

**THE DEVELOPMENT AND ASSESSMENT OF TECHNIQUES FOR  
DAILY RAINFALL DISAGGREGATION IN SOUTH AFRICA**

Darryn Marc Knoesen

DISSERTATION

Submitted in partial fulfilment of the  
requirements for the degree of MSc

School of Bioresources Engineering and Environmental Hydrology

University of KwaZulu-Natal

Pietermaritzburg

June 2005

## ABSTRACT

The temporal distribution of rainfall, *viz.* the distribution of rainfall intensity during a storm, is an important factor affecting the timing and magnitude of peak flow from a catchment and hence the flood-generating potential of rainfall events. It is also one of the primary inputs into hydrological models used for hydraulic design purposes. The use of short duration rainfall data inherently accounts for the temporal distribution of rainfall, however, there is a relative paucity of short duration data when compared to the more abundantly available daily data. One method of overcoming this is to disaggregate coarser-scale data to a finer resolution, e.g. daily to hourly.

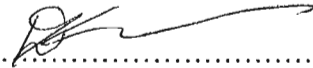
A daily to hourly rainfall disaggregation model developed by Boughton (2000b) in Australia has been modified and applied in South Africa. The primary part of the model is the distribution of  $R$ , which is the fraction of the daily total that occurs in the hour of maximum rainfall. A random number is used to sample from the distribution of  $R$  at the site of interest. The sample value of  $R$  determines the other 23 values, which then undergo a clustering procedure. This clustered sequence is then arranged into 1 of 24 possible temporal arrangements, depending when the hour the maximum rainfall occurs. The structure of the model allows for the production of 480 different temporal distributions with variation between uniform and non-uniform rainfall. The model was then regionalised to allow for application at sites where daily rainfall data, but no short duration data, were available.

The model was evaluated at 15 different locations in differing climatic regions in South Africa. At each location, observed hourly rainfall data were aggregated to yield 24-hour values and these were then disaggregated using the methodology. Results show that the model is able to retain the daily total and most of the characteristics of the hourly rainfall at the site, for when both at-site and regional information are used. The model, however, is less capable of simulating statistics related to the sequencing of hourly rainfalls, e.g. autocorrelations. The model also tends to over-estimate design rainfalls, particularly for the shorter durations.

## PREFACE

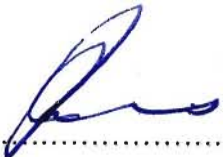
The work described in this dissertation was carried out in the School of Bioresources Engineering and Environmental Hydrology, University of KwaZulu-Natal, Pietermaritzburg, from January 2003 to June 2005, under the supervision of Professor Jeff Smithers.

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to any university. Where use has been made of the work of others it is duly acknowledged in the text.

Signed:  .....

D.M.Knoesen (author)

Date: 8/12/2005

Signed:  .....

J.C.Smithers (supervisor)

Date: 8/12/2005

## ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation and gratitude to the following people and institutions for the assistance rendered during this study:

- Professor Jeff Smithers, for supervising this project and providing invaluable assistance throughout the duration of this study,
- Ms Kershani Chetty, for her support and encouragement throughout the duration this project,
- Dr Walter Boughton, for all the information given to enhance the understanding and operation of his computer program,
- Demetris Koutsoyiannis, Erwin Weinmann, Bellie Sivakumar, Dr Sri Srikanthan, and Dr D Nagesh Kumar, for the relevant contributions they made in the acquisition of current literature,
- Mr Shaun Thornton-Dibb and Mr Ragesh Nundlall, for the time spent assisting with computer programming,
- The Water Research Commission (WRC), for funding the WRC Project K5/1318 titled “The Development of a Continuous Simulation Modelling System for Design Flood Estimation in South Africa”, to which parts of this dissertation contributed,
- The school of Bioresources Engineering and Environmental Hydrology, for providing a working environment and resources making this study possible.

# TABLE OF CONTENTS

	Page
ABSTRACT	i
PREFACE	ii
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	vi
LIST OF TABLES	viii
1. INTRODUCTION	1
2. METHODS TO ESTIMATE THE TEMPORAL DISTRIBUTION OF RAINFALL	4
2.1 Design Hyetographs Derived from Intensity-Duration-Frequency Relationships	4
2.1.1 Temporal storm distributions	5
2.1.2 SCS-SA temporal storm distributions	6
2.2 Hyetographs Derived from Observed Rainfall Data	8
2.2.1 Triangular rainfall distribution	8
2.2.2 Huff curves	10
2.2.3 The average variability method	11
2.2.4 Sampling from historical records	12
2.3 Rainfall Disaggregation Models	13
2.3.1 Stochastic models based on dimensionless hyetographs	14
2.3.2 Rectangular pulse cluster based stochastic rainfall models	15
2.3.3 Multiple site rainfall disaggregation models	16
2.3.4 Artificial neural networks	18
2.3.4.1 Advantages and disadvantages of ANNs	19
2.3.4.2 Neural network applications	20
2.4 Discussion and Conclusions	21
3. DATA UTILISED IN THE STUDY	24
3.1 Selection of Hidden Stations for Model Testing	24
3.2 Data Selection	27
4. THE DISAGGREGATION MODEL	28
4.1 Structure of the Model	28

## TABLE OF CONTENTS (continued)

	Page	
4.2	Distribution of R	28
4.3	Calculating the Other 23 Hourly Fractions	30
4.4	Clustering of Hourly Rainfalls	32
4.5	Daily Temporal Patterns of Hourly Rainfalls	35
4.6	Chapter Conclusions	38
5.	REGIONALISATION	39
6.	MODEL TESTING AND RESULTS	45
6.1	Selecting Data Threshold for Model Development	46
6.2	Moments and Statistics Using At-Site Short Duration Data	48
6.3	Extreme Rainfall Events Using At-Site Short Duration Data	54
6.4	Moments and Statistics When Using Regionalised Information	58
6.5	Extreme Rainfall Events When Using Regionalised Information	62
6.6	Chapter Conclusions	64
7.	DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS	65
7.1	Short Duration Data Used	65
7.2	Methodology	66
7.3	Regionalising the Methodology	66
7.4	Application of the Disaggregation Model	67
7.4.1	Application of the Model Using At-site Short Duration Data	67
7.4.2	Application of the Model Using Regionalised Input	68
7.5	Summary	68
7.6	Recommendations for Future Research	69
8.	REFERENCES	71
APPENDIX A		
SITE CHARACTERISTICS OF STATIONS USED IN MODEL DEVELOPMENT		
		79
APPENDIX B		
AVERAGE RANKED SERIES OF HOURLY FRACTIONS		
		86
APPENDIX C		
CLUSTERED SEQUENCES OF HOURLY FRACTIONS		
		89

## LIST OF FIGURES

		Page
Figure 2-1	Schematic diagram depicting two broad categories into which the methodology for deriving temporal rainfall distributions may be classed	4
Figure 2-2	Synthetic design rainfall distributions for South Africa (after Schmidt and Schulze, 1987)	7
Figure 2-3	Spatial distribution of design rainfall intensity types over South Africa (after Weddepohl, 1988)	7
Figure 2-4	The triangular rainfall distribution (after Lambourne and Stephenson, 1986)	9
Figure 2-5	An example of first quartile Huff curves (Chow <i>et al.</i> , 1988)	10
Figure 2-6	Artificial neuron or processing unit (after Burian <i>et al.</i> , 2000), where $X_i$ denotes the inputs and $W_i$ denotes the weights	18
Figure 3-1	Locations of the 15 “hidden” stations used for model testing	25
Figure 3-2	Locations of stations used for model development	26
Figure 3-3	Distribution of record lengths for the data used in model development	26
Figure 4-1	Example of a single day’s hourly rainfall at Ntabamhlope	29
Figure 4-2	Frequency distributions of R at Stations Jnk19a (in the W. Cape) and N23 (in KZN), using the ranges of R shown in Table 4-1	30
Figure 4-3	Averaged ranked series of hourly fractions for selected ranges of R, calculated using all 157 stations	32
Figure 4-4	Frequency distributions of the hour of maximum rainfall at Station Jnk19a (Jonkershoek, in the W. Cape) and Station N23 (Ntabamhlope, in KZN)	36
Figure 4-5	Samples of the different temporal distributions that are generated by the disaggregation model	37
Figure 5-1	Stations that fell within Range 8 but were included in Range 9	40
Figure 5-2	Distributions of R for sites included with those in Range 9	40
Figure 5-3	Stations that fell within Range 13 but were included in Range 12	41
Figure 5-4	Distributions of R for sites included with those in Range 12	41
Figure 5-5	Regionalised distributions of R	42
Figure 5-6	Regionalised distributions of the hour of maximum rainfall	43
Figure 5-7	Regionalised map of $R_{\text{mean}}$	43

## LIST OF FIGURES (continued)

		Page
Figure 6-1	Locations of stations used for analysis of the 5 variations of the disaggregation model	46
Figure 6-2	Simulated performance of the disaggregation model at Station 0435019	50
Figure 6-3	Simulated performance of the disaggregation model at Station Sacfs	51
Figure 6-4	Design rainfall estimated using disaggregated data for Station 0028748 and Station 0474680	56
Figure 6-5	Simulated performance of the disaggregation model at Station 0435019, when regionalised input is used	59
Figure 6-6	Simulated performance of the disaggregation model at Station Sacfs, when regionalised input is used	60
Figure 6-7	Single-site and regionalised distributions for the hour of maximum rainfall applied at Station 0106880	61
Figure 6-8	Design rainfall estimated using disaggregated data for Station 0028748 and Station 0474680, when regionalised input is used	63



## LIST OF TABLES

		<b>Page</b>
Table 3-1	Details of the 15 “hidden” stations used for model testing	25
Table 4-1	Ranges used when collating R values	29
Table 4-2	Average highest 1-hour, 2-hour, 3-hour, 6-hour, and 12-hour fraction for each range of R, calculated using all 157 stations	34
Table 4-3	24 Samples of temporal arrangements of the hourly rainfalls (after Boughton, 2000b)	37
Table 5-1	Revised ranges used when collating $R_{\text{mean}}$ values	42
Table 6-1	Mean absolute relative error computed for hourly rainfall values for the 5 variations of the disaggregation model tested at 5 different locations	47
Table 6-2	Number of months with complete data in a 20 year record at Station 0435019	48
Table 6-3	<i>MARE</i> and <i>MARE_AD</i> for the 15 test stations	52
Table 6-4	<i>MARE_LAG</i> for autocorrelation lags 1-10 at Station 0435019 and Station Sacfs	53
Table 6-5	Number of months with complete data in a 20 year record at Station Sacfs	53
Table 6-6	<i>MARE</i> and <i>MARE_AD</i> for design rainfalls at the 15 test stations	57
Table 6-7	Mean values of R and <i>MARE</i> for the 15 test stations	57
Table 6-8	<i>MARE</i> and <i>MARE_AD</i> for the 15 test stations using the regionalised methodology	58
Table 6-9	<i>MARE</i> and <i>MARE_AD</i> for design rainfalls at the 15 test stations using the regionalised methodology	62

## 1. INTRODUCTION

Engineers and hydrologists involved in the design of hydraulic structures, such as dams, bridges and culverts, need to accurately assess the frequency and magnitude of extreme hydrological events. Current techniques for design flood estimation in South Africa need to be updated, regional approaches need to be evaluated and new techniques have to be developed and applied (Smithers and Schulze, 2001). One input that would improve the estimation of design floods is to account for the temporal distribution of rainfall.

The temporal distribution of rainfall, *viz.* the distribution of rainfall intensity during a storm, is an important factor affecting the timing and magnitude of peak flow from a catchment and hence the flood-generating potential of rainfall events (Weddepohl, 1988). It is also one of the primary inputs into hydrological models used for the design of hydraulic structures. The temporal distribution of rainfall events may be influenced by many factors that need to be reflected in design temporal distributions. These factors include, *inter alia*, location, storm duration, storm depth, and season of storm occurrence (Hoang *et al.*, 1999).

The intensity distribution of a storm may be estimated by the use of a temporal distribution curve, which may be synthetically derived or obtained from observed hyetographs (Chow *et al.*, 1988). The use of temporal distributions is usually applicable to event-based models, such as the SCS-SA (Schulze *et al.*, 1992). However, they can also be used in rainfall disaggregation approaches (Boughton, 2000b).

Rainfall disaggregation refers to producing high-resolution data that can be aggregated to give values equal to known coarser-scale totals. The use of high-resolution rainfall data inherently accounts for the temporal distribution of rainfall intensity. This is because the incremental time-steps are small enough, *i.e.* hourly or sub-hourly, so as to represent different intensities. High-resolution rainfall data are often required as input into continuous simulation hydrological models. It is important to note that disaggregation is not synonymous with downscaling. Downscaling aims at producing finer scale time-series with the required statistics, like disaggregation, but do not necessarily add up to any coarser-scale totals (Koutsoyiannis, 2003). Downscaling is, in particular, used for hydrological applications of general circulation models.

