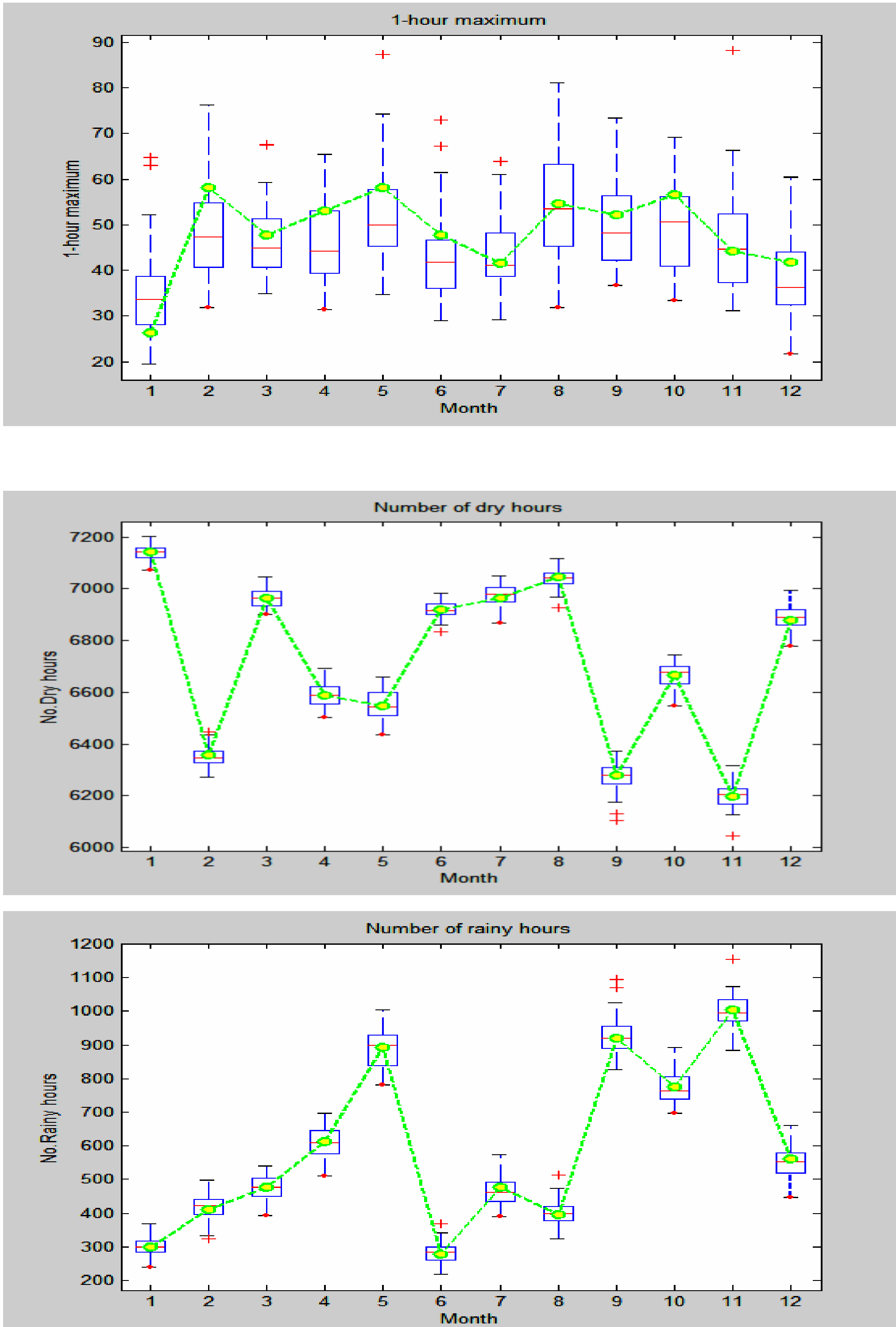


Figure 4.31: Monthly Physical Properties of 1-Hour Rainfall (in mm) using hourly MCME model

4.3.6 Properties of Simulated and Observed Series

Following the comparison of the MCME empirical and simulated parameters, the statistical and physical properties of 50 simulations and the observed hourly rainfall series were compared as shown in the following.

4.3.6.1 Statistical Properties

The statistical properties to be evaluated in this study include mean, standard deviation, coefficient of skewness and correlogram of 1-hour rainfall series evaluated on monthly basis, as shown from Figures 4.29 to 4.30. The observed is represented by the dots connected by dashed lines and the simulated is represented by the box plots. The one-hour rainfall mean has close agreement with the medians of the box plots for the whole year. Similar results obtained for the standard deviation for the one-hour rainfall where the model performed very well with the observed has comparable values with the medians of the simulated box plots throughout the years except in February. The model preserved the one-hour rainfall coefficients of skewness accurately for some of the months. The time dependence characteristics of the hourly rainfall series is basically presented using the correlogram (*r(k*) which is a plot of lag-*k* autocorrelation versus the *k*. The correlograms, *r(k)* values were fairly well reproduced by the model except for lags 1 and 4. Overall, the seasonal variability and trend of the observed properties of rainfall are comparable to the simulated properties. Therefore, the hourly statistical properties of observed rainfalls could be described well by the hourly MCME model.

4.3.6.2 Physical Properties

The physical properties include the hourly maximum and number of dry or rainy hours evaluated on monthly basis. Figure 4.31 shows the physical properties of the observed and simulated hourly rainfall series. For the one-hour maximum rainfall, the observed values were contained in the middle 50% of the box plots for most of the months except in January, February and June. The hourly number of rainy hours and dry hours has shown excellent agreements between the observed values and the medians of simulated properties. The number of rainy hours was the highest in November, followed by September, while the lowest was in June. The number of dry hours was highest in January. The seasonal trends of the properties were well preserved. Therefore, the physical properties of the hourly rainfall were preserved well by the hourly MCME simulation.

4.3.6.3 Lumping to daily rainfall series

A further analysis was conducted to see whether the hourly rainfall data could be “lumped” to form a 24 hourly or a daily equivalent. This analysis was done to determine whether the hourly MCME model is able to reproduce accurately the properties of the rainfall series for daily scale by lumping the hourly data. The simulated statistical and physical properties of the 24-hour rainfalls were then compared to the observed daily data from the same period. Figures 4.32 to 4.33 show the statistical properties of the simulated 24-hour box plots and the observed daily. It can be seen that the simulated 24-hour mean of rainfall has an excellent agreement with observed with the median of the simulated mean has an almost equal value to observed daily mean. The standard deviations of the observed daily rainfall was preserved only fairly well by the model. The coefficients of skewness of daily rainfall was underestimated in April and July but the daily rainfall autocorrelations was underestimated in most of the months by the hourly MCME model. The daily correlogram only shows an excellent fit with the medians of the simulated 24-hour correlogram at lags 4, 10, and 15.

In general, the lumped daily performance was not as good as the hourly performance in preserving the statistical properties of the observed. Nevertheless, the seasonal trends of the daily rainfall properties were very well preserved by the hourly MCME model.

Figure 4.34 shows the physical properties of the 24-hour simulated series and the daily observed series. The daily maximum rainfall of the observed values were managed to be captured in the middle 50% of simulated in the box plots in February, March, May, August, October and November only. However, this is not true for the simulation of the number of rainy days and the number of dry days, where the middle 50% of the simulated values in the box plots do not manage to capture the observed values. While the physical properties of the observed daily series were unable to be matched accurately in the hourly MCME simulations, the seasonal trend of the observed daily properties was very well preserved.

4.3.6.4 Lumping to monthly rainfall series

Following the lumping of the hourly to form a daily series, a further lumping was done to form a 1-month scale data. This analysis was done to determine whether the MCME hourly model is able reproduce accurately the monthly properties of the rainfall series as well. The simulated statistical and physical properties of the 1-month rainfalls amounts were then compared to the observed monthly data for the same period. Figure 4.35 shows the statistical properties of the simulated lumped monthly box plots and the observed monthly properties. It can be seen that the simulated monthly mean of rainfall has an excellent agreement with the observed. However, the standard deviation of the observed monthly rainfall series were unable to be captured in the middle 50% of the box plots for the whole year. The simulated monthly maximum and minimum or rainfall were also unable to capture the observed in the middle 50% in most of the months. Hence, the properties of 1-month scale rainfall amount were not able to be preserved accurately by the MCME hourly simulations. Nevertheless, the hourly MCME model preserved the seasonal trends of the observed monthly rainfall series.

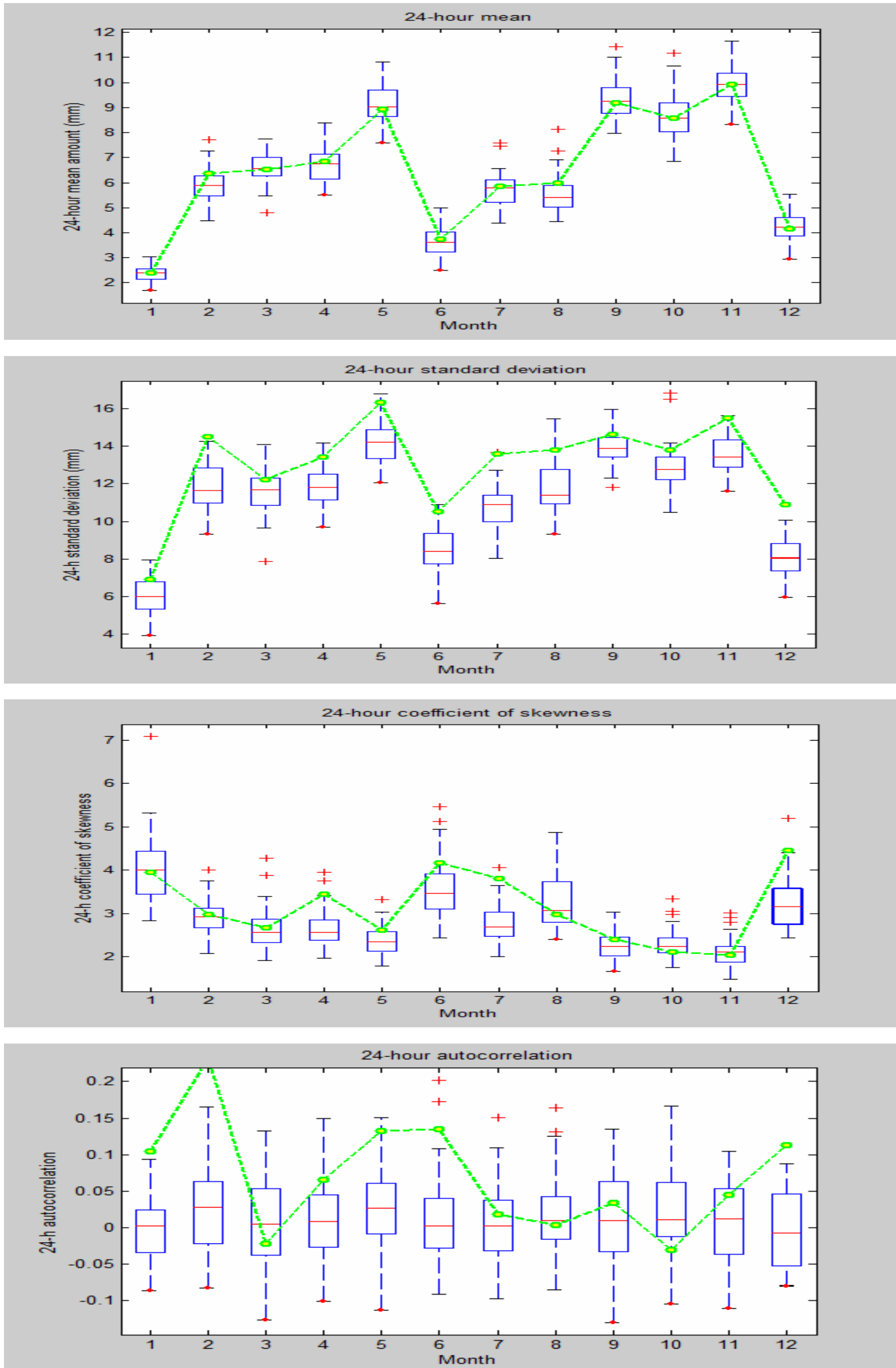


Figure 4.32: Monthly Statistical Properties of 24-Hour Rainfall (in mm) using hourly MCME model

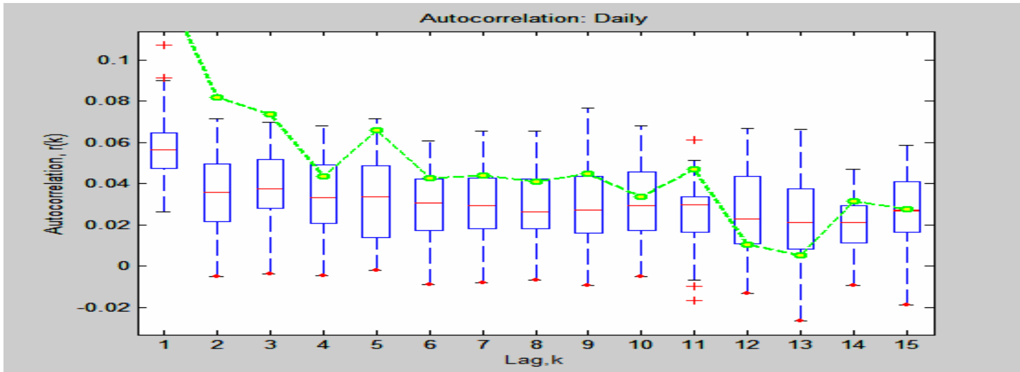


Figure 4.33: Comparison of observed and simulated correlograms of 24 hourly rainfall series using hourly MCME model

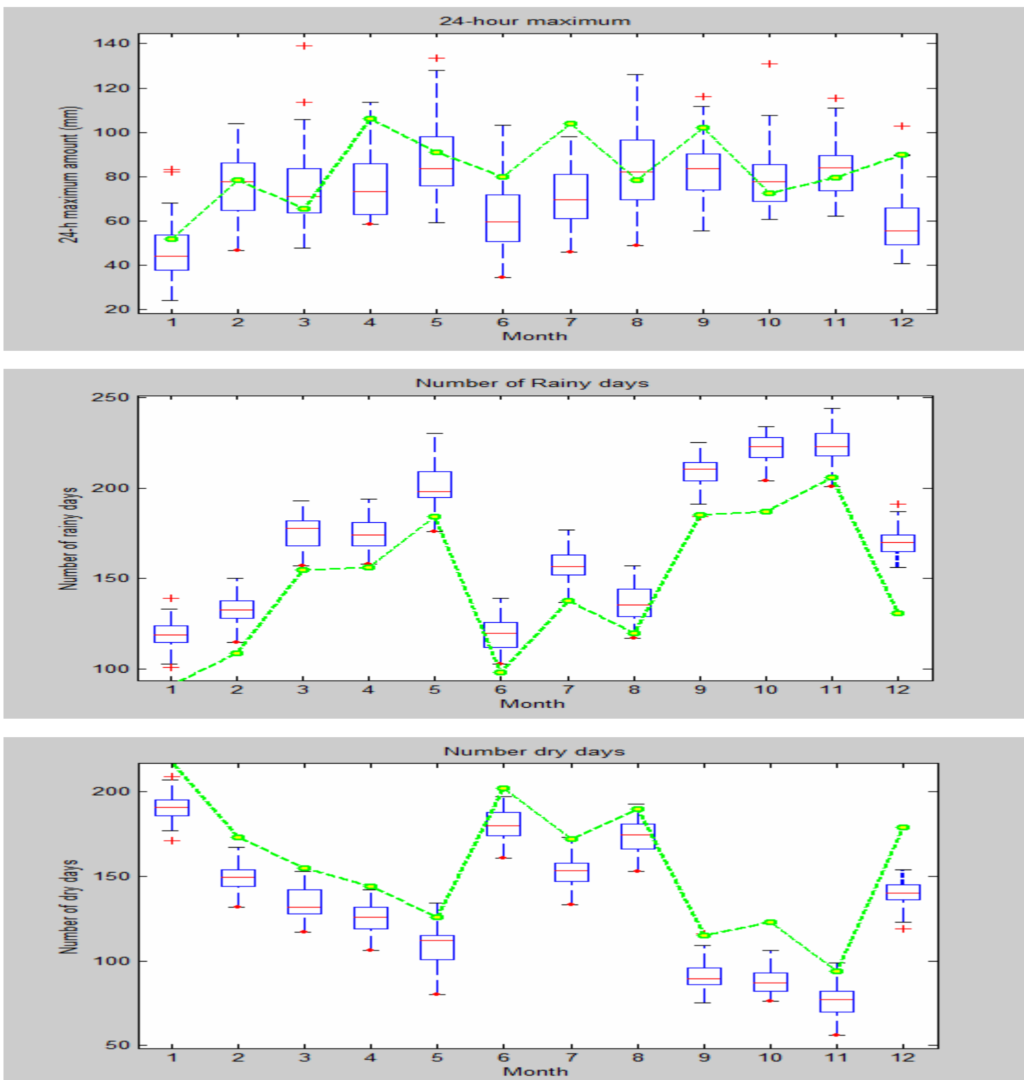


Figure 4.34: Monthly Physical Properties of 24-Hour Rainfall (in mm) using hourly MCME model

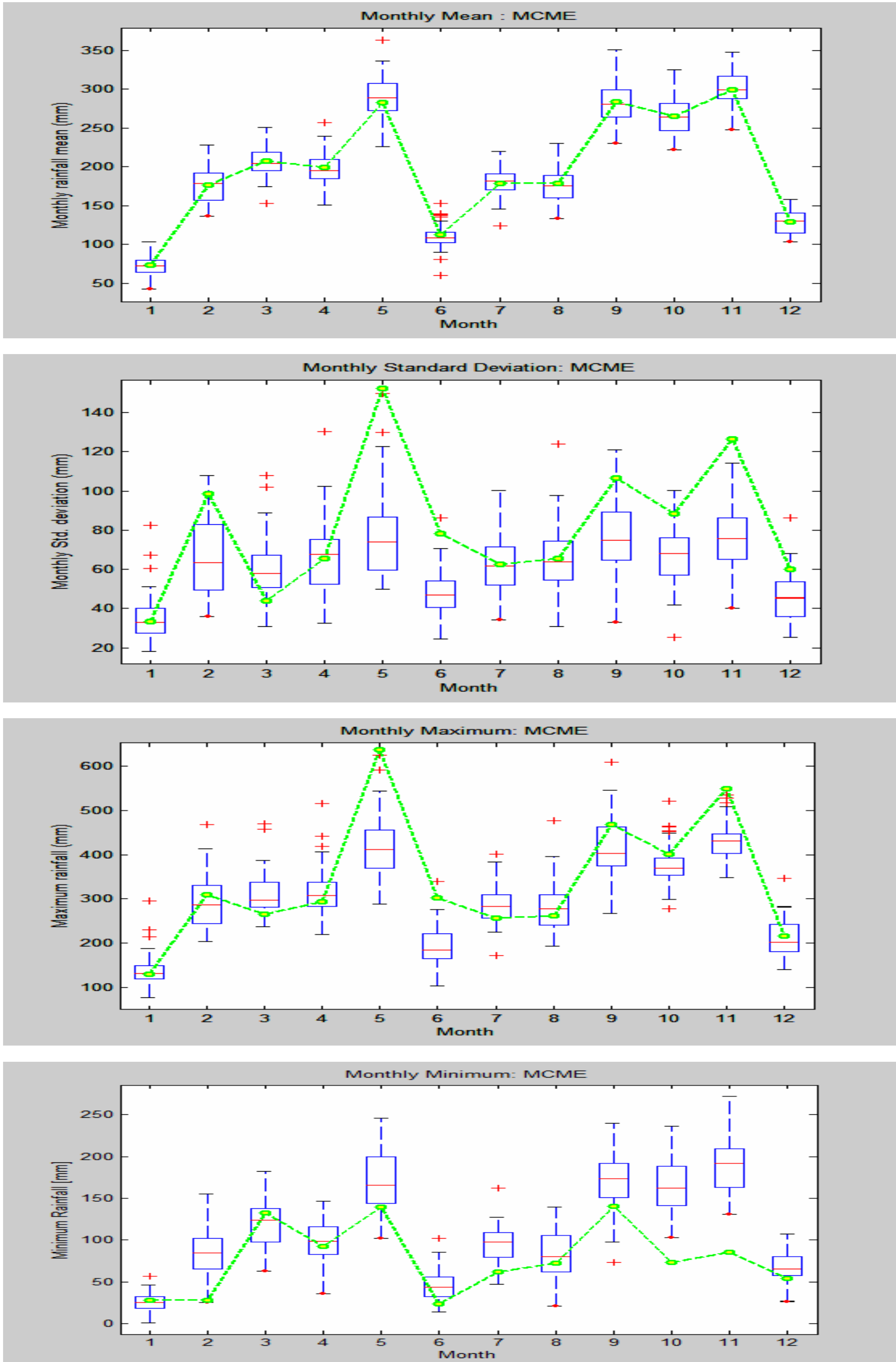


Figure 4.35: Monthly Properties of 1-Month Rainfall (in mm) using hourly MCME model

4.3.7 Validation of the NSRP and MCME models

In assessing the predictive ability of the models, the simulations of both models were extended to 20 years. The first 10 years were used to assess the models' descriptive ability which has been covered in Chapter 4, while the last 10 years was used to assess their predictive ability. The last 10 years simulation was compared with the observed series from 1991 to 2000.

4.3.7.1 Validation of the NSRP model

The descriptive ability of the NSRP model has been discussed. It was found that the NSRP model with mixed exponential distribution to represent the rain cell intensity and combined with the use of the transition probabilities of rainfall occurrences in the estimation of parameters procedures also referred, as the MEXPTRAN was the best model to represent the NSRP. Therefore, to determine the strength of the model in extrapolating beyond the data points (1981-1990), the predictive ability is to be assessed to ensure that the model has the ability to be used in simulating the hourly rainfall series at any length and at any data points.

Figure 4.36 shows the comparison between the observed and the simulated statistical properties of rainfall for the one-hour scale. The one-hour observed rainfall mean was matched very well only in January, July and October. However, the standard deviation of one-hour rainfall was matched within the range of the maximum and minimum value of the box plots in most months except in February, June and July. The coefficients of skewness of one- hour rainfall could be adequately preserved for January, February, May and July. However, the autocorrelation of one-hour rainfall was underestimated in most of the months.

Figure 4.37 shows the comparison between the observed and the simulated physical properties of rainfall for the one-hour scale. The maximum of one-hour rainfall was matched well and within the range in all months except in September. However, the probability of dry hours of rainfall series was poorly matched and underestimated in most of the months.

Figure 4.38 shows the comparison between the observed and the simulated statistical properties of rainfall at the 24-hour scale. The mean of 24-hour rainfall was matched excellently in January, July and October. The variances, autocorrelations and coefficients of skewness of the 24-hour rainfall were fairly matched within the range of the simulated properties.

Figure 4.39 shows the comparison between the observed and the simulated physical properties of rainfall at the 24-hour scale. The maximum 24-hour rainfall. was matched fairly well for the whole year. However, the probability of dry days was either overestimated or underestimated for some months.

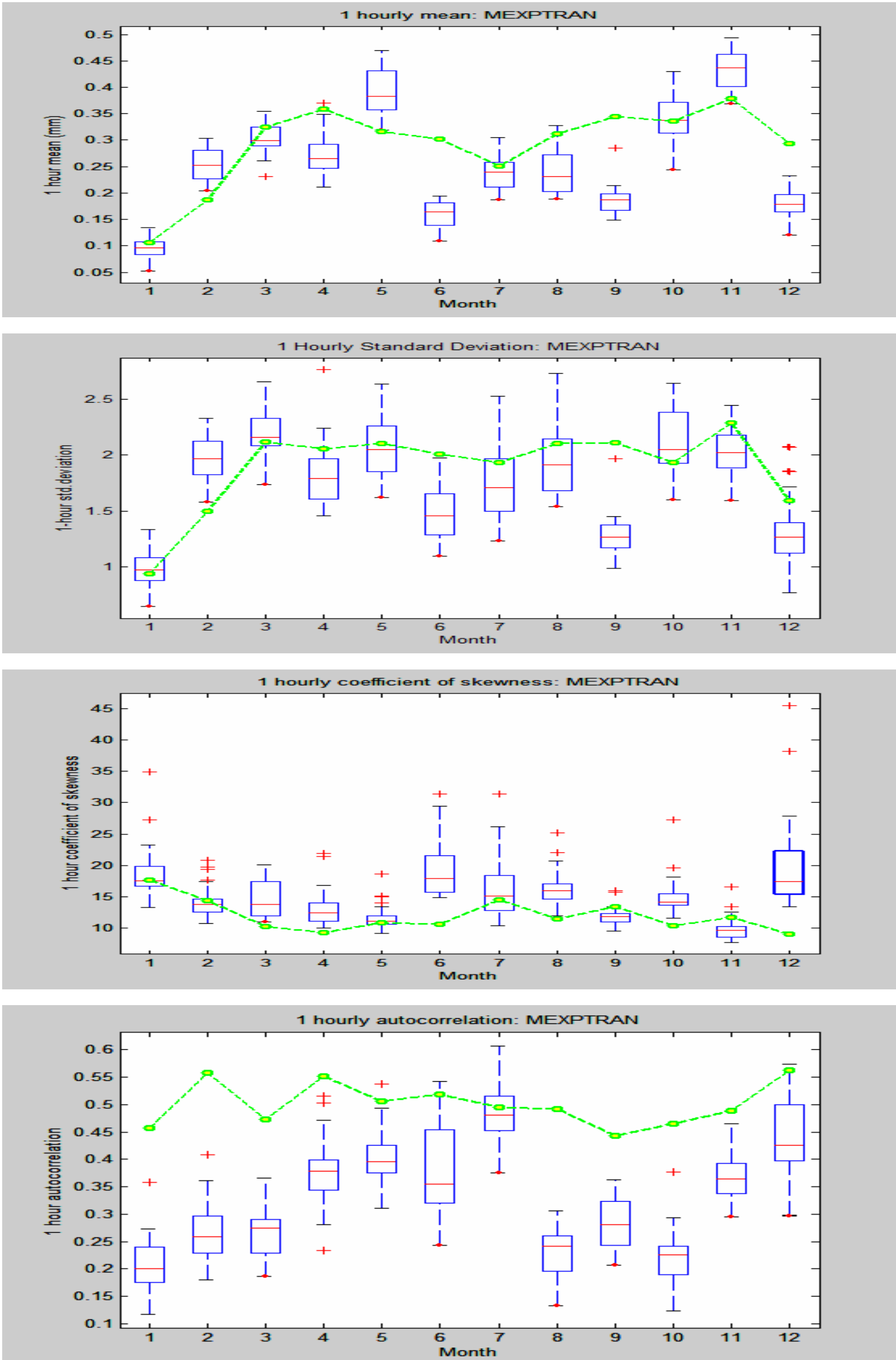


Figure 4.36: Validation of Monthly Statistical Properties of 1-hour Rainfall (mm) of MEXPTRAN model

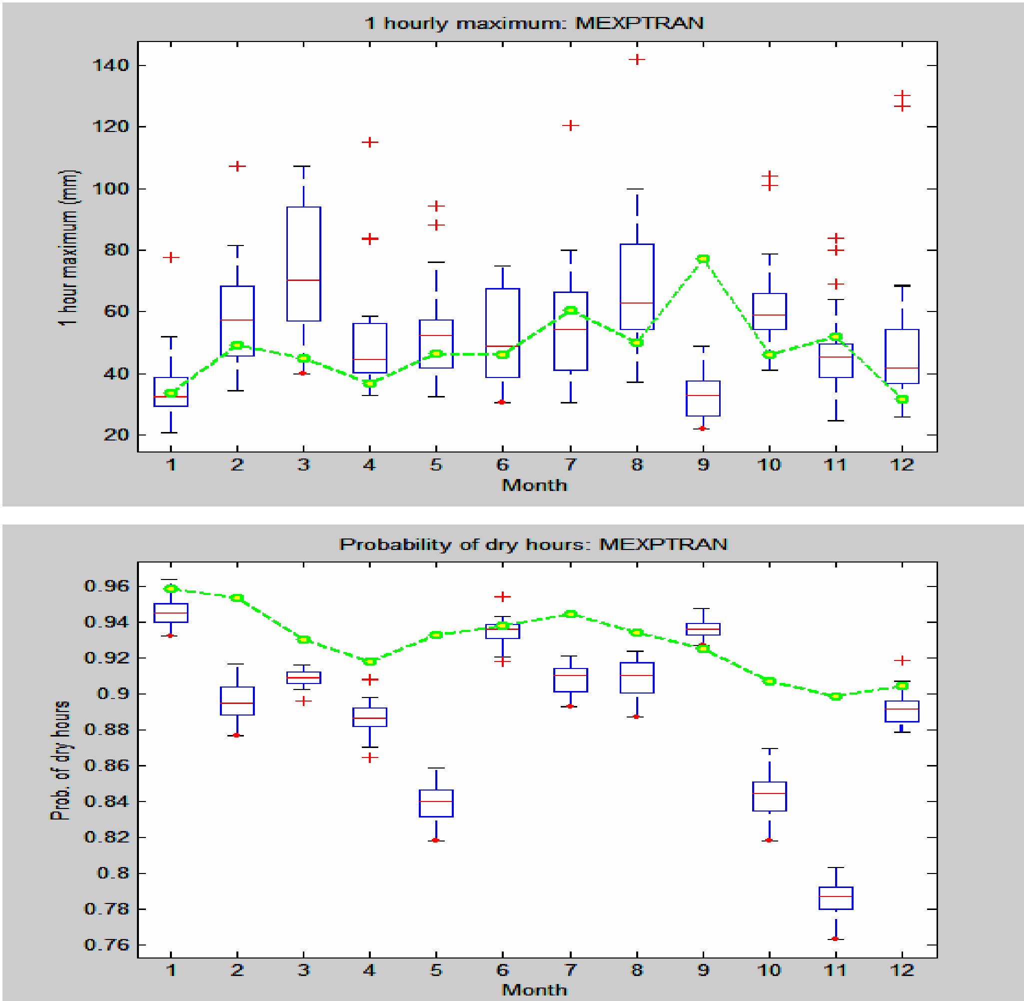


Figure 4.37: Validation of Monthly Physical Properties of 1-hour Rainfall (mm) of MEXPTRAN model

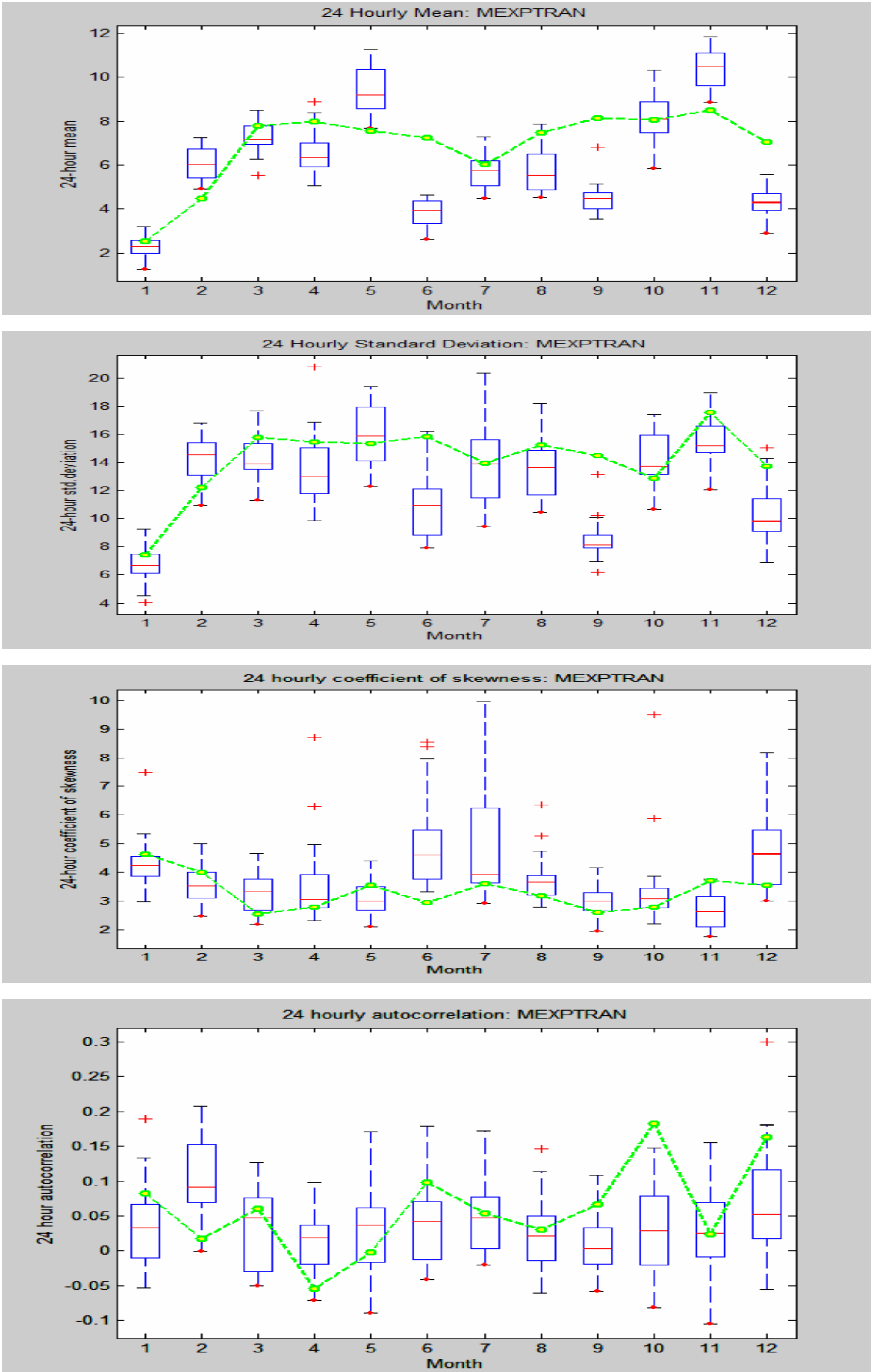


Figure 4.38: Validation of Monthly Statistical Properties of 24-Hour Rainfall of MEXPTRAN model

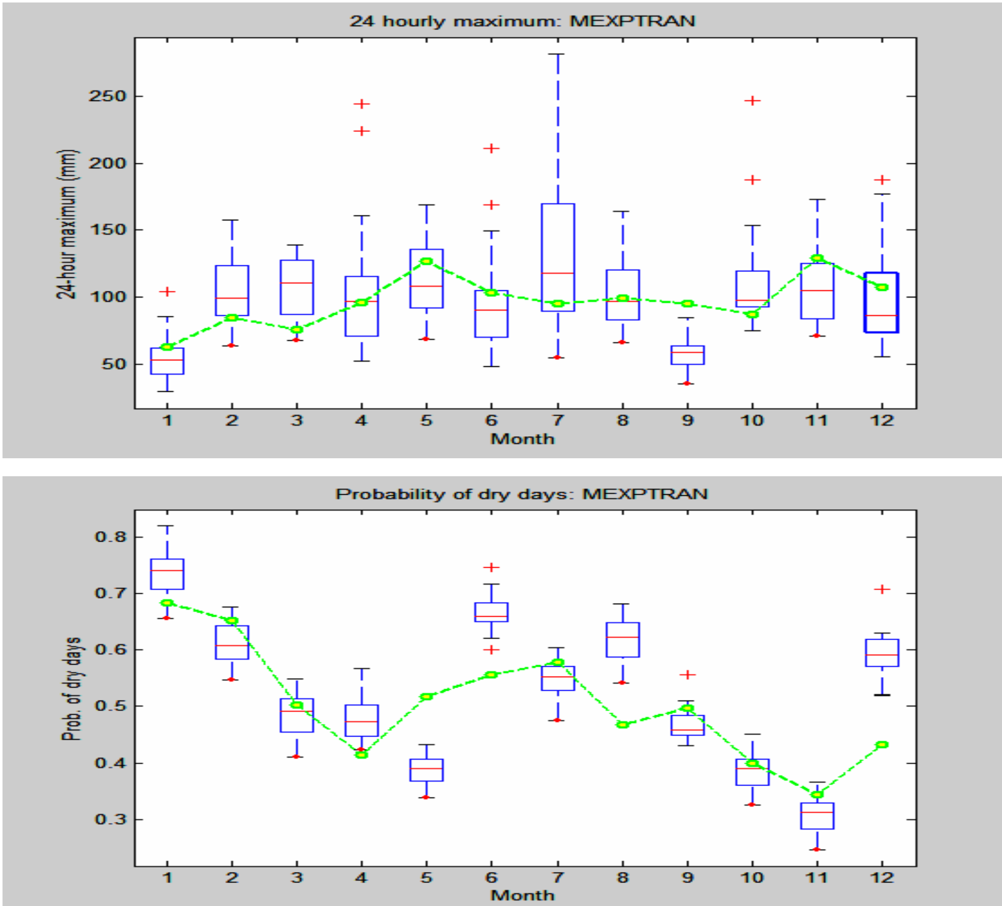


Figure 4.39: Validation of Monthly Physical Properties of 24-Hour Rainfall (mm) of MEXPTRAN

In general, the performance of the NSRP model in the validation period was not as good as in the calibration period. The model was unable to predict the properties of the observed rainfall accurately at various timescales. However, the model preserved the seasonal trends of the observed rainfall properties

4.3.7.2 Validation of the MCME model

The validation of this model was done using the hourly and the daily simulation of the MCME model. Figure 4.40 shows the comparison between the observed and the simulated statistical properties of rainfall for the one-hour scale. The means, variances and coefficients of skewness of one-hour rainfall were fairly matched in which most of the observed rainfall properties fall outside the box but within the range of the simulated values. However, the lag-1 autocorrelation of one-hour rainfall was much underestimated.

Figure 4.41 shows the comparison between the observed and the simulated physical properties of rainfall for the one-hour scale. The probability of dry hours was poorly matched in most of the months. However, the maximum one-hour rainfall was matched very well in most of the months except in September where the observed values were underestimated.

Figures 4.42 show the comparison between the observed and the simulated statistical properties for the 24-hour scale (lumped daily). Similar performance as in the 1-hour scale was seen for the mean, variances and coefficients of skewness of the 24-hour rainfalls where all these properties were only fairly matched. However, the autocorrelations of 24-hour rainfall could be preserved fairly well by the model simulation.

Figure 4.43 shows the comparison between the observed and the simulated physical properties for the 24-hour scale (lumped daily). The maximum 24-hour rainfall was only fairly matched with the medians of the box plots for the whole year. The probability of dry days was seen to be either underestimated or overestimated in some of the months.

Therefore, the performance of the hourly MCME model in the validation period was also not as good as in the calibration period. The model was unable to preserve the properties of the observed rainfall accurately at one-hour scale as well as at 24-hour scale. However, the model could preserve the seasonal trends of the rainfall series.

The validation of the MCME model was also done using the daily model. Figure 4.44 shows the comparison between the observed and the simulated statistical properties of rainfall at daily scale. The means, standard deviations, coefficients of skewness and autocorrelations of the daily rainfalls were only fairly matched in most of the months.

Figures 4.45 shows the comparison between the observed and the simulated physical properties of rainfall at daily scale. The probability of dry days was also fairly matched in most of the months. However, the daily maximum rainfall could be adequately preserved for the whole year.

In general, the performance of the daily MCME model in describing the daily rainfall process during the validation period was not as good as in the calibration period. However, the predictive ability of the daily MCME model in predicting the daily rainfall process can be considered more accurate than the lumped daily from the hourly MCME model. This may be justified by the RMSE evaluated from the monthly square errors between the observed and the medians for the daily properties as given in Table 4.14. It clearly shows that the daily MCME simulation has smaller RMSE in all properties considered. Therefore, the daily MCME model has better ability in predicting the properties of daily rainfall series.

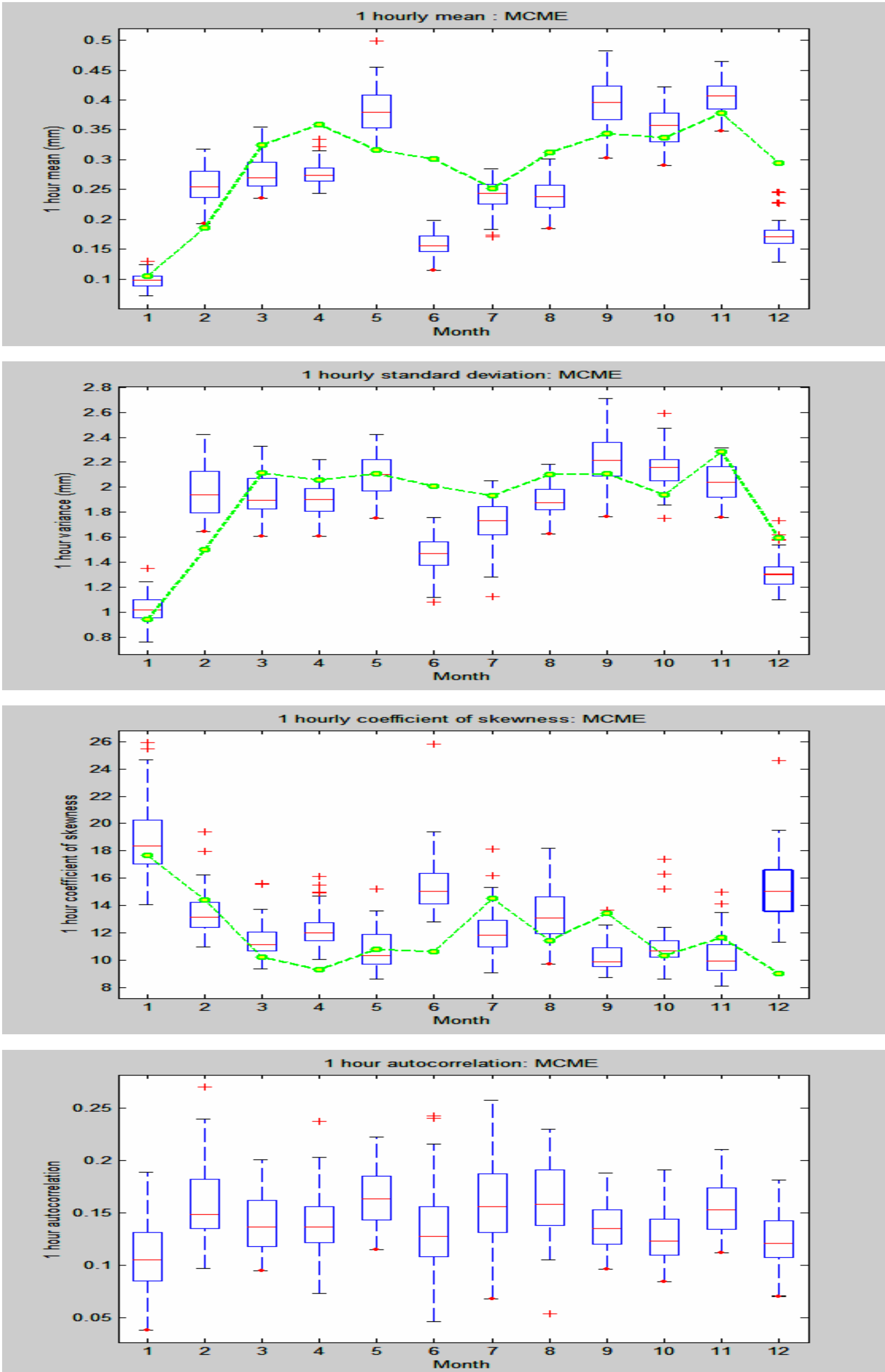


Figure 4.40: Validation of Monthly Statistical Properties of 1-hour Rainfall (mm) of hourly MCME model

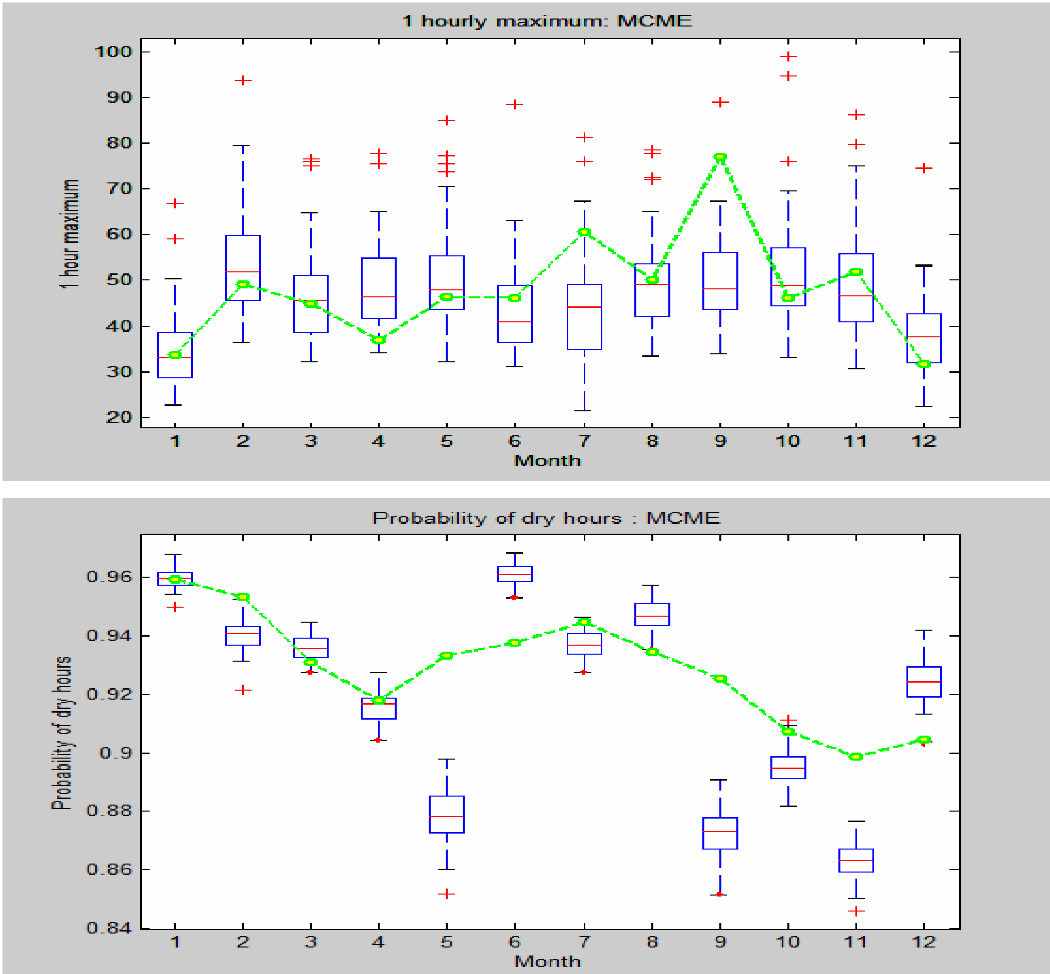


Figure 4.41: Validation of Monthly Physical Properties of 1-hour Rainfall (mm) of hourly MCME model

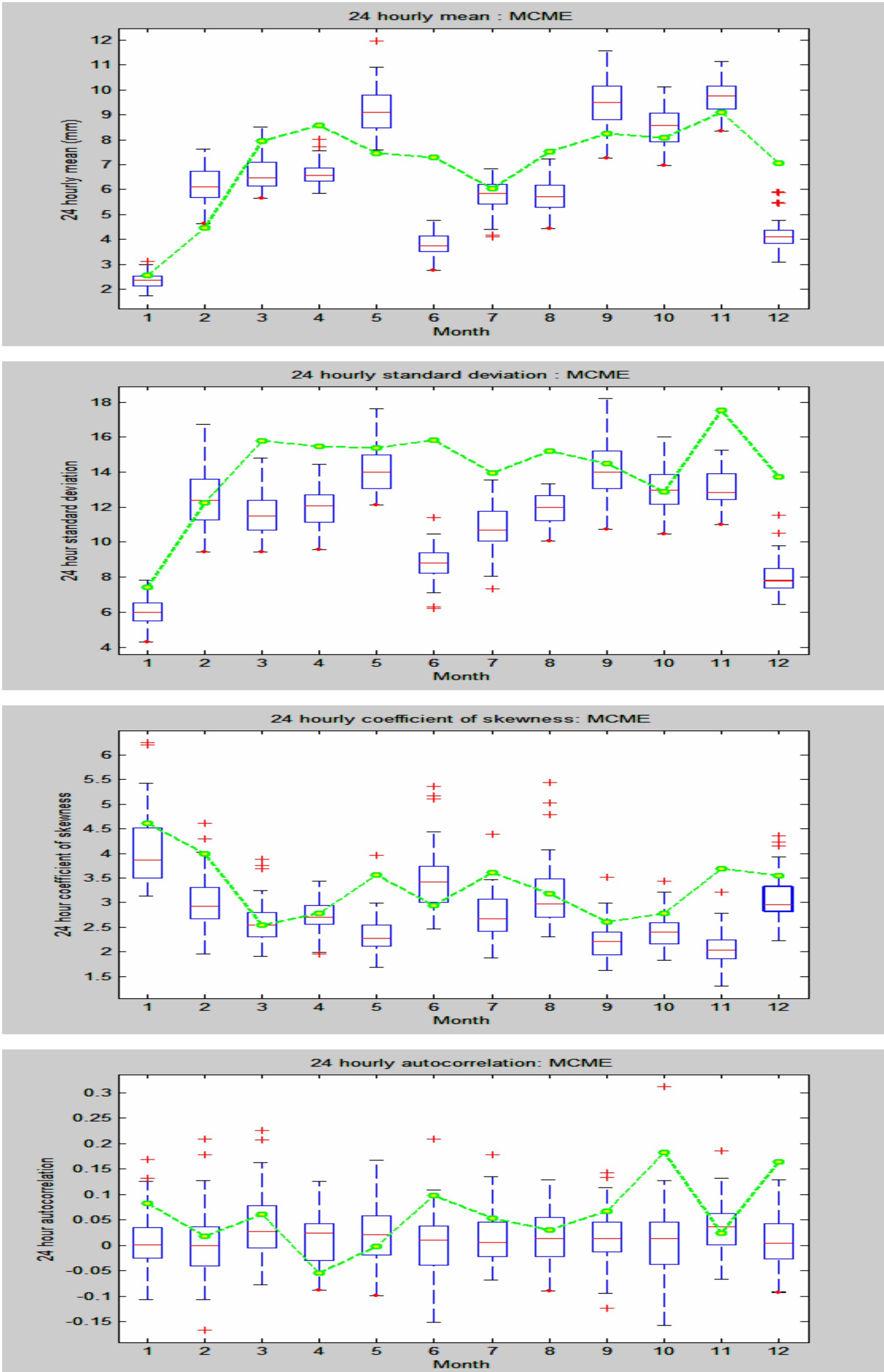


Figure 4.42: Validation of Monthly Statistical Properties of 24-hour Rainfall (mm) of hourly MCME model

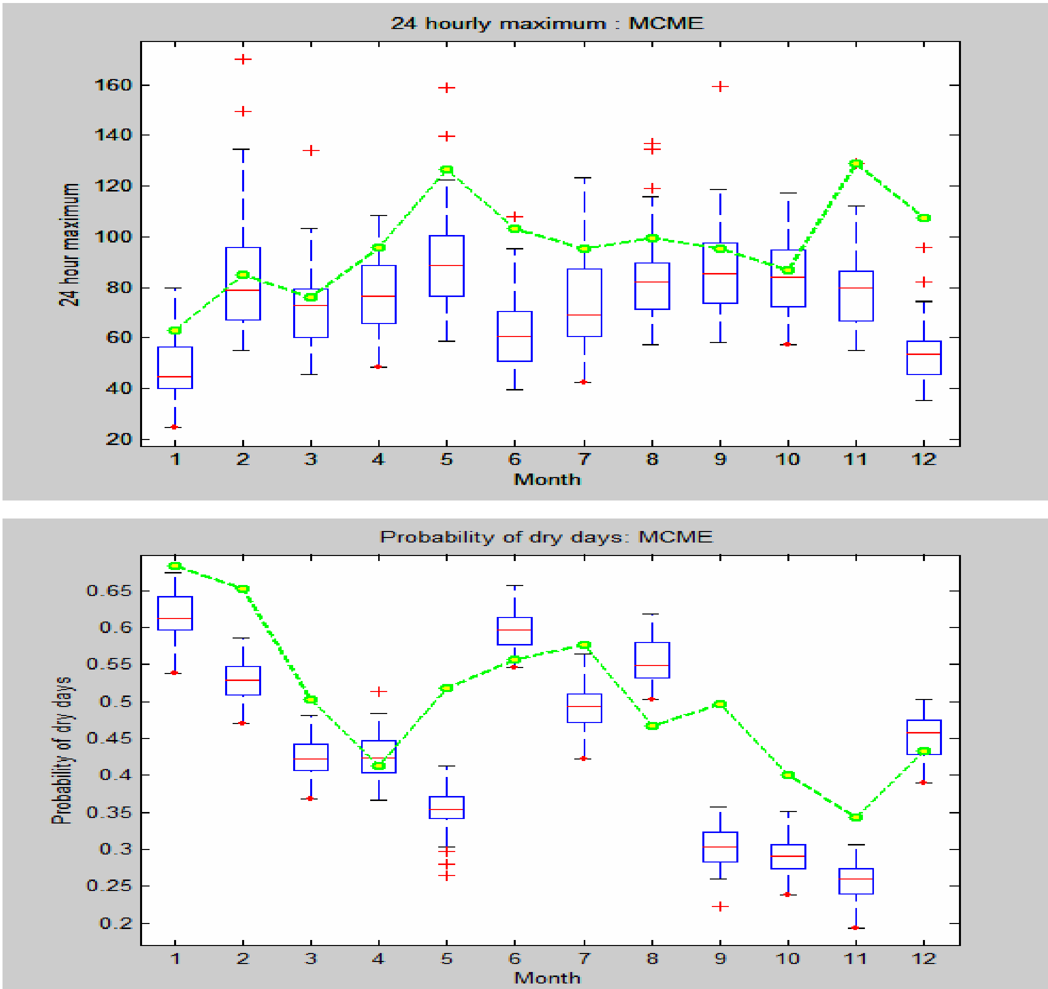


Figure 4.43: Validation of Monthly Physical Properties of 1-hour Rainfall (mm) of hourly MCME model

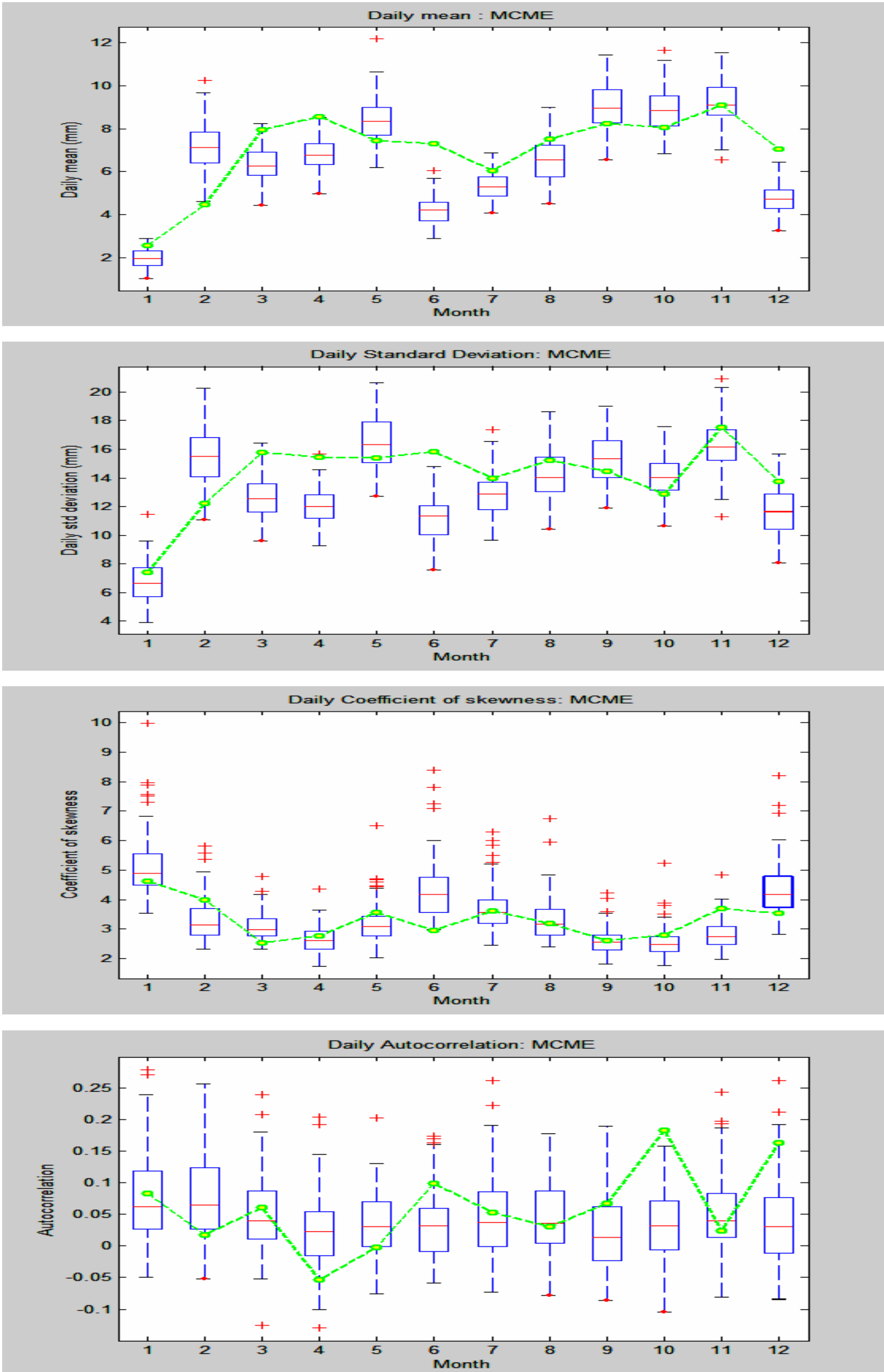
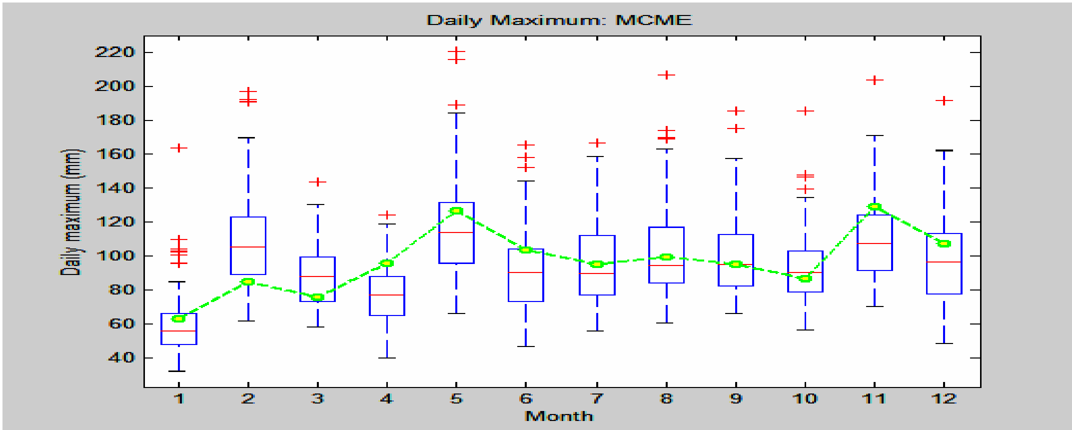


