Point process models for diurnal variation rainfall data

A thesis presented in partial fulfilment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY
IN STATISTICS

at Massey University
Albany (Auckland), New Zealand.

Norazlina binti Ismail

September 2014

Abstract

The theoretical basis of the point process rainfall models were developed for midlatitude rainfall that have different temporal characteristics from the tropical rainfall. The diurnal cycle, a prominent feature in the tropical rainfall, is not represented in the point process models. An extension of the point process models were developed to address the diurnal variation in rainfall. An observed indicator of the rainfall, X is added to the point process models. Two point process models, Poisson white noise (PWN) and Neyman-Scott white noise (NSWN) model were used as the main rainfall event, Y. The rainfall is modelled assuming two cases for the variable X, independent and dependent. Bernoulli trials with Markov dependence are used for the dependent assumption. To allow the model to display the diurnal variation and correlation between hours, the model was fitted to monthly rainfall data by using the properties of two hour blocks for each month of the year. However, the main point process models were assumed the same for each of the 12 blocks, thus having only one set of point process parameters for the models for each month. There are 12 rainfall occurrence parameters and 12 Markov dependence parameters, one for each block. A total of six models were fitted to the hourly rainfall data from 1974 to 2008 taken from a rain site in Empangan Genting Klang, Malaysia.

The PWN and NSWN models with X were first fitted with the assumption that the rainfall indicators are independent between the hours within the two hour block. Simulation studies showed the model does not fit the moments properties adequently. The models were then modified based on a dependence assumption between the hours within the two hour block. These models are known as the Markov X-PWN and Markov X-NSWN models. Both models improve the fit of the moment properties. However, having only one point process model to represent the rainfall events for Malaysia rainfall data was not sufficient. Since tropical rainfall consists of two types of rain, convective and stratiform, the PWN and Markov X-NSWN model were superposed to represent the two types of rainfall. A simple method by assuming non-homogenous PWN process for every two hour block did not fit well the daily diurnal variation. A comparison between the six models show that the superposed PWN and Markov X-NSWN model improved the fitting of mean, variance and autocorrelation. The superposed model was then simplified to an 8-block model to reduce the number of parameters. This modification to the point process models succeeded in describing the diurnal variation in the rainfall, but some of

the models were not able to fit other properties that were not included in the parameter estimation process such as the extreme values.

Acknowledgement

Alhamdulillah. The perfect, most beautiful praise is only for ALLAH. Thank you ALLAH.

2:286 On no soul doth ALLAH place a burden greater than it can bear. It gets every good that it earns, and it suffers every ill that it earns.

2:216 ... But perhaps you hate a thing and it is good for you; and perhaps you love a thing and it is bad for you. And ALLAH knows, while you know not.

I would like to express my sincere appreciation to my supervisors, Dr Barry McDonald and Associate Professor Dr Paul Cowpertwait, for their expertise, assistance, guidance, and patience throughout my PhD studies. It was a real privilage for me to have them both as my supervisors, Dr Paul, a well known researcher in his field and Dr Barry, who is willing to learn with me. Thank you so much Dr Barry for your constant support, availability, constructive suggestions, and careful editing which were determinant for the accomplishment of the work presented in this thesis.

I would like to extend my sincere gratitude to Freda, Annette, Yan Ou, Mike Yap and Lyn for their help and support with the basic needs to survive as a student at Massey University. To Dr Peter Kay, thank you for the guidance in my C programming.

To my dear husband, Mohd Nihra Haruzuan, thank you for supporting my dream and thank you for all the sacrifices you have made to see me reach my goals. To my sons, Aqil and Wafi, you guys are a precious gifts from ALLAH, the smiles and laughters brings mama joy everyday! To my parents Ismail and Zainab, my brothers and sisters, Nizam, Ani (Hemly too), Yan (& Adi) and Iwan, thank you for the support and encouragement. Not forgetting my parent in-law and my sister's in-law. I am blessed with a happy and loving family.

Completing this work would have been all the more difficult were it not for the support and friendship provided by the Malaysian community in Albany, the other fellow post-graduates and also the members of the Statistics Department. Special dedication to John Xie, for his help and advice.

Sincere appreciation is also due to the staff of Malaysia Meteorological Department and Drainage and Irrigation Department for providing the daily rainfall data. Finally, I would like to thank UTM and Ministry of Higher Education of Malaysia under the SLAI program for giving me the opportunity and funding which allowed me to further my study in New Zealand.

65:2-3 And whoever fears ALLAH - He will make for him a way out. And will provide for him from where he does not expect. And whoever relies upon ALLAH - then He is sufficient for him. Indeed, ALLAH will accomplish His purpose. ALLAH has already set for everything a [decreed] extent.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Rainfall is a natural feature of the earth's weather system. It is the main resource for fresh water for human and animal consumption, industrial operations and as irrigation for agriculture. The rainfall amount, intensity and frequency varies over the years, affecting the environment, agriculture and society. Rain that falls onto the land is absorbed into the soil and benefits plants, groundwater and conservation. The same amount of rainfall in a short period of time may cause flooding and runoff, thus damage property and the environment, and temporarily disrupt essential services such as transportation, telecommunications, energy and water supplies. In management of agriculture, the rainfall pattern is important to healthy living plants. Too little rainfall can cause drought, killing the crops. Heavy rainfall leads to high runoff that causes erosion and nutrient transport. In hydrology, rainfall analysis includes an analysis of storm events and of the intensity-duration frequency curve using the rainfall data. Information on the frequency of rainfall occurrence and its intensity is important for planning, design, and management of various water resources systems. Unfortunately the records of rainfall data in some locations may not be available, or maybe incomplete or insufficient in length and spatial coverage. There is a need to be able to simulate realistic sequences of daily or sub-daily rainfall not only for hydrological purposes, but also as a source of input data for other models. To obtain the full synthetic sequences of rainfall data, rainfall modeling has been used to simulate the rainfall data. The physical process of the observed rainfall, and its statistical properties, are emulated by the rainfall models.

A rainfall process involves complex atmospheric processes that progress continuously over space and time. To model the underlying atmospheric processes would be very complicated and it is not very useful in hydrologic or agriculture applications. But by understanding the structure of the rainfall process the rainfall modeling has evolved throughout the years. However, the rainfall physical mechanisms vary in different areas in the world. In the equatorial region, the main type of rainfall is the convective rainfall. It is caused by

the strong heating from the sun which causes significant warm air to rise, before cooling and condensing in the atmosphere to form comulonimbus clouds. The rainfall is heavy and usually accompanied by thunder and lightning. The mid latitude areas have moderate levels of rainfall. This type of rainfall is associated with the development of frontal depressions which form when the warm air and cold air collide, and is known as stratiform rainfall. The rainfall is light, but lasts longer than the convective rainfall. The rainfall patterns are also different depending on the seasons. Regions closest to the equator may have two wet seasons and two dry seasons, while regions further away, may have only one wet and dry season in a year. In mid latitude areas, winter have more rainfall than in summer. Rainfall models that would allow parameters to characterize these variabilities in the rainfall, would benefit agriculture, water resources management and hydrology.

An examination of the research literature has shown that the rainfall models currently available involve a mathematical description of the physical process, with large systems of equations as well as a modest number of parameters to represent the rainfall process and also the seasonal variation in rainfall. Cox & Isham (1994) and Onof *et al.* (2000) classified four categories of rainfall model based on the amount of physical realism incorporated into the model structure; models of dynamic meteorology, multi-scaling models, empirical statistical models, and point process models. In dynamic meteorology models, the physical processes are represented by nonlinear partial differential equations. These are usually used for weather forecasting. In multi-scaling models, the spatial evolution of the rainfall process is described in a scale-independent method. The empirical statistical models are fitted to the sequence of annual, monthly and daily rainfall and climate data. For rainfall data at short time intervals such as hourly, the point process models are more suitable. These models have fewer parameters compared to the other models and the parameters relate to the idealisations or analogues of the physical rainfall process such as rain cells, rain band and cell clusters.

The rainfall models available are not only profiling the structure of the rainfall process, but also include the dependence structure (i.e clustering dependence) shown by rainfall. This is the main feature in the point process models. These models are preferable compared to the other models because they have only a few physically interpretable parameters and also provide a way of simulating rainfall data. Since rainfall data are usually collected in discrete time, but the model is built in countinuous time, there is a slight discrepancy. To overcome this, the model considers the rainfall aggregated over discrete intervals. This will also allow the model to predict the main features of rainfall at scales other than those used in the fitting to historical data. Therefore, the model is versatile in terms of the ability to perform at different time scales.

Rainfall is the main input to the water resource system. The need for data and knowledge of rainfall processes is important in the design and operation of water resource systems, design of urban drainage systems, and land management for events such as floods or drought. Through rainfall modeling, the simulated synthetic rainfall series resembles the observations statistically, and therefore can be used for the purpose of understanding the physical process, data reduction or short term prediction, long term prediction and planning. However, the modeling of rainfall is not an easy task, due to the wide range of temporal and spatial variability. The choice between the different types of rainfall models depends on the main purpose behind the analysis.

1.2 Rationale of the research

The point process model for rainfall modeling was first introduced by Kavvas & Delleur (1975). Waymire & Gupta (1981a), Waymire & Gupta (1981b) and Waymire & Gupta (1981c) did a review on the theory of the point process with applications to rainfall data. The simplest point process models are the Poisson models. They are the Poisson white noise model with instantaneous pulse for the rain fall depth, and the Poisson rectangular pulse model. In the latter, the rainfall duration and depth are attached to the Poisson process, giving rise to the term rectangular pulse. The Poisson cluster model is represented by the Neyman-Scott or Bartlett-Lewis process. The Neyman-Scott white noise model takes the rain cells as being an instantaneous pulse. The Neyman-Scott rectangular pulse and Bartlett-Lewis rectangular pulse models assume cells have a random duration and a random intensity constant throughout the cell duration. These model were developed by Rodriguez-Iturbe *et al.* (1987a). Since the introduction of the Neyman-Scott rectangular pulse model and the Bartlett-Lewis model, many improvements have been made to the models. For example, adding a jitter process to the Bartlett-Lewis model (Rodriguez-Iturbe *et al.* 1987a) or using different distribution and assumptions for the rain cell depth and duration (Yusof *et al.* 2007, Onof *et al.* 2000, Northrop & Stone 2005).

One main aspect of the refinements was to incorporate the different types of rainfall in the model. These was done by having multiple cell types Cowpertwait (1994) and superposed point processes (Cowpertwait 2004, Cowpertwait *et al.* 2007, Morrissey 1993). These models were limited to two types of rainfall and two processes only in order to keep parameter numbers manageable. The generalized Neyman-Scott rectangular pulse model has rain cells of n types. The type of rainfall was based on the heavy, short duration convective rainfall and the light long duration stratiform rainfall. A two cell type Bartlett-Lewis model was developed by Hanaish *et al.* (2011b) to classify the rainfall data in Malaysia. Even though the two cell type type Bartlett-Lewis was superior to the original Bartlett-Lewis rectangular pulse model, the modified Bartlett-Lewis model (Rodriguez-

Iturbe *et al.* 1987a) was able to reproduce the statistical properties of the data better than the other two models. The same two types of rainfall were used in the superposed models. Each of the point processes represent a different type of rainfall. The process can be the same (e.g. in Cowpertwait *et al.* (2007), they superposed the Bartlett-Lewis pulse model with another Bartlett Lewis pulse model) or from different type of processes (e.g. in Cowpertwait (2004), Neyman-Scott rectangular pulse model was superposed with a Poisson rectangular pulse model). Morrissey (1993) characterized the random nature of small convective events and other events like the occurrence of large tropical convective cloud clusters, in a superposed model using Poisson white noise and Neyman-Scott rectangular pulse models, respectively. Morrissey's data was taken from the northern equatorial Pacific, an area with tropical rainfall. The superposed model performed better than the independent Neyman-Scott rectangular pulse model or the Poisson white noise model.

The majority of point process rainfall modeling by Foufoula-Georgiou & Lettermaier (1986), Rodriguez-Iturbe *et al.* (1987a), Onof *et al.* (2000), Cowpertwait (1994), Cowpertwait *et al.* (2007), to name a few, was applied to the mid latitude type of rainfall. A few have analyzed tropical rainfall, for example, Morrissey (1993), Smithers *et al.* (2002) and Lu & Qin (2012). For the Malaysia rainfall data, the only publicly available studies are Yusof *et al.* (2007), Yusof *et al.* (2008) and Hanaish *et al.* (2011b). Morrissey (1993) emphasizes the issue of the suitability of the temporal point process models for the tropics as these models were developed for midlatitude conditions. The major concern was that the point process does not feature a diurnal cycle. It is known that tropical rainfall has a significant diurnal variation due to the afternoon maximum rainfall from the daytime surface heating. Rodriguez-Iturbe *et al.* (1987b) also highlighted this type of rainfall and advised adding a daily cycle in the storm arrivals rate in the point process model to represent the diurnal cycle. Other types of rainfall models that modeled the diurnal pattern are the chain-dependent process of Katz (1995) and the hybrid models of Gyasi-Agyei (2001) and Gyasi-Agyei & Willgoose (1997). In the chain-dependent process, the rainfall occurrence and the rainfall amounts were modeled separately and the diurnal variation was captured in the rainfall occurrence model. The hybrid model also involved two types of process that are products of each other. Both models show the diurnal variation is evident in the transition probabilities of the Markov chain model used in the models.

The focus of this thesis is on modelling Malaysia rainfall data. Malaysia lies on the equator. The climate of Malaysia is influenced by the southwest monsoon and northeast monsoon. The rainfall distribution pattern is determined by the seasonal wind flow patterns together with the local topographic features. There are two periods of maximum rainfall which occur in October to November and April to May. Two periods of minimum rainfall occurs in January to February and June to July. The daily rainfall pattern shows a significant diurnal variation, with a peak during late afternoon. Therefore, from review-

ing the literature, there are gaps in the point process models for capturing this diurnal variation in the rainfall data. Although the Malaysian researchers (Yusof *et al.* (2007) and Hanaish *et al.* (2011b)) obtain good results using the basic Neyman-Scott rectangular pulse model or the modified Bartlett-Lewis model rectangular pulse model, the effect of a strong diurnal cycle in Malaysia rainfall data cannot be ignored. The superposed model using the Neyman-Scott process has not been explored for the Malaysia rainfall data. In summary, the rationale of this research lies in having a diurnal variation incorporated within the point process model, which will help improve the modeling of tropical rainfall data. The simulated rainfall series will be useful in water resources management and hydrologic applications.

1.3 Research objectives

The aim of this study is to develop a point process model for diurnal variation rainfall data in Malaysia. To achieve this aim, several objectives were defined as follows:

1. To apply the nonhomogenous Poisson white noise model to the hourly data.
2. To modify the Poisson white noise model to include an indicator variable with an assumption of independence and dependence between hourly rainfall.
3. To modify the Neyman Scott white noise model to include the indicator variable with an assumption of independence and dependence between hours.
4. To improve the modified models' performance by superposing the modified Poisson white noise and the modified Neyman-Scott white noise model.
5. To assess the feasibility of simulating the hourly rainfall process using the proposed modified models and the superposed models.
6. To reduce the number of parameters in the superposed models.

1.4 Research approach

Two point process models were adopted in this study, the Poisson white noise model and Neyman Scott white noise model. Both models were modified by adding a diurnal variation parameter to the main model. The corresponding indicator variable is assumed to follow the diurnal variation in the daily rainfall data. Therefore, the point process models were assumed stationary within each calendar month and the indicator variable was fitted for every two hour block throughout the day. The modeling involves pooling the observed data for each calendar month. Treating calendar months separately allows for seasonal variation. But the sample moments, mean, variance and autocorrelation of the hourly data in the two hour block scale are to be evaluated for each month. The annual extreme values

are also evaluated. To estimate the model parameters, the sample moments are fitted to the model's first-order and second-order properties. The Nelder-Mead optimization algorithm was used to estimate the parameters. These parameters are then used to simulate the hourly series. The simulated statistics are compared with the historical values at the hourly or the two hour block time scale using numerical analysis and graphical representations. The two point process models are then superposed. The superposed and other models are evaluated and compared. Reducing the number of parameters in the models will conclude the study.

1.5 Scope of study

This study focuses on the hourly rainfall series and single-site rainfall modeling only. Therefore, a complete hourly rainfall series at a site is essential. On this basis, a rainfall station located in the Kuala Lumpur region was chosen for this study. Kuala Lumpur is the Malaysia capital city and it is the most urbanized and industrially developed city in Malaysia. Flash-flooding occurs frequently in this region due to the high occurrence of convectional rainfalls and a poor drainage system. There are 13 rainfall stations located in this area. The Empangan Genting Klang rainfall station has the most complete series and the longest with 36 years data from 1972-2008. Only 2.5% of the data is missing.

1.6 Thesis outline

This thesis consists of eight chapters. The outline for each chapter is discussed below.

Chapter 2: Literature Review
This chapter presents the relevant literature review on the empirical statistical models and the point process models. The review on empirical statistical models focuses on the two-part rainfall models which consist of an occurrence model and a model for the rainfall amounts. The point process models focus on the white noise model, rectangular pulse models, and also the previous point process models using Malaysian rainfall data. Stochastic rainfall models for data with diurnal variation are also reviewed in this chapter.

Chapter 3: Methodology
In this chapter there is a detailed description of the models used in this study. There is a full description of the methodology of the white noise model, Bernoulli trial with dependence structure, and modification to capture the diurnal variation. The model parametrisation and fitting plus the simulation process and the model assessment are explained in detail. An exploratory analysis is carried out for the Malaysia rainfall data.

Chapter 4: Non-homogenous Poisson white noise model
In this chapter, the Poisson white noise (PWN) model was fitted to every 2 hour block throughout the day. The non-homogeneous PWN model specification, properties and the parameter estimation process are explained in details. Model verification was done by comparing the two hour moments and the monthly moments.

Chapter 5: X-PWN model & Markov X-PWN model
The modification to the PWN is formulated in this chapter. An indicator variable and dependent variable are introduced to add to the original PWN model. Two new models, the X-PWN model and Markov X-PWN model, based on the assumption of independence and dependence, respectively are investigated in detail. The model performance comparison results are presented which include moment properties, extreme values and the mean square error.

Chapter 6: X-NSWN model & Markov X-NSWN model
Using the same modification in chapter 5, the Neyman-Scott white noise (NSWN) was tested in this chapter. Two new modified models were discussed in this chapter, the X-NSWN model and the Markov X-NSWN model. The NSWN process was chosen because of the simplicity of the model in describing the rainfall process. The model performances are compared between the NSWN modification models and the PWN modification models based on the derived model properties, extreme values and the mean square error.

Chapter 7: Superposed Model and 8-block Model
The Markov X-NSWN model was superposed with PWN process in chapter 7. The superposed model was intended to improve the fit in terms of the moments properties and the extreme values. However, the superposed model has too many parameters, and therefore in the second part of this chapter, the 12-block superposed model was reduced to an 8-block model.

Chapter 8: Conclusions and recommendations
Finally, the last chapter gives an overall conclusion of the major research results in the preceeding chapters and a few suggestions for future research.

Throughout this study, the open source statistical package R (R Development Core Team 2008) and the open source C language package (The GCC team 2007) were used for processing the raw data, statistical analysis and the simulation process.

Chapter 2

Literature review

2.1 Introduction

The need for a model of the rainfall process that is mathematically managable for operational purposes, has motivated the treatment of rainfall as a stochastic process (Foufoula-Georgiou & Georgakakos 1991). The four types of rainfall models, the dynamic meteorology model, multi-scaling models, empirical statistical models, and point process models, discussed in Chapter 1 have different roles, and the choice between them depends on the purpose of the analysis of the rainfall data. In this chapter, the literature review focuses on empirical statistical models and the point process model.

Srikanthan & McMahon (2001) review the empirical statistical models for different time scales (annual, monthly and daily), type of data (rainfall and climate data) and site (single and a number of sites over an area). Fewer studies looked at the annual and monthly rainfall data compared to the daily rainfall data. Since rainfall is generally recorded at the daily time scale and this forms the basis for monthly and annual rainfall series, more research has been done on the modelling of the daily rainfall process. There are four different types of models for the empirical statisticals model: two-parts models, transition probability matrix models, resampling models, and time series models of the ARMA type. The two-parts models will be discussed further in this chapter.

A point process is a stochastic process that consists of a sequence of points, with each point being characterized in time or space (Faofoula-Georgiou 1985). For example, if the cumulative rainfall amount exceeding a specified threshold over a one day period in a selected area is recorded, then the sequence of daily rainfall forms a point process (Faofoula-Georgiou 1985). A detailed and mathematical treatment of point process theory is given by Cox & Isham (1980). To link point process theory to rainfall, two appoaches have been suggested in the literature (Ramesh 1998). The first is by defining an event of rainfall as an interval with measurable precipitation and describing the probability stucture of the sequence of rainfall events as discrete time point process models. The second

approach is by assuming that the outcome of an underlying continuous-time rainfall generating mechanism is only observed as the integral of the continuous process over the given sampling interval.

In the simplest stochastic point process model, storms arrive according to a Poisson process. Each point of a Poisson process of rate λ per unit time is associated with a pulse (often assumed to be rectangular) (Rodriguez-Iturbe *et al.* 1987a). The pulses for each cell have random height (intensity) and length (duration) that are independent and constant throughout the cell duration. Rodriguez-Iturbe *et al.* (1987a) fit the hourly rainfall data at Denver, Colorado to the Poisson process, but was not able to represent rainfall data for different time scales. To overcome this inadequacy, they used a Poisson cluster based model.

The Neyman-Scott cluster model was developed by Neyman & Scott (1958) to describe the cluster distribution of the galaxies. It is considered as the classical model in point process theory that became a basic foundation in many different area of sciences. Modeling daily rainfall events using a Neyman-Scott cluster process was introduced by Kavvas & Delleur (1975). They symbolized the rain cell as an instantaneous burst and associated the depth of the rain cell with a distribution. This was known as Neyman-Scott white noise model. Waymire & Gupta (1981a, 1981b, 1981c) also demonstrated how the counting process of Neyman-Scott could be used to model rainfall events.

The rainfall process is modeled as a sequence of storms origins that consists of a cluster of rain cells with pulses. The total storm intensity is the summation of intensities of all the raincells at that time. Two well known Poisson cluster models reported in the literature are the Neyman-Scott process (Rodriguez-Iturbe *et al.* 1987a, Cowpertwait 1991) and the Bartlett-Lewis process (Rodriguez-Iturbe *et al.* 1987a). The storms arrive according to a Poisson process for both these models. The difference between the models is in the arrivals of cell origins. In the Neyman-Scott process, the cell arrivals are measured from the storm origin and are independent and exponentially distributed (Onof *et al.* 2000). For Bartlett Lewis model, the storm origin is followed by another Poisson process of cells origin. The lifetime of the cells is exponentially distributed and all of the cells are mutually independent. But Cowpertwait (1998) showed that both models have the same statistical properties up to second moments. Stochastic rainfall models have evolved further in past decades. Different approaches to the modelling and simulation of precipitation have been progressing significantly. The evolution of the models is presented as follows.

2.2 Stochastic rainfall models

2.2.1 Two-part models

In a rainfall model, two main processes of the rainfall are the occurrence and the intensity. Both of these processes are usually modeled separately and then superposed to describe the complete rainfall process. The sequence of wet and dry hours (or days) can be viewed as a binary series of ones and zeroes. The wet hours correspond to ones and dry hours to zeroes. The Bernoulli process is the simplest probability model to describe the binary series. The sequence is assumed independent and Foufoula-Georgiou & Georgakakos (1991) claim that this is not suitable to describe the clustering existing in the short interval time of rainfall occurrences. Klotz (1973) added a Markov dependence parameter to the Bernoulli trials. The next simplest model to fit the daily or hourly rainfall occurrence with a relation between the condition of the current day and the condition of the preceding days is the case with Markov chain models. The order of the Markov chain refers to the number of previous days taken into account. Gabriel & Neumann (1962) first used a two-state Markov chain for the winter daily rainfall in Tel-Aviv. The first order Markov chain is preferable than the higher order because it involves less parameters. Coe & Stern (1982) compared the first- and second-order Markov chain with different transition probabilities according to the the time of year. The Akaike Information criterion (AIC) developed by Akaike (1974) can also be used to determine the order of Markov chains.

Smith & Karr (1983) developed a point process model using a Cox process for the rate of the rainfall occurrence which alternates between two states, zero for no event and a positive state a for rainfall event. During the state of zero intensity, no events can occur. The model selection for the rainfall occurrence over the summer season in the Potomac River basin was based on the number of events in an interval and the interarrival time statistics. Between the renewal processes, Cox processes and Poison cluster processes, it was shown that the most suitable model for the rainfall data was the intersection of the renewal processes and Cox processes. This model is called a renewal Cox process with Markovian intensity (RCM). The RCM process and the occurrences of a rate zero suggest that the clustering of the rainfall is based on the sequence of wet and dry periods.

A Markov renewal model for rainfall occurrences was developed by Foufoula-Georgiou & Guttorp (1987). They assumed two different geometric distributions for the time between rainfall occurrences and the Markov chain was used for the transition between the two distributions. Their model was built based on the primary or secondary rainfall events that exhibit the clustering dependencies in the daily rainfall occurrence which the Markov chain does not fit well. The probability of a wet day for the Markov renewal model does not depend on the previous day condition but on the number of days since the last rainfall event. The model was fitted separately for five seasons of daily rainfall data from

Washington, USA. The model was able to preserve the short-term structure of the rainfall occurrence process and also the distributional properties of the seasonal rainfall amounts.

Smith (1987) developed the Markov Bernoulli model, a discrete point process model for daily rainfall occurrences. The process comprises a sequence of Bernoulli trials with randomized success probabilities based on the first-order, two-state Markov chain. A more complex model of wet-dry sequences can be generalized from the Markov Bernoulli process. The main feature they developed was a recursive formula for calculating the Markov Bernoulli stochastic intensity. The point process approach was the main source to develop the procedures for parameter estimation and model selection. Thus the parameters were estimated using maximum likelihood, and selecting the best model (from Markov chain, Bernoulli or Markov Bernoulli processes) and using likelihood ratio tests. The model selection method, using the inference procedures, was tested using a wet-dry sequence from Washington, D.C.. Three different precipitation boundaries, 0.01, 0.10 and 1.00 inches have different types of model fitted. The Markov Bernoulli was superior to the other models for the moderate precipitation boundaries.

The most common approach to model the rainfall amounts is to assume that rainfall amounts on wet days are independent and fit some theoretical distribution (Todorovic & Woolhiser 1975). Theoretical distributions used for the daily rainfall amounts are Gamma distribution (Katz 1995, Coe & Stern 1982), Exponential distribution (Todorovic & Woolhiser 1975) and Mixed Exponential distribution (Foufoula-Georgiou & Guttorp 1987, Yusof 2007). Another approach for modeling the rainfall amount is the chain-dependent process (Katz 1995), where the rainfall is assumed independent but the distribution function depends on whether the preceding day was wet or dry.

2.2.2 White noise model

The white noise model refers to the Poisson white noise (PWN) model and the Neyman-Scott white noise (NSWN) model. Both models assume that in a rainfall event, the rainfall amount (or the pulse depth) occurs as instantaneous bursts or pulses. The PWN model associates the rainfall occurrence with the instantaneous pulses and assumes the rainfall occurrence to follow the Poisson process. The NSWN model is a Poisson cluster process that assumes the rainfall occurrence follows the Neyman-Scott process with the instantaneous pulses. In the Neyman-Scott process, the storm arrivals are Poisson distributed. Each storm arrival is associated with the random number of rain cell origins. The waiting time for the cell origins from the storm origins are assumed independent and identically distributed. The NSWN process and the PWN process integrates the actual rainfall process in the model.

Kavvas & Delleur (1975) and Kavvas & Delleur (1981) observed the clustering of the rainfalls in Indiana, USA and derived the probability generating function for the NSWN model based on two processes, the rainfall generating mechanism and the rainfall events. Rodriguez-Iturbe *et al.* (1984) used the method of moments instead, by finding the mean, variance and correlation function based on the aggregated rainfall depths. They overcame the discrepancy between the use of a continuous time model and the discrete time series of the observed rainfall that Foufoula-Georgiou & Lettermaier (1986) identified with an earlier white noise model, by aggregating the rainfall amount at a fixed time interval that is identical with the historical rainfall series. Rodriguez-Iturbe *et al.* (1984) fit the PWN model, NSWN model and the Poisson rectangular pulse (PRP) model to hourly and daily rainfall data at Denver, USA and Agua Fria, Venezuela. They found that the PWN and PRP model have different parameter estimates for the hourly and daily rainfall. Only the NSWN model is not scale dependent and produces the same parameter estimate using the hourly and daily data. Thus the NSWN process was superior to the other two models, and which is the same conclusion as reached by Kavvas & Delleur (1981).

Valdes *et al.* (1985) also compared the three models and found that the NSWN model performed better for time scales from, 1 to 24 hours. But all of the models fail to reproduce the extreme values. Smith & Karr (1985) showed that for a Neyman-Scott process, it could be represented as a Cox process with the cluster size distribution as Poisson and the distance between cluster center and cluster members as exponential distributed. They also estimated the parameters using maximum likelihood procedures. But it was pointed out that using maximum likelihood procedures in estimation and model selection for space-time rainfall models would be difficult. Foufoula-Georgiou & Guttorp (1986, 1987) also found difficulty estimating the NSWN parameters using the maximum likelihood and also used the method of moments. The model was fit to different time scales from hourly to daily rainfall data. They showed that the model cannot produce consistent parameter estimates at all the time scales.

2.2.3 Rectangular pulse model

Apart from the white noise model, Rodriguez-Iturbe *et al.* (1984) also studied the Poisson rectangular pulse model. The model process starts with the rainfall occurrence process that follow a Poisson process. Each rainfall occurrence has a pulse with a rectangular shape that characterizes the rainfall intensity for the height and the duration of the rainfall event for the length. The intensity and the duration are assumed identically and independent distributed. They derived the moment properties of the aggregated rainfall and used it for hourly and daily rainfall data. Similarly to the PWN model, the model cannot preserve the moments at different time scales.

The Poisson rectangular pulse model was further studied by Rodriguez-Iturbe *et al.* (1987a). They derived another property, the probability of a dry period. The hourly rainfall data from Denver, Colorado was used to fit the model. Comparison was made with different level of aggregation (1, 6, 12, and 24 hour). Again, the model was not able to fit across different time scales other than the hourly data that was used in the estimation process. Therefore Rodriguez-Iturbe *et al.* (1987a) developed rainfall models that are time scale variant based on the clustered point process of the rectangular pulse. Two point processes were chosen, the Neyman-Scott process and the Bartlett-Lewis process.

The Neyman-Scott rectangular pulse (NSRP) model was developed to model the rainfall process at a single site. The NSRP process begins with the generating mechanism called the storm origin that is assumed Poisson process with rate λ. The storm origin may be passing fronts or some other criteria for convection storms that generates a random number of rain cells. There are no rain cell located at the storm origin. The distribution of the number of rain cells is Poisson distributed or geometric distributed with mean μ_C. The cell origins are independently and exponentially distributed with parameter β from the storm origin. A rectangular pulse associated independently with each cell origin represents the random duration and intensity. Both are exponential distributed with parameter η and ξ respectively. The total intensity at a time point is the sum of the intensities of the rain cells. The description of the NSRP model gives five parameters, λ, μ_C, β, η and ξ .

The Bartlett-Lewis rectangular pulse (BLRP) model also has five parameters, λ, γ, β, ξ and η. The difference between NSRP and BLRP process is in the location of the rain cell origins relative to storm origins. It is assumed that there is a cell origin at the storm origin. The rain cells arrival follows another Poisson process with rate β. The duration of the storm is exponentially distributed with mean γ. The number of cells is geometric distributed with $\mu_C = 1 + \beta/\gamma$.

Both models can generate simulations in continuous time to allow the aggregation of the properties and simulations at different timescales. Rodriguez-Iturbe *et al.* (1987b) carried out full analysis for NSRP and BLRP models using rainfall data from Denver, Colorado. Both models were able to preserve moments up to second order, including the extreme values for different time scales. But both models overestimated the proportion of dry days. To overcome this problem, Rodriguez-Iturbe *et al.* (1988) randomised the parameter for the cell duration distribution in the BLRP model, varying between storms giving a significant improvement in results. The same results were also obtained when Onof & Wheater (1993) applied the random parameter BLRP model to British rainfall data. But the model did not produce a good fit of the extreme values and also overestimated the daily autocorrelations for some months. Burlando & Rosso (1991) proved that the modified Bartlett Lewis model studied by Islam *et al.* (1990) reproduced the same ob-

served statistics as the original BLRP model. The proportion of dry periods also provides a closer fit to the observed values than does the modified model. But they used different minimization algorithms to estimate the parameters, indicating that the results may be affected by the method of estimation.

Islam *et al.* (1990) also did a parameter sensitivity analysis of the modified Bartlett Lewis model. The analysis was to determine which storm characteristics influence the statistical process of rainfall. Results show that the mean cell duration and its interstorm variance have the most significant effect on the rainfall process. Khaliq & Cunnane (1996) also analyzed the parameter sensitivity, finding that the expected intensity of a cell is the least variable. The most variable parameter was the gamma scale parameter which controls the cellular structure of each storm.

The NSRP model was modified by Entekhabi *et al.* (1989) by assuming the cell duration as a random variable as opposed to a constant parameter in the original rectangular pulse model. The probability density function is assumed to be a two-gamma distribution. The model was then able to preserve the proportion of dry periods from the scale of an hour to several days. This is an important variable in hydrological application. The probability of a dry period may also be used in fitting the model or for model validation. Cowpertwait (1991) also derived the probability that an arbitrary interval is dry for the original NSRP model. He also derived the distribution function of the storm duration that can be used to explore the sequence of wet and dry days. He fitted the model to hourly data from Blackpool, England with the expression for the proportion of dry included in the parameter estimation procedure. The NSRP model was able to preserve the main features of the rainfall and also the proportion of dry days for hourly and daily rainfall data.

To improve the results for the BLRP model, a jitter process was included to the model. Rodriguez-Iturbe *et al.* (1987a) introduced a high frequency jitter process to overcome irregular rainfall. Onof & Wheater (1994b) applied two types of jitters, adding to the continuous time process or to each cell to improve the autocorrelation structure in the BLRP model. The number of parameters increased to 7 (from 5 parameters in the original Bartlett-Lewis model), resulting in difficulty to estimate the parameters. They also used a Gamma distribution for depth, improving the reproduction of extreme rainfall depth. This modified version is termed the Bartlett-Lewis rectangular pulse gamma model (BLRPGM). Cameron *et al.* (2001) used a generalised Pareto distribution for the depth of intensity of raincells in a modified Bartlett-Lewis model. The model is used for extreme rainfall simulation of 1 hour and 24 hour data. Comparing the original Barlett-Lewis model, the modified Barlett-Lewis model and the BLRPGM model, all of these models underestimate the extreme values for short duration rainfall (Onof & Wheater 1993).

Smithers *et al.* (2002) compared the modified Bartlett-Lewis and BLRPGM model for South African short rainfall data. They suggested using BLRPGM when only daily rainfall data are available.

The NSRP and BLRP models only define one type of raincell. But according to Cowpertwait (1995) there are two types of storm, heavy convective storms and light stratiform storms. The heavy convective storms have raincells with shorter expected lifetimes than the light cells. This differentation is important because different types of storms correspond to different relations between the depth and the duration. He then developed the generalized Neyman-Scott rectangular pulses (GNSRP) model that allows the generated cell to be of n types. He considered n types of cells to be two, for the convective and stratiform rainfalls. The model can also be applied to multisite rainfall data. The model gives a good fit to the extreme values but was not applied to the proportion of dry periods. The GNSRP model was further studied by Cowpertwait & O'Connell (1997) but using the harmonic estimates from the regression method to reduce the number of parameters.

Another method to allow for different types of rainfall is by a superposed process (Cowpertwait 2004), (Cowpertwait *et al.* 2007). Cowpertwait (2004) assumed the NSRP process represents clusters of convective cells and the Poisson rectangular pulse model represents the isolated convective cells. He superposed the NSRP model with a Poisson rectangular pulse model to make use of the existing NSRP functions that have been derived. The superposed model was able to fit sample properties up to third moment over a range of time scales. The third moment function for the NSRP model was introduced by Cowpertwait (1998). This function was added to the fitting procedure and the model showed a good fit to the observed extreme values at all of the aggregation levels. If the third moment function was not included in fitting the model, the extreme values plot shows a lack of fit. The latest method to allow for different types of rainfall in the model is by assuming a continuum of storm types of random type Z by Cowpertwait (2010). Z is assumed a uniformly distributed and NSRP properties were re-derived to be functions of Z. This model superposed different types of storm within the same storm process.

Other modifications to the model include the assumption of dependence between the cell depth and duration Onof *et al.* (2000). Evin & Favre (2008) developed the NSRP model that considered correlated raincell depth and duration using cubic copula. Northrop & Stone (2005) changed to a more physically realistic form of the rainfall depth structure, from constant rectangular pulse to truncated Gaussian form.

2.2.4 Pulse model

Cowpertwait *et al.* (2007) developed the original Bartlett-Lewis model that can be applied to finer resolution data for example 5 minute rainfall data. The constant cell intensity is replaced by a sequence of instantaneous rainfall pulses. Thus the name Bartlett-Lewis Pulse (BLP) model. Moments up to third order were developed and used for both model fitting and the assessment of model adequency. The BLP model is defined and parameterised as follow.

Storm arrivals occur randomly according to a Poisson process T_i with rate λ. Each of the storm origins generated cell origins in another Poisson process of rate β. Storm durations are independent and exponentially distributed with parameter γ. The cell duration is assumed to be exponentially distributed with parameter η. Each rain cell initiates a sequence of rainfall pulses in a Poisson process of rate ξ. There are no rain cell at the storm origin and at the cell origin, so a storm or rain cell may have no rainfall. This assumption differs from the BLRP model. The pulse intensity is assumed exponentially distributed with parameter θ. This is a marked point process and follow the Bartlett-Lewis process (Cox & Isham 1980). The pulse process stops with the cell or storm duration, whichever is the sooner. The pulse depths from a single cell are dependent, and from distinct cells the model assumed depths are independent.