Continuous simulation hydrological models are important tools when analysing complex hydrologic or hydraulic problems where detrimental effects may occur at different timescales, for example in flood prediction and the assessment of water quality (Mikkelsen *et al.*, 1998). These models require detailed rainfall data, *viz.* hourly or sub-hourly. The advantage of such a time-series is that they reflect all relevant rainfall characteristics from peak intensities with short duration to variations in annual rainfall (Mikkelsen *et al.*, 1998). However, data are generally only widely available at more aggregated levels of the model time-step, such as daily. Koutsoyiannis and Onof (2001) note that in many countries, the number of raingauges providing hourly or sub-hourly resolution data is smaller than the number of daily gauges by about an order of magnitude. This situation reflects a general relative paucity of rainfall data for timescales of one hour or less, both in numbers of gauges and length of the recorded series (Koutsoyiannis and Onof, 2001). This, too, is the case in South Africa where it is reported that there are 172 recording gauges with at least 10 years of breakpoint data (Smithers and Schulze, 2000a), compared to 1806 daily rainfall stations with at least 40 years of data (Smithers and Schulze, 2000b). The need for a model to disaggregate daily rainfall into a sequence of individual storms of finer timescale cannot be overemphasised (Gyasi-Agyei, 1999).

The objectives of this study are:

- to identify techniques for the disaggregation of daily rainfall to produce hourly increments which aggregate to equal the observed daily values; and
- to select and apply one of the abovementioned techniques and assess the applicability of the method in South Africa.

In order to identify techniques for the disaggregation of daily rainfall a literature review is conducted in Chapter 2. After the selection of an appropriate methodology, the selection of various locations to test the methodology is made, and is detailed in Chapter 3. An in-depth description of the employed methodology is given in Chapter 4. This, however, enables the disaggregation of only those sites for which short duration rainfall data are available. In order to apply the methodology at locations where no short duration data are available, but daily rainfall records exist, it becomes necessary to regionalise the methodology. Chapter 5 provides a detailed description of the techniques used to regionalise the disaggregation model. Chapter 6 is a presentation and discussion of the various results obtained during the

application of the disaggregation model for when at-site short duration are available, as well as when regionalised input is used. Finally, conclusions drawn from the study and recommendations for future research are presented in Chapter 7.

## 2. METHODS TO ESTIMATE THE TEMPORAL DISTRIBUTION OF RAINFALL

There are a number of methods for estimating the temporal distribution of rainfall. This chapter is divided into two parts, the first of which is a review of techniques for the development of temporal rainfall distributions and the second part is a review of various approaches for disaggregating course scale rainfall.

From the literature, there are two broad categories into which a methodology for deriving temporal rainfall distributions may be classed, shown in Figure 2-1. As discussed in Section 2.1, one method is to develop the distributions with the use of Intensity-Duration-Frequency (IDF) curves, while methodologies based on observed hyetographs are reviewed in Section 2.2.

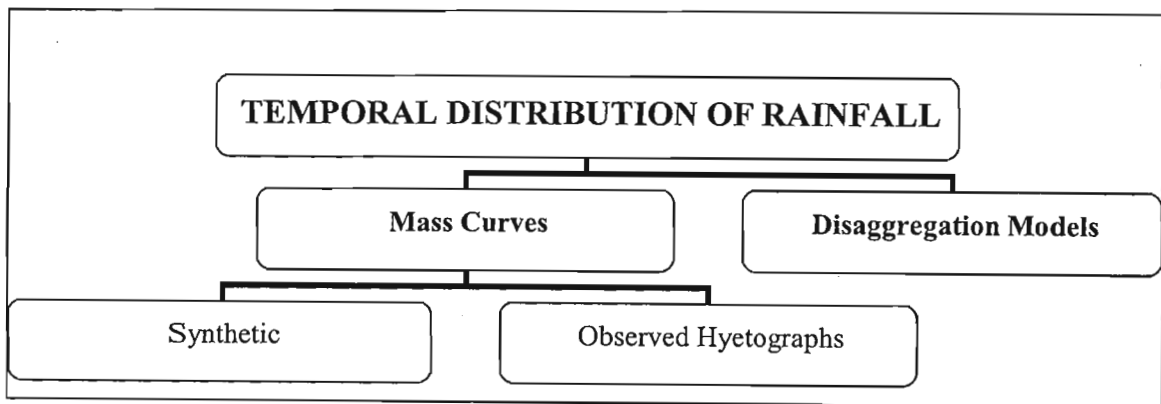


Figure 2-1 Schematic diagram depicting two broad categories into which the methodology for deriving temporal rainfall distributions may be classed

Several different types of models for disaggregating course scale rainfall and their applications are described in Section 2.3. This includes an alternative modelling methodology for rainfall disaggregation known as artificial neural networks (ANNs), which have been described as “elegant and powerful tools for solving problems” (Burian *et al.*, 2000).

### 2.1 Design Hyetographs Derived from Intensity-Duration-Frequency Relationships

The determination of the intensities or depths for selected return periods, which are inversely related to the exceedance probability, of extreme storm events for a variety of storm durations

are necessary for design flood estimation in South Africa. Weddepohl (1988), Smithers and Schulze (2000a) and Görgens (2002) cite several rainfall IDF studies in southern Africa, which include the widely applied Midgley and Pitman (1978) co-axial regional Depth-Duration-Frequency-diagram and the more recently developed regional approach, developed by Smithers and Schulze (2003), which utilises the scaling characteristics of rainfall.

In order to develop a temporal distribution for design rainfall, methods based on design IDF curves have been proposed and applied (Hoang, 2001). These methods include, *inter alia*, the alternating block method and the instantaneous intensity method (Chow *et al.*, 1988), a slightly modified alternating block method used by Boughton (2000a) and the South African version of the Soil Conservation Service (SCS-SA) method (Schulze *et al.*, 1992).

### 2.1.1 Temporal storm distributions

A simple means of developing a design hyetograph from an IDF curve is by way of the alternating block method (Chow *et al.*, 1988). In this method, the storm duration ( $T_d$ ) is divided into  $n$  equal time increments of  $\Delta t$  over a total duration  $T_d = n\Delta t$ . For a selected return period, the rainfall intensity is obtained from the IDF curve for each of the durations  $\Delta t$ ,  $2\Delta t$ ,  $3\Delta t$ , ...,  $n\Delta t$ , and the corresponding rainfall depth computed (Chow *et al.*, 1988). By taking differences between successive rainfall depths, the amount of rainfall to be added for each additional unit of time  $\Delta t$  is found. These incremental rainfall depths are subsequently reordered such that the maximum depth occurs at the centre of the storm duration  $T_d$ . Finally, the remaining incremental rainfall depths are arranged in descending order alternately on either side of the maximum depth to form the design hyetograph (Chow *et al.*, 1988).

A variation of the alternating block method is the instantaneous intensity method. The instantaneous intensity method employs a similar principle to that employed in the alternating block method, in that the rainfall depth for a time interval around the storm peak is equal to the depth given by the IDF curves (Chow *et al.*, 1988). The only difference is that the rainfall intensity is considered to vary continuously throughout the storm (Chow *et al.*, 1988).

In the development of a model to disaggregate daily rainfall to hourly rainfall, Boughton (2000a) also modified the alternating block method to construct temporal distributions. Prior

to the development of the methodology, Boughton (2000a) analysed data from recording gauges at the site of interest and identified the main characteristics of the temporal patterns, which formed the basis for modifications to the alternating block method.

### 2.1.2 SCS-SA temporal storm distributions

To adequately design hydraulic structures, long periods of observed streamflow data are required to estimate design peak discharges and volumes. When observed streamflow data are not available or are inadequate at the site of interest, rainfall/runoff models may be employed to estimate the streamflow. One such event-based model that has become accepted and established for use on small catchments in South Africa is the SCS method (Schulze *et al.*, 1992). The distribution of rainfall intensities is one of the primary inputs into such a model.

The temporal distribution types used in the SCS model have evolved somewhat over the years. Originally, two 24-hour storm distributions (Type I and Type II) were developed in the USA (Chow *et al.*, 1988). These were subsequently re-evaluated by Cronshey (1982) and more intense rainfall distributions were identified. The two original SCS distributions were provisionally adopted for use in South Africa (Schulze and Arnold, 1979). Schulze (1984) made revisions to the adopted distributions used in southern Africa, which gave rise to four revised synthetic distributions, termed Types 1, 2, 3 and 4, and were regionalised for South Africa. As shown in Figure 2-2, the Type 1 distribution contains the lowest design intensities, typifying frontal or general rainfall, while the Type 4 distribution contains the highest design intensities, typifying convective thunderstorms (Schulze *et al.*, 1992). The regionalised maps of the distributions were subsequently revised (Figure 2-3) by Weddepohl (1988) with the use of a database of digitised rainfall data for 40 rainfall stations located throughout South Africa (Schmidt and Schulze, 1987).

This methodology is considered to be a conservative approach (Görgens, 2002) in that the distributions are made up of extreme rainfall depths for each sub-duration centred on the middle of 24 hours. This is because it is unlikely that, for different durations, individual rainfall intensities will correspond to the design intensities (Schulze *et al.*, 1992).

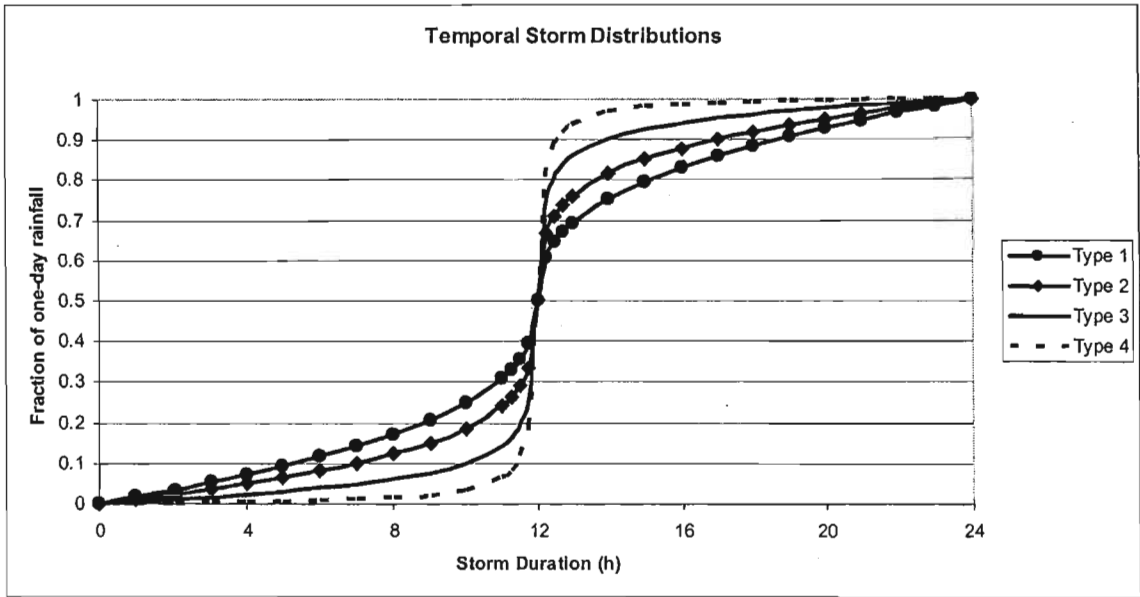


Figure 2-2 Synthetic design rainfall distributions for South Africa (after Schmidt and Schulze, 1987)