Figure 4.44: Validation of Monthly Statistical Properties of Daily Rainfall (mm) of daily MCME model



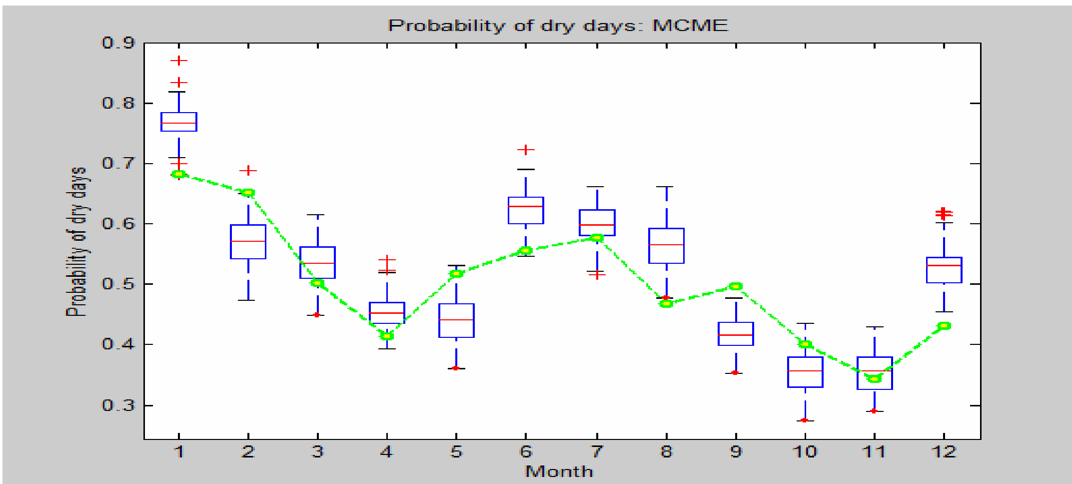


Figure 4.45: Validation of Monthly Physical Properties of 1-hour Rainfall (mm) of daily MCME model

Table 4.14: The RMSE of the MCME models at 24-hour scale in the validation period (1991-2000)

24-Hour Mean

Month	Hourly MCME	Daily MCME
Jan	0.0259	0.3444
Feb	2.5948	6.9634
Mar	1.7067	2.2995
Apr	2.0155	1.5293
May	2.3201	0.6087
Jun	12.1441	9.1436
Jul	0.0313	0.5505
Aug	3.1964	0.9155
Sep	1.8324	0.6572
Oct	0.2392	0.5778
Nov	1.6205	0.3758
Dec	8.6367	5.4281
SSE	3.0303	2.4495
RMSE	**1.7408**	**1.5651**

24-Hour Autocorrelation

Month	Hourly MCME	Daily MCME
Jan	0.0067	0.0004
Feb	0.0003	0.0022
Mar	0.0011	0.0004
Apr	0.0061	0.0058
May	0.0005	0.0010
Jun	0.0079	0.0045
Jul	0.0023	0.0003
Aug	0.0003	0.0000
Sep	0.0029	0.0028
Oct	0.0284	0.0225
Nov	0.0002	0.0003
Dec	0.0252	0.0177
SSE	0.0068	0.0048
RMSE	**0.0826**	**0.0695**

24-Hour Standard Deviation

Month	Hourly MCME	Daily MCME
Jan	2.0274	0.6192
Feb	0.0170	10.6289
Mar	18.3808	10.3509
Apr	11.5406	12.1602
May	1.8458	0.9436
Jun	49.6250	20.5073
Jul	10.5633	1.1452
Aug	10.6138	1.3874
Sep	0.2263	0.7298
Oct	0.0123	1.3482
Nov	22.1177	1.9125
Dec	35.3403	4.4554
SSE	13.5259	5.5157
RMSE	**3.6778**	**2.3486**

24-Hour Maximum

Month	Hourly MCME	Daily MCME
Jan	331.2400	51.8400
Feb	33.9889	424.3600
Mar	11.6964	142.3249
Apr	375.9721	342.2500
May	1437.1681	166.4100
Jun	1840.4100	170.3025
Jul	676.0000	31.6969
Aug	299.6361	25.5025
Sep	89.4916	0.0900
Oct	7.6729	10.3041
Nov	2424.5776	470.8900
Dec	2893.3641	123.2100
SSE	868.4348	163.2651
RMSE	**29.4692**	**12.7775**

24-Hour Coefficient of Skewness

Month	Hourly MCME	Daily MCME
Jan	0.5660	0.0711
Feb	1.1179	0.7333
Mar	0.0001	0.2049
Apr	0.0052	0.0241
May	1.6479	0.2150
Jun	0.2169	1.5146
Jul	0.8843	0.0028
Aug	0.0456	0.0001
Sep	0.1547	0.0024
Oct	0.1504	0.0985

Probability of Dry Days

Month	Hourly MCME	Daily MCME
Jan	0.0049	0.0071
Feb	0.0154	0.0064
Mar	0.0065	0.0010
Apr	0.0001	0.0016
May	0.0267	0.0058
Jun	0.0016	0.0052
Jul	0.0070	0.0004
Aug	0.0065	0.0097
Sep	0.0374	0.0064
Oct	0.0120	0.0019

Nov	2.7492	0.9045
Dec	0.3511	0.4128
SSE	0.6574	0.3487
RMSE	**0.8108**	**0.5905**

Nov	0.0069	0.0002
Dec	0.0007	0.0100
SSE	0.0105	0.0046
RMSE	**0.1024**	**0.0682**

Table 4.15: The RMSE of the Hourly MCME and the MEXPTRAN models at one-hour scale in the validation period (1991-2000)

One-Hour Mean

Month	MEXPTRAN	Hourly MCME
Jan	8.51E-05	4.46E-05
Feb	4.35E-03	4.59E-03
Mar	6.38E-04	2.96E-03
Apr	8.70E-03	7.26E-03
May	4.55E-03	4.03E-03
Jun	1.90E-02	2.11E-02
Jul	1.27E-04	5.50E-05
Aug	6.64E-03	0.0055
Sep	2.46E-02	0.0027
Oct	1.94E-07	0.0004
Nov	3.32E-03	0.0008
Dec	1.32E-02	0.0150
SSE	7.10E-03	0.0054
RMSE	**8.43E-02**	**0.0733**

One-Hour Autocorrelation

Month	MEXPTRAN	Hourly MCME
Jan	0.0654	1.24E-01
Feb	0.0890	1.67E-01
Mar	0.0395	1.13E-01
Apr	0.0302	1.73E-01
May	0.0120	1.17E-01
Jun	0.0268	1.53E-01
Jul	0.0002	1.15E-01
Aug	0.0626	0.1116
Sep	0.0260	0.0947
Oct	0.0569	0.1162
Nov	0.0155	0.1127
Dec	0.0189	0.1955
SSE	0.0369	0.1328
RMSE	**0.1922**	**0.3644**

One-Hour Standard Deviation

Month	MEXPTRAN	Hourly MCME
Jan	1.3E-03	6.58E-03
Feb	2.2E-01	1.93E-01
Mar	1.8E-03	4.91E-02
Apr	7.2E-02	2.52E-02
May	3.6E-03	2.73E-06
Jun	3.1E-01	2.89E-01
Jul	4.9E-02	4.11E-02
Aug	3.61E-02	0.0511
Sep	7.13E-01	0.0109
Oct	1.31E-02	0.0482
Nov	6.92E-02	0.0610
Dec	1.06E-01	0.0866
SSE	1.33E-01	0.0718
RMSE	**3.64E-01**	**0.2680**

One-Hour Maximum

Month	MEXPTRAN	Hourly MCME
Jan	1.3059	3.97E-01
Feb	69.5681	7.08E+00
Mar	639.1593	2.60E-01
Apr	60.4633	9.14E+01
May	35.5229	1.93E+00
Jun	7.1642	2.69E+01
Jul	40.6068	2.70E+02
Aug	167.3517	0.7396
Sep	1947.0774	839.8404
Oct	159.8663	7.3984
Nov	41.7323	28.4089
Dec	101.4613	33.4084
SSE	272.6066	108.9505
RMSE	**16.5108**	**10.4379**

One-Hour Coefficient of Skewness

Month	MEXPTRAN	Hourly MCME
Jan	0.0415	4.68E-01
Feb	0.4754	1.52E+00
Mar	12.7860	9.32E-01
Apr	9.4990	7.22E+00
May	0.1182	1.82E-01
Jun	52.4319	1.94E+01
Jul	0.2846	7.35E+00

Probability of Dry Hours

Month	MEXPTRAN	Hourly MCME
Jan	1.81E-04	4.05E-07
Feb	3.39E-03	1.65E-04
Mar	4.59E-04	2.47E-05
Apr	1.00E-03	1.44E-06
May	8.62E-03	3.02E-03
Jun	2.68E-06	5.09E-04
Jul	1.17E-03	6.48E-05

Aug	20.9572	2.8529
Sep	2.6364	12.8155
Oct	14.4105	0.1189
Nov	3.8816	2.9280
Dec	69.5820	36.2565
SSE	15.5920	7.6685
RMSE	**3.9487**	**2.7692**

Aug	5.79E-04	0.0002
Sep	1.20E-04	0.0028
Oct	3.88E-03	0.0002
Nov	1.24E-02	0.0013
Dec	1.64E-04	0.0004
SSE	2.67E-03	0.0007
RMSE	**5.16E-02**	**0.0266**

Table 4.16: The RMSE of the Hourly MCME and MEXPTRAN models at 24-hour scale in the validation period (1991-2000)

24-Hour Mean

Month	MEXPTRAN	Hourly MCME
Jan	0.0492	0.0259
Feb	2.4612	2.5948
Mar	0.3678	1.7067
Apr	0.1500	2.0155
May	2.6218	2.3201
Jun	10.9352	12.1441
Jul	0.0729	0.0313
Aug	3.8253	3.1964
Sep	13.3249	1.8324
Oct	0.0001	0.2392
Nov	3.9482	1.6205
Dec	7.6001	8.6367
SSE	3.7797	3.0303
RMSE	**1.9442**	**1.7408**

24-Hour Autocorrelation

Month	MEXPTRAN	Hourly MCME
Jan	0.00247	0.0067
Feb	0.00550	0.0003
Mar	0.00016	0.0011
Apr	0.00533	0.0061
May	0.00148	0.0005
Jun	0.00315	0.0079
Jul	0.00004	0.0023
Aug	0.00009	0.0003
Sep	0.00408	0.0029
Oct	0.02343	0.0284
Nov	0.00000	0.0002
Dec	0.01229	0.0252
SSE	0.00483	0.0068
RMSE	**0.06953**	**0.0826**

24-Hour Standard Deviation

Month	MEXPTRAN	HourlyMCME
Jan	0.55	2.0274
Feb	5.20	0.0170
Mar	3.64	18.3808
Apr	6.19	11.5406
May	0.25	1.8458
Jun	24.49	49.6250
Jul	0.00	10.5633
Aug	2.650	10.6138
Sep	40.663	0.2263
Oct	0.707	0.0123
Nov	5.584	22.1177
Dec	15.482	35.3403
SSE	8.783	13.5259
RMSE	**2.964**	**3.6778**

24-Hour Maximum

Month	MEXPTRAN	Hourly MCME
Jan	89.8310	331.2400
Feb	204.9284	33.9889
Mar	1213.0987	11.6964
Apr	1.0896	375.9721
May	344.0778	1437.1681
Jun	166.3316	1840.4100
Jul	527.9208	676.0000
Aug	7.9073	299.6361
Sep	1321.4356	89.4916
Oct	114.1783	7.6729
Nov	595.1729	2424.5776
Dec	447.3336	2893.3641
SSE	419.4421	868.4348
RMSE	**20.4803**	**29.4692**

24-Hour Coefficient of Skewness

Month	MEXPTRAN	Hourly MCME
Jan	0.1470	0.5660
Feb	0.2178	1.1179
Mar	0.6230	0.0001

Probability of Dry Days

Month	MEXPTRAN	Hourly MCME
Jan	0.00	0.0049
Feb	0.00	0.0154
Mar	0.00	0.0065

Apr	0.0692	0.0052	Apr	0.00	0.0001
May	0.3266	1.6479	May	0.02	0.0267
Jun	2.7179	0.2169	Jun	0.01	0.0016
Jul	0.1023	0.8843	Jul	0.00	0.0070
Aug	0.2163	0.0456	Aug	0.024	0.0065
Sep	0.1405	0.1547	Sep	0.001	0.0374
Oct	0.0799	0.1504	Oct	0.000	0.0120
Nov	1.1461	2.7492	Nov	0.001	0.0069
Dec	1.1760	0.3511	Dec	0.025	0.0007
SSE	0.5802	0.6574	SSE	0.007	0.0105
RMSE	**0.7617**	**0.8108**	**RMSE**	**0.086**	**0.1024**

Table 4.17: The RMSE of the Daily MCME and MEXPTRAN models at daily scale in the validation period (1991-2000)

24-Hour Mean

Month	MEXPTRAN	Daily MCME
Jan	0.0492	0.3444
Feb	2.4612	6.9634
Mar	0.3678	2.2995
Apr	0.1500	1.5293
May	2.6218	0.6087
Jun	10.9352	9.1436
Jul	0.0729	0.5505
Aug	3.8253	0.9155
Sep	13.3249	0.6572
Oct	0.0001	0.5778
Nov	3.9482	0.3758
Dec	7.6001	5.4281
SSE	3.7797	2.4495
RMSE	**1.9442**	**1.5651**

24-Hour Autocorrelation

Month	MEXPTRAN	Daily MCME
Jan	0.00247	0.0004
Feb	0.00550	0.0022
Mar	0.00016	0.0004
Apr	0.00533	0.0058
May	0.00148	0.0010
Jun	0.00315	0.0045
Jul	0.00004	0.0003
Aug	0.00009	0.0000
Sep	0.00408	0.0028
Oct	0.02343	0.0225
Nov	0.00000	0.0003
Dec	0.01229	0.0177
SSE	0.00483	0.0048
RMSE	**0.06953**	**0.0695**

24-Hour Standard Deviation

Month	MEXPTRAN	Daily MCME
Jan	0.55	0.6192
Feb	5.20	10.6289
Mar	3.64	10.3509
Apr	6.19	12.1602
May	0.25	0.9436
Jun	24.49	20.5073
Jul	0.00	1.1452
Aug	2.650	1.3874
Sep	40.663	0.7298
Oct	0.707	1.3482
Nov	5.584	1.9125
Dec	15.482	4.4554
SSE	8.783	5.5157
RMSE	**2.964**	**2.3486**

24-Hour Maximum

Month	MEXPTRAN	Daily MCME
Jan	89.8310	51.8400
Feb	204.9284	424.3600
Mar	1213.0987	142.3249
Apr	1.0896	342.2500
May	344.0778	166.4100
Jun	166.3316	170.3025
Jul	527.9208	31.6969
Aug	7.9073	25.5025
Sep	1321.4356	0.0900
Oct	114.1783	10.3041
Nov	595.1729	470.8900
Dec	447.3336	123.2100
SSE	419.4421	163.2651
RMSE	**20.4803**	**12.7775**

24-Hour Coefficient of Skewness

Probability of Dry Days

Month	MEXPTRAN	Daily MCME
Jan	0.1470	0.0711
Feb	0.2178	0.7333
Mar	0.6230	0.2049
Apr	0.0692	0.0241
May	0.3266	0.2150
Jun	2.7179	1.5146
Jul	0.1023	0.0028
Aug	0.2163	0.0001
Sep	0.1405	0.0024
Oct	0.0799	0.0985
Nov	1.1461	0.9045
Dec	1.1760	0.4128
SSE	0.5802	0.3487
RMSE	**0.7617**	**0.5905**

Month	MEXPTRAN	Daily MCME
Jan	0.00	0.0071
Feb	0.00	0.0064
Mar	0.00	0.0010
Apr	0.00	0.0016
May	0.02	0.0058
Jun	0.01	0.0052
Jul	0.00	0.0004
Aug	0.024	0.0097
Sep	0.001	0.0064
Oct	0.000	0.0019
Nov	0.001	0.0002
Dec	0.025	0.0100
SSE	0.007	0.0046
RMSE	**0.086**	**0.0682**

Table 4.18: The summary of the RMSE for the NSRP and MCME models in the Validation Period (1991-2000)

Property	**1-hour Mean**	**1-hour Standard Devaition**	**1-hour Coeff. of Skewness**	**1-hour Maximum**	**1-hour Autocorr.**	**Prob. Dry Hours**
Hourly MCME	**0.07326**	**0.26799**	**2.76921**	**10.43794**	0.36440	**0.02663**
MEXPTRAN	0.08427	0.36430	3.94867	16.51080	**0.19215**	0.05163

Property	**Daily Mean**	**Daily Std.Dev**	**Daily Coeff. of Skewness**	**Daily Max.**	**Daily Autocorr.**	**Prob. Dry day**
Hourly MCME	1.7408	3.6778	0.8108	29.4692	0.0826	0.1024
Daily MCME	1.5651	2.3486	0.5905	12.7775	0.0695	0.0682
NSRP	1.9442	2.9636	0.7617	20.4803	0.0695	0.0857

4.3.8 Numerical Comparison between the MCME and the MEXPTRAN models

Qualitatively, the performance of both models were not as good as in the calibration period. Nevertheless, the comparisons between both models were also done numerically using RMSE. Table 4.15 and 4.16 show the RMSE evaluated from the monthly square errors between the observed and the medians of the simulated properties for the MEXPTRAN and the hourly MCME model. Table 4.17 shows the RMSE evaluated between the daily MCME and the MEXPTRAN for the daily rainfall properties and Table 4.18 gives the summary of the RMSE for all properties considered. For the one-hour scale, the hourly MCME has smaller RMSE values than the MEXPTRAN in all properties considered, except for the one-hour autocorrelation of rainfalls. When the hourly series were lumped to daily series, the MEXPTRAN model could provide smaller RMSE values than the hourly MCME model for most of the daily rainfall properties considered in the study. The daily MCME model can provide even smaller RMSE values than the MEXPTRAN for all the properties considered for the daily scale.

From the numerical analysis results, it can be concluded that both models have the same predictive ability. The predictive ability of the MCME hourly model was found to be better than the NSRP in predicting the hourly rainfall process. When the hourly series were lumped to daily series, the NSRP model performed better than the hourly MCME model in predicting the daily rainfall process. However, the predictive ability of daily MCME model was even better than the NSRP in predicting the daily rainfall process. While both models did not perform as well as in the calibration period, both were able to preserve the seasonal trends of the observed rainfall properties.

4.3.9 Summary

In assessing the descriptive ability of the model using the hourly and daily observed series from year 1981-1990, the performance of both models discussed in Chapter 4 and 5 was compared using the qualitative and numerical analysis. From the RMSE values obtained for all properties considered, it was found that the NSRP (MEXPTRAN) model has the ability to describe the properties of the observed at various timescales, especially at one-hour and daily scales, even though the model only generate hourly rainfall series. The MCME model was found to describe better the properties of the observed when their parameters were estimated using data at the same scale as the observed. But when the series were lumped to other scales, the performance fails to maintain. Therefore, the hourly and daily MCME models do preserve well the observed properties at the respective scales.

In assessing the predictive ability of both models, the hourly and daily series from the same station from year 1991-2000 was used in the validation process. In general, both NSRP and MCME models were found to have the same predictive ability. While both models did not perform as well as in the calibration period, both were able to preserve the seasonal trends of the observed rainfall properties. However, the predictive ability of the daily MCME model was found to be better than the predictive ability of the NSRP and hourly MCME model in predicting the daily rainfall process.

4.4 Short-term Forecast of Rainfall in Lembah Klang

The data analyzed were an hourly rainfall intensity data. The technique employed was a short-term forecasting technique where the prediction was only for a one-hour ahead. According to Burlando (1996), a forecast lead time of a couple of hours, which was close to the response time of the drainage system to surface runoff, could be useful in view of an efficient control of pumping stations and hydraulic control of gates that may prevent flash flood. This can also reduce overflow volumes of water in tanks and channels of the sewer system, and prevent the water from any damages and pollutions.

4.4.2 Stations Selection Criterion

In the current study, the stations were selected based on the analysis of station-to-station correlations, performed on the available rain gage data. As suggested in the literature, the correlation criterion for the correct pairing of stations can be used if long historical records of data are available. This is to make sure that the stations are truly correlated. In other words, they are not correlated in only a short period of time. Although this method of selecting pairing stations is not as good as the one that is based on the storm movements, this is the only way to select stations for the current study due to of lack of information on Malaysia storm movements.

The correlation coefficient $\rho_{X,Y}$ between two random variables X and Y with expected values μ_X and μ_Y, and standard deviations σ_X and σ_Y can be written as

$$\begin{aligned}\rho_{XY} &= \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{E[(X-\mu_X)^2]E[(Y-\mu_Y)^2]}} \\ &= \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}\end{aligned} \tag{4.1}$$

If we have a series of n measurements of X and Y written as x_t and y_t where $t = 1, 2, \ldots, n$, then the Pearson product-moment correlation coefficient, denoted as r_{XY}, can be used to estimate the correlation of X and Y. The formula is

$$\begin{aligned}r_{XY} &= \frac{s_{XY}}{\sqrt{s_X}\sqrt{s_Y}} \\ &= \frac{\sum_{t=1}^{n}(x_t-\bar{x})(y_t-\bar{y})}{\sqrt{\sum_{t=1}^{n}(x_t-\bar{x})^2}\sqrt{\sum_{t=1}^{n}(y_t-\bar{y})^2}}\end{aligned} \tag{4.2}$$

where

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

$$\bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t$$

$$s_X = \frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2$$

$$s_Y = \frac{1}{n}\sum_{t=1}^{n}(y_t-\bar{y})^2$$

$$s_{XY} = \frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})$$

are the sample means, samples variances and sample covariances.

The correlation is defined only if both of the standard deviations are finite and both of them are nonzero. It is a corollary of the Cauchy-Schwarz inequality that the correlation does not exceed one in absolute value. The correlation is one in the case of an increasing linear relationship, negative one in the case of a decreasing linear relationship, and some values in between in all other cases, indicating the degree of linear dependence between the variables. The closer the coefficient is to either negative one or positive one, the stronger the correlation between the variables. If the variables are independent then the correlation is zero, but the converse is not true because the correlation coefficient detects only linear dependencies between two variables.

The results from the analysis of station-to-station correlations for all the stations are shown in Table 4.19. From this table, the two stations that were highly correlated compared to other station combinations were station Empangan Genting Kelang with station Km.11 Gombak, followed by station Empangan Genting Kelang with station Kampung Kuala Saleh. The correlation values were calculated using hourly rainfall data from 1st April 2002 till 29th April 2002 as recorded in each station by the rain gages. This data have been taken because it was during the intermonsoon season where the convective rains always occurred during this monsoon seasons. It is also because there was no missing data during this period.

One reason why the stations were highly correlated was because the distances between the stations were near. Another reason was the storm movements. This can be concluded from the sample radar maps for the storms on 6th April 2006 and 10th May 2006 where both were during the inter-monsoon season.

Table 4.19: Analysis of station-to-station correlation for all the stations listed.

Station Number	3015001	3116003	3116006	3216001	3216004	3217001	3217002	3217003	3217004	3317001	3317004
3015001	1	0.165326	0.233653	-0.01111	0.060295	0.163399	0.158303	0.102992	0.111002	0.021069	0.031556
3116003	0.165326	1	0.372016	0.016552	0.327303	0.248401	0.093928	0.251626	0.11503	0.252736	0.085266
3116006	0.233653	0.372016	1	-0.00803	0.461648	0.192583	0.185403	0.181281	0.094348	0.269431	0.068528
3216001	-0.01111	0.016552	-0.00803	1	0.028524	0.226348	0.022059	0.414294	-0.00926	0.110088	0.01727
3216004	0.060295	0.327303	0.461648	0.028524	1	0.148068	0.044899	0.148925	0.023438	0.205013	0.041727
3217001	0.163399	0.248401	0.192583	0.226348	0.148068	1	0.186133	0.328398	0.073964	0.401712	0.169044
3217002	0.158303	0.093928	0.185403	0.022059	0.044899	0.186133	1	**0.653569**	**0.526572**	0.117254	0.215985
3217003	0.102992	0.251626	0.181281	0.414294	0.148925	0.328398	**0.653569**	1	0.297141	0.284231	0.154582
3217004	0.111002	0.11503	0.094348	-0.00926	0.023438	0.073964	**0.526572**	0.297141	1	0.069179	0.15312
3317001	0.021069	0.252736	0.269431	0.110088	0.205013	0.401712	0.117254	0.284231	0.069179	1	0.197104
3317004	0.031556	0.085266	0.068528	0.01727	0.041727	0.169044	0.215985	0.154582	0.15312	0.197104	1

For the reasons stated earlier, the MARIMA model was then employed using rainfalls data from rain gages taken from two pairing study areas. The first pairing study area was station Empangan Genting Kelang with station Km.11 Gombak and the second pairing study area was station Empangan Genting Kelang with station Kampung Kuala Saleh.

4.4.2 Data Modeling

The process starts with model definition and identification, followed by the process of parameter estimation. The MARIMA model obtained will then be used to forecast future values for the rainfalls intensity.

4.4.2.2 Data Analysis

Real-time prediction of rainfall by means of stochastic models can be viewed as a questionable point which is due to the limited persistence of the rainfall intensity as observed at usual temporal aggregation scales, for example 1 hour. This can be looked as an effect of the intrinsic unpredictability of rainfall, which can be argued based on the small decorrelation time that exhibits rainfall when it is observed at a point in space (Zawadzki, 1987). The dynamics of the rainfall process can explain this effect by looking at the evolution of the rainfall process at a point in space as a result of two intertwined mechanisms.

The first mechanism concerns the intrinsic evolution of the storm as observed from a coordinate system connected to the storm movement. The persistence in this system can be described by the Lagrangian space-time correlation structure of the process (Burlando, 1996). The second one concerns the storm movement, that originates the storm modification normally observed at a fixed point as a result of the continuous shifting of the rainfall field in the spatial domain. The combination of these two mechanisms leads to a rapid decay of the Eulerian time correlation at a point in space, which is generally smaller than the Lagrangian space-time correlation. This property comes out from the analysis of actual data (Bacchi and Borga, 1994), as well as of rainfall fields simulated by space-time models (Waymire et al., 1984).

It is thus expected that a stochastic model of the autoregressive type would be more successful if based on data recorded by a rain gage hypothetically moving jointly with the storm, that is based on the Lagrangian cross-correlation structure detected by radar measurements. However, radar maps do not provide reliable quantitative estimations of rainfall intensity, which is better estimated based on rain gage measurements. Rainfall data based only on the use of radar maps could therefore be misleading in the estimation of the effective depth. On the other hand, forecasting models based only on the Eulerian cross-correlation analysis of rain data are affected by a weaker persistence effect than the one that could be observed from a reference system linked to the storm. Forecasting models based on the Eulerian cross-correlation would thus benefit of poorer information, thus resulting in poorer performance (Burlando, 1996).

Accordingly, a successful forecasting model should combine rain gage data and radar maps in order to reduce the limitations that affect both these types of the measurements. In this view, MARIMA models represent an interesting tool, because they allow to forecast rainfall intensity at a point in space, that is the rain gage station, as a function of current and past rainfall occurrences observed at several points in the basin, including the point itself. They account thus for both the Eulerian and the Lagrangian correlations of the process (Burlando, 1996).

Setting up this type of model to forecast rainfall at a rain gage station would therefore require selecting those stations where current and past rainfall occurrences show the highest level of cross-correlation with the ones observed at the forecasting site. Such a selection can be better afforded on the basis of the kinematic behaviour of the storm that can be detected from radar maps. When the storm speed has been estimated, the time lag for the evaluation of the cross-correlation between rainfall records observed at different sites along the storm trajectory can be selected equal or close to the time the storm takes to travel from those sites to the forecasting one (Burlando, 1996).

However, as mentioned earlier, because of lack of technologies in Malaysian Meteorological Department, not enough information for the storm movement could be obtained. Hence, the stations were selected based on the analysis of correlation between two stations.

MARIMA models allow the computations of future occurrences of a time series as a linear combination of

(i) past occurrences of the time series itself and of time series which are cross-correlated to it; that is the autoregressive component
(ii) the present and past occurrences of a random white noise component; that is the moving average component.

The general form of a MARIMA model with p autoregressive terms and q moving average terms can be written as

$$\begin{aligned} \boldsymbol{\alpha}(B)\boldsymbol{Y}_t &= \boldsymbol{\beta}(B)\boldsymbol{\varepsilon}_t \\ \boldsymbol{Y}_t &= \sum_{i=1}^{p} \boldsymbol{\alpha}_i \boldsymbol{Y}_{t-i} + \sum_{j=0}^{q} \boldsymbol{\beta}_j \boldsymbol{\varepsilon}_{t-j} \end{aligned} \tag{4.3}$$

where $\boldsymbol{Y}_t = (\boldsymbol{I} - \boldsymbol{B})^d \boldsymbol{X}_t$.

$\boldsymbol{X}_t$ is the stochastic process under study, where in this case is the rainfall intensity. $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{B}$ is the backward shift operator, and d is the differencing order of the model. The vector $\boldsymbol{\varepsilon}_t$ consists of N uncorrelated shocks (white noise) of zero mean and unit variance, and $\boldsymbol{\varepsilon}_t$ being uncorrelated with $\boldsymbol{Y}_\tau$, for $\tau < t$. Both $\boldsymbol{Y}_t$ and $\boldsymbol{X}_t$ are N-dimension column matrices where N is the number of series considered in the time series problem. Both should have zero mean, although $\boldsymbol{X}_t$ is allowed to have non-zero mean if $d > 0$ (Box and Jenkins, 1976). $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the $N \times N$ autoregressive and moving average parameters matrices of the model.

4.4.2.2 Model Identification

A change of the values of p and q allows to formulate models of different orders, each one characterized by different correlation structure and number of parameters. The model defined by (4.3) is characterized by a number of parameters which is larger with increasing orders p and q. This can be regarded as a major limitation with respect to analytical tractability, and to parameter estimation in those cases where a limited number of actual observations are available for being used in the estimation process. This is just the case of an event based parameter estimation, which is generally recognized to lead to better performance of the forecasting model as compared to the ones obtained from the model estimated using the raw historical continuous precipitation data set (Burlando et al., 1993). Accordingly, the values of the orders p and q, as well as the number of series, N, which are considered by the model, should be selected as a compromise result between the conflicting needs of descriptiveness and of mathematical tractability.

Consider the autoregressive first order model, MARIMA (1,1,0) which can be written as $p = 1, d = 1, q = 0$, applied to only two time series requiring the estimation of 4 parameters. An event-based estimation of the model would therefore require a minimum number of current observations, being necessary to increase this minimum as the number of parameters increases. A high number of parameters would therefore limit the benefit from the use of the forecasting model only to long lasting events. Moreover, the time required for the estimation and forecasting procedure should be negligible with respect to the lead time of the forecast, which is generally constrained by the flood forecasting and warning systems. Thus, the need for an operational tool, which can be suitable for practical purposes, suggested to limit this study to the first order autoregressive model, MARIMA(1,1,0), as applied to a two-sites time series, with the purpose of forecasting rainfall in one of them.

4.4.2.4 Parameter Estimation

As mentioned earlier, an event-based estimation approach was carried out in this study. According to this approach, each storm event regardless of the month or season is considered separately for parameter estimation. A different parameter set is therefore determined for each storm event considered. Moreover, the data used for estimation are only those which become available as the storm event evolves in time. For this reason the model can be run only when the number of current event observations is sufficient to allow the effective estimation of the parameters. As a new observation becomes available from the monitoring system, the estimation procedure is repeated and the updated parameter set is used by the model to issue a new rainfall forecast for the designed lead time. It is thus expected that forecasts become much more reliable as the event evolves in time.

The estimation procedure can be set-up following two traditional procedures, that is the method of moments and the least squares method. In view of the operational purposes that motivate this study, the method of moments has been preferred, especially because it performs more rapidly than the latter one, so that a possible use within a forecasting system could benefit by saving computing time.

For the autoregressive first order model, MARIMA(1,1,0),

$$\boldsymbol{Y}_t = \boldsymbol{\alpha}\boldsymbol{Y}_{t-1} + \boldsymbol{\varepsilon}_t \qquad (4.4)$$

where

$$\begin{aligned} \boldsymbol{Y}_t &= (\boldsymbol{I} - \boldsymbol{B})\boldsymbol{X}_t \\ &= \boldsymbol{I}\boldsymbol{X}_t - \boldsymbol{B}\boldsymbol{X}_t \\ &= \boldsymbol{X}_t - \boldsymbol{X}_{t-1} \end{aligned}$$

Therefore,

$$\begin{aligned} \boldsymbol{X}_t - \boldsymbol{X}_{t-1} &= \boldsymbol{\alpha}(\boldsymbol{X}_{t-1} - \boldsymbol{X}_{t-2}) + \boldsymbol{\varepsilon}_t \\ \boldsymbol{X}_t &= (\boldsymbol{I} + \boldsymbol{\alpha})\boldsymbol{X}_{t-1} - \boldsymbol{\alpha}\boldsymbol{X}_{t-2} + \boldsymbol{\varepsilon}_t \end{aligned} \tag{4.5}$$

where $\boldsymbol{\varepsilon}_t$ is assumed to be a white noise, and stationarity is assumed to hold, the parameter estimation has been performed by the method of moments. This consists of solving the system

$$\boldsymbol{\alpha} = \boldsymbol{M}_1 \boldsymbol{M}_0^{-1} \tag{4.6}$$

where $\boldsymbol{M}_0$ and $\boldsymbol{M}_1$ denote respectively the covariance and the lag 1 covariance matrices.

Estimating the parameters for every one hour prediction for each station repeatedly is a tedious task. To simplify the task, a computer program has been written using the Microsoft Visual C++. By running the program, the new parameter values will be estimated each time before a forecast is made.

4.4.3 Performance Measure

For the purposes of measuring the forecast performance, two performance measures will be used that is the average values of the residuals and the root mean square error (RMSE). Both performance measures are the ways to quantify the amount by which an estimator differs from the true value of the quantity being estimated.

This performance measure also can be used to compare the models that been used for an estimation or forecast process. By comparing the value of both performance measure for each model, we can determine the best model in terms of the error of the estimator, where the best model have the lowest performance measure values.

The average values of the residuals can be denoted as μ_{ε}, where

$$\mu_{\varepsilon} = \frac{\sum_{i=1}^{n} \left| \hat{\theta}_i - \theta_i \right|}{n}$$

while the root mean square error (RMSE) can be written as

$$\text{RMSE}(\varepsilon) = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{\theta}_i - \theta_i \right)^2}{n}}$$

where

$\hat{\theta}_i = i^{\text{th}}$ estimated value

$\theta_i = i^{\text{th}}$ observed value

n = sample size

4.5 Prediction of Rainfalls Using the MARIMA Model

After estimating the parameters, the forecast value can be calculated using the MARIMA (1,1,0) model. However this is only an hour ahead forecasts for each station. Before the second hour forecast can be made, the parameters need to be estimated again. This technique was repeated each time before a forecast was produced to ensure that the model was used correctly. These tedious works have been simplified by writing a computer program using the Microsoft Visual C++.

By using the program, the forecast values were automatically calculated by the model. There was a problem encountered where some of the forecast values obtained were negative (less than zero). Since rainfalls intensity is never less than zero, it is considers that there is no rain for that hour.

The scatter plots for the past rainfalls intensity data for each station and the forecast values cannot be plotted using this program because there were more than 1000 data. Therefore, to solve this problem, the Microsoft Office Excel was used to do the scatter plots.

4.5.1 Study Area 1

First, we will forecast the rainfalls intensities for station Empangan Genting Kelang with station Km.11 Gombak. The lead time of the forecast has been assumed to be equal to one hour. Results for these stations are shown in Figures 4.46, 4.47, 4.48 and 4.49. These figures show the hyetographs of observed rainfall intensity and corresponding forecasts of one-hour ahead, the observed and forecasted cumulative rainfall intensities.

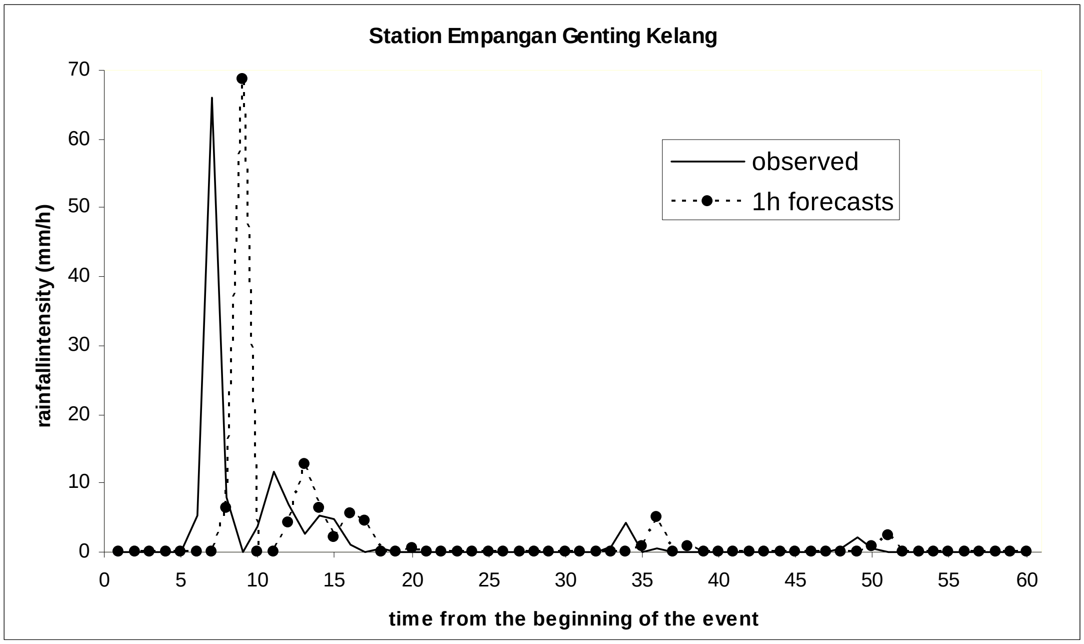


Figure 4.46: The hyetographs of observed rainfall intensity and MARIMA one-hour ahead forecast for station Empangan Genting Kelang.

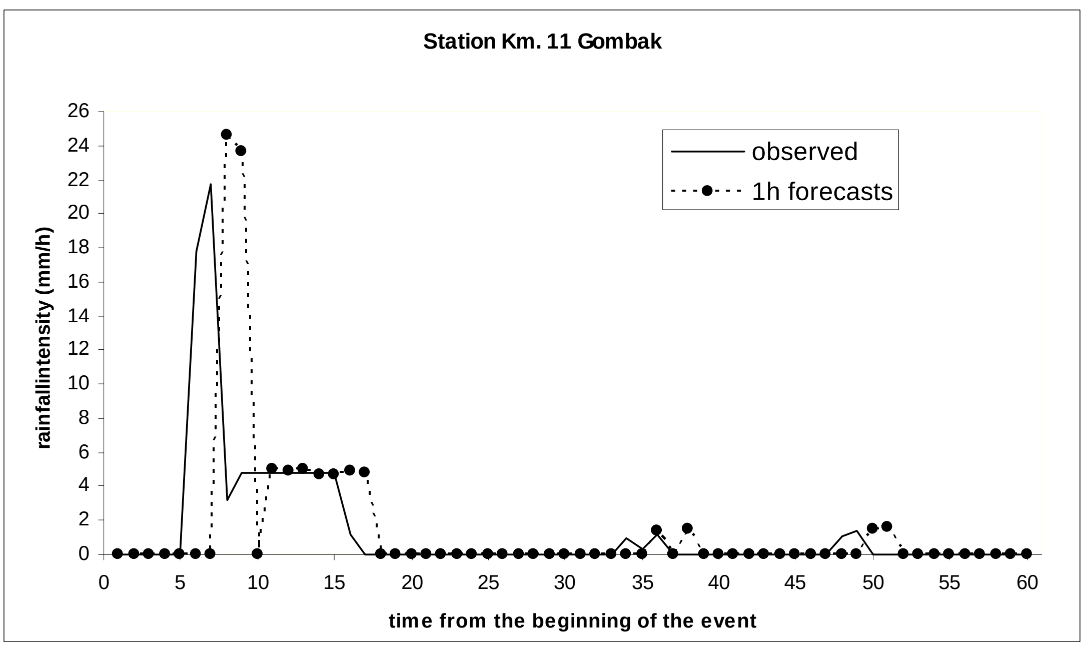


Figure 4.47: The hyetographs of observed rainfall intensity and MARIMA one-hour ahead forecast for station Km.11 Gombak.

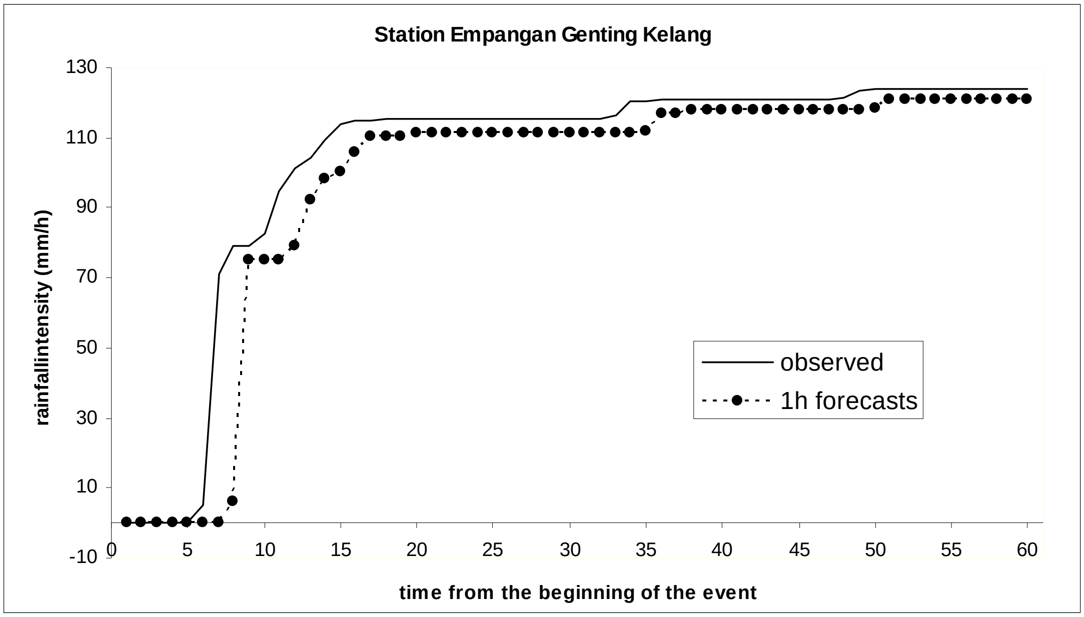


Figure 4.48: The observed and MARIMA one-hour ahead forecast cumulative rainfall intensity for station Empangan Genting Kelang.

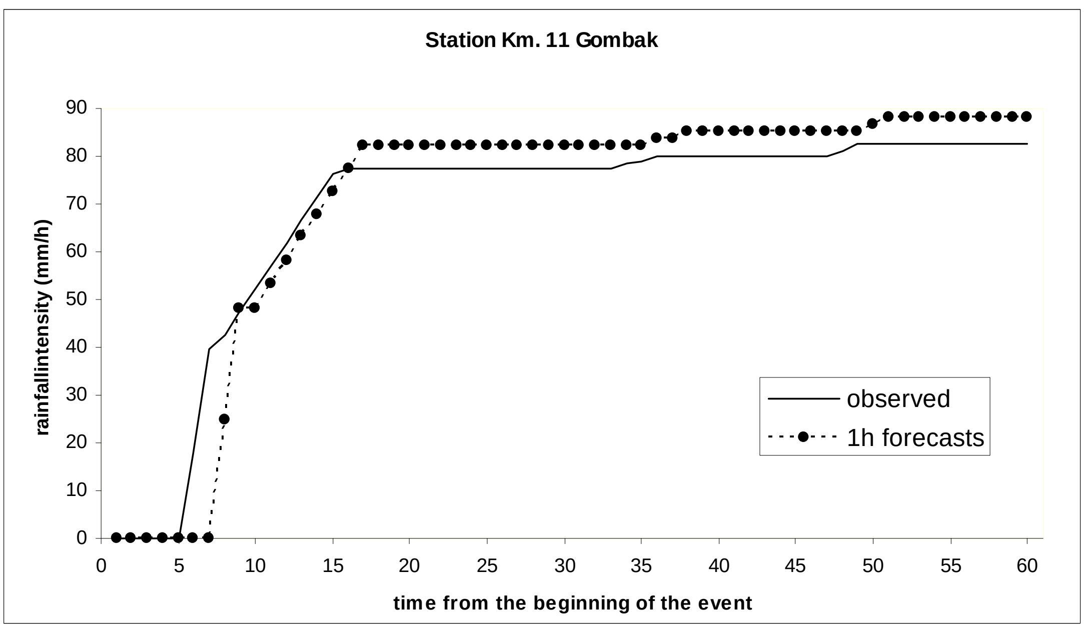


Figure 4.49: The observed and MARIMA one-hour ahead forecast cumulative rainfall intensity for station Km.11 Gombak.

From Figures 4.46 and 4.47, we can see that the forecasted values for both stations are most likely the same as the observed values especially for the zero value observed data. However, if we see the pattern of the forecasted values, those values were much influenced by the values of its last two hours. The cumulative forecast values for both station shown in Figures 4.48 and 4.49 shows that it is most likely the same as the cumulative observed values.

Table 4.20 shows the numerical results for this forecast. In this table, "E" represents station Empangan Genting Kelang and "G" stands for station Km.11 Gombak where it includes the observed values (obs.), forecasted values (pre.), the error between the observed values and the forecasted values (error), the squared error between the observed values and the forecasted values (error^2), the cumulative observed values (cum) and the cumulative forecasted values (pre.cum) for both stations. The estimated parameters for these study area will be shown in the appendices.

Table 4.20: Results for MARIMA model forecast of rainfalls intensity for station Empangan Genting Kelang and station Km.11 Gombak

days	E(obs)	E(pre)	G(obs)	G(pre)	E(error)	E(error^2)	G(error)	G(error^2)	E(cum)	E(pre.cum)	G(cum)	G(pre.cum)
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
29	5.2000	0.0000	17.8000	0.0000	-5.2000	27.0400	-17.800	316.8400	5.2000	0.0000	17.8000	0.0000
29	66.0000	0.0000	21.7000	0.0000	-66.000	4356.0000	-21.700	470.8900	71.2000	0.0000	39.5000	0.0000
29	7.9000	6.2725	3.2000	24.6598	-1.6275	2.6488	21.4598	460.5230	79.1000	6.2725	42.7000	24.6598
29	0.0000	68.7757	4.8000	23.6321	68.7757	4730.0969	18.8321	354.6480	79.1000	75.0482	47.5000	48.2919
29	3.8000	0.0000	4.8000	0.0000	-3.8000	14.4400	-4.8000	23.0400	82.9000	75.0482	52.3000	48.2919
29	11.6000	0.0000	4.8000	4.9856	-11.600	134.5600	0.1856	0.0344	94.5000	75.0482	57.1000	53.2775
29	7.0000	4.3131	4.8000	4.9137	-2.6869	7.2194	0.1137	0.0129	101.5000	79.3613	61.9000	58.1912
29	2.7000	12.6704	4.8000	5.0290	9.9704	99.4089	0.2290	0.0524	104.2000	92.0317	66.7000	63.2202
29	5.2000	6.3066	4.8000	4.6485	1.1066	1.2246	-0.1515	0.0230	109.4000	98.3383	71.5000	67.8687
29	4.7000	2.0401	4.8000	4.6420	-2.6599	7.0751	-0.1580	0.0250	114.1000	100.3784	76.3000	72.5107
29	1.0000	5.5846	1.2000	4.8904	4.5846	21.0186	3.6904	13.6191	115.1000	105.9630	77.5000	77.4011
29	0.0000	4.6217	0.0000	4.7813	4.6217	21.3601	4.7813	22.8608	115.1000	110.5847	77.5000	82.1824
30	0.5000	0.0000	0.0000	0.0000	-0.5000	0.2500	0.0000	0.0000	115.6000	110.5847	77.5000	82.1824
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	110.5847	77.5000	82.1824
30	0.0000	0.5875	0.0000	0.0188	0.5875	0.3452	0.0188	0.0004	115.6000	111.1722	77.5000	82.2012
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012

days	E(obs)	E(pre)	G(obs)	G(pre)	E(error)	E(error^2)	G(error)	G(error^2)	E(cum)	E(pre.cum)	G(cum)	G(pre.cum)
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	111.1722	77.5000	82.2012
30	0.8000	0.0000	0.0000	0.0000	-0.8000	0.6400	0.0000	0.0000	116.4000	111.1722	77.5000	82.2012
30	4.2000	0.0000	1.0000	0.0000	-4.2000	17.6400	-1.0000	1.0000	120.6000	111.1722	78.5000	82.2012
30	0.0000	0.9257	0.3000	0.0302	0.9257	0.8569	-0.2698	0.0728	120.6000	112.0979	78.8000	82.2314
30	0.5000	4.9288	1.2000	1.4296	4.4288	19.6143	0.2296	0.0527	121.1000	117.0267	80.0000	83.6610
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	117.0267	80.0000	83.6610
30	0.0000	0.7529	0.0000	1.4901	0.7529	0.5669	1.4901	2.2204	121.1000	117.7796	80.0000	85.1511
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	117.7796	80.0000	85.1511
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	117.7796	80.0000	85.1511
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	117.7796	80.0000	85.1511
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	117.7796	80.0000	85.1511
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	117.7796	80.0000	85.1511
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	117.7796	80.0000	85.1511
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	117.7796	80.0000	85.1511
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	117.7796	80.0000	85.1511
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	117.7796	80.0000	85.1511
1	0.5000	0.0000	1.1000	0.0000	-0.5000	0.2500	-1.1000	1.2100	121.6000	117.7796	81.1000	85.1511
1	2.0000	0.0000	1.4000	0.0000	-2.0000	4.0000	-1.4000	1.9600	123.6000	117.7796	82.5000	85.1511
1	0.5000	0.7916	0.0000	1.4504	0.2916	0.0850	1.4504	2.1037	124.1000	118.5712	82.5000	86.6015
1	0.0000	2.2938	0.0000	1.5466	2.2938	5.2615	1.5466	2.3920	124.1000	120.8650	82.5000	88.1481
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	120.8650	82.5000	88.1481
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	120.8650	82.5000	88.1481
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	120.8650	82.5000	88.1481
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	120.8650	82.5000	88.1481
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	120.8650	82.5000	88.1481
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	120.8650	82.5000	88.1481
days	**E(obs)**	**E(pre)**	**G(obs)**	**G(pre)**	**E(error)**	**E(error^2)**	**G(error)**	**G(error^2)**	**E(cum)**	**E(pre.cum)**	**G(cum)**	**G(pre.cum)**
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	120.8650	82.5000	88.1481

1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	120.8650	82.5000	88.1481
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	120.8650	82.5000	88.1481

E(obs) = Observed value for station Empangan Genting Kelang

E(pre) = Forecasted value for station Empangan Genting Kelang

G(obs) = Observed value for station Km.11 Gombak

G(pre) = Forecasted value for station Km.11 Gombak

E(error) = E(obs) - E(pre)

G(error) = G(obs) - G(pre)

E($error^2$) = $(E(obs) - E(pre))^2$

G($error^2$)= $(G(obs) - G(pre))^2$

E(cum) = Cumulative observed value for station Empangan Genting Kelang

E(pre.cum) = Cumulative forecasted value for station Empangan Genting Kelang

G(cum) = Cumulative observed value for station Km.11 Gombak

G(pre.cum) = Cumulative forecasted value for station Km.11 Gombak

4.5.2 Study Area 2

To ensure that this model can fit to other study area too, we then applied the MARIMA(1,1,0) to predict the rainfalls for station Empangan Genting Kelang with station Kampung Kuala Saleh. Results for these stations are shown in Figures 4.50 4.51, 4.52 and 4.53.

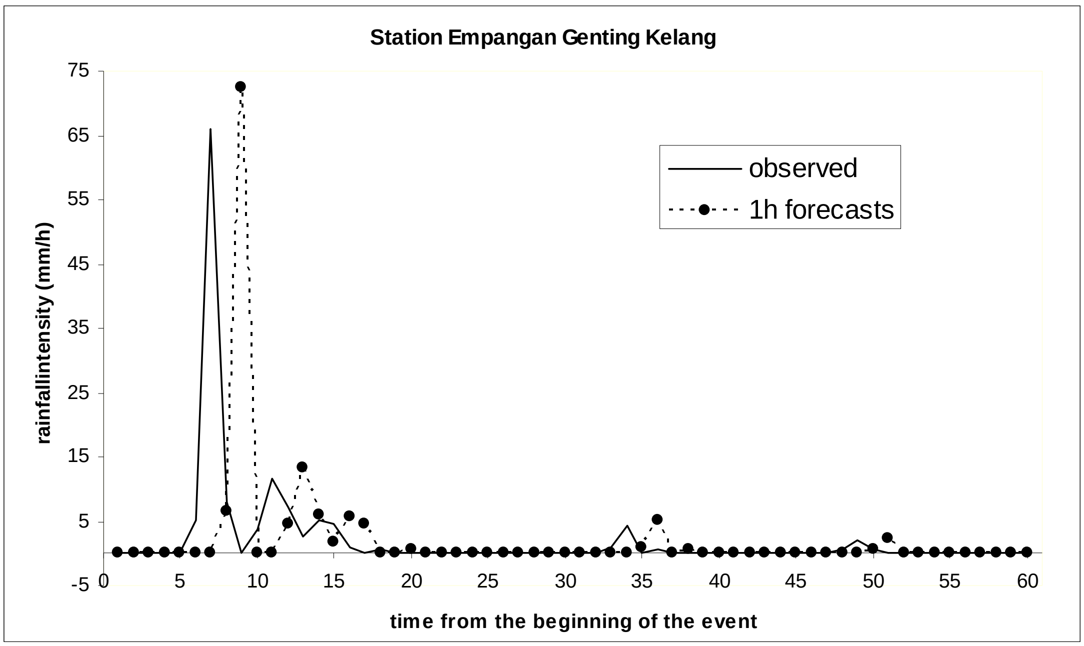


Figure 4.50: The hyetographs of observed rainfall intensity and MARIMA one-hour ahead forecast for station Empangan Genting Kelang.