The superposed independent processes were used to define different storms, convective and stratiform rains. There are a total of 11 parameters, with the same depth for both storms for the pulse model. An exact fit of the BLP model was obtained from 5 minutes up to daily rainfall for all of the statistical properties using the 5-minute Kelburn, New Zealand rainfall data. An exception was for the extreme values for the 5 minutes aggregation level. The BLP model also overestimated the proportion of dry for 1 hour interval. Cowpertwait *et al.* (2011) and Xie (2010) assumed the pulse depths within the same cell to be dependent. The revised BLP model added a conditional exponential distribution to the pulse depths. Using the same rainfall data, the modification improved the fitting of the extreme values and also the probability of dry intervals. The BLP model was further developed by Kaczmarska (2011) by having the random cell duration parameter to vary randomly between storms. This allow the durations of cells within a single storm to be dependent. With only seven parameters, the model succeeded in fitting the 5-minute rainfall data from Bochum, Germany, and performed better than the other Bartlett-Lewis models.

Xie (2010) formulated the BLP models with continuous distributions of storm type using the same approach as Cowpertwait (2010). Moments up to third order were derived and the continuous storm types BLP models were compared to the orginal BLP model by Cowpertwait *et al.* (2007). However, the model fitted poorly to the corresponding historical values. The superposed BLP model with the same number of parameters of the

continuous-storm-types BLP models, eight, performs better in fitting the rainfall data. Even the superposed continuous-storm-types BLP models achieved insignificant improvements in fitting the extreme rainfall values and the proportion of dry periods. These results suggest that the assumption of equal occurrence probability for all types of storms in the continuous-storm-types BLP models is not reasonable.

2.3 Malaysia point process model

The point process models from Rodriguez-Iturbe *et al.* (1984), Onof & Wheater (1993), Cowpertwait (1991), Cowpertwait *et al.* (2007) to Kaczmarska (2011) (to name a few researchers), have been successfully used to describe the rainfall processes. The majority of these developments were for midlatitude conditions. Tropical rainfall has different temporal characteristics and Malaysia rainfall is influenced by the monsoon seasons and consists of heavy rainfall in short duration and light rainfall in long duration. Not much research has been done on the application of point process models to the Malaysia rainfall data. Yusof *et al.* (2007) applied the NSRP model to hourly data from Kuala Lumpur. They compared two types of distributions for the rain cell intensity, exponential and mixed exponential. Both models perform about equally well but based on the root mean square errors, the NSRP model with mixed exponential distribution had a smaller error. They also applied the Shuffle Complex Evolution optimization method compared to the Nelder-Mead optimization technique that is usually used by Cowpertwait (1991). The seasonal variation of the Malaysia rainfall data was captured by using the fourier series in NSRP model (Yusof *et al.* 2008).

Hanaish *et al.* (2011b) developed a Bartlett Lewis model with two cell types to classify the rainfall data in Malaysia as convective and stratiform. They compared three types of Bartlett-Lewis model, the original BLRP, the modified BLRP and the two cells BLRP for hourly rainfall data from Petaling Jaya, Malaysia. Comparisons were made for the dry and wet periods only, July and November. The two cell BLRP model performed better than the original BLRP model, but the modified BLRP model was the best model to fit the rainfall data. The Nelder-Mead optimization technique was used to estimate the parameters. Hanaish *et al.* (2012) studied the modified BLRP model in reproducing the variance over six different aggregated time scales. The model can fit the hourly data and performs better at the 6 and 12 hour time scale compared to the 2 and 3 hours aggregated level. The modified BLRP model was also used by Hanaish *et al.* (2011a) to disaggregate daily rainfall to hourly data. A validation of the model using the extreme values shows that the model did not give a good fit to the time scale of 1, 6 and 24 hour levels of extreme values.

Both point process models, NSRP and BLRP models, have been tested for Malaysia rainfall data. Although both researchers, Yusof *et al.* (2007) and Hanaish *et al.* (2011b) used different rainfall stations, they are nonetheless in the same area, Kuala Lumpur. Although both models can fit the Malaysia rainfall data, the results are not as successful compared to other researchers that applied them to the midlatitude conditions. Another area in South East Asia with the same type of rainfall where researchers applied the point process model is Singapore (Lu & Qin 2012). They compared 5 types of disaggregation model, the Poisson retangular pulse model, BLRP, two cell BLRP, BLRP with cell depth distribution dependent on duration and NSRP model. The rainfall in Singapore is mainly from the convective rainfall and all of the Bartlett-Lewis models gave a good disaggregation performance. The assumption of the independent and identically distributed intervals between the rain cells of the BLRP model describe the local rainfall patterns more accurately than the assumption of the NSRP model.

2.4 Diurnal variation

Diurnal variability of rainfall is a prominent occurrence event in the humid tropics. The variations depend on the location, season and precipitation type (Reading, A. and Millington, A.C. and Thompson, R.D. 1995). In Malaysia, the seasonal differences of climate caused by the monsoon season are small and limited to winds and precipitation only (Nieuwolt 1968). In some part of Malaysia, the diurnal weather processes are influenced by the monsoons and show a strong seasonal difference (Nieuwolt 1968).

According to Nieuwolt (1968), there are two types of diurnal rainfall in Malaysia, the inland type and the marine type. The peak rainfall in the late morning and afternoon due to convection of intense surface heating of the land is the inland type whereas the maximum rainfall occurring during the night or early morning is the marine type. Nieuwolt (1968) also divided the spatial variation in the diurnal cycle of precipitation into three regional types, the west coast, the east coast and the inland. The inland regime includes Kuala Lumpur with an afternoon maximum of rainfall (Ramage 1964, Oki & Musiake 1994).

Ramage (1964) identified five different diurnal rainfall patterns categorized according to the Malaysia regimes for August rainfall. The differences are caused by interactions among the synoptic wind, local land, sea, anabatic and kanabatic winds and the topography. Kuala Lumpur is part of the inland-mountain regime with warm dry air during the night as katabatic winds set in and convectional heating during the day accounts for the afternoon rainfall maximum.

Since the modelling of rainfall data mainly concerns the midlatitude type of weather, the summer rainfall data there may have a prominent diurnal variation. Some researchers (for examples Rodriguez-Iturbe *et al.* (1987b) and Morrissey (1993)) have found the diurnal cycle to have minimum effect on the autocorrelations and chose to ignore the diurnal cycle and model the rainfall data using the original model without modification. Katz (1995) added an extension to a chain-dependent process and by modelling the diurnal cycles in a parsimonious way using Fourier series techniques. The chain-dependent process is a combined process of the occurrence process and the intensity process. In the occurrence process, the transition probabilities of a two-state, first order Markov chain were represented as cosine waves to model the diurnal cycles. Another approach is the hybrid model explored by Gyasi-Agyei (2001) and Gyasi-Agyei & Willgoose (1997). The hybrid model is a product of the occurrence process and an intensity process. Gyasi-Agyei (2001) also used a Markov chain up to third order for the occurrence process and the periodic discrete autoregressive. The rainfall amounts were assumed to be log-normal distributed. Fourier series were also fitted in the hybrid model to reduce the number of parameters. In both the models by Katz (1995) and Gyasi-Agyei (2001), the diurnal cycle in the occurrence process is reflected in the transition probabilities.

In the point process model, Rodriguez-Iturbe *et al.* (1987b) acknowledge the tropical rainfall, with a high percentage of precipitation in the late afternoon and evening, will have higher daily autocorrelation than the 12 hour autocorrelation. They suggested adding a daily cycle in the storm arrivals rate to represent the diurnal cycle. Obeysekera *et al.* (1987) propose, until an add-in diurnal periodicity is developed, to fit the model to just half-day rainfall data during the summer months when there is prominent diurnal variation. Morrissey (1993) also pointed out that the point process models available do not incorporate diurnal cycle and thus will not be able to replicate the autocorrelation form. Morrissey (1993) and Morrissey (2009) superposed the NSRP and PWN model to represent the tropical rainfall of Majuro and Truk and the tropical rain rates of Tonga. The NSRP model represent the large tropical convective cloud cluster rainfall events and the PWN model represents the random nature of the small convective rainfall event. Although the diurnal variation feature was not added to the model, the superposed model was able to fit the tropical rainfall data.

2.5 Spatial rainfall

The poisson process model and the cluster model have been used for single site models with the rainfall intensity at a single point in space. For a single site model, the statistic properties mean, variance, auto-correlation and extreme values must be observed at different level of aggregation. For multi site models, these statistics and also the cross-

correlation must be considered for each site within the region. A simple spatial temporal model where storm origins arrive in a Poisson process in space and time was developed by Cox & Isham (1988). Cowpertwait (1994) generated a simple multi site model using the generalized NSRP model for multiple cell types. Cowpertwait (1995) and Cowpertwait (2006) then derived spatial temporal model properties based on the Neyman Scott model. The spatial temporal model introduces a scale factor to allow for the effects of orography. Two types of rain cell with different parameters were used in the model, so that the random variables of any particular cell, radius and intensity, are correlated. But the rain cells assumed zero velocity which limited the applications of the model.

Cowpertwait *et al.* (2002) further extend the space time Neyman Scott model by using the third order and cross correlation function in the fitting procedure. A parameter which represents cell radius was added and only single type of cell was used in the model. The result shows that the spatial temporal model is able to represent regional extreme values of nine site from Arno Basin, Italy.

Northrop (1998) developed a spatial temporal model based on the Bartlett Lewis model. The cells are assumed to have an elliptical shape because the spatial autocorrelation plots of radar data used to fit the model have elliptical contours. Spatial temporal rainfall models are usually developed either for the radar data or raingauge data (Leonard *et al.* 2008).

The spatial NSRP model does not need radar data for calibration (Cowpertwait 2006). The model can also be used to infill missing data, disaggregate daily data to hourly data, simulate data for sites that do not have historical data and also to extend the rainfall records at any site by simulating additional hourly data (Cowpertwait 2006). Cowpertwait (2006) also used the spatial temporal NSRP model to disaggregate hourly rainfall data to five minutes data from six sites across Auckland City. The results showed that the disaggregation algorithm can be used with confidence.

The literature on spatial temporal models have been used on small regions with the assumption that the storms cover the whole region. But the storms may only partially cover the region. Leonard *et al.* (2008) presented a spatial temporal NSRP model with storm extent to analyze rainfall data over large areas (more than 100km x 100km), for which the space time model by Cowpertwait *et al.* (2002) may give unrealistic cross correlation. Only first and second order properties were used to fit the rainfall statistics, which resulted in more realistic cross correlation values between sites.

2.6 Concluding remarks

This chapter has considered empirical statistical and point process models. The empirical statistical models are a combination of modelling the occurrence of the rainfall and the rainfall amount. These types of rainfall models have been widely used because of the simplicity of the fitted model. But the major disadvantage of the empirical models is that the models do not represent the physical process of the rainfall. On the other hand, the point process models were developed based on an understanding of the rainfall structure from the storm event, rain cells, to the pulse of the intensity. This makes the point process models a more precise mathematical representation of the rainfall.

Both models, the empirical statistical models and point process models, are well developed in the literature. However over the past years, there has been a lack of research on tropical rainfall data, and especially with the diurnal variation problem in the stochastic rainfall models. Even though from the early times of the development of point process models, Rodriguez-Iturbe *et al.* (1987b) has highlighted this issue with the tropical rainfall, still 26 years later, it is yet not fully resolved. This has motivated our research on the stochastic rainfall models for tropical rainfall, especially using the Malaysia rainfall data which has not been fully explored, and to develop point process models that included a diurnal cycle.

Chapter 3

Methodology, data description & exploratory analysis

This chapter is divided into three broad areas. The first section deals with the point process model and will present the Poisson white noise (PWN) model and Neyman-Scott white noise (NSWN) model. It will also describes the new model developed for the diurnal variation rainfall data and briefly explains the superposed model for which results are given in Chapter 4-7. The final part of this section outlines the techniques for estimating the parameters and fitting the models. The section in data description, contains a summary of the Malaysian data. The exploratory analysis will give an understanding of the rainfall distribution of the Genting Klang, Malaysia data.

3.1 Point process model

The point process for modelling rainfall data describes the physical structure of the rainfall process. The sequence of the occurrence or position of rainfalls in time or space forms the point process. If the rainfall depth or intensity is added with the occurrence, then the process is called a marked point process. Rodriguez-Iturbe *et al.* (1984) illustrates two types of marked process for rainfall, one assuming the instantaneous rainfall occurrence and the other assuming a rectangular pulse associated with the occurrence. Two types of point process model for instantaneous rainfall are the Poisson white noise and the Neyman-Scott white noise. Two well known rectangular pulse rainfall models are the Neyman-Scott rectangular pulse model and the Bartlett-Lewis rectangular pulse model developed and analyzed in Rodriguez-Iturbe *et al.* (1987a) and Rodriguez-Iturbe *et al.* (1988).

3.1.1 Poisson white noise model

The PWN model is the simplest point process rainfall model which assumes the rainfall occurrence of the instantaneous pulses, $N(t)$, follows a Poisson process with rate λ per unit time. Each pulse that occurs has a random depth or magnitude, U_i, $i = 0,1,2,...$ that is statistically independent and follows an exponential distribution (parameter η). The rainfall intensity can be written as

$$\xi(t)dt = U(t)dN(t) \tag{3.1}$$

where $dN(t)$ is 1 if a burst occurs in an interval $(t, t+\delta t)$. U is independent of $N(t)$. The accumulated rainfall depth in a fixed time interval is

$$N_{U(t)} = \int_0^t \xi(s)ds = \sum_{i=0}^{N(t)} U_i \quad t > 0 \tag{3.2}$$

where $U_0 = 0$. The analytical expression for the accumulated rainfall sequence, Y_i, $i = 1, 2, ...$ over nonoverlapping time interval of length t (for example hours, days) is

$$Y_i = N_U[it] - N_U[(i-1)t] \;\; i = 1, 2, ... \tag{3.3}$$

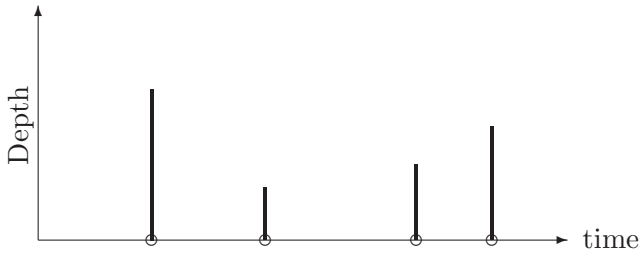
The Y_i are independent and identically distributed. Based on equation (3.3) , Rodriguez-Iturbe *et al.* (1984) give the second-order moments

$$E[Y] = E[U]\lambda t \tag{3.4}$$

$$Var[Y] = E[U^2]\lambda t \tag{3.5}$$

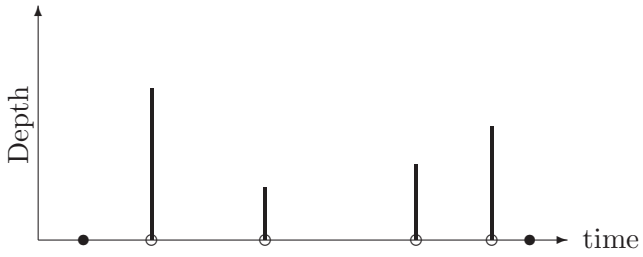
$$Cov[Y_i, Y_j] = 0 \quad i \neq j \tag{3.6}$$

The time t can represent a minute, an hour, a day, etc. It may be observed from (3.4) and (3.5) that there are 2 unknowns parameters : λ and $E[U]$. With the assumption that rainfall depths have an exponential distribution, $E[U^2] = 2E^2[U]$.



∘ rainfall occurence times

Figure 3.1: PWN plot



• storm origins

∘ rainfall occurence times

Figure 3.2: NSWN plot

3.1.2 Neyman-Scott white noise model

The NSWN model was first used by Kavvas & Delleur (1975) to fit the daily rainfall data in Indiana, USA. In the NSWN process, the storm origins occur in a Poisson process of rate λ. Each storm origin is accompanied by a number of instantaneous pulses, ν defined by a Poisson process. Figure 3.2 shows a schematic of the NSWN process. The pulse depths, U can be seen as the spike for each pulse. It is a random variable with an exponential distribution (parameter η). The waiting time of the pulses after the storm origin are assumed to be exponential distributed (parameter β) and independent and identically distributed. The counting process of the pulse occurrence, $N(t)$ is called the Neyman-Scott process or also known as a Poisson cluster process.

The rainfall intensity, $\xi(t)$ is defined as

$$\xi(t)dt = X(t)dN(t) \tag{3.7}$$

where $X(t)$ is the pulse depth at time t and $dN(t)$ is 1 if the rainfall pulses occur in an interval $(t, t+\delta t)$ and 0 otherwise.

The continuous time rainfall process is observed as an aggregated process in discrete time, Y_i^h defined as

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} X(t)dN(t) \tag{3.8}$$

where h is the time scale.

The second order properties of the aggregated process are invoked from (Rodriguez-Iturbe *et al.* 1984):

$$\begin{aligned} E[Y_i^{(h)}] &= \int_0^h \xi(s)ds \\ &= \int_0^h E[X(s)]E[dN(s)] \\ &= h\lambda E[U]E[\nu] \end{aligned} \tag{3.9}$$

To find the variance, the covariance of $\xi(t)$ is

$$Cov[\xi(t), \xi(t+\tau)] = \theta_1\delta(\tau) + \theta_2 e^{-\beta|\tau|} \qquad -\infty < \tau < \infty \tag{3.10}$$

where $\delta(\tau)$ is the Dirac delta function with

$$\int_{-\infty}^{+\infty} \delta(x) = 1 \tag{3.11}$$

$$\theta_1 = 2\lambda E[U]^2 E(\nu) \tag{3.12}$$

$$\theta_2 = \lambda\beta E[U]^2 E(\nu(\nu-1))/2 \tag{3.13}$$

Therefore the variance and covariance of Y are

$$\begin{aligned} Var[Y_i^{(h)}] &= \int_0^h dt \int_0^h [\theta_1\delta(t-s) + \theta_2 e^{-\beta|t-s|}]ds \\ &= \theta_1 h + \frac{2\theta_2}{\beta^2}[\beta h - 1 - e^{-\beta h}] \end{aligned} \tag{3.14}$$

$$\begin{aligned} Cov[Y_i^{(h)}, Y_k^{(h)}] &= \int_0^h dt \int_{(k-1)h}^{kh} [\theta_1\delta(t-s) + \theta_2 e^{-\beta|t-s|}]ds \\ &= \frac{\theta_2}{\beta^2}(1-e^{-\beta h})^2 e^{-\beta(k-1)h} \quad k \geq 1 \end{aligned} \tag{3.15}$$

θ_1 and θ_2 are the same as above. Four parameters, λ, ν, β and $E[U]$ are needed to specify the NSWN model.

3.1.3 Bernoulli trials with Markov dependence

Suppose a trial can yield just one of two possible outcomes. The results can be classified as either a 'success' or as a 'failure'. The probability of getting a success result is p which remains constant from trial to trial. If we let X equal to 1 if the outcome is a success and 0 if is a failure, then X is a Bernoulli random variable.

Klotz (1973) developed a model for Bernoulli trials with Markov dependence between successive observations. An outcome of an experiment that depends only on the outcome of the previous experiment and not on the sequence of events that preceded it, is a Markov process. When combining the Bernoulli trial with the Markov process, apart from the success probability, p, we have a dependence parameter, γ. The probabilities are defined as follows,

$$P[X_i = 1] = \alpha \qquad i = 1, 2, ..., n, \tag{3.16}$$

$$P[X_i = 1 | X_{i-1} = 1] = \gamma \qquad i = 2, 3, ..., n. \tag{3.17}$$

where n is the sample size. From (3.16) and (3.17), it follows that

$$P[X_i = 0 | X_{i-1} = 1] = 1 - \gamma \tag{3.18}$$

$$P[X_i = 1 | X_{i-1} = 0] = \frac{(1-\gamma)\alpha}{1-\alpha} \tag{3.19}$$

$$P[X_i = 0 | X_{i-1} = 0] = \frac{1 - 2\alpha + \alpha\gamma}{1-\alpha} \tag{3.20}$$

The transition probabilities are between zero and one with the condition of $0 \leqq \alpha \leqq 1$ and $\max(0, (2\alpha - 1)/\alpha) \leqq \gamma \leqq 1$. The model simplifies to independent Bernoulli trials if $\gamma = \alpha$. A clustering exists, among the ones and among the zeros, if $\gamma > \alpha$ or a lack of clustering if $\gamma < \alpha$.

The transition probabilities matrix is given by

$$P = \begin{pmatrix} (1 - 2\alpha + \gamma\alpha)/(1-\alpha) & (1-\gamma)\alpha/(1-\alpha) \\ 1 - \gamma & \gamma \end{pmatrix} \tag{3.21}$$

P is diagonalized using its right and left eigen vector and its eigen values, as

$$\begin{pmatrix} 1 & \alpha \\ 1 & -q \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (\gamma - \alpha)/q \end{pmatrix} \begin{pmatrix} q & \alpha \\ 1 & -1 \end{pmatrix} \tag{3.22}$$

with $q = 1 - \alpha$. The kth step transition probability matrix is

$$\begin{pmatrix} 1 & \alpha \\ 1 & -q \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & [(\gamma - \alpha)/q]^k \end{pmatrix} \begin{pmatrix} q & \alpha \\ 1 & -1 \end{pmatrix} \tag{3.23}$$

3.1.4 The X latent point process model

In the point process model the daily rainfall occurrence starts with a thunderstorm or cyclone as the main process. The actual rainfall pouring down is the output of the main process. The total rainfall intensity is determined by the storm arrival, the number of rainfalls and their time positions within that storm. So the point process model incorporates meteorological knowledge into the model to make it more physically meaningful, with successful results.

For our modification, we start by making an observation whether there is a rainfall (or not) in the ith period of length h, with probability α_i. Let X_i^h be the indicator for observing the rainfall in ith interval of width h as

$$X_i^h = \begin{cases} 1 & \text{with probability } \alpha_i \\ 0 & \text{with probability 1-}\alpha_i \end{cases}$$

For every hour, X_i^h is 1 if rainfall process is being observed and 0 otherwise. The probability of observing a rainfall in each hour may be different. In the tropical rainfall, the chance of rain in the late afternoon is higher than in the morning. Therefore, we are taking into account the diurnal variation in the new model.

With every observation of the rainfall, we estimate the total intensity by using the point process model. Here, Y_i^h is assumed to be the latent total rainfall during the ith interval of width h using the PWN and NSWN model. The total observed rainfall, Z_i^h over an interval of length h, is $X_i^h Y_i^h$.

If $h = 1$, then the total rainfall in the ith hour,

$$z_i = X_i Y_i \tag{3.24}$$

z_i can be 0 if either $X_i = 0$ or $Y_i = 0$ or both. For a daily rainfall,

$$\begin{aligned} \sum_{i=1}^{24} z_i &= \sum_{i=1}^{24} X_i Y_i \\ &= Z^{24} \end{aligned} \tag{3.25}$$

The expectation for X is $E(X_i) = \alpha_i$ and $E(X_i^2) = \alpha_i$. We assumed X and Y are independent. The moments for the daily rainfall, Z^{24} are

$$\begin{aligned} E[Z^{24}] &= \sum_{i=1}^{24} E(X_i)E(Y_i) \\ &= E(Y)[\alpha_1 + \alpha_2 + ... + \alpha_{24}] \end{aligned} \tag{3.26}$$

With the variance

$$\begin{aligned} Var[Z^{24}] &= \sum_{i=1}^{24}\sum_{j=1}^{24} Cov[(X_iY_i, X_jY_j)] \\ &= \sum\sum_{i\neq j} E(X_i)E(X_j)E(Y_iY_j) - E(X_i)E(X_j)E(Y_i)E(Y_j) + \quad (3.27) \\ &\quad \sum\sum_{i=j} E(X_i^2)E(Y_i^2) - (E(X_i)^2)(E(Y_i)^2) \quad (3.28) \\ &= \sum\sum_{i\neq j} \alpha_i\alpha_j E(Y_iY_j) - \alpha_i\alpha_j E(Y_i)E(Y_j) + \quad (3.29) \\ &\quad \sum\sum_{i=j} \alpha_i E(Y_i^2) - (\alpha_i^2)(E(Y_i)^2) \quad (3.30) \\ &= \sum\sum_{i\neq j} \alpha_i\alpha_j Cov(Y_iY_j) + \sum_{i=j} \alpha_i Var(Y_i) \\ &\quad + \sum_{i=j} (\alpha_i - \alpha_i^2)(E^2(Y_i)) \quad (3.31) \end{aligned}$$

To find the covariance assuming $\tau \neq 0$

$$\begin{aligned} Cov(Z_i^{24}, Z_{i+\tau}^{24}) &= E(Z_i^{24}Z_{i+\tau}^{24}) - E(Z_i^{24})E(Z_{i+\tau}^{24}) \quad (3.32) \\ &= \sum_{i=1}^{24} E(X_i)E(X_{i+\tau})Cov(Y_i, Y_{i+\tau}) \quad (3.33) \\ &= \sum_{i=1}^{24} \alpha_i\alpha_{i+\tau}Cov(Y_i, Y_{i+\tau}) \quad (3.34) \end{aligned}$$

The rainfall is modelled assuming X_i and $X_{i+\tau}$ are independent and also assuming dependence. The models now composed of the indicator variable with the latent variable will be referred to as X-PWN model, X-NSWN model, Markov X-PWN model and Markov X-NSWN model.

3.1.5 Moments for two hour blocks

In this study, to minimize the number of parameters for every month, we fit the new models to hourly data in two hour blocks. Parameters are estimated for the 12 blocks separately for every month. Properties up to second order are derived in this section.

Let Y be the NSWN process and X_i $(i = 1, 2, ..., 12)$ the indicator variable for every two hours block.
The expected value of the hourly rainfall in a two hour block is

$$E[X_iY] = \alpha_i E(Y) \quad (3.35)$$

Table 3.1: Parameters Description

Random Variable	Parameter	Sampling Distribution	Point Process
Storm origins	λ	Poisson	PWN/NSWN
Number of pulses	ν	Poisson	NSWN
Pulses arrival	β	Exponential	NSWN
Pulses depth	η	Exponential	PWN/NSWN
Probability a rainfall is observed	α	(-)	PWN*/NSWN*
Transition probability	γ	(-)	PWN**/NSWM**

* : X is assumed independent
** : X is assumed dependent

Assuming X_i and Y are independent, the variance is

$$\begin{aligned} Var[X_iY] &= E[X_i^2]E[Y^2] - E^2[X_i]E^2[Y] && (3.36)\\ &= \alpha_i(Var[Y] + E^2[Y]) - \alpha_i^2 E^2[Y] && (3.37)\end{aligned}$$

If Y^1 and Y^2 are the NSWN process in the first and second hour of the block and X^1 and X^2 are the indicator variables for the first and second hour in a block, which are assumed independent, then the covariance between hours within the block is

$$\begin{aligned} Cov(X^1Y^1, X^2Y^2) &= E(X^1)E(X^2)[E(Y^1Y^2) - E(Y^1)E(Y^2)] && (3.38)\\ &= \alpha_i^2 Cov(Y^1, Y^2) \quad i = 1, 2, ..., 12 && (3.39)\end{aligned}$$

The expectation for X is $E(X_i) = \alpha_i$ with $i = 1, 2, ..., 12$. The covariance, $Cov(Y^1, Y^2)$ is calculcated from 3.15.

On the other hand, suppose X_i and $X_{i+\tau}$ are dependent. We will assumed they are Bernoulli trials with Markov dependence. A dependence parameter $\gamma = P[X_i = 1|X_{i-1}] = 1$ is added to the Markov chains. From (Klotz 1973), the

$$E(X_iX_{i+\tau}) = \left(\alpha^2 + \alpha(1-\alpha)\left[\frac{\gamma-\alpha}{1-\alpha}\right]^\tau\right) \quad (3.40)$$

Therefore, the covariance is

$$\begin{aligned} Cov(X^1Y^1, X^2Y^2) &= E(X^1X^2)E(Y^1Y^2) - E(X^1)E(Y^1)E(X^2)E(Y^2)\\ &= \left(\alpha_i^2 + \alpha_i(1-\alpha_i)\left[\frac{\gamma-\alpha_1}{1-\alpha_1}\right]\right)[Cov(Y^1, Y^2) + E^2(Y_i)]\\ &\quad -\alpha_i^2 E^2(Y_i) && (3.41)\end{aligned}$$

The parameters involved in the PWN model, X-PWN model, Markov X-PWN model, X-NSWN model and Markov X-NSWN model are shown in Table 3.1.

3.1.6 Superposed model

In a superposition operation, a number of separate point process are superimposed or merged onto each other. If there are n independent processes superposed, the sum $N(A) = N(A_1) + N(A_2) + ... + N(A_n)$ (Cox & Isham 1980). The use of the superposition of the point process in the rainfall model is to provide additional parameterisation for different types of storm, and to provide flexibility (Cowpertwait 2004). Cowpertwait (2004) superposed two single processes to represents the isolated convective cells and the clusters of convective cells. Two types of rainfall in meteorology, the stratiform and convective, can also be modelled by the superposed model (Cowpertwait *et al.* 2007, Morrissey 2009). The stratiform rainfall is usually from mid-latitude frontal systems with small raindrops or drizzle. The rain drops are larger in convective rain giving heavier rainfall.

The superposed model moments are the sum of the analytical moments from each of the independent point process model (the expression below is for superposition of NSWN (X_i) and PWN (X_j)),

$$E[X_i + X_j] = E[X_i] + E[X_j] \tag{3.42}$$

$$Var[X_i + X_j] = Var[X_i] + Var[X_j] \tag{3.43}$$

$$Cov[X_i + X_j, X_{i+n} + X_{j+n}] = Cov[X_i, X_{i+n}] \tag{3.44}$$

since the covariance for PWN process is zero.

3.1.7 Model parametrisation and fitting

The majority of the statistical literature (for example Cowpertwait *et al.* (2011), Onof & Wheater (1994a)) estimates the parameters for every month separately to take account of seasonal effects. By pooling the observations from the same month, the non-stationarity in the data is assumed weak. In this study, the parameters are estimated for every two hour block across all days, for each month. The two hour blocks are actually hourly data grouped by every two hours. The main reason for using two hour blocks is to summarized the diurnal variation that is apparent in the hourly data. Obeysekera *et al.* (1987) suggested modelling the diurnal periodic data into two time zones with different parameters for each zone. However, in our case diurnal variation is not the same across all months. In one month morning rainfall is more common. So the present model has been developed to give more flexibility for representing the monthly diurnal patterns.

The parameters are estimated using the method of moments, which is the commonly used for the point process models. This is due to the difficulty of finding the likelihood function. The statistical properties of the observed data are equated to their theoretical expression. There are different combination of moments used, without any rules to follow. Fourier models may also be used to estimate the parameters. The Fourier

models will reduce the number of parameters while still describing the seasonal variation (Cowpertwait 1994, Yusof *et al.* 2008).

If more than one unknown parameter is to be estimated simultaneously, the nonlinear equations of method of moments are solved by minimization of an objective function, Z.

$$Z = \sum_{i=1}^{m} \left(1 - \frac{z_i}{\hat{z}_i}\right)^2 \tag{3.45}$$

where z_i is the theoretical moment, $\hat{z}_i$ is the equivalent observed moment and m is the number of moments properties used for Z. The hourly mean, variance and lag 1 autocorrelation in the two hour block are used in the fitting procedure. The lag 1 autocorrelation is $\dfrac{\mathrm{Cov}[Y_j^h, Y_{j+1}^h]}{\mathrm{Var}[Y_j^h]}$.

The objective function is minimized using the simplex method developed by Nelder & Mead (1965). The optimization is constrained with boundary values for each of the unknown parameters. For example, the α's and γ's were set to be between 0 and 1 since both are probabilities. The boundaries set the limit of unrealistic estimates.

During the process of estimating the parameters, we did not have any sufficient problem with the convergence. However, we did have problems with the estimation hitting the boundary limits. This was fixed by changing the boundary limits. Since this was not a major issue in the parametrisation process, the Nelder-Mead optimization technique was considered sufficent for the research work.

Usually a point process rainfall model estimates the parameters for all the months simultaneously (Cowpertwait *et al.* 2011, Onof & Wheater 1994a). This approach was applied here but the models developed involved a lot of parameters. The X-PWN model has 14 parameters, Markov X-PWN has 26 parameters, X-NSWN model has 16 parameters and Markov X-NSWN model has 28 parameters. The estimated parameters tend to hit the boundaries of realism. Fixing the boundary values for all of the months became a tedious process. To ease the estimation procedure, the parameters were estimated separately for each month. The function optim() in R program was used in the estimation process.

3.1.8 Simulation

To simulate rainfall data, a C program was designed. The X-PWN and X-NSWN models have the same simulation process. The monthly parameters for the PWN or NSWN models were combined to simulate the hourly rainfall data. There were 100 simulation samples with each sample being a 36 years period. Then the diurnal pattern of the two hour blocks become apparent when applying the probability of rainfall for every two hour block. For

the Markov X-PWN and Markov X-NSWN models, the simulation start with the Markov chain. A valid wet hour would continue with the PWN or NSWN process being used for the rainfall depth. The final step is to generate the diurnal pattern in the two hours data. The superposed model simulation process is the same as the other models when generating the rainfall magnitude. But the two models involved are simulated separately before they are combined together to proceed with the final step. All the rainfall data is generated by each month.

3.1.9 Model assessment

The primary method to assess the performance of the models in this study is by graphical analysis and mean square error (MSE). Graphical methods compare the historical statistics with the simuation values and also the examination of the extreme values plot. Since the extreme values were not used in the fitting procedure, a good fitting of the historical values by the simulation values will show a good performance of the model. The comparison is made by plotting the annual maximum rainfalls against the reduced Gumbel variate. The annual maximum total rainfall is extracted at different time scales (e.g. hourly or daily). The reduced Gumbel variate is given by

$$-log(-log[p_i]), i = 1, 2, ..., n$$

with n the number of samples and i the variate ranks.

$$p_i = \frac{i - 0.3}{n + 0.4}$$

is taken from Bury (1999). The number of samples, n is 36, corresponding to the 36-year sampling period. If plotting $-log(-log[p_i])$ versus the ordered maximum values show a linear pattern, then the observed extreme values can be modelled by a Gumbel distribution. We can compare the 36 observed historical values with the simulation values which are obtained from the 100 simulation samples.

A horizontal scale corresponding to return period is also added to the Gumbel plot. The return period scale for rainfall events gives the average number of years to the occurrence of an exceedance event (Bury 1999). For example, if a flood occurs on average ten times in 100 years, the return period is 10 years. The annual maximum rainfall is commonly used to find the return period since the annual maximum rainfall events in consecutive years are assumed independent.

3.2 Data description

Malaysia lies in the equatorial zone between 1^0 and 6^0N with a humid tropical climate throughout the year (in Figure 3.3). The temperature is between 25.5^0 to 32^0 celcius with high humidity(81%) and heavy rainfall. The number of wet days annually is between 150-200 days with an average amount of rainfall ranging from 2000mm to 4000mm (Suhaila & Jemain 2007). The distribution of rainfall is influenced by two types of monsoon, the Southwest monsoon from May to August and Northeast monsoon from November to February. The Southwest monsoon brings heavy rainfall to the west coast region whereas the northeastern region receives heavy rainfall during the Northeast monsoon.

Kuala Lumpur is the capital city of Malaysia. The geography of Kuala Lumpur is characterized by a valley known as Klang Valley that is bordered by the Titiwangsa Mountains in the east. The northeast monsoon covers a wide area of Kuala Lumpur with convective rains and thunderstorms, followed with lightning and strong wind (Desa & Niemczynowicz 1996).The southwest monsoon has a similiar rainfall pattern to the northeast monsoon but the rainfall depth can be higher. The inter-monsoon periods between the two season have the wettest months, April to May and September to October. In this period, the rain is mainly a convective type. The storms which build up from the intense local heating in warm, moist and unstable atmosphere only last between half an hour to an hour with heavy downpours (Yusof 2007).

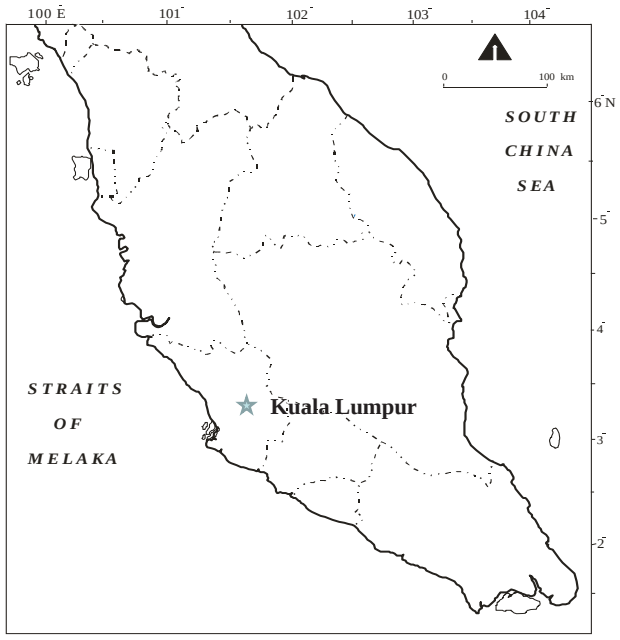


Figure 3.3: Peninsular Malaysia

Figure 3.4: Kuala Lumpur rainfall stations

3.3 Exploratory analysis

There are more than 30 rainfall stations in Kuala Lumpur (in Figure 3.4). The selection of rainfall station for this study is subject to the availability of a complete set and the longest series of 5 minutes, hourly and daily data. The data used for this study is from the Genting Klang rainfall station. The 5 minutes, hourly and daily data for this station was available from 1973 to 2008. Less than 2.5% of the data is missing.

Figure 3.5 is an annual rainfall totals time plot of the Genting Klang data. The line represent annual rainfall totals. The horizontal dashed line is the overall average annual rainfall, 1990.5mm, for the period of 1973-2008. A wetter period occurred in the late 90's to the new millenium. This is also true in 1973, around 23 years apart. The drier period occurred for almost 22 years, starting from 1974 to 1995, with an exception for 1978 and 1985.