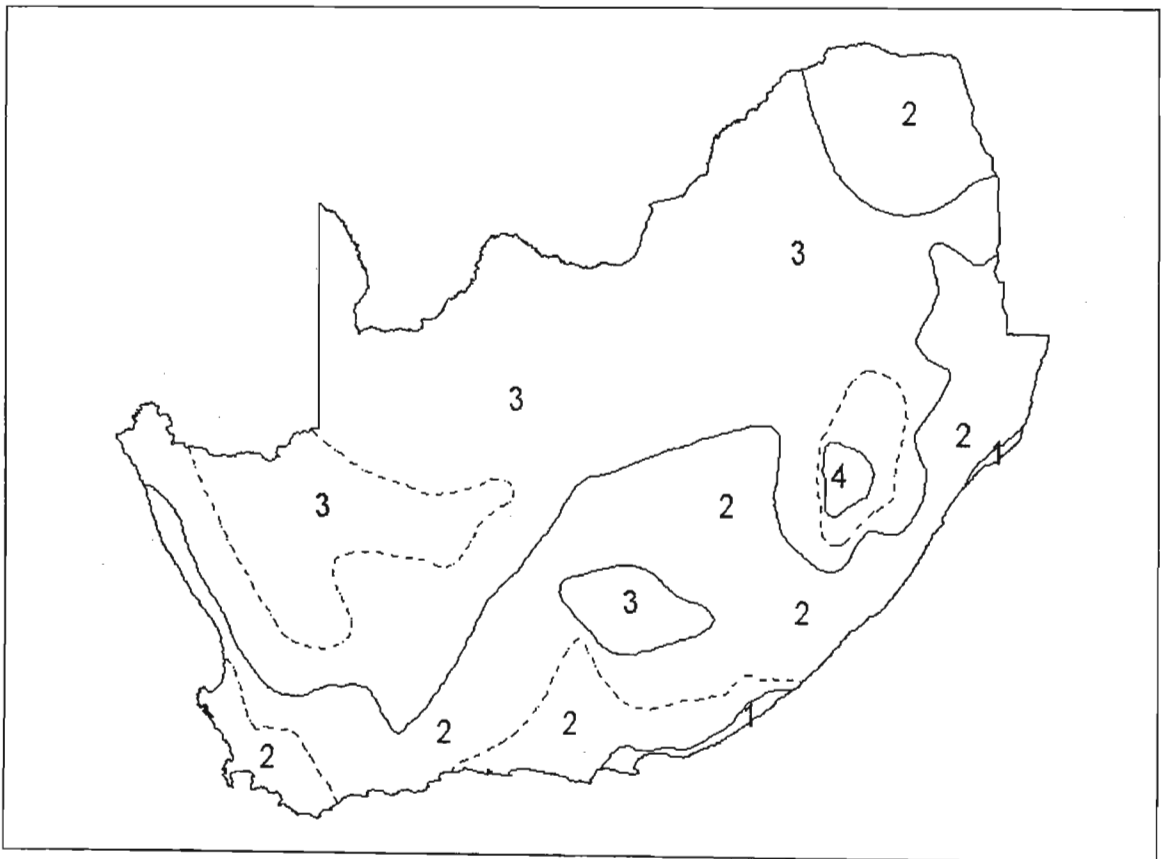


Figure 2-3 Spatial distribution of design rainfall intensity distribution types over South Africa (after Weddepohl, 1988)



Although methods for developing design temporal distributions from IDF curves are simple, there are disadvantages to this approach. Firstly, the patterns are derived from a series of unrelated intensities from a variety of storms and are, therefore, unrealistic of an actual storm (Pilgrim and Cordery, 1975; Cordery *et al.*, 1984; Hoang, 2001). Secondly, these methods fail to represent the variability of rainfall intensity within real storms (Boughton, 2000b; Hoang, 2001). Connolly *et al.* (1998) state that, while useful for estimating design runoff, methods based on rainfall IDF relationships are not appropriate for disaggregating long daily records.

Since the abovementioned methods have been criticised as being unrepresentative of actual storms, a solution may be to base the methodology for deriving temporal rainfall distributions on analyses of observed rainfall events. Methods that are based on analyses of observed hyetographs are reviewed in Section 2.2.

## **2.2 Hyetographs Derived from Observed Rainfall Data**

By analysing observed storm events, the temporal distribution of typical storms can be determined (Chow *et al.*, 1988). Using this approach, patterns of complete storms rather than intense bursts of rainfall have been derived (Pilgrim and Cordery, 1975; Cordery *et al.*, 1984). Pilgrim and Cordery (1975) cite many researchers who have adopted this approach. The methods discussed in this chapter include a triangular rainfall distribution (Yen and Chow, 1980), Huff curves (Huff, 1967), the average variability method (Pilgrim and Cordery, 1975) and the sampling of historical records.

### **2.2.1 Triangular rainfall distribution**

Although many variations of this methodology exist (Lambourne and Stephenson, 1986), Yen and Chow (1980) were the first to propose the triangular distribution for storm rainfall, as shown in Figure 2-4. Yen and Chow (1980) analysed 9869 storms at four locations in the USA and found the triangular rainfall distributions for most heavy storms to be nearly identical in shape. The method was based on the principle that any distribution can be determined once the precipitation depth and duration are known. In order to determine the location of the peak intensity, a storm advancement coefficient  $r$ , defined as the ratio of the

time before the peak to the total storm duration, is used. A suitable value of  $r$  is determined as the mean of the observed values of  $r$  computed for a series of storms of various durations, weighted according to the duration of each storm event (Yen and Chow, 1980).

Lambourne and Stephenson (1986) analysed a series of synthetic rainfall distributions, which included the uniform synthetic rainfall distribution, the Chicago synthetic hyetograph (Keifer and Chu, 1957) and the triangular rainfall distribution and concluded that the triangular distribution tends to represent actual storms more correctly than either the uniform synthetic rainfall distribution or the Chicago synthetic hyetograph.

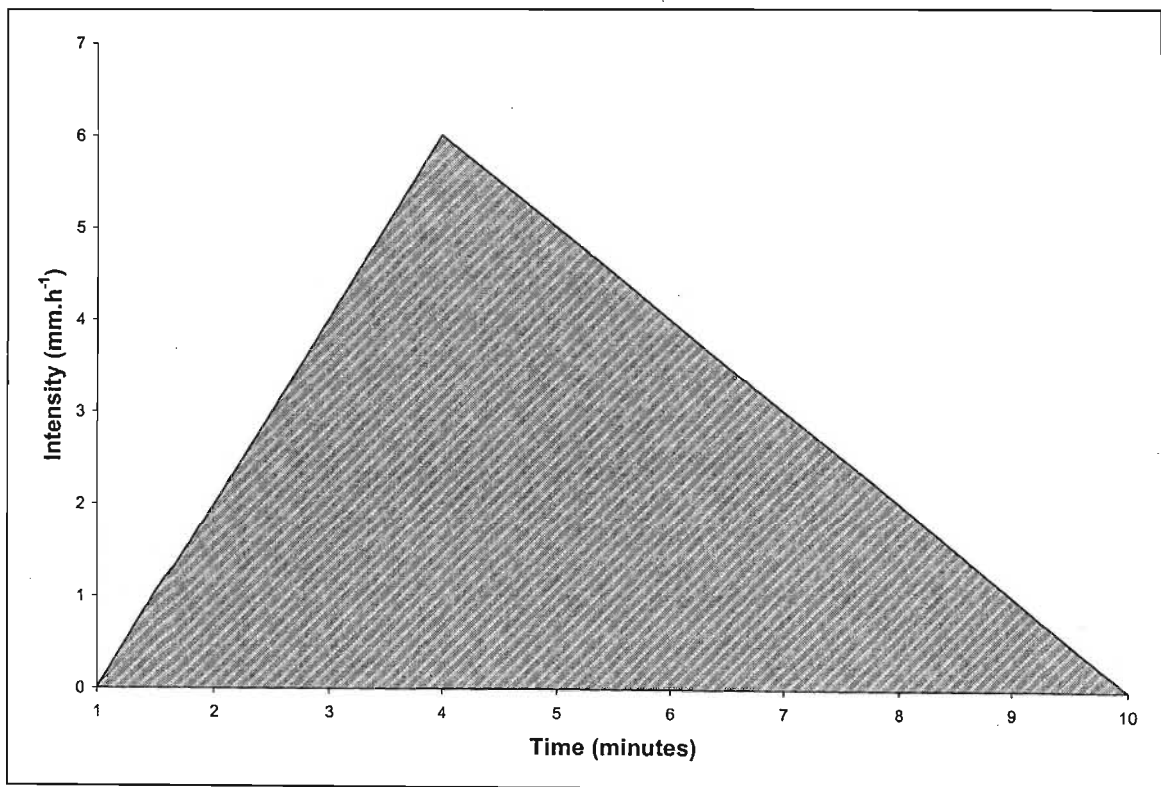


Figure 2-4 The triangular rainfall distribution (after Lambourne and Stephenson, 1986)

The main advantages of this method are the simplicity of use and relatively few parameters are required to establish the distribution (Lambourne and Stephenson, 1986). Hoang (2001), however, states that despite the simplicity of the method, simple hyetograph shapes are inadequate to represent the actual variation of rainfall intensity in typical rainfall events.

### 2.2.2 Huff curves

Huff (1967) analysed storm precipitation data using a dense network of raingauges in East Central Illinois in the USA from which 261 storms ranging in duration from 3 to 48 hours were studied. This study was undertaken primarily to provide information applicable to existing urban design problems (Huff, 1990). The resulting time distribution models were presented as probability distributions, as shown in Figure 2-5, to provide measures of both inter-storm variability and the general time pattern of rainfall. Huff (1967) found that of the total rainfall event, the major portion of rainfall occurs in a relatively small time increment of the total event time, regardless of duration, area or the number of identifiable separate bursts of rain. Measured storm hyetographs were separated into four groups based on the occurrence of the maximum storm intensity in one of four quartiles. These curves became known as “Huff curves”. Huff (1967) found that short duration storms of less than 24 hours duration were generally of the first or second quartile type, while storms of the duration greater than 24 hours were generally of the fourth quartile type.

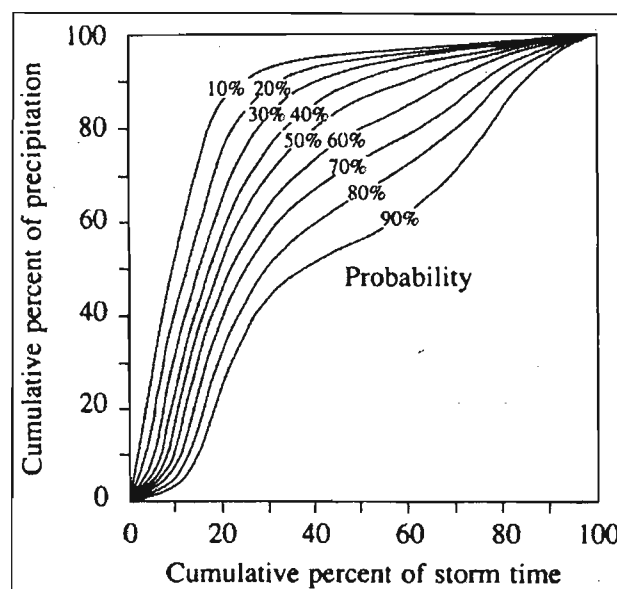


Figure 2-5 An example of first quartile Huff curves (Chow *et al.*, 1988)

Bonta and Rao (1987) investigated the effects of different methods of storm identification, sampling interval of precipitation data and season of year on the development of Huff curves. In general, it was found that Huff curves show little response to both changes in sampling interval and the critical duration between storms (Bonta and Rao, 1987). This is important

because it implies that the Huff curves can be developed from more widely available courser-scale rainfall data. However, individual Huff curves can show moderate influence of temporal pattern due to sampling interval, especially for less frequent (e.g. 10%) patterns (Bonta and Rao, 1987). It was also found that, as the probability associated with individual Huff curves within a quartile decreases, the differences between seasonal mass curves increase, with seasons having higher intensity rainfall displaying more variability within storms. However, after investigating the effects of seasonality on the development of Huff curves, Walker and Tsubo (2003) found that although there were some differences between the rainy and non-rainy seasons, statistical tests showed these differences were not statistically significant.

Bonta and Rao (1988) evaluated several mass curve approaches, *viz.* triangular, two superimposed triangles, three superimposed trapezoids, and Huff curves, in terms of their ability to represent the wide variability in measured hyetographs. Their results showed Huff curves to be the most flexible of the four hyetograph representations investigated. This can be attributed to their multiple-peaked nature and the manner in which the curves were developed (Bonta and Rao, 1988). A disadvantage of Huff curves, however, is that a curve associated with a selected probability must be selected.

Bonta and Rao (1989) investigated the possibility of regionalising Huff curves in the USA. Their findings showed that, despite the use of different methods to develop Huff curves for Ohio, Illinois and Texas, Huff curves from these three states showed remarkable similarity. These findings are supported by Walker and Tsubo (2003), who investigated the distribution of rainfall intensities within rainfall events using Huff curves, and made comparisons based on two different locations in South Africa (Bloemfontein and Pretoria). Results from this study indicated that one set of Huff curves could be used to represent a large portion of the South African Highveld (Walker and Tsubo, 2003).

### **2.2.3 The average variability method**

In most studies where the development of temporal distributions is based on observed raingauge records, it appears that considerable smoothing of the derived distribution occurs due to the averaging of rainfall intensities when considering an entire storm (Pilgrim and

Cordery, 1975). However, average patterns of rainfall are unlike those of most individual storms.

Pilgrim and Cordery (1975) thus derived a method for determining design temporal distributions based on the analysis of intense bursts of various durations, instead of considering complete storms. They developed a hyetograph analysis method which is based on ranking the time intervals in a storm by the amount of rainfall occurring in each interval, and repeating this procedure for many storms in the study area. By summing the ranks for each interval, a typical hyetograph can be derived. The design temporal patterns obtained in this manner represents the average variability of intense bursts of rain (Pilgrim and Cordery, 1975). The average variability method is the basis on which design temporal patterns currently used in Australia are derived (Hoang, 2001).

Regardless of its conceptual simplicity, the average variability method does not represent the variability of observed temporal patterns (Hoang, 2001). Furthermore, the method only accounts for intense bursts of rainfall and not for complete storms (Pilgrim and Cordery, 1975).