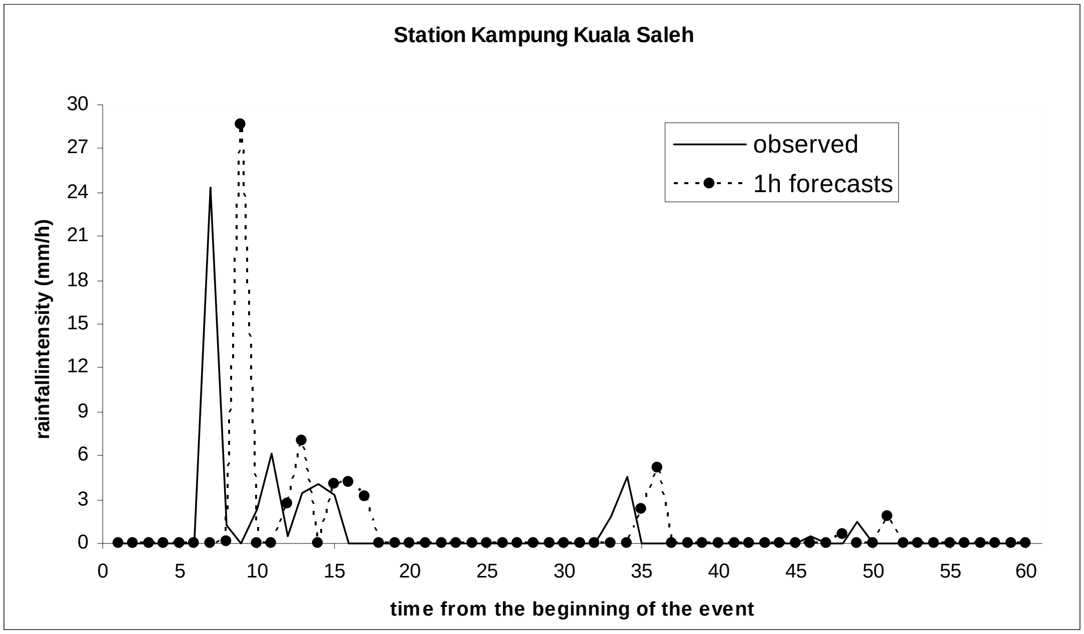


Figure 4.51: The hyetographs of observed rainfall intensity and MARIMA one-hour ahead forecast for station Kampung Kuala Saleh.

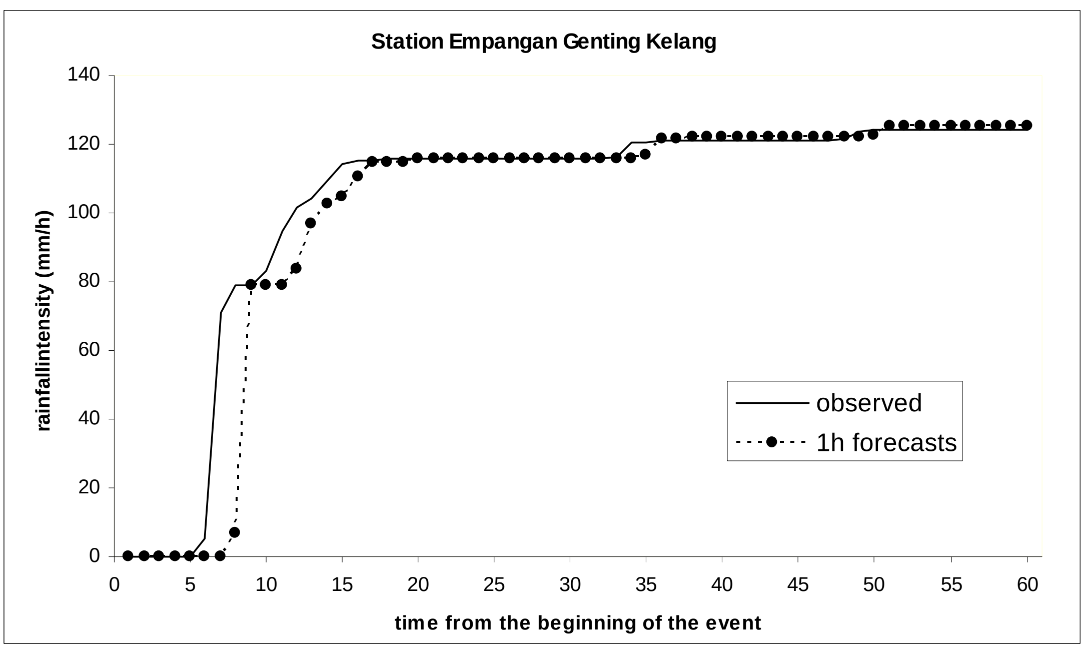


Figure 4.52: The observed and MARIMA one-hour ahead forecast cumulative rainfall intensity for station Empangan Genting Kelang.

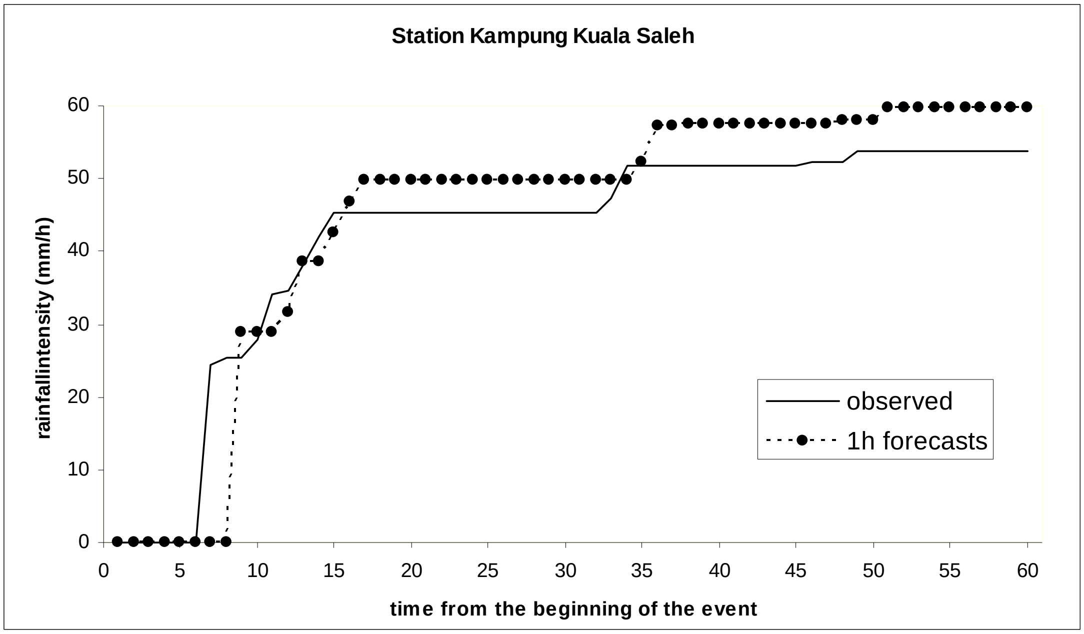


Figure 4.53: The observed and MARIMA one-hour ahead forecast cumulative rainfall intensity for station Kampung Kuala Saleh.

Figure 4.50 show that the forecasted values were more the same as the observed values however some of the forecasted values is much the same as its past two hour observed values. The same results also go for Figure 4.51. However, Figure 4.52 shows that the cumulative forecasted values for station Empangan Genting Kelang that were jointly modeled with station Kampung Kuala Saleh were more likely the same as the cumulative observed values compared to the cumulative forecasted values for station Empangan Genting Kelang that were jointly modeled with station Km.11 Gombak shown in Figure 4.48. Figure 4.53 shows that the cumulative forecasted values were slightly differed from the cumulative observed values.

For these study area, the numerical results are shown in Table 4.21. In this table, “E” represents station Empangan Genting Kelang and “K” stands for station Kampung Kuala Saleh where it includes the observed values (obs), forecasted values (pre), the error between the observed values and the forecasted values (error), the squared error between the observed values and the forecasted values (error^2), the cumulative observed values (cum) and the cumulative forecasted values (pre.cum) for both stations. The estimated parameters for these study area will be shown in the appendices.

Table 4.21: Results for MARIMA model forecast of rainfalls intensity for station Empangan Genting Kelang and station Kampung Kuala Saleh

days	E(obs)	E(pre)	K(obs)	K(pre)	E(error)	E(error^2)	K(error)	K(error^2)	E(cum)	E(pre.cum)	K(cum)	K(pre.cum)
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
29	5.2000	0.0000	0.0000	0.0000	-5.2000	27.0400	0.0000	0.0000	5.2000	0.0000	0.0000	0.0000
29	66.0000	0.0000	24.3000	0.0000	-66.000	4356.0000	-24.300	590.4900	71.2000	0.0000	24.3000	0.0000
29	7.9000	6.6008	1.2000	0.0990	-1.2992	1.6879	-1.1010	1.2122	79.1000	6.6008	25.5000	0.0990
29	0.0000	72.3596	0.0000	28.6702	72.3596	5235.9117	28.6702	821.9804	79.1000	78.9604	25.5000	28.7692
29	3.8000	0.0000	2.3000	0.0000	-3.8000	14.4400	-2.3000	5.2900	82.9000	78.9604	27.8000	28.7692
29	11.6000	0.0000	6.2000	0.0000	-11.600	134.5600	-6.2000	38.4400	94.5000	78.9604	34.0000	28.7692
29	7.0000	4.6056	0.5000	2.7657	-2.3944	5.7332	2.2657	5.1334	101.5000	83.5660	34.5000	31.5349
29	2.7000	13.2479	3.5000	7.0210	10.5479	111.2582	3.5210	12.3974	104.2000	96.8139	38.0000	38.5559
29	5.2000	5.8543	4.0000	0.0000	0.6543	0.4281	-4.0000	16.0000	109.4000	102.6682	42.0000	38.5559
29	4.7000	1.8842	3.3000	4.0555	-2.8158	7.9287	0.7555	0.5708	114.1000	104.5524	45.3000	42.6114
29	1.0000	5.7478	0.0000	4.1267	4.7478	22.5416	4.1267	17.0297	115.1000	110.3002	45.3000	46.7381
29	0.0000	4.5688	0.0000	3.1509	4.5688	20.8739	3.1509	9.9282	115.1000	114.8690	45.3000	49.8890
30	0.5000	0.0895	0.0000	0.0000	-0.4105	0.1685	0.0000	0.0000	115.6000	114.9585	45.3000	49.8890
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	114.9585	45.3000	49.8890
30	0.0000	0.6081	0.0000	0.0052	0.6081	0.3698	0.0052	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942

days	E(obs)	E(pre)	K(obs)	K(pre)	E(error)	E(error^2)	K(error)	K(error^2)	E(cum)	E(pre.cum)	K(cum)	K(pre.cum)
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	115.6000	115.5666	45.3000	49.8942
30	0.8000	0.0000	1.9000	0.0000	-0.8000	0.6400	-1.9000	3.6100	116.4000	115.5666	47.2000	49.8942
30	4.2000	0.0000	4.6000	0.0000	-4.2000	17.6400	-4.6000	21.1600	120.6000	115.5666	51.8000	49.8942
30	0.0000	1.0368	0.0000	2.2990	1.0368	1.0750	2.2990	5.2854	120.6000	116.6034	51.8000	52.1932
30	0.5000	5.0248	0.0000	5.1891	4.5248	20.4738	5.1891	26.9268	121.1000	121.6282	51.8000	57.3823
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	121.6282	51.8000	57.3823
30	0.0000	0.6078	0.0000	0.0049	0.6078	0.3694	0.0049	0.0000	121.1000	122.2360	51.8000	57.3872
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	122.2360	51.8000	57.3872
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	122.2360	51.8000	57.3872
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	122.2360	51.8000	57.3872
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	122.2360	51.8000	57.3872
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	122.2360	51.8000	57.3872
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	122.2360	51.8000	57.3872
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	122.2360	51.8000	57.3872
1	0.0000	0.0000	0.5000	0.0000	0.0000	0.0000	-0.5000	0.2500	121.1000	122.2360	52.3000	57.3872
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	121.1000	122.2360	52.3000	57.3872
1	0.5000	0.0168	0.0000	0.6028	-0.4832	0.2335	0.6028	0.3634	121.6000	122.2528	52.3000	57.9900
1	2.0000	0.0000	1.5000	0.0000	-2.0000	4.0000	-1.5000	2.2500	123.6000	122.2528	53.8000	57.9900
1	0.5000	0.6079	0.0000	0.0049	0.1079	0.0116	0.0049	0.0000	124.1000	122.8607	53.8000	57.9949
1	0.0000	2.3737	0.0000	1.8226	2.3737	5.6345	1.8226	3.3219	124.1000	125.2344	53.8000	59.8175
1	0.0000	0.1262	0.0000	0.0000	0.1262	0.0159	0.0000	0.0000	124.1000	125.3606	53.8000	59.8175
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	125.3606	53.8000	59.8175
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	125.3606	53.8000	59.8175
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	125.3606	53.8000	59.8175
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	125.3606	53.8000	59.8175
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	125.3606	53.8000	59.8175
days	**E(obs)**	**E(pre)**	**K(obs)**	**K(pre)**	**E(error)**	**E(error^2)**	**K(error)**	**K(error^2)**	**E(cum)**	**E(pre.cum)**	**K(cum)**	**K(pre.cum)**
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	125.3606	53.8000	59.8175

1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	125.3606	53.8000	59.8175
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	124.1000	125.3606	53.8000	59.8175

E(obs) = Observed value for station Empangan Genting Kelang

E(pre) = Forecasted value for station Empangan Genting Kelang

K(obs) = Observed value for station Kampung Kuala Saleh

K(pre) = Forecasted value for station Kampung Kuala Saleh

E(error) = E(obs) - E(pre)

K(error) = K(obs) - K(pre)

E($error^2$) = $(E(obs) - E(pre))^2$

K($error^2$) = $(K(obs) - K(pre))^2$

E(cum) = Cumulative observed value for station Empangan Genting Kelang

E(pre.cum) = Cumulative forecasted value for station Empangan Genting Kelang

K(cum) = Cumulative observed value for station Kampung Kuala Saleh

K(pre.cum) = Cumulative forecasted value for station Kampung Kuala Saleh

4.6 Forecasting Rainfalls Using the ARMA Models

To evaluate the performances of the MARIMA model, ARMA model was also used to forecast the rainfalls data. Since the ARMA model is a univariate Box-Jenkins model, the data need to be analyzed individually. Since this is a one-hour ahead forecast process, we need to repeatedly estimate the parameters for the chosen ARMA model every time we do the forecast.

To produce forecasts using the ARMA model, the MINITAB 14 software was used. However, since this is a one-hour prediction with the purpose of comparing the results with the ones obtained from using the MARIMA model, the scatter plots were obtained using the Microsoft Office Excel software.

The data from station Empangan Genting Kelang, station Km.11 Gombak and finally station Kampung Kuala Saleh were analyzed and forecasted separately. The best ARMA model for those stations were ARMA(1,1) model. The results are shown in Figures 4.54., 4.55, 4.56, 4.57, 4.58 and 4.59.

Figures 4.54, 4.56 and 4.58 show the hyetographs of observed rainfall intensities and corresponding forecasts of one-hour ahead. Figure 4.54 shows that most of the forecasted values were almost zero and one of the forecasted values is negative which we can assume as zero or in other words there was no rain. Figure 4.56 show that the forecasted values much more influence by its last observed value where its have four negative forecast values. For station Kampung Kuala Saleh, it showed in Figure 4.58 that the forecast values are most likely equal. Therefore we can assume that the forecast values for this station were just like the average value of its own two or three hours past data.

Figures 4.55, 4.57 and 4.59 show the observed and forecasted cumulative rainfalls. From those figures, we could see that the cumulative forecasted values were much differs from the cumulative observed values. This were caused by the negative forecast values and for station Kampung Kuala Saleh, this also caused by its own forecasted values that were just like the average value of its own two or three hours past data.

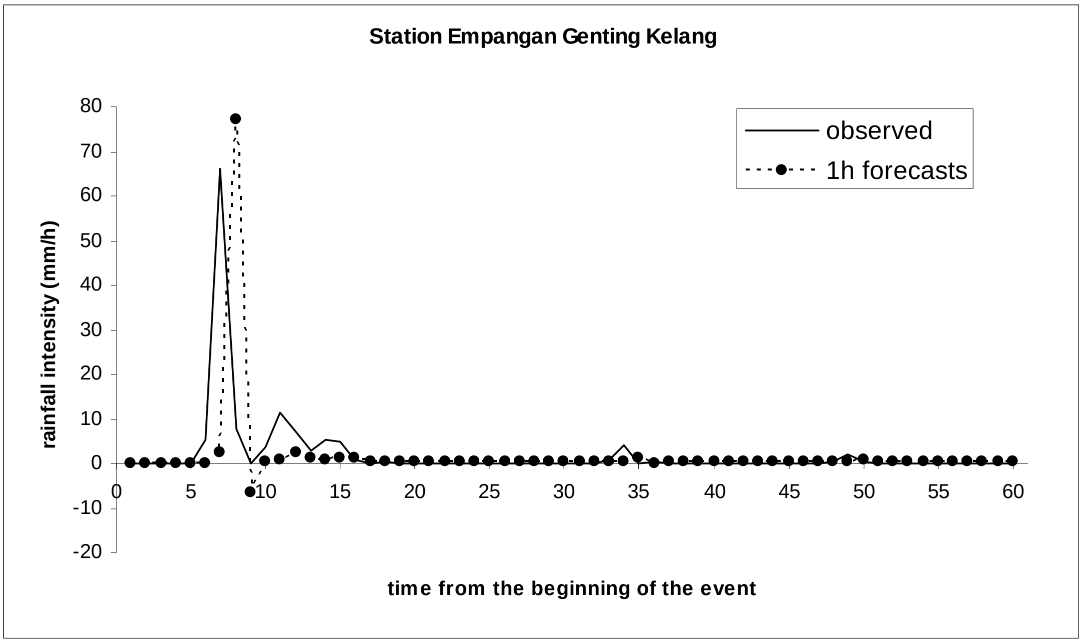


Figure 4.54: The hyetographs of observed rainfall intensity and ARMA(1,1) one-hour ahead forecast for station Empangan Genting Kelang.

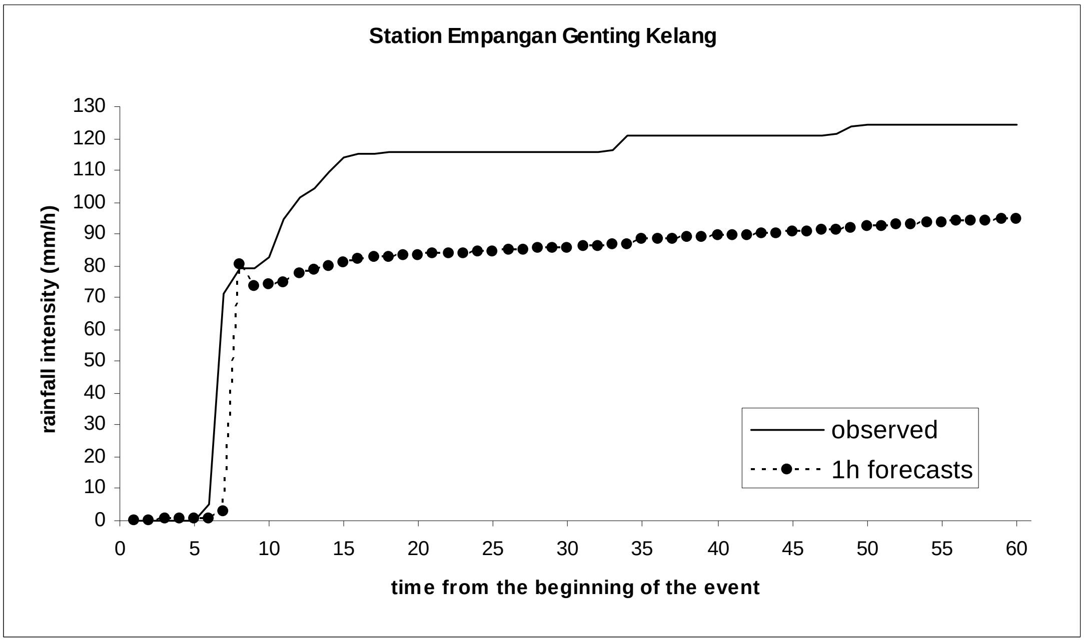


Figure 4.55: The observed and ARMA(1,1) one-hour ahead forecast cumulative rainfall intensity for station Empangan Genting Kelang.

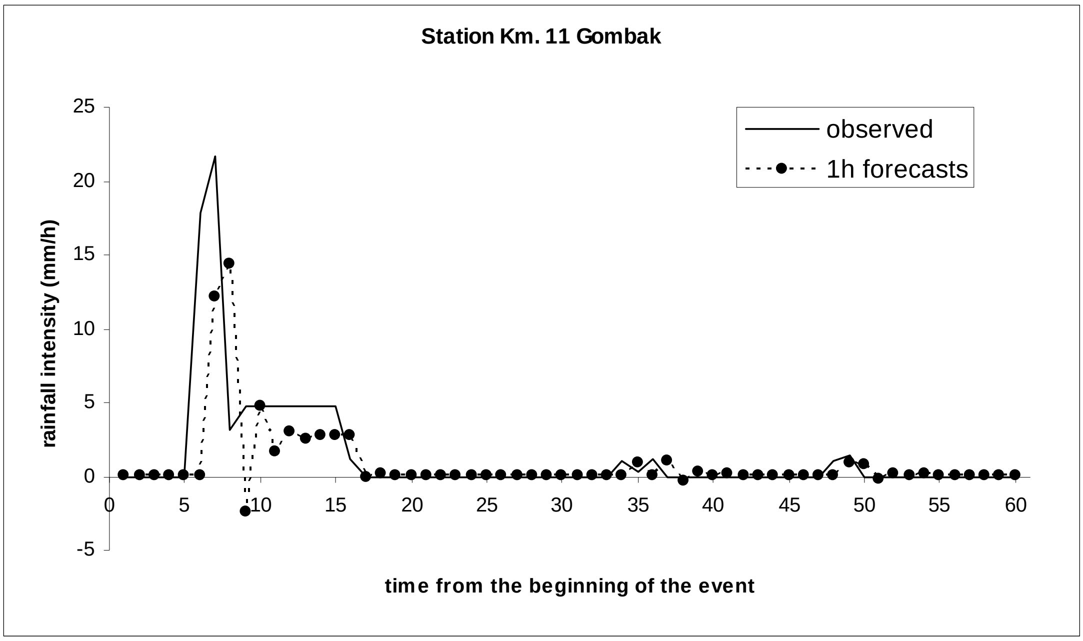


Figure 4.56: The hyetographs of observed rainfall intensity and ARMA(1,1) one-hour ahead forecast for station Km.11 Gombak.

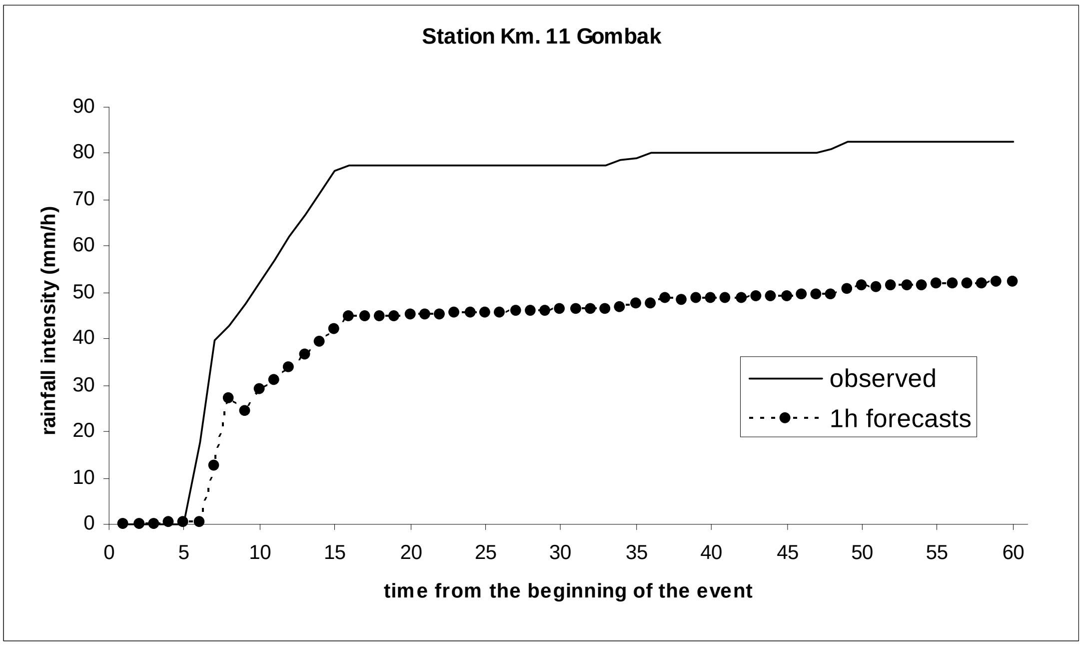


Figure 4.57: The observed and ARMA(1,1) one-hour ahead forecast cumulative rainfall intensity for station Km.11 Gombak.

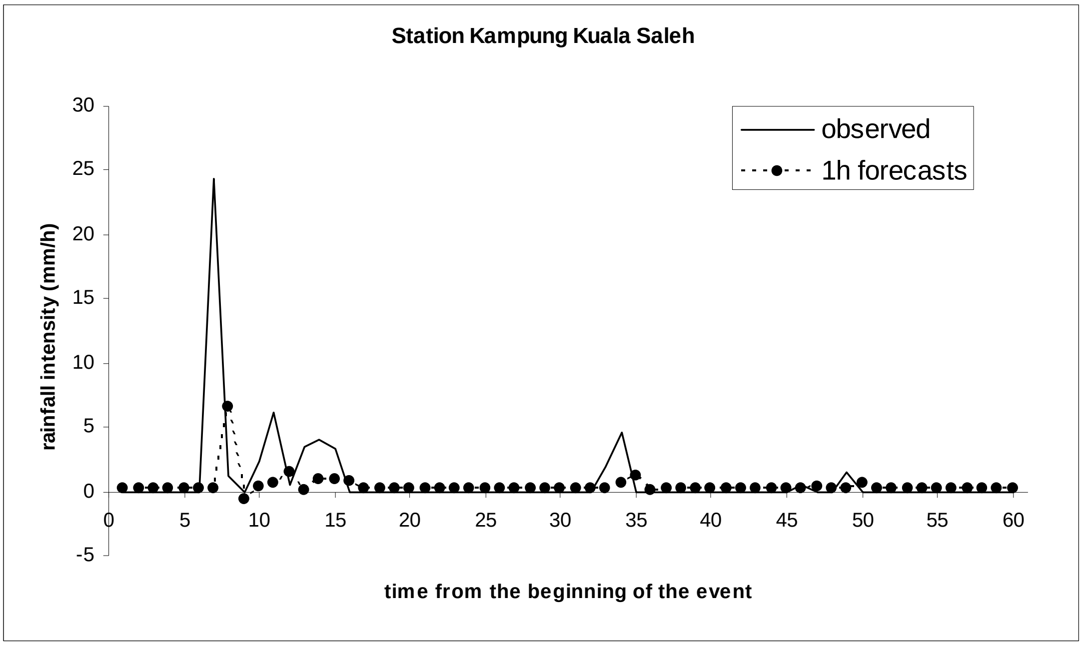


Figure 4.58: The hyetographs of observed rainfall intensity and ARMA(1,1) one-hour ahead forecast for station Kampung Kuala Saleh.

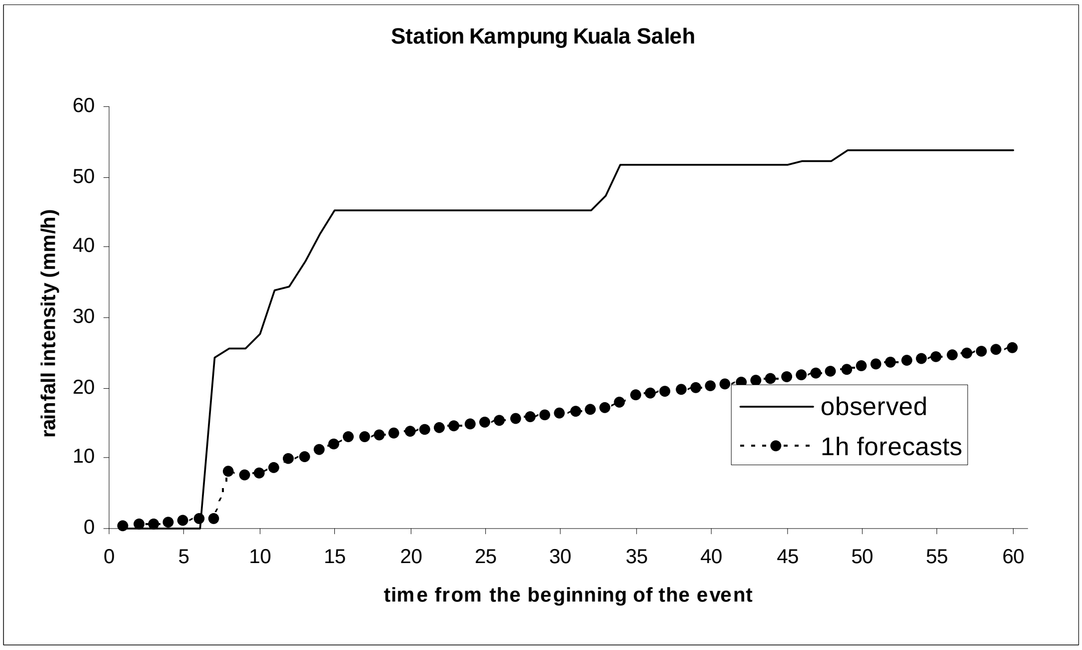


Figure 4.59: The observed and ARMA(1,1) one-hour ahead forecast cumulative rainfall intensity for station Kampung Kuala Saleh.

In Tables 4.22, 4.23 and 4.24**,** the numerical results for the prediction of the rainfalls for all the three stations using ARMA(1,1) are presented. In these tables, the observed values (obs), forecasted values (pre), the error between the observed values and the forecasted values (error), the squared error between the observed values and the forecasted values (error^2), the cumulative observed values (cum) and the cumulative forecasted values (pre.cum) are tabulated.

Table 4.22: Results for ARMA(1,1) model forecast of rainfalls intensity for station Empangan Genting Kelang.

days	E(obs)	E(pre)	E(error)	E(error^2)	E(cum)	E(pre.cum)
29	0.0000	0.102	0.1020	0.0104	0.0000	0.1020
29	0.0000	0.1019	0.1019	0.0104	0.0000	0.2039
29	0.0000	0.1017	0.1017	0.0103	0.0000	0.3056
29	0.0000	0.1016	0.1016	0.0103	0.0000	0.4072
29	0.0000	0.1014	0.1014	0.0103	0.0000	0.5086
29	5.2000	0.1013	-5.0987	25.9967	5.2000	0.6099
29	66.0000	2.3138	-63.6862	4055.9321	71.2000	2.9237
29	7.9000	77.2511	69.3511	4809.5751	79.1000	80.1748
29	0.0000	-6.4969	-6.4969	42.2097	79.1000	73.6779
29	3.8000	0.4148	-3.3852	11.4596	82.9000	74.0927
29	11.6000	0.8561	-10.7439	115.4314	94.5000	74.9488
29	7.0000	2.6457	-4.3543	18.9599	101.5000	77.5945
29	2.7000	1.3537	-1.3463	1.8125	104.2000	78.9482
29	5.2000	0.6486	-4.5514	20.7152	109.4000	79.5968
29	4.7000	1.3865	-3.3135	10.9793	114.1000	80.9833
29	1.0000	1.1585	0.1585	0.0251	115.1000	82.1418
29	0.0000	0.3228	0.3228	0.1042	115.1000	82.4646
30	0.5000	0.233	-0.2670	0.0713	115.6000	82.6976
30	0.0000	0.3679	0.3679	0.1354	115.6000	83.0655
30	0.0000	0.2247	0.2247	0.0505	115.6000	83.2902
30	0.0000	0.2498	0.2498	0.0624	115.6000	83.5400
30	0.0000	0.2449	0.2449	0.0600	115.6000	83.7849
30	0.0000	0.2453	0.2453	0.0602	115.6000	84.0302
30	0.0000	0.2448	0.2448	0.0599	115.6000	84.2750
30	0.0000	0.2445	0.2445	0.0598	115.6000	84.5195
30	0.0000	0.2442	0.2442	0.0596	115.6000	84.7637
30	0.0000	0.2438	0.2438	0.0594	115.6000	85.0075
30	0.0000	0.2434	0.2434	0.0592	115.6000	85.2509
30	0.0000	0.2431	0.2431	0.0591	115.6000	85.4940
30	0.0000	0.2428	0.2428	0.0590	115.6000	85.7368
30	0.0000	0.2424	0.2424	0.0588	115.6000	85.9792
30	0.0000	0.2421	0.2421	0.0586	115.6000	86.2213
30	0.8000	0.2417	-0.5583	0.3117	116.4000	86.4630
30	4.2000	0.4325	-3.7675	14.1941	120.6000	86.8955
30	0.0000	1.2312	1.2312	1.5159	120.6000	88.1267
30	0.5000	0.0742	-0.4258	0.1813	121.1000	88.2009
30	0.0000	0.3956	0.3956	0.1565	121.1000	88.5965
days	**E(obs)**	**E(pre)**	**E(error)**	**E(error^2)**	**E(cum)**	**E(pre.cum)**
30	0.0000	0.22	0.2200	0.0484	121.1000	88.8165
30	0.0000	0.2506	0.2506	0.0628	121.1000	89.0671

30	0.0000	0.2448	0.2448	0.0599	121.1000	89.3119
30	0.0000	0.2454	0.2454	0.0602	121.1000	89.5573
1	0.0000	0.2449	0.2449	0.0600	121.1000	89.8022
1	0.0000	0.2446	0.2446	0.0598	121.1000	90.0468
1	0.0000	0.2442	0.2442	0.0596	121.1000	90.2910
1	0.0000	0.2439	0.2439	0.0595	121.1000	90.5349
1	0.0000	0.2435	0.2435	0.0593	121.1000	90.7784
1	0.0000	0.2432	0.2432	0.0591	121.1000	91.0216
1	0.5000	0.2429	-0.2571	0.0661	121.6000	91.2645
1	2.0000	0.3617	-1.6383	2.6840	123.6000	91.6262
1	0.5000	0.6986	0.1986	0.0394	124.1000	92.3248
1	0.0000	0.2835	0.2835	0.0804	124.1000	92.6083
1	0.0000	0.2381	0.2381	0.0567	124.1000	92.8464
1	0.0000	0.2457	0.2457	0.0604	124.1000	93.0921
1	0.0000	0.2439	0.2439	0.0595	124.1000	93.3360
1	0.0000	0.2439	0.2439	0.0595	124.1000	93.5799
1	0.0000	0.2435	0.2435	0.0593	124.1000	93.8234
1	0.0000	0.2431	0.2431	0.0591	124.1000	94.0665
1	0.0000	0.2428	0.2428	0.0590	124.1000	94.3093
1	0.0000	0.2425	0.2425	0.0588	124.1000	94.5518
1	0.0000	0.2421	0.2421	0.0586	124.1000	94.7939

Table 4.23: Results for ARMA(1,1) model forecast of rainfalls intensity for station Km.11 Gombak.

days	G(obs)	G(pre)	G(error)	G(error^2)	G(cum)	G(pre.cum)
29	0.0000	0.0656	0.0656	0.0043	0.0000	0.0656
29	0.0000	0.0655	0.0655	0.0043	0.0000	0.1311
29	0.0000	0.0654	0.0654	0.0043	0.0000	0.1965
29	0.0000	0.0653	0.0653	0.0043	0.0000	0.2618
29	0.0000	0.0652	0.0652	0.0043	0.0000	0.3270
29	17.8000	0.0651	-17.7349	314.5267	17.8000	0.3921
29	21.7000	12.1489	-9.5511	91.2235	39.5000	12.5410
29	3.2000	14.4196	11.2196	125.8794	42.7000	26.9606
29	4.8000	-2.4233	-7.2233	52.1761	47.5000	24.5373
29	4.8000	4.6985	-0.1015	0.0103	52.3000	29.2358
29	4.8000	1.6676	-3.1324	9.8119	57.1000	30.9034
29	4.8000	2.9809	-1.8191	3.3091	61.9000	33.8843
29	4.8000	2.5472	-2.2528	5.0751	66.7000	36.4315
29	4.8000	2.7708	-2.0292	4.1177	71.5000	39.2023
29	4.8000	2.7262	-2.0738	4.3006	76.3000	41.9285
29	1.2000	2.7819	1.5819	2.5024	77.5000	44.7104
29	0.0000	-0.0198	-0.0198	0.0004	77.5000	44.6906
30	0.0000	0.1669	0.1669	0.0279	77.5000	44.8575
30	0.0000	0.0924	0.0924	0.0085	77.5000	44.9499
30	0.0000	0.1218	0.1218	0.0148	77.5000	45.0717
30	0.0000	0.1098	0.1098	0.0121	77.5000	45.1815
days	G(obs)	G(pre)	G(error)	G(error^2)	G(cum)	G(pre.cum)
30	0.0000	0.1143	0.1143	0.0131	77.5000	45.2958
30	0.0000	0.1123	0.1123	0.0126	77.5000	45.4081
30	0.0000	0.1129	0.1129	0.0127	77.5000	45.5210

30	0.0000	0.1124	0.1124	0.0126	77.5000	45.6334
30	0.0000	0.1124	0.1124	0.0126	77.5000	45.7458
30	0.0000	0.1122	0.1122	0.0126	77.5000	45.8580
30	0.0000	0.1120	0.1120	0.0125	77.5000	45.9700
30	0.0000	0.1118	0.1118	0.0125	77.5000	46.0818
30	0.0000	0.1117	0.1117	0.0125	77.5000	46.1935
30	0.0000	0.1115	0.1115	0.0124	77.5000	46.3050
30	0.0000	0.1114	0.1114	0.0124	77.5000	46.4164
30	0.0000	0.1112	0.1112	0.0124	77.5000	46.5276
30	1.0000	0.1110	-0.8890	0.7903	78.5000	46.6386
30	0.3000	0.8821	0.5821	0.3388	78.8000	47.5207
30	1.2000	0.0359	-1.1641	1.3551	80.0000	47.5566
30	0.0000	1.0648	1.0648	1.1338	80.0000	48.6214
30	0.0000	-0.2632	-0.2632	0.0693	80.0000	48.3582
30	0.0000	0.2596	0.2596	0.0674	80.0000	48.6178
30	0.0000	0.0538	0.0538	0.0029	80.0000	48.6716
30	0.0000	0.1345	0.1345	0.0181	80.0000	48.8061
1	0.0000	0.1025	0.1025	0.0105	80.0000	48.9086
1	0.0000	0.1149	0.1149	0.0132	80.0000	49.0235
1	0.0000	0.1098	0.1098	0.0121	80.0000	49.1333
1	0.0000	0.1115	0.1115	0.0124	80.0000	49.2448
1	0.0000	0.1106	0.1106	0.0122	80.0000	49.3554
1	0.0000	0.1108	0.1108	0.0123	80.0000	49.4662
1	1.1000	0.1105	-0.9895	0.9791	81.1000	49.5767
1	1.4000	0.9552	-0.4448	0.1978	82.5000	50.5319
1	0.0000	0.8538	0.8538	0.7290	82.5000	51.3857
1	0.0000	-0.1810	-0.1810	0.0328	82.5000	51.2047
1	0.0000	0.2272	0.2272	0.0516	82.5000	51.4319
1	0.0000	0.0659	0.0659	0.0043	82.5000	51.4978
1	0.0000	0.1293	0.1293	0.0167	82.5000	51.6271
1	0.0000	0.1040	0.1040	0.0108	82.5000	51.7311
1	0.0000	0.1138	0.1138	0.0130	82.5000	51.8449
1	0.0000	0.1097	0.1097	0.0120	82.5000	51.9546
1	0.0000	0.1111	0.1111	0.0123	82.5000	52.0657
1	0.0000	0.1103	0.1103	0.0122	82.5000	52.1760
1	0.0000	0.1104	0.1104	0.0122	82.5000	52.2864

Table 4.24: Results for ARMA(1,1) model forecast of rainfalls intensity for station Kampung Kuala Saleh.

days	K(obs)	K(pre)	K(error)	K($error^2$)	K(cum)	K(pre.cum)
29	0.0000	0.1998	0.1998	0.0399	0.0000	0.1998
29	0.0000	0.1995	0.1995	0.0398	0.0000	0.3993
29	0.0000	0.1992	0.1992	0.0397	0.0000	0.5985
29	0.0000	0.1989	0.1989	0.0396	0.0000	0.7974
29	0.0000	0.1986	0.1986	0.0394	0.0000	0.9960
days	**K(obs)**	**K(pre)**	**K(error)**	**K($error^2$)**	**K(cum)**	**K(pre.cum)**
29	0.0000	0.1983	0.1983	0.0393	0.0000	1.1943
29	24.3000	0.1980	-24.1020	580.9064	24.3000	1.3923
29	1.2000	6.6212	5.4212	29.3894	25.5000	8.0135
29	0.0000	-0.5657	-0.5657	0.3200	25.5000	7.4478

29	2.3000	0.3595	-1.9405	3.7655	27.8000	7.8073
29	6.2000	0.6753	-5.5247	30.5223	34.0000	8.4826
29	0.5000	1.4638	0.9638	0.9289	34.5000	9.9464
29	3.5000	0.1182	-3.3818	11.4366	38.0000	10.0646
29	4.0000	0.9898	-3.0102	9.0613	42.0000	11.0544
29	3.3000	0.9685	-2.3315	5.4359	45.3000	12.0229
29	0.0000	0.8475	0.8475	0.7183	45.3000	12.8704
29	0.0000	0.1626	0.1626	0.0264	45.3000	13.0330
30	0.0000	0.2719	0.2719	0.0739	45.3000	13.3049
30	0.0000	0.2537	0.2537	0.0644	45.3000	13.5586
30	0.0000	0.2562	0.2562	0.0656	45.3000	13.8148
30	0.0000	0.2553	0.2553	0.0652	45.3000	14.0701
30	0.0000	0.2550	0.2550	0.0650	45.3000	14.3251
30	0.0000	0.2547	0.2547	0.0649	45.3000	14.5798
30	0.0000	0.2543	0.2543	0.0647	45.3000	14.8341
30	0.0000	0.2539	0.2539	0.0645	45.3000	15.0880
30	0.0000	0.2535	0.2535	0.0643	45.3000	15.3415
30	0.0000	0.2532	0.2532	0.0641	45.3000	15.5947
30	0.0000	0.2528	0.2528	0.0639	45.3000	15.8475
30	0.0000	0.2524	0.2524	0.0637	45.3000	16.0999
30	0.0000	0.2521	0.2521	0.0636	45.3000	16.3520
30	0.0000	0.2517	0.2517	0.0634	45.3000	16.6037
30	0.0000	0.2513	0.2513	0.0632	45.3000	16.8550
30	1.9000	0.2510	-1.6490	2.7192	47.2000	17.1060
30	4.6000	0.6584	-3.9416	15.5362	51.8000	17.7644
30	0.0000	1.1846	1.1846	1.4033	51.8000	18.9490
30	0.0000	0.1052	0.1052	0.0111	51.8000	19.0542
30	0.0000	0.2826	0.2826	0.0799	51.8000	19.3368
30	0.0000	0.2525	0.2525	0.0638	51.8000	19.5893
30	0.0000	0.2571	0.2571	0.0661	51.8000	19.8464
30	0.0000	0.2559	0.2559	0.0655	51.8000	20.1023
30	0.0000	0.2556	0.2556	0.0653	51.8000	20.3579
1	0.0000	0.2553	0.2553	0.0652	51.8000	20.6132
1	0.0000	0.2549	0.2549	0.0650	51.8000	20.8681
1	0.0000	0.2545	0.2545	0.0648	51.8000	21.1226
1	0.0000	0.2542	0.2542	0.0646	51.8000	21.3768
1	0.5000	0.2538	-0.2462	0.0606	52.3000	21.6306
1	0.0000	0.3613	0.3613	0.1305	52.3000	21.9919
1	0.0000	0.2360	0.2360	0.0557	52.3000	22.2279
1	1.5000	0.2563	-1.2437	1.5468	53.8000	22.4842
1	0.0000	0.5757	0.5757	0.3314	53.8000	23.0599
1	0.0000	0.2013	0.2013	0.0405	53.8000	23.2612
1	0.0000	0.2630	0.2630	0.0692	53.8000	23.5242
1	0.0000	0.2523	0.2523	0.0637	53.8000	23.7765
1	0.0000	0.2536	0.2536	0.0643	53.8000	24.0301
1	0.0000	0.2530	0.2530	0.0640	53.8000	24.2831
1	0.0000	0.2527	0.2527	0.0639	53.8000	24.5358
days	**K(obs)**	**K(pre)**	**K(error)**	**K(error^2)**	**K(cum)**	**K(pre.cum)**
1	0.0000	0.2523	0.2523	0.0637	53.8000	24.7881
1	0.0000	0.2520	0.2520	0.0635	53.8000	25.0401
1	0.0000	0.2516	0.2516	0.0633	53.8000	25.2917
1	0.0000	0.2513	0.2513	0.0632	53.8000	25.5430

4.7 Comparison Between MARIMA and ARMA Models in Forecasting Rainfalls Data

Scatter diagrams have been plotted to illustrate the differences between the forecasts obtained using the MARIMA and the ARMA(1,1) models. Figures 4.60, 4.62, 4.64 and 4.66 show the hyetographs of observed rainfall intensities and corresponding one-hour ahead forecasts for both MARIMA and ARIMA models. Figures 4.61, 4.63 , 4.65 and 4.67 show the observed and both models forecasted cumulative rainfalls.

Comparing the forecasted values using the MARIMA model and the ARMA(1,1) model in Figures 4.60 and 4.62, we could see that the forecast values using the ARMA(1,1) model were more influenced by an hour past data while the MARIMA model were influenced by the last two hours data. It goes the same for station Km.11 Gombak as shown in Figure 4.64 however we could see that most of the forecasted value by the MARIMA model were most likely the same as the observed values compared to the ARMA(1,1) model. For station Kampung Kuala Saleh, forecasted values for both models were differs than the observed values where both models could only be considered good in forecasting zero values data. Figures 4.61, 4.63, 4.65 and 4.67 shows that the cumulative forecasted values using the MARIMA model were slightly differs than cumulative observed values compared to the ARMA(1,1) model. The cumulative forecasted values for the ARMA(1,1) model were differs form the cumulative observed values were caused by the negative values of the forecast for the ARMA(1,1) model.

The numerical results for the prediction rainfalls values using both models are given in Tables 4.25, 4.26 and 4.27. Table 4.24 lists the values for station Empangan Genting Kelang where E(MG) are the forecast values using the MARIMA model (with station Km.11 Gombak), E(MK) are the forecast values using the MARIMA model (with station Kampung Kuala Saleh), and E(ARMA) are the forecast values using the ARMA(1,1) model. Their cumulative rainfall values are denoted by (cum). Tables 4.26 and 4.27 are for station Km.11 Gombak and station Kampung Kuala Saleh where G(M) and K(M) are the forecast values using the MARIMA model, G(A) and K(A) are the forecast values using the ARMA(1,1) model and their cumulative rainfall values are denoted by (cum).

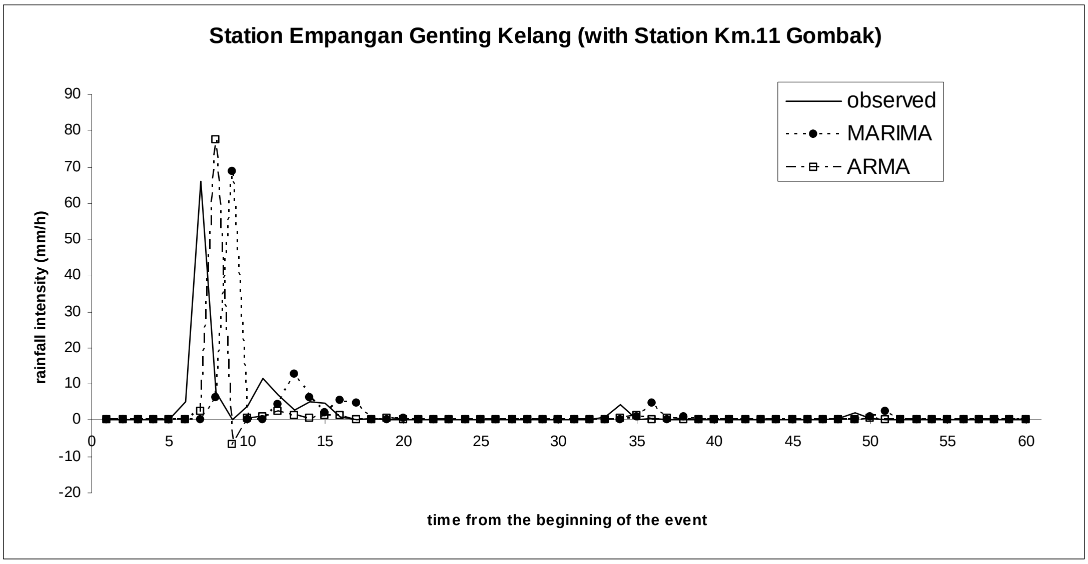


Figure 4.60: The hyetographs of observed rainfall intensity, MARIMA and ARMA(1,1) one-hour ahead forecast for station Empangan Genting Kelang (with station Km.11 Gombak).

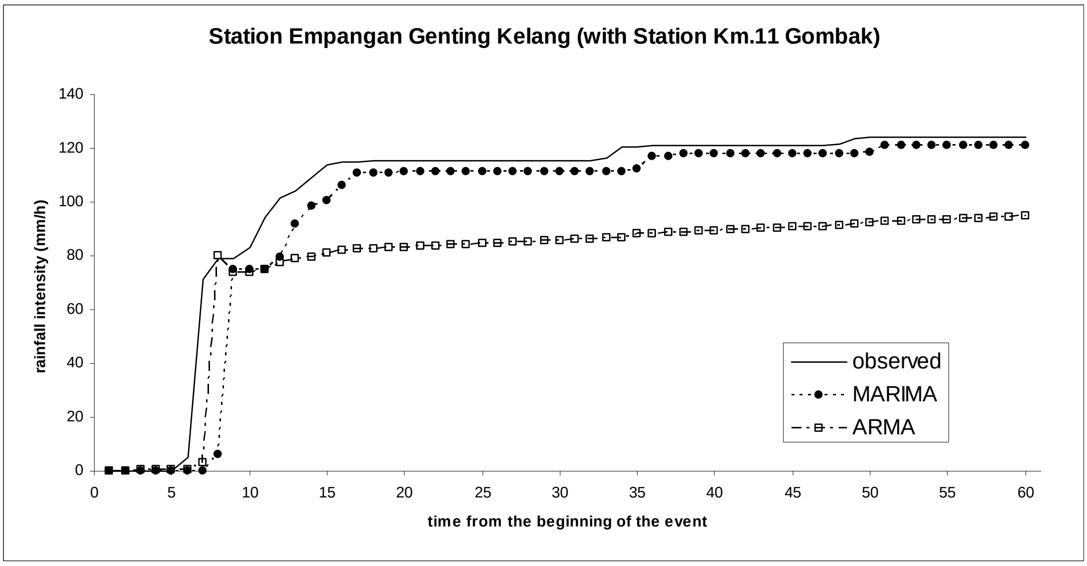


Figure 4.61: The observed, MARIMA and ARMA(1,1) one-hour ahead forecast cumulative rainfall intensity for station Empangan Genting Kelang (with station Km.11 Gombak).

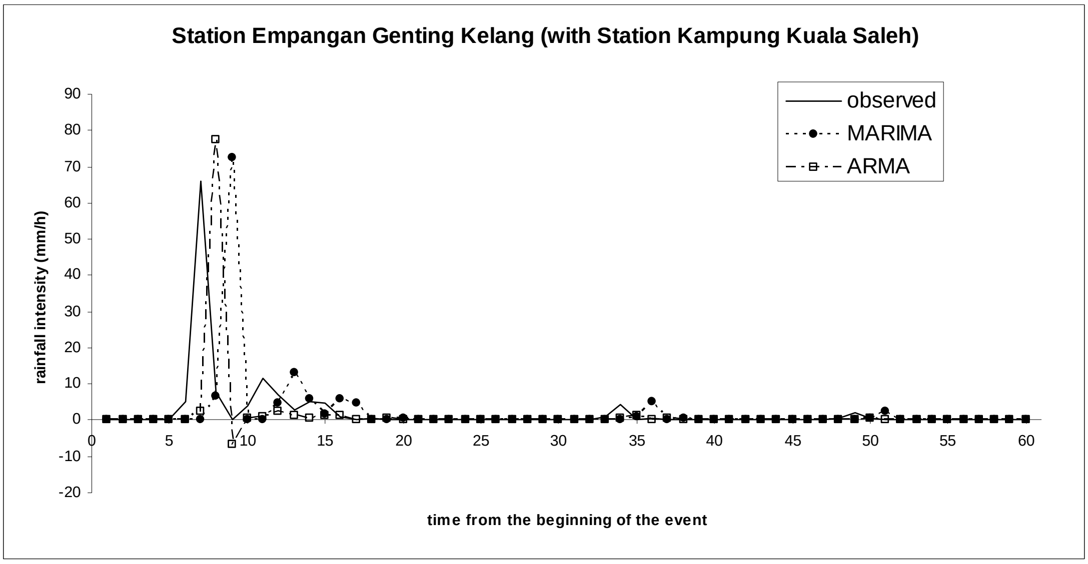


Figure 4.62: The hyetographs of observed rainfall intensity, MARIMA and ARIMA(1,1) one-hour ahead forecast for station Empangan Genting Kelang (with station Kampung Kuala Saleh).

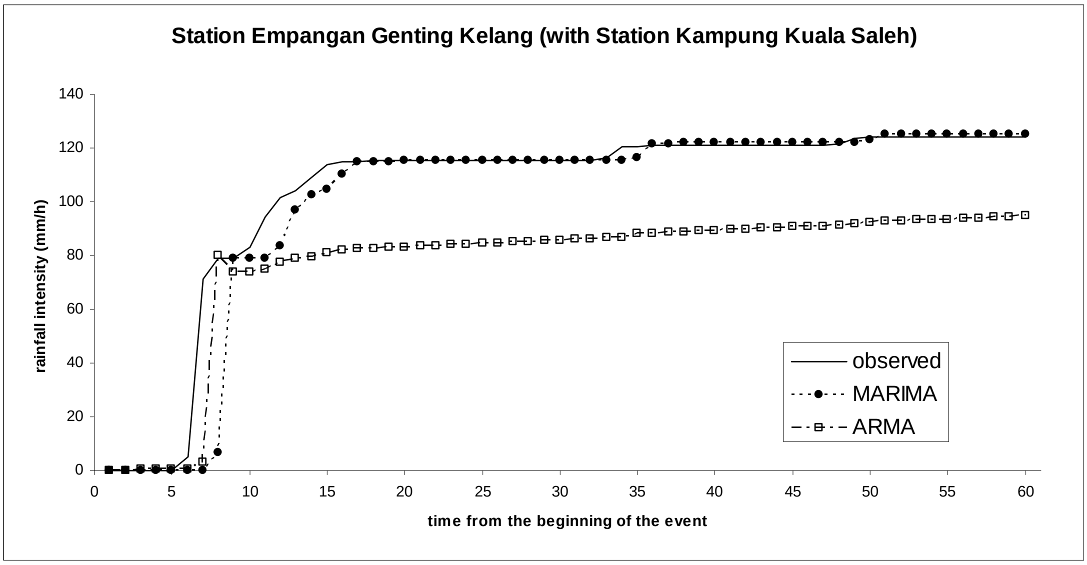


Figure 4.63: The observed, MARIMA and ARIMA(1,1) one-hour ahead forecast cumulative rainfall intensity for station Empangan Genting Kelang (with station Kampung Kuala Saleh).