Figure 3.6 shows a time plot of the monthly mean rainfall. The line represent the monthly rainfall averages over the 36 years sampling period. The mean monthly rainfall totals is represented by the dashed horizontal line. The plot exhibits a strong seasonal pattern. The rainfall averages are higher in the inter monsoon season that occur in April

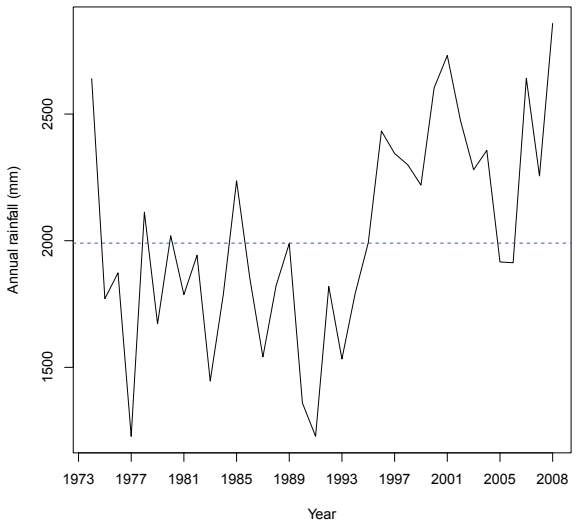


Figure 3.5: Annual rainfall totals: the Genting Klang data (1973-2008)

to May and also in October to November, the beginning of the northeast monsoon season. This pattern is also exhibited in the daily rainfall average plot in Figure 3.7. The wettest day is in April, followed by days in October. More than 13mm of rainfall were recorded. The driest month is January with an average less than 100mm of rainfall.

The hourly plot in Figure 3.8 show a strong diurnal pattern. The hours start from 1 am to 12 midnight. The maximum rainfall occurs in the late afternoon from 4 to 7 pm. During other hours, the average hourly rainfall distribution is the same with an average rainfall less than 0.15mm. By examining the diurnal plot by months, in Figure 3.9, all of the months exhibit the same diurnal pattern regardless of the season. Even the driest and wettest month, the diurnal variation is a permanent features in the Genting Klang hourly rainfall data. In February and September, there are two peaks, in the morning and in the late afternoon. But the highest average rainfall is still in the late afternoon. The autocorrelation values also gives strong evidence of a diurnal variation in the rainfall data. The 24 hour autocorrelation values are mostly bigger than the 12 hour values. This is shown in Table 3.2.

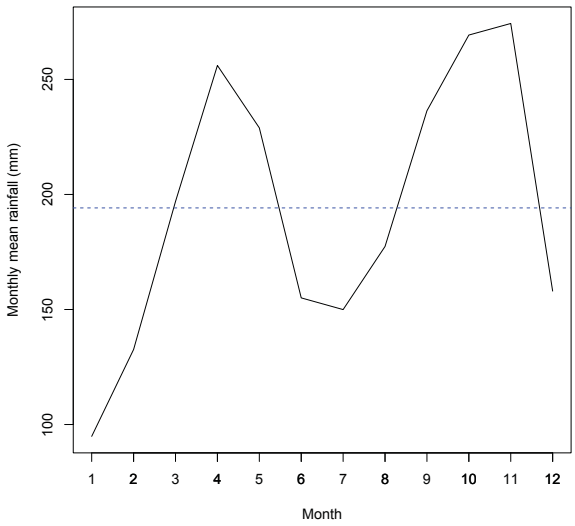


Figure 3.6: The monthly mean rainfall. The Southwest monsoon from May to August and Northeast monsoon from November to February.

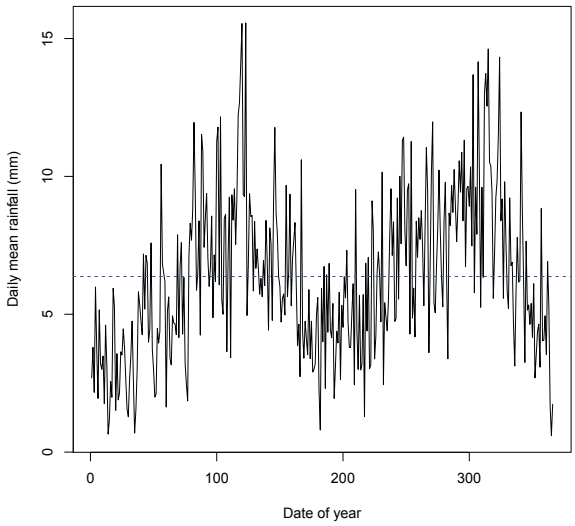


Figure 3.7: Daily mean rainfall

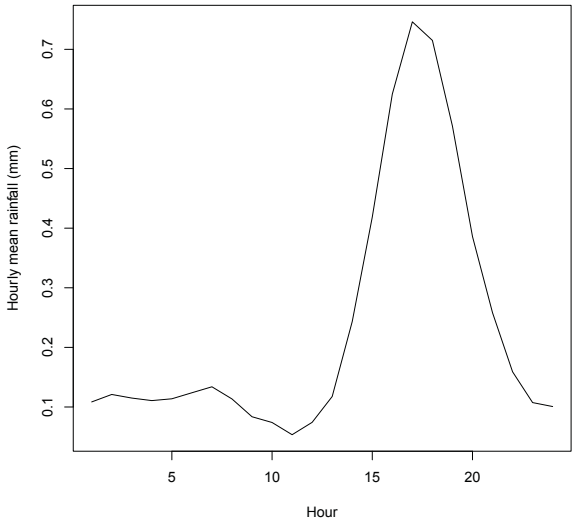


Figure 3.8: Hourly mean rainfall

Table 3.2: Sample statistic of lag 1 autocorrelation for pooled subsamples by month.

Month	ac5m	ac1h	ac2h	ac6h	ac12h	ac24h
Jan	0.7572	0.2559	0.1492	0.1069	0.0927	0.1459
Feb	0.6648	0.3309	0.1854	0.1022	0.0165	0.1384
Mar	0.7186	0.2852	0.1659	0.0352	-0.0436	0.0452
Apr	0.7739	0.3068	0.1636	0.0011	-0.0664	-0.0211
May	0.7104	0.3059	0.1546	0.0504	-0.0450	0.0709
Jun	0.7614	0.2921	0.1542	0.0601	0.0012	0.1371
Jul	0.7285	0.3491	0.2122	0.0316	-0.0206	0.0033
Aug	0.7324	0.3080	0.1525	0.0445	-0.0223	0.1116
Sep	0.7256	0.3485	0.1941	0.0137	-0.0314	0.0330
Oct	0.7319	0.3528	0.1971	0.0330	-0.0432	0.1430
Nov	0.7332	0.3386	0.2064	0.0963	-0.0404	0.1100
Dec	0.7225	0.3066	0.2526	0.1283	0.0064	0.1571

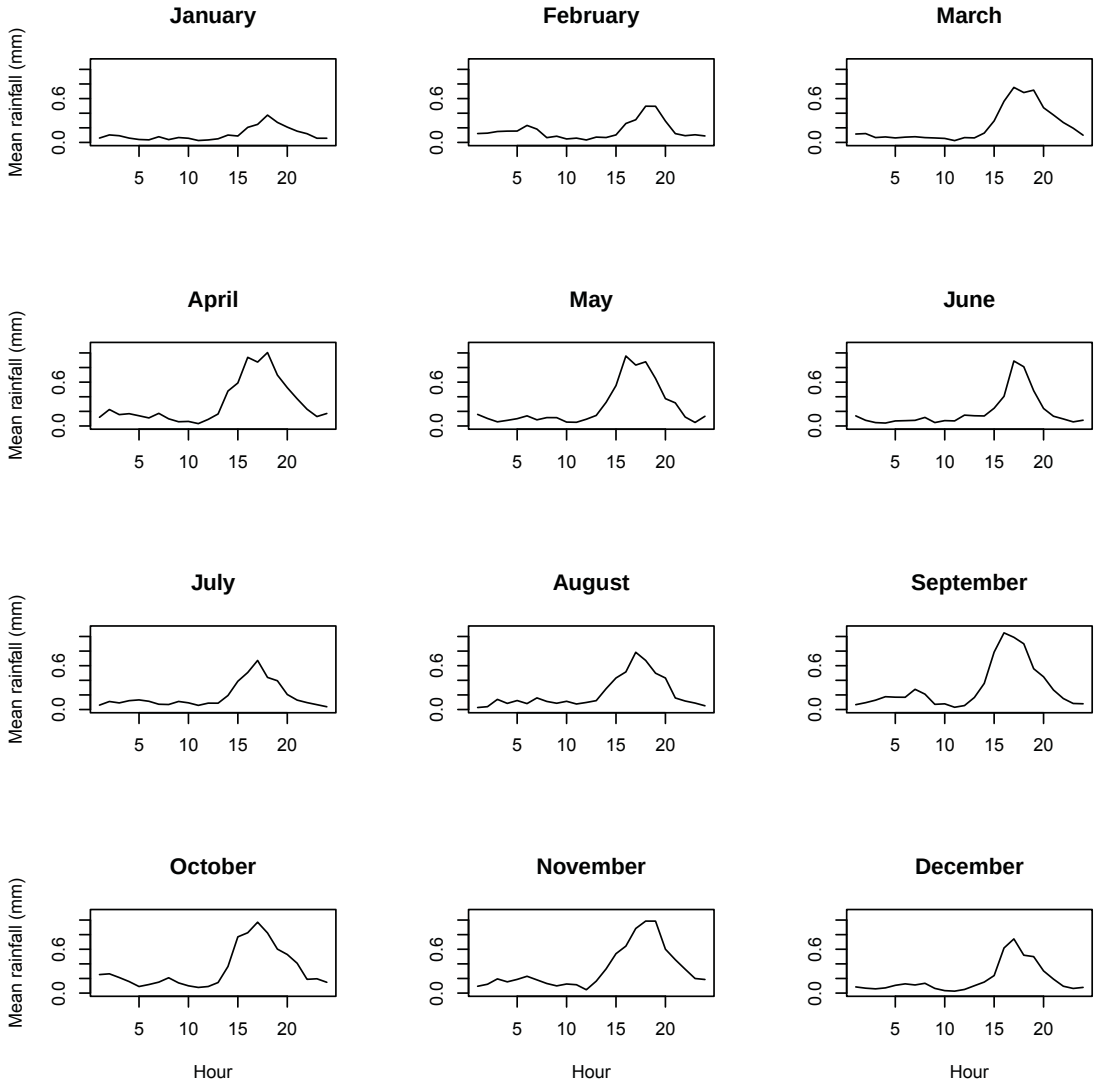


Figure 3.9: Diurnal rainfall pattern for every month

3.4 Summary

The point process rainfall models are well developed for the midlatitude climate type (Cowpertwait *et al.* 2011, Onof & Wheater 1994a). Although the same basic models have been used for the tropical rainfall (Yusof *et al.* 2007, Yusof *et al.* 2008, Hanaish *et al.* 2011a, Hanaish *et al.* 2011b) with good results, nevertheless as pointed out by Morrissey (1993), it does not account for the diurnal pattern. In Malaysia, there is a strong diurnal pattern in the hourly data. To describe the diurnal variation in the point process rainfall model, we modify the existing model by adding the rainfall indicator and the Markov chain. Two basic point process models, the PWN and NSWN models were selected to attempt the modification. Both models were chosen because of the simplicity of the models in term of the instantaneous pulses instead of the rectangular pulses. Two assumptions were also made, namely dependence and independence of the hours within the two hour block. Therefore, four new models are analyzed in the next part of the study, the X-PWN, Markov X-PWN, X-NSWN and Markov X-NSWN. The final part is to superpose the PWN and NSWN models to improve the results. The parameter estimation process and the simulation process will be discussed comprehensively in the following chapters.

Chapter 4

Non-homogenous PWN

4.1 Introduction

In Chapter 3, the theoretical properties of the Poisson white noise (PWN) model has been introduced. Rodriguez-Iturbe *et al.* (1984) used the PWN model to analyze the rainy season in Denver, USA. Two issues arise from the study when comparing the hourly and daily time scales results. First, the parameter estimates are significantly different and second, the correlation function is not zero for both time scales. The poor performance of this model is probably due to the model not accomodating the correlation function. Nevertheless the PWN model will be tried out in the present study because of the simplicity of the model. The process starts with the rainfall event that determines the average number of rainfall occurrences per hour, and the instantaneous burst is associated with each rainfall occurrence. This reflects the convective rainfall in the tropics which is a 'burst' of heavy rainfall in the afternoon.

4.2 PWN model specification & properties

In the rainfall event, the rainfall occurrence is assumed to be a Poisson process with rate λ. The rainfall amount is assumed to occur as an instantaneous burst. The process of pulses is independent. On average, the pulse depth is $E[U]$ per unit time interval. The total amount of rainfall in an interval is the integral of the intensities during that period. Based on equation (3.3), the formulas for mean, variance and autocovariance are cited from (Rodriguez-Iturbe *et al.* 1984) and given below.

$$E[Y] = E[U]\lambda t \tag{4.1}$$

$$Var[Y] = E[U^2]\lambda t \tag{4.2}$$

$$Cov[Y_i, Y_j] = 0 \quad i \neq j \tag{4.3}$$

Only two unknown parameters need to be estimated, λ and $E[U]$. For the pulse depth, it is assumed as an independent exponential (η) random variable. Therefore $E[U^2] = 2E^2[U] = 2\eta^2$.

4.3 Fitting procedure

Rodriguez-Iturbe *et al.* (1984) evaluate the performance of the Poisson model by fitting the model to the month chosen, for example April and September. The process is assumed the same throughout the whole month. Therefore the parameters are homogeneous for every hour of the day. In this study, the parameters are estimated for every two hour block in each month. Instead of having only one parameter for the λ and η for every month, there are 12 values of λ and η within every month of the year. The parameter varies for the two hour blocks. Thus, it is a non-homogeneous PWN model. The PWN model is fit to the two hour block data instead of the hourly data to reduce the number of parameters. However, the diurnal pattern in the two hour block data are the same as in the hourly data as can be seen by comparing graph (3.6) with (4.1) - (4.3). It is assumed that blocks are independent.

The parameters are estimated for every month. The hourly data in that particular month is divided into two hour blocks. There will be 12 blocks, starting from midnight (0 hour). The PWN model is fitted to each block within each month, with data pooled over 36 years. The first model assumed different λ and η for every two hour block. This assumption is to allow the block with the maximum rainfall to have its own rate and depth. However, the simulation results for the month chosen to fit the model showed that it did not perform well. The simulated means and variances did not match well with the historical values. Even though there are 24 parameters (λ_i and η_i with $i = 1, 2, ..., 12$ for each block), the PWN process is not suitable for a short duration time interval. This model was not further investigated and will not be discussed further here.

The next model has a different λ for every two hour block, but η is fixed constant throughout the 12 blocks. The average pulse depth is the same for the whole month. By just allowing the block to have different rainfall occurrence rate with the same depth throughout the day, we assumed the diurnal pattern is from the λ. The model also has less number of parameters per month to define the PWN process a total of 13 parameters per month compared with 24 parameters with the failed model before. A trial examination of this model shows that by allowing the PWN process to have stationary depth throughout the month, the simulated properties fit well with the observed values. The nonhomogenous PWN with fixed η is chosen to be further analyzed.

The parameters are estimated using the method of moments. The hourly mean, variance and lag-1 autocorrelation are chosen to fit the PWN model to data. Expressions for the first two moments using the two hour blocks are

$$E[Y_i] = E[U_i]\lambda_i t \quad (4.4)$$
$$Var[Y_i] = E[U_i^2]\lambda_i t \quad (4.5)$$

with $i = 1, 2, ..., 12$. The autocovariance is equal to zero, thus not included in the fitting procedure. The fitting procedure is based on minimising the objective function in 3.45.

4.4 Analysis

4.4.1 Parameter estimates

For every month, the data is divided by 12 blocks. Each block contains 2 hour data. There are 12 λ for every month and one η. The monthly estimates of η are shown in Table 4.1. The block by block estimates of λ are in Table 4.2.

The parameters of PWN are conceptually based on the rainfall process, with λ as the mean rate of rainfall bursts and η the magnitude of the burst. The parameters are estimated for each month individually, so the rainfall depths do not represent the hierarchy of the month intensity. However, the parameter λ shows the highest rate of rainfall in block 8 to 10. These time blocks have the maximum amount of rainfall of the day and are where the diurnal variation is most clearly seen. Block 9 for hour 18-19 (4-5pm), is the peak hour of heaviest rainfall, thus the rainfall rate is the highest value for every month. It is followed by block 10 and block 8. But for the other blocks, the λ values do not reflect the diurnal pattern of the hourly data.

4.4.2 Moments

The mean and variance for each month are plotted separately in Figure 4.1, Figure 4.2 and Figure 4.3. Historical values are represented by points and simulation results represented by box plots. Based on the estimated model parameters, 100 simulation samples are generated from the PWN model. The model performance is evaluated by fitting the moments of rainfall and compared with the simulation results. Since only the mean and variance are available from the model, only two plots are compared.

λ and η are estimated from the optimization method. Therefore the mean and variance values from the simulation do not have an exact fit with the historical values. This can be seen for the months with high diurnal pattern in the hours from 4-7pm. From March until December, excluding April, the simulation mean is over estimated compared to the historical mean. Therefore the variance is under estimated for those months. Somehow,

Table 4.1: Parameter estimates for the pulse depth. The units are mm.

Month	η
1	0.1731
2	0.1388
3	0.1673
4	0.1106
5	0.1699
6	0.1585
7	0.1879
8	0.1686
9	0.1708
10	0.1475
11	0.1808
12	0.2383

the model is producing more rainfalls during the peak hours. During hours with low mean rainfall, from eight at night until one in the afternoon, the mean values have been matched well enough.

The underestimation and overestimation of the mean and variance suggests that the independence assumption is at fault. A model allowing for positive correlation between rainfall events within blocks would be expected to give a larger variance for the block total rainfall. So we consider the next model, which allows for positive correlation in the rainfall process.

Table 4.2: Parameter estimates for the rate of rainfall occurrences. The units are hour^{-1}

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
λ_1	0.0133	0.0174	0.0186	0.0206	0.028	0.0157	0.0171	0.0038	0.0111	0.0422	0.0157	0.0113
λ_2	0.0086	0.0248	0.0094	0.0134	0.0065	0.0032	0.0212	0.0174	0.0237	0.0254	0.0315	0.0111
λ_3	0.0037	0.0312	0.0081	0.0118	0.0157	0.0105	0.0233	0.0166	0.0267	0.0122	0.0318	0.029
λ_4	0.0116	0.0193	0.0111	0.0165	0.0143	0.0152	0.0106	0.0242	0.0472	0.0275	0.0267	0.0352
λ_5	0.0127	0.0109	0.0076	0.0064	0.0107	0.0085	0.0195	0.0154	0.0115	0.0173	0.0178	0.0093
λ_6	0.0032	0.0043	0.0071	0.0058	0.0132	0.0185	0.0129	0.0117	0.0073	0.0102	0.0139	0.0055
λ_7	0.0137	0.0083	0.0127	0.0407	0.0461	0.0229	0.026	0.0378	0.0481	0.035	0.0446	0.0342
λ_8	0.0302	0.0208	0.0834	0.0869	0.1512	0.0543	0.1008	0.0932	0.1876	0.1365	0.1073	0.1245
λ_9	0.0636	0.0543	0.136	0.0947	0.1629	0.1566	0.1274	0.1422	0.1892	0.1422	0.189	0.1798
λ_{10}	0.049	0.0523	0.1118	0.0668	0.0985	0.0592	0.0665	0.0895	0.0954	0.0909	0.1683	0.1151
λ_{11}	0.0224	0.0086	0.0606	0.0343	0.0427	0.0171	0.0213	0.0205	0.0313	0.0404	0.0737	0.0321
λ_{12}	0.0055	0.0097	0.0211	0.0141	0.0176	0.0062	0.0069	0.0099	0.0091	0.0251	0.0358	0.0132

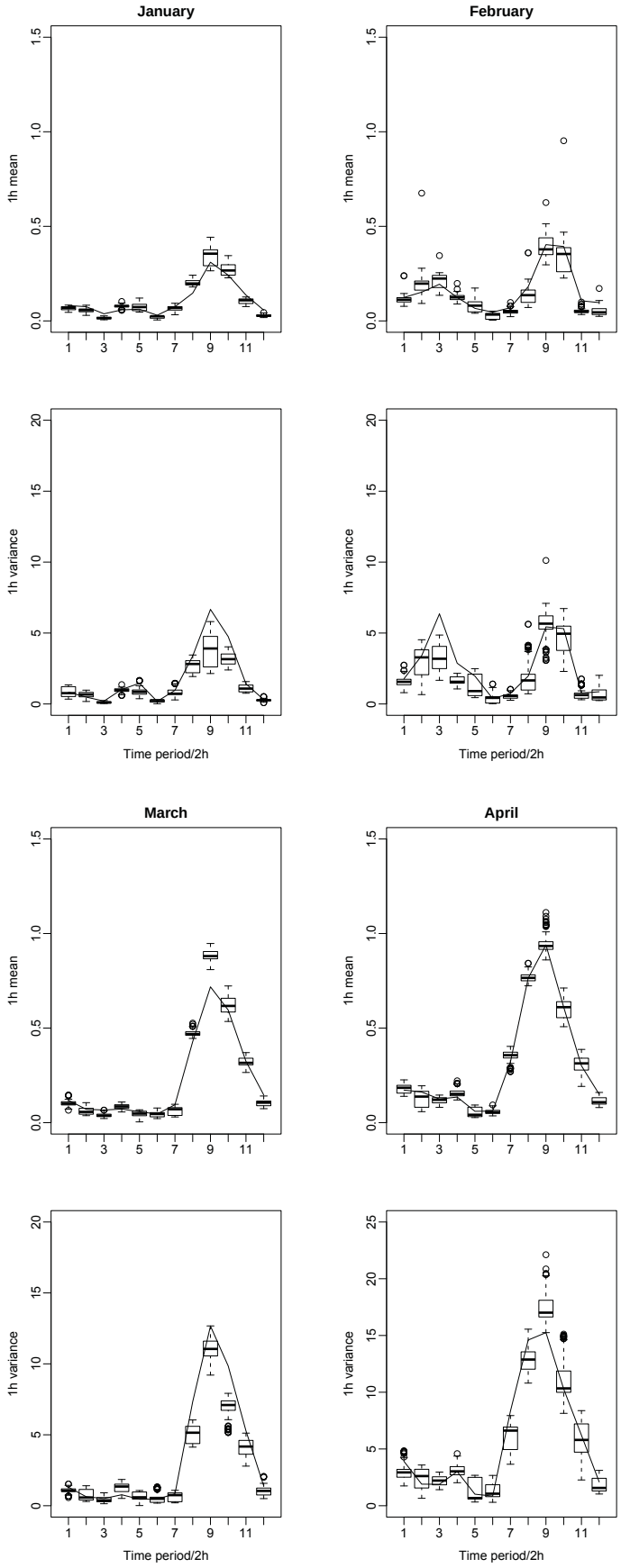


Figure 4.1: The mean and variance of January, February, March and April

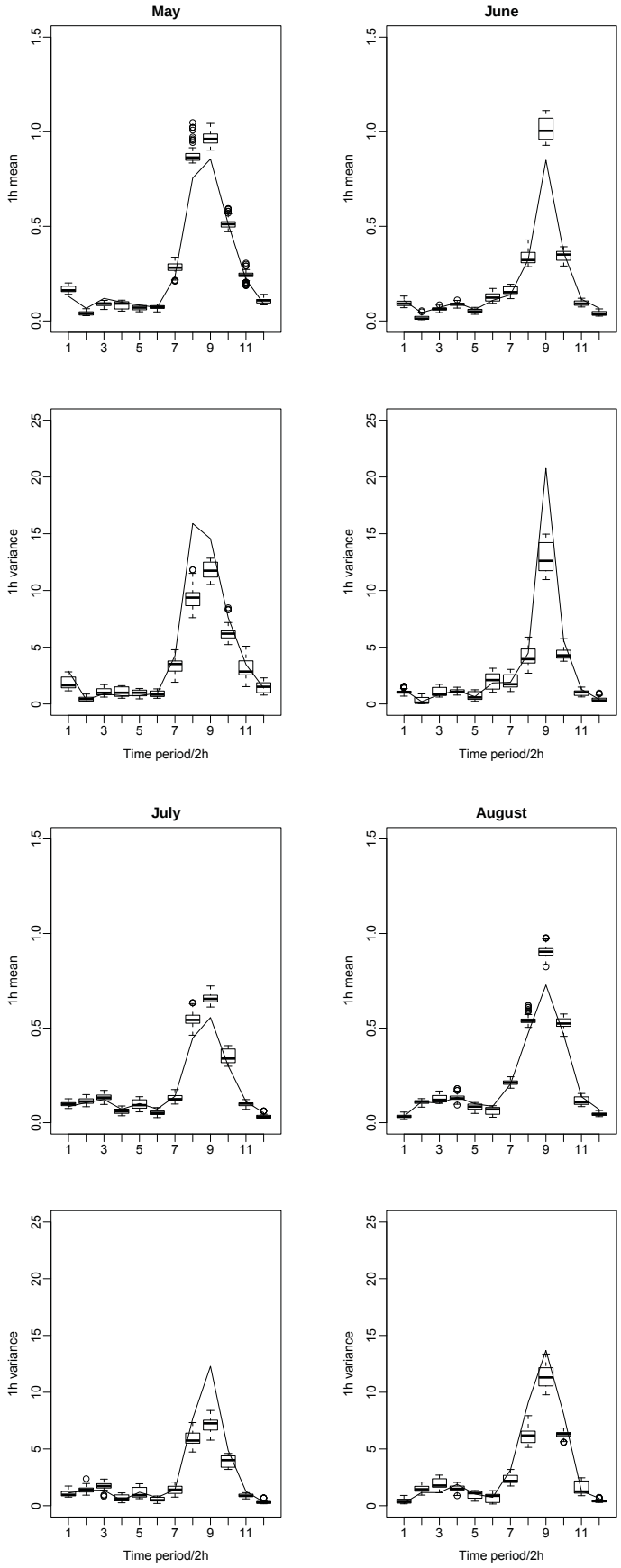


Figure 4.2: The mean and variance of May, June, July and August

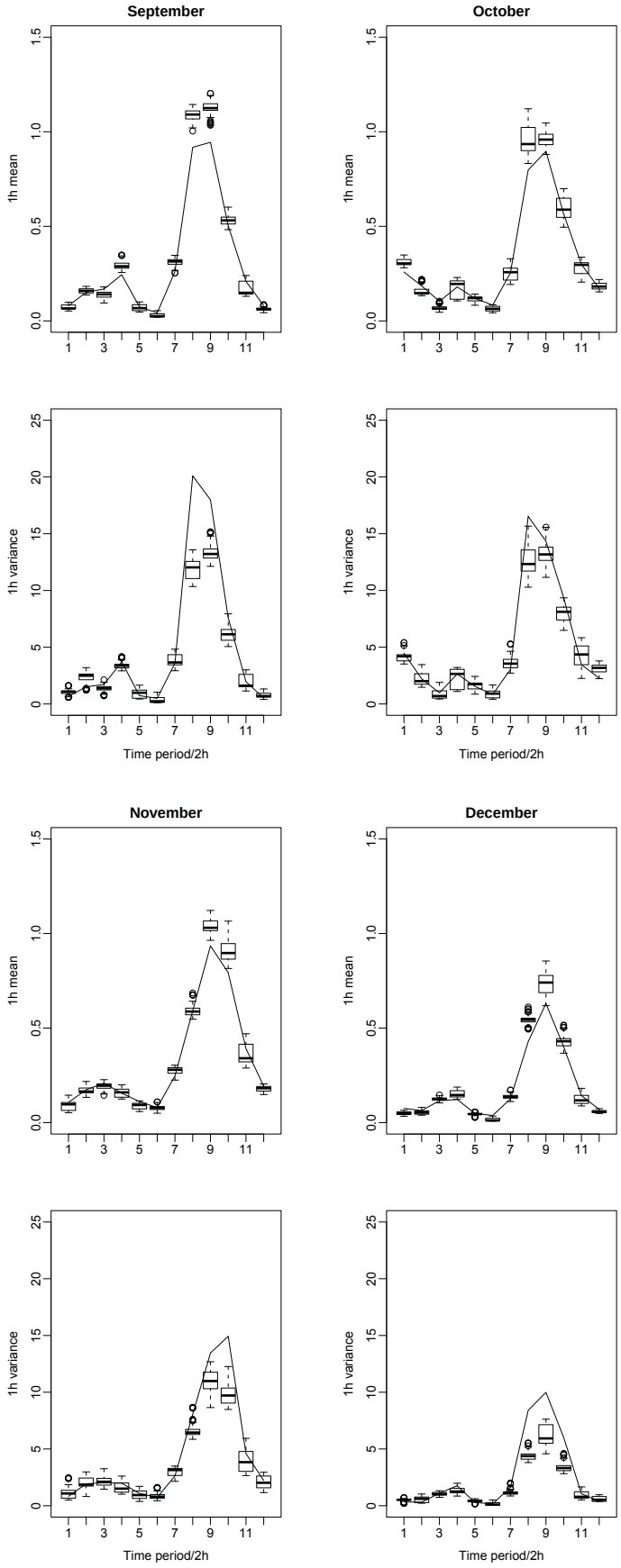


Figure 4.3: The mean and variance of September, October, November and December

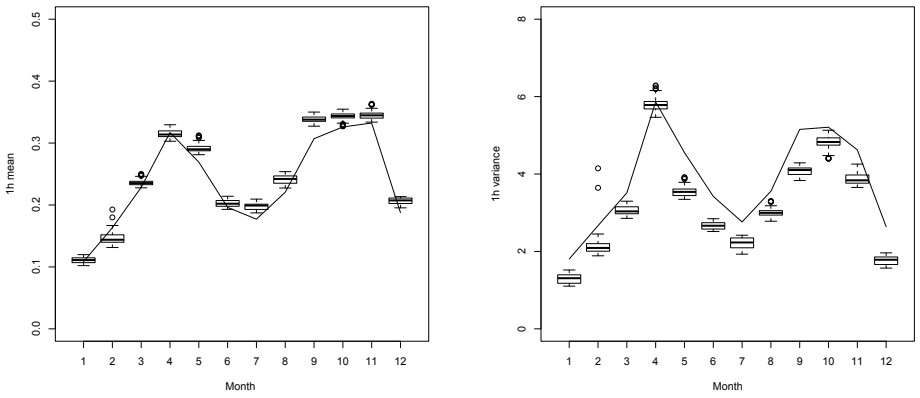


Figure 4.4: The hourly mean and variance for every month

4.4.3 Monthly moments

The hourly mean and variance for every month are calculated and plotted against the historical values. The line in the plots in Figure 4.4 present the hourly mean and variance of the historical data and the boxplot represent the simulation values. The mean has a good fit except for a few months that are overestimated which reflect the block mean. For example, April has a good mean fit in the hourly mean block and the monthly mean is also a good fit. The variance is underestimated for all the months except April but the overall shape is following the historical pattern.

4.5 Summary and conclusion

The PWN model was chosen to be used in the hourly data because of the simplicity of the model. With only two parameters that defined the rate of the storm and the depth of the rainfall, we wanted to see if having the two parameters would be enough to handle the diurnal pattern in the hourly rainfall. Data was divided into 12 blocks of 2 hours, with the model fitted to each month separately. This is to ensure the diurnal and seasonal patterns are accounted for.

The parameter η, for the depth of the rainfall is fixed to be the same throughout the 12 blocks within each month to make the model more parsimonius. So that the rate of the storm is different throughout the day, but with the same depth of pulse. The results for the λ give values that closely resemble the mean rainfall pattern of the day. We have a high rate of storm during the peak hours of rainfall, which is during 4-5pm. However, the results still need improvement. The model overestimated the mean rainfall during the heavy rainfall which is from 2-7pm and underestimated the variance. Since when estimating the parameters for every two block we assumed rainfall events are independent, this seems to be the obvious assumption to liberalise. This is the topic for Chapter 5.

Chapter 5

X-PWN model & Markov X-PWN model

5.1 Introduction

The PWN process models the rainfall event as a sequence of instantaneous bursts of rain. It reflects the convective rainfall that occurs in Malaysia, a short-lived storm with heavy downpours. In Chapter 4, the PWN model was non-homogenous model that fitted the hourly rainfall data in two hour blocks which allows extra parameters to capture the diurnal pattern. In one month, there are 12 blocks of two hour data. In every block, there are different rainfall occurrences rate, λ, with the same depth η per block of the day. A total of 13 parameters per month are needed to simulate the rainfall data. However the results from the modified PWN model were not encouraging. This poor performance may be due to the assumption of the independence between the blocks. The λ values in Table 4.2 shows that it does not follow the diurnal pattern in the two hour block. Therefore by allowing the rate of rainfall occurrence to be different for every two hour block, the diurnal variation in the rainfall data is still not fitted properly.

In order to provide a more realistic characterization of the rain process in the two hour blocks, an indicator variable and a dependence parameter as defined on page 29, are added to the main model. The PWN model is fitted to each month separately to account for the monthly seasonal effects. The PWN process is the same throughout the whole month. There will be only one value of λ and η. The indicator variable α represent the diurnal variation in the rainfall data while the dependent variable γ links the hourly rainfall in the two hour block. The new modification models still maintain the PWN process while acknowledging the diurnal cycle feature.

In this chapter two models are explored. The X-PWN model which includes the indicator variable and Markov X-PWN model with the dependent variable.

5.2 X-PWN model specification & properties

The X-PWN model has been described in Chapter 3. The X symbolizes a rainfall event in the two hour blocks. If there is a rainfall, then the intensity will be calculated from then PWN model. The expected value of the hourly rainfall in a two hour block is $\alpha E(Y)$, where Y is the PWN process. The variance in the two hour block is

$$Var[XY] = \alpha_i(Var[Y] + E^2[Y]) - \alpha_i^2 E^2[Y] \quad i = 1, 2, ..., 12 \tag{5.1}$$

If Y^1 and Y^2 are the PWN process in the first and second hour of the block and X^1 and X^2 are the indicator variables for the first and second hour in a block, where the Xs and Ys are assumed independent, then the covariance between hours within the block is

$$Cov(X^1Y^1, X^2Y^2) = \alpha_i^2 Cov(Y^1, Y^2) \quad i = 1, 2, ..., 12 \tag{5.2}$$

where α_i is the probability of the rainfall event for every block. Since $Cov(Y^1, Y^2)$ is from the PWN process, the covariance is zero.

5.3 Fitting procedure & simulation

For model fitting, the method of moments was used by using the mean, variance and covariance of the two hour block. The fitting procedure is based on minimising the following sum of squares

$$\sum_{i=1}^{36} \left(1 - \frac{z_i}{\hat{z}_i}\right)^2 \tag{5.3}$$

where z_i is the theoretical moment and $\hat{z}_i$ is the equivalent observed moment of the hourly mean, variance and autocorrelation for every block. Since α is the probability of a rainfall event, in a 2 hour block and the probability of an event happening is different in each block, we have 12 different α values. But the PWN process is assumed the same throughout the day. So there is only a single value for the rate of the pulse λ, and the pulse depth η, for every month. The parameter estimates are used to simulate rainfall data. There are 100 simulations each over a 36 years period. The simulation start with the PWN process, then the hourly rainfall distribution of the two hour block is constructed using the probability α.

5.4 Analysis

5.4.1 Parameter estimates

For each month, there are 14 parameter estimates, which consist of twelve α, λ and η. The resulting λ and η are shown in Table 5.1. The wettest month, November has the biggest λ, which implies more pulses of rainfall compared to the other months. January, the driest month, has the smallest value of λ. But the pulse depths are randomly different

Table 5.1: Parameter estimates for the pulse rate λ and depth η. The units hour^{-1} for λ and mm for η.

Month	λ	η
1	0.2559	5.3969
2	0.3555	5.9935
3	0.5460	4.5302
4	0.5313	7.0238
5	0.3790	5.7686
6	0.3701	5.2448
7	0.3871	5.0117
8	0.5329	4.8443
9	0.5996	4.6987
10	1.0867	4.5903
11	1.1890	3.7027
12	0.4723	3.6125

for every month, with no distinguishing pattern for wet or dry month.

The estimates of parameter α are given in Table 5.2. There are twelve α, each for the two hour blocks starting from 12am to 1.59p.m. In keeping with the diurnal pattern the α values are larger in blocks 8-10 (2pm-8pm). For the other α values, the pattern are similar with the diurnal cycle. The diurnal pattern can be seen in the hourly mean plot in Figure 3.9.

Table 5.2: Parameter estimates of alpha. The units are hour^{-1}

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
α_1	0.0538	0.0588	0.0469	0.0513	0.0699	0.0540	0.0460	0.0083	0.0227	0.0565	0.0185	0.0258
α_2	0.0347	0.0833	0.0247	0.0333	0.0152	0.0104	0.0564	0.0396	0.0483	0.0336	0.0386	0.0249
α_3	0.0145	0.1097	0.0207	0.0307	0.0355	0.0358	0.0612	0.0381	0.0545	0.0160	0.0428	0.0706
α_4	0.0483	0.0683	0.0281	0.0400	0.0291	0.0496	0.0237	0.0553	0.0956	0.0368	0.0369	0.0841
α_5	0.0537	0.0374	0.0184	0.0157	0.0288	0.0285	0.0520	0.0350	0.0234	0.0230	0.0228	0.0216
α_6	0.0130	0.0149	0.0173	0.0143	0.0321	0.0633	0.0337	0.0268	0.0148	0.0135	0.0157	0.0129
α_7	0.0561	0.02782	0.0329	0.0992	0.1200	0.0754	0.0663	0.0873	0.0988	0.0466	0.0538	0.0848
α_8	0.1273	0.0732	0.2004	0.2160	0.4025	0.1777	0.2646	0.2146	0.3855	0.1847	0.1458	0.3006
α_9	0.2666	0.1876	0.3379	0.2477	0.4359	0.5200	0.3407	0.3271	0.3914	0.1939	0.2364	0.4365
α_{10}	0.2055	0.1831	0.2771	0.1617	0.2467	0.2059	0.1749	0.2058	0.1976	0.1227	0.2108	0.2773
α_{11}	0.0911	0.0307	0.1500	0.0865	0.1078	0.0562	0.0560	0.0459	0.0642	0.0545	0.0902	0.0756
α_{12}	0.0222	0.0336	0.0535	0.0350	0.0438	0.0201	0.0168	0.0231	0.0183	0.0333	0.0437	0.0310

5.4.2 Moments

Figures in this chapter have the same presentation. Historical values are represented by lines and simulation results or model fitted values represented by a box plot. Based on the estimated parameters, 100 simulation samples of 36 years were generated from the X-PWN model. Model performance in fitting the moments mean and variance are compared with the simulation results. If the observed mean or variance are in the boxplot, we conclude that the model is well fitted to the data.