#### **2.2.4 Sampling from historical records**

To generate a design temporal distribution for a given design storm depth and duration, a very simple method is to sample temporal distributions from historical records (Hoang, 2001). This requires data with adequate lengths of record and at the required temporal resolution so that a large sample of all observed temporal distributions for the specified duration can be determined. It is then possible to randomly select the design temporal distribution for the defined event from the sample of dimensionless temporal patterns for the corresponding duration (Hoang, 2001).

Rahman *et al.* (2001) applied this method in an attempt to determine flood frequency curves resulting from events defined as storm cores, which are the period of maximum intensity within a storm. This project characterised the temporal time distribution by dimensionless mass curves divided into 10 equal time increments. The results showed that temporal patterns of rainfall depth for storm cores are not dependent on the season or total storm depth. Thus

dimensionless temporal patterns from different seasons and for different rainfall depths could be grouped. However, the patterns were found to be dependant on storm duration, yielding two groups: (1) up to 12 hours duration, and (2) greater than 12 hours duration (Rahman *et al.*, 2001).

The sampling of temporal distributions from historical records is a promising method because it is simple and it can model the variation of temporal patterns from event to event. However, one problem with this method is that it can only characterise patterns that are actually observed, but not the patterns that could have equally likely occurred (Hoang, 2001).

In this section various techniques for deriving temporal distributions through the analysis of observed rainfall events have been discussed. However, as mentioned, each of these methods has their inherent disadvantages. Most of these disadvantages can be overcome by applying rainfall disaggregation models, which have the ability to generate long sequences of short duration synthetic rainfall data while preserving important statistics pertaining to the observed courser-scale rainfall data for the respective site. Many different types of rainfall disaggregation models have been developed, some of which are reviewed in Section 2.3.

### **2.3 Rainfall Disaggregation Models**

When the main aim of hydrological modelling was to supplement design of various engineering structures, simple approximations such as the design storm approach proved to be sufficient (Arnbjerg Nielsen *et al.*, 1998). However, the recent shift of focus to sustainability and environmental protection requires that larger and more complex systems are analysed and thus long and accurate rainfall records are required. The use of single design storms are too simplistic as the detrimental effects in question may occur on many different time-scales (Arnbjerg Nielsen *et al.*, 1998).

Compared to the number of raingauges for recording daily rainfall, there is a relative paucity of rainfall data at sub-daily time-scales (Koutsoyiannis and Onof, 2001). The need for sub-daily rainfall data suggests the use of a disaggregation model to utilise available daily information and provide the user with possible realisations of sub-daily rainfall information, which aggregate up to the given daily data (Koutsoyiannis and Onof, 2001). Such a model

would be useful in continuous simulation applications, which in turn is important in design flood estimation (Hingray *et al.*, 2002).

Due to the limited success in deterministic rainfall modelling, stochastic point representations of rainfall have become more established in hydrology (Heneker *et al.*, 2001). The modelling of rainfall using stochastic techniques has a wide range of potential for hydrological applications which vary from hydrological design to the disaggregation of large time interval data into shorter durations (Onof and Wheater, 1993; Onof and Wheater, 1994). Many methodologies have been developed for disaggregating rainfall data temporally, some of which are discussed in the subsequent subsections.

### **2.3.1 Stochastic models based on dimensionless hyetographs**

Boughton (2000b) describes a stochastic model that disaggregates daily rainfall to hourly rainfall, while retaining the daily total and contains the statistical characteristics of hourly rainfalls at the gauge site. The primary part of the model is the distribution  $R$ , which is the fraction of the daily total that occurs in the hour of maximum rainfall. A random number is used to sample from the distribution of  $R$ . The sample value of  $R$  determines the other 23 values, which then undergo a clustering procedure, together with the value for  $R$ , in order to best maintain the statistics for 2, 3, 6 and 12 hour durations. This clustered sequence is then arranged into 1 of 24 possible temporal arrangements, depending on when the hour the maximum rainfall occurs. The structure of the model allows for the production of 480 different temporal distributions with variation between uniform and non-uniform rainfall.

Boughton (2000b) states that the model does not attempt to reproduce the hourly rainfalls that formed an actual daily total of rain, but rather is intended to be used in design flood estimation procedures, particularly in combination with a daily rainfall generation model. Tests of the model performance showed the methodology to be adequate in reproducing IFD statistics at each of the stations tested (Boughton, 2000b).

Woolhiser and Osborn (1985) presented a stochastic model for the disaggregation of an individual storm's depth into fractional depths, each corresponding to one tenth of the storm's duration. Their scheme was based on a dimensionless Markov process, resulting from

successive transformations of the real rainfall process (Koutsoyiannis, 2003). Garcia-Guzman and Aranda-Oliver (1993) proposed a model to disaggregate the total depth of a rainfall event of a specified duration into hourly rainfall depths. The temporal distribution was characterised by a dimensionless storm mass curve, whose ordinates were assumed to be ordered samples from a beta distribution, the parameters of which were estimated from observed storm data. Similar techniques using dimensionless hyetographs were also applied by Hershenhorn and Woolhiser (1987), Heneker *et al.* (2001), Loukas (2002) and Walker and Tsubo (2003).

### 2.3.2 Rectangular pulse cluster based stochastic rainfall models

Stochastic rainfall models based on point processes have been one of the most widespread and useful tools in the analysis and modelling of rainfall (Koutsoyiannis and Mamassis, 2001). Among them are cluster based models such as the Neyman-Scott rectangular pulse (NSRP) model and the Bartlett-Lewis rectangular pulse (BLRP) model, on which considerable research on cluster based modelling has been focussed (Rodriguez-Iturbe *et al.*, 1987; Cowpertwait, 1991; Onof and Wheater, 1993; Smithers and Schulze, 2000a; Gyasi-Agyei, 2005). Cluster based models have proved to be an elegant and physically realistic way to describe temporal rainfall (Olsson and Burlando, 2002). In cluster based models, storms are modelled as clusters of rain cells and each cell is a pulse with a random duration and random intensity, which is constant for the duration of the pulse (Smithers and Schulze, 2000a; Frost *et al.*, 2004). In these models, both storms and the origin of cells for each storm arrive follow a Poisson process (Koutsoyiannis and Mamassis, 2001; Koutsoyiannis, 2003). The two abovementioned cluster based models have slight differences (Entekhabi *et al.*, 1989). Frost *et al.* (2004) cited (Cowpertwait, 1998) who showed analytically that that the two models were equivalent up to second order statistics. Smithers and Schulze (2000a) found that the modifications of the BLRP model were able to reproduce the characteristics of shorter duration rainfall in South Africa, even when the model was calibrated using only daily rainfall data.

Glasbey *et al.* (1995) developed a system based on the BLRP model to estimate realistic hourly increments using daily rainfall. They achieved this by stochastically generating rainfall values at an hourly time-step and then choosing a three-day time-series, from the generated series, that best matched the measured daily rainfall prior to, on and following the day to be



disaggregated. Although this method produced satisfactory results, it has been pointed out that it is not theoretically justified as three-day periods of rainfall are not independent, which raises the issue of the applicability of the method to other data (Koutsoyiannis, 2003).

Koutsoyiannis and Onof (2000) developed a rainfall disaggregation model (HYETOS), which is a combination of the BLRP model, repetition techniques and adjusting procedures developed by Koutsoyiannis and Manetas (1996), the purpose of which was to generate and adjust hourly values of rainfall so as to sum to daily values. To validate the model Koutsoyiannis and Onof (2001) performed two case studies for two raingauges with extremely different climatic conditions, viz. a wet region (MAP > 600 mm) and a semi-arid region (MAP < 300 mm). The results of the case studies indicate a good performance of the methodology in preserving the most important statistical characteristics of the rainfall process, including intermittency (Koutsoyiannis and Onof, 2001; Koutsoyiannis, 2003).

Gyasi-Agyei (2005) developed a disaggregation model which incorporates the abovementioned repetition techniques and adjusting procedures (Koutsoyiannis and Onof, 2000; Koutsoyiannis and Onof, 2001) into a regionalised hybrid model. A case study at one location showed that the model was capable of reproducing near perfect dry probability, variance, autocorrelation, and IFD curves for all months. The author notes that some of the techniques may be seen as ad hoc and more consistent techniques need to be explored. Furthermore, the model needs to be tested in different climatic regions for general applicability (Gyasi-Agyei, 2005).

### **2.3.3 Multiple site rainfall disaggregation models**

The spatial distribution of rainfall is important for agricultural purposes, in the evaluation of regional hydrological behaviour and estimation of catchment flows, and in the assessment and simulation of flood events (Kottegoda *et al.*, 2003). Multiple site rainfall disaggregation, as a means of spatial and temporal disaggregation, is a relatively new approach to rainfall disaggregation. It presents significant differences from that of single-site disaggregation, including increased mathematical complexity (Koutsoyiannis, 2003).

Socolofsky *et al.* (2001) presented a simple method to disaggregate daily rainfall for use in continuous simulation modelling. The disaggregation model makes use of the 8000 hourly gauges and 25000 daily gauges in the Charles River catchment, Massachusetts, USA. This method relies on measured hourly data from the same climatological regime as the daily data to be disaggregated and samples the measured hourly data directly, as opposed to sampling analytically derived statistical distributions. Despite its simplicity, the method is shown to perform well in reproducing intermittency and in preserving the characteristics of the seasonal rainfall process (Socolofsky *et al.*, 2001). Using the disaggregation technique, the efficiency of the runoff model (ranging from  $-\infty$  to 1, where 1 is perfect agreement and  $-\infty$  is very poor agreement) for annual runoff volume improved from 0.36, obtained when using hourly data outside the catchment, to 0.81 when using disaggregated daily rainfall data from inside the catchment (Socolofsky *et al.*, 2001).

Jennings *et al.* (2002) modified the point rainfall model DRIP (Disaggregated Rainfall Intensity Pulse), developed by Heneker *et al.* (2001). The resultant model is capable of simulating long-term synthetic high-resolution rainfall data at numerous sites using short-term historical data (Jennings *et al.*, 2002).

Kottegoda *et al.* (2003) set out to develop a parsimonious model to simulate daily rainfall data and disaggregate these values to hourly values. Using a multiple site approach, the authors disaggregated daily rainfall values into hourly values through dimensionless accumulated hourly amounts generated by a beta distribution. Application of the model to the Tiber River catchment in central Italy showed this approach to satisfactorily reproduce extremes and other statistical properties of daily and hourly rainfall (Kottegoda *et al.*, 2003).

Koutsoyiannis *et al.* (2001) developed a model called MuDRain (Multivariate Disaggregation of Rainfall) for spatial-temporal disaggregation of rainfall. MuDRain involves the combination of several univariate and multivariate rainfall models operating at different time-scales in a disaggregation framework that can appropriately modify outputs of finer time-scale models so as to become consistent with given coarser time-scale series. Application of this model showed good preservation of important properties of the rainfall process, including good reproduction of actual hyetographs (Koutsoyiannis, 2003).

### 2.3.4 Artificial neural networks

Another approach to rainfall disaggregation modelling is to use Artificial Neural Networks (ANNs). These models have become frequently used in solving complex non-linear problems in many professions. ANNs are elegant and powerful tools for solving difficult problems (Burian *et al.*, 2000). They are massive parallel distributed processors made up of simple processing units (artificial neurons), shown in Figure 2-6, which have a natural propensity for storing experimental knowledge (Nagesh Kumar, 2003; Walker and Tsubo, 2003). In the past, a persistent criticism of ANNs has been that they are black boxes whose inner mechanisms are not well understood. However, properly constructed ANNs are no more mysterious than the countless other regression and optimisation tools commonly used in everyday engineering practice (Burian *et al.*, 2000).

Before neural networks can be applied to a specific problem, certain decisions need to be made. Firstly, an appropriate neural network type must be chosen; and secondly, an appropriate training algorithm, suitable training periods and an appropriate network structure must be determined (Dawson and Wilby, 2001).