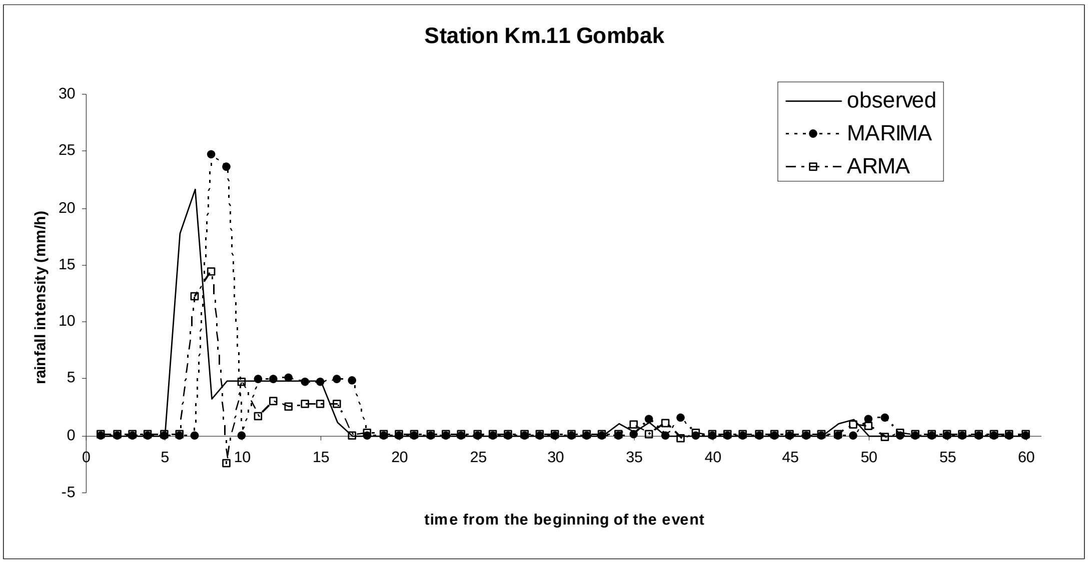


Figure 4.64: The hyetographs of observed rainfall intensity, MARIMA and ARMA(1,1) one-hour ahead forecast for station Km.11 Gombak.

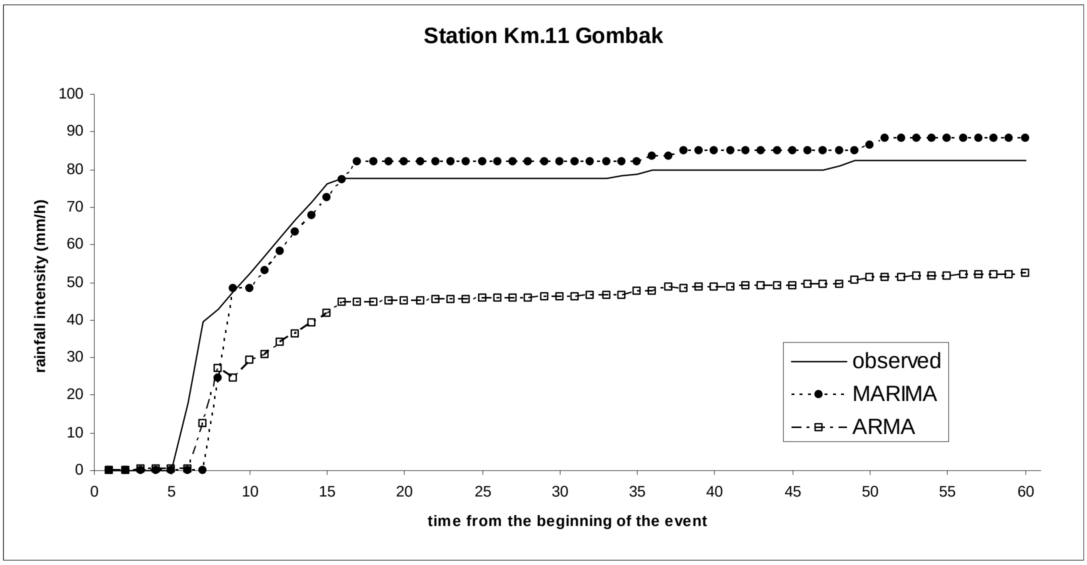


Figure 4.65: The observed, MARIMA and ARMA(1,1) one-hour ahead forecast cumulative rainfall intensity for station Km.11 Gombak.

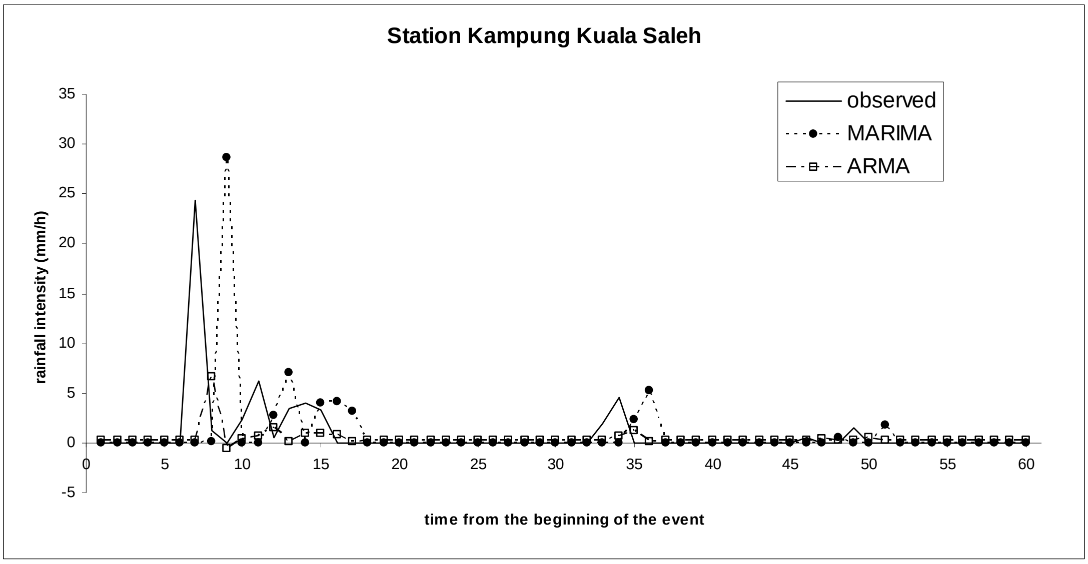


Figure 4.66: The hyetographs of observed rainfall intensity, MARIMA and ARMA(1,1) one-hour ahead forecast for station Kampung Kuala Saleh.

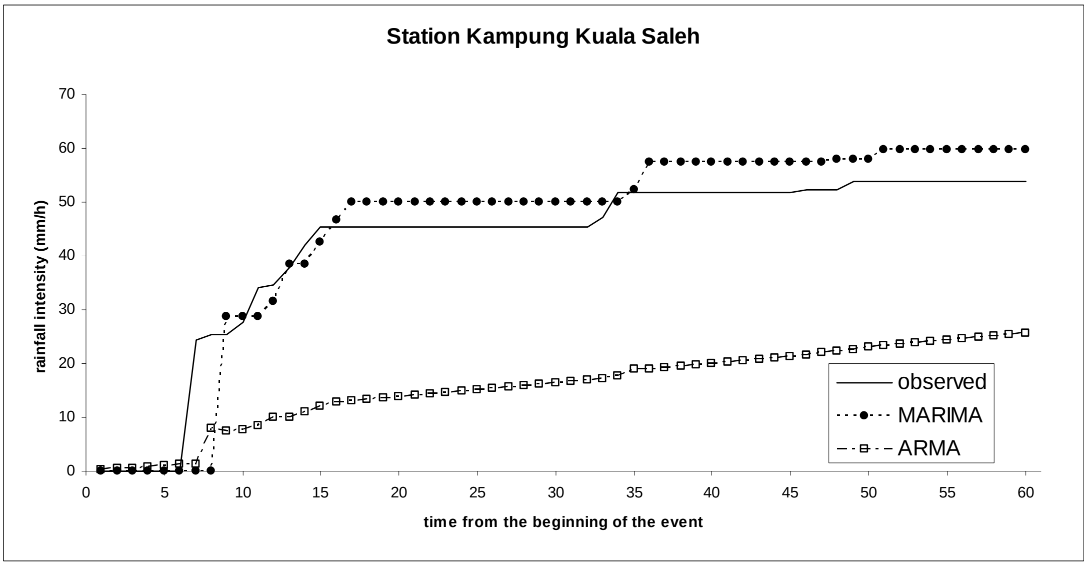


Figure 4.67: The observed, MARIMA and ARMA(1,1) one-hour ahead forecast cumulative rainfall intensity for station Kampung Kuala Saleh.

Table 4.25: Comparison of rainfalls intensity forecast value from MARIMA model and ARMA(1,1) model for station Empangan Genting Kelang.

days	E(obs)	E(MG)	E(MK)	E(ARMA)	E(cum)	E(MGcum)	E(MKcum)	E(Acum)
29	0.0000	0.0000	0.0000	0.102	0.0000	0	0	0.102
29	0.0000	0.0000	0.0000	0.1019	0.0000	0	0	0.2039
29	0.0000	0.0000	0.0000	0.1017	0.0000	0	0	0.3056
29	0.0000	0.0000	0.0000	0.1016	0.0000	0	0	0.4072
29	0.0000	0.0000	0.0000	0.1014	0.0000	0	0	0.5086
29	5.2000	0.0000	0.0000	0.1013	5.2000	0	0	0.6099
29	66.0000	0.0000	0.0000	2.3138	71.2000	0	0	2.9237
29	7.9000	6.2725	6.6008	77.2511	79.1000	6.2725	6.6008	80.1748
29	0.0000	68.7757	72.3596	-6.4969	79.1000	75.0482	78.9604	73.6779
29	3.8000	0.0000	0.0000	0.4148	82.9000	75.0482	78.9604	74.0927
29	11.6000	0.0000	0.0000	0.8561	94.5000	75.0482	78.9604	74.9488
29	7.0000	4.3131	4.6056	2.6457	101.5000	79.3613	83.566	77.5945
29	2.7000	12.6704	13.2479	1.3537	104.2000	92.0317	96.8139	78.9482
29	5.2000	6.3066	5.8543	0.6486	109.4000	98.3383	102.6682	79.5968
29	4.7000	2.0401	1.8842	1.3865	114.1000	100.3784	104.5524	80.9833
29	1.0000	5.5846	5.7478	1.1585	115.1000	105.963	110.3002	82.1418
29	0.0000	4.6217	4.5688	0.3228	115.1000	110.5847	114.869	82.4646
30	0.5000	0.0000	0.0895	0.233	115.6000	110.5847	114.9585	82.6976
30	0.0000	0.0000	0.0000	0.3679	115.6000	110.5847	114.9585	83.0655
30	0.0000	0.5875	0.6081	0.2247	115.6000	111.1722	115.5666	83.2902
30	0.0000	0.0000	0.0000	0.2498	115.6000	111.1722	115.5666	83.54
30	0.0000	0.0000	0.0000	0.2449	115.6000	111.1722	115.5666	83.7849
30	0.0000	0.0000	0.0000	0.2453	115.6000	111.1722	115.5666	84.0302
30	0.0000	0.0000	0.0000	0.2448	115.6000	111.1722	115.5666	84.275
30	0.0000	0.0000	0.0000	0.2445	115.6000	111.1722	115.5666	84.5195
30	0.0000	0.0000	0.0000	0.2442	115.6000	111.1722	115.5666	84.7637
30	0.0000	0.0000	0.0000	0.2438	115.6000	111.1722	115.5666	85.0075
30	0.0000	0.0000	0.0000	0.2434	115.6000	111.1722	115.5666	85.2509
days	E(obs)	E(MG)	E(MK)	E(ARMA)	E(cum)	E(MGcum)	E(MKcum)	E(Acum)
30	0.0000	0.0000	0.0000	0.2431	115.6000	111.1722	115.5666	85.494
30	0.0000	0.0000	0.0000	0.2428	115.6000	111.1722	115.5666	85.7368
30	0.0000	0.0000	0.0000	0.2424	115.6000	111.1722	115.5666	85.9792
30	0.0000	0.0000	0.0000	0.2421	115.6000	111.1722	115.5666	86.2213
30	0.8000	0.0000	0.0000	0.2417	116.4000	111.1722	115.5666	86.463
30	4.2000	0.0000	0.0000	0.4325	120.6000	111.1722	115.5666	86.8955
30	0.0000	0.9257	1.0368	1.2312	120.6000	112.0979	116.6034	88.1267
30	0.5000	4.9288	5.0248	0.0742	121.1000	117.0267	121.6282	88.2009
30	0.0000	0.0000	0.0000	0.3956	121.1000	117.0267	121.6282	88.5965
30	0.0000	0.7529	0.6078	0.2200	121.1000	117.7796	122.236	88.8165
30	0.0000	0.0000	0.0000	0.2506	121.1000	117.7796	122.236	89.0671
30	0.0000	0.0000	0.0000	0.2448	121.1000	117.7796	122.236	89.3119
30	0.0000	0.0000	0.0000	0.2454	121.1000	117.7796	122.236	89.5573
1	0.0000	0.0000	0.0000	0.2449	121.1000	117.7796	122.236	89.8022
1	0.0000	0.0000	0.0000	0.2446	121.1000	117.7796	122.236	90.0468
1	0.0000	0.0000	0.0000	0.2442	121.1000	117.7796	122.236	90.291
1	0.0000	0.0000	0.0000	0.2439	121.1000	117.7796	122.236	90.5349
1	0.0000	0.0000	0.0000	0.2435	121.1000	117.7796	122.236	90.7784
1	0.0000	0.0000	0.0000	0.2432	121.1000	117.7796	122.236	91.0216

1	0.5000	0.0000	0.0168	0.2429	121.6000	117.7796	122.2528	91.2645
1	2.0000	0.0000	0.0000	0.3617	123.6000	117.7796	122.2528	91.6262
1	0.5000	0.7916	0.6079	0.6986	124.1000	118.5712	122.8607	92.3248
1	0.0000	2.2938	2.3737	0.2835	124.1000	120.865	125.2344	92.6083
1	0.0000	0.0000	0.1262	0.2381	124.1000	120.865	125.3606	92.8464
1	0.0000	0.0000	0.0000	0.2457	124.1000	120.865	125.3606	93.0921
1	0.0000	0.0000	0.0000	0.2439	124.1000	120.865	125.3606	93.336
1	0.0000	0.0000	0.0000	0.2439	124.1000	120.865	125.3606	93.5799
1	0.0000	0.0000	0.0000	0.2435	124.1000	120.865	125.3606	93.8234
1	0.0000	0.0000	0.0000	0.2431	124.1000	120.865	125.3606	94.0665
days	**E(obs)**	**E(MG)**	**E(MK)**	**E(ARMA)**	**E(cum)**	**E(MGcum)**	**E(MKcum)**	**E(Acum)**
1	0.0000	0.0000	0.0000	0.2428	124.1000	120.865	125.3606	94.3093
1	0.0000	0.0000	0.0000	0.2425	124.1000	120.865	125.3606	94.5518
1	0.0000	0.0000	0.0000	0.2421	124.1000	120.865	125.3606	94.7939

E(obs) = Observed values for station Empangan Genting Kelang

E(MG) = Forecasted values for station Empangan Genting Kelang using the MARIMA model (with station Km.11 Gombak)

E(MK) = Forecasted values for station Empangan Genting Kelang using the MARIMA model (with station Kampung Kuala Saleh)

E(ARMA) = Forecasted values for station Empangan Genting Kelang using the ARMA model

E(cum) = Cumulative observed values for station Empangan Genting Kelang

E(MGcum)= Cumulative forecasted values for station Empangan Genting Kelang using the MARIMA model (with station Km.11 Gombak)

E(MKcum)= Cumulative forecasted values for station Empangan Genting Kelang using the MARIMA model (with station Kg. Kuala Saleh)

E(Acum) = Cumulative forecasted values for station Empangan Genting Kelang using the ARMA model

Table 4.26: Comparison of rainfalls intensity forecast value from MARIMA model and ARMA(1,1) model for station Km.11 Gombak.

days	G(obs)	G(M)	G(A)	G(cum)	G(Mcum)	G(Acum)
29	0.0000	0.0000	0.0656	0.0000	0.0000	0.0656
29	0.0000	0.0000	0.0655	0.0000	0.0000	0.1311
29	0.0000	0.0000	0.0654	0.0000	0.0000	0.1965
29	0.0000	0.0000	0.0653	0.0000	0.0000	0.2618
29	0.0000	0.0000	0.0652	0.0000	0.0000	0.3270
29	17.8000	0.0000	0.0651	17.8000	0.0000	0.3921
29	21.7000	0.0000	12.1489	39.5000	0.0000	12.5410
29	3.2000	24.6598	14.4196	42.7000	24.6598	26.9606
29	4.8000	23.6321	-2.4233	47.5000	48.2919	24.5373
29	4.8000	0.0000	4.6985	52.3000	48.2919	29.2358
29	4.8000	4.9856	1.6676	57.1000	53.2775	30.9034
29	4.8000	4.9137	2.9809	61.9000	58.1912	33.8843
29	4.8000	5.0290	2.5472	66.7000	63.2202	36.4315
29	4.8000	4.6485	2.7708	71.5000	67.8687	39.2023
29	4.8000	4.6420	2.7262	76.3000	72.5107	41.9285
29	1.2000	4.8904	2.7819	77.5000	77.4011	44.7104
29	0.0000	4.7813	-0.0198	77.5000	82.1824	44.6906
30	0.0000	0.0000	0.1669	77.5000	82.1824	44.8575
30	0.0000	0.0000	0.0924	77.5000	82.1824	44.9499
30	0.0000	0.0188	0.1218	77.5000	82.2012	45.0717
30	0.0000	0.0000	0.1098	77.5000	82.2012	45.1815
30	0.0000	0.0000	0.1143	77.5000	82.2012	45.2958
30	0.0000	0.0000	0.1123	77.5000	82.2012	45.4081
30	0.0000	0.0000	0.1129	77.5000	82.2012	45.5210
30	0.0000	0.0000	0.1124	77.5000	82.2012	45.6334
30	0.0000	0.0000	0.1124	77.5000	82.2012	45.7458
30	0.0000	0.0000	0.1122	77.5000	82.2012	45.8580
30	0.0000	0.0000	0.1120	77.5000	82.2012	45.9700
30	0.0000	0.0000	0.1118	77.5000	82.2012	46.0818
30	0.0000	0.0000	0.1117	77.5000	82.2012	46.1935
30	0.0000	0.0000	0.1115	77.5000	82.2012	46.3050
30	0.0000	0.0000	0.1114	77.5000	82.2012	46.4164
30	0.0000	0.0000	0.1112	77.5000	82.2012	46.5276
30	1.0000	0.0000	0.1110	78.5000	82.2012	46.6386
30	0.3000	0.0302	0.8821	78.8000	82.2314	47.5207
30	1.2000	1.4296	0.0359	80.0000	83.6610	47.5566
30	0.0000	0.0000	1.0648	80.0000	83.6610	48.6214
30	0.0000	1.4901	-0.2632	80.0000	85.1511	48.3582
30	0.0000	0.0000	0.2596	80.0000	85.1511	48.6178
30	0.0000	0.0000	0.0538	80.0000	85.1511	48.6716
30	0.0000	0.0000	0.1345	80.0000	85.1511	48.8061
1	0.0000	0.0000	0.1025	80.0000	85.1511	48.9086
1	0.0000	0.0000	0.1149	80.0000	85.1511	49.0235
1	0.0000	0.0000	0.1098	80.0000	85.1511	49.1333

1	0.0000	0.0000	0.1115	80.0000	85.1511	49.2448
1	0.0000	0.0000	0.1106	80.0000	85.1511	49.3554
1	0.0000	0.0000	0.1108	80.0000	85.1511	49.4662
1	1.1000	0.0000	0.1105	81.1000	85.1511	49.5767
days	**G(obs)**	**G(M)**	**G(A)**	**G(cum)**	**G(Mcum)**	**G(Acum)**
1	1.4000	0.0000	0.9552	82.5000	85.1511	50.5319
1	0.0000	1.4504	0.8538	82.5000	86.6015	51.3857
1	0.0000	1.5466	-0.1810	82.5000	88.1481	51.2047
1	0.0000	0.0000	0.2272	82.5000	88.1481	51.4319
1	0.0000	0.0000	0.0659	82.5000	88.1481	51.4978
1	0.0000	0.0000	0.1293	82.5000	88.1481	51.6271
1	0.0000	0.0000	0.1040	82.5000	88.1481	51.7311
1	0.0000	0.0000	0.1138	82.5000	88.1481	51.8449
1	0.0000	0.0000	0.1097	82.5000	88.1481	51.9546
1	0.0000	0.0000	0.1111	82.5000	88.1481	52.0657
1	0.0000	0.0000	0.1103	82.5000	88.1481	52.1760
1	0.0000	0.0000	0.1104	82.5000	88.1481	52.2864

Table 4.27: Comparison of rainfalls intensity forecast value from MARIMA model and ARMA(1,1) model for station Kampung Kuala Saleh.

days	K(obs)	K(M)	K(A)	K(cum)	K(Mcum)	K(Acum)
29	0.0000	0.0000	0.1998	0.0000	0.0000	0.1998
29	0.0000	0.0000	0.1995	0.0000	0.0000	0.3993
29	0.0000	0.0000	0.1992	0.0000	0.0000	0.5985
29	0.0000	0.0000	0.1989	0.0000	0.0000	0.7974
29	0.0000	0.0000	0.1986	0.0000	0.0000	0.9960
29	0.0000	0.0000	0.1983	0.0000	0.0000	1.1943
29	24.3000	0.0000	0.1980	24.3000	0.0000	1.3923
29	1.2000	0.0990	6.6212	25.5000	0.0990	8.0135
29	0.0000	28.6702	-0.5657	25.5000	28.7692	7.4478
29	2.3000	0.0000	0.3595	27.8000	28.7692	7.8073
29	6.2000	0.0000	0.6753	34.0000	28.7692	8.4826
29	0.5000	2.7657	1.4638	34.5000	31.5349	9.9464
29	3.5000	7.0210	0.1182	38.0000	38.5559	10.0646
29	4.0000	0.0000	0.9898	42.0000	38.5559	11.0544
29	3.3000	4.0555	0.9685	45.3000	42.6114	12.0229
29	0.0000	4.1267	0.8475	45.3000	46.7381	12.8704
29	0.0000	3.1509	0.1626	45.3000	49.8890	13.0330
30	0.0000	0.0000	0.2719	45.3000	49.8890	13.3049
30	0.0000	0.0000	0.2537	45.3000	49.8890	13.5586
30	0.0000	0.0052	0.2562	45.3000	49.8942	13.8148
30	0.0000	0.0000	0.2553	45.3000	49.8942	14.0701
30	0.0000	0.0000	0.2550	45.3000	49.8942	14.3251
30	0.0000	0.0000	0.2547	45.3000	49.8942	14.5798
30	0.0000	0.0000	0.2543	45.3000	49.8942	14.8341
days	**K(obs)**	**K(M)**	**K(A)**	**K(cum)**	**K(Mcum)**	**K(Acum)**
30	0.0000	0.0000	0.2539	45.3000	49.8942	15.0880

30	0.0000	0.0000	0.2535	45.3000	49.8942	15.3415
30	0.0000	0.0000	0.2532	45.3000	49.8942	15.5947
30	0.0000	0.0000	0.2528	45.3000	49.8942	15.8475
30	0.0000	0.0000	0.2524	45.3000	49.8942	16.0999
30	0.0000	0.0000	0.2521	45.3000	49.8942	16.3520
30	0.0000	0.0000	0.2517	45.3000	49.8942	16.6037
30	0.0000	0.0000	0.2513	45.3000	49.8942	16.8550
30	1.9000	0.0000	0.2510	47.2000	49.8942	17.1060
30	4.6000	0.0000	0.6584	51.8000	49.8942	17.7644
30	0.0000	2.2990	1.1846	51.8000	52.1932	18.9490
30	0.0000	5.1891	0.1052	51.8000	57.3823	19.0542
30	0.0000	0.0000	0.2826	51.8000	57.3823	19.3368
30	0.0000	0.0049	0.2525	51.8000	57.3872	19.5893
30	0.0000	0.0000	0.2571	51.8000	57.3872	19.8464
30	0.0000	0.0000	0.2559	51.8000	57.3872	20.1023
30	0.0000	0.0000	0.2556	51.8000	57.3872	20.3579
1	0.0000	0.0000	0.2553	51.8000	57.3872	20.6132
1	0.0000	0.0000	0.2549	51.8000	57.3872	20.8681
1	0.0000	0.0000	0.2545	51.8000	57.3872	21.1226
1	0.0000	0.0000	0.2542	51.8000	57.3872	21.3768
1	0.5000	0.0000	0.2538	52.3000	57.3872	21.6306
1	0.0000	0.0000	0.3613	52.3000	57.3872	21.9919
1	0.0000	0.6028	0.2360	52.3000	57.9900	22.2279
1	1.5000	0.0000	0.2563	53.8000	57.9900	22.4842
1	0.0000	0.0049	0.5757	53.8000	57.9949	23.0599
1	0.0000	1.8226	0.2013	53.8000	59.8175	23.2612
1	0.0000	0.0000	0.2630	53.8000	59.8175	23.5242
1	0.0000	0.0000	0.2523	53.8000	59.8175	23.7765
1	0.0000	0.0000	0.2536	53.8000	59.8175	24.0301
1	0.0000	0.0000	0.2530	53.8000	59.8175	24.2831
1	0.0000	0.0000	0.2527	53.8000	59.8175	24.5358
1	0.0000	0.0000	0.2523	53.8000	59.8175	24.7881
1	0.0000	0.0000	0.2520	53.8000	59.8175	25.0401
1	0.0000	0.0000	0.2516	53.8000	59.8175	25.2917
1	0.0000	0.0000	0.2513	53.8000	59.8175	25.5430

An overall evaluation of the forecast performance of both the MARIMA and ARMA models are summarized in Table 4.28 for station Empangan Genting Kelang with station Km.11 Gombak and station Kampung Kuala Saleh. Tables 4.29 and 4.30 tabulate the forecast performance for station Km.11 Gombak and station Kampung Kuala Saleh respectively.

Table 4.28: Performance measure of the forecast for station Empangan Genting Kelang.

Statistic	**MARIMA**		**ARMA (1,1)**
	With station Km.11 Gombak	**With station Kampung Kuala Saleh**	
μ_{ε_t}, (mm)	3.3319	3.38778	3.1745783
$\text{RMSE}(\varepsilon_t)$,(mm)	12.5642	12.902865	12.338675

Table 4.29: Performance measure of the forecast for station Km.11 Gombak.

Statistic	**MARIMA**	**ARMA(1,1)**
μ_{ε_t}, (mm)	1.7068	1.1587567
$\text{RMSE}(\varepsilon_t)$,(mm)	5.2814	3.2122195

Table 4.30: Performance measure of the forecast for station Kampung Kuala Saleh.

Statistic	**MARIMA**	**ARMA(1,1)**
μ_{ε_t}, (mm)	1.6469917	1.1269467
$\text{RMSE}(\varepsilon_t)$,(mm)	5.1342631	3.4076801

From Table 4.28, it can be concluded that the MARIMA models was outperformed by the ARMA (1,1) models for station Empangan Genting Kelang. The average value of the residuals (error) of hourly forecasts, μ_{ε_t}, for the data from this station are slightly better by using the ARMA(1,1) models compared to the MARIMA models in both study areas (with station Km.11 Gombak and station Kampung Kuala Sleh) where the differences are approximately equal to one. The root mean square error (RMSE) for the ARMA models were also slightly smaller compared to the MARIMA model. However, if we compared the both study areas that using the MARIMA model, we can see that study area one, that is the jointly modeled with station Km.11 Gombak, were better than the other one. From this, its can conclude that the higher correlated stations produced a better forecast results compared to the lower correlated stations.

It goes the same too for station Km.11 Gombak and station Kampung Kuala Saleh, where from Tables 4.29 and 4.30 it can be concluded that the ARMA models performed better than the MARIMA models. However, the root mean square error were much more better or much more less in both study area by using the ARMA(1,1) models compared to the MARIMA models.

4.9 Forecast Error Normality Check

Finally, the forecast error for both study area that consists of three stations which are station Empangan Genting Kelang, station Km.11 Gombak and station Kampung Kuala Saleh had been analyze to check whether those errors were normally distributed or does not normally distributed. This checking were to ensure whether the MARIMA models were suitable or not to be used to forecast the rainfalls intensity.

By using the Minitab 14 software, those forecast errors had been analyzed separately for each study areas and for each station. Figures 4.68, 4.69, 4.70 and 4.71 show the normal probability plot for all the stations based on the Anderson-Darling test of normality.

By the Anderson-Darling test of normality, we just need to check whether the P-Value is more than 0.05 that ensure the forecast error is normally distributed or not. From the results of the normality check, we could see that the P-Value for all the stations from Figures 4.68, 4.69, 4.70 and 4.71 were less than 0.05. Therefore it suggests that the MARIMA models were not suitable to be used to forecast the rainfall intensity.

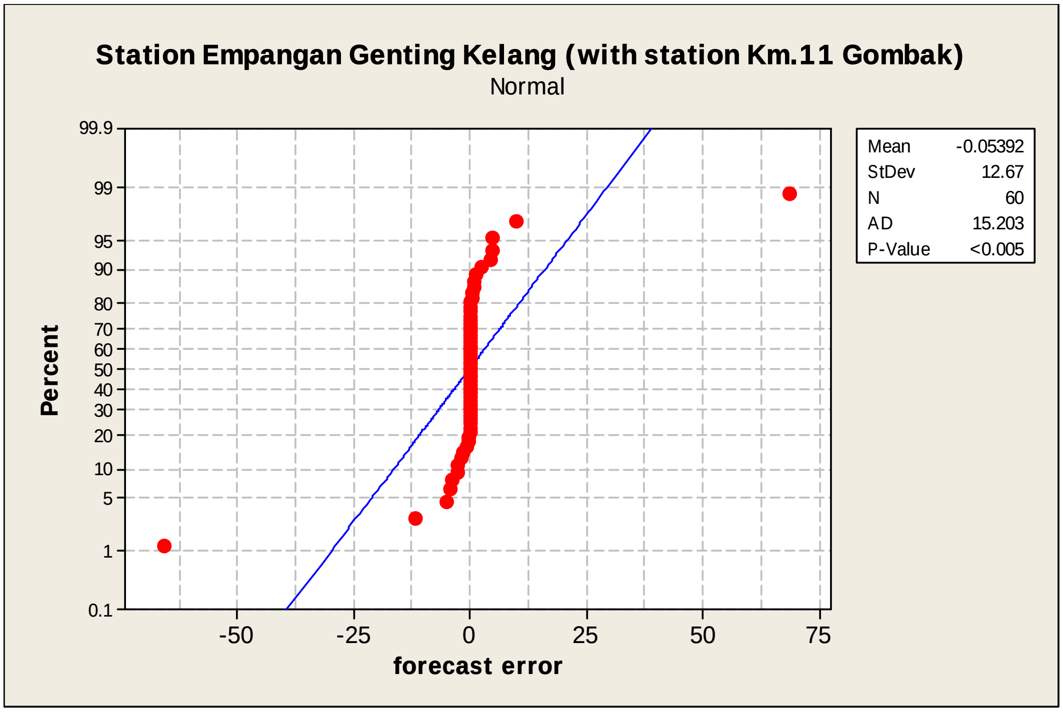


Figure 4.68: Normal probability plot of the forecast errors for station Empangan Genting Kelang (with station Km.11 Gombak) using the MARIMA models.

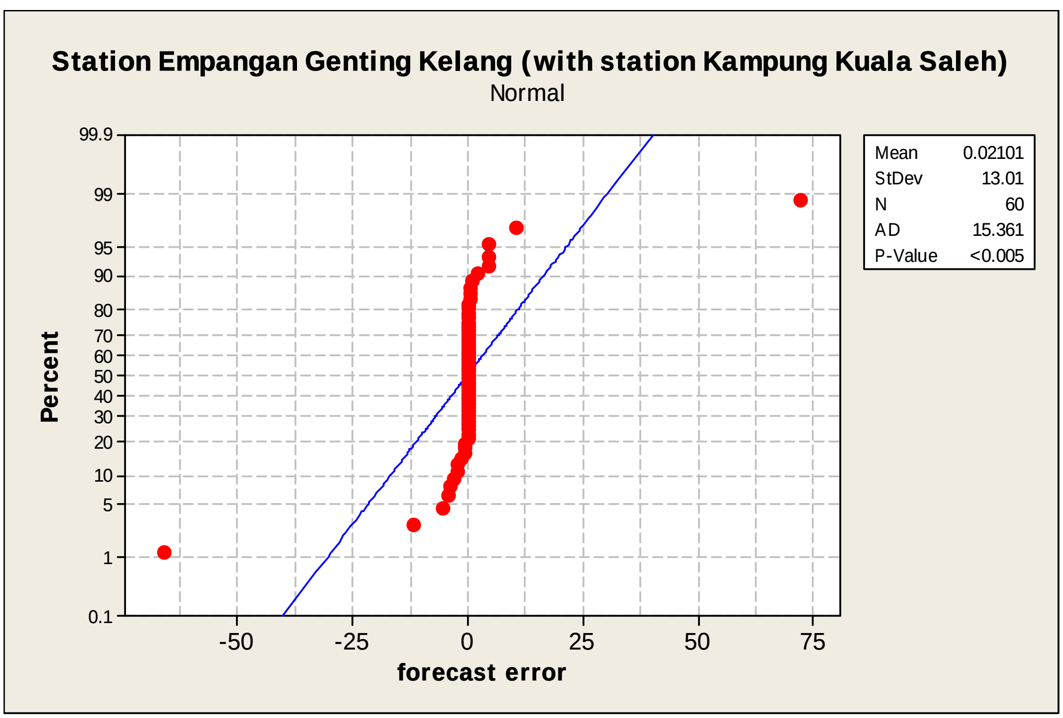


Figure 4.69: Normal probability plot of the forecast errors for station Empangan Genting Kelang (with station Kapung Kuala Saleh) using the MARIMA models.

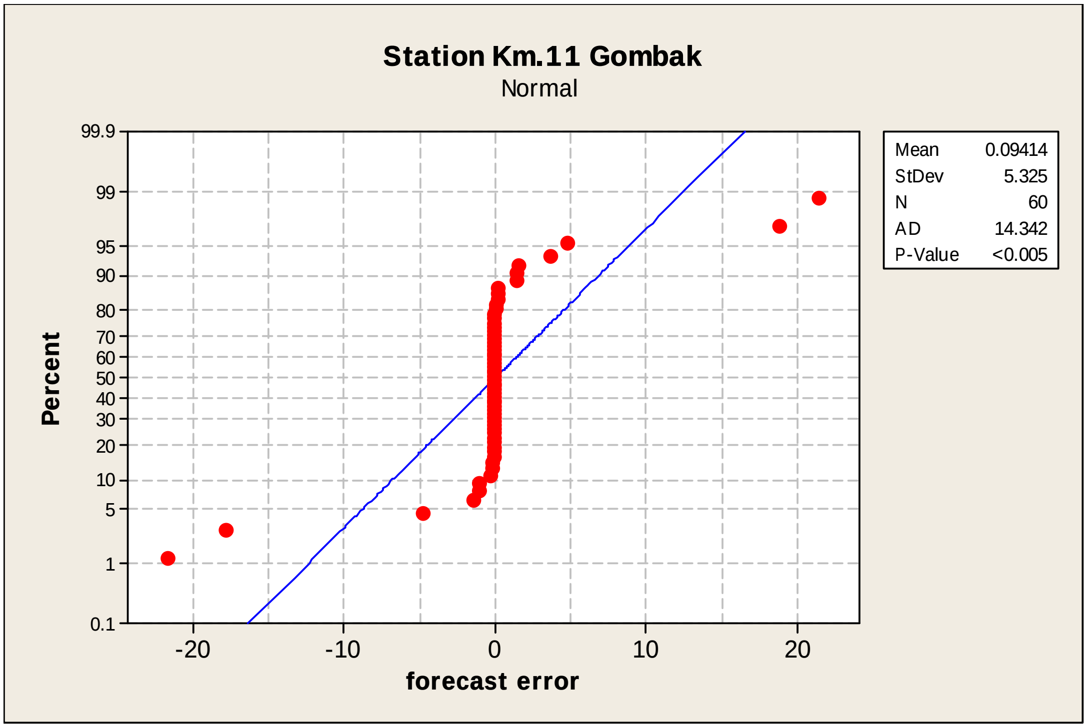


Figure 4.70: Normal probability plot of the forecast errors for station Km.11 Gombak using the MARIMA models.

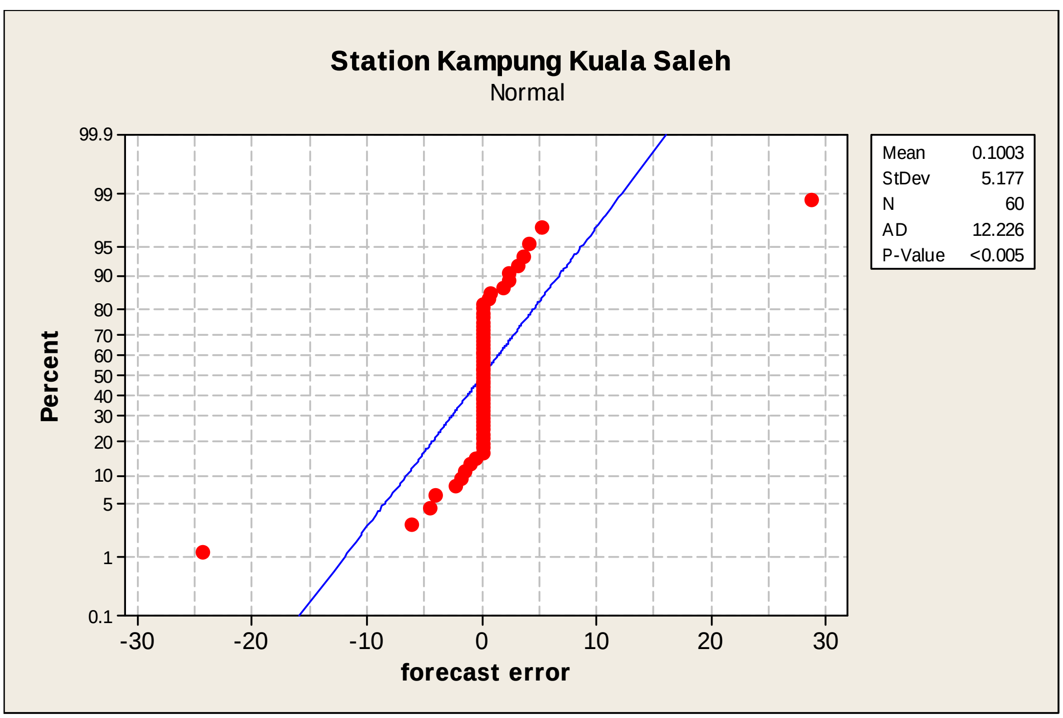


Figure 4.71: Normal probability plot of the forecast errors for station Kampung Kuala Saleh using the MARIMA mode

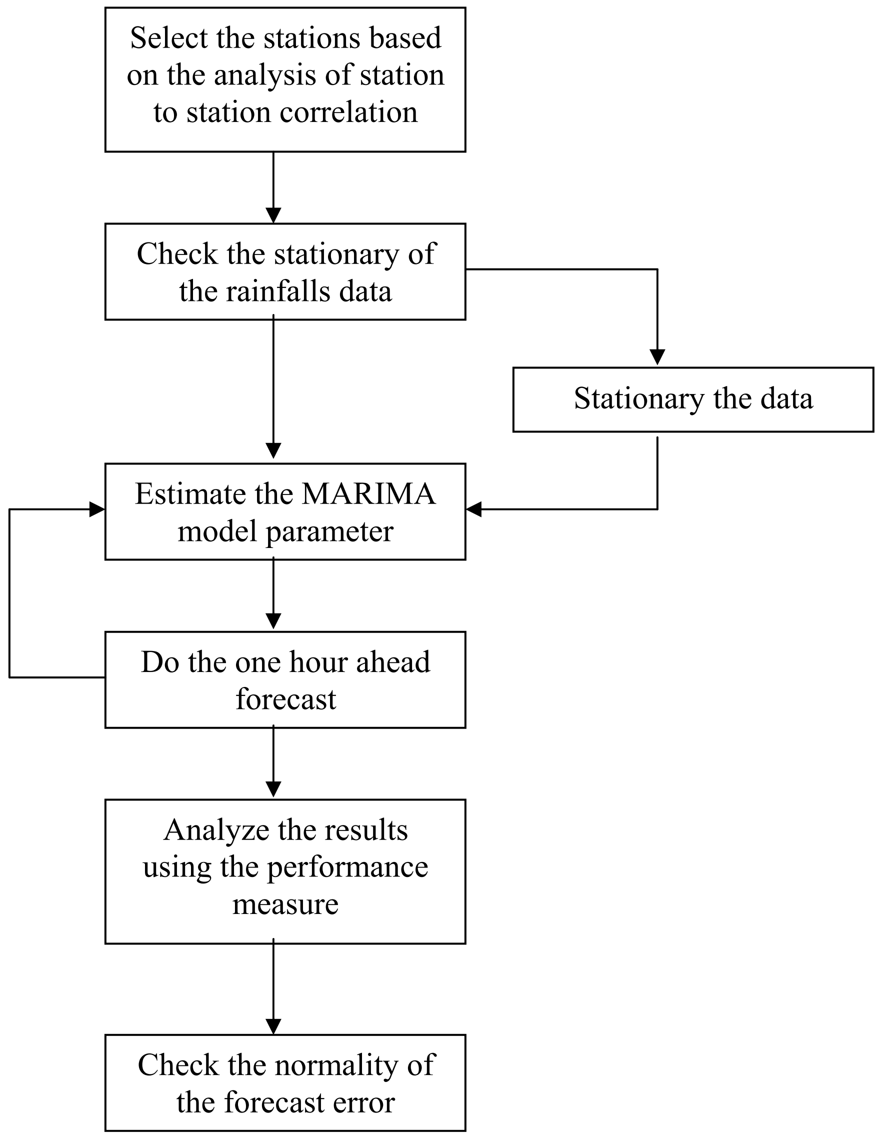


Figure 4.72: Flowchart of the methodology

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

1. Characteristics of Convective Rain Based on Short Rainfall Duration Data

The diurnal and monthly distribution of rainfall (greater than 5mm) in 2004 at Station 3117070 was discussed in Chapter IV. The results show that the bulk of the rains fall in the afternoon, between 13:00 and 19:00 which makes up about 75 % of the total rainfall. This type of storm can be classified as convectional storms. Convective storms are predominant and are an active component of the tropical weather system. A Minimum Interevent Time (MIT) of 3 hours was used to separate storm events. Convective rain occurred most frequently in November and the highest frequency of convective storm happened in the intermonsoon months which made up about 44% was of the storms. This is due to light variable winds and an unstable atmosphere which favor strong convective activity. This results in thunderstorms and heavy rains especially in the late afternoons and early evenings. Over five years, the highest intensity was 384 mm/hr occurred in 2003. These characteristics were discussed in Chapter IV where a great variety of storm shape is evident and the patterns show that most of the convective events occurred over short durations, ranging from 15 to 90 minutes.

2. Classification of Convective Events

A classification of episodes based on β parameter was discussed in Chapter IV. This classification is according to their greater or lesser convective character (Llasat, 2001). The classification of the convective storm into slightly, moderately and strongly convective indicates that the highest proportion is for the moderately convective class, which makes up 63.8% of the total convective events. It seems that a 35 mm/hr threshold intensity is appropriate for separating convection from non convective storms for local conditions. However, this analysis needs to be replicated to cover more rainfall stations.

3. Comparison of Spatial Distribution of Convective Rainfall between Radar and Ground Rainfall

Comparison of spatial distribution between radar and surface rainfall were examined in terms of intensity, areal coverage, storm movements and depth-area relationship. The intensity values between raingauge and radar show large differences. The main difficully in determining the *Z-R* (with Z in mm^6/m^3 and R in mm/hr) relationship arises from the fact that radar measures precipitation in the atmosphere while gages measure it at the ground. In addition, precipitation may evaporate before reaching the ground, especially in the tropics. Winds may also carry precipitation away from beneath the producing cloud.

As for the storm intensity, out of four storms, only one showed reasonably good match in the contour patterns between radar and raingauge. This might be due to inadequate number of raingauge and missing data which limit the ability of Kriging methods.

The aerial rainfall for each interval of isohyets between radar and surface rainfall was compared using GIS software. The ground rainfall data produced remarkably different areal rainfall for various intervals of isohyets. Overall, the areas derived from raingauge are bigger than those derived from radar.

Each storm is unique in term of the movement of the storm cell. Some have long paths while others are circling within a limited path.

Depth-area relationships of six storms were examined. Each storm display quite different areal reduction curve. However, in general rainfall the depth decreases with increasing catchment area. The ARF curve was compared with the ARFs from other areas. The present study introduced quite similar ARF values obtained by Yan and Lin, (1986). The ARF values derived from smaller areas were different from this study. Therefore, the shapes of such curves can only be compared if the temporal and spatial resolutions of the measurements are similar. However, the agreement between the relationships derived for convective storms cells in Klang Valley and the entire Peninsular Malaysia (Yan and Lin, 1986) can be explained in term of similarity in the climatic condition.

4. From four candidate distribution functions for hourly rainfall amount (exponential, gamma, Weibull, and mixed-exponential), the mixed-exponential was found to be the best model based on the numerical (goodness-of-fit tests) and graphical comparisons. This distribution function is expected to explain well both the small and large amounts of hourly rainfall amounts.

5. The present study has proposed a new NSRP model that used the mixed-exponential distribution for describing the rain cell intensities. Results of the calibration and validation of the proposed model have indicated its superior performance in preserving more accurately the statistical and physical properties of the underlying observed hourly rainfall series as compared to the traditional NSRP model using the exponential for rain cell intensity distribution

6. In the calibration of NSRP models, it has been shown that the use of the transition probabilities of rainfall occurrences rather than the autocorrelations of rainfall amounts can provide more accurate description of the observed rainfall properties. In particular, the modified NSRP model (MEXPTRAN) with mixed-exponential distribution to describe the rain cell intensities and using transition probabilities in the fitting procedures was found to be the best model in terms of its accuracy in preserving the statistical and physical properties of the observed rainfall series.

7. In consideration of rainfall characteristics over different timescales, it was found that the NSRP (MEXPTRAN) model can describe very well many rainfall statistical and physical properties for both one-hour and 24-hour scales. In addition, the model was able to preserve accurately some relevant rainfall physical properties such as the probability of dry days and the daily transition probabilities of rainfall occurrences for the whole year.

8. The first-order two-state Markov Chain (MC) model was found to describe accurately the hourly and daily rainfall occurrence processes.

9. The Fourier series was found to be able to describe accurately the seasonality of the MCME model parameters for hourly and daily rainfall series, especially the transition probabilities. It was also found that the Fourier series fit for the hourly parameters are better than the daily parameters.

10. The MCME hourly model was found to be able to describe adequately the statistical and physical properties of the rainfall process at the hourly scale. However, when the hourly rainfall series were lumped to daily (24-hour) or monthly series the hourly MCME model produced larger errors than those given by the daily MCME model.

11. The MCME daily model was also found to be able to describe adequately the statistical and physical properties of the underlying daily rainfall process. However, when the generated daily series were lumped to monthly series, the daily model could only preserve the monthly rainfall mean, but could not describe well other rainfall properties. Nevertheless, the daily MCME model produced smaller errors than the hourly model in preserving the rainfall properties at the monthly scale.

12. The comparison between the NSRP and the hourly MCME model has shown that both models have comparable performance in preserving the properties of the observed at the hourly scale. But when the generated series were lumped to daily (24-hour) sequences, the NSRP was found to perform better than the MCME in describing the daily rainfall properties. However, the MCME daily model was found to produce smaller errors than the NSRP in describing the rainfall properties at the daily scale. Therefore, the NSRP model has the ability to describe the underlying rainfall processes at both hourly and daily scales. The MCME models, on the other hand, could only preserve the properties of the observed when their parameters were estimated using data at the same scale as the observed data.

13. In general, both NSRP and MCME models were found to have the same predictive ability. While the models did not perform as well as in the calibration period, both were able to preserve the seasonal trend of the observed rainfall properties. The predictive ability of the MCME daily model was found to be better than the predictive ability of the NSRP and MCME hourly model in describing the daily rainfall process. The hourly MCME model is better than the NSRP in preserving the hourly rainfall series but when lumped to daily equivalent, the NSRP was better than the MCME in preserving the properties at the daily scale.

14. By using MARIMA (1,1,0), the model can be written as

$$\boldsymbol{Y}_t = \boldsymbol{\alpha} \boldsymbol{Y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{5.1}$$

where

$$\begin{aligned} \boldsymbol{Y}_t &= (\boldsymbol{I} - \boldsymbol{B})\boldsymbol{X}_t \\ &= \boldsymbol{I}\boldsymbol{X}_t - \boldsymbol{B}\boldsymbol{X}_t \\ &= \boldsymbol{X}_t - \boldsymbol{X}_{t-1} \end{aligned}$$

Therefore,

$$\begin{aligned} \boldsymbol{X}_t - \boldsymbol{X}_{t-1} &= \boldsymbol{\alpha}(\boldsymbol{X}_{t-1} - \boldsymbol{X}_{t-2}) + \boldsymbol{\varepsilon}_t \\ \boldsymbol{X}_t &= (\boldsymbol{I} + \boldsymbol{\alpha})\boldsymbol{X}_{t-1} - \boldsymbol{\alpha}\boldsymbol{X}_{t-2} + \boldsymbol{\varepsilon}_t \end{aligned} \tag{5.2}$$

where $\boldsymbol{\varepsilon}_t$ is assumed to be a white noise. In this study, Lembah Kelang was selected as the study area. We specifically forecasted rainfall intensity data for two study cases which were station Empangan Genting Kelang with station Km.11 Gombak and Empangan Genting Kelang with station Kampung Kuala Sleh.

For comparison purposes, the univariate ARMA model was also employed to forecast rainfall intensity in the above study area. The Box-Jenkins model, ARMA (1,1) model used can be written as

$$X_t = \phi X_{t-1} - \theta\varepsilon_{t-1} + \varepsilon_t \tag{5.3}$$

For these study areas, the root mean square deviation (RMSD), which is a measure of the difference between values predicted by a model and the observed value, and the average value of the residuals (error) of hourly forecasts, μ_{ε_t}, were calculated. Based on these values, it was concluded that for all the selected stations, the MARIMA model have been outperformed by the ARMA (1,1) models where for station Empangan Genting Kelang, the differences only small compared to station Km.11 Gombak and station Kampung Kuala Saleh. However, since the differences were small, the MARIMA models forecasts could be considered as good as the ARMA models.

From the value of the performance measure for station Empangan Genting Kelang that have been jointly modeled with station Km.11 Gombak and with Kampung Kuala Saleh, it proved that a more highly correlated stations could gave a better forecast results when it is been jointly modeled by using the MARIMA models.

15. Looking at the data set, it was possible that the forecasts using the MARIMA models were poorer than the ARMA models maybe due to the 7th data, which was at 2.00 pm, on 29th April 2002, where the rainfall intensity was not normal. The value of the data was very much different from the other stations where the value was 66 mm/h for station Empangan Genting Kelang which can be considered as an outlier. This may caused some interruptions because in using MARIMA, the two stations were jointly modeled. Since the other two stations were lowly correlated, these two stations could not be modeled together. Furthermore, the lack of technologies in Malaysian Meteorological Department in providing radar maps for the storm movements also contributed to this poor forecast results.

16. In general, the MARIMA model is a potential method for forecasting hourly rainfall intensity. Instead of using many variables such as the humidity, temperature and the direction of the wind in the model, several rainfall data series from several stations can be used. This simplifies the process of forecasting rainfalls. Since rain can be forecasted, the results of the current study can help the relevant authorities in manning and preventing possible hazards caused by rains.

5.2 Recommendations for Future Works

1. The first part of the research focused on two major aspects: 1) characterization of convective rain and 2) spatial variation of convective rainfall derived from radar data and surface data. Prior to this study, the approach used to characterize and compare spatial variations between radar and surface rainfall data has not been tested in the tropics. In order to improve future studies, the following research areas are suggested:

a) This study used one station to characterize convective rain. Future studies shall use more rainfall stations to examine the spatial consistently of the characteristics.

b) The number of rainfall stations need to be increased to give a better interpolation in Kriging Method. This is because kriging works best when the input point is large and vice versa when the number of point is small.

c) The influence of wind direction and wind velocity need to be checked in evaluating the storm movement.

d) The difficulties to interpret radar rainfall intensity from JPEG file need to be checked to prevent overestimate or underestimate of rainfall intensity values. This is might be solved by doing a programming to interpret the coding output from radar software or execute a projection using GIS method after get the z coordinate value.

2. Meanwhile, several recommendations may be suggested for improving the modelling of the NSRP and the MCME, such as:

a) This study used one station to characterize convective rain. Future studies can use more rainfall stations to examine weather convective rainfalls do vary spatially.

b) The number of rainfall stations need to be increased to give a better interpolation in Kriging Method. This is because kriging works best when the input point is large and vice versa when the number of point is small.

c) The influence of wind direction and wind velocity need to be checked in evaluating the storm movement.

d) The NSRP and the MCME models could be used in the study of the impacts of climate change on rainfall processes if it is feasible to develop some linkages to link the model parameters with climate variable.

e) The NSRP and the MCME models could be generalized for stochastic simulation of rainfall processes for many sites simultaneously.

f) The NSRP and MCME models could be modified to describe more accurately the extreme rainfall characteristics at any given location. This would require the use of other heavy-tailed distributions to represent the rainfall amounts in MCME model or the rain cell intensities in NSRP model.

g) In estimating parameters for the NSRP with mixed-exponential distribution, there were seven parameters to be estimated for each month independently. This task may be tedious. Hence, the Fourier series may be used to reduce the number of estimated parameters.

3. For the the third part, which involves the development of the short-term foreasting technique of convective rains, several recommendations may be suggested for producing better forecast results, such as:

a) Only two stations can be jointly modeled in this MARIMA model because of the problem of parameters estimation. It is suggested that this problem is overcome so that more than two stations can be jointly modeled in future studies. It is also suggested that future studies use different variables such as the humidity, temperature and altitude in the MARIMA model to forecast the rainfalls.

b) This study may be applicable in a wide range of situations. It is therefore suggested that this study should be replicated in other types of forecasts and in other types of industries so as to determine the potential of the MARIMA models in those situations and industries. As examples:

(i) Different types of forecasts as the subject of the investigation such as demand for materials, cash flow and inventory levels.
(ii) Different types of industries as the subject of investigation such as chemical industry, textile product industry and wood industry.

c) Analyzing multivariate data is a very tedious work. The potential use of statistical software such as SAS, S-PLUS and MINITAB in analyzing the time series multivariate data should be investigated.

d) In this study, only a one-hour ahead prediction was produced. A longer term prediction can give better information to predict a flash flood. Thus it is suggested that a longer term prediction be made in future study

REFERENCES

Adams, B.J., Fraser, H.G., Howard, C.D.D. and Hanafy, M.S. (1986). "Meteorological Data Analysis for Urban Drainage System Design". *Journal of Environmental Engineering*. **112**(5): 827-848.

Akaike, H. (1974). "A New Look at the Statistical Model Identification". *IEEE Transactions on Automatic Control*. **AC-19**: 716-723.

Amorocho, J. and Wu,B. (1977). "Mathematical models for simulation of cyclonic storm sequences and precipitation fields". *Journal of Hydrology*. **32**:329-345.

Austin, P.M.& Houze, R.A. 1972. "Analysis of the structure of precipitation patterns in New England". *J.appl. Meteorology*. **11**:926-935.

Bacchi, B. and Borga, M. (1994). "Spatial Correlation Patterns and Rainfall Field Analysis". *Excerpta*. **7**: 7-40

Bedient, P.B. and Huber, W.C. (2002). *Hydrology and Floodplain Analysis*. 3rd ed. New Jersey: Prentice Hall.

Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis and Control*. Holden Day.

Box, G.E.P., Jenkins, G.M. and Reinsel, G. C. (1994). *Time Series Analysis, Forecasting and Control*. Prentice-Hall, Inc.

Box, G.E.P. and Tiao, G.C. (1977). "A Canonical Analysis of Multiple Time Series. *Biometrika"*. **64**: 355-365.

Bras, R., and I.Rodriguez-Iturbe (1976). "Rainfall Generation: A non-stationary time varying multidimensional model". *Water Resources Research*. **12(3):** 450-456.

Bras, R. L., and Rodriguez-Iturbe, J. (1985). "Random functions and hyrology". Addison-Wesley Publishing Co. Reading, Massachusetts

Brath, A., Burlando, P. and Rosso, R. (1988). "Sensitivity analysis of real-time flood forecasting to on-line rainfall prediction". *Selected Papers from Workshop on Natural Disasters in European Mediterranean Countries, Perugia, Italy.* pp. 469-488.

Brillinger, D.R. (1975). *Time Series: Data Analysis and Theory*. Holt, Rhinehart and Winston.

Brooks, H.E., C.A Doswell III, and R.A. Maddox (1992). "On the use of mesosscale and cloud-scale models in operational forecasting". *Wea. Forecasting*. **7**: 120-132

Buishand, T.A. (1977). "Some remarks on the use of daily rainfall models". *Journal of Hydrology.* **36**:295-308.

Buizza, R. (2001). "Accuracy and potential economic value of categorical and probabilistic forecasts of discrete events". *Monthly Weather Review*. **129:** 2329-2345.