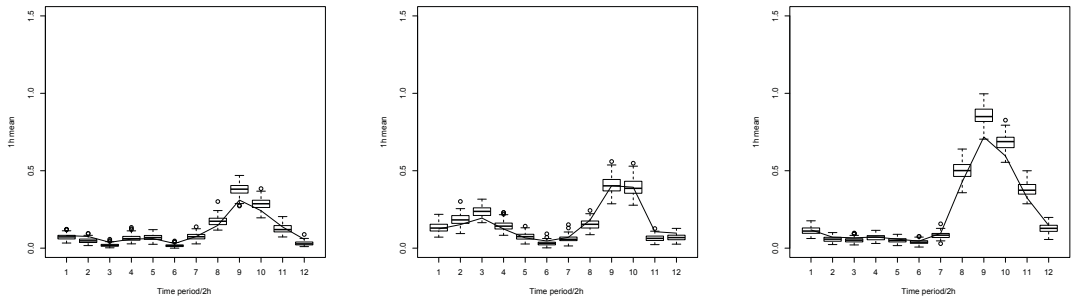
Figure 5.1 shows the mean of 2 hour blocks from January to December. The low rainfall blocks on the left show a good match between the observed and the simulation means. However the maximum mean rainfall that occurs during late afternoon, from 2pm until 7.59pm, is overestimated by the fitted model for most of the months excluding February and April.

The variance plot in Figure 5.2 also shows a good estimate compared with the observed value, especially in blocks 1 to 7 and blocks 11 to 12 where the minimum rainfall occurred. Where there was a good estimation for the mean in February and April, we see the variance is also matched well.

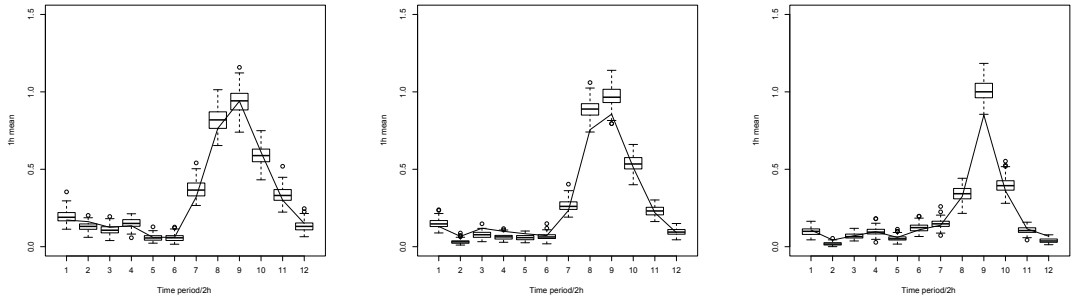
Overall, we conclude that the X-PWN model is not reproducing the moment properties well enough. The model can give a very good fit to normal rainy hours, but during the late afternoon rainfall, the diurnal effect still cannot be fitted from using this model (even though February and April gave a good fit).

5.4.3 Extreme values

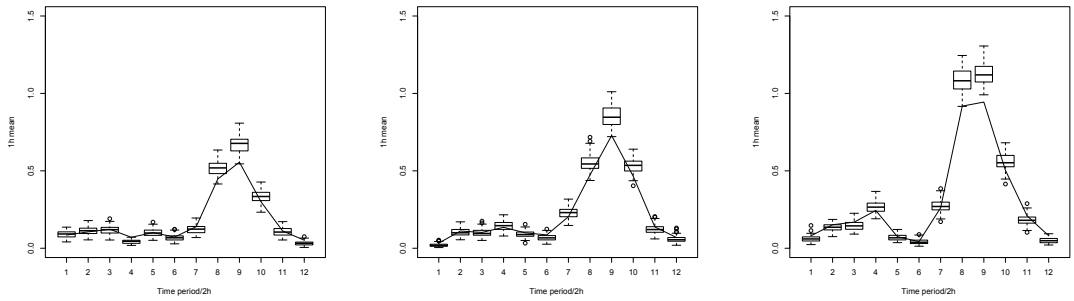
In the fitting procedure of the X-PWN model, only the mean and variance of the two hour blocks are used. The preceding section measured the model performance by comparing the observed moments with simulated moments. Another approach is by plotting the annual extreme values against the reduced Gumbel variate since the extreme values are not included in the fitting procedure. Each simulation is over 36 year sampling period. For every simulation there are 36 observed annual maxima. After placing these in order, they are plotted against the reduced Gumbel variate which is calculated by -log(-log((*i*-0.3)/36.4)), i=1,2,...,36. In Figure 5.3, the line is the historical annual maximum rainfall. The dashes are the simulated annual maxima. If the line plot is going through the dashes plot, then the model has a good fit to the extreme values. Unfortunately, the extreme values are underestimated for all the plots which are the hourly, two hours, six hours and twenty-four hours annual maximum. The inability of the X-PWN model to reproduce the annual extreme values at different time scales, makes the model unsuccessful.



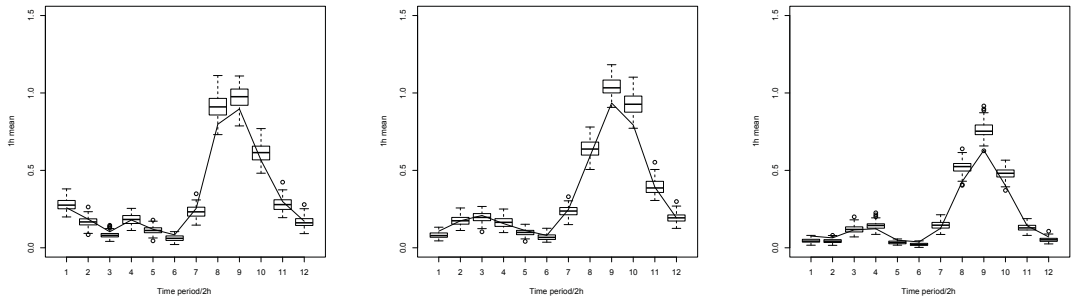
(a) January (b) February (c) March



(d) April (e) May (f) June

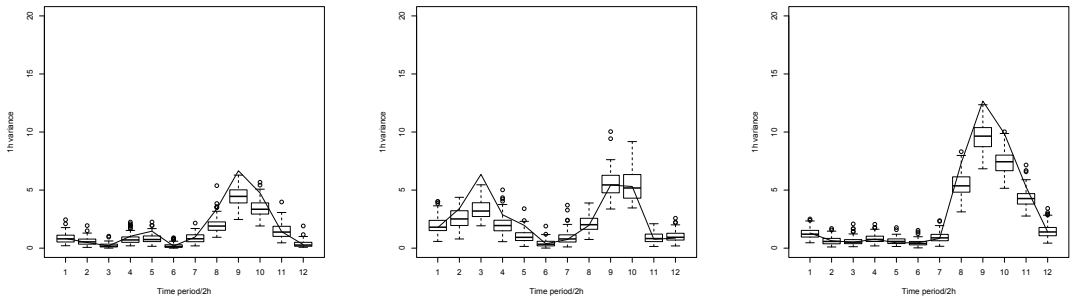


(g) July (h) August (i) September

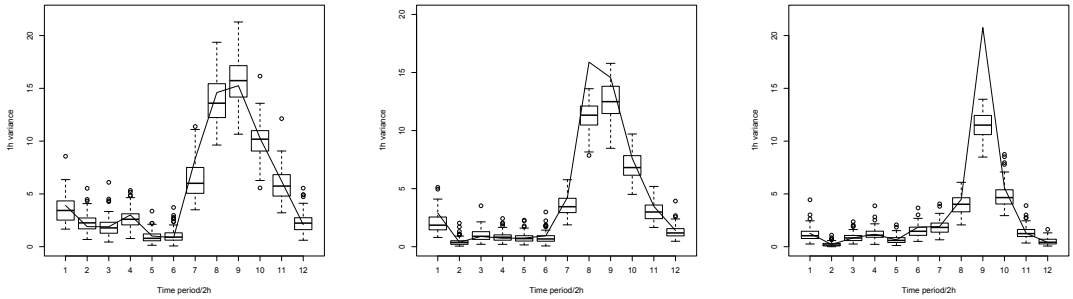


(j) October (k) November (l) December

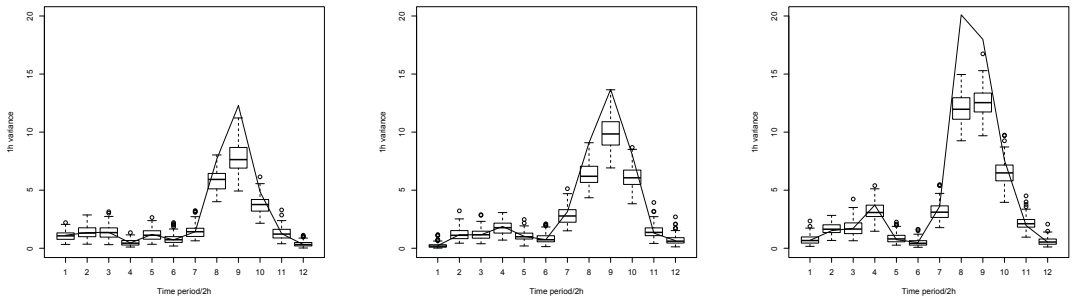
Figure 5.1: The mean of 2 hour blocks for every month



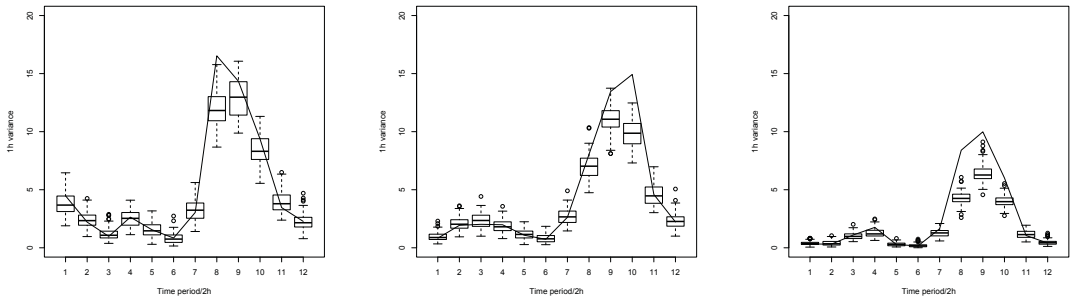
(a) January (b) February (c) March



(d) April (e) May (f) June

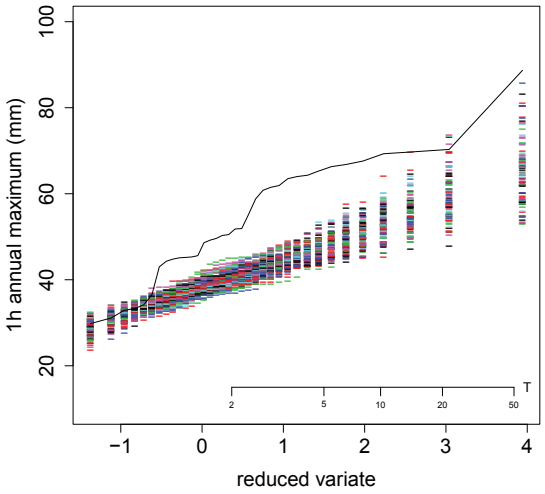


(g) July (h) August (i) September

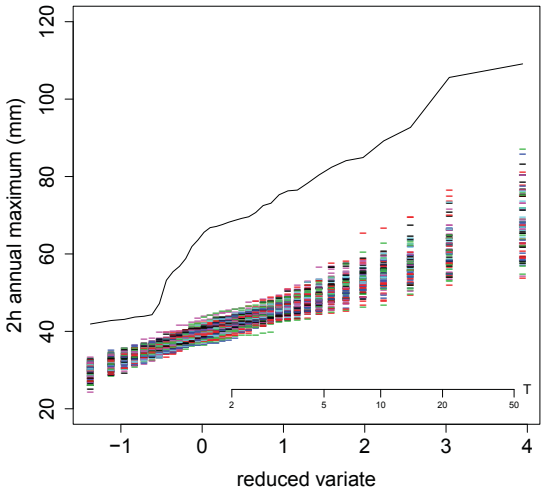


(j) October (k) November (l) December

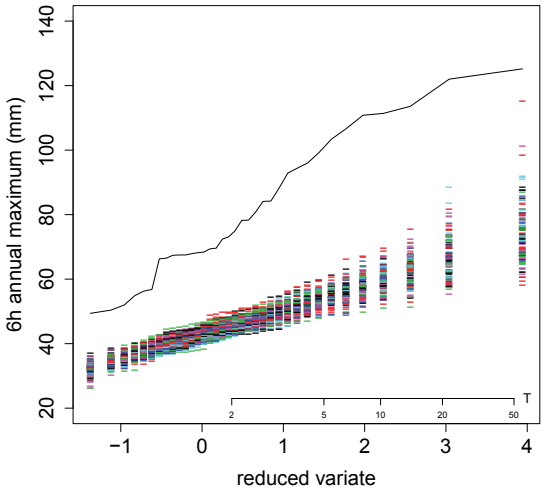
Figure 5.2: The variance of 2 hour blocks for every month



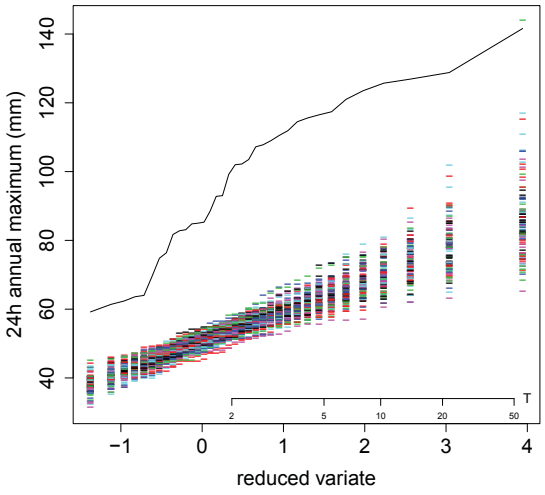
(a) 1 hour



(b) 2 hour



(c) 6 hour



(d) 24 hour

Figure 5.3: Annual extreme values for X-PWN

5.4.4 Monthly moments

In Figure 5.4 the simulation data is aggregated to monthly hourly means and variances. The monthly hourly moments were not used in the parameter estimation process but the model was able to fit the monthly hourly data. However the mean is overestimated for some months and the variance is underestimated for all of the months. The unsatisfactory performance of the monthly moments fitting could contribute to the poor fitting of the extreme values.

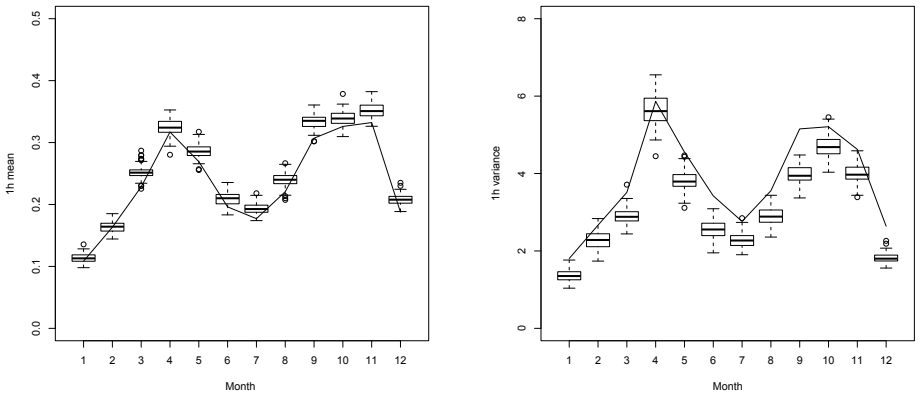


Figure 5.4: The hourly mean and variance for every month

5.5 Markov X-PWN model specification & properties

One of the main assumptions of the X-PWN model is independence between the hours within the blocks. This assumption was made because of the simplicity of the mathematical formulation. However, since the rainfall event could last longer than an hour this assumption is not valid. The unsatisfactory results from the X-PWN model also showed that the assumption of independence between the hours in the two hours block may not be reliable.

To assume dependence between the two hours in the block, Markov dependence is attached to the Bernoulli trials. A dependence parameter γ is added to the Markov chains.

The mean and variance are the same as X-PWN model,

$$E(XY) = \alpha_i E(Y) \qquad i = 1, 2, ..., 12 \tag{5.4}$$

$$Var(XY) = \alpha_i (Var[Y] + E^2[Y]) - \alpha^2 E^2[Y] \quad i = 1, 2, ..., 12 \tag{5.5}$$

with X as the indicator variables and Y the PWN process. The 12 blocks have different α values, but the same PWN process. To calculate the covariance within hours of the block,

$$\begin{aligned} Cov(X^1Y^1, X^2Y^2) &= \left(\alpha_i^2 + \alpha_i(1-\alpha_i)\left[\frac{\gamma-\alpha_1}{1-\alpha_1}\right]\right)[Cov(Y^1,Y^2) + E^2(Y_i)] \\ &\quad -\alpha_i^2 E^2(Y_i) \quad i = 1,2,...,12 \end{aligned} \tag{5.6}$$

where Y^1 and Y^2 are the PWN process in the first and second hour of the block and X^1 and X^2 are the indicator variables for the first and second hour in a block, which are assumed dependent. γ is the dependent variables discuss in 3.1.3.

5.6 Fitting procedure & simulation

The model was fitted to the monthly rainfall data separately. For every month, the mean, variance and autocorrelation are used in the method of moments to estimate the parameters of the pulse rate, λ, the pulse depth, η, the rainfall indicator, α and the dependence parameter, γ. The PWN parameters are assumed to be the same throughout the whole day while the rainfall events are different for every two hours.

In the simulation process, the Markov chains generate the sequence of wet and dry hours in the two hour block. The transition probabilities matrix is

$$P = \begin{pmatrix} (1-2\alpha+\gamma\alpha)/(1-\alpha) & (1-\gamma)\alpha/(1-\alpha) \\ 1-\gamma & \gamma \end{pmatrix} \tag{5.7}$$

The PWN process then generates the intensities for the wet days.

5.7 Analysis

5.7.1 Parameter estimates

The Markov X-PWN model has four parameters. The parameter estimates for λ and η are given in Table 5.3. The rates of the rainfall occurrence are smaller compared to rates from the X-PWN model. But the values do not seem to fit the wet and dry months characteristics. The pulse depth is also smaller than the pulse depth for the X-PWN model. The lack of pattern does not seem to be due to convergence issues.

Estimates for α are listed by column for months and by row for two hours block in Table 5.4. The estimates are following the diurnal pattern. α_9 has the highest value for all of the months, occurring at the time of the highest peak of mean rainfall. Table 5.5 give the estimates for γ. The estimate does not show any pattern throughout the day and

also across the months.

Table 5.3: Parameter estimates for the pulse rate λ and depth η. The units hour^{-1} for λ and mm for η.

Month	λ	η
1	3.3056	2.2340
2	3.9866	2.5088
3	2.4278	2.6740
4	4.6054	3.3247
5	3.5880	2.7020
6	8.6173	1.2480
7	9.9545	1.0833
8	3.3419	2.3983
9	6.5481	1.5733
10	6.5413	1.7799
11	7.1595	1.3661
12	6.4519	1.1010

5.7.2 Moments

Based on the estimated model parameters, 100 simulated samples were generated from the Markov X-PWN model. The historical mean, variance and autocorrelation are compared with the simulation values. In Figure 5.5, the mean at different blocks have been matched well. However block 9 is still overestimated for half of the 12 months.

The variance plot in Figure 5.6 shows the simulation values are still underestimated in block 8-10. For February, these blocks have a good estimate of the variance since the mean is well estimated. However, the variance is underestimated in blocks 2-5 even though the mean has an exact fit. The autocorrelation plot (Figure 5.7) shows a very good fit in all the months with only a few fitted values underestimated. The Markov X-PWN model estimates the autocorrelation value, which the X-PWN model does not.

The mean square error (MSE) values and the number of parameters per month for the PWN, X-PWN and Markov X-PWN model are presented in Table 5.6. The MSE values represent the total squared errors from the differences between the fitted values and the historical values from the pooled monthly properties. MSE totals are calculated from the 144 pooled monthly properties. The X-PWN model does not improve upon PWN model even though it has an extra parameter. The mean has the same fitting but the variance

Table 5.4: Parameter estimates of alpha. The units are hour^{-1}

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
α_1	0.0102	0.0121	0.0179	0.0114	0.0153	0.0094	0.00796	0.0026	0.0054	0.0236	0.0078	0.0061
α_2	0.0066	0.0174	0.0091	0.0066	0.0032	0.0017	0.0097	0.0123	0.0123	0.0141	0.0171	0.0057
α_3	0.0028	0.0228	0.0076	0.0063	0.0073	0.0063	0.0102	0.0118	0.0146	0.0062	0.01799	0.0164
α_4	0.0091	0.0143	0.0108	0.0082	0.0061	0.0083	0.0039	0.0175	0.0268	0.0160	0.0163	0.01992
α_5	0.0101	0.0079	0.0067	0.0031	0.0059	0.0049	0.00895	0.0109	0.0064	0.0095	0.0095	0.00498
α_6	0.0025	0.0029	0.0062	0.0027	0.0068	0.0112	0.00598	0.0083	0.0031	0.0054	0.0068	0.0029
α_7	0.0107	0.0056	0.0119	0.0213	0.0258	0.0131	0.0112	0.0274	0.0251	0.0181	0.0224	0.0199
α_8	0.0238	0.0148	0.0772	0.0433	0.0926	0.0311	0.0466	0.0681	0.1055	0.0790	0.0656	0.0723
α_9	0.0499	0.0378	0.1281	0.0536	0.0989	0.0948	0.0584	0.1043	0.1023	0.0741	0.1045	0.1004
α_{10}	0.0385	0.0390	0.1012	0.0347	0.0509	0.0368	0.0301	0.0649	0.0531	0.0482	0.0932	0.06495
α_{11}	0.0173	0.0062	0.0587	0.0191	0.0234	0.0096	0.0091	0.0143	0.0162	0.0220	0.0388	0.0178
α_{12}	0.0043	0.0069	0.0204	0.0069	0.0094	0.0034	0.0028	0.0072	0.0046	0.0140	0.0186	0.0073

Table 5.5: Parameter estimates of gamma. The units are hour^{-1}

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
γ_1	0.3928	0.7994	0.4378	0.3245	0.5580	0.8185	0.2632	0.1689	0.2735	0.8759	0.4534	0.6315
γ_2	0.9984	0.5196	0.4948	0.8751	0.2958	0.4759	0.3950	0.4758	0.5906	0.4875	0.3760	0.8340
γ_3	0.3425	0.7253	0.8278	0.2240	0.9913	0.5687	0.5178	0.2964	0.9089	0.2645	0.4382	0.4453
γ_4	0.3292	0.1351	0.9754	0.9850	0.7434	0.4669	0.3751	0.8961	0.5262	0.5042	0.5165	0.9059
γ_5	0.9961	0.1697	0.2556	0.2151	0.3845	0.7159	0.7240	0.3918	0.1411	0.5212	0.5471	0.7220
γ_6	0.5215	0.2713	0.4501	0.0235	0.9679	0.1984	0.9059	0.1630	0.1388	0.3891	0.3208	0.1890
γ_7	0.8901	0.7212	0.4985	0.4245	0.7281	0.2632	0.1853	0.2046	0.3438	0.1732	0.1294	0.44878
γ_8	0.2621	0.4326	0.6557	0.2480	0.3410	0.5510	0.3068	0.4216	0.4391	0.5433	0.3476	0.1758
γ_9	0.7105	0.5693	0.4740	0.4738	0.4755	0.4633	0.2948	0.8794	0.4751	0.4970	0.4583	0.3527
γ_{10}	0.1420	0.4532	0.4861	0.5107	0.6547	0.3098	0.4410	0.4441	0.4295	0.3146	0.5547	0.3354
γ_{11}	0.2624	0.2090	0.6302	0.2786	0.5082	0.2767	0.5304	0.8309	0.4015	0.3707	0.4623	0.1699
γ_{12}	0.3804	0.1462	0.3184	0.5622	0.5814	0.2860	0.3106	0.4720	0.6586	0.1078	0.4564	0.2741

has a bigger error. The Markov X-PWN model improves significantly upon X-PWN model but with an increased number of parameters. This implies that the rainfall process is better represented by the Markov X-PWN model due to the assumption of the dependence between the hours in the two hour block.

Table 5.6: Model comparison in terms of MSE

Model	Parameters (per month)	MSE Mean	MSE Variance	MSE Autocorrelation	MSE Total
PWN	13	0.0022	2.5553	-	1.2788
X-PWN	14	0.0021	2.8249	-	1.4135
Markov X-PWN	26	0.0007	2.3693	0.00065	0.7902

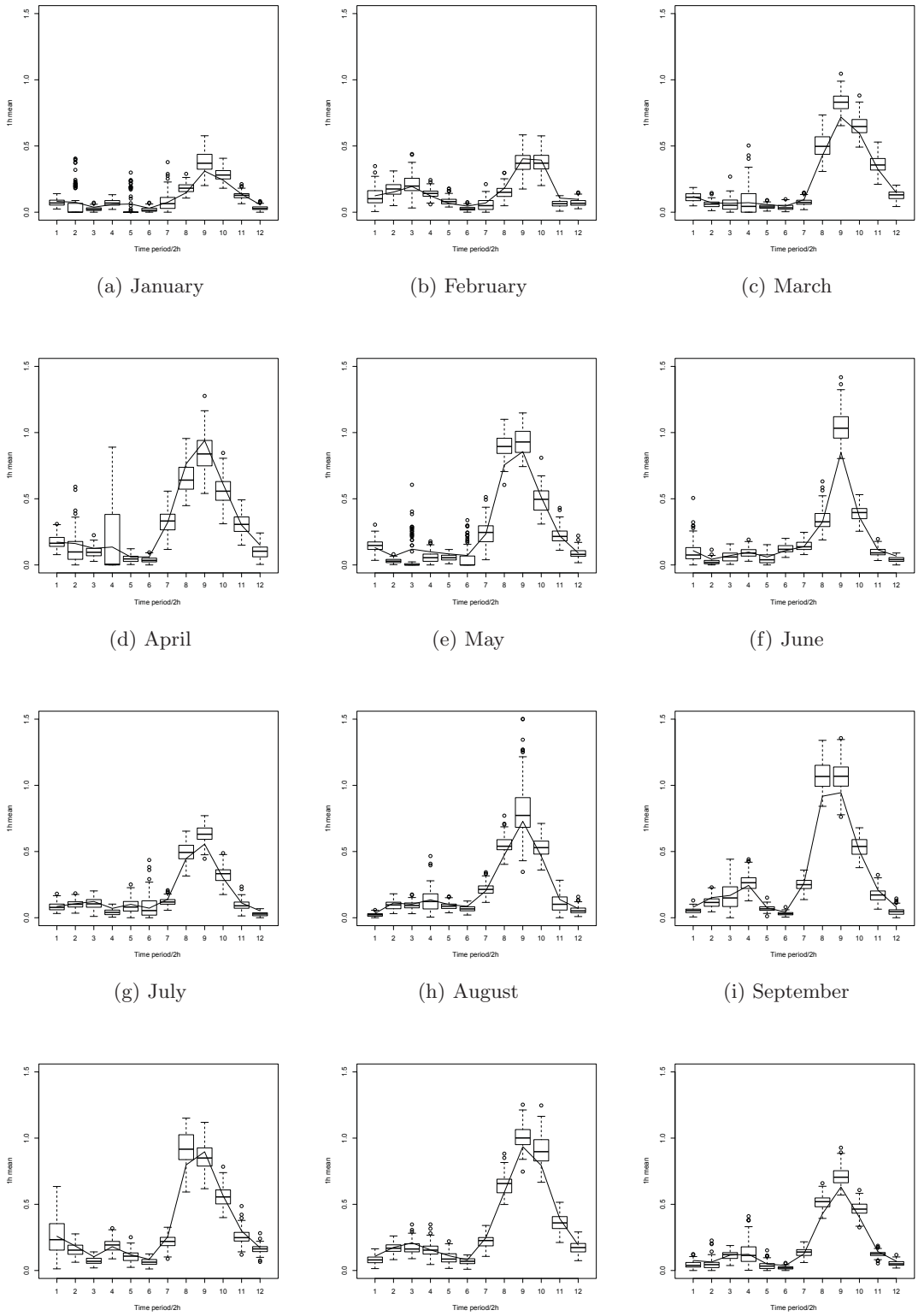
5.7.3 Extreme values

The ordered annual maxima against the reduced variate at four different aggregation time scales, 1, 2, 6 and 24 hours are plotted in Figure 5.8. The historical values are shown as a solid line. The dashes represent the ordered annual maxima from 100 simulation samples.

The model is producing lower extreme values of rainfall comparing with the historical values for all of the aggregation time scales. Since extreme values are not included in the fitting procedure, thus the model fails to fit the adequate annual maxima with the historical values.

5.7.4 Monthly moments

The monthly hour mean, variance and covariance were calculated from the simulation data of Markov X-PWN model. In Figure 5.9, the mean is well estimated. But the variance is slightly underestimated and the autocorrelation is not fitted well. Even though the Markov X-PWN model has more parameters, it did not improved the fitting of the two hour block variance thus affecting the fitting of the monthly variance. Although the autocorrelation is fitted for the Markov X-PWN model, it seems the assumption of dependence between hours in the two hour block, (and also the assumption of stationary from the PWN process) are not enough to simulate the monthly hourly autocorrelation.



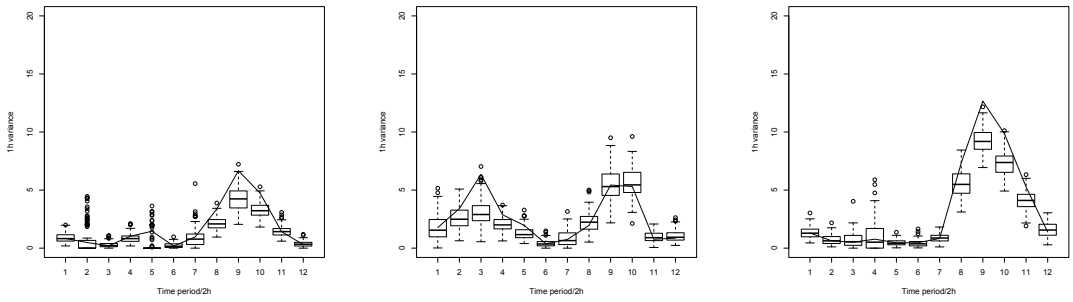
(a) January (b) February (c) March

(d) April (e) May (f) June

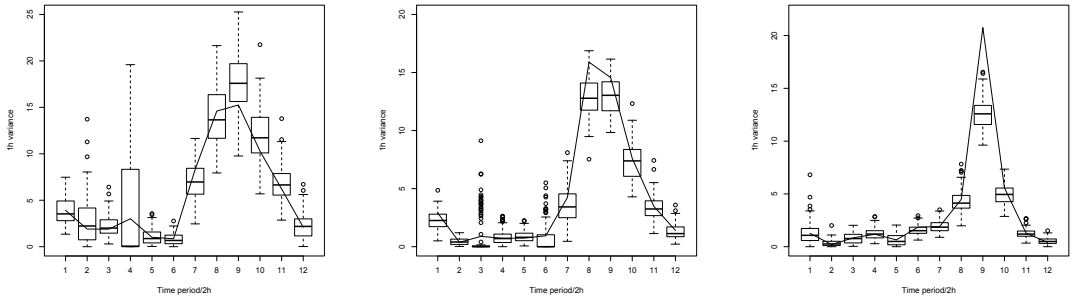
(g) July (h) August (i) September

(j) October (k) November (l) December

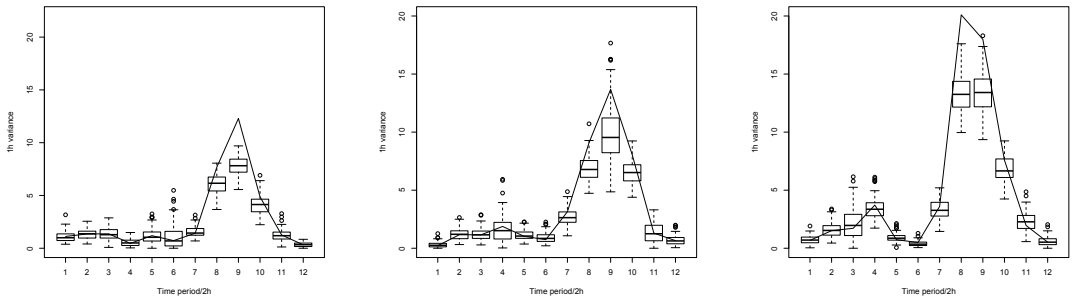
Figure 5.5: The mean of 2 hour blocks for every month



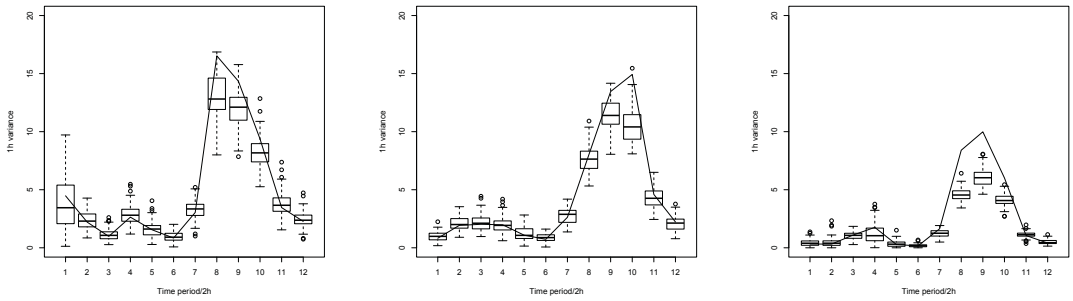
(a) January (b) February (c) March



(d) April (e) May (f) June

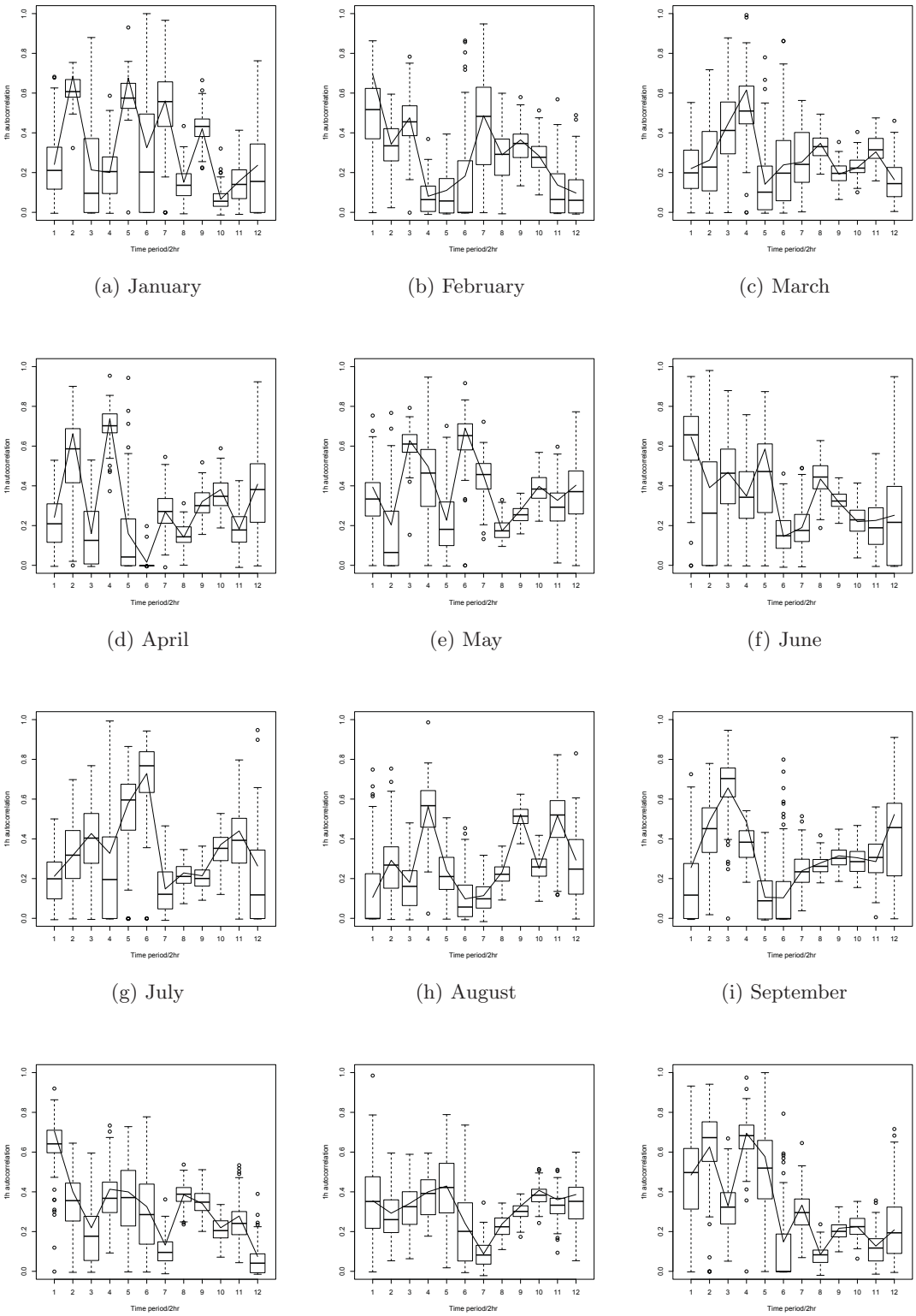


(g) July (h) August (i) September



(j) October (k) November (l) December

Figure 5.6: The variance of 2 hour blocks for every month



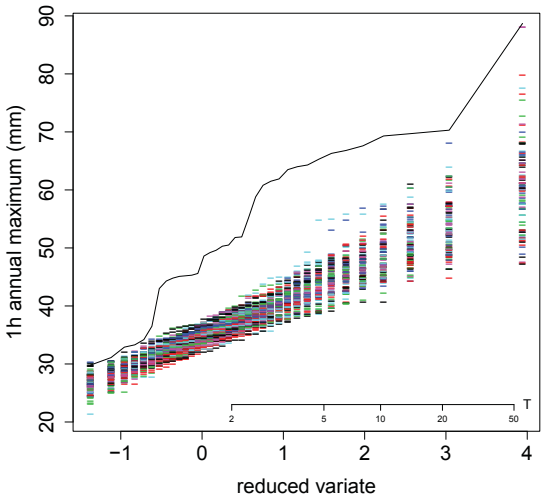
(a) January (b) February (c) March

(d) April (e) May (f) June

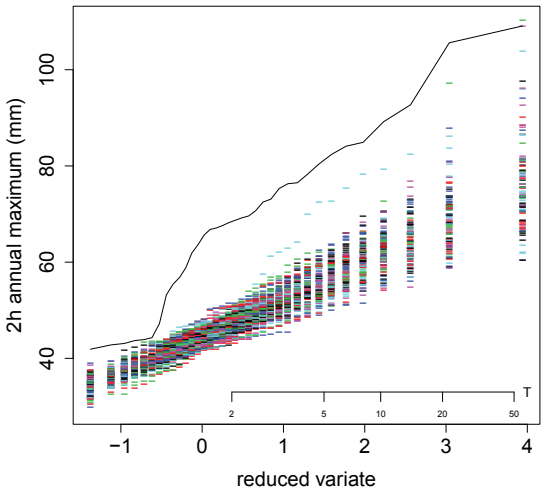
(g) July (h) August (i) September

(j) October (k) November (l) December

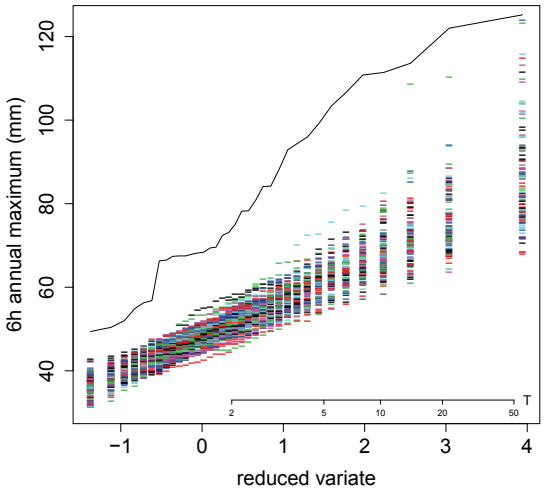
Figure 5.7: The autocorrelation of 2 hour blocks for every month