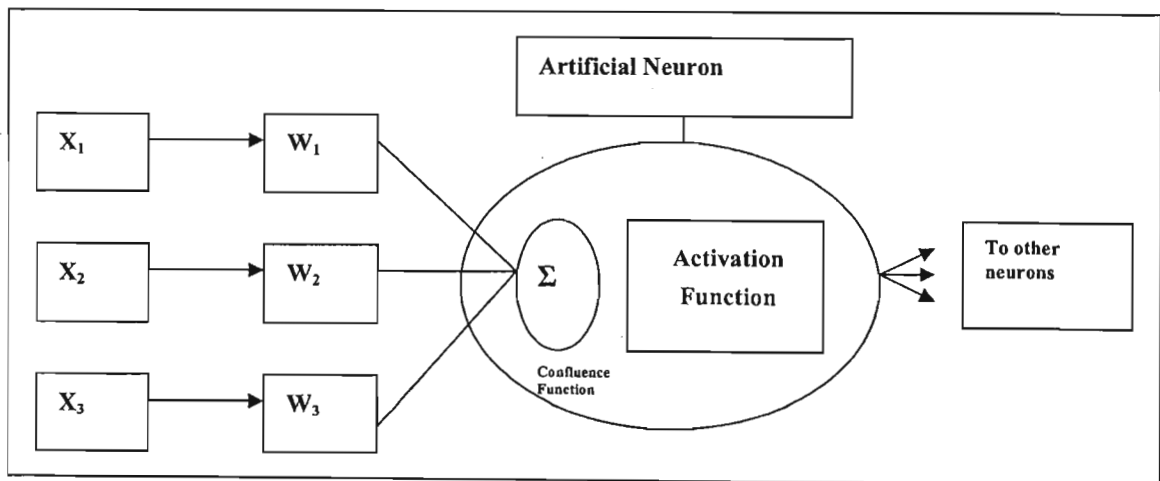


Figure 2-6 Artificial neuron or processing unit (after Burian *et al.*, 2000), where  $X_i$  denotes the inputs and  $W_i$  denotes the weights

ANNs may be described as a network of interconnected neurons, each consisting of inputs and outputs. Each processing unit finds a weighted sum of the inputs using a confluence

function and evaluates an activation function to compute its output (Burian *et al.*, 2000). The network is trained by adjusting the weights that link its neurons (Dawson and Wilby, 2001). The training is applied with a training set of data, where the network is presented with a set of inputs and corresponding outputs, and reduces the error in its output for a given input and target output (Walker and Tsubo, 2003). For more information on hydrological modelling using ANNs, the reader is referred to Dawson and Wilby (2001).

#### **2.3.4.1 Advantages and Disadvantages of ANNs**

ANNs offer the following advantages:

- They are flexible mathematical structures that are capable of identifying a complex non-linear relationship between input and output data sets (Walker and Tsubo, 2003).
- They can handle incomplete and ambiguous data (Maier and Dandy, 1996).
- All the existing complex relationships between various aspects of the process under investigation need not be known (Nagesh Kumar and Srinivasa Raju, 2000).
- They are often cheaper and simpler to implement than their physically based counterparts (Campolo *et al.*, 1999).
- They are well suited to dynamic problems and are parsimonious in terms of information storage within the trained model (Thirumalaiah and Deo, 1998).
- It has been shown that they can be trained using data from a source that is a considerable distance from the station of interest (Burian *et al.*, 2001).

Although ANNs have proven to be potentially useful tools in hydrology, their disadvantages should not be ignored. A major limitation of ANNs is the lack of physical concepts and relations (Nagesh Kumar, 2003). Most ANN applications are unable to explain the way they arrive at a decision. Another issue is that there is no standardised way of selecting network architecture. The choice of network architecture, training algorithm, and definition of error are usually determined by the user's past experience and preference, rather than the physical aspects of the problem (Nagesh Kumar, 2003). Koutsoyiannis (2003) states that, because ANNs are not based on probability and stochastic processes, they may not be appropriate for large length simulations, as they do not perform well in extrapolation. For example they may result in poor estimation of extreme events, where long periods of record are required.

#### 2.3.4.2 Neural network applications

Neural networks have been applied to solve many problems over the last decade. Specific applications to hydrology and water resources engineering include rainfall-runoff modelling, precipitation estimation based on remotely sensed data, river stage forecasting and water demand forecasting (Burian *et al.*, 2000).

Burian *et al.* (2000) developed two ANN models for disaggregating long-term hourly rainfall records into 15-minute increments. One was a feed-forward model trained by back-propagation. The other model used a similar network, but based on the idea of competitive learning, which uses feedback among the hidden layer (Burian *et al.*, 2000). The performance of the trained ANN models were compared to a linear disaggregation method and the continuous-deterministic rainfall disaggregation model proposed by Ormsbee (1989). The performance evaluation showed the ANNs to perform consistently on the same level as the linear and Ormsbee approaches. However, the ANN models were more accurate in predicting specific characteristics of the individual rainfall hyetographs, these being the maximum 15-minute rainfall depth and the time when this rainfall occurred. This is important as the maximum incremental rainfall depth and the time the depth occurs in a storm event has a significant effect on the resulting hydrograph (Burian *et al.*, 2000).

Burian and Durrans (2002) showed how the abovementioned feed-forward model trained by back-propagation (Burian *et al.*, 2000) can improve rainfall-runoff modelling. The runoff hydrographs produced by the ANN model rainfall patterns were compared to the runoff hydrographs produced by the observed 15-minute rainfall patterns, the observed hourly-increment rainfall patterns, and the rainfall patterns from a geometric similarity disaggregation technique. The peak discharges produced by the ANN model rainfall patterns proved to be more accurate than when using the observed hourly increment rainfall and the pattern produced by the geometric similarity rainfall disaggregation model (Burian and Durrans, 2002).

Nagesh Kumar and Srinivasa Raju (2000) used ANNs to disaggregate predicted monsoon (seasonal) rainfall to monthly rainfall for Orissa state, India. The correlation coefficient between the observed and the disaggregated value was 0.703. It was further stated that the ability to disaggregate seasonal rainfall to monthly rainfall could be used for short term

planning and operation of river basin projects, which will enhance the planning for sustainable development.

Walker and Tsubo (2003) applied ANNs to estimate rainfall intensities at 15-minute increments. Meteorological data, which included specific humidity, temperature, wind velocity and direction, were used as an input to estimate the rainfall intensity. After a trial and error process, the best ANN model consisted of one hidden layer, which in turn had 288 artificial neurons. Results from the modelling process reflected some negative values, which were then adjusted to give zero values, which agreed with the actual measured values. However, these adjustments posed a further problem in that the actual daily rainfall was no longer consistent with the daily rainfall sums from the generated rainfall intensities. Walker and Tsubo (2003) found that, more often than not, the simulated daily value was greater than the actual daily value because each simulated rainfall that was originally a negative value had been increased by its absolute value to zero. It was decided that the ANN model was not acceptable as a short-duration rainfall intensity generator at the present time, but showed potential for future modelling of rainfall intensity if the current problems could be overcome (Walker and Tsubo, 2003).

## 2.4 Discussion and Conclusions

Four methods for developing synthetic temporal rainfall distributions were reviewed, *viz.*, the alternating block method, the instantaneous intensity method, a variation of the alternating block method developed by Boughton (2000a) and the SCS-SA (Schulze *et al.*, 1992). These methods are based on IDF data. Although these methods for developing temporal rainfall distributions are simple, they fail to characterise actual storms. This is because they are derived from a series of unrelated design events and fail to represent the variability of rainfall intensity within real storms. However, these methods do not attempt to represent actual storms, but are rather to be used as a tool for design purposes (Schulze *et al.*, 1992; Boughton, 2000a).

Methodologies for deriving temporal rainfall distributions based on observed hyetographs were considered to counter the shortfalls of the abovementioned methodologies. The methods reviewed were the triangular distribution (Yen and Chow, 1980), Huff curves (Huff, 1967), the average variability method (Pilgrim and Cordery, 1975) and the sampling of historical

records. The triangular distribution, although simple to use, is unable to represent the actual variation of rainfall intensity in rainfall events. Huff curves are better in this regard as they are divided into quartiles, depending on when the major burst of rainfall occurs, and are presented as a set of probability curves. They have been shown to be flexible and studies in both North America and South Africa indicate that they can be regionalised. One disadvantage in the use of Huff curves is that a choice needs to be made from a range of curves (Bonta and Rao, 1988), and hence their use is not clear.

The average variability method, by only accounting for intense bursts of rainfall, cannot be representative of entire storms. As the name suggests, this method does not represent the variability of possible distributions, but rather an average variability of many storms (Pilgrim and Cordery, 1975). These deficiencies are well accounted for if distributions are derived from historical records. This methodology results in distributions for entire rainfall events, and accounts for the natural variability over time. However, one possible problem with the average variability method is that it can only characterise patterns that are actually observed, but not the patterns that could have equally likely occurred (Hoang, 2001).

Most of the disadvantages inherent in the abovementioned methods can be overcome by applying rainfall disaggregation models, which have the ability to generate long sequences of synthetic rainfall data for short durations, while preserving important statistics pertaining to the observed longer duration rainfall data for the respective site. Of the rainfall disaggregation techniques discussed, the multiple site models seem to have the most potential in hydrological applications. This is because they can be applied to derive spatially consistent high-resolution rainfall series at sites where only daily data are available.

ANNs are powerful tools with the potential for solving complex problems in hydrology, such as rainfall disaggregation. ANNs have been shown to successfully disaggregate seasonal rainfall to monthly rainfall and hourly rainfall to 15-minute rainfall. Although ANNs have disadvantages, they have the potential to be useful tools in hydrology.

Of the techniques reviewed, Boughton's (2000b) methodology showed promising results when estimating extreme values in Australia. It is also a simple approach and lends itself to the possibility of being utilised as a multiple site model. Boughton's (2000b) methodology for disaggregating daily rainfall to hourly values was selected in this study for use in South

Africa. However, Boughton's (2000b) methodology only accounts for all rainfalls where the daily total is greater than or equal to 15 mm. Hence, the methodology will need to be modified to account for all days on which rainfall occurred.



### 3. DATA UTILISED IN THE STUDY

Smithers and Schulze (2000a) compiled a short duration ( $\leq 24$ h) rainfall database from several different organisations, viz., the University of KwaZulu-Natal (UKZN); the Council for Scientific Research and Industrial Research (CSIR); the South African Sugar Research Institute (SASRI); the South African Weather Service (SAWS); the Cape Town City Engineer's Department (CTCE); Rhodes University (RU) and the University of Zululand (UZ). These data were made available by Smithers and Schulze (2003) as daily blocks of either 24 hourly values, or 96 quarter-hourly values, extracted for 0:00 to 0:00 periods.

#### 3.1 Selection of Hidden Stations for Model Testing

Hourly data from 172 stations, all of which had record lengths greater than 10 years, were available for use. It was necessary to exclude some stations from the model development process in order to independently evaluate the model. One station from each of the 15 relatively homogeneous extreme rainfall clusters, identified and described by Smithers and Schulze (2000a), was removed from the dataset and not used in the development of the disaggregation model. The 172 stations were divided into their respective homogeneous clusters and the station with the median record length in each cluster was removed. This resulted in 15 test stations, the details of which are listed in Table 3-1 and the locality shown in Figure 3-1.

Thus 157 stations remained to be used for model development, and details of these gauges are listed in Appendix A. The locations of the 157 stations used for model development are shown in Figure 3-2 and the distribution of their length of record is shown in Figure 3-3.

Table 3-1 Details of the 15 “hidden” stations used for model testing

Organisation	Station No.	Location	Years Record	Latitude			Longitude			Altitude (m)	Mean Annual Precipitation (mm)
				Deg	Min	Sec	Deg	Min	Sec		
SAWS	0435019	OTTOSDAL	20	26	49	0	26	1	0	1498	559
SAWS	0552581	OUDESTAD	18	25	11	0	29	20	0	953	609
UKZN	C173	CEDARA	20	29	33	50	30	15	0	1143	866
SAWS	0317474	UPINGTON	25	28	24	0	21	16	0	836	176
SAWS	0719370	MARNITZ	27	23	10	0	28	13	0	932	391
SAWS	0023710	ROBERTSON	25	33	50	0	19	54	0	159	272
UZ	304474	KWA-DLANGZWA	12	28	54	0	31	46	0	32	1292
SASRI	Sacfs	UMHLANGA	20	29	43	0	31	3	0	76	915
SAWS	0028748	GEORGE	18	33	58	0	22	25	0	221	606
SAWS	0092288	BEAUFORT WEST	23	32	18	0	22	40	0	893	188
SAWS	0474680	CARLETONVILLE	19	26	20	0	27	23	0	1500	660
SAWS	0261516	BLOEMFONTEIN	31	29	6	0	26	18	0	1351	514
SAWS	0127272	UMTATA WO	21	31	32	0	28	40	0	742	608
SAWS	0258213	DRIELOTTE	29	29	3	0	24	38	0	1120	404
SAWS	0106880	VREDENDAL	35	31	40	0	18	30	0	37	141