Burgueno, A., Austin, J., Vilar, E. and Puigcerver, M. (1987). *Analysis of moderate and intense rainfall rates continuously recorded over half a century and influence on microwave communication planning and rain rate data acquisition*, IEEE Trans. Comm. COM-35, **4:** 382-395

Burgueno, A., Puigcerver, M. and Vilar, E. and 1988. "Influence of the rain gauge integration time on the rain rate statistics used in microwave communications". *Ann Tele'comm*., **43**: 522-527

Burlando, P. and Rosso, R. (1991). "Comment on parameter estimation and sensitivity analysis for the modified Bartlett-Lewis Rectangular Pulses model of rainfall". *J. Geophys. Res*. **96**(D5):9391-9395.

Burlando, P. and Rosso, R. (1996). "Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation". *Journal of Hydrology*. **187**:45-64.

Burlando, P., Montanari, A., and Ranzi, R. (1996). "Forecasting of storm rainfall by combined use of radar, rain gages and linear models." *Atmospheric Research***. 42:** 199-216.

Burlando, P., Rosso, R., Cadavid, L. and Salas, J.D. (1993). "Forecasting of short-term rainfall using ARMA models. *Journal of Hydrology***. 144**: 193-211.

Callado, A. and Pascual, R.(2002). "Storms in front of the mouth rivers in north-eastern coast of the Iberian Peninsula". *4th plinius Conference on Mediterranean Storms, Pollentia (Spain),* 2-4 October 2002

Calenda, G. and Napolitano, F. (1999). “Parameter estimation of Neyman-Scott processes for temporal point rainfall simulation”. *Journal of Hydrology*. **225**:45-66.

Chang, I., Tiao, G.C. and Chen, C. (1988). “Estimation of Time Series Parameters in the Presence of Outliers”. *Technometrics*. **30**: 193-204.

Chatfield, C. (1997). “Forecasting in the 1990s”. *The Statistician*. **46:** 461-473.

Charles A. Doswell III, (1993). “Flash Flood-Producing Convective Storms: Current Understanding and Research”. *Report of the Proceedings of the U.S.-Spain Workshop on Natural Hazards (*Barcelona, Spain), Eds., pp. 97 – 107.

Charles L. Hogue (2007). “Convection and Wind”. Retrieved on November 8, 2007, from http:// costa-rica-uide.com/travel/index.php?option=com_content&task =view&id=365&Itemid=604&limit=1&limitstart=2

Cheng-Lung Chen, M. (1983). “Rainfall intensity-duration-frequency”. *Journal of Hydraul Engineering,* **109**: 1603-1621

Chaudry, F. H., Andrade Filho, A. G. And Cavalheiros, R. V. (1994). “Statistic on Tropical Convective Storms Observed by Radar. In: Niemczyniwicz”. *J. (Ed.) Atmospheric Research, Special Issue: Closing the Gap Between Theory and Practice in Urban Rainfall Applications.* **42:** 1-4, 217-227.

Chin, Edwin H.(1977). “Modeling daily precipitation occurrence process with Markov Chain”. *Water Resources Research*. **13**:949-956.

Choi, B. (1992). *ARMA Model Identification*. Springer-Verlag.

Chow, V.T. (1964). *Handbook of Applied Hydrology: A Compendium of Water-Resources Technology*. McGraw-Hill.

Coe, R. and Stern, R.D. (1982). “Fitting models to daily rainfall data”. *Journal of Applied Meteorology.* **21**:1024-1031.

Cowpertwait, P.S.P. (1991). “Further developments of the Neyman-Scott clustered point process for modeling rainfall”. *Water Resources Research*. **27**:1431-1438.

Cowpertwait, P.S.P. (1992). “Correction to Further developments of the Neyman-Scott clustered point process for modeling rainfall”. *Water Resources Research*. **28**:1497.

Cowpertwait, P.S.P. (1994). “A generalized point process model for rainfall”. *Proc. Roy. Soc. London Ser. A*. **447**(1929):23-37.

Cowpertwait, P.S.P. (1995). “A generalized spatial-temporal model of rainfall based on clustered point process”. *Proc. Roy. Soc. London Ser. A*.**450**(1929):163-175.

Cowpertwait, P.S.P.,O'Connel,P.E.,Metcalfe, A.V., Mawdsley, J.A.(1996). "Stochastic point process modeling of rainfall: 1. Single-site fitting and validation". *Journal of Hyrology*. **175**:17-46.

Cowpertwait, P.S.P.,O'Connell,P.E.,Metcalfe, A.V., Mawdsley, J.A.(1996). "Stochastic point process modeling of rainfall: 1Regionalization and disaggregation". *Journal of Hyrology*. **175**:47-65.

Cowpertwait, P.S.P. and O'Connell, P.E.,(1997). "A regionalized Neyman-Scott model of rainfall with convective and stratiform cells". *Hydrol. Earth Syst. Sci*. **1**:71-80.

Cowpertwait, P.S.P. (1998). "A Poisson-cluster model of rainfall: High-order moments and extreme values". *Proc. Roy. Soc. London Ser. A*. **454**(1929):885-898.

Cowpertwait, P.S.P., Kilsby, C.G., O'Connell, P.E., (2002). "A space-time Neyman-Scott model of rainfall: Empirical analysis of extremes". *Water Resources Research*. **38**:1-14.

Cowpertwait, P.S.P. (2004). "Mixed rectangular pulses models of rainfall". *Hydrol. Earth Syst. Sci.* **8**(5):993-1000.

Cox, D.R. and Isham, V. (1980). "Point Processes". Chapman Hall London.

Cox, D.R. and Isham, V. (1994). "Chapter I: Stochastic models of precipitation" in Barnett, V. and Turkman, K. S. "Statistics for the Environment 2: Water Related Issues". John Wiley &Sons New York. 1-17.

David C. Curtis (1999). "Comparing Spatial Distribution of Rainfall Derived from Rain Gages and Radar". *Journal of Floodplain Management.* Brett S. Clyde NEXRAIN Corporation, Folsom, CA

Delleur, J.W., Tao, P. and Kavvas, M.L. (1976). "An evaluation of the practicality and complexity of some rainfall and runoff series models". *Water Resources Research*. **12**: 953-970.

Desa, M. N, Munira Akhmal H. and Kamsiah A.W. (2005). "Capturing Extreme Rainfall Events in Kerayong Catchment". 10th International Conference on Urban Drainage, Copenhagen, Denmark.

Dille, J. A., Milner, M., Groeteke, J. J., Mortensen, D. A., and Williams, M. M. (2002). "How good is your weed map? A comparison of spatial interpolators". *Weed Sci.* ***51**:44-55.*

Donald, W. W. (1994). "Geostatistics for mapping weeds, with a Canada thistle (*Cirsium arvense*) patch as a case study". *Weed Sci*. **42**:648-657

Dong Kyou Lee and Hyung Woo Kim (2000). *An Observational Study of Heavy Rainfall with Mesoscale Convective Systems over the Korean Peninsular.* Unpublished note. Atmospheric Science Program, Seoul National University.

Doswell, C. A., III, H. Brooks, and R. Maddox (1996). "Flash flood forecasting: An ingredient-based methodology". *Weather. Forecasting*, **11:** 560-581.

Duchon J. (1976) "Splines minimizing rotation invariant seminorms in sobolev spaces, constructive theory of functions of several variables". **1:** 85-100

Dutton, E. J. and Dougherty, H. T. (1979). "Year to year variability of rainfall for microwave applications in the USA". *IEEE Trans. Comm*. COM-27, **5:** 829-832

Duan,Q., Sorooshian,S. and Gupta, V. (1992). "Effective and efficient global Optimization for conceptual rainfall-runoff models". *Water Resources Research*. **28**(4):1015-1031.

Duan,Q., Sorooshian,S. and Gupta, V. (1994). "Optimal use of the SCE-UA global optimization method for calibrating watershed models". *Water Resources Research*. **158**(1994): 265-284.

Engle, R.F. and Granger, C.W.J. (1987). "Co-Integration and Error Correction: Representation, Estimation and Testing". *Econometrica*. **55**: 251-276

Elsner, J.B. and Tsonis, A.A. (1993). "Complexity and Predictability of Hourly Precipitation". *Journal of the Atmospheric Sciences*. **50**(3):400-405.

Entekhabi, D., Rodriguez-Iturbe, I., and Eagleson, P.S. (1989). "Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: Parameter estimation and validation". *Water Resources Research*. **25**:295-302.

Fankhauser, J. C., (1988). "Estimates of thunderstorm precipitation efficiency from field measurements in CCOPE". *Monthly Weather Review*, **116.** 663-684

Fayerherm, A.M., and L.D. Bark (1965). "Statistical Methods for persistent precipitation pattern". *Journal of Applied Meteorology*. **4**:320-328.

Favre, A.-C.(2001). "Single and multi-site modeling of rainfall based on the Neyman-Scott process". Swiss Fed.Inst. of Technol., Lausanne. Ph.D. thesis,

Favre, A.-C., Musy,A., Morgenthaler, S. (2002). "Two-site modeling of rainfall based on the Neyman-Scott process". *Water Resources Research*. **38**(21):43.1-43.7.

Favre, A.-C, Musy, A., Morgenthaler,S.(2004). "Unbiased parameter estimation of Neyman-Scott model for rainfall simulation with related confidence interval". *Journal of Hydrology*. **286**:168-178.

Frank, C., Garg, A., Raheja, A., Sztandera, L. (2003). ”Forecasting women’s apparel sales using mathematical modelling”. *International Journal of Clothing*. **15:** 107-125.

Follansbee, W. A.(1973). “Estimation of Daily Precipitaion over China and the USSR Using Satellite Imagery, NOAA Tech,.Memo”. NESS 81

Foufoula-Georgiou,E. and Guttorp, Peter.(1987). “Assessment of a class of Neyman-Scott Models for Temporal Rainfall”. *Journal of Geophysical Research*. **92**(D8):9679-9682.

Foufoula-Georgiou,E. and Krajewski(1995). “Recent advances in rainfall modelling, estimation, and forecasting”. *Reviews of Geophysics*. **Supplemen**t:1125-1137.

Gabriel,K.R., and J. Neumann (1962). “A Markov chain model for daily rainfall occurrence at Tel Aviv”. *Q.J.R. Meteorology. Soc*. **88**:90-95

Gates, P. and Tong, H. (1976). “On markov Chain Modeling to some weather data”. *Journal of Applied Meteorology*. **15**:1145-1151.

Georgakakos, K.P. and Krajewski, W.F. (1991). “Worth of radar data in the real-time prediction of mean areal rainfall by nonadvective physically based models”. *Water Resources Research.* **27**: 185-197.

Ghil, M., Benzi, R. and Parisi, G. (1985). *Turbulence and predictability in geophysical fluid dynamics and climate dynamics*. North Holland.

Ghilardi, P. and Rosso, R. (1990). Comment on “Chaos in Rainfall” by Rodriguez Iturbe et al. *Water Resources Research*. **26:** 1837-1839.

Gray, H.L., Kelley, G.D. and McIntire, D.D. (1978). “A New Approach to ARMA Modeling”. *Communications in Statistics*. **7:** 1-77.

Griffith, C. G., W. L. Woodley and P. C. Grube (1978). “Rain Estimation from Geosynchronous Satellite Imagery-Visible and Infrared Studies”. *Monthly Weather Review*. **106**(8):1153-1171

Guhathakurta, P. (2006). “Long-range monsoon rainfall prediction of 2005 for the districts and sub-division Kerala with artificial neural network.” *Current Science*. **90:** 773-779.

Gupta,V.K., and E.C. Waymire (1979). “A stochastic kinematic study of subsynoptic space-time rainfall”. *Water Resources Research*. **15**(3):637-644.

Gregory,J.M.,Wigley.T.M.L.,Jones,P.D. (1992). "Determining and Interpreting the order of a two-state Markov Chain: Application to Models of Daily Precipitation". *Water Resources Research.* **28**(5):1443-1446.

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.

Haan, C.T.,Johnson, H.P. and Brakensiek, D.L. (1982). "Hydrologic modelling of small watersheds". *American Society of Agricultural Engineers*.

Han, S-Y, (2001). "Stochastic Modeling of Rainfall processes: A Comparative Study using data from different climatic conditions". McGill University, Montreal, Quebec, Canada. Master thesis.

Hannan, E.J. (1980). "The Estimation of the Order of an ARMA Process". *The Annals of Statistics*. **8:** 1071-1081.

Hara, Yoshikane and Kimura, (2006). "Mechanism of Diurnal Cycle of Convective Activity Over Borneo Island". *2006 7th WRF User's Workshop,* 17-22 June 2006, Boulder , Colorado, United States

Hastenrath, S (1991). *Climate Dynamics of the Tropics. Updated Edition from Climate and Circulation of the Tropics.* Kluwer Academic Publishers.

Hewlett, J. D. (1969). *Principle at Forest Hydrology.* The University Georgia Press, 183pp.

Houze RA Jr (1993): *Cloud Dynamics*. Academic Press: New York

Hydroscience, Inc. (1979). *A Statistical Method for Assessment of Urban Stormwater Loads-Impacts-Controls*. Washington: United States Environmental Protection Agency.

Ichikawa, A. and Sakakibara, T. (1994). "Control measures for combined sewer overflows in Japan". *Proc. VII Conference Water Quality International '94, Budapest*.

Isaaks, E. H., and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, Oxford, UK.

Islam, S., Entekhabi, D., Bras, R. L. & Rodriguez-Iturbe, I. (1990). "Parameter estimation and sensitivity analysis for the modified Bartlett-Lewis rectangular pulses model of rainfall". *J. Geophys.Res*. **95**:2093-2100.

Ison, N.T., Feyerherm, A.M., Bark, L.D.(1971). “Wet Period Precipitation and the Gamma Distribution”. *Journal of Applied Meteorology*. **10**:658-665.

Jeff Haby, (2003). “Dynamic Precipitation VS Convective Precipitataion”. Retrieved on November 8, 2007, from http://www.theweatherprediction.com/habyhints/336

Johnson, J. T., P. L. MacKeen, A. Witt, E.D. Mitchell, G. J. Stumpf, M. D. Eilts and K. W. Thomas (1998). “The Storm Cell Identification and Tracking (SCIT) Algorithm: An Enhanced WSR-88D Algorithm”. *Weather and Forecasting*. **13**:263-276

Johnson, E.R. and Bras, R.L. (1980). “Multivariate short-term rainfall prediction”. *Water Resources Research*. **16**: 173-185.

Johnson, R.A. and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc.

Jon Goodall and David R. Maidment. (2002). “The Geostatistical Analyst. Center for Research in Water Resources”. University of Texas at Austin. Retrieved from http://www.ce.utexas.edu/prof/MAIDMET/giswr2005/geostat/Geostatistic Excercise.htm

Katz.R.W.,(1977). “Precipitation as a chain-dependent process”. *Journal of Applied* Meteorology.**16**:671-676.

Katz.R.W., Parlange,M.C.(1995). “Generalizations of chain-dependent processes: Appliocation to hourly precipitation”. *Water Resources Research*. **31**(5):1331-1341.

Khaliq, M. and Cunnane,C. (1996). “Modelling point rainfall occurrences with the modified Bartlett-Lewis Rectangular pulses model”. *Journal of Hydrology*.**180**:109-138.

Kavvas, M.L. & Delleur, J.W. (1975). “The Stochastic and Chronological structure of rainfall sequences: application to Indiana”. Water Resource Center, Rep. 57. Purdue University, West Lafayette, Indiana.

Kavvas, M.L. & Delleur, J.W. (1981). “A stochastic cluster model of daily rainfall sequences”. *Water Resources Research*. **17**:1151-1160.

Keckler, Doug (1995). *Surfer for Windows*. Golden Software, Inc., Golden, CO.

Keith Browning, Alan Blyth and Peter Clark, (2004). ‘Thunderstorm are’. Planet Earth Winter 2004 in www.nerc.ac.uk

Kim,S and Kavvas,M.L. (2006). “Stochastic Point Rainfall Modeling for Correlated Rain Cell Intensity and Duration”. J. *Hydrologic Engineering*. **11**(1):29-36.

Konstantin Krivoruchko (2006). “Introduction to Modeling Spatial Processes Using Geostatistical Analyst”. Retrieved on November 13, 2007, from http://www.esri.com/library/whitepapers/pdfs/intro-modeling.pdf

Lawrence, A.J., (1972). “Some models for stationary series of univariate events, in Stochastic Point Process: Statistical Analysis Theory and Applications” .1st ed., edited by P.A.W. Lewis, Wiley-Interscience, New York.

Llasat, M.C. (2001). “An objective classification of rainfall events on the basis of their convective features. Application to rainfall intensity in the North-East of Spain”. *International Journal of Climatology*, **21**(11): 1385 – 1400

Llasat, M. C. and Barnolas, M. (2007). “A Flood Geodatabase and its Climatological Applications: the Case of Catalonia for the Last Century”. *Natural Hazards and Earth System Sciences.* **7**: 271-281

Llasat, M. C. and Barriendos, M., Barrera, T., and Rigo, T. (2005). “Floods in Catalonia (NE Spains) since the 14th century. Climatological and meteorological aspects from historical documentary sources and instrumental records”. *Journal of Hydrology*. **313**:32-47

Llasat and Puigcerver. (1997). “Total Rainfall and Convective Rainfall in Catalonia, Spain”. *International Journal of Climatology***, 17:** 1683 – 1695

Loke, K.W. (1994). “Clustering Rainfall Frequencies for Klang River Basin”. Universiti Teknologi Malaysia: Master Thesis.

Maidment, D.R. (1993). *Handbook of Hydrology*. McGraw-Hill.

Main, C. L., Robinson, D. K., McElroy, J. S., Mueller, T. C., and Wilkerson, J. B. (2004). “A guide to predicting spatial distribution of weed emergence using geographic information systems (GIS)”. Online. Applied Turfgrass Science doi:10.1094/ATS-2004-1025-01-DG.

Manning, M.R., Pearman, G.I., Etheridge, D.M., Fraser, P.J., Lowe, D.C. and Steele, L.P. (1996). “The changing composition of the atmosphere”. In: Bouma, W.J., Pearman, G.I. and Manning, M.R. (eds) *Greenhouse:coping with climate change*. CSIRO, Collingwood, Australia. pp. 3-26

Matheron, G. (1963). “Principles of geostatistics”. *Eco. Geol*., **58**:1246-1266

Matthew, A.L., Linda, M.K., Charles, R.S., Jonathan, E.T. and George, A.W. (2003). “Antarctic Satellite Meteorology: Applications for Weather Forecasting”. *Journal of Climate*. **131:** 371-383.

Mielke, P.W. (1973). “Another family of distribution for describing and analyzing precipitation”. *Journal of Applied Meteorology*. **12**: 275-280.

Mohd. Nor Mohd. Desa (1997). "Characterization of Urban Rainfall in Kuala Lumpur Malaysia". Department of Water Resources Engineering, Lund University, Lund, Sweden: Ph.D Thesis.

Mohd. Nor Mohd. Desa and Zalina Mohd.Daud (1999). "Interpretation of spatial and temporal properties of annual and monthly rainfall in Selangor, Malaysia." Proc. Colloquium on Hydrology and Water Management in the Humid Tropics. Panama

Mohd Nor, M.D. and Zalina, M.D. (1999). "On spatial and temporal properties of rainfall in Selangor". *Proc. International Seminar on Hydrology and Water Resources in SEA and the Pacific, Taegu, Korea*. pp. 347-356.

Montanari, A., Burlando, P. and Rosso, R. (1994). "Forecasting of short-term rainfall using multivariate ARMA models (abstract)". *XIX EGS General Assembly, Grenoble. Annales Geophysicae*, **12 (Special Issue),** C325, C409.

Murphy, A.H. (1993). "What is a good forecast? An essay on the nature of goodness in weather forecasting". *Weather and Forecasting*. **8:** 281-293.

Murphy, J. (1999). "An Evaluation of Statistical and Dynamical Techniques for Downscaling Local Climate". *Journal of Climate*. **12:** 2256-2284.

Nelder, J.A. and Mead, R. (1963). "A simplex method for function minimization". *The Computer Journal.* 308-312.

Neyman, J. and Scott, E.L. (1958). "Statistical approach to problems of cosmology". *Journal of Royal Statistical Society Series B*.**20**:1-43.

Niemczynowicz, J. and Dahlblom, P. (1984). "Dynamics properties of Rainfall in Lund". *Nordic Hydrol*., **15:** 9 – 24.

Nix, S. J. (1994). *Urban Stromwater Modeling and Simulation*. Lewis Publishers/CRC Press, 224 pp.

Norhan Mat Yusoff and Mazizn Hahin (2007). "Change Detection Analysis of Urban Forest in Klang Valley using Multiemporal Remote sensing Data: some Preliminary Result". *GIS Development*. Retrieved August 25, 2007, from http://www.gisdevelopment.net/aars/acrs/1997/ps2/ps4015.asp

Nordila A.,Zulkifli Y. and Zalina MD.(2006). "Characterization of Convective Rain in Klang Valley". In Proceedings of National Conference on Water for Sustainable Development towards a Developed Nation by 2020

Nguyen, V.T.V., McPherson, M.B. and Rousselle, J. (1978). Technical Memo 35. *Urban Water Resources Research Program, American Society Civil Engineering*.

Nguyen, V.T.V and Rousselle, J. (1981). "A stochastic Model for the time distribution of hourly rainfall depth". *Water Resources Research*. **17**(2) :399-409.

Nguyen, V.T.V and Mayabi, A. (1990). "Probabilistic Analysis of Summer Daily Rainfall for the Montreal Region". *Canadian Water Resources Journal*. **15**(3).

Ohsawa, T., H. Ueda, T. Hayashi, A. Watanabe and J. Matsumoto, (2001). "Diurnal Variations of Convective Activity and Rainfall in Tropical Asia". *Journal of the Meteorological Society of Japan*, **79:** 333-352

Oliver, V. J., and R. A. Scofield (1976). "Estimation of Rainfall from Satellite Imagery", *First Conference on Hyrdometeorology, Ft. Worth, Tex*., American Meteorology Society. April 20-22, pp. 98-101

Onof, C. and Wheater, H.S. (1993). "Modelling of British rainfall using a random parameter Bartlett-Lewis rectangular pulse model". *Journal of Hydrology*. **149**:67-95.

Onof, C. and Wheater, H.S. (1994). "Improvements to the modeling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model". *Journal of* Hydrolog.**157**: 177-195.

Onof,C., Chandler,R.E., Kakou,K., Northrop,P., Wheater, H.S., and Isham, V. (2000). "Rainfall modelling using Poisson-cluster processes: A review of developments". *Stochastic Environmental Research and Risk Assessment* .**14**: 384-411.

Pascual, R., Callado, A. and Berenguer, M. (2004). "Convective Storm Initiation in Central Catalonia". *Proceedings of ERAD (2004)*. Copernicus, 464-468

Pattison, A. (1965). "Synthesis of hourly rainfall data". *Water Resources Research*. **1**:489-498.

Phanartzis, C.A. (1979). "Rainfall prediction, progress report". *Wastewater Program, City and Country of San Francisco,* CA.

PennState (2001). "Convective Rainfall. Topic 4". Penn State Personal Web Server, in www.personal.psu.edu/users/m/s/mss298/Meteo482/topic4.html

Peyron, N. and Nguyen, V.T.V. (2004). "A systematic assessment of global optimization methods for conceptual hydrologic model calibration". Proceedings of the 6th International Conference on Hydroinformatics, Singapore, World Scientific Publishing Company.

Priestly, M.B. (1981). *Spectral Analysis and Time Series*. Academic Press, Inc.

Quenouille, M.H. (1957). *The Analysis of Multiple Time Series*. Griffin.

Rascko, P., Szeidl., and Semenov, M. (1991). "A Serial Approach to Local Stochastic Weather Models". *Ecological Modelling*. **57**: 27-41.

Ray K. Linsey, Jr., Max A. Kohler and Joseph L. H. Paulhus (1988). *Hydrology for Engineers*. (SI Metric Edition) Singapore: Mc Graw Hill.

Richardson, C.W. (1981). "Stochastic simulation of daily precipitation, temperature, and solar radiation". *Water Resources Research*. **17**(1):182-190.

Richardson, C.W., and D.A Wright.(1984). "WGEN: A model for generating daily weather variables". U.S. Department of Agriculture Research Service, ARS-8.

Rigo T. and Llasat M. C. (2002). "Analysis of Convective Structures that Produce Heavy Rainfall Events in Catalonia (NE of Spain), Using Meteorological Radar". *Proceedings of ERAD*. Copernicus, 45-48.

Rodriguez-Iturbe, I., (1986). "Scale of fluctuation of rainfall models". *Water Resources Research*. **22**:15S-37S.

Rodriguez-Iturbe, I., Cox, D.R., and Isham, V. (1987a). "Some models for rainfall based on stochastic point processes". *Proc. Roy. Soc. London Ser. A*. **410**(1839):269-288.

Rodriguez-Iturbe, I., Cox, D.R., and Isham, V. (1988). "A point processes model for rainfall: further development". *Proc. Roy. Soc. London Ser. A*. **417**(1853):283-298.

Rodriguez-Iturbe, I.,Febres De Power, B., and Valdes, J. (1987b). "Rectangular pulses point process models for rainfall: analysis of empirical data". *J. Geophys. Res*. **92**(D8):9645-9656.

Rodriguez-Iturbe, I.,Gupta, V., and Waymire, E. (1984). "Scale considerations in modeling of temporal rainfall". *Water Resources Research*. **20**:1611-1619.

Rodriguez Iturbe, I., Febres de Power, B., Sharifi, M.B. and Georgakakos, K.P. (1989). "Chaos in Rainfall". *Water Resources Research*. **25:** 1667-1675.

Roldan, J. and Woolhiser, D.A. (1982). "Stochastic Daily precipitation Models 1. A comparison of Occurrence Processes". *Water Resources Research.* **18**(5): 1451-1459.

Ronald L. Holle and Andrew I. Watson. (1983). "Duration of Convective Events Related to Visible Cloud, Convergence, Radar and Rain Gage Parameters in South Florida". *Monthly Weather Review* **111**(5):1046-1051.

Salas, J.D., Delleur,J.W., Yevjevich,V., Lane,W.L. (1980). "Applied modeling of Hydrologic Time Series". Water Resources Publications.

Shaw, E. M. (1989). *Hydrology in Practice*. 2nd Edition, VNR International

Shaw, S.R (1983). "An investigation of the cellular structure of storms using correlation techniques". *Journal of Hydrology*. **62**:63-79.

Schilling, W. (1986). "Urban runoff quality management by real-time control". *Nato ASI Series*, Vol. G10, pp. 765-817.

Smith, R.E., and H.A. Schreiber (1973). "Point processes of seasonal thunderstorm rainfall, 1. Distribution of rainfall events". *Water Resources Research*. **9**(4), 871-884.

Smith, J.A. and Karr, A.F. (1985a). "Parameter estimation for a model of space-time rainfall". *Water Resources Research*. **21**:1251-1257.

Smith, J.A. and Karr, A.F. (1985b). "Statistical inference for point process models of rainfall". *Water Resources Research*, **21**:73-79.

Sirangelo, B.(1992). "Intensity-duration-frequency curves from rainfall models with beta-shaped pulses". In Sixth IAHR International Symposium on Stochastic Hydraulics, Taipei.

Steiner M, Houze RA Jr, Yuter S (1995). Climatological Characterization of Three-dimensional Storm Structure from Operational Radar and Raingauge Data. *Journal of Applied Meteorology*. **34**: 1978-2006

Tiao, G.C. and Box, G.E.P. (1981). "Modeling multiple time series with Applications". *Journal of the American Statistical Association*. **76:** 802-816.

Todorovic,P., and D.Woolhiser (1975). "A stochastic model of n-day precipitation". *J. Applied Meteorology*. **14**(1), 17-24.

Toth, E., Brath, A. and Montanari, A (2000). "Comparison of short-term rainfall prediction models for real-time flood forecasting". *Journal of Hydrology*. **239**: 132-147.

Tsay, R.S. (2000). "Time series and forecasting: Brief history and future research". *Journal of the American Statistical Association*. **95:** 638-643.

Tsay, R.S. and Tiao, G.C. (1984). "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models". *Journal of the American Statistical Association*. **79**: 84-96.

Tsonis, A.A. and Elsner, J.B. (1989). "Chaos, strange attractors and weather". *Bulletin of the American Meteorological Society*. **70:** 14-23.

University of Illinois. Cold Front. Retrieved on 2006-10-22.
Vazquez, R. Lorente, J. and Redano, (1987). Curves IDF Barcelona-Fabra, Rev. Ob Publ., **3255**: 91-102.

Valdes, J.B., Rodriguez-Iturbe, I., and Gupta, V.K.(1985). “Approximations of temporal rainfall from multidimensional model”. *Water Resources Research*. **21**:1259-1270.

Vazquez, R. Lorente, J. and Redano, (1987). ‘Curves IDF Barcelona-Fabra’, *Rev. Ob. Publication*., 3255, 91-102

Velghe, T., Troch, P.A., De Troch, F.P., and Van de Velde, J. (1994). “Evaluation of cluster-based rectangular pulses point process models for rainfall”. *Water Resources Research.* **30**:2847-2857.

Vilar, E., Burgueno, A., Puigcerver, M. and Austin, J. (1988). “Analysis of joint rainfall rate and duration statistics: microwave system design implications”, *IEEE Trans. Comm.* COM-36, 650-660.

Watson, P. A., Gunes, M., Potter, B. A., Sathiaseelan, V. and Leitas, J. (1982). “Development of a Climatic Map of Rainfall Attenuation for Europe”, Final Report, European Space Agency. Postgraduate School of Electronics and Electronic Engineering, University of Bradford, Bradford.

Waymire, E. and Gupta, V.K. (1981a). “The mathematical structure of rainfall representations: 1. A review of stochastic rainfall models”. *Water Resources Research*. **17**:1261-1272.

Waymire, E. and Gupta, V.K. (1981b). “The mathematical structure of rainfall representations: 1. A review of the theory of point processes”. *Water Resources Research*. **17**:1273-1285.

Waymire, E. and Gupta, V.K. (1981c). “The mathematical structure of rainfall representations: 1. Some applications of the point process theory to rainfall processes”. *Water Resources Research*. **17**:1287-1294.

Waymire, E., Gupta, V.K. and Rodriguez-Iturbe, I. (1984).”A Spectral Theory of Rainfall Intensity at the meso-β scale”. *Water Resources Research*. **20:** 1453-1465.

Weise (2001). “Kriging – a statistical interpolation method and its applications”. GEOG 516.

Wahba G (1990). “Spline models for observational data.” Philadelphia: Society for Industrial and Applied Mathematics.

WikiAnswers (2007). "Precipitation". Retrieved on November 8, 2007, from http://www.answers.com/topic/precipitation

Wikipedia (2007). "Convection cell". Retrieved on November 9, 2007, from http://en.wikipedia.org/wiki/Convection_cell

Wikipedia (2007). "Mesoscale Convective Systems". Retrieved on November 14, 2007, http://en.wikipedia.org/wiki/Mesoscale_Convective_System

Wikipedia (2007). "Raingauge". Retrieved on November 12, 2007, from http://en.wikipedia.org/wiki/Rain_gauge

Wilks, D.S., and Wilby, R.L (1998). "Interannual variability and extreme-value characteristics of several stochastic daily precipitation models". *Agricultural and Forest Meteorology*. **93**: 153-169.

Woolhiser,D.A. and Roldan,J. (1982). "Stochastic Daily Precipitation Models 2. A comparison of Distributions of Amounts". *Water Resources Research.* 18(5):1461-1468.

Woolhiser,D.A. and Roldan,J. (1986). "Seasonal and Regional Variability of Parameters for Stochastic Daily Precipitation Models:South Dakota, U.S.A*". Water Resources Research*. **22**(6):965-978.

Woolhiser, D.A. and Pergram, G.G.S. (1979). "Maximum Likelihood Estimation of Fourier Coefficients to describe Seasonal Variations of Parameters in Stochastic Daily precipitation models". *Journal of Applied Meteorology*. **18**:34-42.

Yan, O. C. & Lin, L. W. (1986). "Variation of rainfall with area in Peninsular Malaysia. Water Resources Publication". No 17. Bahagian Parit dan Taliair Kementerian Pertanian Malaysia.

Yeboah, G-A. and Garry R,W. (1997). "A hybrid model for point rainfall modelling". *Water Resources Research*. **33**(7): 1699-1706

Yoo, C., Valdes, J.B. and North, G.R. (1996). "Stochastic modelling of multidimensional precipitation fields considering spectral structure". *Water Resources Research*. 3**2**(7): 2175-2187.

Zawadski, I. (1987). "Fractal Structure and Exponential Decorrelation in Rain". *Journal of Geophysical Research*. **92:** 9586-9591.

APPENDIX A

PROCESS OF DIGITIZE RADAR IMAGE

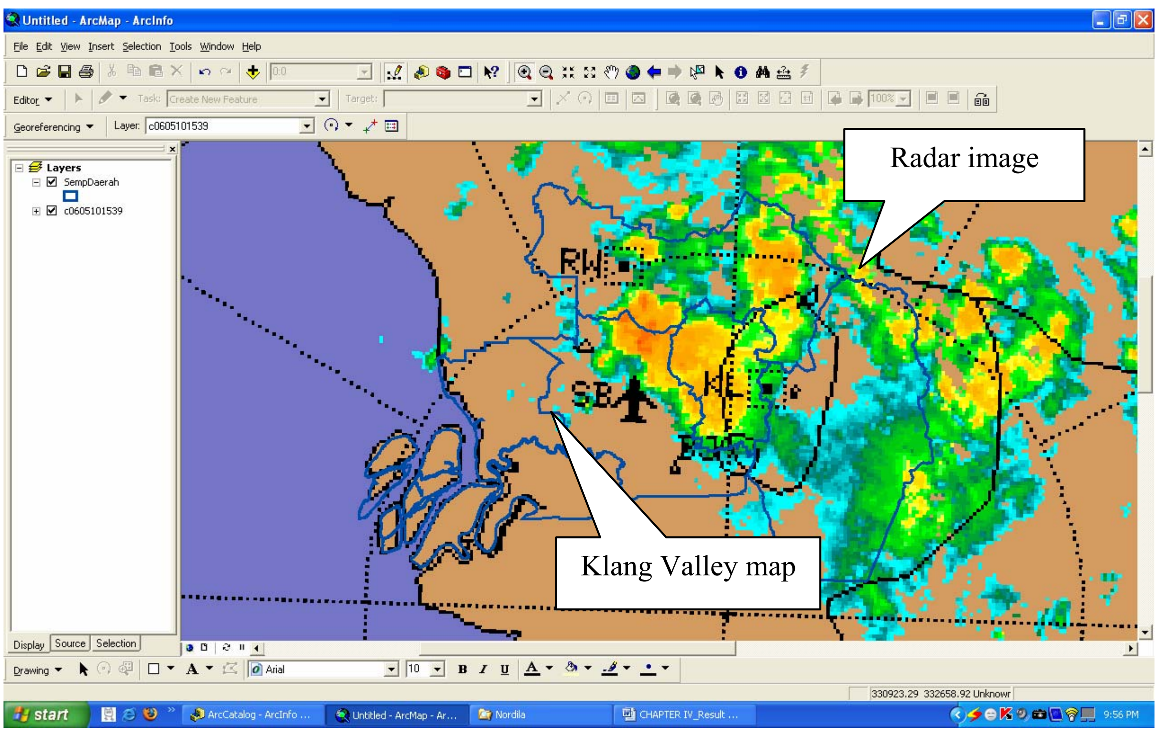


Figure A1 : Radar image is rectified with Klang Valley map

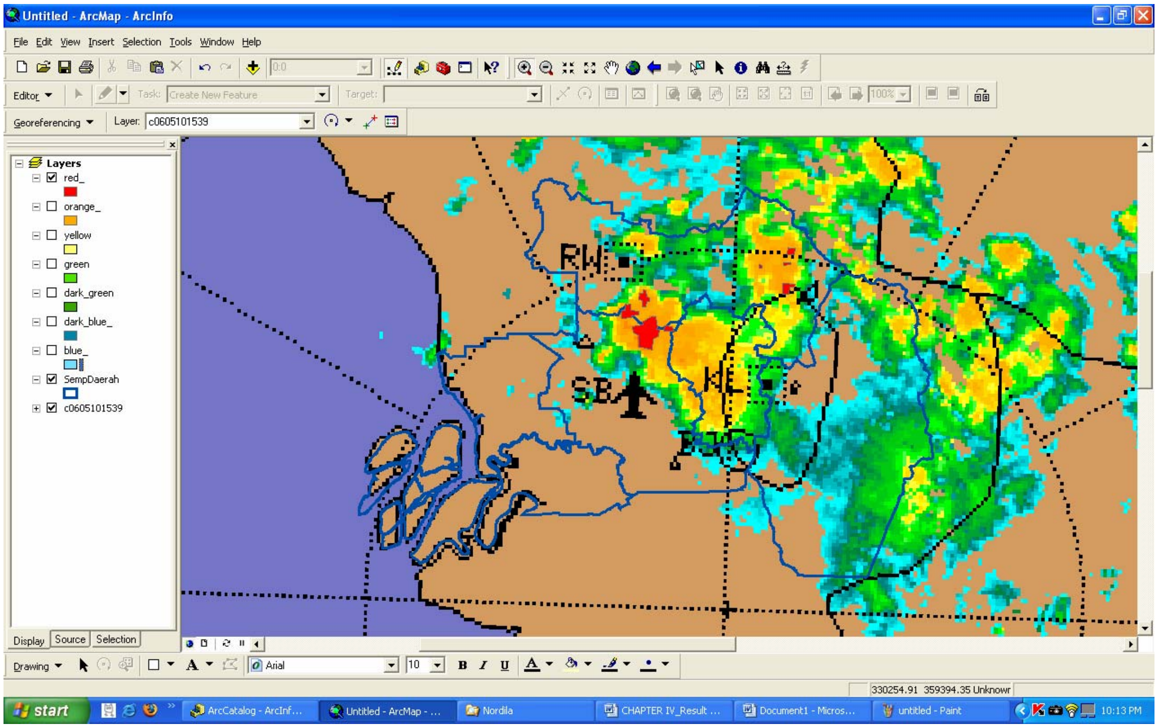


Figure A2 : Digitizing of radar image for intensity 80 – 100 mm/hr (red layer)

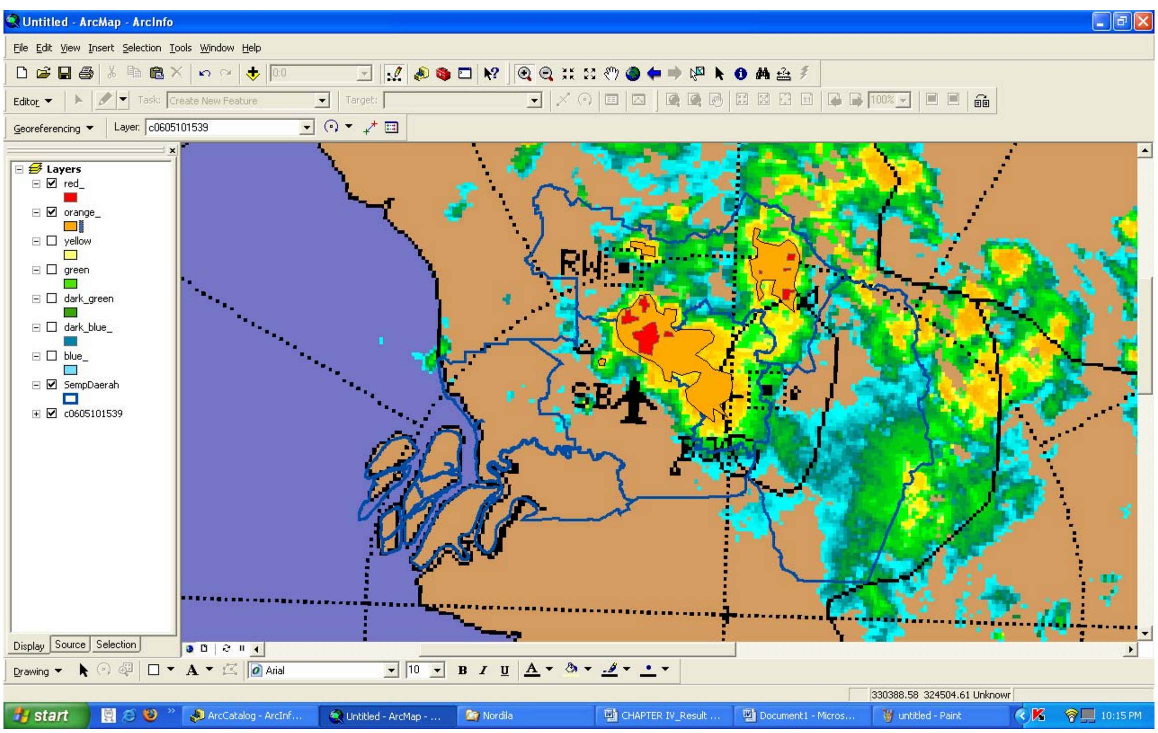


Figure A3 : Digitizing of radar image for intensity 35 – 80 mm/hr (orange layer)

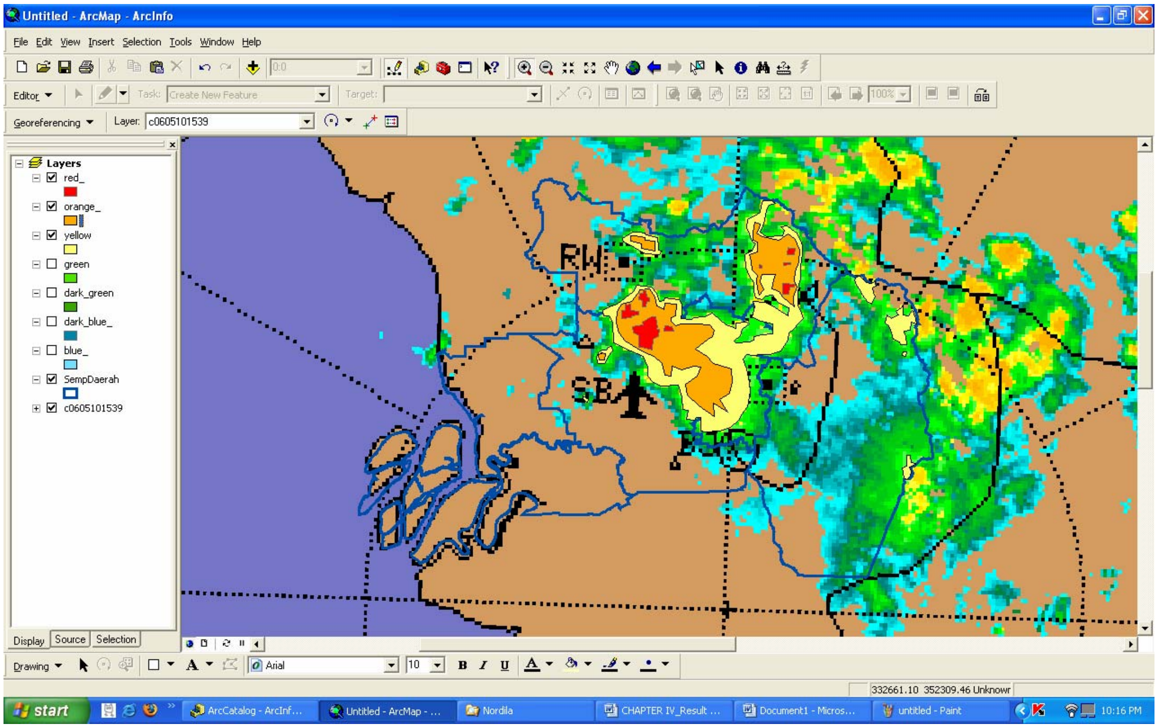


Figure A4 : Digitizing of radar image for intensity 8 – 35 mm/hr (yellow layer)

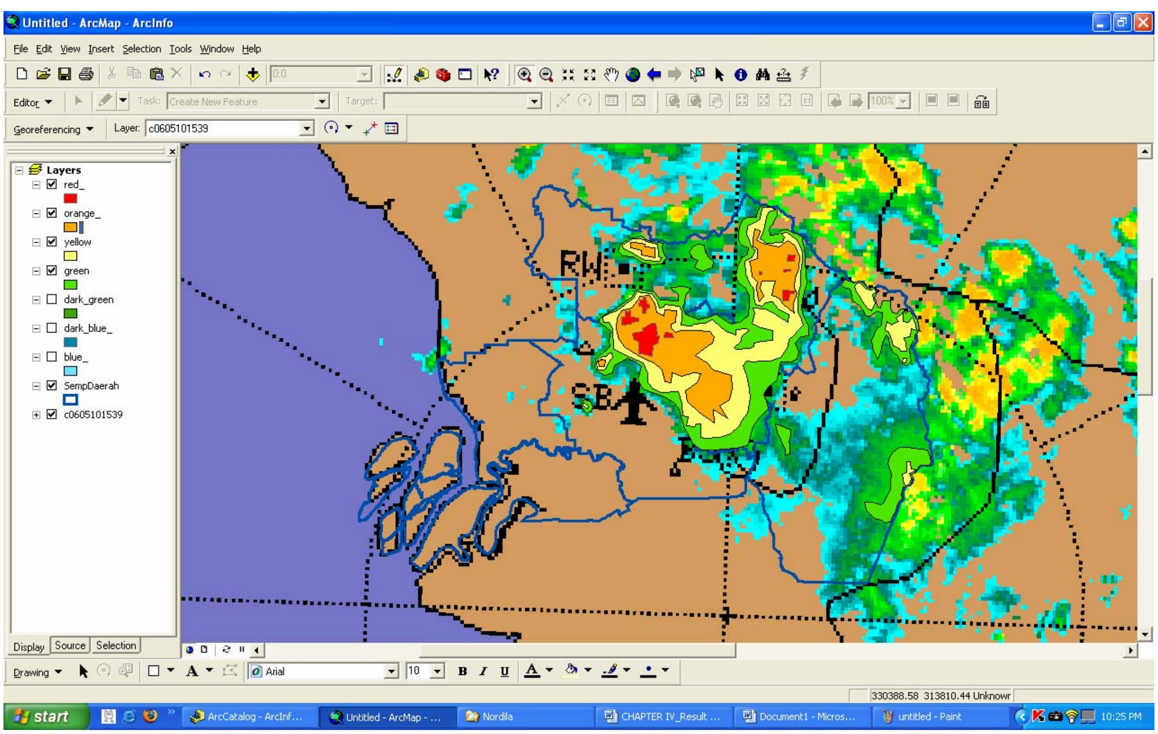


Figure A5 : Digitizing of radar image for intensity 3 – 8 mm/hr (green layer)

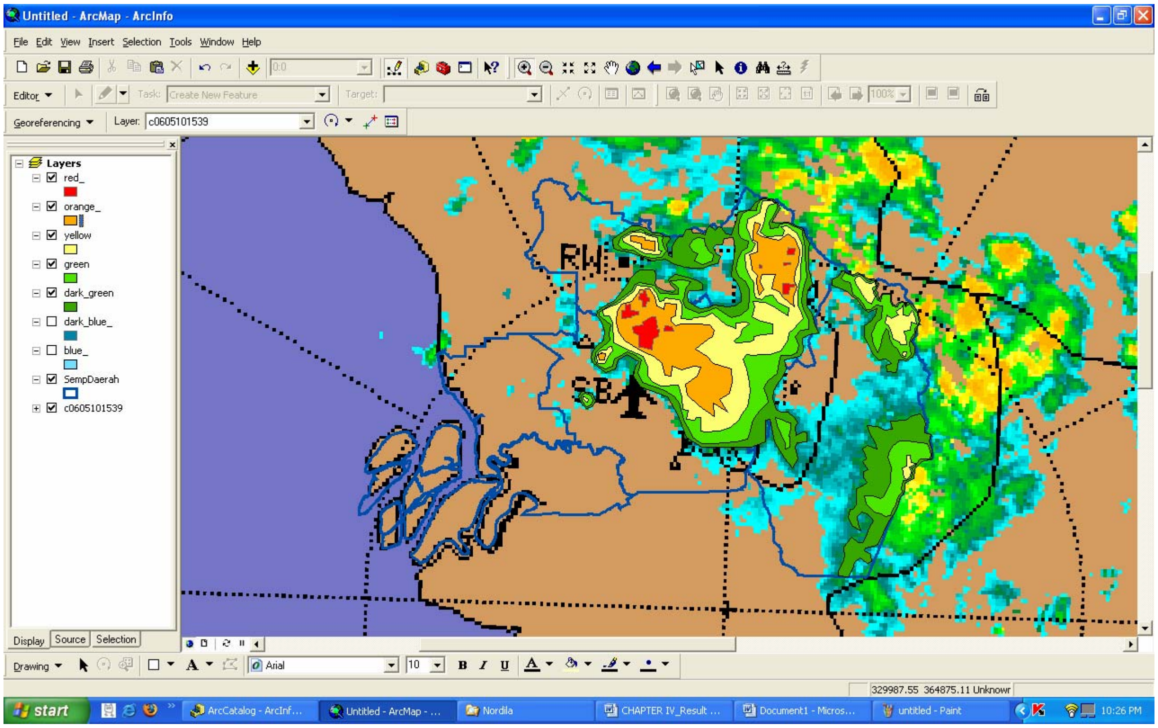


Figure A6 : Digitizing of radar image for intensity 0.9 – 3 mm/hr (dark green layer)

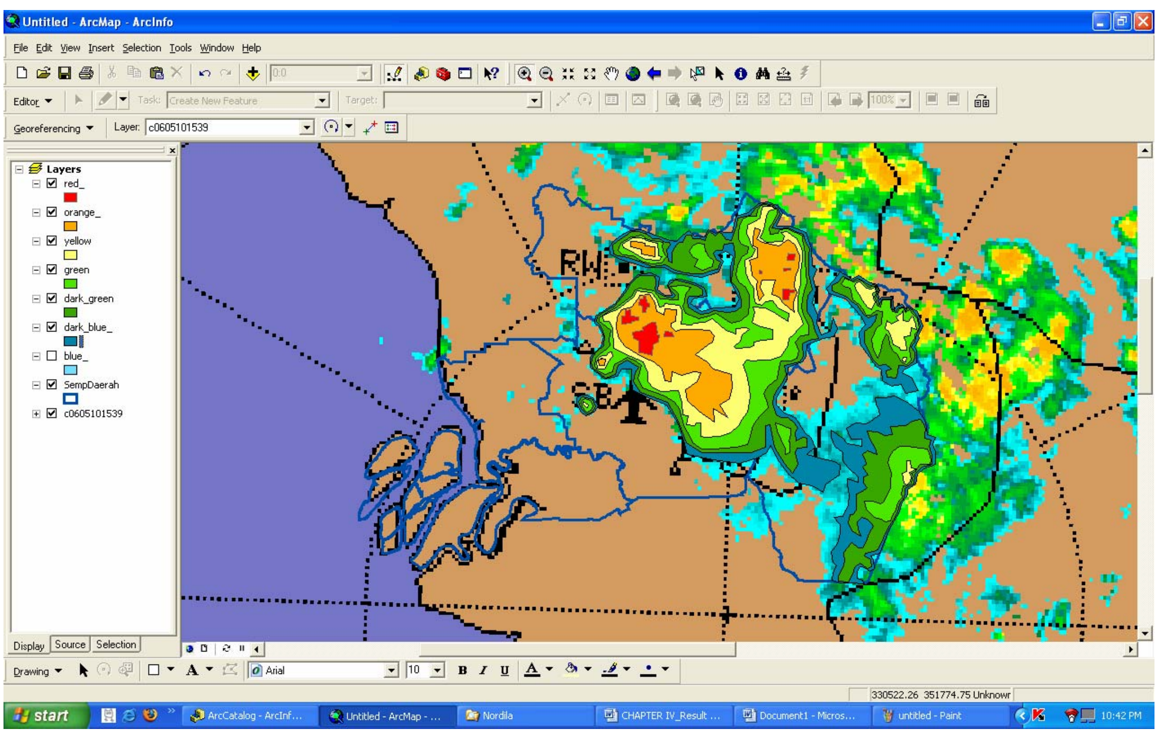


Figure A7 : Digitizing of radar image for intensity 0.5 – 0.9 mm/hr (dark blue layer)

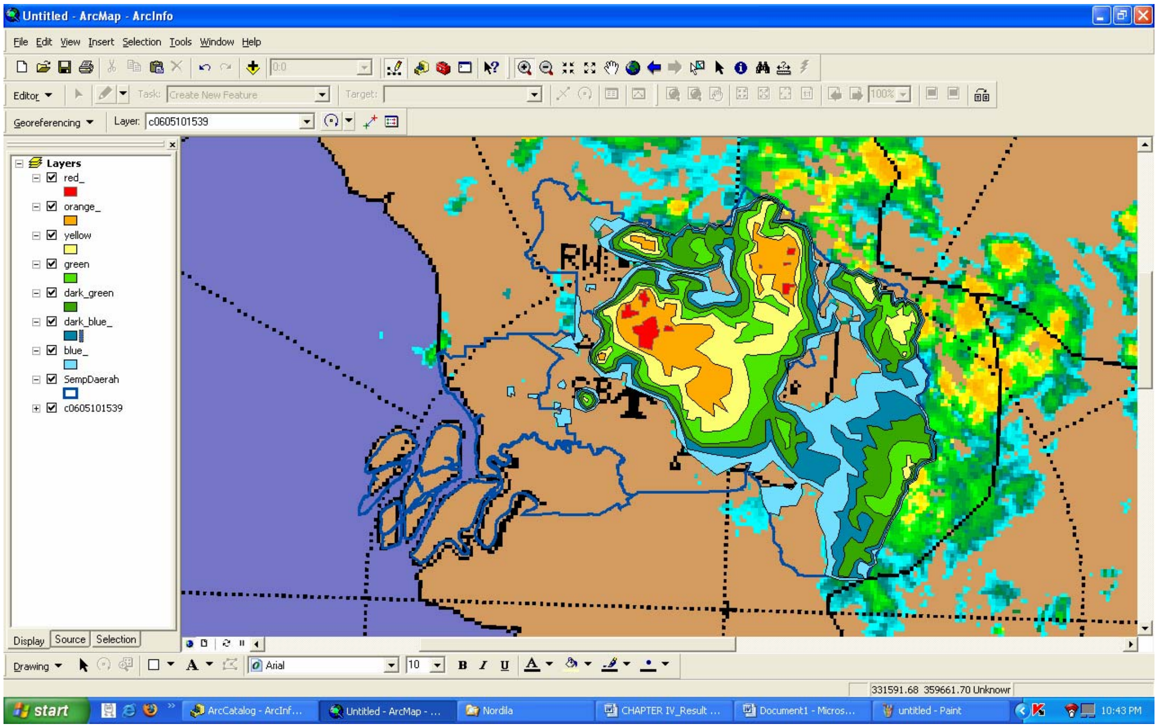


Figure A8 : Digitizing of radar image for intensity 0.3 – 0.5 mm/hr (blue layer)

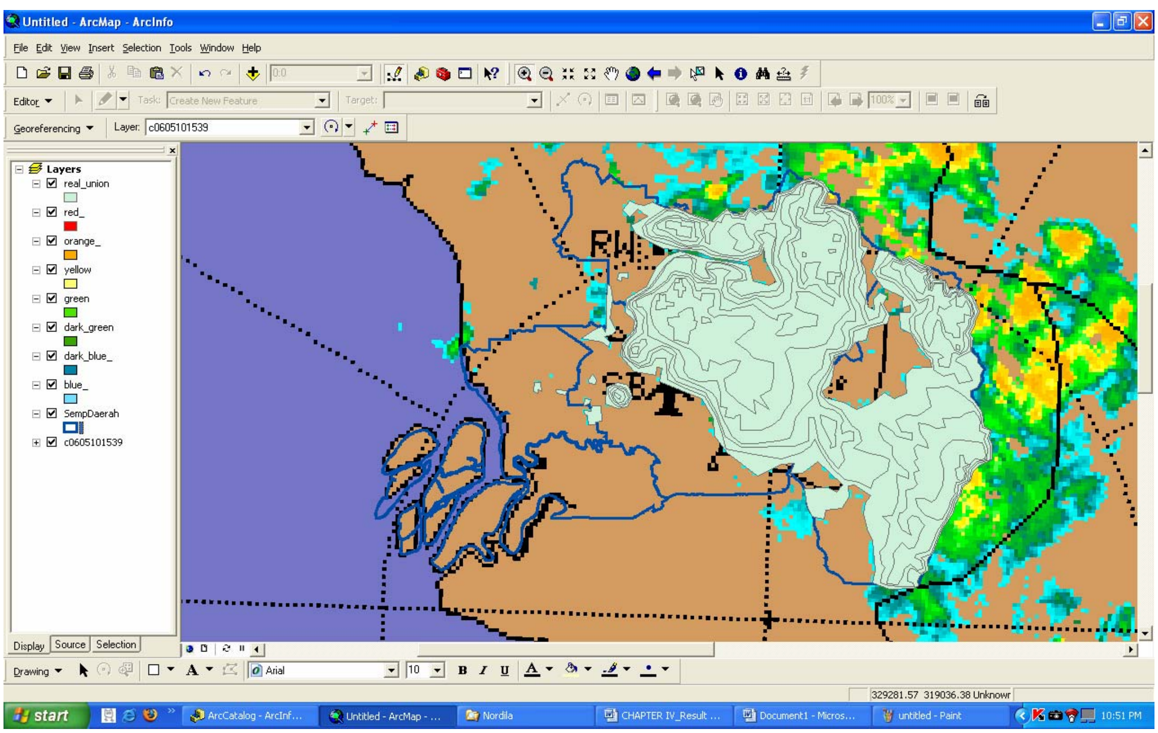


Figure A9 : Union process (merged all layers)