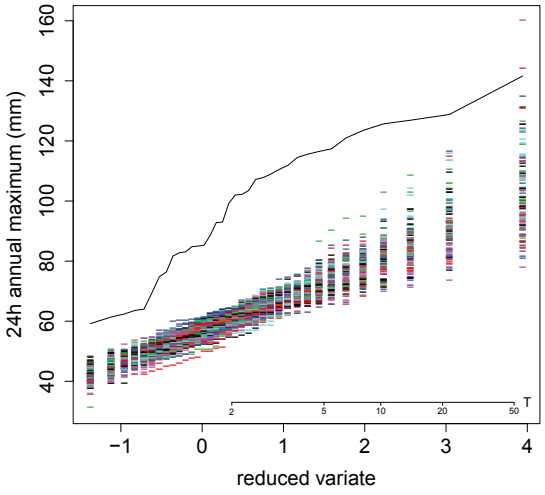
(a) 1 hour



(b) 2 hour



(c) 6 hour



(d) 24 hour

Figure 5.8: Annual extreme values for Markov X-PWN

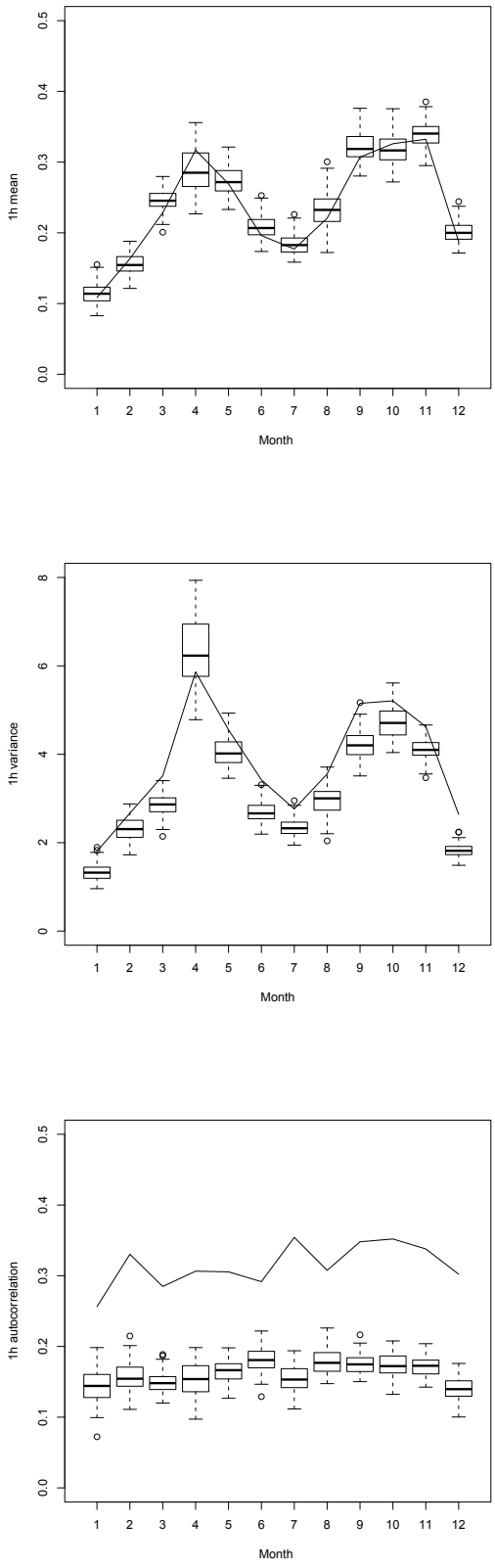


Figure 5.9: The hourly mean, variance and autocorrelation for every month

5.8 Summary and conclusion

In this chapter, a PWN model with an additional indicator for rainfall events is formulated to represent the diurnal pattern in rainfall data. The modification allows the occurrence of a rainfall event to be modelled by a binary random variable with probability α, a probability which is allowed to vary every two hours throughout the day. Since in the tropical rainfall, the rainfall event in the afternoon is higher than in the morning, therefore the parameter α will be able to capture the variation in rainfall especially the wettest hours. Two assumptions were tested about how the two hours within a two hour block are related: one that rainfall events are independent, the other that events are dependent.

The X-PWN model assumed independence between the hours within blocks. Results were satisfactory when comparing the historical means and variance with the simulation values. However in blocks 8-10, where the maximum rainfall occurs, the mean was overestimated. This might due to the independence between the hours in the two hour block not allowing the rainfall to fall in the next hours. Thus, the Markov X-PWN model was developed to include the dependence within the blocks. There is a slight improvement in the simulation values of mean and variance, but the mean is still overestimated in blocks 8-10. The Markov X-PWN model also allows one to calculate the covariance between hourly rainfalls within a block, which gave very good results comparing with the historical values.

The comparison between the PWN, X-PWN and Markov X-PWN model show that, given 12 more parameters specified for the dependence assumption in a model, the Markov X-PWN model is the best model. However, all fitted models are very poor in fitting the statistical properties in blocks 8-10, where the maximum rainfall occurs.

To validate the model, extreme values of rainfall total were plotted against the reduced Gumbel variate, since the sample extreme value is not used in the fitting procedure. However, both models do not perform well when fitting the extreme values. In conclusion, the research results of this chapter show that the rainfall indicator does effectively model the diurnal pattern and the dependence variable effectively models the covariance between hours within blocks. But since the X-PWN model and Markov X-PWN model are far from adequate, a different type of point process is tested in the next chapter.

Chapter 6

X-NSWN model & Markov X-NSWN model

6.1 Introduction

In Chapter 5, a method to capture a diurnal pattern in the point process model was introduced. A binary variable used as a rainfall indicator was added to the PWN model to observe the rainfall event in a two hours block. This will allow the maximum rainfall during late afternoon to be captured in the model. However, due to the simplicity of the PWN model with only two parameters to describe the rainfall event, the model was not able to fit the statistical properties of the maximum rainfall hours. Since the inclusion of the rainfall indicator and the Markov chain successfully captured the diurnal pattern of the two hours rainfall, the method is further investigated using the Neyman-Scott white noise (NSWN) model. The NSWN model was chosen because it still produces instantaneous pulses as in the PWN model, but the NSWN model is more elaborate in describing features of the overall rainfall event. The Neyman-Scott process for rainfall occurrence are associated with a cluster of rain cells. By better fitting the non-diurnal aspects of the rainfall it will give scope to better model the diurnal aspects as well. In this chapter the X-NSWN model and Markov X-NSWN model are introduced and analyzed by comparing with the X-PWN and Markov X-PWN model.

6.2 X-NSWN model specification & properties

The NSWN model has been introduced in Chapter 3. It is a Poisson cluster process with instantaneous pulses. The storm arrival with the Poisson process of rate λ is followed by a random number of rain cell origins, ν, with waiting times that are exponentially distributed with rate β. The rain cells are assumed to be instantaneous pulses with random depth, U, which is exponentially distributed with parameter η. The formulas of mean, variance

and covariance from Rodriguez-Iturbe *et al.* (1984) are given below.

$$E[Y] = h\lambda E[U]E[\nu] \tag{6.1}$$

$$Var[Y] = \theta_1 h + \frac{2\theta_2}{\beta^2}[\beta h - 1 - e^{-\beta h}] \tag{6.2}$$

$$Cov[Y_i, Y_k] = \frac{\theta_2}{\beta^2}(1 - e^{-\beta h})^2 e^{-\beta(k-1)h} \quad k \geq 1 \tag{6.3}$$

with $\theta_1 = 2\lambda E[U]^2 E(\nu)$ and $\theta_2 = \lambda\beta E[U]^2 E(\nu(\nu-1))/2$.

Let Y be the NSWN process and X_i $(i = 1, 2, ..., 12)$ the binary nonzero rainfall indicator variable for every two hour block. The mean and variance for X-NSWN model are

$$E[X_i Y] = \alpha_i E(Y) \tag{6.4}$$

$$Var[X_i Y] = \alpha_i(Var[Y] + E^2[Y]) - \alpha_i^2 E^2[Y] \tag{6.5}$$

If Y^1 and Y^2 are the NSWN process in the first and second hour of the block and X^1 and X^2 are the indicator variables for the first and second hour in a block, which are assumed independent, then the covariance between hours within the block is

$$Cov(X_i^1, Y^1 X_i^2 Y^2) = \alpha_i^2 Cov(Y^1, Y^2) \tag{6.6}$$

with $i = 1, 2, ..., 12$

6.3 Fitting procedure

The X-NSWN model was fitted to the two hour block data for every month. In minimising the sum of squares, the sample estimates were calculated by using the hourly mean, variance and autocorrelation of the two hour block data. This was repeated for every month using data pooled from all the years available. The NSWN process is assumed to be the same for the 12 two hour blocks but the rainfall indicators are different. This assumption was made to allow the indicator variables to follow the diurnal pattern in the rainfall data. 12 blocks were chosen again, instead of 24 blocks, to reduce the total number of parameters. The fitting procedure was done separately for every month.

The optimization method estimated the parameters λ, ν, β, η and α. Using the estimated parameters, 100 sets of 36 years of data were simulated by the NWSN process. Then the hourly rainfall in the two hour block patterns follow the distribution of the indicator variable.

6.4 Analysis

6.4.1 Parameter estimates

Each month there are 12 blocks, each representing the two hours data starting from 0000 to 2200 hours. The X-NSWN model has four parameters from the NSWN process and one parameter for the diurnal pattern. A total of 16 parameters are estimated for each month. The parameter estimates for the NSWN model are given in Table 6.1 and the binary variable representing the rainfall indicator parameters, α are given in Table 6.2.

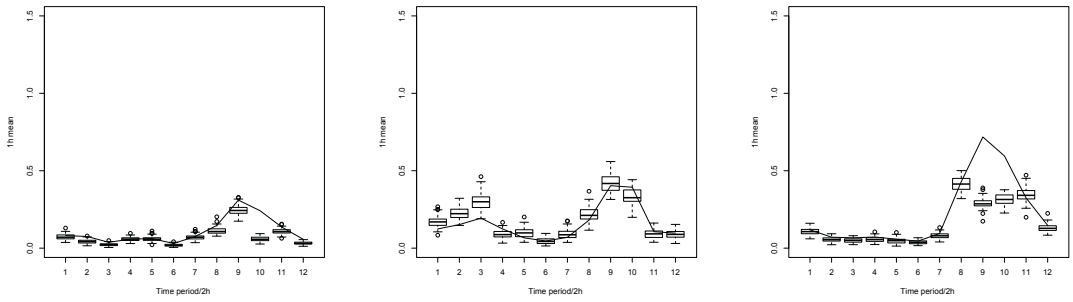
The parameter estimates β are bigger than λ which is consistent with the assumption that waiting times for the rain cells are shorter than for the storm arrival. By looking at the ν values, which represent the number of pulses, they are big which suggests that NSWN models is capturing the instantaneous pulses. But, although the estimation process is done separately for each month, the parameter estimates for the NSWN process do not adequately represent the wet and dry months behaviour. Although January is the driest month, the estimate for λ is not the lowest rate compared to the other months. On the other hand November has the highest rate and it is the wettest month.

The parameter estimates α have 12 estimated values for every month. α_9 should be the highest value for all the months since it is the highest peak in the daily two hours rainfall. That is, we expect the α's to follow the rainfall pattern. But except for January, February, August and December they do not follow the expected pattern. This is a big disadvantage. In the model estimation process, different initial values in the optimization or changes in parameter space limits can all give different parameter estimates. The results given here are the best estimated values without hitting the boundaries of the limits given for each parameter.

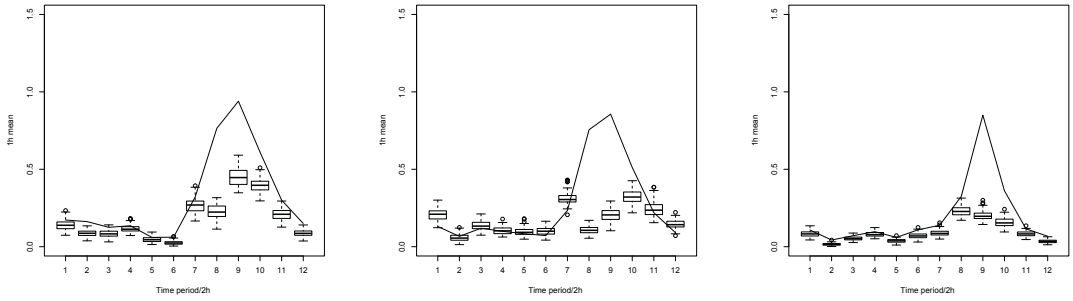
6.4.2 Moments

In Figure 6.1, 6.2 and 6.3, simulated moments, mean, variance and autocorrelation, are compared with the historical values. In all these figures, the historical values are represented by the solid line. The box plots represent the derived values of model X-NSWN. In Figure 6.1, the mean of two hour blocks are plotted for every month. The X-NSWN underestimated the mean value in blocks 8 to 10 for the majority of months. This is likely due to the estimated α's not following the diurnal pattern. In the other blocks, the estimated mean has an exact fit with the historical mean, but a few underestimated mean (for example block 2 and 3 in April) and overestimated mean (in block 2 and 3 in February).

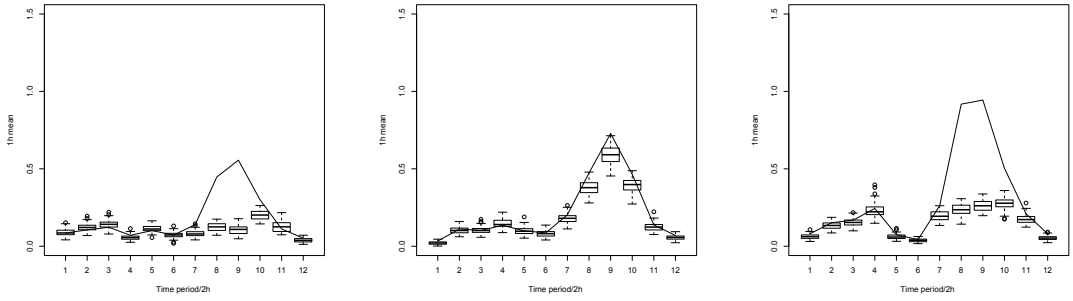
Figure 6.2 show the comparisons of two hours variance by months. An underestimated variance in blocks 8-10 is due to the underestimated mean in the diurnal blocks. However



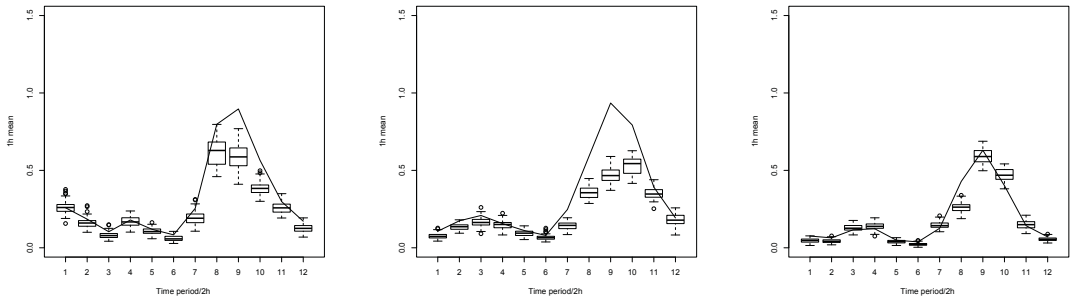
(a) January (b) February (c) March



(d) April (e) May (f) June



(g) July (h) August (i) September



(j) October (k) November (l) December

Figure 6.1: The mean of 2 hour blocks for every month

Table 6.1: Parameter estimates for the pulse depth. The units are hour^{-1} for λ and β with η the unit is mm and ν for number of pulses.

Month	λ	β	ν	η
1	0.0144	0.4284	29.3172	0.8732
2	0.0575	0.5216	20.9879	0.9793
3	0.0129	0.2522	62.3983	0.9357
4	0.0103	0.2743	95.7065	0.8269
5	0.0077	0.6328	36.8565	1.1290
6	0.0075	0.2926	50.6844	0.8219
7	0.0043	0.2294	69.48474	0.9615
8	0.0137	0.3017	58.6051	0.9668
9	0.0066	0.2005	72.0822	0.9074
10	0.0112	0.2165	90.4966	0.9156
11	0.0251	0.7082	25.4053	0.8914
12	0.0575	0.5216	20.9879	0.9793

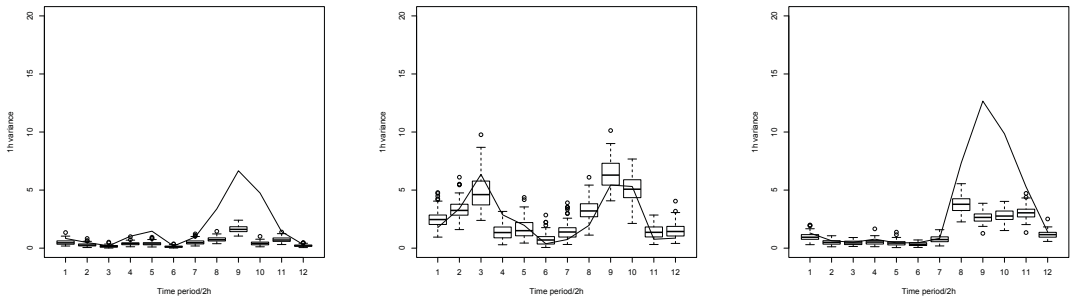
for the other blocks, the simulation values are following the historical patterns with a slight underestimation. Only in February, most of the historical values are well located within the simulation box plots.

The autocorrelation plot in Figure 6.3 shows the model does not reproduce a good estimate of the autocorrelation. The autocorrelation in the first seven blocks is underestimated for all of the months and is not following the rainfall pattern. For the rest of the blocks, the simulation values are following the diurnal pattern with values mostly matching the historical autocorrelation and just a few underestimated values in some months. Even for February with a good fit to the observed mean and variance, the autocorrelation is not well estimated.

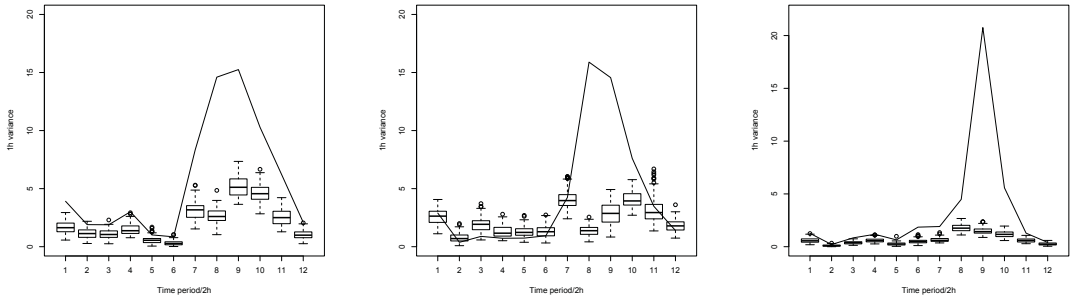
Overall, we conclude that the X-NSWN model does not reproduce the moments properties well enough for all of the months. Even though there are two more parameters per month compared with the X-PWN model, it does not improve the results. Although the X-PWN model does not have autocorrelation to compare, the mean and variance plots were better than the X-NSWN model plot. The autocorrelation plot for the X-NSWN model shows that the assumption of independence between hourly data in the two hours block is invalid.

Table 6.2: Parameter estimates of α. The units are hour^{-1}

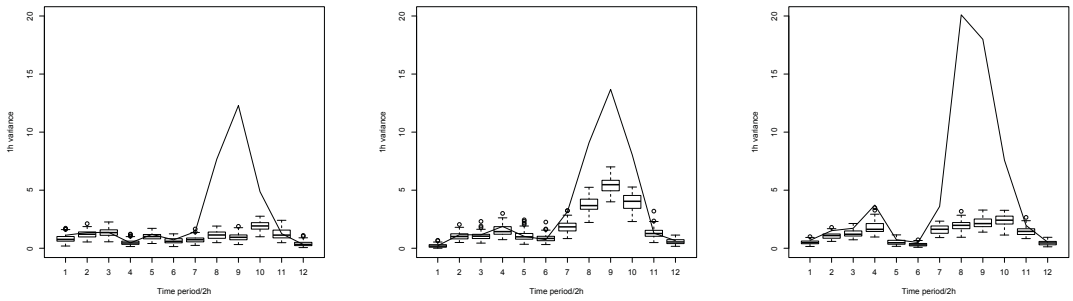
	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
α_1	0.1947	0.2801	0.1492	0.1773	0.6501	0.2569	0.3113	0.0302	0.1535	0.2853	0.1291	0.0401
α_2	0.1277	0.3857	0.0781	0.1071	0.1921	0.0488	0.4039	0.1392	0.3192	0.1739	0.2729	0.0380
α_3	0.0589	0.5107	0.0671	0.1060	0.4273	0.1698	0.4591	0.1365	0.3579	0.0828	0.3067	0.1094
α_4	0.1616	0.1571	0.0748	0.1313	0.3510	0.2495	0.1927	0.1926	0.5808	0.1873	0.2569	0.1146
α_5	0.1695	0.1590	0.0589	0.0525	0.3274	0.1304	0.3921	0.1219	0.1461	0.1147	0.1561	0.0333
α_6	0.0512	0.0728	0.0482	0.0303	0.3411	0.2581	0.2536	0.0971	0.0961	0.0675	0.1088	0.0194
α_7	0.1846	0.1315	0.1033	0.3304	0.9562	0.3113	0.2977	0.2228	0.4512	0.2112	0.2411	0.1282
α_8	0.3178	0.3445	0.5472	0.2803	0.3629	0.7431	0.4354	0.4715	0.5307	0.6785	0.6340	0.2335
α_9	0.7032	0.6474	0.3850	0.5852	0.6051	0.6382	0.3887	0.7830	0.6046	0.6368	0.8741	0.5226
α_{10}	0.1544	0.5352	0.4217	0.5174	0.9614	0.4594	0.7281	0.4991	0.6165	0.4158	0.9951	0.4052
α_{11}	0.2858	0.1499	0.4453	0.2692	0.7564	0.2691	0.4236	0.1610	0.3996	0.2793	0.6164	0.1238
α_{12}	0.0881	0.1461	0.1824	0.1163	0.4586	0.1034	0.1361	0.0810	0.1220	0.1343	0.3078	0.0479



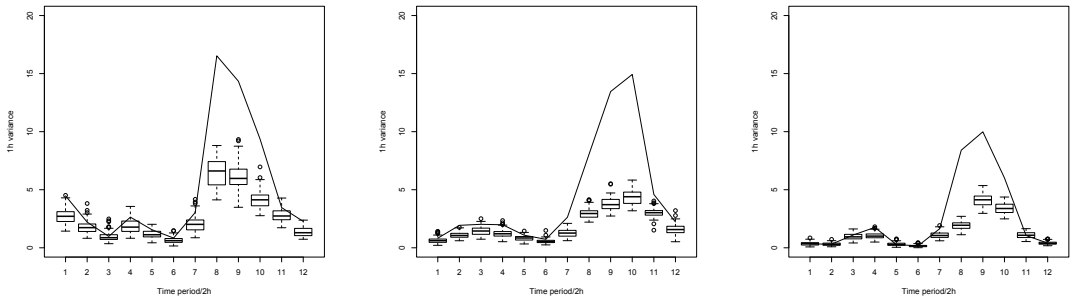
(a) January (b) February (c) March



(d) April (e) May (f) June

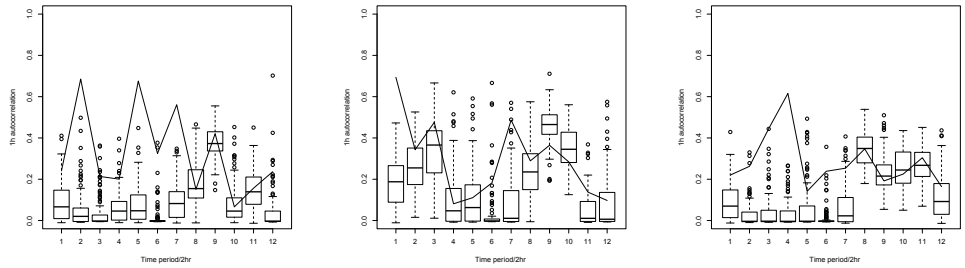


(g) July (h) August (i) September

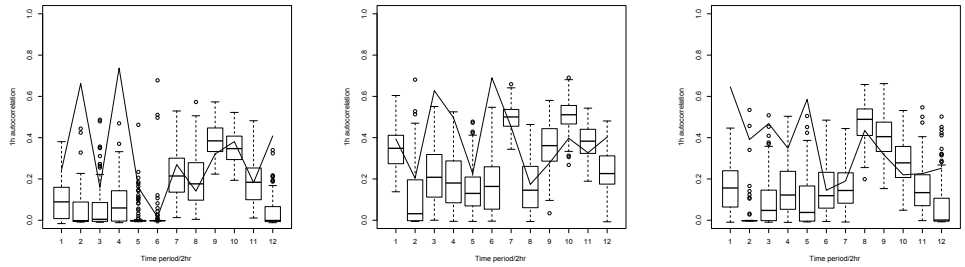


(j) October (k) November (l) December

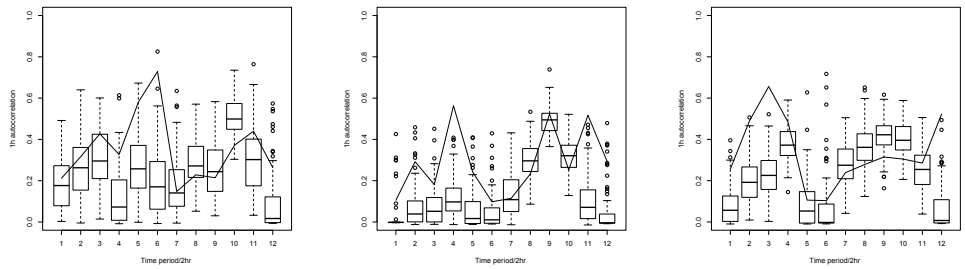
Figure 6.2: The variance of 2 hour blocks for every month



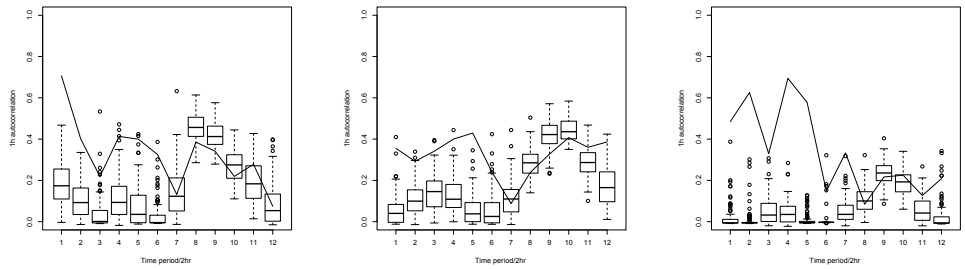
(a) January (b) February (c) March



(d) April (e) May (f) June



(g) July (h) August (i) September



(j) October (k) November (l) December

Figure 6.3: The autocorrelation of 2 hour blocks for every month

6.4.3 Extreme values

Figure 6.4 presents the annual maxima plots at four different aggregation time scales, 1, 2, 6 and 24 hours. In each plot, the 36 observed annual maxima from the 36 year of sampling period are ordered and then plotted against reduced Gumbel variate. The values from the historical series are shown as solid line. The annual maxima calculated from 100 simulation samples are shown as dashed lines.

In Chapter 5, both X-PWN and Markov X-PWN do not fit extreme values sufficiently. Both models underestimated the extreme values at all four aggregation time scales. X-NSWN has similarly poor performance. The model does not fit the historical values at the 1, 2, 6 and 24 hours aggregation levels. Since the X-NSWN simulations were not able to preserve the mean, variance and covariance that were used in the parameter estimation procedures, therefore it is to be expected that the model would not give a good results in simulating the extreme values.

6.4.4 Monthly moments

The hourly mean, variance and autocorrelation are calculated from the simulation data for each months and plotted against the historical values. The plots are in Figure 6.5. The solid line represents the historical values and the box plot is the simulation values. All the plots show the model underestimated the hourly mean, variance and autocorrelation. Again, this results is consistent with the trends of the model being unable to reproduce the statistical properties reasonably.

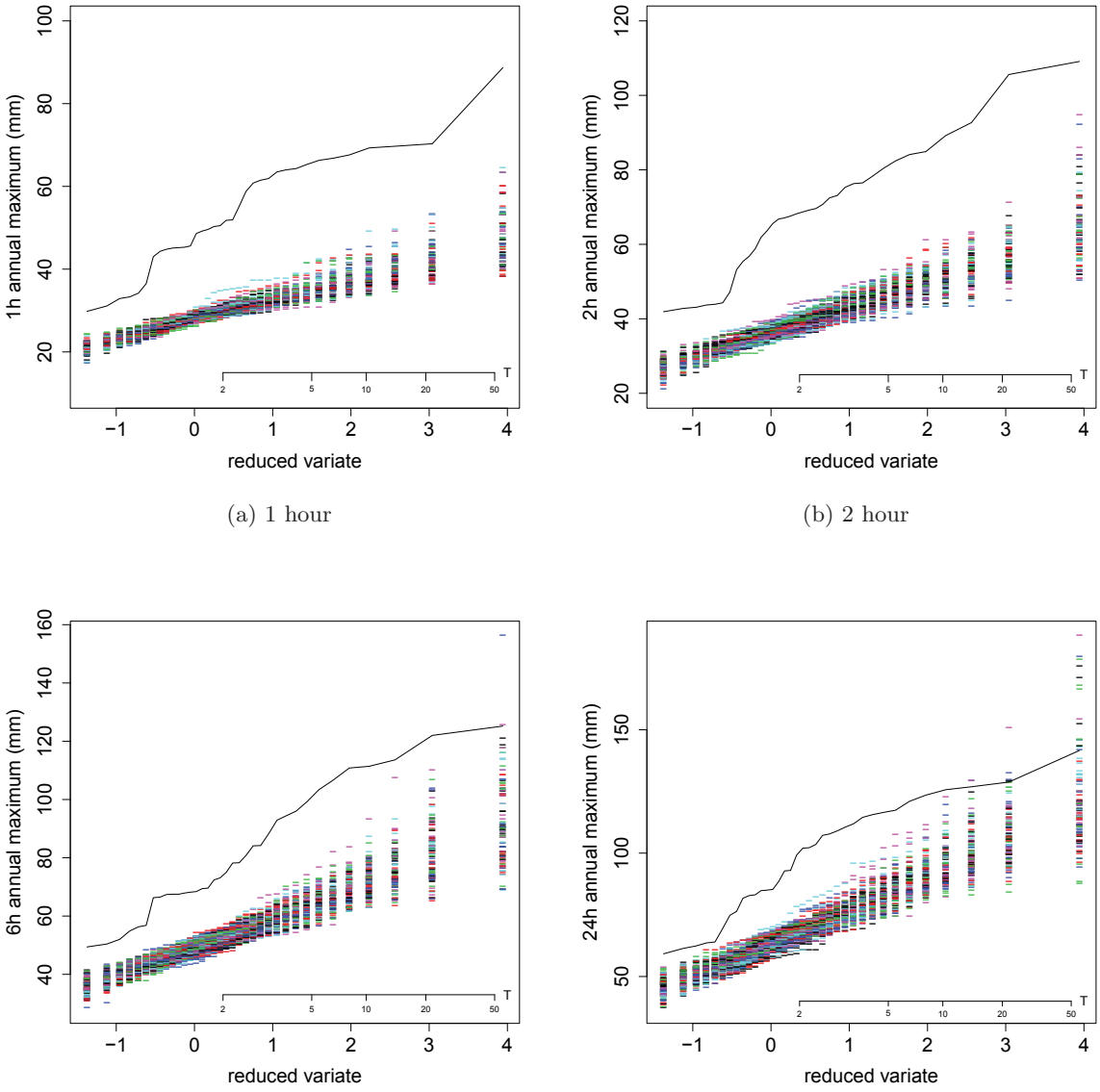
6.5 Markov X-NSWN model specification & properties

The X-NSWN model assumed independence between the hourly data in the two hours block, then the Markov X-NSWN model assumed dependence. In the simulations, the dependency allows the binary rainfall indicator variable, α to obtain the rainfall data in the first hour and continue in the second hour, thus capturing all the rainfall in the block.

The mean and variance are the same as in the X-NSWN model. To calculate the covariance, we have Y^1 and Y^2 as the NSWN process in the first and second hour of the block and X^1 and X^2 as the indicator variables for the first and second hour in a block, which are assumed dependent. The covariance is

$$\begin{aligned} Cov(X_1^1 Y^1, X_i^2 Y^2) &= \left(\alpha_i^2 + \alpha_i(1-\alpha_i)\left[\frac{\gamma_i - \alpha_i}{1-\alpha_i}\right]^\tau\right)[Cov(Y^1, Y^2) + E^2(Y)] \\ &\quad -\alpha_i^2 E^2(Y) \quad i = 1, 2, ..., 12 \end{aligned} \tag{6.7}$$

The covariance for the NSWN process is from (6.3).



(a) 1 hour

(b) 2 hour

(c) 6 hour

(d) 24 hour

Figure 6.4: Annual extreme values for Markov X-PWN

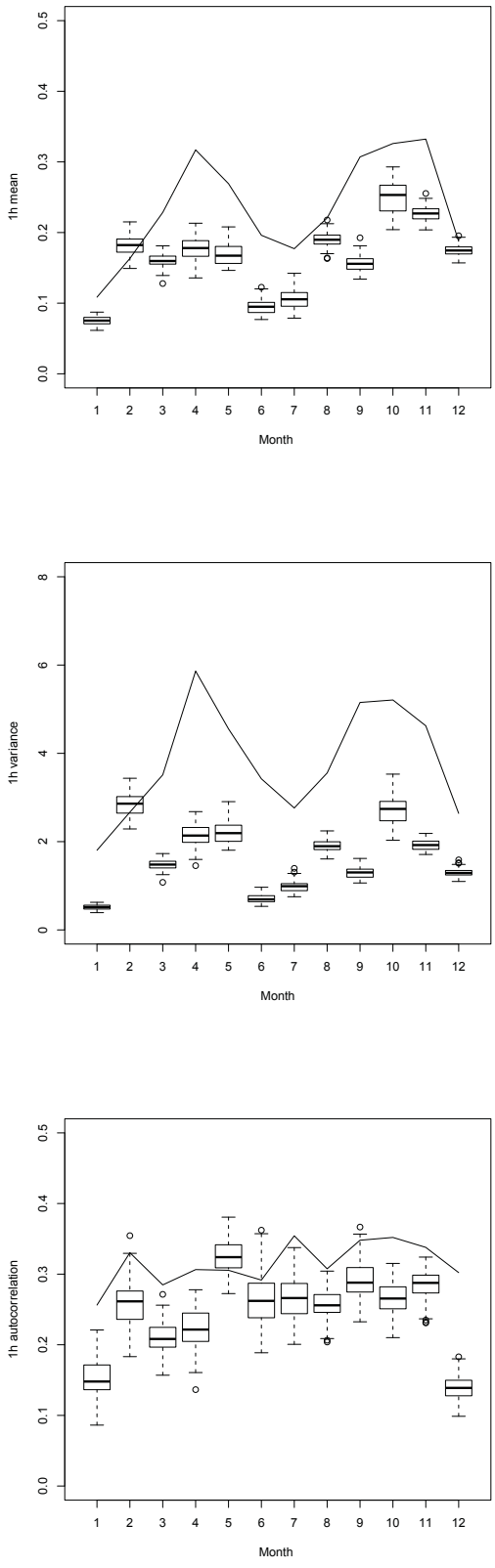


Figure 6.5: The hourly mean, variance and autocorrelation for every month

6.6 Fitting procedure

The fitting procedure is the same as for the X-NSWN process. Sample estimates used in the fitting procedure are the mean, variance and autocorrelation. The model is fitted to the hourly data in two hour blocks. The parameters for the Markov X-NSWN model are estimated simultaneously but separately for every month and each two hour block.

For simulation, the Markov process determines the wet or dry hours in each of the blocks. If it is a wet period, then the rainfall depths are produced using the NSWN parameters.

6.7 Analysis

6.7.1 Parameter estimates

The parameters of the Markov X-NSWN model are estimated for each month with 12 different α and γ, with 12 represent the total number of two hour blocks in a month. The NSWN process is assumed the same throughout the 12 blocks for each month. There are 28 parameters for each month.