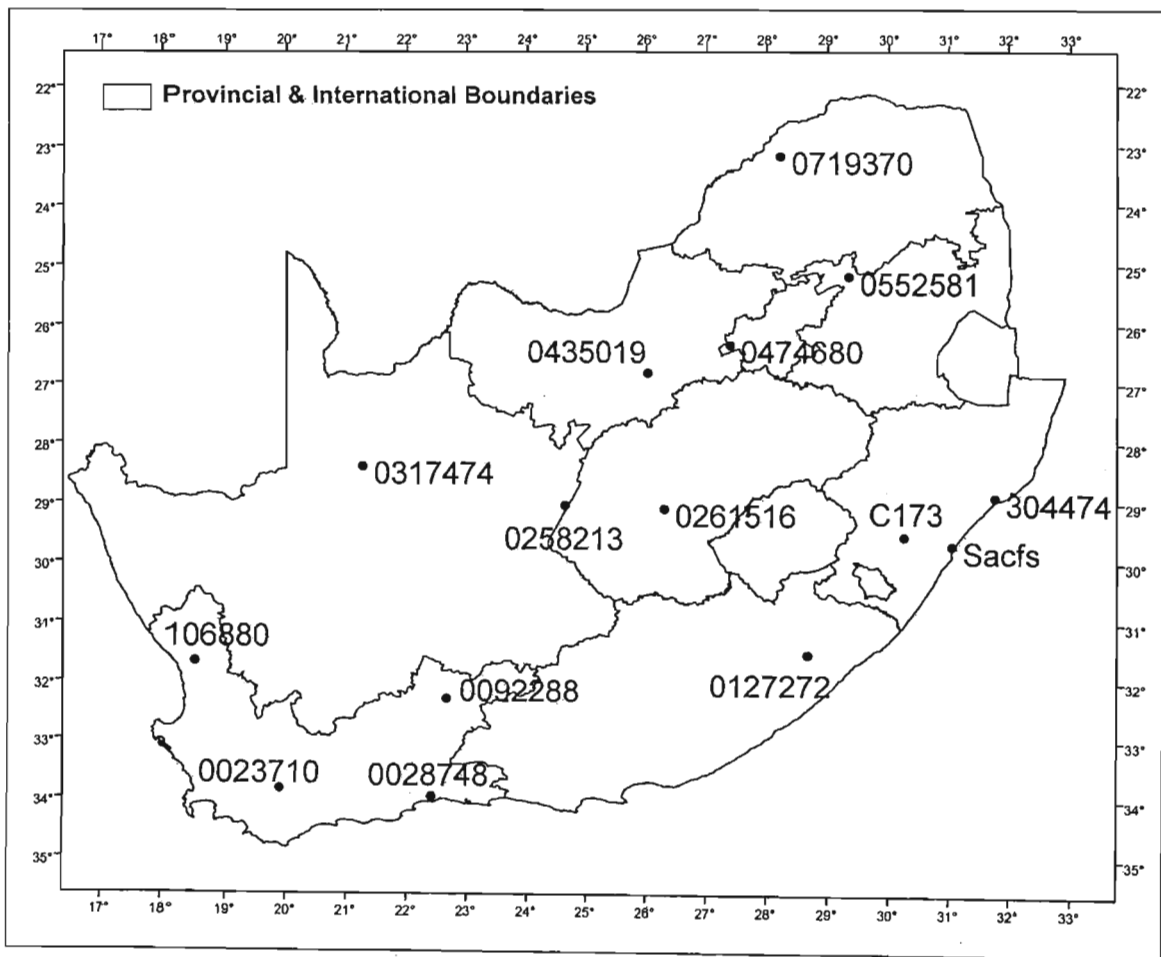


Figure 3-1 Locations of the 15 “hidden” stations used for model testing

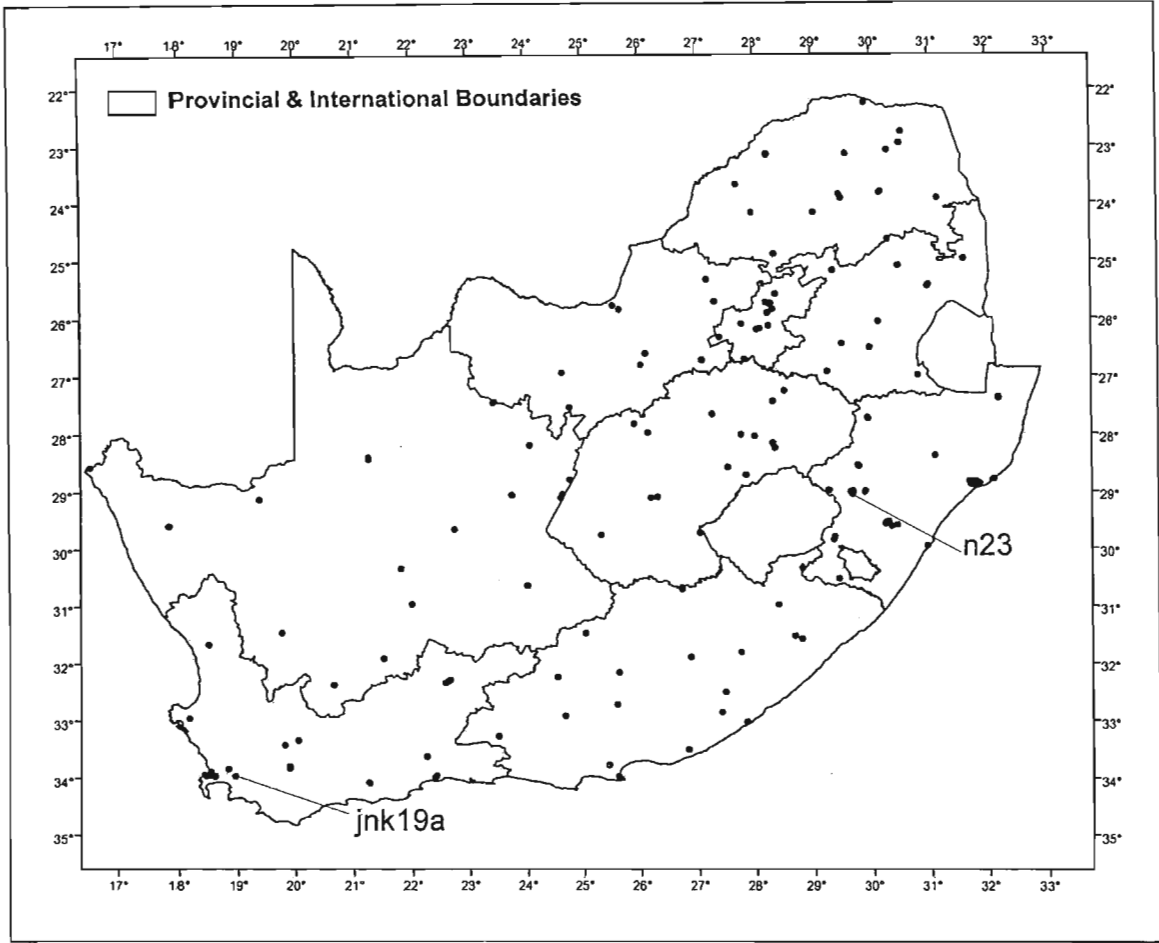


Figure 3-2 Locations of stations used for model development

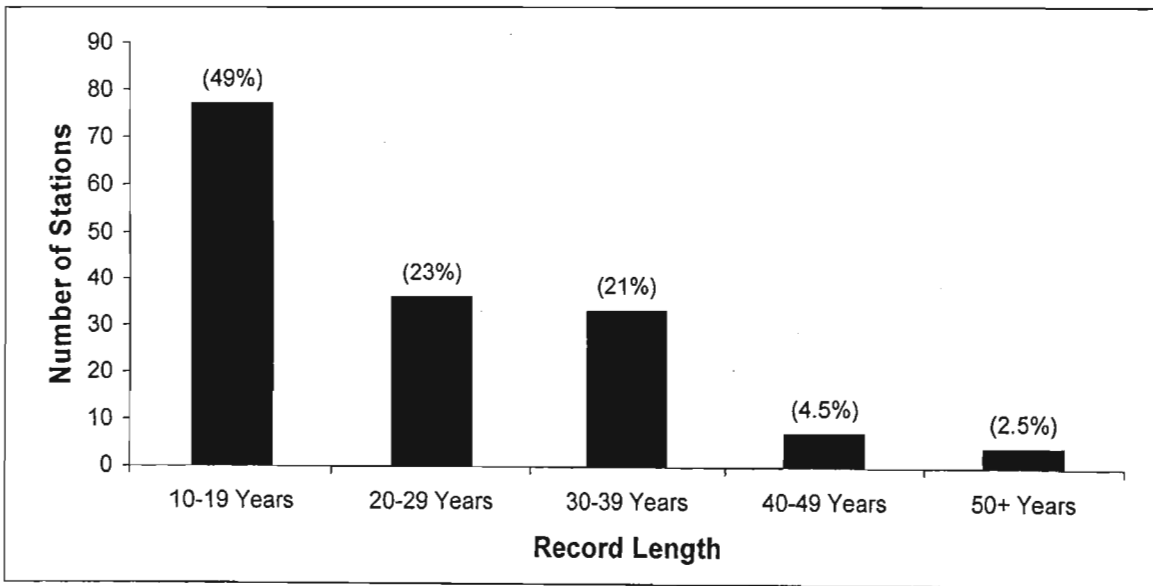


Figure 3-3 Distribution of record lengths for the data used in model development

### 3.2 Data Selection

Unlike the work done by Boughton (2000b), which was to disaggregate only the larger daily rainfalls considered important in flood studies, the purpose of this study was to disaggregate all daily rainfalls. In Boughton's (2000b) study only rainfall data greater than, or equal to, 15 mm were used to develop his model. However, Boughton (2005) states that the methodology can be applied to all daily non-zero rainfalls.

For the purpose of this study it was decided to develop the model in order to disaggregate all daily non-zero rainfalls. All hourly data, from the remaining 157 stations, for which the daily rainfall total was greater than or equal to 1 mm were used in the development of the disaggregation model, i.e. distributions calculated using all data for which the daily rainfall total was greater than or equal to 1 mm will be used to disaggregate all daily non-zero values. The choice of using all data for which the daily rainfall total was greater than or equal to 1 mm to develop the model is justified in Section 6.1. Furthermore, only data where all 24 hours of the day were successfully recorded were used, i.e. all days that contained missing hours or infilled data were discarded. Boughton (2000b) states that the model treats each day as an independent event for disaggregation, so the selection of days for analysis does not require whole months or whole years of data. Details of the model developed by Boughton (2000b) are contained in the following chapter.

## 4. THE DISAGGREGATION MODEL

The daily to hourly disaggregation model used and modified in this study is based largely on the work done by Boughton (2000b). The details of the model developed by Boughton (2000b) are described in this chapter and changes to the methodology developed by Boughton (2000b) are highlighted.

### 4.1 Structure of the Model

The model is comprised of 4 main parts:

- a) The distribution of the fraction of the daily total,  $R$ , that occurs in the hour of maximum rainfall.
- b) For each value of  $R$  there is an average set of values for the other 23 hourly fractions of the daily total.
- c) Given the 24 fractions from above, the values are clustered to maintain the observed average highest 2-hour, 3-hour, 6-hour and 12-hour fractions.
- d) These clusters are then arranged into random patterns so as to reproduce the variations in daily temporal patterns while retaining the abovementioned statistics.

### 4.2 Distribution of $R$

The primary part of the disaggregation model is the fraction,  $R$ , of the daily total that occurs in the hour of maximum rainfall. A value of  $R = 1.0$  indicates that all of the rainfall on the day fell in a single hour. This is the upper limit of  $R$  and is the boundary of non-uniformity. Completely uniform rainfall throughout a day would yield  $R = 0.04167$  (i.e.  $1/24$  of the daily total). This is the lower limit of  $R$ .

A single day's rainfall in hourly increments at Raingauge N23 at Ntabamhlope in the KwaZulu-Natal (KZN) midlands is shown in Figure 4-1. The daily total was 84.4 mm and the hour in which the most rainfall fell contained 40.9 mm. This yields a ratio of  $R = 40.9/84.4 = 0.48$  for the day.

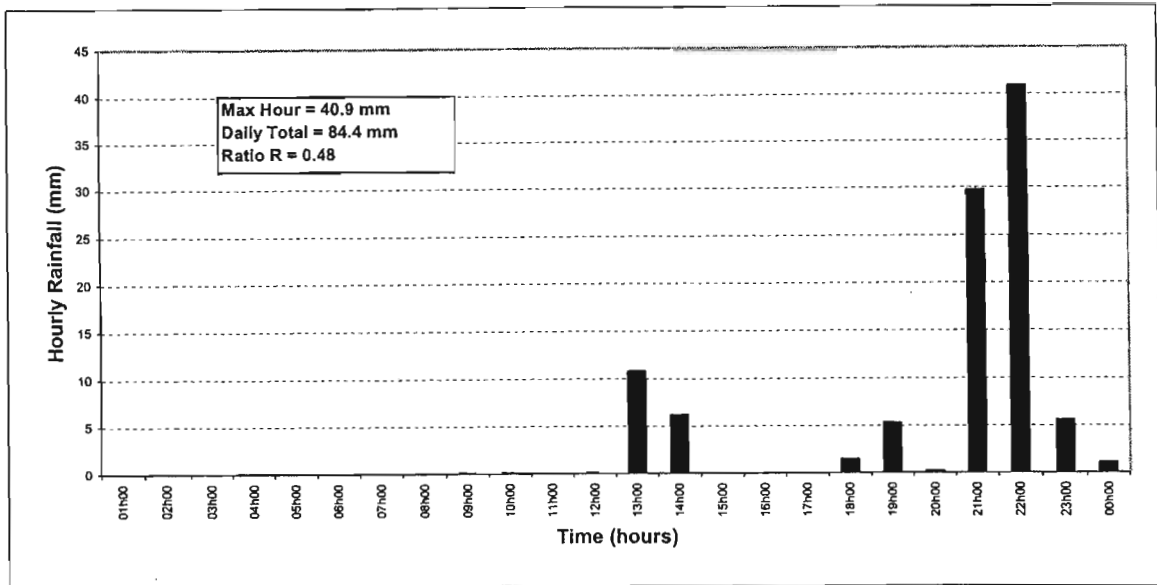


Figure 4-1 Example of a single day's hourly rainfall at Ntabamhlope

If the ratio R is determined for all days with 1 mm of rainfall or more, the distribution of R has a pattern that is a major characteristic of hourly rainfall at the site. The distribution of R for a particular site is created by extracting all the values of R at the site for days where the rainfall was greater than or equal to 1 mm, for the entire length of record. The computed R values are then collated into 20 ranges, which were chosen by Boughton (2000b) and are shown in Table 4-1. The distribution of R thus shows the proportion of all values of R in each of the ranges.