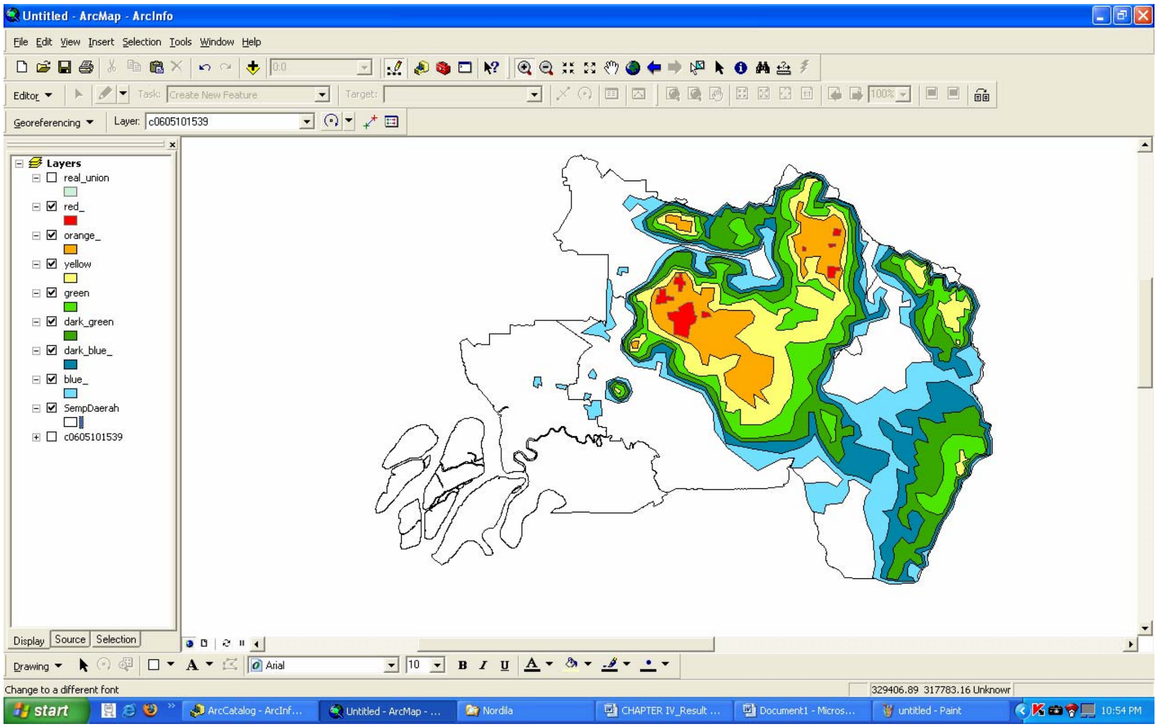


Figure A10 : Digitized image

APPENDIX B

STEPS TO MAKE RAINFALL CONTOURS DERIVED BY KRIGING METHOD USING GEOSTATISTICAL ANALYST

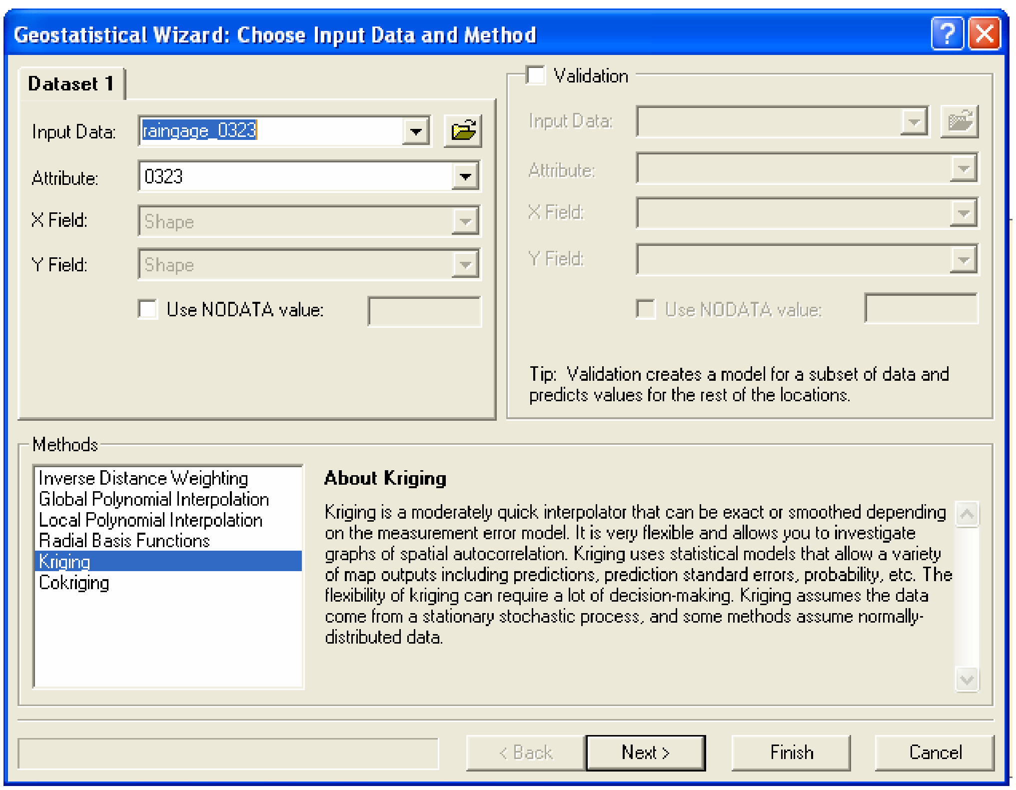


Figure B1 : Choose input data and method

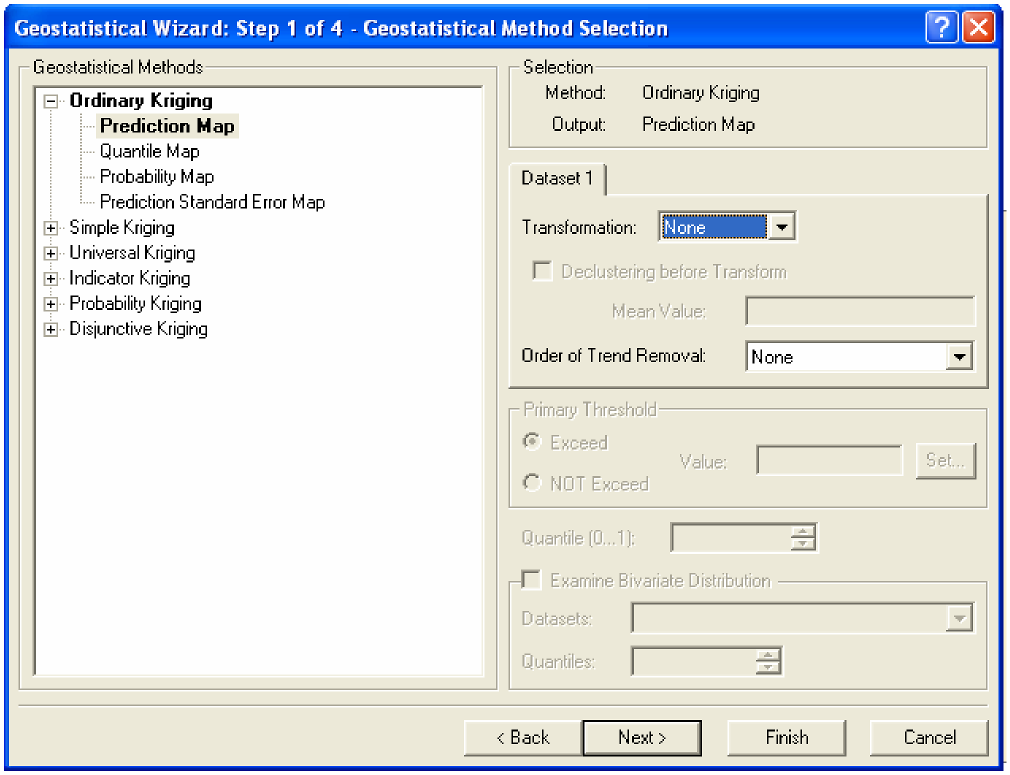


Figure B2 : Geostatistical method selection

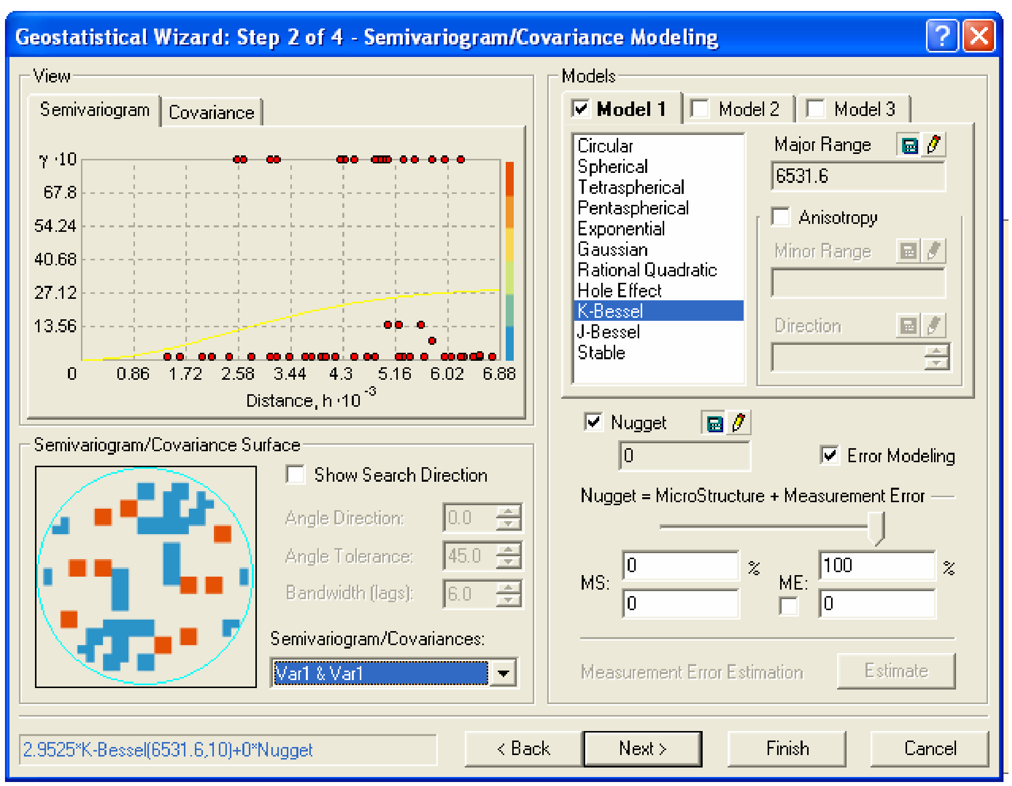


Figure B3 : Semivariogram / Covariance modeling

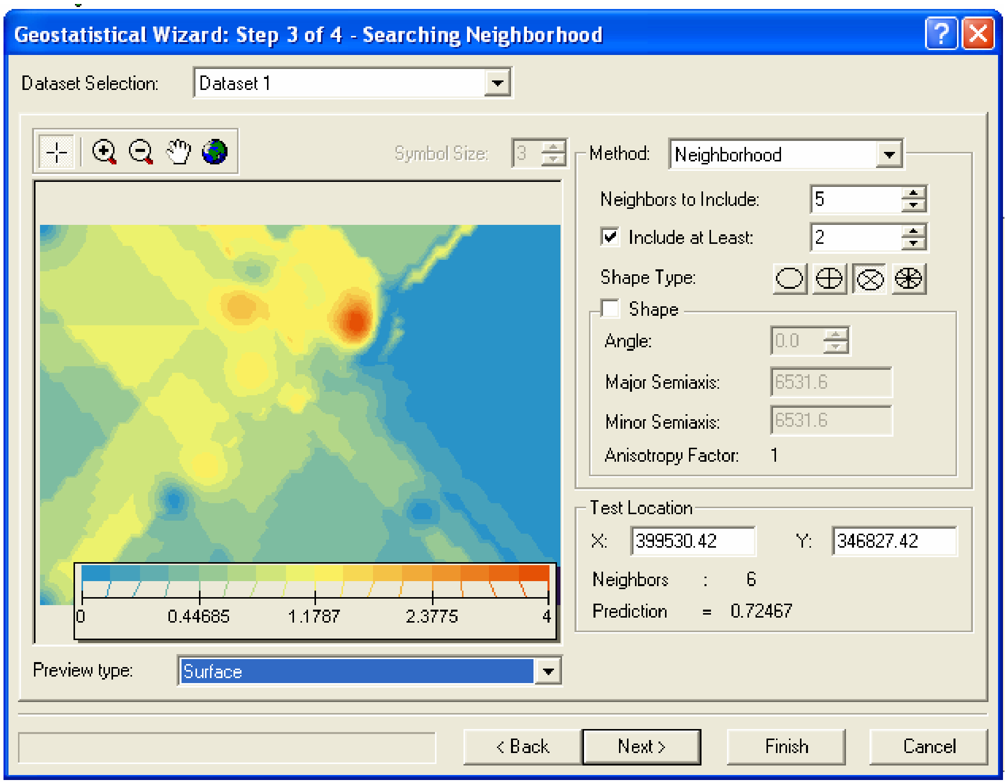


Figure B4 : Searching neighborhood

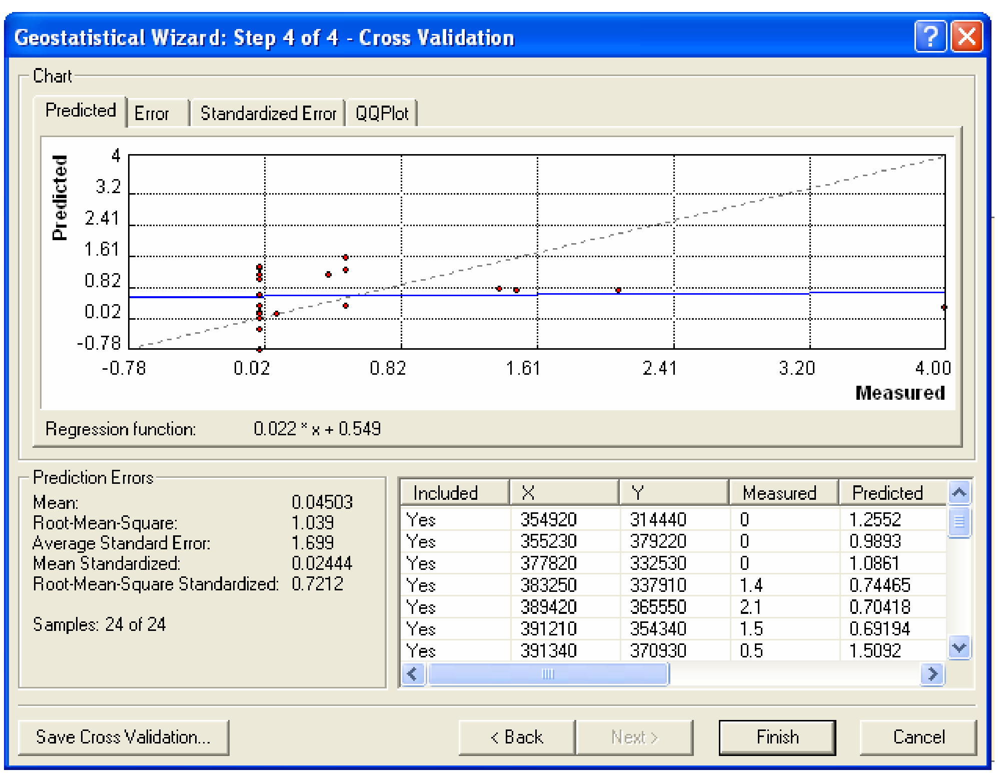


Figure B5 : Cross validation

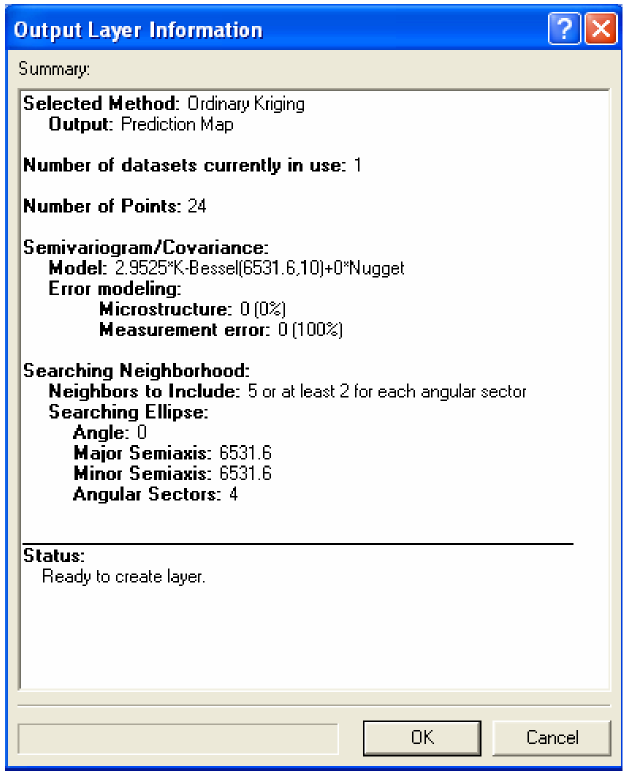


Figure B6 : Output layer information

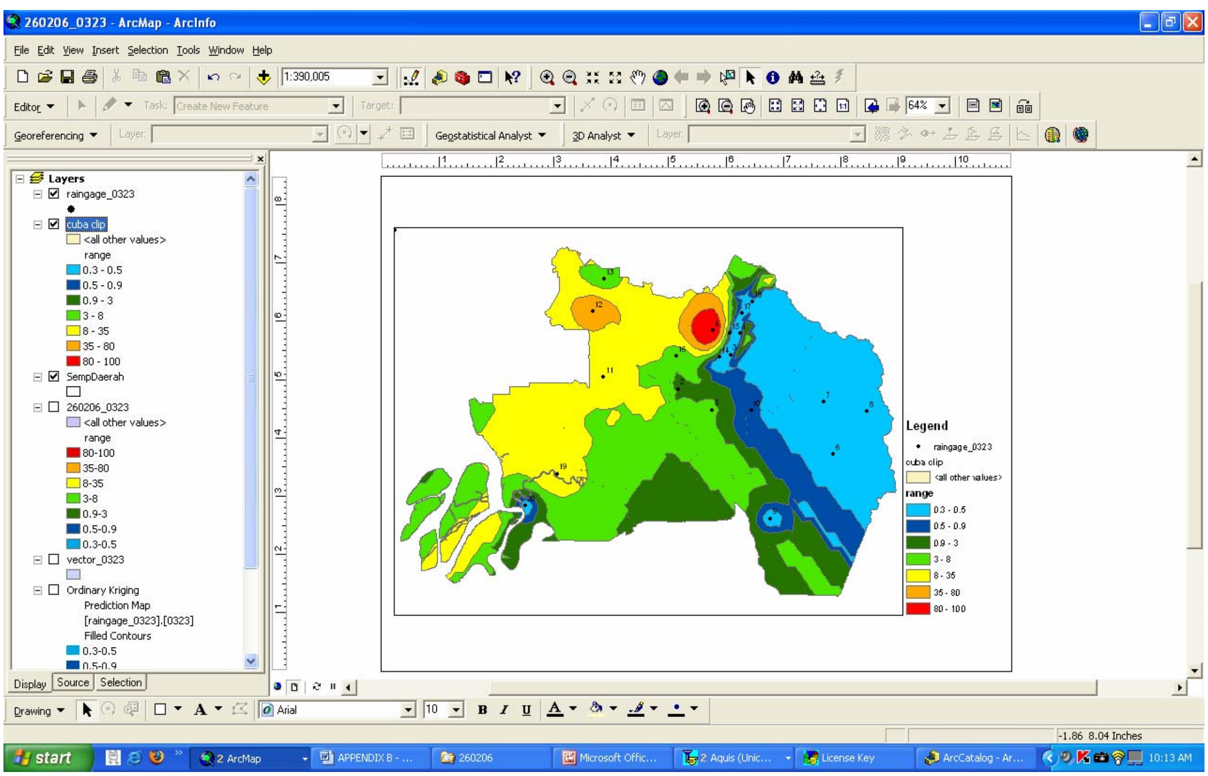
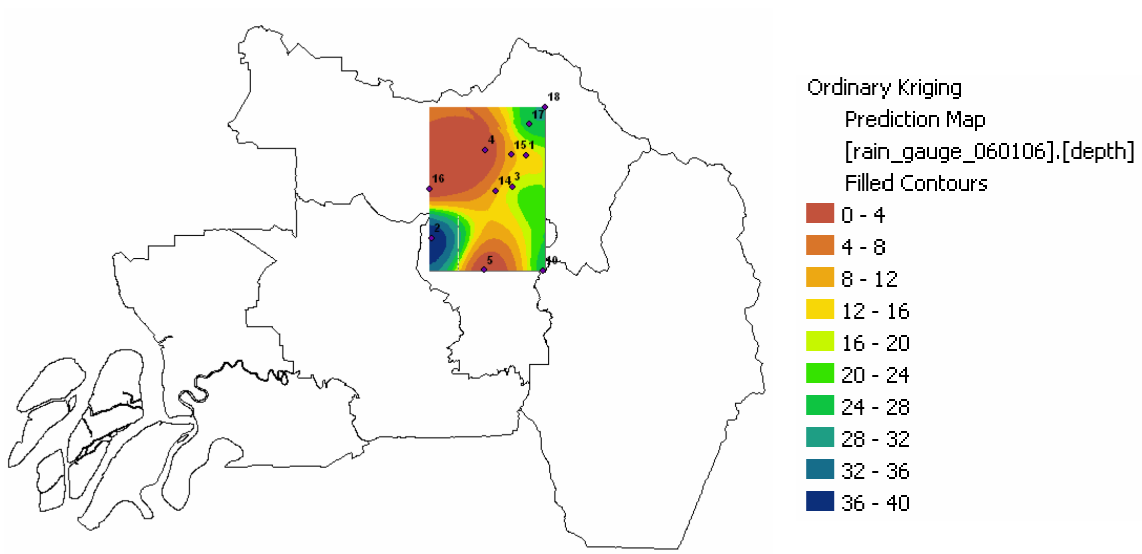


Figure B7 : Rainfall contour derived from Kriging

APPENDIX C

CALCULATION TO PRODUCE AREAL REDUCTION CURVE

Event on January 6, 2006



Classes	Value_Min	Value_Max	F_AREA	km_square	range
0	0	5.732849	64988555.5251	64.99	0 - 4
1	5.732849	9.867464	31596596.3039	31.6	4 - 8
2	9.867464	12.849408	28627410.7376	28.63	8 - 12
3	12.849408	15.000031	34384762.8536	34.38	12 - 16
4	15.000031	16.551090	21701199.8258	21.7	16 - 20
5	16.551090	18.701712	26224774.7132	26.22	20 - 24
6	18.701712	21.683657	13813125.8888	13.81	24 - 28
7	21.683657	25.818270	7139504.17534	7.14	28 - 32
8	25.818270	31.551119	6549684.33259	6.55	32 - 36
9	31.551119	39.5	6315137.79596	6.32	36 - 40

Percentage reduction (%) of storm depth (event on January 6, 2006)

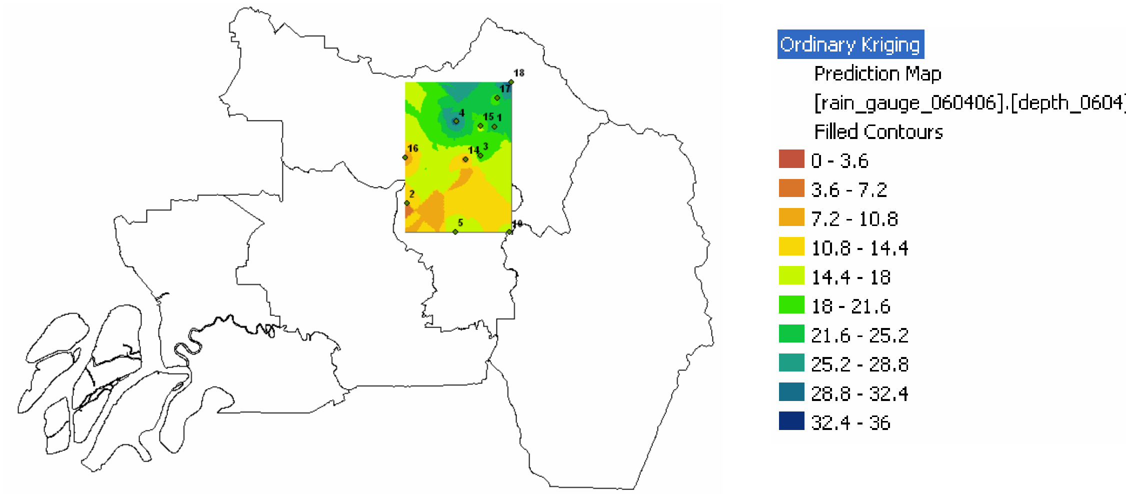
1	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP) = (average between isohyet x area between isohyet) / total areas between all pairs of neighbouring isohyets	(36 + 40) / 2= **38** 6.32+ 0 = **6.32** (38 x 6.32)/6.32 = **38**
2	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(32 + 36)/2 = **34** 6.32 + 6.55 **= 12.87** [(38 x 6.32) + (34 x 12.87)] / 12.87 **= 36**
3	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(28 + 32) /2 = **30** 6.32 + 6.55 + 7.14 = **20.01** [(38 x 6.32) + (34 x 12.87) + (30 x 20.01)] / 20.01 = **33.8**
4	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(24 + 28) /2 = **26** 6.32 + 6.55 + 7.14 + 13.81 = **33.82** [(38 x 6.32) + (34 x 12.87) + (30 x 20.01) + (26x33.82)] / 33.82 = **30.6**
5	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(20 + 24) /2 = **22** 6.32 + 6.55 + 7.14 + 13.81 + 21.7 = **55.52** [(38 x 6.32) + (34 x 12.87) + (30 x 20.01) + (26x33.82) + (22x 55.52)] / 55.52 = **27.3**
6	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(16 + 20)/2 = **18** 6.32 + 6.55 + 7.14 + 13.81 + 1.7 + 26.22 = **81.74** [(38 x 6.32) + (34 x 12.87) + (30 x 20.01) + (26x33.82) + (22x 55.52) + (18 x 81.74)] / 81.74 = **24.3**

7	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(12 + 16)/2 = **14** 6.32 + 6.55 + 7.14 + 13.81 + 21.7 + 26.22 + 28.63 = **110.37** [(38 x 6.32) + (34 x 12.87) + (30 x 20.01) + (26x33.82) + (22x 55.52) + (18 x 81.74) + (14 x 110.37)] / 110.37 = **21.6**
8	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(8 + 12) /2= 10 6.32 + 6.55 + 7.14 + 13.81 + 21.7 + 26.22 + 28.63 + 31.6 = **141.97** [(38 x 6.32) + (34 x 12.87) + (30 x 20.01) + (26x33.82) + (22x 55.52) + (18 x 81.74) + (14 x 110.37) + (10 x 141.97)] / 141.97 = **19.0**
9	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(4 + 8) /2 = **6** 6.32 + 6.55 + 7.14 + 13.81 + 21.7 + 26.22 + 28.63 + 31.6 + 34.38 = **176.35** [(38 x 6.32) + (34 x 12.87) + (30 x 20.01) + (26x33.82) + (22x 55.52) + (18 x 81.74) + (14 x 110.37) + (10 x 141.97) + (6 x 176.35)] / 176.35 = **16.5**
10	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(0 + 4) /2= 2 6.32 + 6.55 + 7.14 + 13.81 + 21.7 + 26.22 + 28.63 + 31.6 + 34.38 + 64.99 = **241.34** [(38 x 6.32) + (34 x 12.87) + (30 x 20.01) + (26x33.82) + (22x 55.52) + (18 x 81.74) + (14 x 110.37) + (10 x 141.97) + (6 x 176.35) + (2 x 241.34)] / 241.34 = **12.6**

Percentage reduction (%) of storm depth	**=** (Mean Area Precipitation, (MAP) / storm maximum)* 100
storm maximum (reference gauge)	= 39.5 mm

No.	**Percentage reduction (%) of storm depth**		
1	38 / 39.5 * 100 = 96.2 %	6	24.3 / 39.5 * 100 = 61.5%
2	36 / 39.5 * 100 = 91 %	7	21.6 / 39.5 * 100 = 54.7 %
3	33.8 / 39.5 * 100 = 85.7 %	8	19.0 / 39.5 * 100 = 48.2 %
4	30.6 / 39.5 * 100 = 77.6 %	9	16.5 / 39.5 * 100 = 41.8 %
5	27.3 / 39.5 * 100 = 69 %	10	12.6 / 39.5 * 100 = 31.9 %

Event on April 6, 2006



Classes	Value_Min	Value_Max	F_AREA	km_square	range	new_range
0	0	2.531383	179.519956	0	0 - 4	0 - 3.6
1	2.531383	4.587440	1185586.49282	1.19	4 - 8	3.6 - 7.2
2	4.587440	6.257425	18469874.1837	18.47	8 - 12	7.2 - 10.8
3	6.257425	8.313482	57384348.9834	57.38	12 - 16	10.8 - 14.4
4	8.313482	10.844866	73482386.0629	73.48	16 - 20	14.4 - 18
5	10.844866	13.961461	39794589.7728	39.79	20 - 24	18 - 21.6
6	13.961461	17.798561	35797833.7429	35.8	24 - 28	21.6 - 25.2
7	17.798561	22.522732	13698741.8142	13.7	28 - 32	25.2 - 28.8
8	22.522732	28.339050	1502809.22176	1.5	32 - 36	28.8 - 32.4
9	28.339050	35.5	24402.357480	0.02	36 - 40	32.4 - 36

Percentage reduction (%) of storm depth (event on April 6, 2006

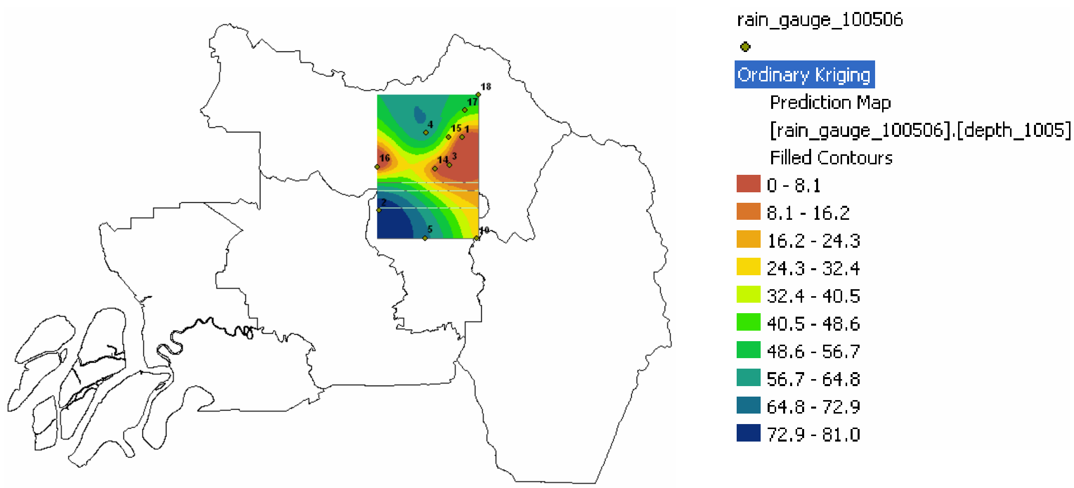
1	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP) = (average between isohyet x area between isohyet) / total areas between all pairs of neighbouring isohyets	(32.4 + 36) /2 = **34.2** 0.02 + 0 = **0.02** (34.2 x 0.02)/0.02 = **34.2**
2	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(28.8 + 32.4) / 2 = **30.6** 0.02 + 1.50 **= 1.52** [(34.2 x 0.02) + (30.6 x 1.52)] / 1.52 **= 30.6**
3	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(25.2 + 28.8) / 2 = **27** 0.02 + 1.50 + 13.7 = **15.22** [(34.2 x 0.02) + (30.6 x 1.52+ (27 x 15.22)] / 15.22 = **27.4**
4	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(21.6 + 25.2) / 2 = **23.4** 0.02 + 1.50 + 13.7 + 35.8 = **51.02** [(34.2 x 0.02) + (30.6 x 1.52+ (27 x 15.22) + (23.4 x 51.02)] / 51.02 = **24.6**
5 6	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP) Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(18 + 21.6) /2 = **19.8** 0.02 + 1.50 + 13.7 + 35.8 + 39.79 = **90.81** [(34.2 x 0.02) + (30.6 x 1.52+ (27 x 15.22) + (23.4 x 51.02) + (19.8 x 90.81)] / 90.81 = **22.5** (14.4 + 18) / 2 = **16.2** 0.02 + 1.50 + 13.7 + 35.8 + 39.79 + 73.48 = **164.29** [(34.2 x 0.02) + (30.6 x 1.52+ (27 x 15.22) + (23.4 x 51.02) + (19.8 x 90.81)+ (16.2 x 164.29)] / 164.29 = **19.7**

7	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(10.8 + 14.4)/2 = **12.6** 0.02 + 1.50 + 13.7 + 35.8 + 39.79 + 73.48 + 57.38 = **221.67** [(34.2 x 0.02) + (30.6 x 1.52+ (27 x 15.22) + (23.4 x 51.02) + (19.8 x 90.81)+ (16.2 x 164.29) + (12.6 x 221.67)]/ 221.67 = **17.8**
8	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(7.2 + 10.8) /2 = 9 0.02 + 1.50 + 13.7 + 35.8 + 39.79 + 73.48 + 57.38 + 18.47 = **240.14** [(34.2 x 0.02) + (30.6 x 1.52+ (27 x 15.22) + (23.4 x 51.02) + (19.8 x 90.81)+ (16.2 x 164.29) + (12.6 x 221.67) + (9 x 240.14)] / 240.14 = **17.2**
9	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(3.6 + 7.2) /2 = **5.4** 0.02 + 1.50 + 13.7 + 35.8 + 39.79 + 73.48 + 57.38 + 18.47 + 1.19 = **241.33** [(34.2 x 0.02) + (30.6 x 1.52+ (27 x 15.22) + (23.4 x 51.02) + (19.8 x 90.81)+ (16.2 x 164.29) + (12.6 x 221.67) + (9 x 240.14) + (5.4 x 241.33)] / 241.33 = **17.1**
10	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(0 + 3.6) /2 = 1.8 0.02 + 1.50 + 13.7 + 35.8 + 39.79 + 73.48 + 57.38 + 18.47 + 1.19 + 0 = **241.33** [(34.2 x 0.02) + (30.6 x 1.52) + (27 x 15.22) + (23.4 x 51.02) + (19.8 x 90.81)+ (16.2 x 164.29) + (12.6 x 221.67) + (9 x 240.14) + (5.4 x 241.33) + (1.8 x 241.34)] / 241.33 = **17.1**

Percentage reduction (%) of storm depth	**=** (Mean Area Precipitation, (MAP) / storm maximum)* 100
storm maximum (reference gauge)	= 35.5 mm

No.		Percentage reduction (%) of storm depth			
1	34.2 / 35.5 * 100	= 96.3 %	6	19.7 / 35.5 * 100	= 55.4%
2	30.6 / 35.5 * 100	= 86.3 %	7	17.8 / 35.5 * 100	= 50.3 %
3	27.4 / 35.5 * 100	= 77.1 %	8	17.2 / 35.5 * 100	= 48.3 %
4	24.6 / 35.5 * 100	= 69.2 %	9	17.1 / 35.5 * 100	= 48.2 %
5	22.5 / 35.5 * 100	= 63.3 %	10	17.1 / 35.5 * 100	= 48.2 %

Event on May 10, 2006



Classes	Value_Min	Value_Max	F_AREA	km_square	range	new_range
0	0	1.839708	26507794.1499	26.51	0 - 9	0 - 8.1
1	1.839708	4.243395	11369254.3081	11.37	9 - 18	8.1 - 16.2
2	4.243395	7.383955	15102385.9386	15.1	18 - 27	16.2 - 24.3
3	7.383955	11.487280	24609646.9948	24.61	27 - 36	24.3 - 32.4
4	11.487280	16.848516	25057030.7491	25.06	36 - 45	32.4 - 40.5
5	16.848516	23.853291	25714463.6504	25.71	45 - 54	40.5 - 48.6
6	23.853291	33.005440	35086115.7999	35.09	54 - 63	48.6 - 56.7
7	33.005440	44.963257	43615625.0581	43.62	63 - 72	56.7 - 64.8
8	44.963257	60.586853	15055284.0318	15.06	72 - 81	64.8 - 72.9
9	60.586853	81	19223151.4712	19.22	81 - 90	72.9 - 81.0

Percentage reduction (%) of storm depth (event on May 10, 2006)

1	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP) = (average between isohyet x area between isohyet) / total areas between all pairs of neighbouring isohyets	(72.9 + 81)/2 = **76.95** 19.22 + 0 = **19.22** (76.95 x 19.22)/ 19.22 = **76.95**
2	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(64.8 + 72.9)/2 = **68.85** 19.22 + 15.06 = **34.28** [(76.95 x 19.22) + (68.85 x 34.28)] / 34.28 = **73.4**
3	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(56.7 + 64.8) /2 = **60.75** 19.22 + 15.06+ 43.62 = **77.9** [(76.95 x 19.22) + (68.85 x 34.28) + (60.75 x 77.9)] / 77.9 = **66.3**
4	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(48.6 + 56.7)/2 = **52.65** 19.22 + 15.06+ 43.62 + 35.09 = **112.99** [(76.95 x 19.22) + (68.85 x 34.28) + (60.75 x 77.9) + (52.65 x 112.99)] / 112.99 = **62.1**
5	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(40.5 + 48.6)/2 = **44.55** 19.22 + 15.06+ 43.62 + 35.09 + 25.71 = **138.7** [(76.95 x 19.22) + (68.85 x 34.28) + (60.75 x 77.9) + (52.65 x 112.99) + (44.55 x 138.7)] / 138.7 = **58.8**
6	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(32.4 + 40.5)/2 = **36.45** 19.22 + 15.06+ 43.62 + 35.09 + 25.71 + 25.06 = **163.76** [(76.95 x 19.22) + (68.85 x 34.28) + (60.75 x 77.9) + (52.65 x 112.99) + (44.55 x 138.7))+ (36.45 x 163.76)] / 163.76 = **55.4**
7	Average between isohyet	(24.3 + 32.4)/2 = **28.35**

	Total Areas between isohyet Mean Area Precipitation, (MAP)	19.22 + 15.06+ 43.62 + 35.09 + 25.71 + 25.06+ 24.61 = **188.37** [(76.95 x 19.22) + (68.85 x 34.28) + (60.75 x 77.9) + (52.65 x 112.99) + (44.55 x 138.7))+ (36.45 x 163.76) + (28.35 x 188.37)]/ 188.37 = **51.9**
8	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(16.2 + 24.3)/2 = 20.25 19.22 + 15.06+ 43.62 + 35.09 + 25.71 + 25.06+ 24.61 + 15.1 = **203.47** [(76.95 x 19.22) + (68.85 x 34.28) + (60.75 x 77.9) + (52.65 x 112.99) + (44.55 x 138.7))+ (36.45 x 163.76) + (28.35 x 188.37) + (20.25 x 203.47)] / 203.47 = **49.5**
9	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(8.1 + 16.2)/2 = **12.15** 19.22 + 15.06+ 43.62 + 35.09 + 25.71 + 25.06+ 24.61 + 15.1 + 11.37 = **214.84** [(76.95 x 19.22) + (68.85 x 34.28) + (60.75 x 77.9) + (52.65 x 112.99) + (44.55 x 138.7))+ (36.45 x 163.76) + (28.35 x 188.37) + (20.25 x 203.47) + (12.15 x 214.84)] / 214.84 = **47.5**
10	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(0 + 8.1)/2 = 8.1 19.22 + 15.06+ 43.62 + 35.09 + 25.71 + 25.06+ 24.61 + 15.1 + 11.37 + 26.51 = **241.35** [(76.95 x 19.22) + (68.85 x 34.28) + (60.75 x 77.9) + (52.65 x 112.99) + (44.55 x 138.7))+ (36.45 x 163.76) + (28.35 x 188.37) + (20.25 x 203.47) + (12.15 x 214.84+ (8.1 x 241.35)] / 241.35 = **42.8**

Percentage reduction (%) of storm depth	**=** (Mean Area Precipitation, (MAP) / storm maximum)* 100
storm maximum (reference gauge)	= 81.0 mm

No.	Percentage reduction (%) of storm depth		
1	76.95 / 81.0 * 100 = 95.0 %	6	55.4 / 81.0 * 100 = 68.4%
2	73.4 / 81.0 * 100 = 90.6 %	7	51.9 / 81.0 * 100 = 64.0 %
3	66.3 / 81.0 * 100 = 81.9 %	8	49.5 / 81.0 * 100 = 61.1 %
4	62.1 / 81.0 * 100 = 76.6 %	9	47.5 / 81.0 * 100 = 58.7 %
5	58.8 / 81.0 * 100 = 72.6 %	10	42.8 / 81.0 * 100 = 52.8 %

Percentage reduction (%) of storm depth (event on June 10, 2003)

1	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP) = (average between isohyet x area between isohyet) / total areas between all pairs of neighbouring isohyets	(117 + 130)/2 = **123.5** **0.02** (123.5 x 0.02) / 0.02 = **123.5**
2	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(104 + 117)/2 = **110.5** 0.02 + 0 **= 0.02** [(123.5 x 0.02) + (110.5 x 0.02)] / 0.02 = **123.5**
3	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(91 + 106)/2 = **98.5** 0.02 + 0 + 1.25 = **1.27** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27)] / 1.27 = **98.9**
4	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(78 + 91) /2 = **84.5** 0.02 + 0 + 1.25+ 24.35 = **25.62** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27) + (84.5 x 25.62)] / 25.62 = **85.2**
5	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(65 + 78) /2 = **71.5** 0.02 + 0 + 1.25+ 24.35 + 77.35 = **102.97** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27) + (84.5 x 25.62) + (71.5 x 102.97)] / 102.97 = **74.9**
6	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(52 + 65)/2 = **58.5** 0.02 + 0 + 1.25+ 24.35 + 77.35 + 80.59 = **183.56** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27) + (84.5 x 25.62) + (71.5 x 102.97)+ (58.5 x 183.56)] / 183.56 = **67.7**
7	Average between isohyet	(39 + 52)/2 = **45.5**

	Total Areas between isohyet Mean Area Precipitation, (MAP)	0.02 + 0 + 1.25+ 24.35 + 77.35 + 80.59 + 36.34 = **219.9** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27) + (84.5 x 25.62) + (71.5 x 102.97)+ (58.5 x 183.56) + (219.9 x 45.5)]/ 219.9 = **64.0**
8	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(26 + 39)/2 = 32.5 0.02 + 0 + 1.25+ 24.35 + 77.35 + 80.59 + 36.34 + 19.41 = **239.31** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27) + (84.5 x 25.62) + (71.5 x 102.97)+ (58.5 x 183.56) + (219.9 x 45.5)+ (32.5 x 239.31)] / 239.31 = **61.5**
9	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(13 + 26) /2 = **19.5** 0.02 + 0 + 1.25+ 24.35 + 77.35 + 80.59 + 36.34 + 19.41 + 2.06 = **241.37** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27) + (84.5 x 25.62) + (71.5 x 102.97)+ (58.5 x 183.56) + (219.9 x 45.5)+ (32.5 x 239.31) + (19.5x 241.37)] / 241.37= **61.1**
10	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(0 + 13)/2 = 6.5 0.02 + 0 + 1.25+ 24.35 + 77.35 + 80.59 + 36.34 + 19.41 + 2.06 + 0 = **241.37** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27) + (84.5 x 25.62) + (71.5 x 102.97)+ (58.5 x 183.56) + (219.9 x 45.5)+ (32.5 x 239.31) + (19.5x 241.37) + (6.5 x 241.37)] / 241.37 = **61.1**

Percentage reduction (%) of storm depth	**=** (Mean Area Precipitation, (MAP) / storm maximum)* 100
storm maximum (reference gauge)	= 129.5 mm

No.	**Percentage reduction (%) of storm depth**		
1	123.5 / 129.5 * 100 = 95.4 %	6	67.7 / 129.5 * 100 = 52.3 %
2	123.5 / 129.5 * 100 = 95.4 %	7	64.0 / 129.5 * 100 = 49.4 %
3	98.9 / 129.5 * 100 = 76.4 %	8	61.5 / 129.5 * 100 = 47.5 %
4	85.2 / 129.5 * 100 = 65.8 %	9	61.1 / 129.5 * 100 = 47.2 %
5	74.9 / 129.5 * 100 = 57.8 %	10	61.1 / 129.5 * 100 = 47.2 %

Percentage reduction (%) of storm depth (event on February, 26 2006)

1	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP) = (average between isohyet x area between isohyet) / total areas between all pairs of neighbouring isohyets	(67.5 + 75)/2 = **71.25** **19.7** (71.25 x 19.7) / 19.7 = **71.25**
2	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(60 + 67.5)/2 = **110.5** 19.7 + 36.49 **= 56.19** [(71.25 x 19.7)) + (110.5 x 56.19)] / 56.19 **= 66.4**
3	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(52.5 + 60)/2 = **56.25** 19.7 + 36.49 + 24.86 = **81.05** [(71.25 x 19.7)) + (110.5 x 56.19)+ (56.25 x 81.05)] / 81.05 = **63.3**
4	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(45 + 52.5) /2 = **48.75** 19.7 + 36.49 + 24.86 + 20.92 = **101.97** [(71.25 x 19.7)) + (110.5 x 56.19)+ (56.25 x 81.05) + (48.75 x 101.97)] / 101.97 = **60.3**
5	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(37.5 + 45) /2 = **41.25** 19.7 + 36.49 + 24.86 + 20.92 + 31.82 = **133.79** [(71.25 x 19.7)) + (110.5 x 56.19)+ (56.25 x 81.05) + (48.75 x 101.97)+ (41.25 x 133.79)] / 133.79 = **55.8**
6	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(30 + 37.5)/2 = **33.75** 19.7 + 36.49 + 24.86 + 20.92 + 31.82 + 40.39 = **174.18** [(71.25 x 19.7)) + (110.5 x 56.19)+ (56.25 x 81.05) + (48.75 x 101.97)+ (41.25 x 133.79)+ (33.75 x 174.18)] / 174.18 = **50.7**
7	Average between isohyet	(22.5 + 30)/2 = **26.25**

	Total Areas between isohyet Mean Area Precipitation, (MAP)	19.7 + 36.49 + 24.86 + 20.92 + 31.82 + 40.39 + 43.85 = **218.03** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27) + (84.5 x 25.62) + (71.5 x 102.97)+ (58.5 x 183.56) + (219.9 x 45.5)]/ 219.9 = **64.0**
8	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(15 + 22.5)/2 = 18.75 19.7 + 36.49 + 24.86 + 20.92 + 31.82 + 40.39 + 43.85 + 20.6 = **238.63** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27) + (84.5 x 25.62) + (71.5 x 102.97)+ (58.5 x 183.56) + (219.9 x 45.5)+ (18.75 x 238.63)] / 238.63 = **43.4**
9	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(7.5 + 15) /2 = **11.25** 19.7 + 36.49 + 24.86 + 20.92 + 31.82 + 40.39 + 43.85 + 20.6 + 2.43 = **241.06** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27) + (84.5 x 25.62) + (71.5 x 102.97)+ (58.5 x 183.56) + (219.9 x 45.5)+ (18.75 x 238.63)+ (11.25x 241.06)] / 241.06 = **43.1**
10	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(0 + 7.5) / 2 = 3.75 19.7 + 36.49 + 24.86 + 20.92 + 31.82 + 40.39 + 43.85 + 20.6 + 2.43 + 0.28 = **241.34** [(123.5 x 0.02) + (110.5 x 0.02) + (98.5 x 1.27) + (84.5 x 25.62) + (71.5 x 102.97)+ (58.5 x 183.56) + (219.9 x 45.5)+ (18.75 x 238.63)+ (11.25x 241.06)+ (3.75 x 241.34)] / 241.34 = **43.0**

Percentage reduction (%) of storm depth	**=** (Mean Area Precipitation, (MAP) / storm maximum)* 100
storm maximum (reference gauge)	= 72.5 mm

No.	**Percentage reduction (%) of storm depth**		
1	71.25 / 72.5 * 100 = 98.3 %	6	50.7 / 72.5 * 100 = 69.9 %
2	66.4 / 72.5 * 100 = 91.6 %	7	45.8 / 72.5 * 100 = 63.1 %
3	63.3 / 72.5 * 100 = 87.3 %	8	43.4 / 72.5 * 100 = 59.9 %
4	60.3 / 72.5 * 100 = 83.2 %	9	43.1 / 72.5 * 100 = 59.4 %
5	55.8 / 72.5 * 100 = 76.9 %	10	43 / 72.5 * 100 = 59.4 %

Percentage reduction (%) of storm depth (event on November 5, 2004)

1	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP) = (average between isohyet x area between isohyet) / total areas between all pairs of neighbouring isohyets	(85.5 + 95)/2 = **90.25** **0.02** (85.5 x 0.02) / 0.02 = **90.25**
2	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(76 + 85.5)/2 = **80.75** 0.02 + 11.01 **= 11.03** [(85.5 x 0.02) + (80.75 x 11.03)] / 11.03 **= 80.8**
3	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(66.5 + 76)/2 = **71.25** 0.02 + 11.01 + 13.78 = **24.81** [(85.5 x 0.02) + (80.75 x 11.03)+ (71.25 x 24.81)] / 24.81 = **75.5**
4	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(57 + 66.5) /2 = **61.75** 0.02 + 11.01 + 13.78 + 21.63 = **46.44** [(85.5 x 0.02) + (80.75 x 11.03)+ (71.25 x 24.81) + (61.75 x 46.44)] / 46.44 = **69.1**
5	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(47.5 + 57) /2 = **52.25** 0.02 + 11.01 + 13.78 + 21.63 + 41.58 = **88.02** [(85.5 x 0.02) + (80.75 x 11.03)+ (71.25 x 24.81) + (61.75 x 46.44)+ (52.25x 88.02)] / 88.02 = **61.1**
6	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(38 + 47.5)/2 = **42.75** 0.02 + 11.01 + 13.78 + 21.63 + 41.58 + 36.41 = **124.23** [(85.5 x 0.02) + (80.75 x 11.03)+ (71.25 x 24.81) + (61.75 x 46.44)+ (52.25x 88.02)+ (22.75 x 124.23)] / 124.23 = **55.8**
7	Average between isohyet	(28.5 + 38)/2 = **33.25**

	Total Areas between isohyet Mean Area Precipitation, (MAP)	0.02 + 11.01 + 13.78 + 21.63 + 41.58 + 36.41 + 32.04 = **156.27** [(85.5 x 0.02) + (80.75 x 11.03)+ (71.25 x 24.81) + (61.75 x 46.44)+ (52.25x 88.02)+ (22.75 x 124.23) + (33.25 x 156.27)]/ 156.27 = **51.2**
8	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(19 + 28.5)/2 = 23.75 0.02 + 11.01 + 13.78 + 21.63 + 41.58 + 36.41 + 32.04 + 34.84 = **191.11** [(85.5 x 0.02) + (80.75 x 11.03)+ (71.25 x 24.81) + (61.75 x 46.44)+ (52.25x 88.02)+ (22.75 x 124.23) + (33.25 x 156.27)+ (23.75 x 191.11)] / 191.11 = **46.2**
9	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(9.5 + 19) /2 = **12.25** 0.02 + 11.01 + 13.78 + 21.63 + 41.58 + 36.41 + 32.04 + 34.84 + 45.75 = **236.86** [(85.5 x 0.02) + (80.75 x 11.03)+ (71.25 x 24.81) + (61.75 x 46.44)+ (52.25x 88.02)+ (22.75 x 124.23) + (33.25 x 156.27)+ (23.75 x 191.11)+ (12.25x 236.86)] / 236.86 = **40.0**
10	Average between isohyet Total Areas between isohyet Mean Area Precipitation, (MAP)	(0 + 9.5) / 2 = 4.75 0.02 + 11.01 + 13.78 + 21.63 + 41.58 + 36.41 + 32.04 + 34.84 + 45.75 + 4.5 = **241.36** [(85.5 x 0.02) + (80.75 x 11.03)+ (71.25 x 24.81) + (61.75 x 46.44)+ (52.25x 88.02)+ (22.75 x 124.23) + (33.25 x 156.27)+ (23.75 x 191.11)+ (12.25x 236.86)+ (4.75 x 241.36)] / 241.36 = **39.3**

Percentage reduction (%) of storm depth	**=** (Mean Area Precipitation, (MAP) / storm maximum)* 100
storm maximum (reference gauge)	= 92 mm

No.	**Percentage reduction (%) of storm depth**		
1	90.25 / 92* 100 = 98.1 %	6	55.8 / 92* 100 = 60.6 %
2	80.8 / 92* 100 = 87.8 %	7	51.2 / 92* 100 = 55.6 %
3	75.5 / 92* 100 = 82.0 %	8	46.2 / 92* 100 = 50.2 %
4	69.1 / 92* 100 = 75.1 %	9	40.0 / 92* 100 = 43.5 %
5	61.1 / 92* 100 = 66.4 %	10	39.3 / 92* 100 = 42.8 %

APPENDIX D

SAMPLE MOMENTS

1. The sample moments used autocorrelations in the estimation of parameters procedures.

Months	1-hr-mean $\hat{\mu}(1)$	1-hr-var $\hat{\gamma}(1)$	1-hr-auto $\hat{\rho}(1,1)$	6-hr-var $\hat{\gamma}(6)$	6-hr-auto $\hat{\rho}(6,1)$	24-hr-var $\hat{\gamma}(24)$	24-hr-auto $\hat{\rho}(24,1)$	Dry-Prob $\hat{\phi}(24)$
Jan	0.099	1.034	0.0322	11.19	0.0038	47.68	0.0105	0.71
Feb	0.261	4.144	0.3336	42.46	0.0936	209.02	0.2485	0.61
March	0.279	3.842	0.3478	40.77	0.0039	149.42	0.0212	0.50
April	0.277	3.532	0.3637	41.34	0.0067	180.16	0.0850	0.48
May	0.380	4.312	0.4125	48.76	0.1616	265.50	0.1326	0.41
June	0.156	2.085	0.4211	23.58	0.1012	110.04	0.1354	0.67
July	0.240	3.045	0.5101	39.44	0.0718	185.87	0.0178	0.55
August	0.240	3.874	0.3173	36.82	0.0679	189.67	0.0033	0.61
Sept	0.394	4.996	0.3582	51.49	0.0652	214.54	0.0348	0.38
Oct	0.356	4.796	0.3143	44.19	0.0560	191.81	0.0320	0.40
Nov	0.416	4.130	0.4061	48.90	0.1192	241.48	0.0510	0.31
Dec	0.170	1.662	0.4104	18.80	0.1771	118.59	0.1132	0.58

2. The transition probabilities used the fitting procedures

Months	*p00*-hourly $\hat{\phi}_{DD}(1)$	*p11*-hourly $\hat{\phi}_{WW}(1)$	*p00*-daily $\hat{\phi}_{DD}(24)$	*p11*-daily $\hat{\phi}_{WW}(24)$
Jan	0.9811	0.5485	0.7798	0.4615
Feb	0.97498	0.6141	0.7399	0.5872
March	0.96639	0.50943	0.5742	0.5779
April	0.9671	0.6454	0.5556	0.5833
May	0.9612	0.71445	0.4762	0.6393
June	0.97962	0.4946	0.7376	0.4592
July	0.97214	0.5924	0.614	0.5217
August	0.977	0.58987	0.709	0.5417
Sept	0.9552	0.73833	0.4087	0.627
Oct	0.95274	0.593023	0.5203	0.6882
Nov	0.948515	0.682	0.4574	0.7476
Dec	0.96903	0.62032	0.6927	0.5846