The NSWN parameters are listed by column for each month. The resulting estimates are given in Table 6.3. Since the Markov X-NSWN model can be related to the physical process of rainfall, the estimates should reflect that intended physical meaning. But the estimation process was done separately by months, and therefore comparing the estimated values between the months does not seem possible. λ for January is the smallest, which is consistent with it being the driest month, but λ for November is not the largest despite it being the wettest month. The number of pulses is bigger compared to the X-NSWN model. This shows that the Markov X-NSWN model is representing more instantaneous rainfall.

The parameter estimates for α are in Table 6.4. The estimates correspond to the diurnal pattern of the rainfall. Block 9 has the maximum rainfall for all of the months and α_9 reflects this, except for September where it is slightly underestimated. Table 6.5 shows that the γ estimates have no patterns across the months and within the blocks. It does not imitate the diurnal pattern as α and different hours of the day have highest or lowest dependency throughout the months.

Table 6.3: Parameter estimates for the pulse depth. The units are hour^{-1} for λ and β with η the unit is mm and ν for number of pulses.

Month	λ	β	ν	η
1	0.0184	0.1477	117.1484	0.9552
2	0.0990	0.5297	202.7804	0.2107
3	0.0368	0.1295	170.2235	0.6184
4	0.0350	0.1304	148.8782	1.0824
5	0.0270	0.1230	146.5259	0.9314
6	0.0643	0.2219	170.7501	0.3925
7	0.0479	0.1777	212.3239	0.3764
8	0.0854	0.2272	71.1119	0.7863
9	0.0658	0.2084	113.8338	0.6000
10	0.0428	0.1035	163.257	0.8182
11	0.0799	0.1170	95.8072	0.8111
12	0.0829	0.2140	96.3639	0.4558

6.7.2 Moments

The derived statistical properties mean, variance and autocorrelation of the Markov X-NSWN model are plotted against the historical moments to check the model fits. The comparison are presented in Figure 6.6 to Figure 6.8. Again, the solid line represent the historical values and the box plots represent the simulation values.

In Figure 6.6, the mean of the two hours are plotted for the 12 months. There is a very good fit for the month April. Even for February, with morning and afternoon maximum rainfall, the model was able to adequately capture the mean in all of the blocks with slightly underestimated mean in block 11. Blocks with minimum rainfalls are well fitted for all of the months. But for blocks 9 and 10, the Markov X-NSWN model overestimates the historical values.

Figure 6.7 shows the variance of two hours for each month. Only for the month April does the model fit the historical value well. Again, the derived variance for blocks with minimum rainfall have a very good fit with the historical values, but the variance is underestimated in the other blocks.

Figure 6.8 shows that Markov X-NSWN model fits the autocorrelation for all of the months. A few of the blocks have underestimates, but the historical values are still within the box-plot.

Table 6.4: Parameter estimates of alpha. The units are hour^{-1}

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
α_1	0.0366	0.0289	0.0298	0.0337	0.0409	0.0247	0.0237	0.0044	0.0139	0.0498	0.01272	0.0120
α_2	0.0238	0.0415	0.0157	0.0219	0.0093	0.0047	0.0301	0.0216	0.0300	0.0298	0.0266	0.0116
α_3	0.0103	0.0555	0.0132	0.0205	0.0209	0.0163	0.0318	0.0203	0.0337	0.01417	0.0293	0.0330
α_4	0.0324	0.0341	0.0178	0.0266	0.0173	0.0222	0.0127	0.0299	0.0592	0.03201	0.0255	0.0394
α_5	0.0352	0.0189	0.0115	0.0107	0.017	0.0132	0.0266	0.0191	0.01448	0.0199	0.0152	0.0101
α_6	0.0089	0.0075	0.0109	0.0091	0.0189	0.0282	0.0173	0.0143	0.0093	0.01175	0.0108	0.0060
α_7	0.0375	0.0139	0.0206	0.0647	0.0714	0.0347	0.0344	0.0473	0.0621	0.0410	0.0374	0.0396
α_8	0.0848	0.0369	0.1271	0.1436	0.2404	0.0819	0.1350	0.1162	0.2432	0.1613	0.1016	0.1405
α_9	0.1781	0.0933	0.2166	0.1639	0.2608	0.2391	0.1741	0.1770	0.2426	0.1704	0.1659	0.2044
α_{10}	0.1382	0.0911	0.1748	0.1082	0.1455	0.0921	0.0916	0.1107	0.1240	0.1075	0.1476	0.1297
α_{11}	0.0619	0.0154	0.0954	0.0568	0.064049	0.0253	0.0285	0.0249	0.0396	0.0471	0.0626	0.0356
α_{12}	0.0149	0.0168	0.0337	0.0228	0.0261	0.0100	0.0100	0.0126	0.0113	0.02927	0.0301	0.0145

Table 6.5: Parameter estimates of gamma. The units are hour^{-1}

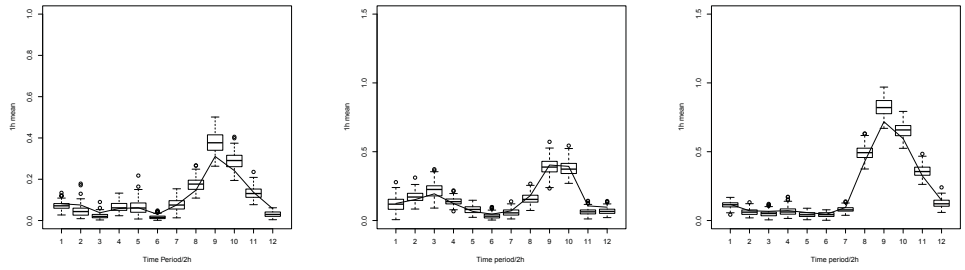
	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
γ_1	0.3181	0.8829	0.2659	0.3015	0.4864	0.7759	0.2519	0.1324	0.3123	0.8400	0.4232	0.5849
γ_2	0.8733	0.4522	0.3132	0.7904	0.2451	0.4331	0.3815	0.3720	0.5930	0.4770	0.3492	0.7523
γ_3	0.2726	0.6252	0.5231	0.1969	0.7782	0.5222	0.5001	0.2332	0.772	0.2599	0.4195	0.4059
γ_4	0.2625	0.1161	0.7275	0.8932	0.6147	0.3996	0.3658	0.7021	0.5941	0.4858	0.4821	0.8376
γ_5	0.8622	0.1483	0.1695	0.1933	0.2781	0.7013	0.6299	0.3031	0.1310	0.4758	0.5139	0.6953
γ_6	0.4149	0.2368	0.2830	0.0220	0.8515	0.1703	0.8429	0.1256	0.1268	0.3834	0.28912	0.1798
γ_7	0.7214	0.6199	0.3016	0.3469	0.5603	0.2192	0.1827	0.1617	0.3026	0.1708	0.1242	0.4124
γ_8	0.2099	0.3802	0.4361	0.2163	0.2803	0.5143	0.2999	0.3329	0.4070	0.4960	0.3283	0.1640
γ_9	0.5579	0.4945	0.2943	0.4316	0.4022	0.4464	0.2924	0.6837	0.4442	0.4561	0.4552	0.3379
γ_{10}	0.1137	0.3941	0.3163	0.4874	0.5216	0.2759	0.4483	0.3492	0.3993	0.2940	0.5258	0.3188
γ_{11}	0.2118	0.1816	0.3860	0.2348	0.4196	0.2702	0.5132	0.6484	0.3507	0.3396	0.4526	0.1676
γ_{12}	0.3078	0.1299	0.2016	0.4972	0.4975	0.2827	0.3017	0.3669	0.6201	0.0997	0.4667	0.2559

In general, the Markov X-NSWN model performed very well in preserving the observed mean and variance of rainfalls in blocks 1-8 and blocks 11-12 for all of the months. The model has also managed to reproduce the autocorrelation successfully for all of the blocks for every months. Comparing the total mean square errors of the X-NSWN model with the Markov X-NSWN model in Table 6.6, the latter model has the smallest value. This shows that the assumption of dependence between the hours in the two hours block is more relevant for the Malaysian rainfall data.

Measured by the mean square errors, Markov X-PWN has the smallest error overall. The X-NSWN model performed the worst compared to the other models. Therefore, the independence assumption within the hours in the block cannot be justified. The variance is better fit using the Markov X-PWN than the Markov X-NSWN. Only the autocorrelation from the Markov X-NSWN has an improved fit based upon the MSE value. Even though the number of parameters increases for the NSWN models, both models do not perform better as expected.

Table 6.6: Model comparison in terms of MSE

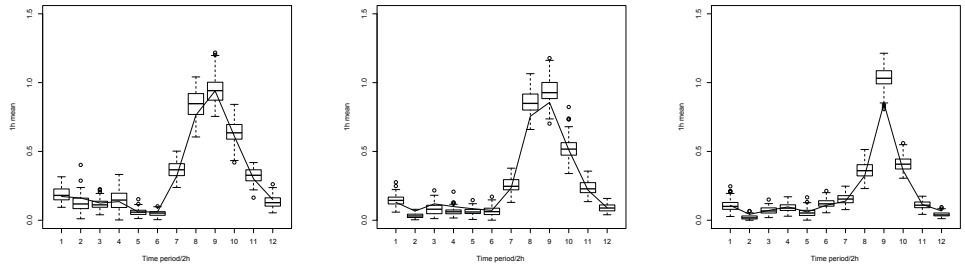
Model	Parameters (per month)	MSE Mean	MSE Variance	MSE Autocorrelation
X-PWN	14	0.0021	2.8249	-
Markov X-PWN	26	0.0007	2.3693	0.00065
X-NSWN	16	0.0225	14.5496	0.0617
Markov X-NSWN	28	0.0022	6.0011	0.00004



(a) January

(b) February

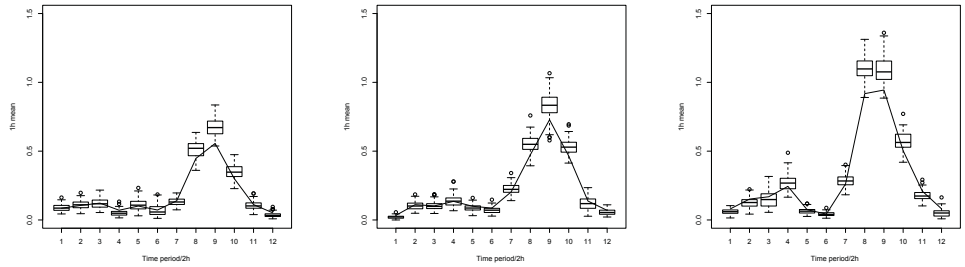
(c) March



(d) April

(e) May

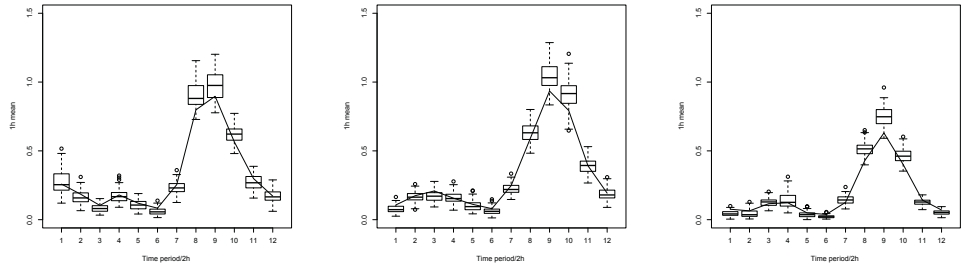
(f) June



(g) July

(h) August

(i) September

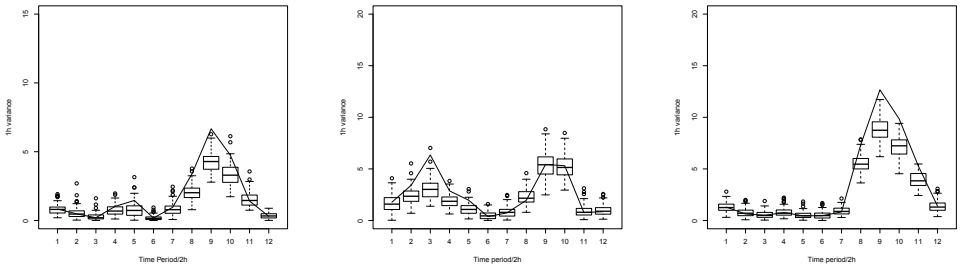


(j) October

(k) November

(l) December

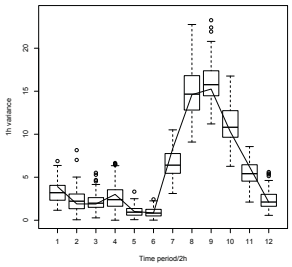
Figure 6.6: The mean of 2 hour blocks for every month



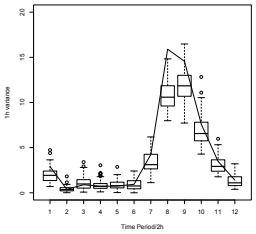
(a) January

(b) February

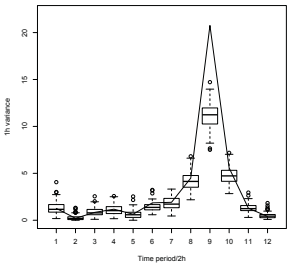
(c) March



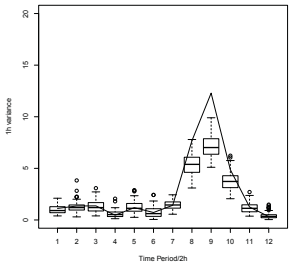
(d) April



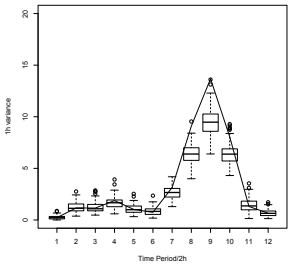
(e) May



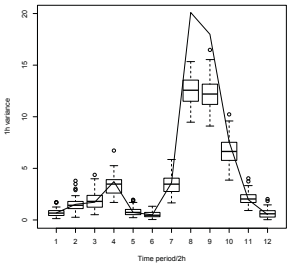
(f) June



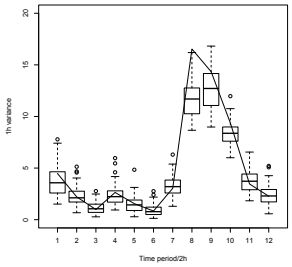
(g) July



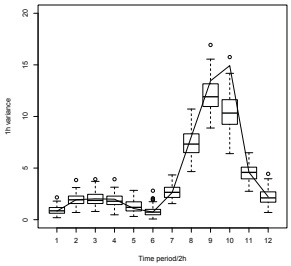
(h) August



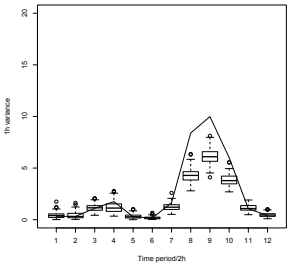
(i) September



(j) October

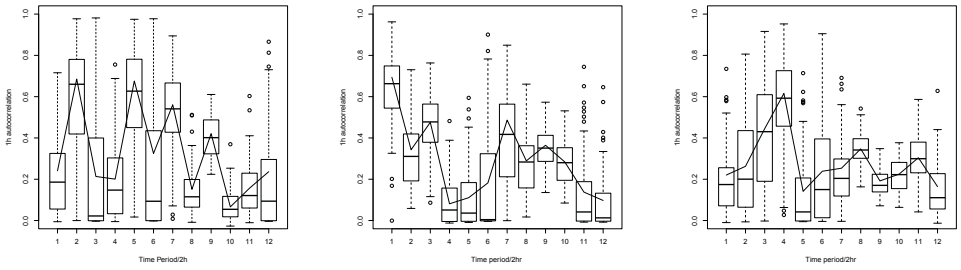


(k) November

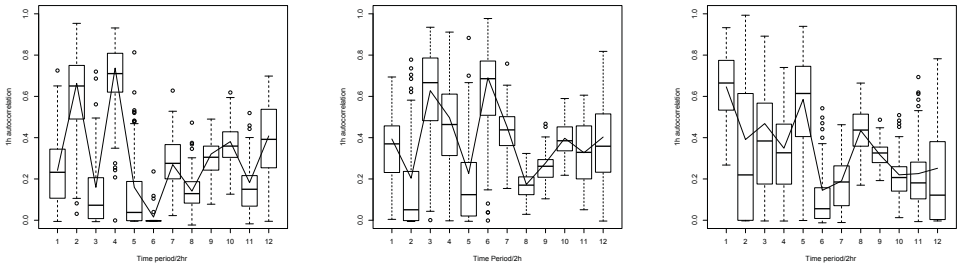


(l) December

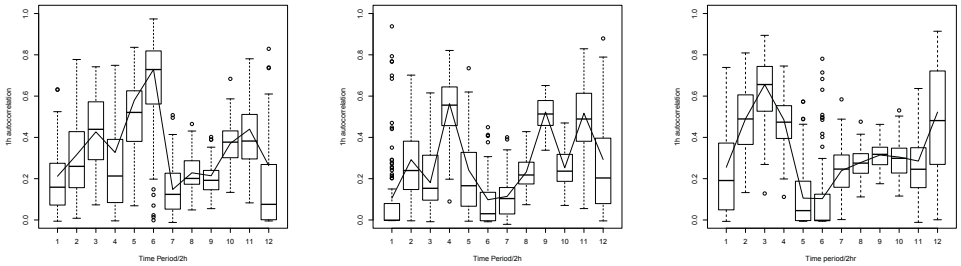
Figure 6.7: The variance of 2 hour blocks for every month



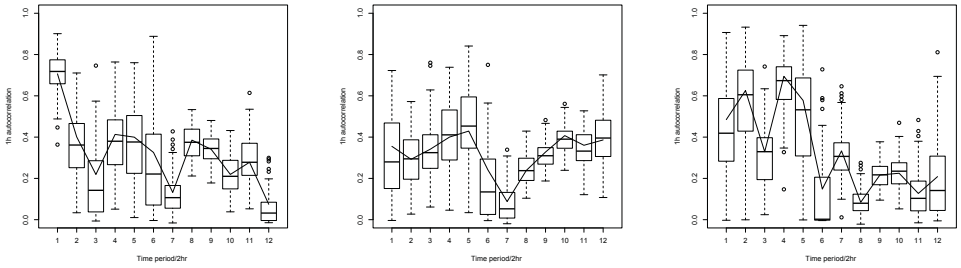
(a) January (b) February (c) March



(d) April (e) May (f) June

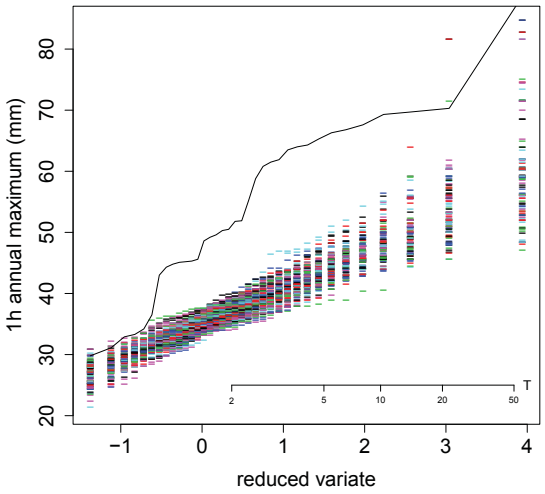


(g) July (h) August (i) September

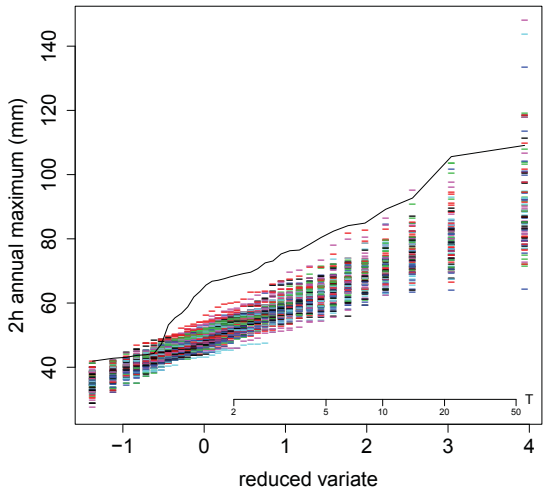


(j) October (k) November (l) December

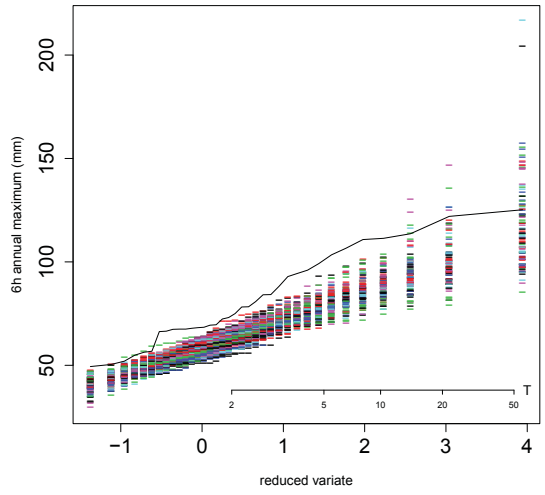
Figure 6.8: The autocorrelation of 2 hour blocks for every month



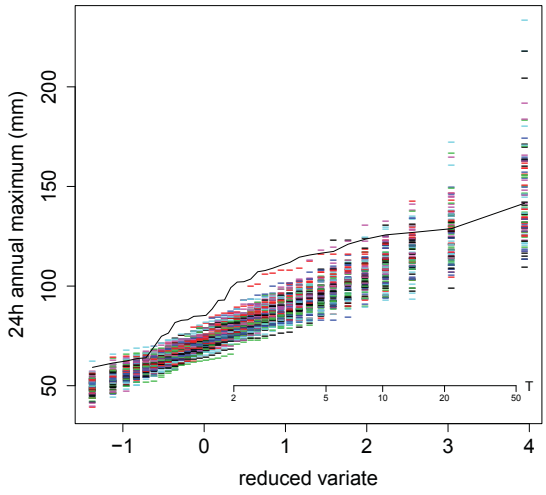
(a) 1 hour



(b) 2 hour



(c) 6 hour



(d) 24 hour

Figure 6.9: Annual extreme values for Markov X-PWN

6.7.3 Extreme values

Figure 6.9 examines the annual maxima plots at four different aggregation time scales, 1, 2, 6 and 24 hours. The ordered annual extreme values are plotted against the reduced Gumbel variate. As before, the solid line represents the historical values, the dashed lines are the annual maximum calculated from the 100 simulated samples.

At all of the aggregation levels, the model underestimates the historical values. There is an improvement at the 2, 6, and 24 hour aggregation level compared with the X-NSWN, X-PWN and Markov X-PWN models. At those three levels of aggregation, the model only slightly underestimated the annual maxima.

6.7.4 Monthly moments

The 1 hour statistical properties are calculated for each month. The properties are plotted against the historical values. Figure 6.10 shows that there is a good fit to the hourly mean values. Even though the parameter estimation process only includes the two hourly blocks properties, nevertheless the monthly hourly mean is well estimated by the Markov X-NSWN model. For the fitting of variance, there is a slight underestimation of the historic variance values. The autocorrelation is not a good fit, being underestimated and also not following the historical distribution. This could be the reason for the lack-of-fit of the extreme values.

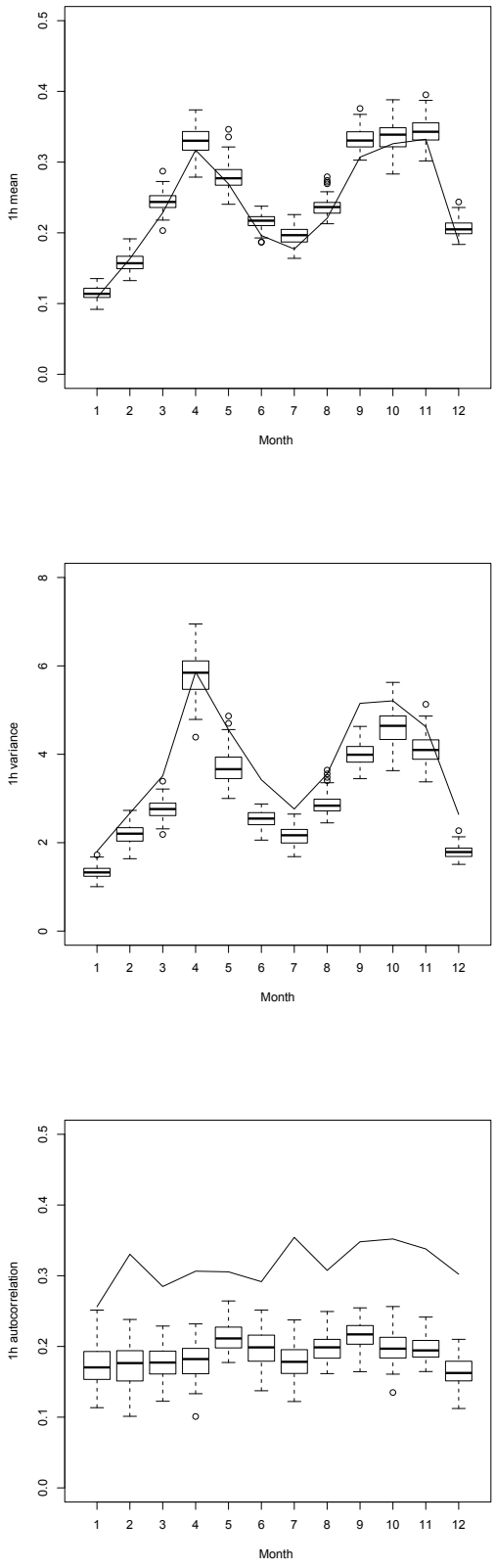


Figure 6.10: The hourly mean, variance and autocorrelation for every month

6.8 Summary and conclusion

The NSWN process provides a better mathematical representation of rainfall compared to the PWN model. Thus it was chosen as the main rainfall process to generate the rainfall depth over the month and to further investigate the function of the random variable X for the rainfall occurrence. The NSWN process was assumed the same throughout the whole month, allowing the two hours rainfall pattern to be shaped by the rainfall indicator parameter.

The two models developed in this chapter are the X-NSWM model and Markov X-NSWN model. Both model are compared based on the model properties and also with the results from the X-PWN model and the later with Markov X-PWN model. The comparison shows that assuming dependency between the hours in the two hour blocks, improved the fitting of the derived properties to the related historical values. The independence assumption was violated with a poor fit of covariance from the X-NSWN model.

The poor fitting of the X-NSWN model was also due to the estimated α values not being consistent with the two hours rainfall pattern. This was true for blocks 8 to 10. For the other blocks, with the corresponding estimates, there was a good fit for the mean, variance and covariance with the historical values. This emphasizes the importance of the parameter X in recognizing the diurnal pattern in the two hour rainfall data.

The NSWN process was explored in this chapter to improve the result from using the PWN process. But based on the mean square errors, the Markov X-PWN perform better than the X-NSWN and Markov X-NSWN model. Even though the Markov X-NSWN model has more parameters than the Markov X-PWN model, the model is not a good fit to the Malaysia diurnal rainfall data. However, both of the Markov models also do not fit well the properties for the blocks with the most rainfall. This shows that the NSWN or PWN process alone, might not be enough to represent the diurnal events in Malaysia. Since Malaysia has a convective and stratiform types of rainfall, a superposed model might be used to model the two types of rainfall together (Cowpertwait *et al.* 2007, Morrissey 2009). Therefore, the next step is to investigate a superposed Markov X-NSWN and Markov X-PWN process.

Chapter 7

Superposed model and 8-block model

7.1 Introduction

The main objective of the research is to integrate the parameters for capturing diurnal patterns in the Malaysia rainfall data with the point process rainfall models. This was achieved by introducing the rainfall occurrence binary variable and its associated (diurnal) parameters. In the last two chapters, it has been shown that by allowing the indicator parameter to be non-homogeneous for every two hour block, one can simulate the diurnal pattern of the Malaysia rainfall. Although the probability of the rainfall occurrence differs by the two hour blocks, the point process itself was assumed to be the same throughout the months. Two point processes were chosen from the white noise family, the PWN model and NSRP model. Both models assume instantaneous pulses of rainfall that seem to well describe the Malaysia rainfall data especially during the late afternoon when there is maximum rainfall.

The Markov X-PWN and Markov X-NSWN model tended overestimate the mean in time blocks 8 and 9. The variance was also not well estimated. Those results are discussed in Chapter 5 and 6. The research literature Morrissey (1993) and Morrissey (2009) has shown that the tropical rainfall consists of two types of rainfalls, the convective and stratiform. This implies that having only one point process model to represent all the rainfall events for Malaysia rainfall data may not be sufficient. Furthermore, both of the model parameters do not account well for the variability of the rainfall. The parameter λ improves the overall mean rainfall and parameter γ accounts for the autocorrelation of the rainfall. To improve the variability of the rainfall pattern, the models needs to add another parameter.

Results in the last two chapters shows that the Markov X-PWN and Markov X-NSWN model can reasonably describe the Malaysian diurnal pattern in the rainfall. We therefore superposed the models to add the variability of the rainfalls that comes from the two types of storm. The PWM models the convective rainfall, or showery precipitation that have shorter duration of light to heavy rainfall (Morrissey 1993, Morrissey 2009). The NSWN models the cloud cluster events of rainfall, the stratiform events. Combining both of these events will allow the different types of rainfall to be included in the model.

Later, the results will show that the superposed model improved the fitting of the variance. But there are too many parameters per month. To reduce the number of parameters, the two hour blocks are modified to three hour blocks, thus the name 8-block model. By changing the interval hour for the block, it still succeeds in preserving the diurnal variation in the rainfall data. In this chapter these two models will be discussed, the superposed model and the 8-block model.

7.2 Superposed Markov X-NSWN and PWN model specification & properties

The superposed Markov X-NSWN and PWN model is assuming the stratiform rainfall as the main event, producing large amounts of rainfall, and allowing the rainfall to spread out over the two hours period in a block. The PWN process is assumed as a secondary rainfall event that occur frequently throughout the day. The NSWN and PWN process can be assumed independent events because the trade wind showers do not contribute a significantly amount of rainfall to the total amount of rainfall produced by the cloud cluster system (Morrissey 2009). The possibilities of both events to happen at the same time and place are quite small.

To superpose the Markov X-NSWN and Markov X-PWN model would require too many parameters. So different combinations of the superposed models with different simplifying assumptions were considered (there is no testing here). One approach was to superpose NSWN and PWN processes and then add the Markov chain or superposed Markov X-PWN model to the NSWN. Results from this combinations did not perform well, therefore it will not be discussed further here. The last combination, the Markov X-NSWN with PWN model gave a better results compared with the others. Thus the superposed Markov X-NSWN and PWN model is developed here.

Let Y be the NSWN process with X_i $(i = 1, 2, ..., 12)$ the indicator variable for every two hours block (later, every three hours block) and Z the PWN process. The mean and variance for the superposed X-NSWN and PWN model are

$$E[X_iY + Z] = \alpha_i E(Y) + E[Z] \tag{7.1}$$

$$Var[X_iY + Z] = \alpha_i(Var[Y] + E^2[Y]) - \alpha_i^2 E^2[Y] + Var[Z^2] \tag{7.2}$$

Since there is no covariance for the PWN model, the covariance for the superposed model is

$$\begin{aligned} Cov(X_1^1Y^1, X_i^2Y^2) = & \left(\alpha_i^2 + \alpha_i(1-\alpha_i)\left[\frac{\gamma_i - \alpha_i}{1-\alpha_i}\right]\right)[Cov(Y^1, Y^2) + E^2(Y)] \\ & -\alpha_i^2 E^2(Y) \end{aligned} \tag{7.3}$$

$E[Z]$ is from (4.1), $Var[Z]$ is from (4.2) and $Cov(Y^1, Y^2)$ is from (6.3).

7.3 Fitting procedure & simulation

To estimate the parameters, the mean, variance and covariance for every two hour blocks are used in the method of moments. The parameters for the superposed model are estimated simultaneously for every month. For every month there are 30 parameters. They are estimated separately by month.

In simulation, the process starts with simulating the 36 years rainfall data by the NSWN process. Then, the probability of occurrence is used to reconstruct the hourly rainfall data for every block. The PWN process simulated another set of 36 years data, and was combined with the rainfall data from the Markov X-NSWN process. Figure 7.1 is an example of 1 month simulation of the different facets of the models and their superposition. The first graph shows hourly rainfall downpours from the Markov X-NSWN process. Although the rainfall events does not occur frequently these are the main events for the afternoon rainfall. The next graph shows the PWN process that is assumed as the trade wind showers of light rainfall, occur frequently throughout the day. The amount of rainfall in the hours with minimum rainfall is contributed by the PWN process. The last graph shows the total amount of rainfall from both event.

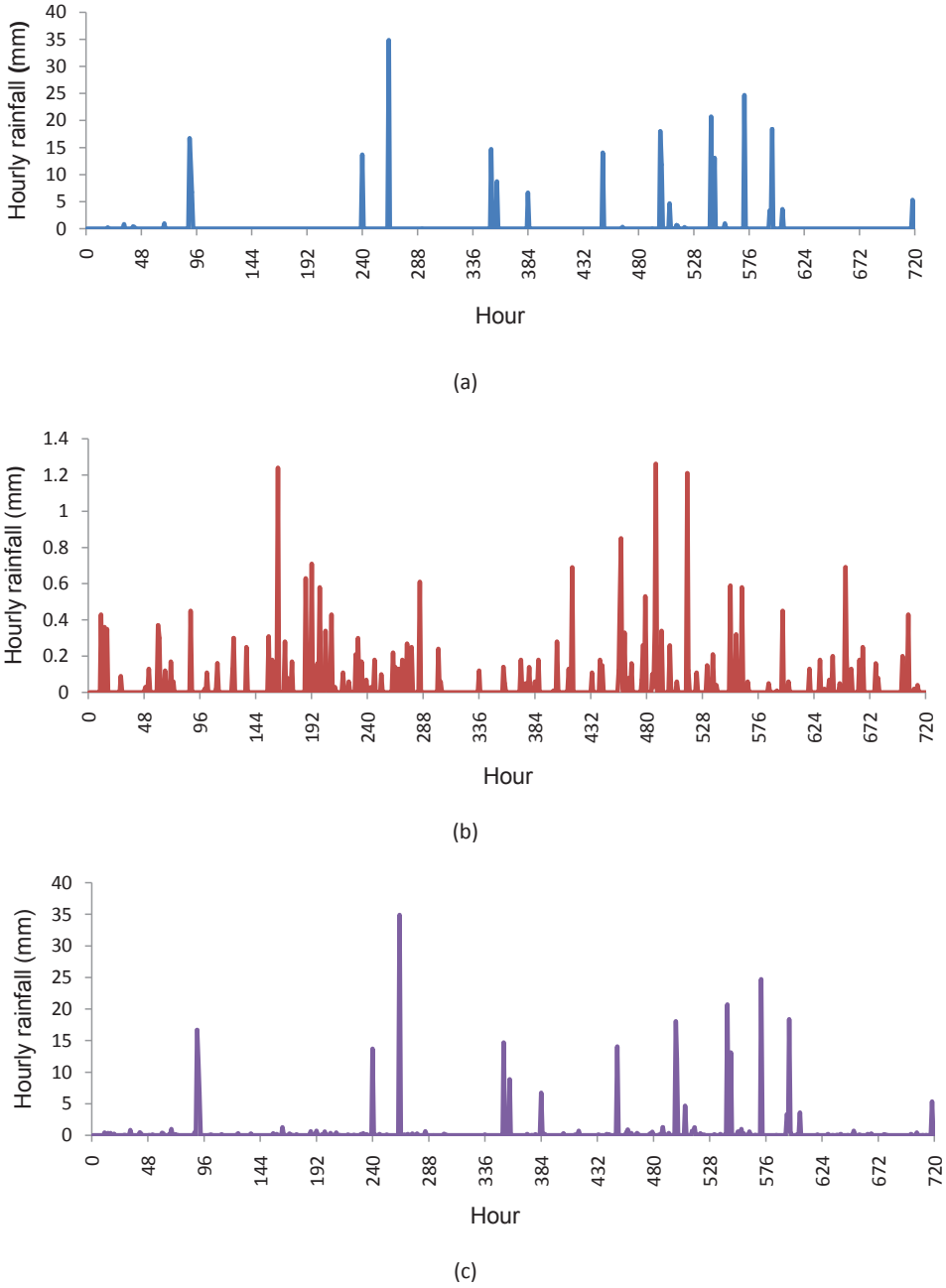


Figure 7.1: The hourly rainfall from the (a) Markov X-NSWN process (b) PWN process (c) superposed process

7.4 Analysis

7.4.1 Parameter estimates

The parameters from the NSWN model and PWN model (in Table 7.1), α (in Table 7.2) and γ (in Table 7.3). The first four parameters in Table 7.1 represent the NSWN process and the last two for the PWN process. Both types of storm contribute to the total rainfall depth. The parameter estimates are close to the estimates from the Markov X-NSWN model.

The α values are consistent with the trend of the diurnal pattern in the two hour blocks. The probability of having the most rainfall is in block 9. The estimated values are smaller than the values estimated from the Markov X-NSWN model. This implies that the total rainfalls is contributed to not only from the Markov X-NSWN process, but also from the PWN process. The γ values are random with no trends between the months.

Table 7.1: Parameter estimates for the pulse depth. The units are hour^{-1} for all estimates, except η the unit is mm and ν for number of pulses.