Table 4-1 Ranges used when collating R values

No.	Range	No.	Range	No.	Range	No.	Range
1	[0.0417-0.075)	6	[0.275-0.325)	11	[0.525-0.575)	16	[0.775-0.825)
2	[0.075-0.125)	7	[0.325-0.375)	12	[0.575-0.625)	17	[0.825-0.875)
3	[0.125-0.175)	8	[0.375-0.425)	13	[0.625-0.675)	18	[0.875-0.925)
4	[0.175-0.225)	9	[0.425-0.475)	14	[0.675-0.725)	19	[0.925-0.975)
5	[0.225-0.275)	10	[0.475-0.525)	15	[0.725-0.775)	20	[0.975-1.000]

The distributions of R for two sites in differing climates are shown in Figure 4-2. Jonkershoek (Station Jnk19a), in the Western Cape, is located in a winter rainfall region whereas Ntabamhlope (Station N23), which lies inland in KZN, is located in a summer rainfall region. The locations of these stations are highlighted in Figure 3-2.

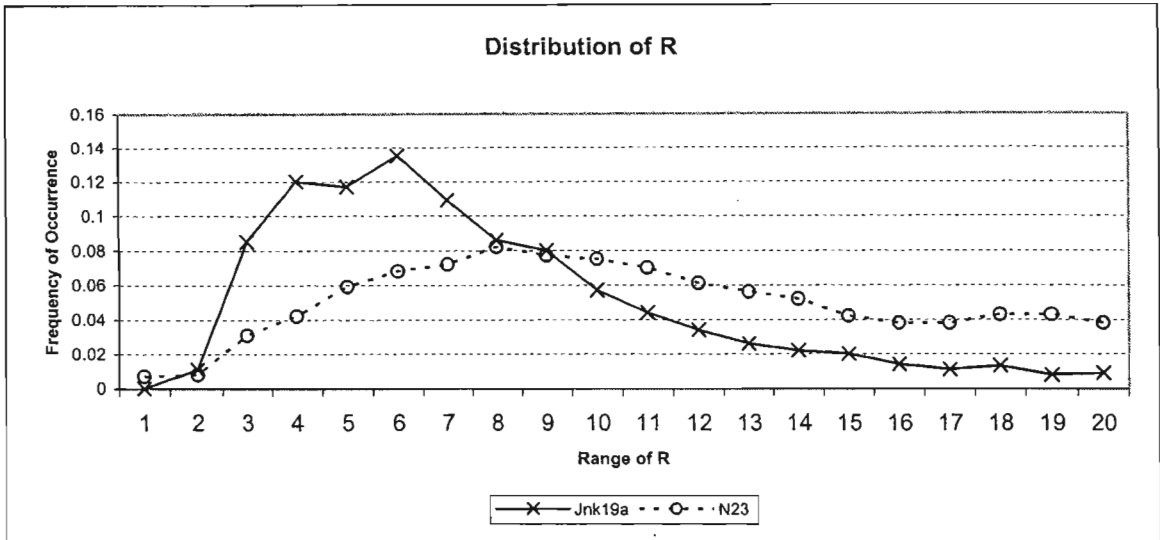


Figure 4-2 Frequency distributions of R at Stations Jnk19a (in the W. Cape) and N23 (in KZN), using the ranges of R shown in Table 4-1

From Figure 4-2 it is evident that the majority of the days at Jonkershoek fall into Range 6 in Table 4-1 and have many low values of R (mean R = 0.385) indicating that there is a tendency for more uniform rainfall. The distribution of R at Ntabamhlope (mean R = 0.537) shows a larger proportion of the days having larger values for R. This indicates that at Ntabamhlope large portions of the daily rain fall in a single hour, which is typical of the convective storms in the summer rainfall region.

### 4.3 Calculating the Other 23 Hourly Fractions

If  $R = 1.0$  then all of the rainfall fell in a single hour, hence the other 23 hourly fractions must be 0. If  $R = 0.04167$  then the other 23 hourly fractions must equal 0.04167. If, however, R is slightly less than 1.0 it is probable that the rest of the day's rainfall fell in 1 or 2 other hours, resulting in the remaining 21 or 22 hours having zero rainfall. Conversely, if R is slightly greater than 0.04167 then the other 23 values will be slightly less than, but close to, 0.04167. This is important to note as it indicates that the value of R has a strong influence in determining the other 23 hourly fractions of rainfall.

In order to determine the other 23 hourly fractions, the 24 hourly fractions for every day on record were ranked in order of magnitude, with R being the largest value on each day. This was done for each of the 157 stations. These ranked series, from all 157 sites, were then

pooled together and averaged in each of the 20 ranges of R shown in Table 4-1. This resulted in 20 averaged ranked series of hourly fractions, one for each range of R, examples of which are shown in Figure 4-3. The averaged ranked series for all 20 ranges are listed in Appendix B. The procedure followed when calculating the 20 averaged ranked series is detailed in Equations 4.1 to 4.6.

$$\text{Let } p_{(h,k)} = \text{rainfall that occurred in hour } h \text{ on day } k, \quad \dots 4.1$$

$$P_{(k)} = \text{the daily total rainfall on day } k = \sum_{h=1}^{24} p_{(h,k)} \quad \dots 4.2$$

$$\text{Therefore, the hourly fraction of the daily total for hour } h \text{ on day } k = f_{(h,k)} = \frac{p_{(h,k)}}{P_{(k)}} \quad \dots 4.3$$

The 24 hourly fractions of the daily total for a particular day can therefore be represented by the following matrix:

$$f_{(k)} = [f_{(1,k)}, f_{(2,k)}, f_{(3,k)}, \dots, f_{(24,k)}] \quad \dots 4.4$$

The elements of  $f_{(k)}$  are then ranked in order of magnitude to give:

$$r_{(k)} = \text{descending ranked series of } f_{(k)}, \text{ i.e. } r_{(1,k)} = \text{MAX}(f_{(k)}), r_{(24,k)} = \text{MIN}(f_{(k)}) \quad \dots 4.5$$

where

$$r_{(j,k)} = \text{ranked hourly fraction of the daily total rainfall for rank } j, \text{ day } k.$$

These ranked series are then separated into 20 different groups of data. This is achieved using the value of R on the respective day and the 20 ranges shown in Table 4-1.

$$\overline{r_{(i,j)}} = \frac{1}{N_{(i)}} \sum_{k=1}^{N_{(i)}} r_{(j,k)} \quad \dots 4.6$$

where

$$\overline{r_{(i,j)}} = \text{average hourly fraction of the daily total rainfall for range } i \text{ and rank } j,$$

$$\overline{r_{(1,j)}} = \text{average maximum 1-hour fraction ( } AM_{(i)}^{(1)} \text{), i.e. } \overline{R}, \text{ for range } i, \text{ and}$$

$$N_{(i)} = \text{total number of days in range } i.$$



Once all 24 averaged hourly fractions have been determined for each range of R they can be used to create daily temporal patterns of rainfall. The following two sections contain a description of how these 24 hourly fractions are arranged to recreate possible realisations of the temporal distribution of daily rainfall.

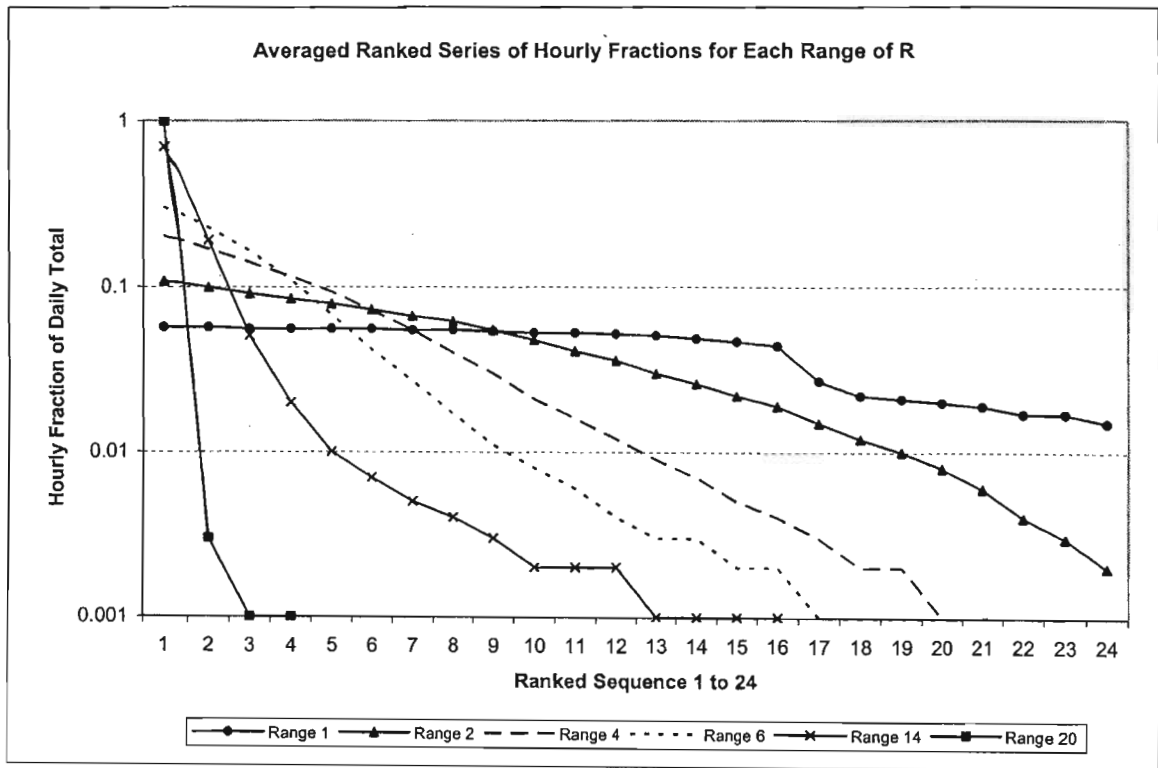


Figure 4-3 Averaged ranked series of hourly fractions for selected ranges of R, calculated using all 157 stations

#### 4.4 Clustering of Hourly Rainfalls

In order to cluster the 24 hourly fractions, the data from all stations were again processed to calculate the highest 2-hour fraction of the daily total ( $T_{(k)}^{(2)}$ ), the highest 3-hour fraction ( $T_{(k)}^{(3)}$ ), the highest 6-hour fraction ( $T_{(k)}^{(6)}$ ) and the highest 12-hour fraction ( $T_{(k)}^{(12)}$ ) on each day. As for the ranked series in Section 4.3, all of these fractions were then averaged within the range of R in which they occurred. This resulted in an average maximum 2-hour fraction ( $AM_{(i)}^{(2)}$ ), 3-hour fraction ( $AM_{(i)}^{(3)}$ ), 6-hour fraction ( $AM_{(i)}^{(6)}$ ) and 12-hour fraction ( $AM_{(i)}^{(12)}$ ) of the daily total for each of the 20 ranges of R (Table 4-2). Example equations demonstrating

how the average maximum 3-hour fraction was calculated are shown in Equations 4.7 to 4.10. Similar equations were used when calculating  $AM_{(i)}^{(2)}$ ,  $AM_{(i)}^{(6)}$  and  $AM_{(i)}^{(12)}$ .