APPENDIX E

DESCRIPTIVE STATISTICS

1. Hourly Descriptive Statistics for 10-year period (1981-1990)

Months	Jan	Feb	Mac	Apr	May	Jun
Count	7440	6768	7440	7200	7440	7200
Sum	734.2	1764.9	2074.1	1992.1	2827.6	1122.1
Mean	0.0987	0.2608	0.2788	0.2767	0.3801	0.1558
StdDev	1.0167	2.0357	1.9601	1.8793	2.0766	1.4439
Kurtosis	318.9105	338.2578	193.9259	210.0779	180.3225	319.4403
Skewness	16.5887	15.6453	12.0520	12.3997	11.0559	15.4455
Maximum	26.4	58.2	47.7	53.1	58.1	47.8
Minimum	0	0	0	0	0	0
Correlation	0.3221	0.3331	0.3478	0.3630	0.4133	0.4205
AutoCovar	0.3330	1.3822	1.3361	1.2817	1.7788	0.8766
AutoCorrel	0.3221	0.3336	0.3478	0.3630	0.4125	0.4205
P(Dry)	0.9598	0.9391	0.9359	0.9150	0.8800	0.9613
Months	July	Aug	Sept	Oct	Nov	Dec
Count	7440	7440	7200	7440	7200	7440
Sum	1788.4	1783.3	2836.1	2648	2994.6	1288.3
Mean	0.2404	0.2397	0.3939	0.3559	0.4159	0.1732
StdDev	1.7450	1.9684	2.2352	2.1899	2.0322	1.2895
Kurtosis	184.8828	289.6727	148.5206	204.2800	111.2652	288.5568
Skewness	12.2051	14.8370	10.5626	12.1538	9.1723	14.6195
Maximum	41.6	54.7	52.1	56.5	44.3	41.7
Minimum	0	0	0	0	0	0
Correlation	0.5101	0.3173	0.3577	0.3143	0.4055	0.4104
AutoCovar	1.5532	1.2292	1.7869	1.5073	1.6745	0.6823
AutoCorrel	0.5101	0.3173	0.3577	0.3143	0.4055	0.4104
P(Dry)	0.9360	0.9469	0.8721	0.8960	0.8607	0.9246

2. Daily Descriptive Statistics for 10-year period (1981-1990)

Months	Jan	Feb	Mac	Apr	May	Jun
Count	310	282	310	300	310	300
Sum	734.2	1764.9	2074.1	1992.1	2827.6	1122.1
Mean	2.368	6.259	6.691	6.640	9.121	3.740
StdDev	6.950	13.655	12.366	13.650	15.898	10.476
Kurtosis	19.196	9.389	7.519	32.541	7.109	20.004
Skewness	4.123	2.956	2.640	4.607	2.561	4.106
Maximum	54.6	74.5	65.5	138	91.4	79.3
Minimum	0	0	0	0	0	0
Q1	0	0	0	0	0	0
Q3	0.5	5.225	8.475	8.55	10.35	1.2
Correlation	0.0717	0.2035	-0.0443	-0.0095	0.1638	0.1252
AutoCovar	3.4526	38.8845	-6.6286	-1.7599	40.5327	13.6901
AutoCorrel	0.0717	0.2093	-0.0435	-0.0095	0.1609	0.1252
P(Dry)	0.6903	0.5887	0.4516	0.4533	0.3613	0.6633
Months	July	Aug	Sept	Oct	Nov	Dec
Count	310	310	300	310	300	310
Sum	1788.4	1783.3	2836.1	2648	2994.6	1288.3
Mean	5.7690	5.7526	9.4537	8.5419	9.9820	4.1558
StdDev	13.1371	12.8800	14.7976	13.0897	14.8121	9.6160
Kurtosis	18.9313	8.8279	8.5869	5.5017	3.7666	17.3644
Skewness	3.8906	2.9176	2.5594	2.2304	1.9683	3.7832
Maximum	100.2	72.2	101.5	72.1	80.4	75.7
Minimum	0	0	0	0	0	0
Q1	0	0	0	0	0	0
Q3	5.1	3.5	13.2	12.45	13.7	3.6
Correlation	0.0024	0.0134	0.1793	0.0988	0.1041	0.2609
AutoCovar	0.4183	2.2203	38.9845	16.7983	22.7350	24.0414
AutoCorrel	0.0024	0.0134	0.1786	0.0984	0.1040	0.2608
P(Dry)	0.5161	0.6000	0.3167	0.3677	0.3167	0.5355

3. Monthly Descriptive Statistics for 10-year period (1981-1990)

Months	Jan	Feb	Mac	Apr	May	Jun
Count	10	10	10	10	10	10
Sum	734.2	1758.4	2074.1	1992.1	2827.6	1122.1
Mean	73.42	175.84	207.41	199.21	282.76	112.21
StdDev	33.4735	99.5972	44.1827	65.6461	152.4618	78.2572
Kurtosis	-0.9501	-1.4301	-1.0531	-0.8662	2.4504	3.9360
Skewness	0.3406	-0.0944	-0.1810	0.0117	1.4666	1.6950
Maximum	129.9	310.3	266.2	293.7	636	303.3
Minimum	28.2	25.9	133	92.5	139.8	24
Q1	47.475	87.975	183.775	161.15	183.3	64
Q3	95.075	262.15	247.725	248.35	330.7	132.65
Correlation	-0.4054	0.0802	-0.2302	-0.8013	0.7289	-0.3696
AutoCovar	-389.5374	310.7776	-398.4031	-3029.8946	8996.6520	-1956.1404
AutoCorrel	-0.3863	0.0348	-0.2268	-0.7812	0.4300	-0.3549
Months	July	Aug	Sept	Oct	Nov	Dec
Count	10	10	10	10	10	10
Sum	1788.4	1783.3	2836.1	2648	2994.6	1288.3
Mean	178.84	178.33	283.61	264.8	299.46	128.83
StdDev	62.5591	65.3600	106.8322	88.6362	126.4881	59.9286
Kurtosis	-0.3534	-0.8876	-0.8577	1.8503	1.0309	-1.6579
Skewness	-0.4069	-0.4690	0.3758	-0.7333	0.4750	0.1940
Maximum	257.5	262.3	467.8	400.4	548.7	216.8
Minimum	62.1	72.6	140.9	73.6	85.4	54.2
Q1	150.875	140.85	190.175	236.425	233.625	74
Q3	235.975	234.125	341.475	321.975	355.15	179.6
Correlation	-0.2666	0.0000	0.0762	-0.2798	0.3374	0.5385
AutoCovar	-922.2052	0.0450	671.1214	-1966.6590	3894.6064	1608.3358
AutoCorrel	-0.2618	0.0000	0.0653	-0.2781	0.2705	0.4976

4. Hourly Descriptive Statistics for 10-year period (1991-2000)

Month	Jan	Feb	Mar	Apr	May	Jun
Count	7224	6695	7344	7200	7248	6816
Sum	763.3	1249.7	2385.4	2395.7	2288.5	2056.4
Mean	0.1057	0.1867	0.3248	0.3327	0.3157	0.3017
StdDev	0.9399	1.4992	2.1166	1.9631	2.1067	2.0099
Kurtosis	443.1636	295.3575	128.5620	110.4368	151.0954	151.3543
Skewness	17.6961	14.3928	10.2044	9.2921	10.7962	10.6382
Maximum	33.7	49.2	45	36.8	46.5	46.1
Minimum	0	0	0	0	0	0
Correlation	0.4564	0.5569	0.4732	0.5291	0.5051	0.5218
AutoCovar	0.4034	1.2530	2.1210	2.0385	2.2440	2.1129
AutoCorrel	0.4568	0.5576	0.4735	0.5290	0.5057	0.5231
P(Dry)	0.9592	0.9535	0.9308	0.9226	0.9332	0.9382
Month	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Count	7416	7440	7200	7440	7200	7128
Sum	1864.5	2323.4	2443.6	2506	2544.9	2093.8
Mean	0.2514	0.3123	0.3394	0.3368	0.3535	0.2937
StdDev	1.9336	2.1041	2.0804	1.9364	2.1569	1.5933
Kurtosis	299.4920	175.4460	325.2916	151.2159	180.2724	104.5280
Skewness	14.5308	11.4109	13.4449	10.3552	11.6361	9.0187
Maximum	60.5	50	77.1	46.2	51.9	31.8
Minimum	0	0	0	0	0	0
Correlation	0.4956	0.4923	0.4323	0.4644	0.5303	0.5619
AutoCovar	1.8529	2.1791	1.8705	1.7413	2.4667	1.4301
AutoCorrel	0.4957	0.4923	0.4322	0.4644	0.5303	0.5634
P(Dry)	0.9448	0.9344	0.9254	0.9073	0.9024	0.9046

5. Daily Descriptive Statistics for 10-year period (1991-2000)

Month	Jan	Feb	Mar	Apr	May	Jun
Count	1204	1115	1224	1200	1208	1136
Sum	763.3	1249.1	2385.4	2395.7	2288.5	2056.4
Mean	0.6340	1.1203	1.9489	1.9964	1.8945	1.8102
StdDev	3.0949	5.6158	7.2172	6.9531	7.5369	7.4904
Kurtosis	75.4670	88.1790	33.2187	43.8413	64.9904	44.3015
Skewness	7.8223	8.4371	5.3790	5.8574	6.8706	6.1173
Maximum	40.8	83.9	65.9	77.4	113.9	82
Minimum	0	0	0	0	0	0
Q1	0	0	0	0	0	0
Q3	0	0	0	0	0	0
Correlation	0.2176	0.1058	0.0715	0.0587	0.0575	0.0416
AutoCovar	2.0945	3.3860	3.7745	2.8339	3.3584	2.5039
AutoCorrel	0.2188	0.1075	0.0725	0.0587	0.0592	0.0447
P(Dry)	0.8729	0.8574	0.7933	0.7692	0.8071	0.8345
Month	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Count	1236	1240	1200	1240	1200	1188
Sum	1864.5	2323.4	2443.6	2506	2544.9	2093.8
Mean	1.5085	1.8737	2.0363	2.0210	2.1208	1.7625
StdDev	6.9227	7.2516	7.0031	6.6877	8.2763	6.1050
Kurtosis	73.4585	58.8447	48.9746	42.3570	87.1590	43.9447
Skewness	7.6791	6.6143	5.9283	5.6477	8.0546	5.9323
Maximum	95.1	99.5	90.7	83.6	129	65.9
Minimum	0	0	0	0	0	0
Q1	0	0	0	0	0	0
Q3	0	0	0	0.5	0.5	0.5
Correlation	0.0314	0.0522	0.0345	0.0304	0.0441	0.1776
AutoCovar	1.5100	2.7446	1.6929	1.3570	3.0162	6.7517
AutoCorrel	0.0315	0.0522	0.0345	0.0304	0.0441	0.1813
P(Dry)	0.8406	0.7984	0.7933	0.7395	0.7250	0.7407

6. Daily Descriptive Statistics for 10-year period (1991-2000)

Month	Jan	Feb	Mar	Apr	May	Jun
Count	301	278	306	300	302	284
Sum	763.3	1249.1	2385.4	2395.7	2288.5	2056.4
Mean	2.5359	4.4932	7.7954	7.9857	7.5778	7.2408
StdDev	7.2820	11.7783	15.3228	14.0790	15.6431	15.2831
Kurtosis	26.2550	18.1157	6.3706	9.5604	17.6960	10.0186
Skewness	4.6183	3.9893	2.5414	2.7763	3.5557	2.9483
Maximum	62.9	84.9	76.1	95.8	126.8	103.4
Minimum	0	0	0	0	0	0
Q1	0	0	0	0	0	0
Q3	1	2	7.1	11.275	7.3	6.5
Correlation	0.0828	0.1344	0.0418	0.0304	0.0136	0.1453
AutoCovar	4.4482	19.2440	10.5628	5.9968	4.7521	36.7981
AutoCorrel	0.0842	0.1392	0.0451	0.0304	0.0195	0.1581
P(Dry)	0.6578	0.6259	0.4869	0.3867	0.4834	0.5493
Month	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Count	309	310	300	310	300	297
Sum	1864.5	2323.4	2443.6	2506	2544.9	2093.8
Mean	6.0340	7.4948	8.1453	8.0839	8.4830	7.0498
StdDev	13.8981	14.9427	13.9312	14.0516	16.5787	13.7569
Kurtosis	15.1119	12.3061	9.2290	9.1925	17.6850	16.3232
Skewness	3.6093	3.1856	2.6073	2.7909	3.6971	3.5465
Maximum	95.1	99.5	95.1	87	129	107.4
Minimum	0	0	0	0	0	0
Q1	0	0	0	0	0	0
Q3	5.4	7.4	12.525	9.6	8.5	7.5
Correlation	0.0385	0.0033	0.0625	0.0936	0.0452	0.1946
AutoCovar	7.1410	0.7330	12.0705	18.4056	12.3541	38.0475
AutoCorrel	0.0371	0.0033	0.0624	0.0935	0.0451	0.2017
P(Dry)	0.5534	0.4677	0.4733	0.3516	0.3367	0.4007

APPENDIX F

ROOT MEAN SQUARE ERROR (RMSE)

1. Calibration Period (1981-1990)

A. NSRP MODELS

One-Hour Mean

Months	**Jan**	**Feb**	**Mac**	**April**	**May**	**June**	**July**
MEXP	8E-07	1E-04	5E-05	6E-05	2E-05	4E-05	8E-07
EXP	1E-05	2E-06	6E-04	1E-06	3E-07	8E-05	5E-06
EXPTRAN	5E-04	6E-04	1E-03	2E-03	3E-03	2E-04	7E-04
MEXPTRAN	1E-05	2E-04	8E-05	1E-07	1E-09	3E-05	3E-05
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	8E-05	2E-05	1E-04	2E-05	3E-05	5E-05	7E-03
EXP	2E-05	2E-04	2E-06	6E-06	9E-07	8E-05	9E-03
EXPTRAN	1E-03	5E-03	2E-03	1E-03	9E-04	2E-03	4E-02
MEXPTRAN	3E-04	7E-05	5E-04	2E-04	4E-07	1E-04	1E-04

One-Hour Variance

Months	**Jan**	**Feb**	**Mac**	**April**	**May**	**June**	**July**
MEXP	0.0012	0.0034	2.7114	8.2375	0.2155	0.1147	0.2854
EXP	0.0085	0.0970	8.5290	2.9339	0.1100	0.0158	0.0919
EXPTRAN	0.0131	0.5517	0.0025	0.0014	0.0380	0.0358	0.0720
MEXPTRAN	0.0005	0.0013	0.0533	0.0165	0.0773	0.0030	0.1053
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.1010	0.0010	0.2575	0.0048	0.0000	0.9945	0.9972
EXP	0.5745	0.0912	0.0743	0.0043	0.0161	1.0455	1.0225
EXPTRAN	0.3132	0.1498	0.1990	0.0428	0.0485	0.1223	0.3497
MEXPTRAN	0.0358	0.1857	0.0739	0.1526	0.0315	0.0614	0.2478

One-Hour Autocorrelation

Months	Jan	Feb	Mac	April	May	June	July
MEXP	9.9E-05	8.7E-04	3.7E-02	4.3E-02	8.4E-04	3.7E-03	8.2E-04
EXP	5.6E-04	2.1E-04	4.3E-02	3.9E-02	5.5E-05	1.1E-03	1.2E-03
EXPTRAN	1.5E-02	8.5E-04	5.4E-03	8.4E-03	5.0E-04	2.0E-02	3.0E-04
MEXPTRAN	2.4E-03	1.1E-02	4.5E-03	1.9E-04	3.8E-04	8.1E-03	1.0E-03
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.0324	0.0008	0.0000	0.0003	0.0074	0.0106	0.1029
EXP	0.0136	0.0008	0.0002	0.0000	0.0010	0.0084	0.0916
EXPTRAN	0.0005	0.0015	0.0007	0.0004	0.0077	0.0052	0.0718
MEXPTRAN	0.0087	0.0008	0.0125	0.0010	0.0022	0.0044	0.0661

One-Hour Coefficient of Skewness

Months	Jan	Feb	Mac	April	May	June	July
MEXP	5.5188	0.0605	20.8724	16.7787	0.3796	0.9768	0.4892
EXP	2.6004	6.1697	36.1402	7.0940	3.2045	0.5827	0.1084
EXPTRAN	11.1802	19.5990	2.2655	10.2236	10.7896	2.2902	4.1962
MEXPTRAN	3.5994	9.1505	0.0032	0.4543	1.2209	3.5431	2.3760
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	7.2817	0.2640	0.9907	0.0346	2.3593	4.6672	2.1604
EXP	0.1594	0.5105	0.1065	1.1720	7.9385	5.4822	2.3414
EXPTRAN	10.1742	7.6893	8.9439	4.0490	15.8479	8.9374	2.9895
MEXPTRAN	0.6519	0.8343	9.6159	0.0810	4.3255	2.9880	1.7286

P00(1)

Months	Jan	Feb	Mac	April	May	June	July
MEXP	3.1E-06	5.6E-05	7.3E-06	2.4E-05	6.9E-04	2.8E-05	3.2E-06
EXP	1.2E-05	1.0E-04	6.4E-05	5.0E-05	4.1E-04	2.6E-06	2.7E-05
EXPTRAN	3.1E-07	1.6E-05	6.2E-06	6.5E-07	4.7E-05	3.0E-07	7.5E-06
MEXPTRAN	5.1E-05	2.8E-04	4.5E-05	1.0E-04	1.9E-04	6.1E-05	4.5E-05
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.0002	0.0002	0.0000	0.0003	0.0000	0.0001	0.0115
EXP	0.0000	0.0000	0.0000	0.0003	0.0000	0.0001	0.0092
EXPTRAN	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0043
MEXPTRAN	0.0002	0.0002	0.0003	0.0007	0.0001	0.0002	0.0138

P10(1)

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.0119	0.0499	0.0455	0.1067	0.0366	0.0015	0.0019
EXP	0.0167	0.0429	0.0391	0.1157	0.0496	0.0044	0.0034
EXPTRAN	0.0039	0.0001	0.0031	0.0001	0.0026	0.0048	0.0022
MEXPTRAN	0.0042	0.0011	0.0066	0.0008	0.0001	0.0120	0.0042
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.0743	0.0692	0.0247	0.0254	0.0178	0.0388	0.1969
EXP	0.0573	0.0324	0.0183	0.0269	0.0163	0.0353	0.1878
EXPTRAN	0.0001	0.0043	0.0000	0.0001	0.0000	0.0018	0.0422
MEXPTRAN	0.0014	0.0002	0.0017	0.0007	0.0020	0.0029	0.0542

Probabilty of dry hours

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.0001	0.0001	0.0002	0.0011	0.0000	0.0000	0.0001
EXP	0.0002	0.0000	0.0008	0.0009	0.0003	0.0000	0.0003
EXPTRAN	0.0000	0.0000	0.0002	0.0000	0.0000	0.0001	0.0000
MEXPTRAN	0.0005	0.0023	0.0006	0.0010	0.0015	0.0008	0.0008
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	4.6E-06	1.1E-03	6.2E-04	1.9E-04	7.9E-05	3.0E-04	1.7E-02
EXP	1.0E-04	1.6E-03	7.3E-04	2.2E-04	4.0E-04	4.6E-04	2.1E-02
EXPTRAN	1.4E-04	4.9E-06	1.3E-05	1.2E-06	2.0E-06	4.0E-05	6.4E-03
MEXPTRAN	1.8E-03	1.2E-03	2.5E-03	5.0E-03	9.3E-04	1.6E-03	4.0E-02

Six-Hour Mean

Months	Jan	Feb	Mac	April	May	June	July
MEXP	2.8E-05	5.6E-03	1.7E-03	3.1E-03	5.9E-04	1.7E-03	3.2E-05
EXP	5.0E-04	1.9E-05	2.2E-02	2.5E-04	1.1E-05	2.5E-03	1.8E-04
EXPTRAN	1.9E-02	2.1E-02	3.7E-02	9.3E-02	1.1E-01	7.4E-03	2.3E-02
MEXPTRAN	5.3E-04	5.2E-03	3.0E-03	4.8E-05	9.0E-08	8.5E-04	9.8E-04
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.0030	0.0008	0.0054	0.0014	0.0011	0.0020	0.0451
EXP	0.0006	0.0093	0.0001	0.0065	0.0000	0.0035	0.0588
EXPTRAN	0.0501	0.1719	0.0939	0.0832	0.0335	0.0617	0.2484
MEXPTRAN	0.0119	0.0026	0.0163	0.0248	0.0000	0.0055	0.0743

Six-Hour Variance

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.1534	10.1201	13.6186	48.9691	14.8815	8.9813	15.2493
EXP	0.6852	14.4988	123.4543	0.4206	15.3195	0.2249	3.5296
EXPTRAN	0.1477	4.2925	7.2611	5.1217	2.3139	8.7492	1.3521
MEXPTRAN	1.0733	13.2151	19.5160	13.8013	2.3815	1.2803	30.5698
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	32.093	1.4543	61.5103	3.4439	4.3421	17.9014	4.2310
EXP	4.4114	18.7241	81.2255	2.2079	2.3810	22.2569	4.7177
EXPTRAN	0.1119	7.9963	4.4055	0.6754	0.2768	3.5587	1.8864
MEXPTRAN	4.3609	5.8018	8.4801	0.0367	2.7692	8.6072	2.9338

Six-Hour Autocorrelation

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.0008	0.0002	0.0021	0.0007	0.0000	0.0000	0.0007
EXP	0.0000	0.0017	0.0017	0.0012	0.0000	0.0000	0.0007
EXPTRAN	0.0061	0.0103	0.0158	0.0155	0.0007	0.0072	0.0048
MEXPTRAN	0.0049	0.0069	0.0090	0.0099	0.0005	0.0005	0.0058
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.0007	0.0023	0.0008	0.0001	0.0001	0.0007	0.0266
EXP	0.0002	0.0000	0.0002	0.0017	0.0007	0.0007	0.0259
EXPTRAN	0.0105	0.0041	0.0039	0.0000	0.0093	0.0074	0.0857
MEXPTRAN	0.0056	0.0011	0.0047	0.0014	0.0001	0.0042	0.0648

Six-Hour Coefficient of Skewness

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.2140	3.4131	1.6857	0.0183	1.1247	1.2680	0.1410
EXP	0.4808	0.0067	4.1136	0.4528	0.9024	0.2609	0.0210
EXPTRAN	3.7946	1.1651	0.6527	6.1916	0.4360	2.2014	3.2631
MEXPTRAN	0.0713	0.0091	0.0011	1.6932	3.5820	0.3992	0.0233
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	1.6057	0.1565	0.9831	0.1853	0.0130	0.9007	0.9490
EXP	0.7267	0.2712	1.2734	0.0070	0.5131	0.7525	0.8675
EXPTRAN	0.3237	3.4114	0.2011	1.0676	2.5700	2.1065	1.4514
MEXPTRAN	0.4457	0.5699	3.4588	0.0176	3.0796	1.1126	1.0548

24-Hour Mean

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.0005	0.0885	0.0274	0.0509	0.0093	0.0270	0.0005
EXP	0.0080	0.0004	0.3453	0.0039	0.0002	0.0399	0.0030
EXPTRAN	0.2977	0.3358	0.5997	1.4901	1.7127	0.1181	0.3758
MEXPTRAN	0.0086	0.0832	0.0482	0.0008	0.0000	0.0136	0.0161
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.0482	0.0123	0.0859	0.0219	0.0176	0.0325	0.1803
EXP	0.0103	0.1480	0.0014	0.1076	0.0005	0.0557	0.2360
EXPTRAN	0.8035	2.7433	1.3158	1.3410	0.5390	0.9727	0.9863
MEXPTRAN	0.1878	0.0423	0.2610	0.3944	0.0002	0.0880	0.2967

Months	Jan	Feb	Mac	April	May	June	July
MEXP	2.693	187.583	2538.144	429.553	55.767	114.912	124.940
EXP	0.001	146.205	4488.102	1.184	80.673	37.343	15.715
EXPTRAN	4.160	16.354	760.693	1009.116	432.632	4.047	37.881
MEXPTRAN	3.135	97.958	234.979	35.467	500.120	0.042	806.493
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	3024.819	579.459	2301.351	60.428	657.836	839.790	28.979
EXP	593.877	2.407	1374.203	13.664	514.816	605.682	24.611
EXPTRAN	59.870	252.017	819.340	34.689	76.530	292.277	17.096
MEXPTRAN	231.933	201.447	213.215	25.991	401.060	229.320	15.143

24-Hour Autocorrelation

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.0107	0.0026	0.0003	0.0079	0.0000	0.0001	0.0002
EXP	0.0124	0.0059	0.0001	0.0075	0.0000	0.0004	0.0003
EXPTRAN	0.0067	0.0298	0.0009	0.0007	0.0137	0.0086	0.0001
MEXPTRAN	0.0078	0.0278	0.0036	0.0027	0.0113	0.0114	0.0000
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.0005	0.0002	0.0012	0.0002	0.0004	0.0020	0.0450
EXP	0.0000	0.0020	0.0006	0.0001	0.0000	0.0024	0.0495
EXPTRAN	0.0016	0.0001	0.0053	0.0005	0.0013	0.0058	0.0760
MEXPTRAN	0.0001	0.0000	0.0058	0.0000	0.0027	0.0061	0.0781

24-Hour Coefficient of Skewness

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.5301	1.3286	1.0666	0.0726	0.2298	0.3846	0.2846
EXP	0.8604	0.5715	2.0747	0.0121	0.0334	0.0833	0.5058
EXPTRAN	0.1018	0.0187	0.0015	0.7730	0.3627	0.6629	0.1714
MEXPTRAN	0.3570	0.2140	0.1022	0.0449	0.2419	0.0257	0.0546
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.3519	0.1199	0.4871	0.3494	0.7379	0.4952	0.7037
EXP	0.6942	0.0167	0.9357	0.3299	1.3368	0.6212	0.7882
EXPTRAN	0.0001	0.2716	0.0027	0.0001	2.4251	0.3993	0.6319
MEXPTRAN	0.3629	0.0003	1.1490	0.1216	0.0042	0.2232	0.4724

P00(24)

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.0063	0.0006	0.0060	0.0093	0.0162	0.0001	0.0005
EXP	0.0033	0.0001	0.0008	0.0094	0.0135	0.0019	0.0000
EXPTRAN	0.0021	0.0002	0.0003	0.0010	0.0008	0.0060	0.0004
MEXPTRAN	0.0004	0.0005	0.0001	0.0000	0.0003	0.0001	0.0000
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.0512	0.0000	0.0094	0.0072	0.0043	0.0093	0.0963
EXP	0.0231	0.0000	0.0098	0.0056	0.0045	0.0060	0.0775
EXPTRAN	0.0007	0.0018	0.0008	0.0004	0.0126	0.0023	0.0475
MEXPTRAN	0.0000	0.0005	0.0000	0.0003	0.0013	0.0003	0.0175

P10(24)

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.0124	0.0083	0.0004	0.0002	0.0109	0.0289	0.0022
EXP	0.0169	0.0061	0.0178	0.0003	0.0092	0.0112	0.0068
EXPTRAN	0.0062	0.0100	0.0041	0.0024	0.0022	0.0019	0.0041
MEXPTRAN	0.0001	0.0003	0.0000	0.0002	0.0002	0.0006	0.0000
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.0001	0.0075	0.0010	0.0000	0.0009	0.0061	0.0779
EXP	0.0041	0.0000	0.0026	0.0000	0.0004	0.0063	0.0793
EXPTRAN	0.0045	0.0004	0.0071	0.0107	0.0022	0.0046	0.0681
MEXPTRAN	0.0004	0.0043	0.0002	0.0001	0.0001	0.0005	0.0233

Probabilty of dry days

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.00032	0.00229	0.00067	0.00111	0.00038	0.00422	0.00004
EXP	0.00002	0.00344	0.00625	0.00100	0.00002	0.00000	0.00138
EXPTRAN	0.00000	0.00363	0.00163	0.00188	0.00126	0.00811	0.00235
MEXPTRAN	0.00038	0.00110	0.00013	0.00007	0.00000	0.00014	0.00002
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	0.0230	0.0034	0.0005	0.0003	0.0015	0.0031	0.0561
EXP	0.0053	0.0002	0.0002	0.0002	0.0038	0.0018	0.0425
EXPTRAN	0.0030	0.0001	0.0018	0.0047	0.0167	0.0038	0.0613
MEXPTRAN	0.0001	0.0012	0.0000	0.0000	0.0004	0.0003	0.0174

One-Month Mean

Months	Jan	Feb	Mac	April	May	June	July
MEXP	0.437	91.930	26.381	45.688	8.939	24.289	0.453
EXP	7.661	0.093	331.704	3.507	0.184	35.874	2.782
EXPTRAN	286.24	247.812	576.187	1341.828	1642.030	106.060	360.973
MEXPTRAN	8.221	55.832	46.178	0.713	0.001	12.195	15.452
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	46.512	11.105	82.670	20.562	16.901	31.322	5.597
EXP	9.838	133.308	1.281	93.926	0.512	51.722	7.192
EXPTRAN	771.9	2480.48	1263.345	1199.354	517.499	899.475	29.991
MEXPTRAN	180.81	38.185	250.914	357.871	0.209	80.549	8.975

One-Month Standard Deviation

Months	Jan	Feb	Mac	April	May	June	July
MEXP	16.8496	95.1850	1421.490	51.1451	3161.214	13.8940	95.8173
EXP	30.5070	96.8958	1646.266	46.3753	2519.189	159.333	58.8315
EXPTRAN	150.791	619.915	869.8444	87.0374	4369.891	393.680	18.2520
MEXPTRN	28.8132	454.009	874.1512	166.4436	3998.651	428.105	118.956
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	33.3889	387.989	49.5075	1973.535	57.6954	613.143	24.7617
EXP	0.5521	801.914	8.9778	1509.809	6.3979	573.754	23.9532
EXPTRAN	149.320	329.439	130.5474	1657.382	3.7619	731.655	27.0491
MEXPTRN	160.614	868.439	289.7869	1063.720	5.0545	704.729	26.5467

One-Month Maximum

Months	Jan	Feb	Mac	April	May	June	July
MEXP	459.26	1833.08	8193.02	3452.59	36129.12	1224.19	3878.04
EXP	26.02	453.99	3197.01	1136.81	38941.72	6379.22	1027.94
EXPTRAN	1801.4	53.30	6312.43	4250.67	27591.63	5294.90	3158.34
MEXPTRN	112.64	616.06	4692.23	1394.79	39858.84	7313.95	2909.03
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	34.74	1231.68	11.49	13588.32	5.50	5836.75	76.40
EXP	1087.3	3712.46	267.70	10133.72	80.69	5537.05	74.41
EXPTRAN	3031.2	14.98	824.89	5599.24	783.74	4893.06	69.95
MEXPTRAN	2516.9	4347.01	264.07	8061.17	549.40	6053.00	77.80

One-Month Minimum

Months	Jan	Feb	Mac	April	May	June	July
MEXP	41.19	7.79	936.97	43.71	7.35	75.65	892.46
EXP	11.60	419.72	2592.51	175.51	78.22	4.02	42.81
EXPTRAN	3.36	1942.08	1.51	1450.92	5424.03	301.51	1522.14
MEXPTRAN	22.33	1987.89	1234.55	238.04	534.31	280.68	115.69
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
MEXP	81.13	16.00	5810.94	10198.16	95.35	1517.23	38.95
EXP	284.62	601.89	3833.59	8593.09	5.79	1386.95	37.24
EXPTRAN	222.18	3860.15	13529.71	16550.55	62.06	3739.18	61.15
MEXPTRAN	24.92	382.59	5915.76	12167.14	3.77	1908.97	43.69

B. MCME MODELS

One-Hour Scale

MONTHS	Jan	Feb	Mar	Apr	May	Jun	Jul
Mean	4.5E-06	1.2E-05	7.5E-06	1.1E-06	2.3E-05	3.1E-05	2.3E-06
Std.Deviation	4.8E-03	9.9E-03	8.3E-05	5.9E-04	1.9E-03	3.6E-04	7.3E-04
Coeff.of Skewness	5.1E+0	5.3E+0	6.4E-01	4.7E-02	1.1E-01	1.6E+0	2.7E-02
Maximumn	5.5E+0	3.1E+0	2.8E+0	2.2E+0	8.9E+0	6.6E+0	3.8E+00
Autocorrelation	4.6E-02	3.1E-02	4.3E-02	5.0E-02	6.4E-02	8.2E-02	1.3E-01
dry hours	49	64	25	36	64	81	1
rainy hours	56.25	56.25	25	42.25	56.25	81	0.25
MONTHS	Aug	Sep	Oct	Nov	Dec	MSE	RMSE
Mean	1.2E-05	6.1E-06	9.8E-05	7.7E-05	1.3E-06	2.8E-05	5.3E-03
Std.Deviation	2.0E-03	4.4E-03	6.9E-03	1.6E-03	1.8E-04	3.3E-03	5.8E-02
Coeff.of Skewness	2.7E+00	3.0E-01	1.5E+00	2.6E-01	6.8E-02	1.8E+00	1.3E+00
Maximumn	3.1E+0	3.6E+0	7.2E+0	1.6E+0	2.0E+0	4.3E+0	6.6E+00
Autocorrelation	3.0E-02	5.0E-02	3.3E-02	6.3E-02	8.2E-02	7.0E-02	2.7E-01
Dry hours	49	361	225	9	9	97.3	9.8641
Rainy hours	49	380.25	210.25	4	6.25	96.7	9.8336

24-Hour Scale

MONTHS	Jan	Feb	Mar	Apr	May	Jun	Jul
Mean	2.7E-05	2.2E-01	5.1E-04	1.4E-02	5.0E-03	1.5E-02	3.5E-03
Std.Deviation	7.9E-01	8.1E+00	3.1E-01	2.6E+00	4.3E+00	4.3E+00	7.5E+00
Coeff.of Skewness	2.7E-03	2.5E-03	1.6E-02	7.7E-01	7.8E-02	5.0E-01	1.2E+00
Maximum	5.9E+01	7.6E-01	3.2E+01	1.0E+03	5.3E+01	4.2E+02	1.2E+03
Autocorr.	1.0E-02	4.1E-02	7.6E-04	3.3E-03	1.1E-02	1.7E-02	2.5E-04
Rainy	8.4E+02	6.8E+02	5.3E+02	2.9E+02	2.6E+02	7.8E+02	4.8E+02
Dry	7.8E+02	5.3E+02	5.3E+02	3.2E+02	2.0E+02	4.8E+02	3.2E+02
MONTHS	Aug	Sep	Oct	Nov	Dec	MSE	RMSE
Mean	0.320	0.003	0.000	0.000	0.007	0.049	0.222
Std.Deviation	5.723	0.604	1.209	4.450	8.083	3.995	1.999
Coeff.of Skewness	0.009	0.025	0.016	0.005	1.694	0.363	0.603
Maximum	14.669	317.552	25.301	302.064	149.084	302.908	17.404
Autocorr.	0.000	0.001	0.002	0.001	0.015	0.009	0.092
Rainy	529.000	576.000	900.000	324.000	1600.000	649.000	25.475
Dry	225.000	625.000	1296.000	289.000	1521.000	593.833	24.369

Monthly Scale (MCME Hourly)

MONTHS	Jan	Feb	Mar	Apr	May	Jun	Jul
Mean	1.716	5.108	6.300	14.516	31.810	13.032	8.762
Std.Dev.	0.274	1313.135	190.089	4.016	6165.677	985.785	0.607
Maximum	5.29	576	954.81	225	50310.49	13924	702.25
Minimum	7.29	3475.1025	90.25	39.69	686.44	404.01	1288.81
MONTHS	Aug	Sep	Oct	Nov	Dec	MSE	RMSE
Mean	9.797	10.956	1.210	0.002	2.789	8.833	2.972
Std.Dev.	2.074	1023.499	417.231	2571.314	213.411	1073.926	32.771
Maximum	265.69	4316.49	906.01	14042.25	196	7202.0233	84.8647
Minimum	57.76	1062.76	7903.21	11257.21	136.89	2200.7852	46.9125

Monthly scale (Daily MCME)

MONTHS	Jan	Feb	Mar	Apr	May	Jun	Jul
Mean	28.62	66.42	4.20	11.42	2.16	1.96	35.05
Std.Dev.	44.44	408.04	324.47	23.06	2985.53	370.68	112.42
Maximum	49.00	1024.00	1489.96	650.25	37249.00	8010.25	1772.41
Minimum	144.00	894.01	600.25	100.00	10.89	20.70	240.25
MONTHS	Aug	Sep	Oct	Nov	Dec	MSE	RMSE
Mean	8.07	1.17	4.88	10.24	0.90	14.59	3.82
Std.Dev.	255.52	850.31	24.79	1326.42	8.26	561.16	23.69
Maximum	4678.56	4212.01	201.64	12521.61	600.25	6038.25	77.71
Minimum	1.21	510.76	6544.81	6822.76	84.64	1331.19	36.49

Daily Scale

MONTHS	Jan	Feb	Mar	Apr	May	Jun	Jul
Mean	0.0084	0.0035	0.0025	0.0008	0.0681	0.0537	0.0262
Std.Dev,	0.0020	0.2426	0.0006	0.4521	0.0112	0.0016	1.0681
Coeff.of Skewness	0.2338	0.2301	0.0270	0.4971	0.1230	0.1358	0.2654
Maximum	9	930.25	197.4025	526.7025	324	38.44	231.04
Autocorr.	0.0042	0.0248	0.0024	0.0020	0.0126	0.0112	0.0007
Rainy	1	0	4	1	4	0	0.25
Dry	1	0	4	1	4	0	0.25
MONTHS	Aug	Sep	Oct	Nov	Dec	MSE	RMSE
Mean	0.00085	0.00859	0.00064	0.06502	0.00669	0.02041	0.14287
Std.Dev.	0.14267	0.00001	0.56336	0.02576	0.25999	0.23084	0.48046
Coeff.of Skewness	0.29316	0.00032	0.23256	0.24945	0.00286	0.19088	0.43690
Maximum	635.0400	81.9025	336.7225	484.0000	7.0225	316.794	17.7987
Autocorr	0.0035	0.0008	0.0047	0.0006	0.0071	0.0062	0.0788
Rainy	0	16	0	4	4	2.8542	1.6894
Dry	0	16	0	4	4	2.8542	1.6894

2. Validation Period (1991-2000)

A. NSRP Model (MEXPTRAN)

One-Hour Scale

Months	Jan	Feb	Mac	April	May	June	July
Mean	8.5E-05	4.3E-03	6.4E-04	8.7E-03	4.6E-03	1.9E-02	1.3E-04
Variance	4.7E-03	2.6E+00	3.4E-02	1.1E+00	6.1E-02	3.7E+00	6.5E-01
Skewness	4.1E-02	4.8E-01	1.3E+01	9.5E+00	1.2E-01	5.2E+01	2.8E-01
Maximum	1.3E+00	7.0E+01	6.4E+02	6.0E+01	3.6E+01	7.2E+00	4.1E+01
Autocorr.	6.5E-02	8.9E-02	4.0E-02	3.0E-02	1.2E-02	2.7E-02	2.0E-04
Prob.Dry	1.8E-04	3.4E-03	4.6E-04	1.0E-03	8.6E-03	2.7E-06	1.2E-03
Std.Dev.	1.3E-03	2.2E-01	1.8E-03	7.2E-02	3.6E-03	3.1E-01	4.9E-02
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
Mean	6.6E-03	2.5E-02	1.9E-07	3.3E-03	1.3E-02	7.1E-03	8.4E-02
Variance	5.9E-01	8.1E+00	2.1E-01	1.3E+00	8.7E-01	1.6E+00	1.3E+00
Skewness	2.1E+01	2.6E+00	1.4E+01	3.9E+00	7.0E+01	1.6E+01	3.9E+00
Maximum	1.7E+02	1.9E+03	1.6E+02	4.2E+01	1.0E+02	2.7E+02	1.7E+01
Autocorr.	6.3E-02	2.6E-02	5.7E-02	1.6E-02	1.9E-02	3.7E-02	1.9E-01
Prob.Dry	5.8E-04	1.2E-04	3.9E-03	1.2E-02	1.6E-04	2.7E-03	5.2E-02
Std.Dev.	3.6E-02	7.1E-01	1.3E-02	6.9E-02	1.1E-01	1.3E-01	3.6E-01

24-Hour Scale

Months	Jan	Feb	Mac	April	May	June	July
Mean	0.05	2.46	0.37	0.15	2.62	10.94	0.07
Var.	109.00	3720.92	3212.66	5010.08	251.79	17543.08	1.62
Skew	0.15	0.22	0.62	0.07	0.33	2.72	0.10
Max	89.83	204.93	1213.10	1.09	344.08	166.33	527.92
Autocor	0.00	0.01	0.00	0.01	0.00	0.00	0.00
Prob.Dry	0.00	0.00	0.00	0.00	0.02	0.01	0.00
Std.Dev.	0.55	5.20	3.64	6.19	0.25	24.49	0.00
Months	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**MSE**	**RMSE**
Mean	3.825	13.325	0.000	3.948	7.600	3.780	1.944
Var.	2189.1 04	20757.531	506.531	5969.580	8585.491	5654.782	75.198
Skew	0.216	0.140	0.080	1.146	1.176	0.580	0.762
Max	7.907	1321.436	114.178	595.173	447.334	419.442	20.480
Autocor	0.000	0.004	0.023	0.000	0.012	0.005	0.070
Prob.Dry	0.024	0.001	0.000	0.001	0.025	0.007	0.086
Std.Dev.	2.650	40.663	0.707	5.584	15.482	8.783	2.964

B. MCME Models

1-Hour Scale

Month	Jan	Feb	Mac	April	May	June	July
Mean	4.46E-05	4.59E-03	2.96E-03	7.26E-03	4.03E-03	2.11E-02	5.50E-05
Std.Dev	6.58E-03	1.93E-01	4.91E-02	2.52E-02	2.73E-06	2.89E-01	4.11E-02
Skew	4.68E-01	1.52E+00	9.32E-01	7.22E+00	1.82E-01	1.94E+01	7.35E+00
Max	3.97E-01	7.08E+00	2.60E-01	9.14E+01	1.93E+00	2.69E+01	2.70E+02
Autocor	1.24E-01	1.67E-01	1.13E-01	1.73E-01	1.17E-01	1.53E-01	1.15E-01
Prob.Dry	4.05E-07	1.65E-04	2.47E-05	1.44E-06	3.02E-03	5.09E-04	6.48E-05
Month	Aug	Sept	Oct	Nov	Dec	MSE	RMSE
Mean	5.55E-03	2.65E-03	4.15E-04	7.78E-04	1.50E-02	5.37E-03	7.33E-02
Std.Dev	5.11E-02	1.09E-02	4.82E-02	6.10E-02	8.66E-02	7.18E-02	2.68E-01
Skew	2.85E+00	1.28E+01	1.19E-01	2.93E+00	3.63E+01	7.67E+00	2.77E+00
Max	7.40E-01	8.40E+02	7.40E+00	2.84E+01	3.34E+01	1.09E+02	1.04E+01
Autocor	1.12E-01	9.47E-02	1.16E-01	1.13E-01	1.95E-01	1.33E-01	3.64E-01
Prob.Dry	1.54E-04	2.76E-03	1.58E-04	1.26E-03	3.96E-04	7.09E-04	2.66E-02

24-Hour Scale

Month	Jan	Feb	Mac	April	May	June	July
Mean	0.0259	2.5948	1.7067	2.0155	2.3201	12.1441	0.0313
Std.Dev	2.0274	0.0170	18.3808	11.5406	1.8458	49.6250	10.5633
Skew	0.5660	1.1179	0.0001	0.0052	1.6479	0.2169	0.8843
Max	331.2400	33.9889	11.6964	375.9721	1437.1681	1840.4100	676.0000
Auto	0.0067	0.0003	0.0011	0.0061	0.0005	0.0079	0.0023
Prob.Dry	0.0049	0.0154	0.0065	0.0001	0.0267	0.0016	0.0070
Month	Aug	Sept	Oct	Nov	Dec	MSE	RMSE
Mean	3.1964	1.8324	0.2392	1.6205	8.6367	3.0303	1.7408
Std.Dev	10.6138	0.2263	0.0123	22.1177	35.3403	13.5259	3.6778
Skew	0.0456	0.1547	0.1504	2.7492	0.3511	0.6574	0.8108
Max	299.6361	89.4916	7.6729	2424.5776	2893.3641	868.4348	29.4692
Auto	0.0003	0.0029	0.0284	0.0002	0.0252	0.0068	0.0826
Prob.Dry	0.0065	0.0374	0.0120	0.0069	0.0007	0.0105	0.1024

Daily Scale

Month	Jan	Feb	Mac	April	May	June	July
Mean	0.344	6.963	2.300	1.529	0.609	9.144	0.551
Std.Dev	0.619	10.629	10.351	12.160	0.944	20.507	1.145
Skew	0.071	0.733	0.205	0.024	0.215	1.515	0.003
Max	51.840	424.360	142.325	342.250	166.410	170.303	31.697
Auto	0.000	0.002	0.000	0.006	0.001	0.004	0.000
Prob.Dry	0.007	0.006	0.001	0.002	0.006	0.005	0.000
Month	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**	**SSE**	**RMSE**
Mean	0.9155	0.6572	0.5778	0.3758	5.4281	2.4495	**1.5651**
Std.Dev	1.3874	0.7298	1.3482	1.9125	4.4554	5.5157	**2.3486**
Skew	0.0001	0.0024	0.0985	0.9045	0.4128	0.3487	**0.5905**
Max	25.5025	0.0900	10.3041	470.8900	123.2100	163.2651	**12.7775**
Auto	0.0000	0.0028	0.0225	0.0003	0.0177	0.0048	**0.0695**
Prob.Dry	0.0097	0.0064	0.0019	0.0002	0.0100	0.0046	**0.0682**

APPENDIX G

SAMPLE OF COMPUTER PROGRAMS

1. NSRP SIMULATION

```
% NSRP simulation program
%
% input variable:
% storm   - Total number of storm to run NSRP simulation
%
% Parameter:
% lambda  - average waiting time between subsequent storm origins (/hour)
% beta    - average waiting time of the raincells after the storm origin (/hour)
% n       - average cell durations(/hour)
% v       - average number of cells per storm (cell/storm)
% epsilon - average cell intensity (mm/hour)
% theta   - intensities :mix-exponential
% alfa    - intensities: weighT

% Variables:
% ta  - inter-arrival time of storms
% C   - number of rain cells
```

```
% b   - waiting times from storm origin to rain cells
% L   - duration of rain cell
% x   - intensities

% C = [x x x x] --> storm sequence
% b or l or epsilon = [ x x x    | --> cell sequence (C)
%                       x x x    |
%                       x x x ] storm sequence
%
%
% To start simulation, please type NSRP in Matlab command window.
%
%
% Please make sure the files had copy into your ...\Matlab6p1\work before start running
simulation.
%
%
% Note: This file require another function mixexprnd.m to run simulation.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Clear all memory
clear all
% EXAMPLE
% Parameter value
%lambda  = 1/0.0499995;
%n       = 1/1.82966;
%v       = 2.02527;
%epsilon = 4.35301;
%alfa    = 0.957907;
%theta   = 37.7444;
% Time resolution: total points calculation per hour
```

```
sampling_rate = 1000 % /hour
% Set the tmax
clc
fprintf('\n------------------------ NSRP Rainfall Simulation ------------------------\n\n');
%storm = input    ('Please key in the number of storms you want to run NSRP
simulation: ');
total_hours = input('Please key in the total times in hour you want to run NSRP
simulation: '); %total_hours must be integer
```

% Part 1: Generate random waiting time between storm origins (exponential function)

```
%ta = exprnd(lambda,1,storm);
ta1 = [];
i=1;
while sum(ta1) < total_hours
   ta_rnd = exprnd(lambda);
   ta1(i) = ta_rnd;
   i=i+1;
end

storm = length(ta1)-1;
ta = zeros(1,storm);
ta(:) = ta1(1:storm);

clear ta_rnd ta1;
```

% Part 2: Generate random number of rain cells per storm (Poisson Distribution)

```
C = poissrnd(v,1,storm);
```

% Part 3: Generate random waiting times from storm origin to rain cells (exponential distribution)

```
cmax = max(C)   ;        % largest C
b = ones(storm,cmax);    % define mxn maxtrix b
b(1:storm,1:cmax) = -1;
i = 1;
while i <= storm        % generate waiting times
   rain_cell_waiting_time = exprnd(beta,1,C(i));
   b(i,1:C(i)) = rain_cell_waiting_time(1:C(i));
   i=i+1;
end

clear rain_cell_waiting_time;
```

% Part 4: Generate random duration for each rain cell (exponential distribution)

```
L = ones(storm,cmax);    % define mxn maxtrix L
L(1:storm,1:cmax) = -1;
i = 1;
while i <= storm        % generate durations
   rain_cell_duration = exprnd(n,1,C(i));
   L(i,1:C(i)) = rain_cell_duration(1:C(i));
   i=i+1;
end

clear rain_cell_duration;
```

% Part 5: Generate random intensities for each rain cell (exponential distribution)

```
x = ones(storm,cmax);    % define mxn maxtrix x
x(1:storm,1:cmax) = -1;
i = 1;
while i <= storm        % generate intensity
   %rain_cell_intensity = exprnd(epsilon,1,C(i));
   rain_cell_intensity = mixexprnd(alfa,epsilon,theta,C(i));
```

```
   x(i,1:C(i)) = rain_cell_intensity(1:C(i));
   i=i+1;
end

clear rain_cell_intensity;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Calculate storm_position
storm_position = zeros(1,storm);
i = 1;
to = 0;
while i <= storm       % determine the storm's time position (hour)
   to = to + ta(i);
   storm_position(i) = to;
   i=i+1;
end

clear to;
% Calculate rain_cell_position
rain_cell_position_m = zeros(storm,cmax);
i = 1;
while i <= storm
   rain_cell_position_m(i,1:C(i)) = storm_position(i)+b(i,1:C(i)) ;
   i=i+1;
end
rain_cell_position = zeros(1,sum(C));
i =1;
counter = 0;
while i <= storm
   rain_cell_position(counter+1:C(i)+counter) = rain_cell_position_m(i,1:C(i));
   counter = counter + C(i);
   i=i+1;
```

```
end

clear rain_cell_positon_m

% Calculate Total_intensities
duration = zeros(1,sum(C));
i =1;
counter = 0;
while i <= storm
   duration(counter+1:C(i)+counter) = L(i,1:C(i));
   counter = counter + C(i);
   i=i+1;
end

intensity = zeros(1,sum(C));
i =1;
counter = 0;
while i <= storm
   intensity(counter+1:C(i)+counter) = x(i,1:C(i));
   counter = counter + C(i);
   i=i+1;
end

clear counter;

%%%%%%% clear %%%%%%%%%
clear ta x b L;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Calculate Total_intensities  %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
t=0;
```

```
i=0;
clear storm_position;
length_t = total_hours*sampling_rate+1;
Total_intensities = zeros(1,length_t);
   j=1;
   while j <= sum(C)

      tstart = rain_cell_position(j)*sampling_rate+1;
      tstart = round(tstart);
      tstop = (rain_cell_position(j)+duration(j))*sampling_rate+1;
      tstop = round(tstop);

      if tstop > length_t
        tstop = length_t;      % limit the tstop to the longest simulation time
      end

      Total_intensities(tstart:tstop)= Total_intensities(tstart:tstop)+intensity(j);
      j=j+1;

   end

clear rain_cell_position intensity duration

% Calculate intensities per hour
intensities_per_hour = [];
sub_Total_intensities = [];
i=1;
while i <= sampling_rate+1
   sub_Total_intensities(i) = Total_intensities(i);
   i=i+1;
end
```

```
intensities_per_hour(1)= 1/sampling_rate*sum(sub_Total_intensities);

if total_hours > 1
  hour_counter = 2;

  while hour_counter <= total_hours
      sub_Total_intensities = [];
      i=1;

      while i <= sampling_rate
        sub_Total_intensities(i) = Total_intensities(sampling_rate*(hour_counter-1)+i+1);
        i=i+1;
      end

      intensities_per_hour(hour_counter)= 1/sampling_rate*sum(sub_Total_intensities);
      hour_counter=hour_counter+1;

  end
end
%Simulation Result Display
fprintf('\n-------------------------- Simulation Result ---------------------------\n\n');
fprintf('Rainfall amount hourly (mm)\n\n');
hour_counter=1;

fid = fopen('data.txt','w');    % open txt file to save data

while hour_counter <= total_hours

    if intensities_per_hour(hour_counter)~=0
```

```
    fprintf(fid,'%d hour: %2.8g\n',hour_counter, intensities_per_hour(hour_counter));  %
save data to file
    fprintf('%g hour: %g\n',hour_counter, intensities_per_hour(hour_counter))
    end

    hour_counter=hour_counter+1;
end

fclose(fid);
fprintf('\nNote 1: The number of hour not shown is 0mm\n\n');
fprintf('Note 2: The data had been saved to data.txt - Please open it using Wordpad\n\n');
%%%%%%%%%%%%%%%%%%%%%%%%% Program End %%%%%%%%%%%%%%%%%
```

2. MCME Hourly Simulation

```
%% Compares parameters for every month estimated through SCE with
%% parameters of generated precipitation period.

% Initialize


clear
S = rand('state');

load pre17.dat; % Load data file

%%-- Separate to two sets of 15 years --%%

% j = length(pre15);
% no1 = 1;
% no2 = 1;
%
```

```
% for k = 1:j
%     if pre15(k,1)<16
%         first15(no1,:) = pre15(k,:);
%         no1 = no1+1;
%     else
%         last15(no2,:) = pre15(k,:);
%         no2 = no2+1;
%     end
% end
%
% clear j k no1 no2
%%-- Estimate Monthly Transitional Probabilities and
%%-- Mixed Exponential Parameters for 1st 15 years --%%
pre_month = arrange_monthly(pre17);
[result_obs]=stat_descriptive_monthly(pre17);
for i=1:12

    [para(i,:)] = para_SCE(pre_month(:,:,i), [0.2 4 12]); % SCE Optimization
    [pij(i,1), pij(i,2)] = para_transprob(pre_month(:,:,i));

end
parameters = [pij para(:,1:3)];
[parameters]= para_FOURIER(parameters);
%%-- Create Synthetic Time Series Matrix --%%
period_synth = time_sim(10,4,1,1,1);
prev_state = 1;
%%-- Run 100 Simulations and Calculate Parameters for All Runs --%%
result_ans=[];
for run = 1:50
  rand('state',sum(100*clock))
    % Use newton-raphson to approximate rainfall with random number generation
```

```
    precip_synth(:,:,run) = precipsim_newton(period_synth, parameters, prev_state);
        %precip_synth(:,:,run) = precip_sim(period_synth, parameters, prev_state);
    [result]=stat_descriptive_monthly(precip_synth(:,:,run));
    result_ans =[result_ans;result];

    % Compare new parameters
    j = length(precip_synth(:,:,run));
    synth_month = arrange_monthly(precip_synth(1:j,:,run));
      for i=1:12
        [para(i,:)] = para_SCE(synth_month(:,1:5,i), [0.2 1 12]); % SCE Optimization
        [pij(i,1), pij(i,2)] = para_transprob(synth_month(:,1:5,i));
    end

    parameters_synth(:,:,run) = [pij para(:,1:3)];
    %result_synth(:,:,run)= stat_descriptive(precip_synth);
end
BoxPlotStatDes(result_ans,result_obs);
p00(:,:) = parameters_synth(:,1,:);
p10(:,:) = parameters_synth(:,2,:);
p(:,:) = parameters_synth(:,3,:);
u1(:,:) = parameters_synth(:,4,:);
u2(:,:) = parameters_synth(:,5,:);
p00 = p00';
p10 = p10';
p = p';
u1 = u1';
u2 = u2';
sim50 = [];
for run = 1:50
    sim50 = [sim50 precip_synth(:,5,run)];
end
```

```
 boxplot_comp % compare simulated data to observed
 obs = pre17(:,5);
 sim = sim50;
save sim50 sim obs

clear pij para paraML period_synth start leap period pre_month synth_month prev_state
i j run
```

APPENDIX H

CODING FOR MICROSOFT VISUAL C++ PROGRAM TO CALCULATE THE FORECAST OF THE RAINFALLS USING THE MARIMA MODEL

<u>Program.h</u>

```
#include <afxwin.h>
#include <afxcmn.h>
#include <afxdlgs.h>
#include <math.h>
#include "resource.h"
#define IDC_BUTTON 500
#define m 680 //number of data used
#define n 10
#define v 2
#define w 1


class program : public CFrameWnd
{
protected:
	int idc,flag,p;
	double h,*a,*b,*c;
	double cov[v+1][v+1], covlag[v+1][v+1], invcov[v+1][v+1],
tracovlag[v+1][v+1], phi[v+1][v+1], er[v+1],
x[v+1],d[v+1],e[v+1],obs[v+1],r[v+1];
	CListCtrl table,table2;
	CPoint px,pg,home1,home2,end1,end2,hBox1,hBox2;
	CEdit eBox1,eBox2;
	CStatic sBox1,sBox2,sBox3,sBox4,fileBox;
	CString strFile;
	CButton bnDraw;
	CSize BoxSize;
	typedef struct
	{
		double x,y;
	} PT;
	PT *pt,max,min,left,right;
public:
```

```
      program();
      ~program();
      void ShowTable();
      afx_msg void OnPaint();
      afx_msg void OnPolynomial();
      afx_msg void OnFileOpen();
      afx_msg void OnExit();
      afx_msg void OnForecast();
      DECLARE_MESSAGE_MAP()
};

class CMyWinApp : public CWinApp
{
public:
      virtual BOOL InitInstance();
};
```