Month	λ	β	ν	η	λ_P	η_P
1	0.0024	0.2597	99.4215	1.6711	0.0367	1.2452
2	0.0349	0.1029	168.9217	1.1057	0.0319	0.9072
3	0.0289	0.4592	53.294	1.1823	0.1343	0.2372
4	0.0624	0.5841	162.6005	0.3873	0.0107	0.5312
5	0.0074	0.1693	80.5353	2.3365	0.1129	0.451185
6	0.0169	0.1681	89.5293	1.7376	0.0249	0.6878
7	0.0094	0.1204	169.4791	1.4014	0.0418	1.0279
8	0.0412	0.4531	183.4118	0.3858	0.0254	0.7638
9	0.0471	0.1386	140.0734	0.7772	0.0295	1.2427
10	0.0502	0.2870	174.0928	0.5444	0.1418	0.3085
11	0.0461	0.1867	143.9560	0.7125	0.0960	0.4801
12	0.0092	0.2124	179.1527	0.7621	0.0382	0.9692

7.4.2 Moments

The values of mean, variance and autocorrelation for the superposed model are estimated using the parameter estimates. The goodness of fit of the fitted model is tested by comparing the historical values with the derived values. The comparison is made by plotting the values. The results are in Figure 7.2, Figure 7.3 and Figure 7.4. In all figures, the historical values are represented by the solid line. The box plots are representing the

Table 7.2: Parameter estimates of alpha. The units are hour^{-1}

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
α_1	0.0873	0.0139	0.044427	0.0432	0.0734	0.0243	0.0232	0.0023	0.0099	0.0481	0.0114	0.0128
α_2	0.0499	0.0215	0.020637	0.0304	0.0113	0.0022	0.0298	0.0224	0.0212	0.0255	0.0268	0.0127
α_3	0.0161	0.0292	0.0167	0.027934	0.0408	0.0148	0.0318	0.0213	0.0259	0.0113	0.0306	0.0530
α_4	0.0863	0.0167	0.0239	0.0327	0.0291	0.0219	0.0101	0.0340	0.0458	0.0292	0.0252	0.0763
α_5	0.1063	0.0073	0.0145	0.0112	0.02362	0.0107	0.0267	0.0192	0.0099	0.0172	0.0143	0.0110
α_6	0.0130	0.0024	0.0129	0.01014	0.0265	0.0313	0.0153	0.0138	0.0052	0.0102	0.0089	0.0036
α_7	0.1021	0.0055	0.0293	0.087465	0.1377	0.0387	0.0349	0.0588	0.0505	0.0377	0.0394	0.0735
α_8	0.3280	0.0188	0.2272	0.189879	0.4725	0.0997	0.1816	0.1629	0.1801	0.1701	0.1155	0.3481
α_9	0.7301	0.0509	0.3530	0.24496	0.5530	0.3060	0.2566	0.2561	0.1881	0.1705	0.1915	0.4839
α_{10}	0.4955	0.0499	0.3171	0.1410	0.297925	0.1204	0.1154	0.1563	0.1014	0.1073	0.1523	0.2898
α_{11}	0.1657	0.0078	0.1660	0.0752	0.120429	0.0253	0.0291	0.0260	0.02887	0.0432	0.0694	0.0531
α_{12}	0.0284	0.0076	0.0521	0.028	0.039855	0.0106	0.0058	0.0113	0.0075	0.0255	0.0308	0.0162

Table 7.3: Parameter estimates of gamma. The units are hour^{-1}

.

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
γ_1	0.3412	0.8465	0.3395	0.3474	0.5730	0.8998	0.2884	0.2422	0.3011	0.8474	0.450085	0.7907
γ_2	0.9327	0.4207	0.4057	0.9587	0.3308	0.9802	0.4204	0.4115	0.5713	0.5010	0.364326	0.8999
γ_3	0.3768	0.5858	0.6888	0.2336	0.9357	0.7285	0.5729	0.2607	0.8396	0.272939	0.4234	0.4463
γ_4	0.2867	0.1087	0.9431	0.9993	0.7516	0.5172	0.487042	0.7985	0.5819	0.49554	0.4951	0.8944
γ_5	0.878	0.1453	0.2211	0.2403	0.3453	0.9437	0.7650	0.3441	0.1296	0.5011	0.5377	0.8996
γ_6	0.6049	0.2743	0.3729	0.0261	0.9956	0.214773	0.8871	0.1543	0.1293	0.390471	0.3102	0.3815
γ_7	0.7872	0.6468	0.3898	0.3955	0.6593	0.2727	0.1988	0.1653	0.2412	0.1700	0.119794	0.4365
γ_8	0.2181	0.3457	0.5457	0.2386	0.2893	0.6337	0.3089	0.3346	0.3901	0.4909	0.319247	0.1391
γ_9	0.5682	0.4491	0.3397	0.4851	0.4290	0.4783	0.2992	0.7274	0.4353	0.4458	0.44041	0.3078
γ_{10}	0.1003	0.3565	0.3778	0.5537	0.5839	0.3185	0.4835	0.3637	0.3771	0.2878	0.526039	0.3001
γ_{11}	0.2247	0.1809	0.4758	0.2709	0.4760	0.3363	0.5941	0.7205	0.3550	0.3441	0.450735	0.1772
γ_{12}	0.3796	0.1254	0.255491	0.5857	0.5971	0.4485	0.438205	0.4497	0.367903	0.0951	0.4776	0.3235

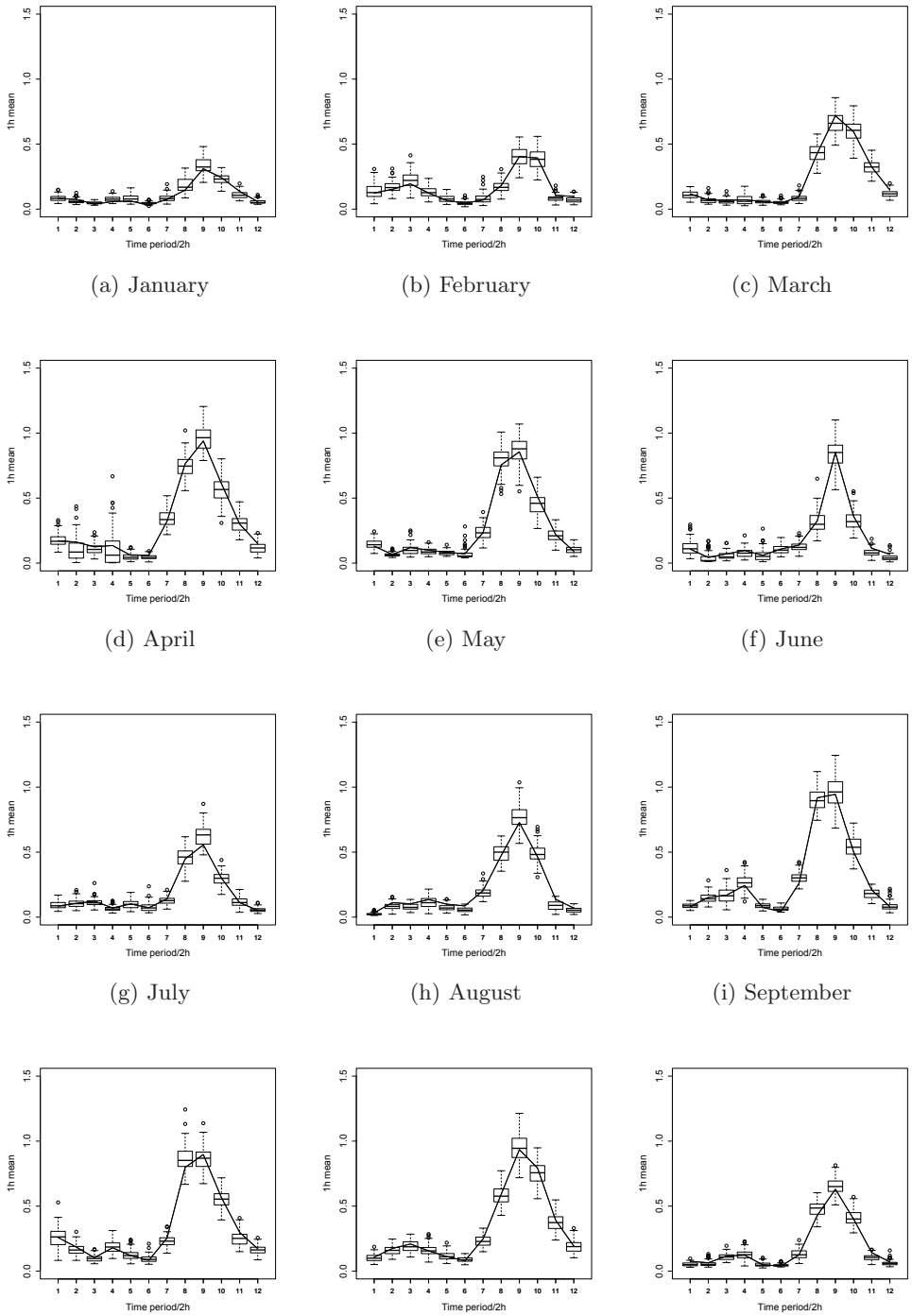
distribution of the simulated values.

Figure 7.2 shows the plot for mean of the two hours blocks for the 12 months. Overall, the model fits the mean, especially for block 9. Except for September, which is slightly overestimated, the mean rainfall has a good fit for block 9. In the Markov X-PWN model and Markov X-NSWN model, both models overestimated the mean for that block. For the other blocks, the results are the same. The variance plot in Figure 7.3 show an improvement compared to the result from the Markov X-PWN model and Markov X-NSWN model. Again, the superposed model is able to derive the variance in block 9 with a good fit in the majority of the months. But the improvement of the derived mean and variance causes the autocorrelation to be slightly underestimated for some of the blocks (for example in block 2 for month June or block 6 in May). Overall, the superposed model can fit the autocorrelation.

The MSE values in Table 7.4 present the numerical method of comparison. The MSE for the variance decreases dramatically from 6 in the Markov X-NSWN model to 0.8625 in the superposed model. It is also smaller than the MSE from Markov X-PWN model. By increasing the number of parameters in the model, the superposed Markov X-NSWN and PWN model improved the fitting for variance. There is a slightly worse fit to autocorrelation but overall the MSE value shows that the superposed model outperforms the Markov X-PWN model and Markov X-NSWN model.

Table 7.4: Model comparison in terms of MSE

Model	Parameters (per month)	MSE Mean	MSE Variance	MSE Autocorrelation
Markov X-PWN	26	0.0007	2.3693	0.00065
Markov X-NSWN	28	0.0022	6.0011	0.00004
Superposed model	30	0.0192	0.8625	0.00077



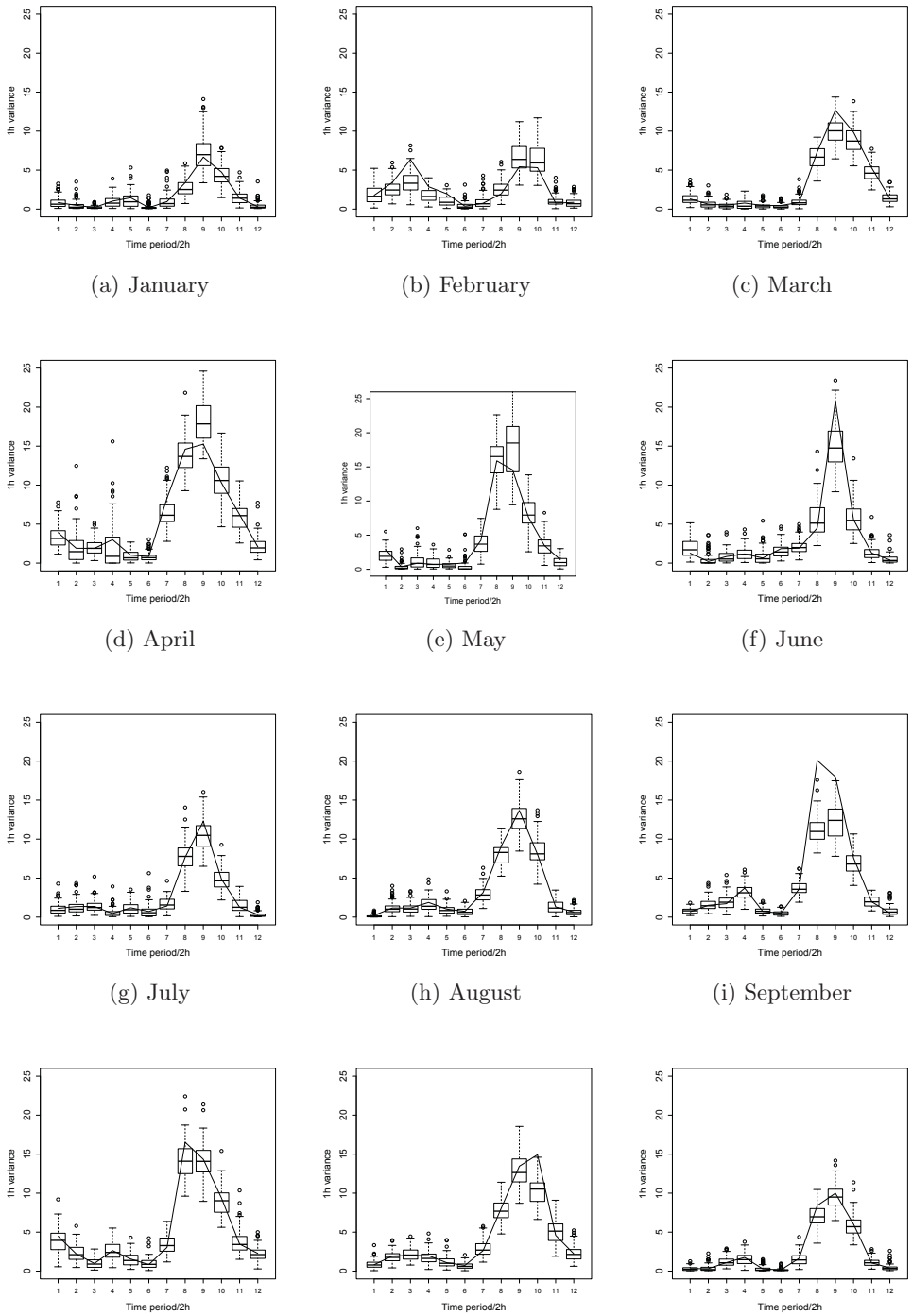
(a) January (b) February (c) March

(d) April (e) May (f) June

(g) July (h) August (i) September

(j) October (k) November (l) December

Figure 7.2: The mean of 2 hour blocks for every month



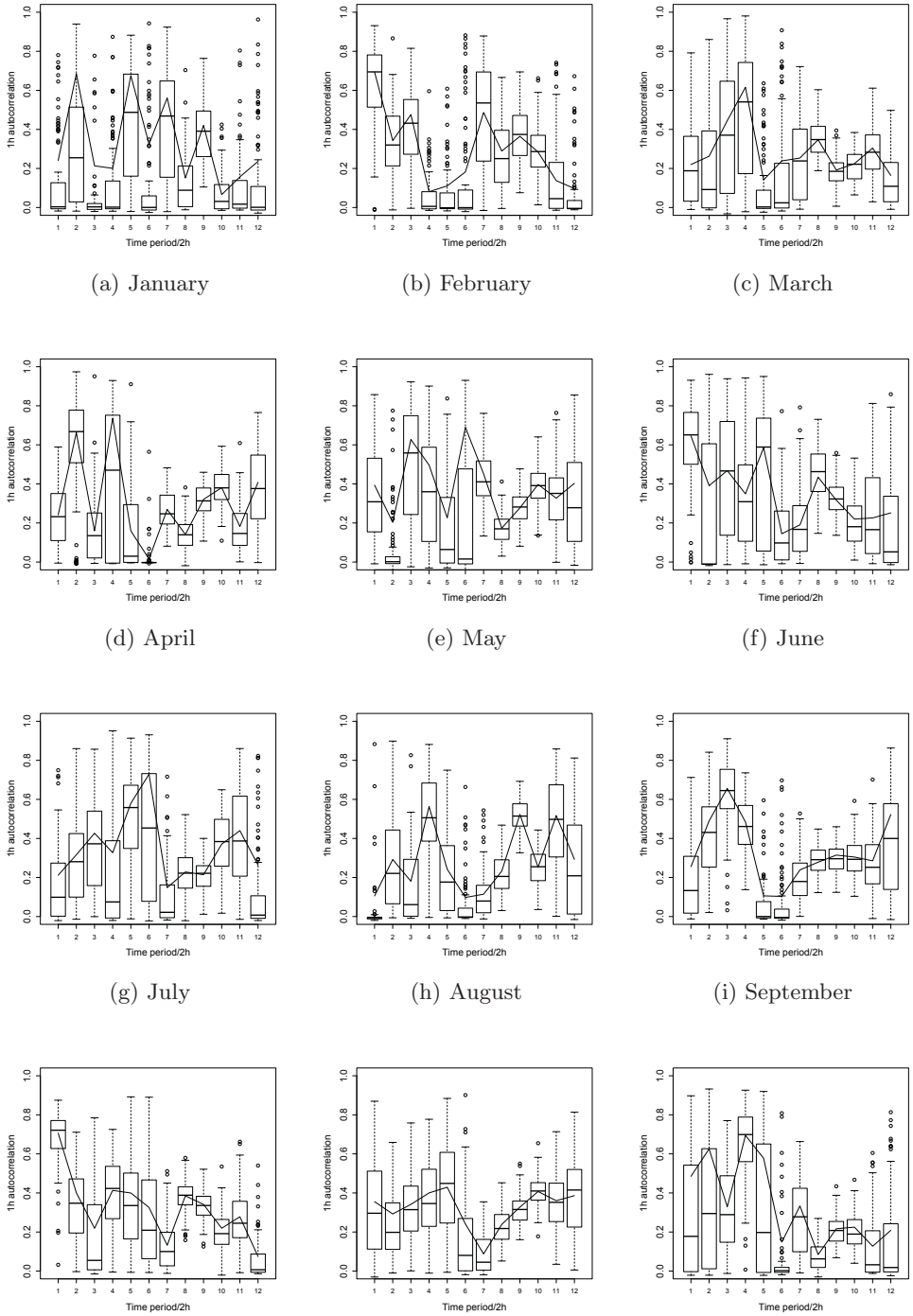
(a) January (b) February (c) March

(d) April (e) May (f) June

(g) July (h) August (i) September

(j) October (k) November (l) December

Figure 7.3: The variance of 2 hour blocks for every month



(a) January (b) February (c) March

(d) April (e) May (f) June

(g) July (h) August (i) September

(j) October (k) November (l) December

Figure 7.4: The autocorrelation of 2 hour blocks for every month

7.4.3 Extreme values

Figure 7.5 examines the annual extreme values pattern at four different aggregation time scales, 1, 2, 6 and 24 hour. The values from the historical data are plotted as a solid line. The simulation annual maximum are plotted as dash lines. The ordered maxima are plotted against the reduce Gumbel variate.

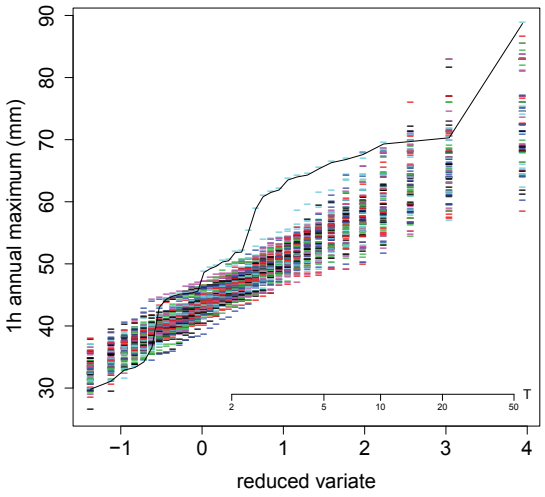
The 1 hour annual maximum plot shows that for return period less than 10, the superposed model slightly underfits the historical values. For over the range of $T > 10$, the superposed model has an adequate fit. For the 2 hour level, the model fits the historical values well for the range of $T > 5$. But does not fit well in other areas.

For the 6 hour plot, the model performs well for all of the return periods but slightly overestimates the maximum for $T < 3$ and $T > 10$. This also shows in the 24 hour plot. When the return period is over the range of $T > 10$, the superposed model overfits the historical values. But for $T < 3$, the model underfits the historical values.

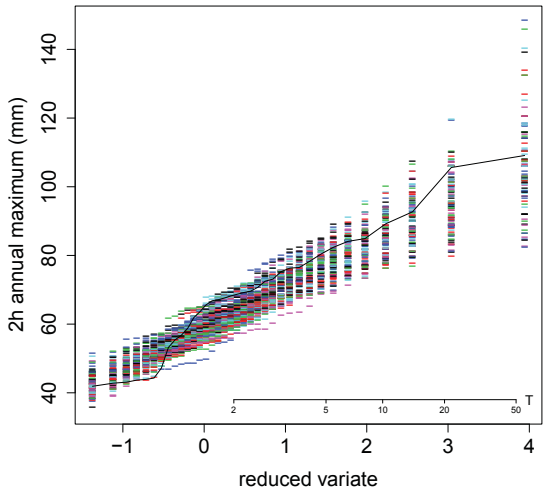
Overall, we conclude that the superposed model has a better performance in fitting the historical extreme values compared to the Markov X-PWN model and Markov X-NSWN model.

7.4.4 Monthly moments

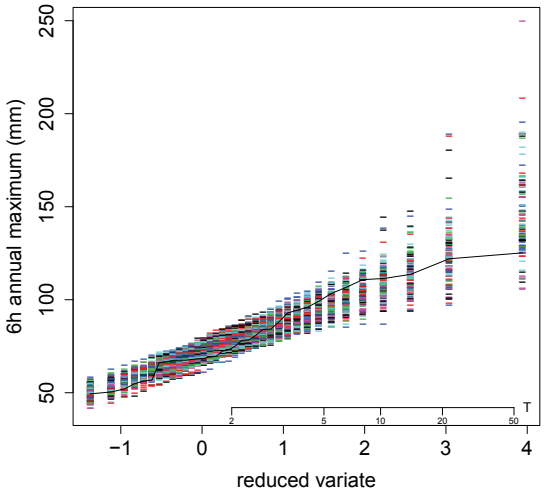
The 1 hour statistical properties are calculated for each month. The properties are plotted against the historical values. Figure 7.6 shows that a good fit for the hourly mean values. Even though the parameter estimates process only includes the hourly block properties, the monthly hourly mean is well estimated by the Markov X-NSWN model. For the fitting of the variance, there is a slight underestimation of the historic values. The autocorrelation is not a good fit, being underestimated and also not following the historical distribution. This could be the reason for the lack-of-fit of the extreme values.



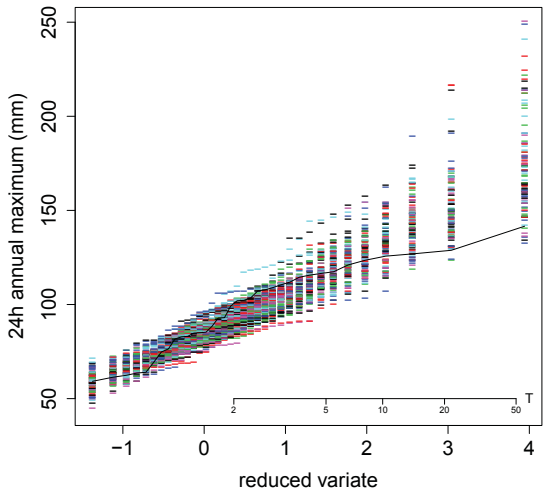
(a) 1 hour



(b) 2 hour



(c) 6 hour



(d) 24 hour

Figure 7.5: Annual extreme values for Markov X-PWN

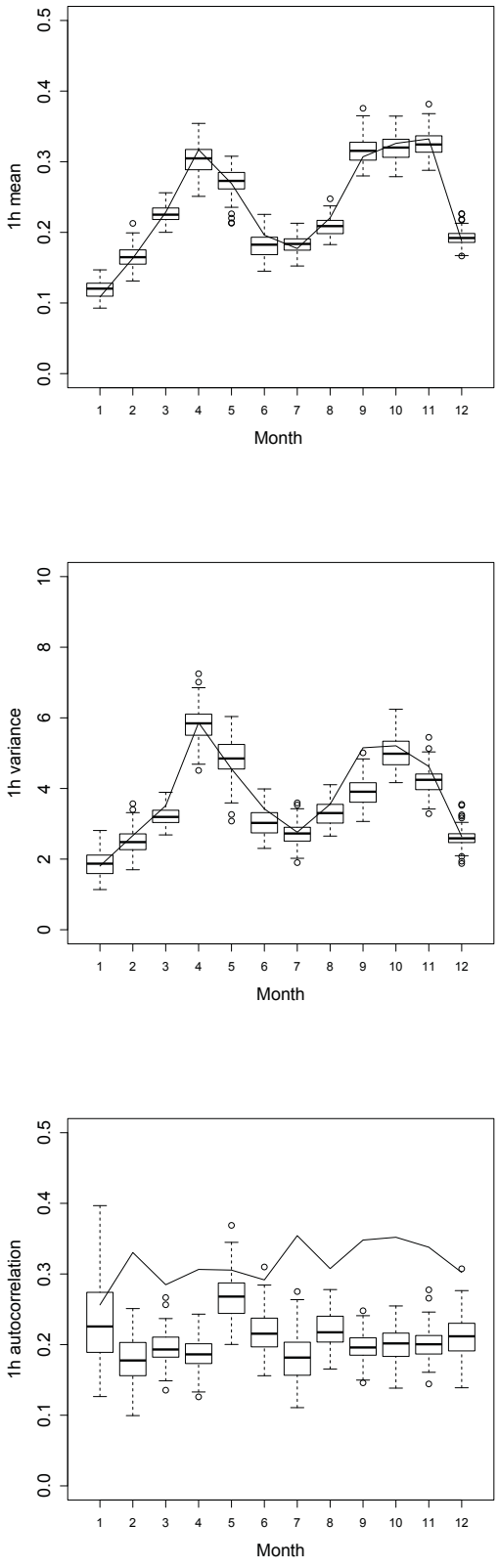


Figure 7.6: The hourly mean, variance and autocorrelation for every month

7.5 8-block model specification & properties

The superposed Markov X-NSWN and PWN model significantly improved the fitting of the variance. Overall results show that the superposed model outperforms the other models. But the superposed model has the most parameters per month with 30 parameters. There are many parameters because the model is constructed for two hour blocks. Since α and γ represent the two hour block for the diurnal pattern of the daily rainfall, it is not possible to combine the 12 different values for each of the parameters.

One approach to reduce the number of parameters, is to change the 2 hour blocks to 3 hour blocks. This will still allow α and γ to represent the diurnal pattern as in Figure 7.7. The NSWN and PWN processes remain the same but the pulse depth is assumed to be the same for both of the models. Therefore, the total parameters for the 8 block model is 21 parameters per month. The properties of the model are the same as the superposed Markov X-NSWN model in 7.1, 7.2 and 7.3.

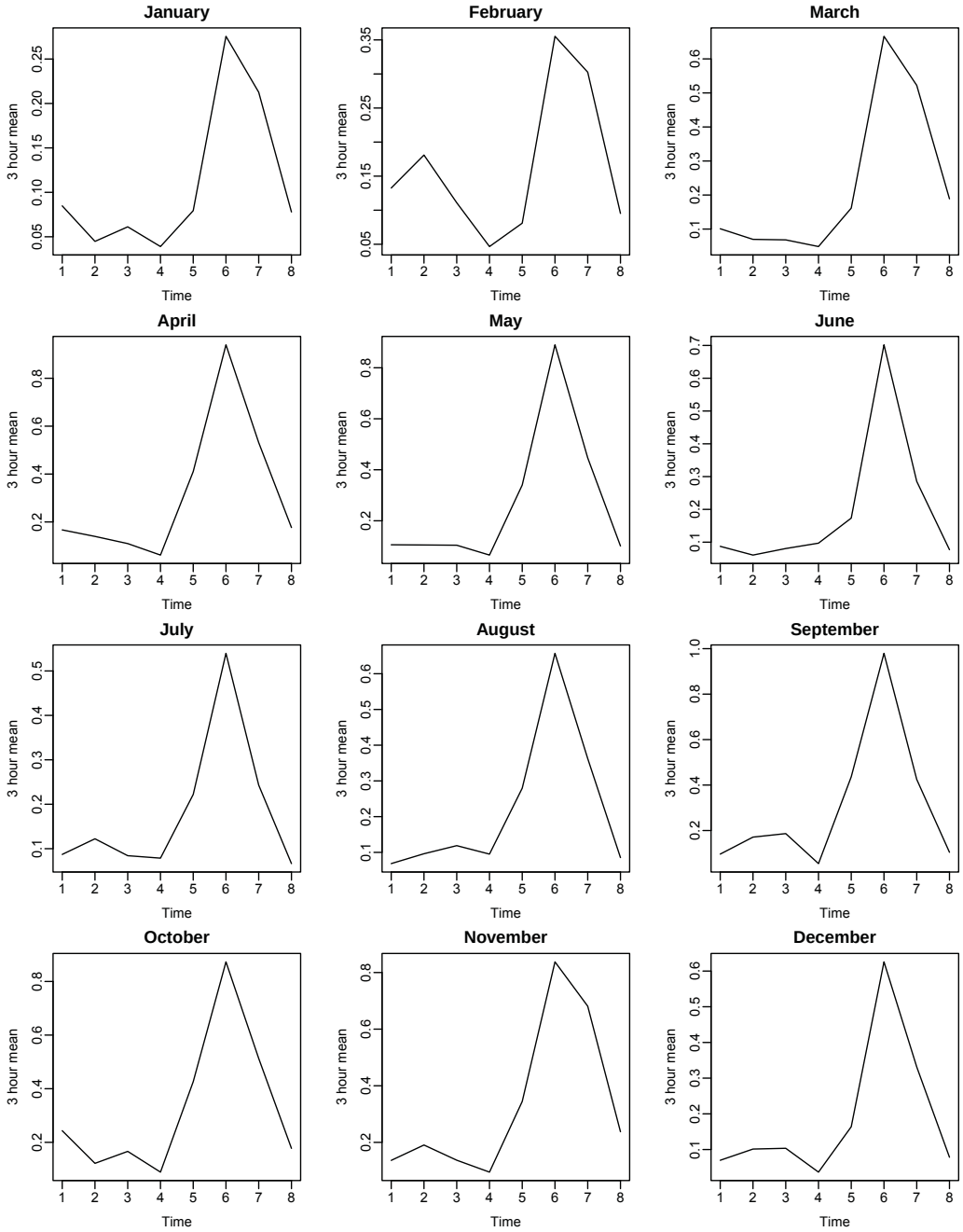


Figure 7.7: The hourly mean of 3 hour blocks for every month

7.6 Fitting procedure

The model is fitted to the hourly data in the 3 hour block for every month. The fitting procedure is the same as the superposed model. The 21 parameters are estimated for each month separately.

In the simulation, the parameter estimates are used to simulate the 36 years data starting with the Markov process. Then the PWN parameter estimates are used to simulate the PWN process. Both data are combined to give the total rainfall in the 3 hour block. A total of 100 simulated data sets are produced.

7.7 Analysis

7.7.1 Parameter estimates

The NSWN and PWN parameters are listed in columns in Table 7.5. The first four columns are the NSWN parameters and the last two columns are the PWN parameters. η and η_P represents the rainfall pulse depth. Since it is assumed the same for both of the models, the depth values are the same. This assumption allowed the different types of rainfall to be classified by the rate of the storm. Table 7.5 shows the values of λ and λ_P for the stratiform and convective rainfalls. November and October are two wettest month, and both have rate of storm for the NSWN process higher than the PWN process. In January and February, the two driest month, the PWN process is slightly higher than the NSWN process. This suggest that one type of rainfall dominates in those months.

Assuming the same rainfall depth for both process also affects the mean number of pulses in the NSWN process. By comparing the ν in Table 7.5 with ν in Table 7.1, the latter has a bigger values. However, this results does not seem to affect the comparable moments.

The parameter α has 8 values for every month (in Table 7.6). The estimates are consistent with the diurnal pattern of the hourly daily data. In Table 7.7 shows the estimated γ values. There are 8 γ values every month in the range from zero to one.

7.7.2 Moments

The parameter estimates are used to simulate 100 datasets consisting of 36 years data. For each simulated dataset, the mean, variance and autocorrelation are derived and combined with the other 99 sets of values and plotted as a boxplot. The historical values are plotted as a solid line. The simulated values of the statistical properties are compared with the historical values to determined the goodness of fit of the 8 blocks model. This is shown in

Table 7.5: Parameter estimates for the NSWN and PWN process. The units are hour^{-1} for all estimates, except η the unit is mm and ν for number of pulses.

Month	λ	β	ν	η	λ_P	η_P
1	0.0042	0.3815	61.7828	1.7579	0.0167	1.7579
2	0.0048	0.2859	51.5314	2.0416	0.0124	2.0416
3	0.0410	0.1151	51.7509	2.4832	0.0137	2.4832
4	0.0722	0.1064	48.6191	2.5542	0.0078	2.554233
5	0.0693	0.1172	63.5381	2.0447	0.0221	2.0447
6	0.0099	0.3365	53.7725	1.9339	0.0209	1.9339
7	0.0399	0.3568	52.2433	1.4405	0.0248	1.4404
8	0.0418	0.4757	49.4175	1.4064	0.0305	1.4063
9	0.0659	0.3017	38.7704	2.0236	0.0221	2.0235
10	0.0736	0.1527	47.5814	2.3410	0.0269	2.3410
11	0.0839	0.1665	39.7886	1.8559	0.0267	1.8558
12	0.0157	0.3242	46.4933	1.7549	0.0221	1.7549

Figure 7.8, Figure 7.9 and Figure 7.10.

Figure 7.8 shows the hourly mean of the 3 hour block for 12 months. All of the months have a very good fitting of the mean. The variance plots in Figure 7.9 shows the actual variance is within the interquartile range of simulated variances. Only in February and November, the variance is not well fitted. Some of the variances of the 3 hour blocks are underestimated and some are overestimated. But by comparing the MSE of the 8 blocks model with the 12 blocks superposed Markov X-NSWN model (in Table 7.8), the 8 blocks model has a smaller values overall and also by mean, variance and autocorrelation. However in Figure 7.10, the autocorrelation plots show that the fitted model does not perform well in fitting the autocorrelation (compared with the mean and variance). The autocorrelation is underestimated in some of the blocks for January, March, May, June, September and October.

Overall, we conclude that the 8 block superposed Markov X-NSWN model reproduces the moment properties well. Reducing the parameters by combining the 3 hour data in a block instead of using 2 hour blocks has no effect in fitting the sample properties.

Table 7.6: Parameter estimates of alpha. The units are hour^{-1}

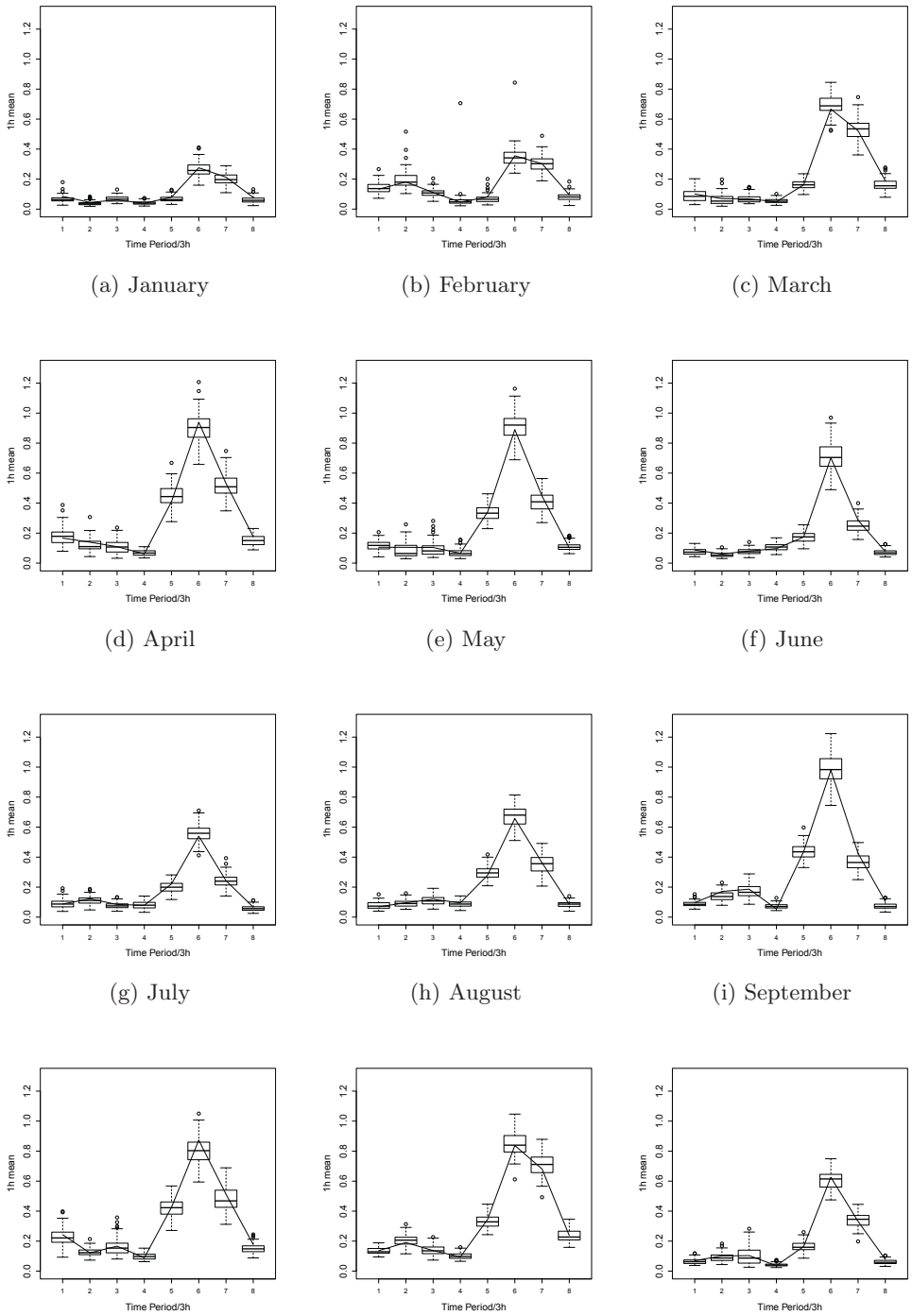
	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
α_1	0.0767	0.2343	0.0115	0.0171	0.0087	0.0398	0.0194	0.0115	0.0073	0.0215	0.0138	0.0119
α_2	0.0222	0.3687	0.0053	0.0107	0.0039	0.0246	0.0259	0.0155	0.0174	0.0071	0.0246	0.0391
α_3	0.0982	0.2154	0.0064	0.0108	0.0047	0.0395	0.0129	0.0247	0.0241	0.0128	0.014822	0.0556
α_4	0.0382	0.0449	0.0036	0.0051	0.0031	0.0633	0.0157	0.0162	0.0046	0.0038	0.008475	0.0036
α_5	0.1025	0.0812	0.0246	0.0470	0.0325	0.1222	0.0568	0.0889	0.0766	0.0442	0.0460	0.1031
α_6	0.5840	0.5698	0.1243	0.0969	0.0955	0.6895	0.1858	0.2095	0.1868	0.0923	0.1291	0.4865
α_7	0.3759	0.4548	0.0958	0.0561	0.0405	0.2097	0.0700	0.1047	0.0649	0.0496	0.1093	0.2186
α_8	0.0689	0.0971	0.0245	0.0147	0.0066	0.0277	0.0094	0.014146	0.0058	0.0107	0.0309	0.0190

.

Table 7.7: Parameter estimates of gamma. The units are hour^{-1}

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
γ_1	0.6009	0.5855	0.8106	0.5796	0.4402	0.8848	0.7868	0.6201	0.4399	0.7284	0.3463	0.8214
γ_2	0.8998	0.6520	0.9976	0.5301	0.8933	0.6276	0.5757	0.4212	0.5315	0.3908	0.4636	0.8017
γ_3	0.3755	0.2514	0.6649	0.7456	0.8674	0.5225	0.4862	0.6786	0.5682	0.7184	0.5832	0.9908
γ_4	0.4016	0.3270	0.3436	0.0343	0.7489	0.2906	0.6422	0.1181	0.2579	0.4052	0.3869	0.8382
γ_5	0.5999	0.6119	0.1448	0.4832	0.2899	0.2297	0.4038	0.2433	0.3563	0.2671	0.2728	0.3621
γ_6	0.3993	0.4705	0.4050	0.4685	0.3665	0.4970	0.4918	0.6106	0.5023	0.4463	0.4514	0.4039
γ_7	0.2593	0.4140	0.3890	0.5535	0.4662	0.3494	0.6071	0.4350	0.4661	0.4726	0.5359	0.3771
γ_8	0.5901	0.6196	0.6633	0.3921	0.4242	0.5103	0.4760	0.6169	0.6397	0.3226	0.4396	0.5108

.



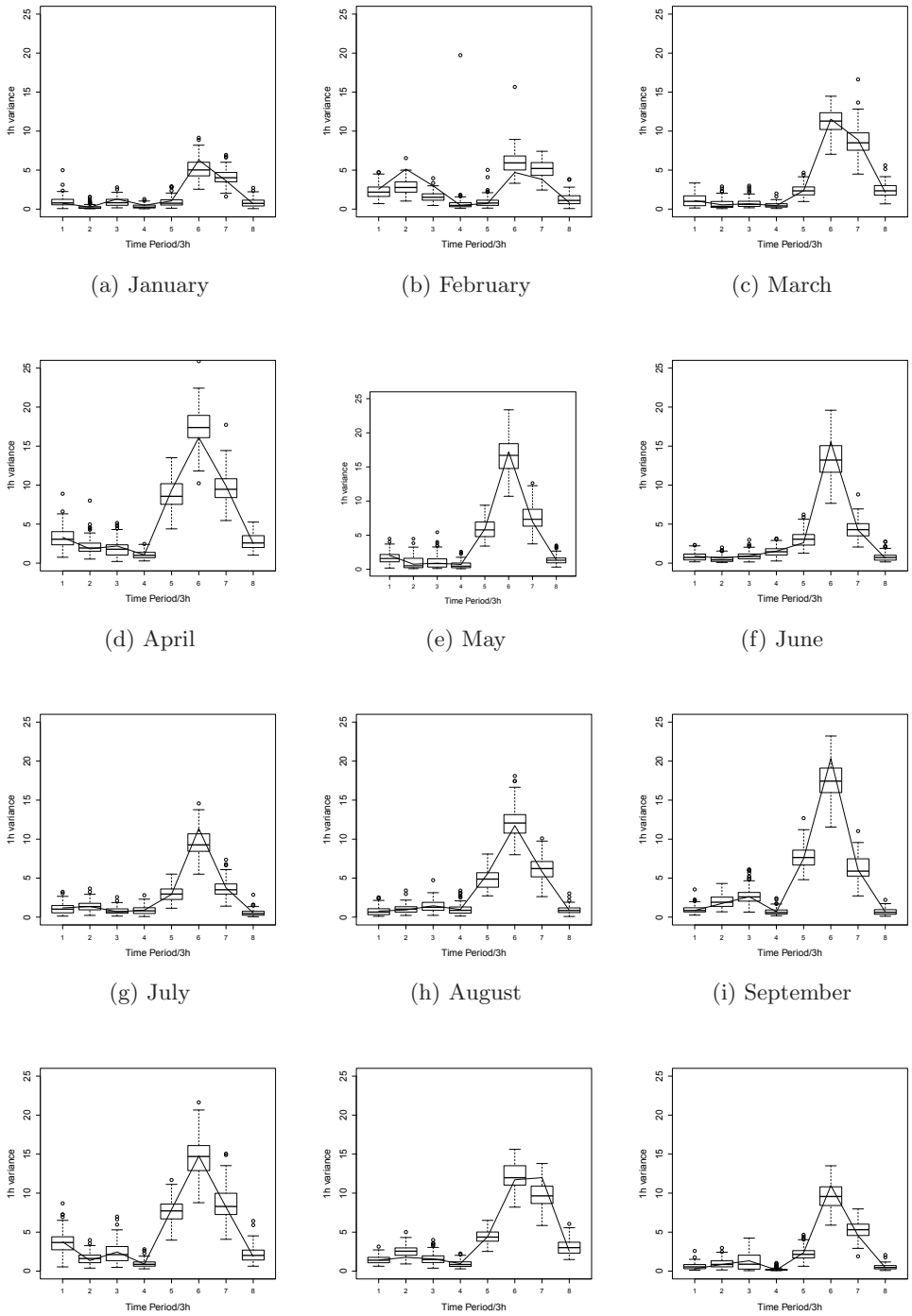
(a) January (b) February (c) March

(d) April (e) May (f) June

(g) July (h) August (i) September

(j) October (k) November (l) December

Figure 7.8: The mean of 3 hour blocks for every month



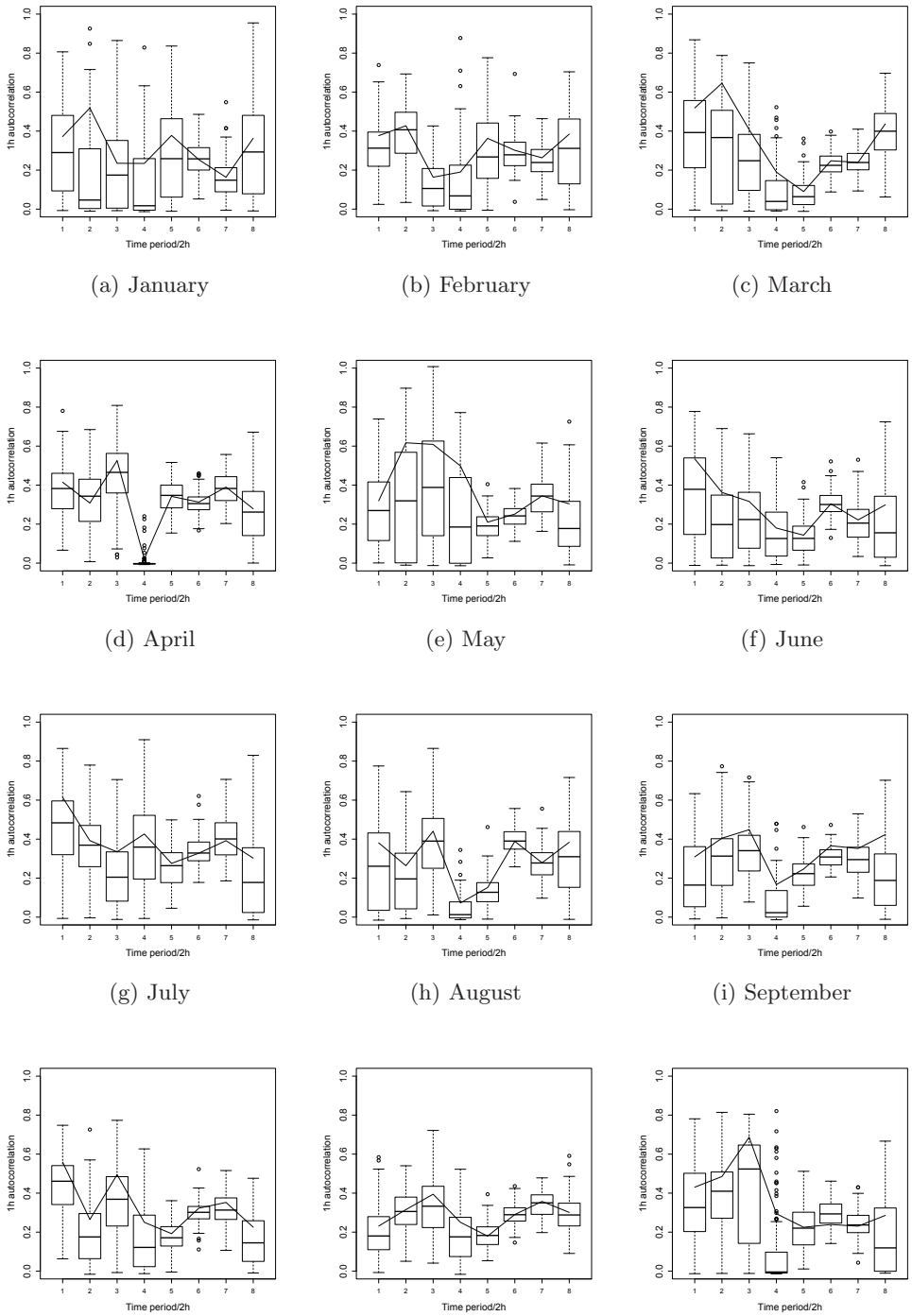
(a) January (b) February (c) March

(d) April (e) May (f) June

(g) July (h) August (i) September

(j) October (k) November (l) December

Figure 7.9: The variance of 3 hour blocks for every month



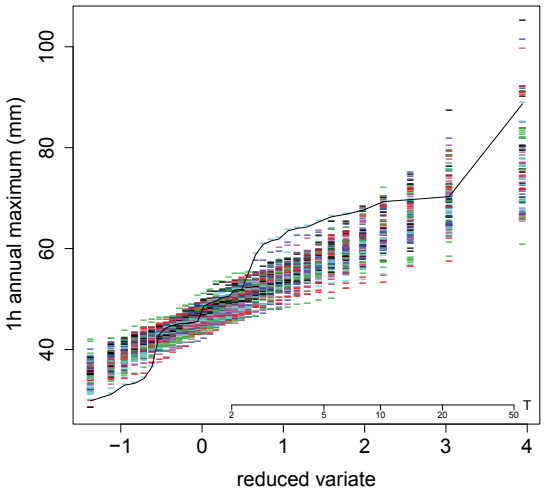
(a) January (b) February (c) March

(d) April (e) May (f) June

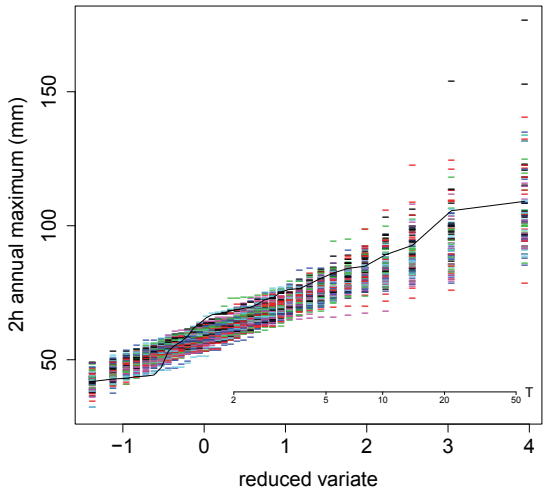
(g) July (h) August (i) September

(j) October (k) November (l) December

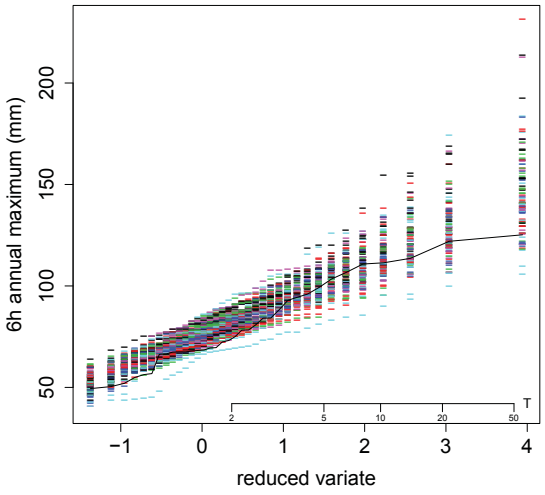
Figure 7.10: The autocorrelation of 3 hour blocks for every month



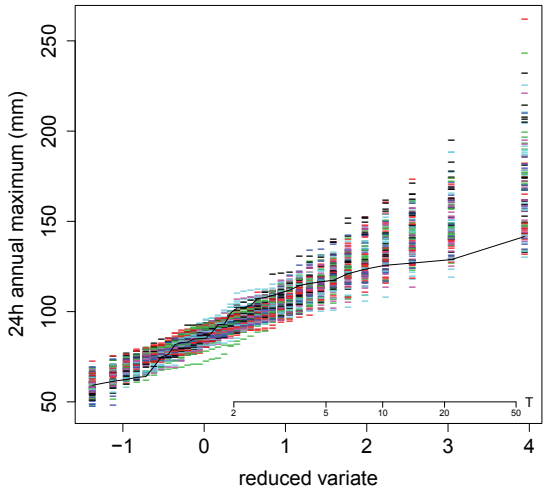
(a) 1 hour



(b) 2 hour



(c) 6 hour



(d) 24 hour

Figure 7.11: Annual extreme values for 8-blocks model

Table 7.8: Model comparison in terms of MSE

Model	Parameters (per month)	MSE Mean	MSE Variance	MSE Autocorrelation
12-block model	30	0.0192	0.8625	0.00077
8-block model	21	0.0005	0.1958	0.00048

7.7.3 Extreme values

The annual extreme values at 1, 2, 6 and 24 hour time scales are plotted in Figure 7.11. The solid line represent the historical annual extreme values of 36 years. The dash lines represent the simulation annual maximum taken from the 100 simulation set, each 36 years. The ordered maxima are plotted agains the reduce Gumbel variate.

The results are the same as the 12 blocks model of superposed Markov X-NSWN model. In the 1 hour annual maximum plot, the model underestimated the extreme values for return period less than 15 years. In the 2 hour level, the model slightly underfit the historical values for the range $T < 10$. For the 6 hour plot, the model overestimated the annual maximum. It is the same for the 24 hour plot in the range of $T > 10$. We conclude, from the extreme value results, that the model is still not perfect.

7.7.4 Monthly moments

Figure 7.12 shows the hourly mean, variance and autocorrelation for every month. Even though the monthly hourly properties were not used in the parameter estimation process, the mean and variance have a good fit for all of the months except in June. It is slightly underestimated. But the model underestimated the autocorrelation values for all of the months.

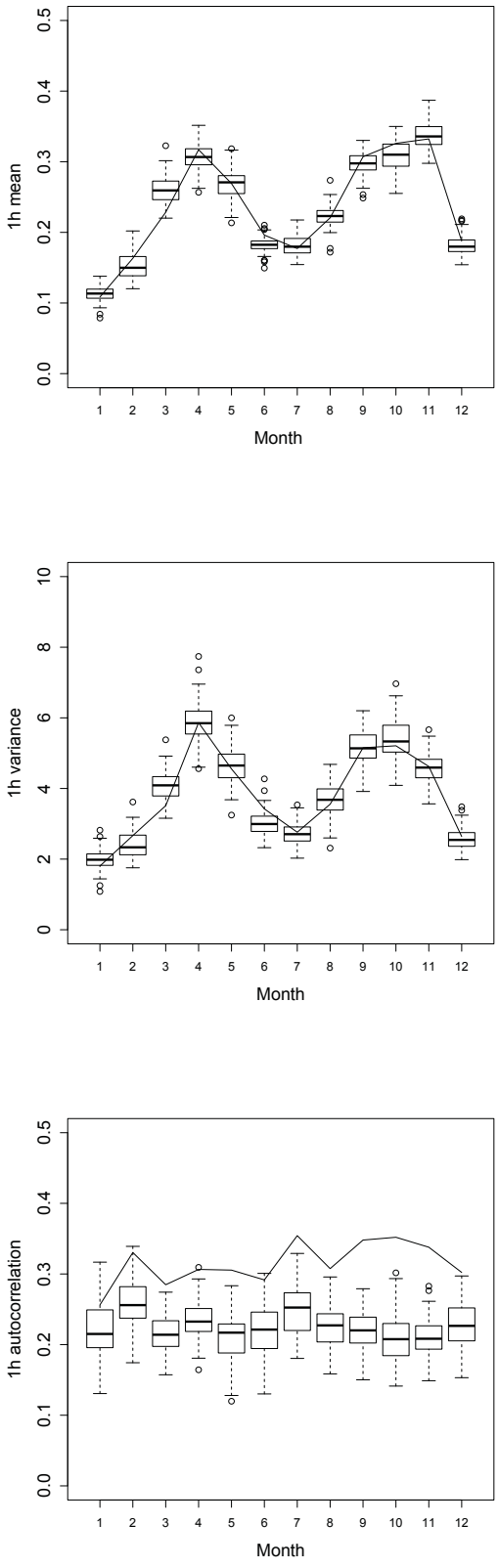


Figure 7.12: The hourly mean, variance and autocorrelation for every month

7.8 Summary and conclusion

The two point processes, NSWN and PWN were superposed to represent the two types of rainfall in Malaysia. The NSWN, a cluster process, represents the stratiform rainfall events. The PWN with only an instantaneous burst, represents the convective rainfall events. In the superposition, the NSWN is assumed as the main event. Therefore it is combined with the rainfall occurrence parameter and also the Markov dependence parameter. The PWN process is added to main rainfall event, and is assumed similar to the convective event, short but intense rainfall.

The main reason to superpose the Markov X-NSWN and PWN is to add to the variability of the rainfall in the process. It was thought this would also help reproduce the historic variance. In the Markov X-PWN model and Markov X-NSWN model, both have similar results for the variance, mostly a good estimate in all of the two hour blocks except for block 9, the hours with maximum rainfall. Both the models underestimated the variance. For the superposed model, it improved the result for variance, especially for block 9. Overall the superposed Markov X-NSWN and PWN improved the fitting for the mean, variance and autocorrelation. Even for the extreme values which were not included in the estimation process, the superposed model fit better than the Markov X-PWN model and Markov X-NSWN model.

The disadvantage of the superposed model is that it involves a lot of parameters. For every month, there are 30 parameters needed to simulate the rainfall data. This may cause some concern for other researchers. However, this chapter's results at least give a preliminary analysis to show that the superposed model can perform adequately. If a model is to describe rainfall events it should also be realistic in portraying the types of rainfall in Malaysia. From the results, the superposed Markov X-NSWN and PWN perform better than the earlier models. The model succeeds in describing the Malaysia rainfall data as a combination of two different rainfall event.

To reduce the number of parameters, it was not possible to combine the 12 α or γ. Although results from combining the 12 values are not discussed here, it is clear theoretically, the α and γ represent the characteristic of the 12 blocks. These parameters shape the diurnal pattern in the hourly rainfall data. Therefore, to scale down the number of α and γ is to decrease the total number of blocks used in the model. The 12 blocks for the superposed model was reduced to 8 blocks. Results show that the 8-block model can performed better than the 12-block model. The diurnal pattern in the hourly data is still visible in the 8 block data. The total number of parameters is only 21 per month compared to 30 per month.

In this chapter, we have shown that the superposed Markov X-NSWN and PWN model is the best model to fit the daily hourly Malaysian rainfall data. The 8-block model is sufficient to fit the monthly rainfall events with a diurnal pattern in the hourly data. The flexibility of the model to fit the 2 hour block or the 3 hour block will allow other researchers to explore the diurnal pattern at different time scales.

Chapter 8

Conclusions and recommendations

8.1 Conclusions

Point process models have been widely used for the rainfall data with variety of modifications to the model to improve its performance in reproducing realistic rainfall simulations. In this thesis, the aim of the study is to build in a diurnal variation in the point process models. This was done by adding the rainfall occurrence parameter.

The PWN and NSWN models were chosen to describe the rainfall process in Malaysia. The instantaneous pulses best described the rainfall events in Malaysia since the afternoon rainfall is usually heavy in a short period of time. The white noise models were considered (instead of the rectangular pulse models) because of the simplicity of the models to formulate the diurnal variation.

Malaysia rainfall data has a significant diurnal variation in the daily data. To allow the model to display the diurnal variation, the model was fitted to the monthly rainfall data by using the properties of two hour blocks instead of the usual norm of using the hourly properties. However, the main point process models were assumed the same for each of the 12 blocks, thus having only one set of parameters for the models for each month with the rainfall data pooled within these blocks over 36 years. On the other hand there are 12 rainfall occurrence parameters, one for each block. It has been shown that the rainfall occurrence parameter follows the diurnal pattern in the daily rainfall.

To present the conclusions of the research described in the thesis, the aim and objectives of the research, outlined in chapter 1, are reviewed and their achievement addressed according to the chapters. These are listed below.

1. The nonhomogenous PWN model was fitted to the two hour blocks separately with the same depth parameter. This was the first approach to model the diurnal variation using the point process model. The PWN process with a different storm rate for

each of the two hour blocks was able to capture the observed diurnal pattern in the rainfall. But this was done with the assumption that the PWN process was independent between the blocks. Even though the dependence between the blocks was ignored, having a nonhomogenous PWN process for each two hour block was chosen to account for the diurnal variation in the daily rainfall data.

2. Instead of having different parameters of the point process models by assuming non-homogeneity for every two hour block, an additional binary variable used as an indicator for rainfall events was formulated to represent the diurnal pattern in the rainfall data. The main process for the rainfall is still the point process model and assumed stationary through out the month. But at the same time the added parameter will allow for excess rainfall during the heavy rainfall period in the late afternoon and reduced rainfall depth in other periods, thus giving the diurnal variation in the rainfall data.

3. The binary rainfall variable is assumed to be independent of the main point process models. So the point processes are assumed stationary for every month, then the indicator variables are assumed non-stationary for every two hour block. The two hour block was chosen to reduce the number of parameters. But using the two hour block provide enough to represent the diurnal variation. Therefore, there are 12 indicator variables with probability α. The block start at 12-1am. The hours in each block were first assumed independent, with a second variant of this model prescribing that they are dependent. A Markov dependence parameter, γ is added to the modified model.

4. The X-PWN model is the modified PWN with an assumption of independence between the hours within the two hour block. The simulated data produced more rainfall in block 8 to 10, where the maximum rainfall occurs. This was slightly improved with the added Markov dependence in the Markov X-PWN model. The comparison between the PWN, X-PWN and Markov X-PWN model based on the mean square error showed that the best model was Markov X-PWN model. However, the extreme value plot showed that the model is not able to preserve other properties of the observed rainfall data that were not included in the parameter estimation process.

5. The X-NSWN model also had a poor fitting in block 8 to 10 which was reflected in the α. This shows that the rainfall indicator is the main parameter for the diurnal variation. The assumption of dependence within the hours of the block is more accurate for the Malaysia rainfall data since the Markov X-NSWN model has better fit results than the X-NSWN model. The model also did not produce a good fit of the extreme values.

6. Although both the Markov X-PWN and Markov X-NSWN models are preferred over the independence model, both models still do not perform well in fitting the afternoon blocks which have the heaviest rainfall. This suggests that it may not be possible to represents Malaysia rainfall events by just one point process.

7. The Markov X-NSWN model was superposed with the PWN process. Only the modified NSWN process causes the diurnal variation in the rainfall data while the PWN process is assumed as the convective rainfall. The superposed model was superior compared to the other models, based on the graphical comparisons and the mean square error. But the superposed model had 29 parameters per month. The extreme values plot had an improved fit compared with the other models.

8. To reduced the number of parameters, the superposed model was fitted to three hour blocks. The 8-block model produced a good fit to the derived properties to the related historical values. The extreme values plot for the 8-block model was the same as the superposed model. The plot shows an improvement to the fitting of the extreme values compared to the other models even though it is still slightly underestimated. The modified models may not be able to produce a sastify extreme values.

9. The modified point process model has shown that the rainfall indicator represents the diurnal variation in the rainfall data, while remaining the point process as the main model for rainfall events. The 8-block superposed model is well fitted to the Malaysia rainfall data. The Markov X-NSWN and PWN process represent Malaysia rainfall events with the assumption of dependence between the hours within the three hour blocks.

8.2 Recommendations for future work

Several recommendations are identified for future work in the modeling of diurnal variation rainfall data and listed below.

1. The point process model was modified with a few simple assumptions as a starting point of the modification. Some other assumptions such as the independence assumption between the blocks could be relaxed by assuming dependence by using a Markov chain or copula model. The rainfall depth was assumed exponential distributed. Other heavy tailed distributions could be used for the rainfall depth such as the mixed exponential distribution. This could improve the extreme rainfall characteristics.

2. In this thesis, only the white noise models were tested for the add-in diurnal variation parameter. Other point process models such as the rectangular pulse models (NSRP

model or BLRP model) or the pulse model (Bartlett-Lewis pulse model) could be used as the main point process models for the rainfall events.

3. The superposed model was developed for a single site only. But the model could be generalized for many sites simultaneously for further development.

4. Even though in the current model many parameters are needed to represent the daily diurnal variation, and it was shown that these parameters were needed to account for the diurnal variation, a model with less parameters for the diurnal rainfall could be preferred, and should always be a goal of modelling.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 19(6): 716–723.

Burlando, P. & Rosso, R. (1991). Comment on parameter estimation and snsitivity analysis for the modified Bartlett-Lewis rectangular pulses model of rainfall. *Journal of Geophysical Research-Atmospheres*. 96(D5).

Bury, K. (1999). *Statistical distributions in engineering*. Cambridge University Press.

Cameron, D., Beven, K. & Tawn, J. (2001). Modelling extreme rainfalls using a modified random pulse Bartlett-Lewis stochastic rainfall model(with uncertainty). *Advances in Water Resources*. 24(2): 203–211.

Coe, R. & Stern, R. (1982). Fitting models to daily rainfall data. *Journal of Applied Meteorology*. 21: 1024–1031.

Cowpertwait, P. (1991). Further developments of the Neyman-Scott clustered point process for modeling rainfall. *Water Resources Research*. 27(7).

Cowpertwait, P. (1994). A generalized point process model for rainfall. In *Proceedings of the Royal Society: Mathematical and Physical Sciences (1990-1995)*. Vol. 447. The Royal Society. 23–37.

Cowpertwait, P. (1995). A generalized spatial-temporal model of rainfall based on a clustered point process. In *Proceedings of the Royal Society: Mathematical and Physical Sciences (1990-1995)*. Vol. 450. The Royal Society. 163–175.

Cowpertwait, P. (1998). A Poisson cluster model of rainfall: Some high-order moments and extreme values. *Proceedings of the Royal Society of London, Series A*. 454: 885–898.

Cowpertwait, P. (2004). Mixed rectangular pulses models of rainfall. *Hydrology and Earth System Sciences*. 8(5): 993–1000.

Cowpertwait, P. (2006). A spatial–temporal point process model of rainfall for the Thames catchment, UK. *Journal of Hydrology*. 330(3-4): 586–595.

Cowpertwait, P., Isham, V. & Onof, C. (2007). Point process models of rainfall: developments for fine-scale structure. *Proceedings of the Royal Society of London, Series A.* 463: 2569–2587.

Cowpertwait, P., Kilsby, C. & O'Connell, P. (2002). A space-time Neyman-Scott model of rainfall: Empirical analysis of extremes. *Water Resources Research.* 38(8): 6.

Cowpertwait, P. & O'Connell, P. (1997). A regionalised Neyman-Scott model of rainfall with convective and stratiform cells. *Hydrology and Earth System Sciences.* 1(1): 71–80.

Cowpertwait, P. S. P. (2010). A Neyman–Scott model with continuous distributions of storm types. In Howlett, P., Nelson, M. & Roberts, A. J. (Eds.). *Proceedings of the 9th Biennial Engineering Mathematics and Applications Conference, EMAC-2009.* Vol. 51 of *ANZIAM J.*. C97–C108.

Cowpertwait, P. S. P., Xie, G., Isham, V., Onof, C. & Walsh, D. C. I. (2011). A fine-scale point process model of rainfall with dependent pulse depths within cells. *Hydrological Sciences Journal.* 56: 1110–1117.

Cox, D. & Isham, V. (1980). *Point processes.* Chapman & Hall/CRC.

Cox, D. & Isham, V. (1988). A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences.* 317–328.

Cox, D. & Isham, V. (1994). Stochastic models of precipitation. *In: V.Barnett and K.F. Turkman (Eds), Statistics for the Environment 2 : Water Related Issues, Wiley, Chichester, UK.* 218.

Desa, M. & Niemczynowicz, J. (1996). Temporal and spatial characteristics of rainfall in Kuala Lumpur, Malaysia. *Atmospheric Research.* 42(1-4): 263–277.

Entekhabi, D., Rodriguez-Iturbe, I. & Eagleson, P. (1989). Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: Parameter estimation and validation. *Water Resources Research.* 25(2): 295–302.

Evin, G. & Favre, A. (2008). A new rainfall model based on the Neyman-Scott process using cubic copulas. *Water Resources Research.* 44(3): W03433.

Faofoula-Georgiou, E. (1985). Discrete time point process models for daily Rainfall. *Water Resources Series Technical Report.* 93.

Foufoula-Georgiou, E. & Georgakakos, J. P. (1991). Hydrologic advances in space-time precipitation modeling and forecasting. In Bowles, D. & O'Connell, P. (Eds.). *Recent Advances in the Modeling of Hydrologic Systems.* 47–65.

Foufoula-Georgiou, E. & Guttorp, P. (1986). Compatibility of continuous rainfall occurrence models with discrete rainfall observations. *Water Resources Research*. 22: 1316–1322.

Foufoula-Georgiou, E. & Guttorp, P. (1987). Assessment of a class of Neyman-Scott models for temporal rainfall. *Journal of Geophysical Research*. 92: 9679–9682.

Foufoula-Georgiou, E. & Lettermaier, D. (1986). Continuous-time versus discrete-time point process models for rainfall occurrence series. *Water Resources Research*. 22: 531–542.

Gabriel, K. & Neumann, J. (1962). A markov chain model for daily rainfall occurrence at Tel Aviv. *Journal of the Royal Meteorology Society*. 90–95.

Gyasi-Agyei, Y. (2001). Modelling diurnal cycles in point rainfall properties. *Hydrological Processes*. 15(4): 595–608.

Gyasi-Agyei, Y. & Willgoose, G. (1997). A hybrid model for point rainfall modeling. *Water Resources Research*. 33(7).

Hanaish, I. S., Ibrahim, K. & Jemain, A. A. (2011a). Daily rainfall disaggregation using HYETOS model for Peninsular Malaysia. *Recent Researches in Applied Mathematics, Simulation and Modelling*. 146–150.

Hanaish, I. S., Ibrahim, K. & Jemain, A. A. (2011b). Stochastic modeling of rainfall in Peninsular Malaysia using Bartlett Lewis rectangular pulses models. *Modelling and Simulation in Engineering*. 2011.

Hanaish, I. S., Ibrahim, K. & Jemain, A. A. (2012). Reproduction of the variance scaling structure for rainfall in Peninsular Malaysia using Bartlett Lewis model. *Applied Mathematical Sciences*. 6: 5305–5313.

Islam, S., Entekhabi, D., Bras, R. & Rodriguez-Iturbe, I. (1990). Parameter estimation and sensitivity analysis for the modified Bartlett-Lewis rectangular pulses model of rainfall. *Journal of Geophysical Research*. 95(D3): 2093–2100.

Kaczmarska, J. (2011). *Further development of Bartlett-Lewis model for fine-resolution rainfall*. Technical Report 312. Department of Statistical Science, University College London.

Katz, R.W., P. M. (1995). Generalizations of chain-dependent processes: application to hourly precipitation. *Water Resources Research*. 31(5): 1331–1341.

Kavvas, M. & Delleur, J. (1975). The Stochastic and chronologic structure of rainfall sequences-application to Indiana. *Water Resources Center*. Rep.57.

Kavvas, M. L. & Delleur, J. W. (1981). A stochastic cluster model of daily rainfall sequences. *Water Resources Research*. 17: 1151–1160.

Khaliq, M. & Cunnane, C. (1996). Modelling point rainfall occurrences with the modified Bartlett-Lewis rectangular pulses model. *Journal of Hydrology*. 180(1-4): 109–138.

Klotz, J. (1973). Statistical inference in Bernoulli trials with dependence. *The Annals of Statistics*. 1(2): 373–379.

Leonard, M., Lambert, M., Metcalfe, A. & Cowpertwait, P. (2008). A space-time Neyman–Scott rainfall model with defined storm extent. *Water Resources Research*. 44(9).

Lu, Y. & Qin, X. (2012). Comparison of stochastic point process models of rainfall in Singapore. *World Academy of Science, Engineering and Technology*. 68: 1493–1497.

Morrissey, M. (2009). Superposition of the Neyman-Scott rectangular pulses model and the Poisson white noise model for the representation of tropical rain rates. *Journal of Hydrometeorology*. 10(2): 395–412.

Morrissey, M. L. (1993). A point process model for tropical rainfall. *Journal of Geophysical Research*. 98(D9): 16639–16652.

Nelder, J. & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*. 7: 308–313.

Neyman, J. & Scott, E. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1–43.

Nieuwolt, S. (1968). Diurnal rainfall variation in Malaya. *Annals of the Association of American Geographers*. 58(2): 313–326.

Northrop, P. (1998). A clustered spatial-temporal model of rainfall. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 454(1975): 1875–1888.

Northrop, P. J. & Stone, T. M. (2005). *A point process model for rainfall with truncated gaussian rain cells*. Technical Report Research Report No.251. Department of Statistical Science, University College London.

Obeysekera, J. T. B., Tabios, G. Q. & Salas, J. D. (1987). On parameter estimation of temporal rainfall models. *Water Resources Research*. 23: 1837–1850.

Oki, T. & Musiake, K. (1994). Seasonal change of the diurnal cycle of precipitation over Japan and Malaysia. *Journal of Applied Meteorology*. 33: 1445–1463.

Onof, C., Chandler, R., Kakou, A., Northrop, P., Wheater, H. & Isham, V. (2000). Rainfall modelling using Poisson-cluster processes: a review of developments. *Stochastic Environmental Research and Risk Assessment (SERRA)*. 14(6): 384–411.

Onof, C. & Wheater, H. (1993). Modelling of British rainfall using a random parameter Bartlett-Lewis rectangular pulse model. *Journal of hydrology (Amsterdam).* 149(1-4): 67–95.

Onof, C. & Wheater, H. (1994a). Improved fitting of the Bartlett-Lewis rectangular pulse model for hourly rainfall. *Hydrological Sciences Journal.* 39(6): 663–680.

Onof, C. & Wheater, H. (1994b). Improvements to the modelling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model. *Journal of hydrology (Amsterdam).* 157(1-4): 177–195.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0.

Ramage, C. (1964). Diurnal variation of summer rainfall of Malaysia. *Journal Tropical Geography.* 19: 62–68.

Ramesh, N. I. (1998). Temporal modelling of short-term rainfall using Cox processes. *Environmetrics.* 9: 629–643.

Reading, A. and Millington, A.C. and Thompson, R.D. (1995). *Humid tropical environments*. Wiley.

Rodriguez-Iturbe, I., Cox, D. & Isham, V. (1987a). Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences.* 269–288.

Rodriguez-Iturbe, I., Cox, D. & Isham, V. (1988). A point process model for rainfall: further developments. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences (1934-1990).* 417(1853): 283–298.

Rodriguez-Iturbe, I., Gupta, V. & Waymire, E. (1984). Scale considerations in the modeling of temporal rainfall. *Water Resources Research.* 20(11): 1611–1619.

Rodriguez-Iturbe, I., Power, B. D. & J.B.Valdes (1987b). Rectangular pulses point process models for rainfall: analysis of empirical data. *Journal of Geophysical Research-Atmospheres.* 92(D8).

Smith, J. A. (1987). Statistical modeling of daily rainfall occurrences. *Water Resources Research.* 23: 885–893.

Smith, J. & Karr, A. (1983). A point procss model of summer season rainfall occurrence. *Water Resources Research.* 19: 95–103.

Smith, J. & Karr, A. (1985). Statistical inference for point process models of rainfall. *Water Resources Research.* 21(1): 73–80.

Smithers, J., Pegram, G. & Schulze, R. (2002). Design rainfall estimation in South Africa using Bartlett–Lewis rectangular pulse rainfall models. *Journal of Hydrology*. 258(1-4): 83–99.

Srikanthan, R. & McMahon, T. (2001). Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth Sciences*. 5(4): 653–670.

Suhaila, J. & Jemain, A. A. (2007). Fitting daily rainfall amount in Malaysia using the normal transform distribution. *Journal of Applied Sciences*. 7(14): 1880–1886.

The GCC team (2007). *GCC: A free C language software package, compiler version 4.1.1.*

Todorovic, P. & Woolhiser, D. (1975). A stochastic model of n-day precipitation. *Journal of Applied Meteorology*. 14(1): 17–24.

Valdes, J., Rodriguez-Iturbe, I. & Gupta, V. (1985). Approximations of temporal rainfall from multidimensional model. *Water Resources Research*. 21: 1259–1270.

Waymire, E. & Gupta, V. (1981a). The mathematical structure of rainfall representations, 1, A review of the stochastic rainfall models. *Water Resources Research*. 17(5): 1261–1272.

Waymire, E. & Gupta, V. (1981b). The mathematical structure of rainfall representations, 2, A review of the theory of point processes. *Water Resources Research*. 17(5): 1273–1285.

Waymire, E. & Gupta, V. (1981c). The mathematical structure of rainfall representations, 3, Some applications of the point process theory to rainfall processes. *Water Resources Research*. 17(5): 1287–1294.

Xie, G. (2010). *Further developments of two point process models for fine-scale time series*. Massey University: Ph. D. Thesis.

Yusof, F. (2007). *Stochastic modeling of hourly rainfall processes*. Universiti Teknologi Malaysia: Ph. D. Thesis.

Yusof, F., Norzaida, A. & Zalina, M. D. (2008). Fourier series in a Neyman-Scott rectangular pulse model. *MATEMATIKA*. 24(2): 243–257.

Yusof, F., Zalina, M. D., Nguyen, V.-T.-V. & Zulkifli, Y. (2007). Performance of mixed exponential an exponential distribution representing rain cell intensity in Neyman-Scott rectangular pulse(NSRP) model. *Malaysian Journal of Civil Engineering*. 19(1): 55–72.