Program.cpp

```
#include "program.h"

CMyWinApp  MyApplication;

BOOL CMyWinApp::InitInstance()
{
    program* pFrame = new program;
    m_pMainWnd = pFrame;
    pFrame->ShowWindow(SW_SHOW);
    pFrame->UpdateWindow();
    return TRUE;
}

BEGIN_MESSAGE_MAP(program,CFrameWnd)
      ON_WM_PAINT()
      ON_COMMAND(ID_FILEOPEN,OnFileOpen)
      ON_COMMAND(ID_EXIT,OnExit)
    ON_BN_CLICKED(IDC_BUTTON,OnPolynomial)
END_MESSAGE_MAP()

program::program()
{
      Create(NULL,"Code25D: Menus and file I/O",WS_OVERLAPPEDWINDOW,
            CRect(0,0,800,600),NULL,MAKEINTRESOURCE(IDR_MENU1));
      pt=new PT [m+1];
      idc=400; flag=0;  p=m-1;
      home1=CPoint(5,50); end1=CPoint(900,250);
      home2=CPoint(5,300); end2=CPoint(900,500);
      hBox1=CPoint(100,550); hBox2=CPoint(200,550);
      BoxSize=CSize(1,1);
      a=new double [p+1];
      b=new double [p+1];
      c=new double [p+1];
      er[1]=0;
      er[2]=0;
}

program::~program()
```

```
{
      delete pt;
}

void program::OnPaint()
{
      CPaintDC dc(this);
      CString str;
      CRect Box;
      CFont fontTimes;

      // show the parameter estimated and the forecast
      if (flag==1 || flag==2 )
      {

            fontTimes.CreatePointFont(90,"Arial");
            dc.SelectObject(fontTimes);
            str.Format("Model Parameters");
            dc.TextOut(home1.x+10,home1.y-30,str);
            str.Format("%lf  %lf",phi[1][1],phi[1][2]);
            dc.TextOut(home1.x+10,home1.y,str);
            str.Format("%lf  %lf",phi[2][1],phi[2][2]);
            dc.TextOut(home1.x+10,home1.y+20,str);
            str.Format("%lf",x[1]);
            sBox3.SetWindowText(str);
            str.Format("%lf",x[2]);
            sBox4.SetWindowText(str);
      }
}

void program::OnPolynomial()
{
      CClientDC dc(this);
      CRect rc;
      CString str;
      int i,j,k;
      double x1[v+1],x2[v+1];
      double meanxm,meanx1m;

      CBrush bkBrush(RGB(255,255,255));
      rc=CRect(home1.x,home1.y-50,end2.x,end2.y+10);
      dc.FillRect(&rc,&bkBrush);

      //covariance and covariance lag 1
            double sum1=0;
            for(i=0;i<=p;i++)
                  sum1+=a[i];
            meanxm=sum1/(p+1);
            double sum2=0;
            for(i=0;i<=p;i++)
                  sum2+=b[i];
            meanx1m=sum2/(p+1);
            double sum3=0;
            for(i=0;i<=p;i++)
                  sum3+=((a[i]-meanxm)*(a[i]-meanxm));
            cov[1][1]=sum3/(p+1);
            double sum4=0;
```

```
		for(i=0;i<=p;i++)
			sum4+=((a[i]-meanxm)*(b[i]-meanx1m));
		cov[1][2]=sum4/(p+1);
		cov[2][1]=cov[1][2];
		double sum5=0;
		for(i=0;i<=p;i++)
			sum5+=((b[i]-meanx1m)*(b[i]-meanx1m));
		cov[2][2]=sum5/(p+1);

		double sum6=0;
		for(i=0;i<=p-1;i++)
			sum6+=((a[i]-meanxm)*(a[i+1]-meanxm));
		covlag[1][1]=sum6/(p+1);
		double sum7=0;
		for(i=0;i<=p-1;i++)
			sum7+=((b[i]-meanx1m)*(a[i+1]-meanxm));
		covlag[1][2]=sum7/(p+1);
		double sum8=0;
		for(i=0;i<=p-1;i++)
			sum8+=((a[i]-meanxm)*(b[i+1]-meanx1m));
		covlag[2][1]=sum8/(p+1);
		double sum9=0;
		for(i=0;i<=p-1;i++)
			sum9+=((b[i]-meanx1m)*(b[i+1]-meanx1m));
		covlag[2][2]=sum9/(p+1);

		//inverse covariance
		invcov[1][1]=cov[2][2]/(cov[1][1]*cov[2][2]-
cov[1][2]*cov[2][1]);
		invcov[1][2]=-cov[1][2]/(cov[1][1]*cov[2][2]-
cov[1][2]*cov[2][1]);
		invcov[2][1]=-cov[2][1]/(cov[1][1]*cov[2][2]-
cov[1][2]*cov[2][1]);
		invcov[2][2]=cov[1][1]/(cov[1][1]*cov[2][2]-
cov[1][2]*cov[2][1]);


		//estimate parameter phi
		for (i=1; i<=v; i++)
		{
			for (j=1;j<=v;j++)
			{
				phi[i][j]=0;
				for (k=1;k<=v;k++)
				phi[i][j] += covlag[i][k]*invcov[k][j];
			}
		}


	//calculate the forecast using the MARIMA model
	if(p>m)
	{
		eBox1.GetWindowText(str); obs[1]=atof(str);
		eBox2.GetWindowText(str); obs[2]=atof(str);

		x1[1]=a[p];
		x1[2]=b[p];
```

```
            x2[1]=a[p-1];
            x2[2]=b[p-1];
            for (i=1; i<=v; i++)
            {
                  d[i]=0;
                  e[i]=0;
                  er[i]=obs[i]-x[i];
                  r[i]=x[i];
                  for (j=1;j<=v;j++)
                  {
                        d[i] += phi[i][j]*x1[j];
                        e[i] += phi[i][j]*x2[j];
                  }
                  x[i]=x1[i]+d[i]-e[i];
                  if(x[i]<0)
                  {
                        x[i]=0;
                  }
            }
            a[p+1]=obs[1];
            b[p+1]=obs[2];
      }
      p=p+1;



      if (flag<=3)
      {
            if (flag==2 || flag==3 )
            {
                  bnDraw.DestroyWindow(); fileBox.DestroyWindow();
                  eBox1.DestroyWindow(); eBox2.DestroyWindow();
                  sBox1.DestroyWindow(); sBox2.DestroyWindow();
                  sBox3.DestroyWindow(); sBox4.DestroyWindow();
                  fileBox.Create(strFile,WS_CHILD | WS_VISIBLE |
SS_CENTER | SS_SIMPLE,
                              CRect(hBox1.x,hBox1.y-
30,hBox1.x+120,hBox1.y-10),this,idc++);
            }
            max.y=0;
            for (i=0;i<=m;i++)
            {
                  if (max.y<a[i])
                        max.y=a[i];
                  if (max.y<b[i])
                        max.y=b[i];
            }
            InvalidateRect(rc);
            ShowTable();

      }
      if (flag==3)
      {
            GetClientRect(&rc);
            CBrush whiteBrush(RGB(255,255,255));
            dc.FillRect(&rc,&whiteBrush);
      }
```

```
}

//show the table
void program::ShowTable()
{
      CString str;
      CRect rcTable=CRect(620,20,840,540);
      table.DestroyWindow();
    table.Create(WS_VISIBLE | WS_CHILD | WS_DLGFRAME | LVS_REPORT
        | LVS_NOSORTHEADER,rcTable,this,idc++);
      table.InsertColumn(0,"day",LVCFMT_CENTER,30);
      table.InsertColumn(1,"station1",LVCFMT_CENTER,80);
      table.InsertColumn(2,"station2",LVCFMT_CENTER,80);

                  for (int i=0;i<=m;i++)
                  {
                        str.Format("%d",c[i]);
table.InsertItem(i,str,0);
                        str.Format("%lf",a[i]);
table.SetItemText(i,1,str);
                        str.Format("%lf",b[i]);
table.SetItemText(i,2,str);
                  }

      if(p==m)
      {
      CRect rcTable2=CRect(10,100,600,500);
      table2.DestroyWindow();
    table2.Create(WS_VISIBLE | WS_CHILD | WS_DLGFRAME | LVS_REPORT
        | LVS_NOSORTHEADER,rcTable2,this,idc++);
      table2.InsertColumn(0,"day",LVCFMT_CENTER,40);
      table2.InsertColumn(1,"station1",LVCFMT_CENTER,90);
      table2.InsertColumn(2,"station2",LVCFMT_CENTER,90);
      table2.InsertColumn(3,"station1(p)",LVCFMT_CENTER,90);
      table2.InsertColumn(4,"station2(p)",LVCFMT_CENTER,90);
      table2.InsertColumn(5,"station1(er)",LVCFMT_CENTER,90);
      table2.InsertColumn(6,"station2(er)",LVCFMT_CENTER,90);
      }
      if(p>m+1)
      {

                  str.Format("obs"); table2.InsertItem(p-m-2,str,0);
                  str.Format("%lf",a[p]); table2.SetItemText(p-m-
2,1,str);
                  str.Format("%lf",b[p]); table2.SetItemText(p-m-
2,2,str);
                  str.Format("%lf",r[1]); table2.SetItemText(p-m-
2,3,str);
                  str.Format("%lf",r[2]); table2.SetItemText(p-m-
2,4,str);
                  str.Format("%lf",er[1]); table2.SetItemText(p-m-
2,5,str);
                  str.Format("%lf",er[2]); table2.SetItemText(p-m-
2,6,str);
      }
}
```

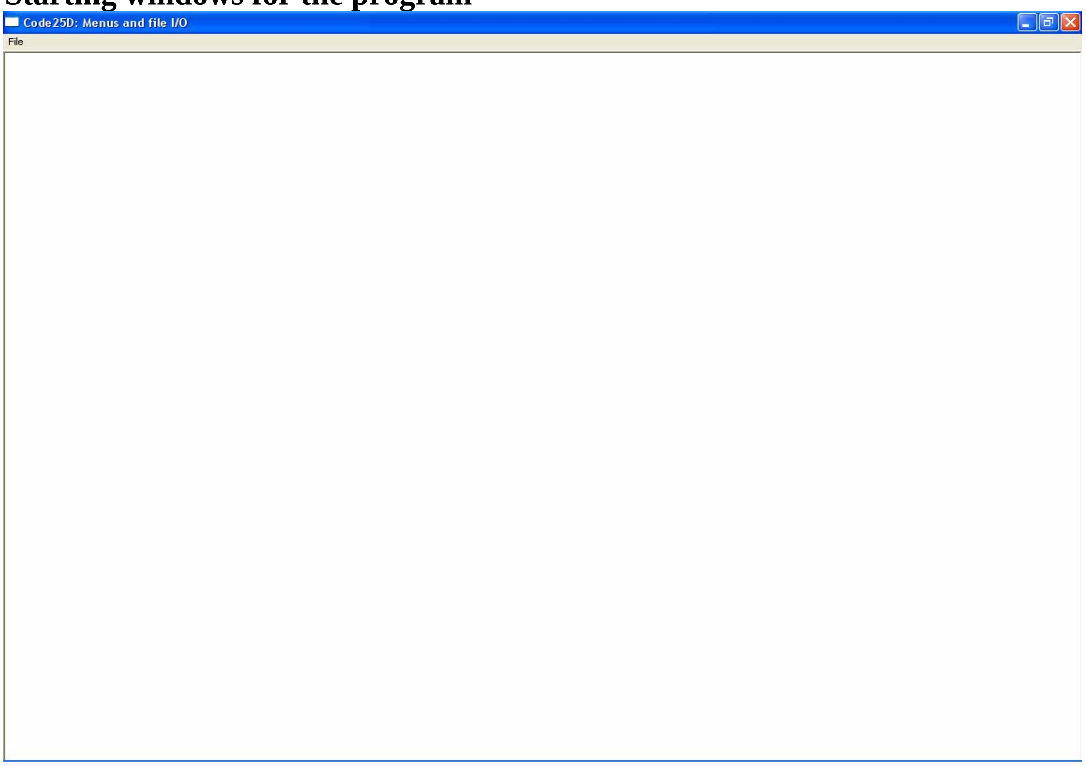
```
void program::OnForecast()
{
	bnDraw.DestroyWindow(); fileBox.DestroyWindow();
	eBox1.DestroyWindow(); eBox2.DestroyWindow();
	sBox1.DestroyWindow(); sBox2.DestroyWindow();
	sBox3.DestroyWindow(); sBox4.DestroyWindow();
	bnDraw.Create("Forecast",WS_CHILD | WS_VISIBLE |
BS_DEFPUSHBUTTON,
	  CRect(300,550,440,580),this,IDC_BUTTON);
	sBox1.Create("station1",WS_CHILD | WS_VISIBLE | SS_SUNKEN |
SS_CENTER,
		CRect(hBox1.x,hBox1.y-30,hBox1.x+60,hBox1.y-
10),this,idc++);
	sBox2.Create("station2",WS_CHILD | WS_VISIBLE | SS_SUNKEN |
SS_CENTER,
		CRect(hBox2.x,hBox2.y-30,hBox2.x+60,hBox2.y-
10),this,idc++);
	sBox3.Create("",WS_CHILD | WS_VISIBLE | SS_SUNKEN | SS_CENTER,

	CRect(hBox1.x,hBox1.y+50,hBox1.x+60,hBox1.y+70),this,idc++);
	sBox4.Create("",WS_CHILD | WS_VISIBLE | SS_SUNKEN | SS_CENTER,

	CRect(hBox2.x,hBox2.y+50,hBox2.x+60,hBox2.y+70),this,idc++);
	eBox1.Create(WS_CHILD | WS_VISIBLE | WS_BORDER,
		CRect(CPoint(hBox1),CSize(70,25)),this,idc++);
	eBox2.Create(WS_CHILD | WS_VISIBLE | WS_BORDER,
		CRect(CPoint(hBox2),CSize(70,25)),this,idc++);
	flag=1;
}

void program::OnFileOpen()
{
	CString strFilter="|*.*|";
	CFileDialog FileDlg(TRUE,"",NULL,0,strFilter);
	FILE *ifp;

	if (FileDlg.DoModal()==IDOK)
	{
		strFile=FileDlg.GetFileName();
		ifp=fopen(strFile,"r");
		for (int i=0;i<=m;i++)
			fscanf(ifp,"%d %lf %lf",&c[i],&a[i],&b[i]);
		fclose(ifp);
		flag=2;
		obs[1]=a[m];
		obs[2]=b[m];
		x[1]=a[m];
		x[2]=b[m];
		OnPolynomial();
		OnForecast();
	}
}
void program::OnExit()
{
	OnExit();
```

APPENDIX I

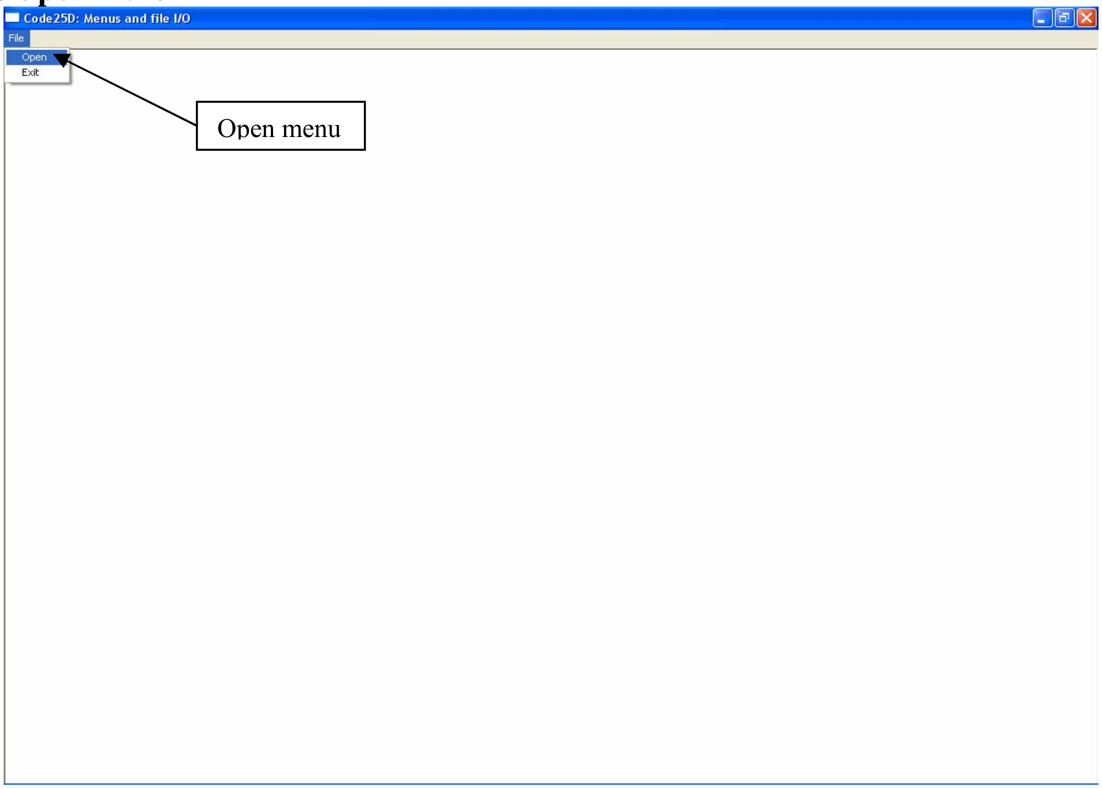
USER INTERFACE FOR THE MICROSOFT VISUAL C++ PROGRAM TO CALCULATE THE FORECAST OF THE RAINFALLS USING THE MARIMA MODE

l

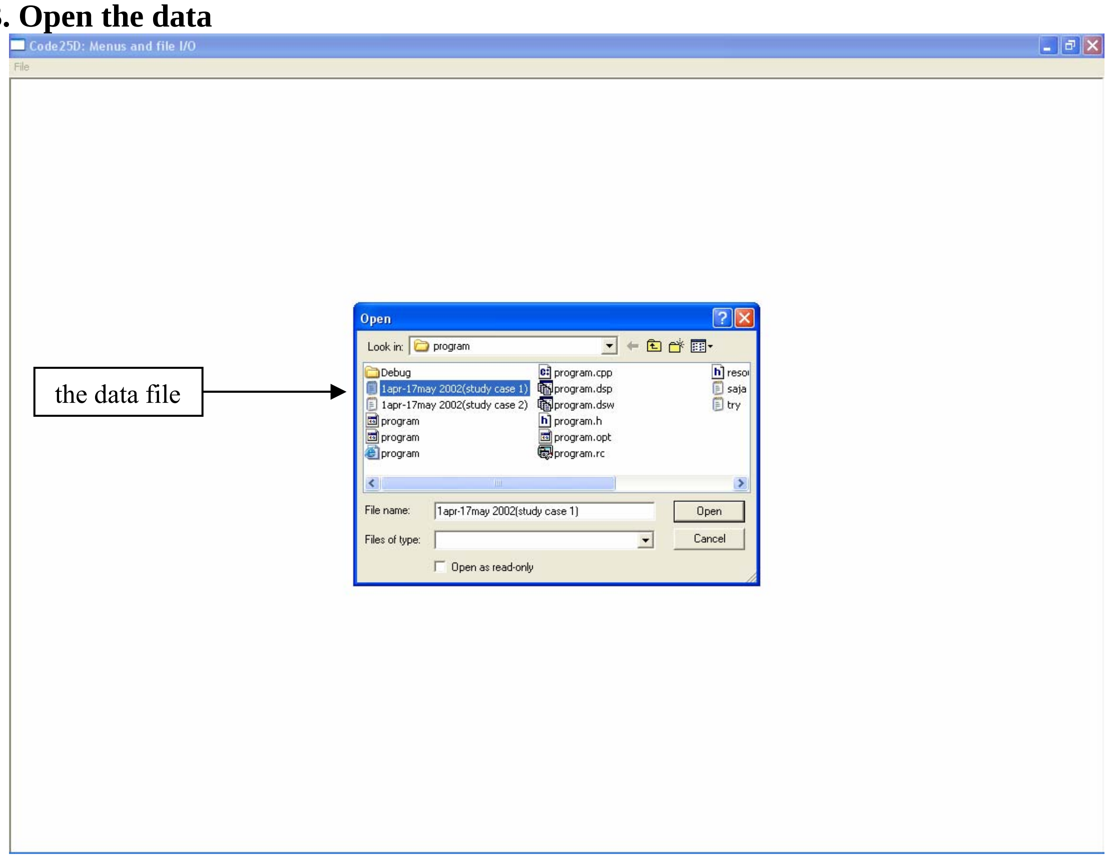
1. Starting windows for the program



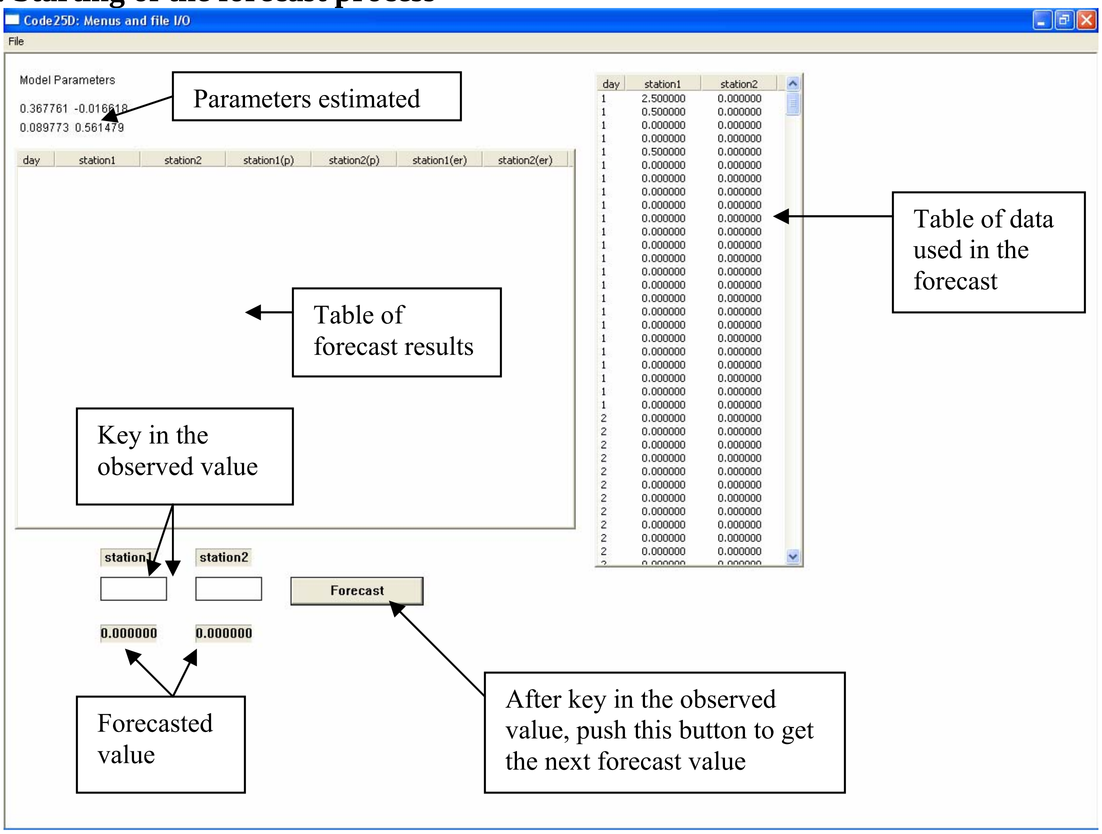
2. Open menu



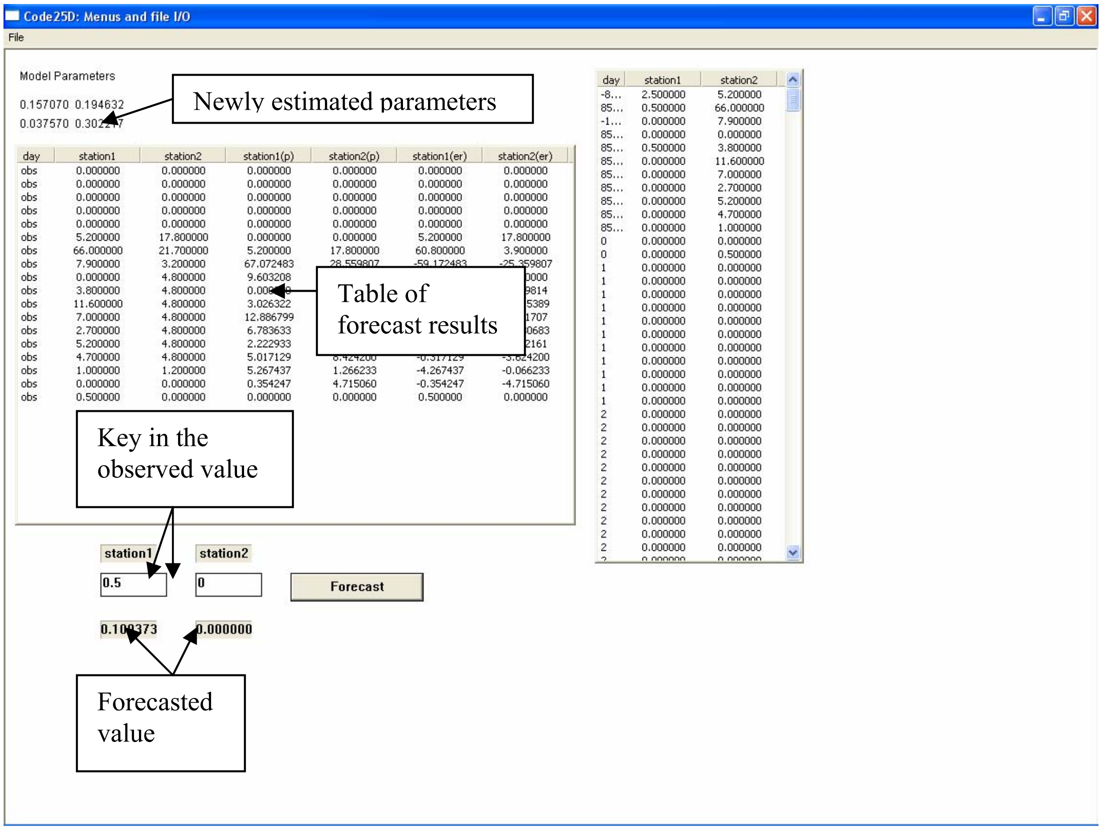
3. Open the data



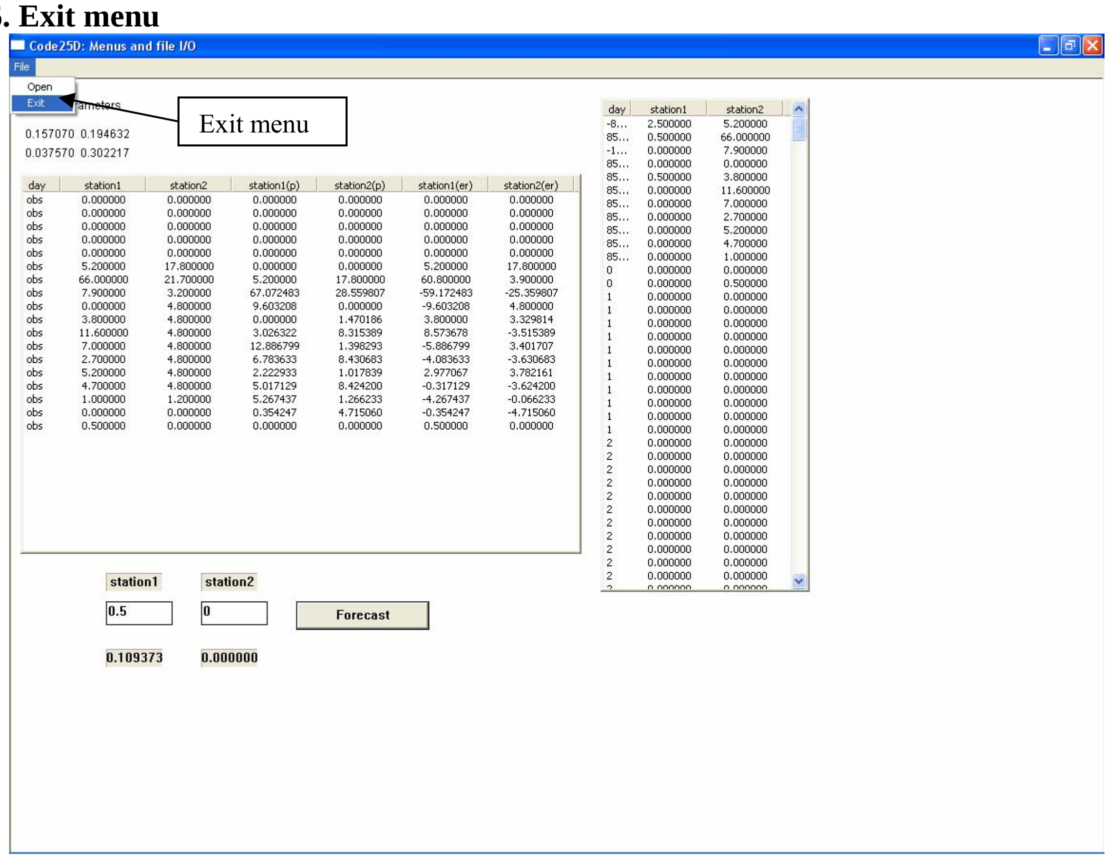
4. Starting of the forecast process



5. Results



6. Exit menu



APPENDIX J

HOURLY RAINFALLS INTENSITY DATA USED TO FORECAST THE RAINFALLS INTENSITY

Study Case 1: Station Empangan Genting Kelang (3217002) and station Km.11 Gombak (3217003) (from 0100 hour, 1st April 2002 to 0800 hour, 29th April 2007)

Days	3217002	3217003	Days	3217002	3217003	Days	3217002	3217003
1	2.5	0	3	0	0	5	1.3	1.1
1	0.5	0	3	0	0	5	1.4	2
1	0	0	3	0	0	5	1.3	0.3
1	0	0	3	0	0	5	0	0
1	0.5	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	1.2
1	0	0	3	0	0	5	2	0.4
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0.5	0.5
1	0	0	3	0	0	5	0.5	0.5
1	0	0	3	0	0	5	1.5	0.5
1	0	0	3	0	0	5	0	1
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	1	0
2	0	0	4	0	0	6	1	0.5
2	0	0	4	0	0	6	0.5	0.4
2	0	0	4	0	0	6	0.5	8
2	0	0	4	0	0	6	0	2.8
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0.5

2	0	0	4	0	0	6	1	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	1	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0.5	1.5	6	0	0

Days	3217002	3217003	Days	3217002	3217003	Days	3217002	3217003
7	0	0	9	0	0	12	0	0
7	0	0	9	7.3	0	12	0	0
7	0	0	9	10.3	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	10	0	0	12	0	0
7	0	0	10	0	1	12	0	0
7	0	0	10	2.9	0.5	12	0	0
7	0	0	10	6.4	0	12	0	0
7	0	0.1	10	0.7	0	12	0	0
7	7	2.9	10	0.5	0	12	0	0
7	0	0.5	10	0	0	12	0	0
7	0	0	10	0	0	12	0	0
7	0	0	10	0	0	12	0	0
7	0	0	10	0	0	12	0	0
7	0.5	0	10	0	0	13	0	0
7	1	1.5	10	0	0	13	0	0
7	0	0	10	0	0	13	0	0
7	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0

8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	14	0	0
8	0	0	11	0	0	14	0	0
8	0	0	11	0	0	14	0	0
8	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	12	0	0	14	6.4	0.9
9	0	0	12	0	0	14	6	1.1
9	0	0	12	0	0	14	0.6	0
9	0	0	12	0	0	14	0	0

Days	**3217002**	**3217003**	**Days**	**3217002**	**3217003**	**Days**	**3217002**	**3217003**
14	0	0	17	0	0	19	0	0
14	0	0	17	0	0	19	0	0
14	0	0	17	0	0	20	0	0
14	0	0	17	0	0	20	0	0
14	0	0	17	0	0	20	0	0
14	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0.5	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	18	0	0	20	0	0
15	0	0	18	0	0	20	0	0
15	0	0	18	0	0	20	0	0
15	0.5	0	18	0	0	20	0	0
15	0	0	18	0	0	20	0	0
15	0	0	18	0	0	20	0	0
15	0	0	18	0	0	20	0	0
15	0	0	18	0	0	20	0	0.2
15	0	0	18	0	0	20	0	0.3
15	0	0	18	0	0	20	0	0
15	0	0	18	0	0	21	0	1
15	0	0	18	0	0	21	0	0.5
15	0	2.8	18	0	0	21	0	0
15	2	2.2	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0

16	0	0	18	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	0	0.5	19	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	5	1	19	0	0	21	0	1
16	0.5	0	19	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	0.5	0.5	19	0	0	21	0	0
16	0	0	19	0	0	22	0	0
16	0	0	19	0	0	22	0	0
16	0	0	19	0	0	22	0	0
16	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0

Days	**3217002**	**3217003**	**Days**	**3217002**	**3217003**	**Days**	**3217002**	**3217003**
22	2	0	25	0	0	27	0	0
22	6.2	11.9	25	0	0	27	0	0
22	1.8	14	25	0	0	27	0	0.3
22	0	11.8	25	0	0	27	1.5	1.2
22	0	0	25	0	0	27	0	0
22	1	0	25	0	0	27	0	0
22	1	0.1	25	0	0	27	0	0
22	0	1.2	25	0	0	27	0	0
22	0	1	25	0	0	28	0	0
22	0	0	25	0	0	28	0	0
22	0	0	25	0	0	28	0	0.5
22	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	29	0	0
23	0	0	26	0	0	29	0	0
23	0	0	26	0	0	29	0	0

23	0	0	26	0	0	29	0	0
24	0	0	26	0.5	0	29	0	0
24	0	0	26	0	0	29	0	0
24	0	0	26	0	0	29	0	0
24	0	0	26	3	0.5	29	0	0
24	0	0	26	0	0			
24	0	0	26	0	0			
24	0	0	26	0	0			
24	0	0	26	0	0			
24	0	0	26	0	0			
24	0	0	26	0.3	2.9			
24	0	0	27	5.2	7.2			
24	0	0	27	0.6	0			
24	0	0	27	0	0			
24	0	0	27	0.5	0			
24	0	0	27	0	0			
24	0	0	27	0	0			
24	0	0	27	0	0			
24	0	0	27	0	0.5			
24	0	0	27	0	0			
24	0	0	27	0	0			
24	0	0	27	0	0			
24	0	0	27	0.5	0			
24	0	0	27	0	1.2			
24	0	0	27	0	3.8			
25	0	0	27	0	0			
25	0	0	27	0	0			

Study Case 2: Station Empangan Genting Kelang (3217002) and station Kampung Kuala Saleh (3217004) (from 0100 hour, 1st April 2002 to 0800 hour, 29th April 2007)

Days	3217002	3217004	Days	3217002	3217004	Days	3217002	3217004
1	2.5	3.4	3	0	0	5	1.3	0.5
1	0.5	0	3	0	0	5	1.4	1
1	0	0.5	3	0	0	5	1.3	1
1	0	0	3	0	0	5	0	0
1	0.5	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0.5
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	2	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0	0
1	0	0	3	0	0	5	0.5	0

1	0	0	3	0	0	5	0.5	0
1	0	0	3	0	0	5	1.5	0
1	0	0	3	0	0	5	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	1	0
2	0	0	4	0	0	6	1	0.2
2	0	0	4	0	0	6	0.5	6.8
2	0	0	4	0	0	6	0.5	0.6
2	0	0	4	0	0	6	0	0.5
2	0	0	4	0	0	6	0	0.5
2	0	0	4	0	0	6	0	1
2	0	0	4	0	0	6	1	0.8
2	0	0	4	0	0	6	0	0.7
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0	0	6	0	0
2	0	0	4	0.5	0.5	6	0	0

Days	3217002	3217004	Days	3217002	3217004	Days	3217002	3217004
7	0	0	9	0	0	12	0	0
7	0	0	9	7.3	9.8	12	0	0
7	0	0	9	10.3	4.2	12	0	0
7	0	0	9	0	0.5	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	9	0	0	12	0	0
7	0	0	10	0	0	12	0	0
7	0	0	10	0	0	12	0	0
7	0	0	10	2.9	4.1	12	0	0
7	0	14.8	10	6.4	6.7	12	0	0
7	0	16.5	10	0.7	1	12	0	0
7	7	0.7	10	0.5	1.5	12	0	0
7	0	0	10	0	0.5	12	0	0
7	0	0	10	0	0	12	0	0
7	0	0	10	0	0	12	0	0
7	0	0.5	10	0	0	12	0	0
7	0.5	0	10	0	0	13	0	0
7	1	0	10	0	0	13	0	0
7	0	0	10	0	0	13	0	0
7	0	0	10	0	0	13	0	0
8	0	1	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0

8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	2	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	10	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	5	11	0	0	13	0	0
8	0	0.5	11	0	0	13	0	0
8	0	1.9	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	13	0	0
8	0	0	11	0	0	14	0	0
8	0	0	11	0	0	14	0	0
8	0	0	11	0	0	14	0	0
8	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	0
9	0	0	11	0	0	14	0	3
9	0	0	11	0	0	14	0	28
9	0	0	12	0	0	14	6.4	0.6
9	0	0	12	0	0	14	6	0
9	0	0	12	0	0	14	0.6	0.5
9	0	0	12	0	0	14	0	0

Days	**3217002**	**3217004**	**Days**	**3217002**	**3217004**	**Days**	**3217002**	**3217004**
14	0	0	17	0	0	19	0	0
14	0	0	17	0	0	19	0	0
14	0	0	17	0	0	20	0	0
14	0	0	17	0	0	20	0	0
14	0	0	17	0	0	20	0	0
14	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0.5	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	0	17	0	0	20	0	0
15	0	8.5	18	0	0	20	0	0
15	0	0.5	18	0	0	20	0	1
15	0	0	18	0	0	20	0	0
15	0.5	0	18	0	0	20	0	0
15	0	0	18	0	0	20	0	0
15	0	0	18	0	0	20	0	0
15	0	0	18	0	0	20	0	0
15	0	0	18	0	0	20	0	0

15	0	0	18	0	0	20	0	0
15	0	0	18	0	0	20	0	0
15	0	0	18	0	0	21	0	0
15	0	0	18	0	0	21	0	0
15	0	0	18	0	0	21	0	0
15	2	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0.5
16	0	0	18	0	0	21	0	0.5
16	0	0	18	0	0	21	0	0
16	0	0.5	18	0	0	21	0	0
16	0	0	18	0	0	21	0	0
16	0	0.5	18	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	0	1	19	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	5	0.5	19	0	0	21	0	0
16	0.5	0	19	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	0	0	19	0	0	21	0	0
16	0.5	0.5	19	0	0	21	0	0
16	0	0	19	0	0	22	0	3.9
16	0	0	19	0	0	22	0	1.5
16	0	0	19	0	0	22	0	3.5
16	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0.1
17	0	0	19	0	0	22	0	0.4
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0
17	0	0	19	0	0	22	0	0

Days	3217002	3217004	Days	3217002	3217004	Days	3217002	3217004
22	2	0	25	0	0	27	0	0
22	6.2	0	25	0	3	27	0	0
22	1.8	0	25	0	2	27	0	0
22	0	0	25	0	0	27	1.5	0
22	0	0	25	0	0	27	0	0
22	1	0	25	0	0	27	0	0
22	1	0	25	0	0	27	0	0
22	0	0	25	0	0	27	0	0
22	0	0	25	0	0	28	0	0
22	0	0	25	0	0	28	0	0
22	0	0	25	0	0	28	0	0
22	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0

23	0	0	25	0	0	28	0	0
23	0	0	25	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0.5	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	28	0	0
23	0	0	26	0	0	29	0	0
23	0	0	26	0	0	29	0	0
23	0	0	26	0	0	29	0	0
23	0	0	26	0	0	29	0	0
24	0	0	26	0.5	0	29	0	0
24	0	0	26	0	0	29	0	0
24	0	0	26	0	0	29	0	0
24	0	0	26	3	0	29	0	0
24	0	0	26	0	0			
24	0	0	26	0	0			
24	0	5	26	0	0			
24	0	1.5	26	0	0			
24	0	0	26	0	0			
24	0	0	26	0.3	0			
24	0	0	27	5.2	0			
24	0	0	27	0.6	0			
24	0	0	27	0	0			
24	0	0	27	0.5	0			
24	0	0	27	0	0			
24	0	0	27	0	0			
24	0	0	27	0	0			
24	0	0	27	0	0			
24	0	0	27	0	0			
24	0	0	27	0	0			
24	0	2	27	0	0			
24	0	13.5	27	0.5	0			
24	0	0	27	0	0			
24	0	0	27	0	0			
25	0	0	27	0	0			
25	0	0	27	0	0			

APPENDIX K

RAINFALLS FORECAST RESULTS (PRE) USING THE MARIMA MODEL WITH OBSERVED VALUE (OBS), FORECAST ERROR (ER) AND THE ESTIMATED PARAMETERS ($\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}$).

Study Case 1: Station Empangan Genting Kelang (E) and station Km.11 Gombak (G)

Days	E(obs)	G(obs)	E(pre)	G(pre)	E(er)	G(er)	α_{11}	α_{12}	α_{21}	α_{22}
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3678	-0.0166	0.0898	0.5615
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3678	-0.0166	0.0898	0.5615
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3678	-0.0166	0.0898	0.5615
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3678	-0.0166	0.0898	0.5615
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3678	-0.0166	0.0898	0.5615
29	5.2000	17.8000	0.0000	0.0000	-5.2000	-17.8000	0.3679	-0.0165	0.0899	0.5615
29	66.0000	21.7000	0.0000	0.0000	-66.0000	-21.7000	0.3679	-0.0165	0.0899	0.5615
29	7.9000	3.2000	6.2725	24.6598	-1.6275	21.4598	0.3687	-0.0475	0.0800	0.3620
29	0.0000	4.8000	68.7757	23.6321	68.7757	18.8321	0.3294	0.1982	0.0202	0.1799
29	3.8000	4.8000	0.0000	0.0000	-3.8000	-4.8000	0.1369	0.1927	0.0247	0.2754
29	11.6000	4.8000	0.0000	4.9856	-11.6000	0.1856	0.1368	0.1919	0.0321	0.2745
29	7.0000	4.8000	4.3131	4.9137	-2.6869	0.1137	0.1350	0.1936	0.0299	0.2766
29	2.7000	4.8000	12.6704	5.0290	9.9704	0.2290	0.1372	0.1950	0.0294	0.2791
29	5.2000	4.8000	6.3066	4.6485	1.1066	-0.1515	0.1507	0.1929	0.0329	0.2906

29	4.7000	4.8000	2.0401	4.6420	-2.6599	-0.1580	0.1535	0.1926	0.0367	0.2942
29	1.0000	1.2000	5.5846	4.8904	4.5846	3.6904	0.1538	0.1940	0.0362	0.2970
29	0.0000	0.0000	4.6217	4.7813	4.6217	4.7813	0.1566	0.1944	0.0374	0.3012
30	0.5000	0.0000	0.0000	0.0000	-0.5000	0.0000	0.1571	0.1947	0.0376	0.3023
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1946	0.0376	0.3022
30	0.0000	0.0000	0.5875	0.0188	0.5875	0.0188	0.1571	0.1946	0.0376	0.3022
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1946	0.0376	0.3022
Days	**E(obs)**	**G(obs)**	**E(pre)**	**G(pre)**	**E(er)**	**G(er)**	α_{11}	α_{12}	α_{21}	α_{22}
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1946	0.0376	0.3022
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1946	0.0376	0.3022
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1946	0.0376	0.3022
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1946	0.0376	0.3023
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1947	0.0376	0.3023
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1947	0.0376	0.3023
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1947	0.0376	0.3023
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1947	0.0376	0.3023
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1947	0.0377	0.3023
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1947	0.0377	0.3023
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1947	0.0377	0.3024
30	0.8000	0.0000	0.0000	0.0000	-0.8000	0.0000	0.1571	0.1947	0.0377	0.3024
30	4.2000	1.0000	0.0000	0.0000	-4.2000	-1.0000	0.1571	0.1947	0.0377	0.3024
30	0.0000	0.3000	0.9257	0.0302	0.9257	-0.2698	0.1571	0.1947	0.0377	0.3023
30	0.5000	1.2000	4.9288	1.4296	4.4288	0.2296	0.1574	0.1937	0.0376	0.3016
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1937	0.0376	0.3016
30	0.0000	0.0000	0.7529	1.4901	0.7529	1.4901	0.1571	0.1937	0.0375	0.3015
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1936	0.0375	0.3014
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1571	0.1936	0.0375	0.3014
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1936	0.0375	0.3015
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1936	0.0375	0.3015
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1937	0.0375	0.3015
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1937	0.0376	0.3015

1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1937	0.0376	0.3015
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1937	0.0376	0.3015
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1937	0.0376	0.3016
1	0.5000	1.1000	0.0000	0.0000	-0.5000	-1.1000	0.1572	0.1937	0.0376	0.3016
1	2.0000	1.4000	0.0000	0.0000	-2.0000	-1.4000	0.1572	0.1937	0.0376	0.3016
1	0.5000	0.0000	0.7916	1.4504	0.2916	1.4504	0.1572	0.1937	0.0376	0.3015
Days	**E(obs)**	**G(obs)**	**E(pre)**	**G(pre)**	**E(er)**	**G(er)**	α_{11}	α_{12}	α_{21}	α_{22}
1	0.0000	0.0000	2.2938	1.5466	2.2938	1.5466	0.1572	0.1936	0.0375	0.3013
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1936	0.0373	0.3014
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1936	0.0373	0.3014
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1937	0.0373	0.3015
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1937	0.0373	0.3015
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1937	0.0373	0.3015
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1937	0.0373	0.3015
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1572	0.1937	0.0373	0.3015
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1573	0.1937	0.0373	0.3015
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1573	0.1937	0.0373	0.3016

Study Case 2: Station Empangan Genting Kelang (E) and station Kampung Kuala Saleh (K)

Days	E(obs)	K(obs)	E(pre)	K(pre)	E(er)	K(er)	α_{11}	α_{12}	α_{21}	α_{22}
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2867	0.1974	0.0204	0.2479
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2867	0.1974	0.0205	0.2479
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2867	0.1974	0.0205	0.2479
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2868	0.1974	0.0206	0.2479
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2868	0.1974	0.0206	0.2479
29	5.2000	0.0000	0.0000	0.0000	-5.2000	0.0000	0.2868	0.1974	0.0207	0.2480
29	66.0000	24.3000	0.0000	0.0000	-66.0000	-24.3000	0.2868	0.1974	0.0207	0.2480
29	7.9000	1.2000	6.6008	0.0990	-1.2992	-1.1010	0.2694	0.1967	0.0190	0.2460

29	0.0000	0.0000	72.3596	28.6702	72.3596	28.6702	0.0916	0.0325	0.0280	0.1097
29	3.8000	2.3000	0.0000	0.0000	-3.8000	-2.3000	0.1938	0.0321	0.0108	0.1870
29	11.6000	6.2000	0.0000	0.0000	-11.6000	-6.2000	0.1934	0.0321	0.0101	0.1871
29	7.0000	0.5000	4.6056	2.7657	-2.3944	2.2657	0.1928	0.0317	0.0098	0.1862
29	2.7000	3.5000	13.2479	7.0210	10.5479	3.5210	0.1957	0.0311	0.0112	0.1881
Days	**E(obs)**	**K(obs)**	**E(pre)**	**K(pre)**	**E(er)**	**K(er)**	α_{11}	α_{12}	α_{21}	α_{22}
29	5.2000	4.0000	5.8543	0.0000	0.6543	-4.0000	0.2084	0.0328	0.0071	0.1988
29	4.7000	3.3000	1.8842	4.0555	-2.8158	0.7555	0.2118	0.0316	0.0103	0.1999
29	1.0000	0.0000	5.7478	4.1267	4.7478	4.1267	0.2126	0.0326	0.0103	0.2017
29	0.0000	0.0000	4.5688	3.1509	4.5688	3.1509	0.2157	0.0334	0.0111	0.2050
30	0.5000	0.0000	0.0895	0.0000	-0.4105	0.0000	0.2162	0.0335	0.0105	0.2055
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2162	0.0335	0.0104	0.2056
30	0.0000	0.0000	0.6081	0.0052	0.6081	0.0052	0.2162	0.0335	0.0104	0.2055
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2161	0.0335	0.0104	0.2056
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2161	0.0335	0.0104	0.2056
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2161	0.0335	0.0104	0.2056
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2162	0.0335	0.0104	0.2056
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2162	0.0336	0.0104	0.2057
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2162	0.0336	0.0104	0.2057
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2162	0.0336	0.0104	0.2057
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2162	0.0336	0.0105	0.2057
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2162	0.0336	0.0105	0.2057
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2162	0.0336	0.0105	0.2058
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2162	0.0336	0.0105	0.2058
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2162	0.0336	0.0105	0.2058
30	0.8000	1.9000	0.0000	0.0000	-0.8000	-1.9000	0.2162	0.0337	0.0105	0.2058
30	4.2000	4.6000	0.0000	0.0000	-4.2000	-4.6000	0.2162	0.0337	0.0105	0.2058
30	0.0000	0.0000	1.0368	2.2990	1.0368	2.2990	0.2162	0.0336	0.0105	0.2056
30	0.5000	0.0000	5.0248	5.1891	4.5248	5.1891	0.2158	0.0337	0.0100	0.2056
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2156	0.0336	0.0097	0.2054
30	0.0000	0.0000	0.6078	0.0049	0.6078	0.0049	0.2156	0.0335	0.0098	0.2053

30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2156	0.0336	0.0097	0.2054
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2156	0.0336	0.0097	0.2054
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2156	0.0336	0.0097	0.2054
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2156	0.0336	0.0098	0.2054
Days	**E(obs)**	**K(obs)**	**E(pre)**	**K(pre)**	**E(er)**	**K(er)**	α_{11}	α_{12}	α_{21}	α_{22}
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2157	0.0336	0.0098	0.2054
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2157	0.0336	0.0098	0.2055
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2157	0.0336	0.0098	0.2055
1	0.0000	0.5000	0.0000	0.0000	0.0000	-0.5000	0.2157	0.0336	0.0098	0.2055
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2157	0.0337	0.0098	0.2055
1	0.5000	0.0000	0.0168	0.6028	-0.4832	0.6028	0.2157	0.0337	0.0098	0.2055
1	2.0000	1.5000	0.0000	0.0000	-2.0000	-1.5000	0.2157	0.0337	0.0099	0.2055
1	0.5000	0.0000	0.6079	0.0049	0.1079	0.0049	0.2157	0.0336	0.0099	0.2055
1	0.0000	0.0000	2.3737	1.8226	2.3737	1.8226	0.2157	0.0334	0.0099	0.2052
1	0.0000	0.0000	0.1262	0.0000	0.1262	0.0000	0.2158	0.0334	0.0097	0.2053
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2158	0.0334	0.0096	0.2053
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2158	0.0334	0.0096	0.2053
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2158	0.0334	0.0097	0.2053
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2158	0.0334	0.0097	0.2053
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2158	0.0335	0.0097	0.2054
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2158	0.0335	0.0097	0.2054
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2158	0.0335	0.0097	0.2054
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2158	0.0335	0.0097	0.2054

APPENDIX L

PUBLICATIONS/PRESENTATIONS

From the material in this report there are, at the time of submission, the following papers were published/presented or submitted for publications/presentations.

Papers Published in National Journals

P1. Fadhilah Yusof, Zalina Mohd Daud, Nguyen V-T-V and Zulkifli Yusop, Performance of Mixed Exponential and Exponential Distribution Representing Rain Cell Intensity in Neyman-Scott Rectangular Pulse (NSRP) Model, *Malaysian Journal of Civil Engineering*, Vol.10, No.1, 2007, pp 55-72.

P2. Fadhilah Yusof, Zalina Mohd Daud, Nguyen V-T-V, Suhaila S. and Zulkifli Yusop, Fitting The Best-Fit Distribution For the Hourly Rainfall amount in the Wilayah Persekutuan, *Jurnal Teknologi*, Vol.46 (C), June 2007, pp 49-58.

Papers published in International Proceedings

P3. Fadhilah Yusof.,Zalina Mohd. Daud., Nguyen V-T-V., Zulkifli Yusop, Evaluation of Neyman-Scott Rectangular Pulse (NSRP) model for modeling hourly rainfall series. *Proceeding of the XIIth Applied Stochastic Models and Data Analysis (ASMDA2007) International Conference,* Chania, Crete, Greece, May 20-June 1, 2007.

P4. Fadhilah Yusof.,Zalina Mohd.Daud, Maizah Hura, Nguyen V-T-V., Zulkifli Yusof, Assessment of the point process following the Neyman-Scott process. *International Conference On Mathematics and Sciences (ICOMS 2007)*, Ibnu Sina Institute, UTM. 28-29 May 2007.

Papers Published in National Proceedings

P5. Fadhilah Yusof.,Zalina MD., Nguyen V-T-V., Zulkifli Y , Stochastic modeling of rainfall using Neyman-Scott Rectangular Pulse (NSRP) Model. *Proceeding of the National Water Conference 2006 on Water for Sustainable Development Towards a Developed Nation by 2020,* Guoman Beach Resort, Port Dickson, Negeri Sembilan, 13-14 July 2006.

P6. Fadhilah Yusof.,Zalina MD., Nguyen V-T-V., Zulkifli Y, Stochastic modeling of hourly rainfall series using Markov Chain Mixed Exponential (MCME) model. Proceeding of Simposium Kebangsaan Sains Matematik ke 15, Concorde Hotel, Shah Alam, Selangor. June 5 -7 2007.

Papers Presented at the International Coferences

P7. Fadhilah Yusof.,Zalina MD., Nguyen V-T-V., Zulkifli Y, Evaluation of Neyman-Scott Rectangular Pulse (NSRP) model for modeling hourly rainfall series. The XIIth Applied Stochastic Models and Data Analysis (ASMDA2007) International Conference, Chania, Crete, Greece, May 20-June 1, 2007.

P8. Fadhilah Yusof.,Zalina MD., Nguyen V-T-V., Zulkifli Y, Assessment of the point process following the Neyman-Scott process. International Conference On Mathematics and Sciences, ICOMS 2007, Ibnu Sina Institute, UTM. 28-29 May 2007.

P9. Fadhilah Yusof.,Zalina MD., Nguyen V-T-V., Zulkifli Y, Using Markov Chain Process in the Fitting Procedure for the Neyman-Scott Rectangular

Pulse (NSRP) model with Mixed Exponential Distribution, The 9th Islamic Countries Conference on Statistical Sciences 2007 (ICCS-IX), Concorde Hotel, Shah Alam, 12-14 December 2007.

Papers Presented at the National Conferences

P10**.** Fadhilah Yusof.,Zalina MD., Nguyen V-T-V., Zulkifli Y, Stochastic modeling of rainfall using Neyman-Scott Rectangular Pulse (NSRP) Model. *National Conference for Sustainable development Towards a Developed Nation by 2020*. Guoman Beach Resort, Port Dickson, Negeri Sembilan. 13-14 July 2006.

P11. Fadhilah Yusof.,Zalina MD., Nguyen V-T-V., Zulkifli Y, Stochastic modeling of hourly rainfall series using Markov Chain Mixed Exponential (MCME) model. *Simposium Kebangsaan Sains Matematik ke 15,* Concorde Hotel, Shah Alam, Selangor. June 5 -7 2007.

P12. Zalina MD, Nurul Huda MA., Fadhilah Y, Maizah Hura A., Robiah A, W.Azli WH., Exploring the anomalies of rainfall events in the Klang Valley. *Water World Day 2007*. 15-16 April 2007.

P13. Nordila A., Zulkifli Y, Zalina MD, Characterization of Convective Storms in Kuala Lumpur. *Water World Day 2007*. 15-16 April 2007.

P14. Nordila A., Zulkifli Y, Zalina MD, Characterization of Convective Rains in the Klang Valley. *National Conference for Sustainable development Towards a Developed Nation by 2020*. Guoman Beach Resort, Port Dickson, Negeri Sembilan. 13-14 July 2006.

Submitted Presentation at International Conference

P15. Fadhilah Yusof.,Zalina MD., Nguyen V-T-V., Zulkifli Y, Comparative Study Between the Cluster-based Model and Markov-Chain Mixed Exponential

Model. In Preparation for *Water Down Under 2008*, Adelaide, South Australia, 12-14 April 2008.

Presentation at UTM Seminars

P16. Fadhilah Yusof.,Zalina MD., Nguyen V-T-V., Zulkifli Y, Using Poisson cluster process in modelling hourly rainfall. *Seminar Jabatan Matematik*, UTM, 27th September 2006.