Let the matrix  $\mathbf{t}_{(x,k)}^{(3)}$  represent the sequence of 3-hour rainfall for day  $k$ ,

$$\mathbf{t}_{(x,k)}^{(3)} = [t_{(1,k)}^{(3)}; t_{(2,k)}^{(3)}; t_{(3,k)}^{(3)}; \dots; t_{(20,k)}^{(3)}; t_{(21,k)}^{(3)}; t_{(22,k)}^{(3)}] \quad \dots 4.7$$

where

$x = 1$  represents the first 3-hour value, and  $x = 22$  represents the last 3-hour value on day  $k$ ,

$$t_{(x,k)}^{(3)} = \sum_{x=h}^{h+2} f_{(x,k)}, \text{ for } h = 1-22 \quad \dots 4.8$$

$$T_{(i,k)}^{(3)} = \text{MAX}(t_{(x,k)}^{(3)}) \quad \dots 4.9$$

$$AM_{(i)}^{(3)} = \frac{1}{N_{(i)}} \sum_{k=1}^{N_{(i)}} T_{(i,k)}^{(3)} \quad \dots 4.10$$

where

- $AM_{(i)}^{(3)}$  = average maximum 3-hour fraction of the daily total for range  $i$ ,
- $T_{(i,k)}^{(3)}$  = maximum value of  $(\mathbf{t}_{(x,k)}^{(3)})$ , and
- $N_{(i)}$  = total number of days in range  $i$ .

Using the averaged ranked sequences computed using Equation 4.6, a computer program was written to check the sum of the first value in the ranked series with each of the other 23 hourly fractions in order to find which of the 23 values, when added to the first value, gave the best match to the average 2-hour fraction for the respective range of  $R$ . After fixing that value as the value to accompany the first value for the highest 2-hour fraction, the program then checks the remaining 22 hourly values to find which value should accompany the 2-hour fraction to form the average highest 3-hour fraction. The program then searches for the next 3 values to form the average highest 6-hour fraction, and then searches for the next 6 values to form the average highest 12-hour fraction. Performing this for each range of  $R$  resulted in 20 clustered sequences, shown in Appendix C. This process is detailed in Equations 4.11 to 4.14.

Table 4-2 Average highest 1-hour, 2-hour, 3-hour, 6-hour, and 12-hour fraction for each range of R, calculated using all 157 stations

	$AM_{(i)}^{(1)} (R)$	$AM_{(i)}^{(2)}$	$AM_{(i)}^{(3)}$	$AM_{(i)}^{(6)}$	$AM_{(i)}^{(12)}$
Range 1	0.057	0.114	0.170	0.335	0.648
Range 2	0.108	0.195	0.271	0.466	0.726
Range 3	0.154	0.266	0.357	0.564	0.783
Range 4	0.202	0.338	0.442	0.660	0.846
Range 5	0.251	0.408	0.520	0.733	0.885
Range 6	0.300	0.475	0.592	0.789	0.909
Range 7	0.350	0.539	0.657	0.826	0.923
Range 8	0.400	0.600	0.711	0.852	0.933
Range 9	0.450	0.662	0.761	0.877	0.944
Range 10	0.500	0.727	0.807	0.900	0.955
Range 11	0.549	0.768	0.837	0.913	0.960
Range 12	0.599	0.797	0.855	0.922	0.965
Range 13	0.649	0.827	0.877	0.934	0.969
Range 14	0.699	0.852	0.894	0.941	0.974
Range 15	0.749	0.877	0.910	0.950	0.977
Range 16	0.800	0.900	0.926	0.958	0.981
Range 17	0.850	0.924	0.942	0.966	0.985
Range 18	0.901	0.947	0.958	0.975	0.989
Range 19	0.950	0.971	0.977	0.987	0.995
Range 20	0.993	0.996	0.997	0.998	1.000

$$AM_{(i)}^{(1)} = \overline{r_{(i,j1)}} \quad \dots 4.11$$

where

$$j1 = 1$$

$$AM_{(i)}^{(2)} \approx \overline{r_{(i,j1)}} + \overline{r_{(i,j2)}} \quad \dots 4.12$$

where

$j1 \neq j2$ , and  $\overline{r_{(i,j2)}}$  is the fraction which, when added to  $\overline{r_{(i,j1)}}$ , closest approximates  $AM_{(i)}^{(2)}$ .

$$AM_{(i)}^{(3)} \approx \overline{r_{(i,j1)}} + \overline{r_{(i,j2)}} + \overline{r_{(i,j3)}} \quad \dots 4.13$$

where

$j_1 \neq j_2 \neq j_3$ , and  $\overline{r_{(i,j_3)}}$  is the fraction which, when added to  $\overline{r_{(i,j_1)}}$  and  $\overline{r_{(i,j_2)}}$ , closest approximates  $AM_{(i)}^{(3)}$ .

The  $AM_{(i)}^{(6)}$  and  $AM_{(i)}^{(12)}$  were approximated in the same way, resulting in a clustered sequence,  $CS_{(i)}$ , for each range of R whereby the average maximum 2-hour, 3-hour, 6-hour and 12-hour fractions of the daily total are approximated.

$$CS_{(i)} = [ \overline{r_{(i,j_1)}}, \overline{r_{(i,j_2)}}, \overline{r_{(i,j_3)}}, \dots, \overline{r_{(i,j_{24})}} ] \quad \dots 4.14$$

The next step was to arrange these clustered sequences into temporal patterns.

#### 4.5 Daily Temporal Patterns of Hourly Rainfalls

Schmidt and Schulze (1987) derived four design rainfall distributions to be used for different regions in South Africa. This suggests that a single distribution can be used to represent the temporal distribution of rainfall for a particular region. This, however, is not realistic and analysis of the rainfall data shows that there are several temporal patterns ranging from nearly uniform rainfall to highly variable rainfall. Furthermore, the peak intensity can occur during any hour of the day, adding to the variability of temporal rainfall patterns. In order to account for the variability of temporal patterns of rainfall, several temporal distributions should be employed.

The hour of day when the highest intensity rainfall occurred was determined for each day that rainfall occurred at each of the 157 stations. The results show a definitive distribution for the timing of peak rainfall occurrence for a particular location. As shown in Figure 4-4 for Station Jnk19a, the hour of maximum rainfall has a somewhat uniform distribution, indicating that the hour of maximum rainfall has a reasonably equal probability of occurring in any hour of a particular day. Station N23, however, has a sinusoidal-like distribution with the majority of days having the peak rainfall falling during the late afternoon and evening.

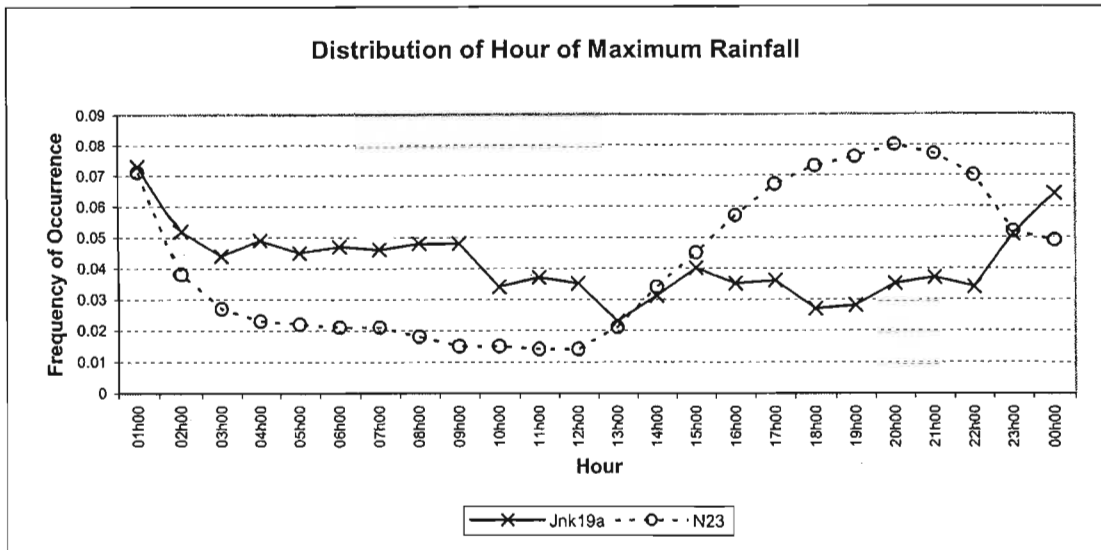


Figure 4-4 Frequency distributions of the hour of maximum rainfall at Station Jnk19a (Jonkershoek, in the W. Cape) and Station N23 (Ntabamhlope, in KZN)

In application, a random number is used to select the hour of maximum rainfall from the distribution of the hour of maximum rain for the site of interest. This differs from the work done by Boughton (2000b) as in that study no distinct distribution was found for the time of maximum rainfall and hence the hour of maximum rainfall was selected at random.

Using the clustered sequences established above and assigning the numerals “1” for the highest fraction ( $\overline{r_{(i,j1)}}$ ), “2” for the fraction that accompanies “1” to form the 2-hour fraction ( $\overline{r_{(i,j2)}}$ ), “3” for the fraction that accompanies “1” and “2” to form the 3-hour fraction ( $\overline{r_{(i,j3)}}$ ), etc., and then accounting for all permutations when the hour of maximum rainfall can occur, 24 arrangements of the clustered sequences can be created, as shown in Table 4-3.

The combination of these 24 arrangements with the 20 possible ranges of R results in 480 different temporal patterns, as opposed to one averaged distribution. These range from uniform to non-uniform with the possibility of the hour of maximum rainfall occurring in any hour of the day. Figure 4-5 contains a sample of the different temporal distributions that the disaggregation model produces.

Table 4-3 24 samples of temporal arrangements of the hourly rainfalls (after Boughton, 2000b)

Sample	Sequence																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
2	2	1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
3	3	2	1	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
4	6	5	4	1	2	3	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
5	6	5	4	2	1	3	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
6	6	5	4	3	2	1	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
7	9	8	7	6	5	4	1	2	3	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
8	9	8	7	6	5	4	2	1	3	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
9	9	8	7	6	5	4	3	2	1	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
10	12	11	10	9	8	7	6	5	4	1	2	3	13	14	15	16	17	18	19	20	21	22	23	24
11	12	11	10	9	8	7	6	5	4	2	1	3	13	14	15	16	17	18	19	20	21	22	23	24
12	12	11	10	9	8	7	6	5	4	3	2	1	13	14	15	16	17	18	19	20	21	22	23	24
13	15	14	13	12	11	10	9	8	7	6	5	4	1	2	3	16	17	18	19	20	21	22	23	24
14	15	14	13	12	11	10	9	8	7	6	5	4	2	1	3	16	17	18	19	20	21	22	23	24
15	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	16	17	18	19	20	21	22	23	24
16	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	1	2	3	19	20	21	22	23	24
17	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	2	1	3	19	20	21	22	23	24
18	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	19	20	21	22	23	24
19	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	1	2	3	22	23	24
20	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	2	1	3	22	23	24
21	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	22	23	24
22	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	1	2	3
23	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	2	1	3
24	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

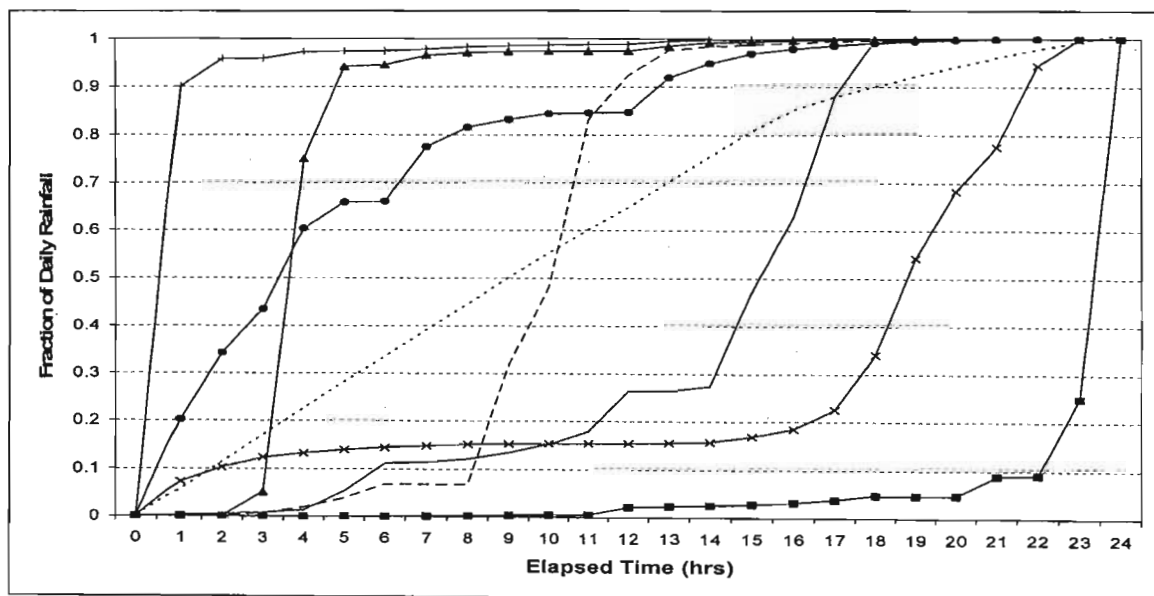


Figure 4-5 Samples of the different temporal distributions that are generated by the disaggregation model

## 4.6 Chapter Conclusions

The disaggregation model developed by Boughton (2000b) was developed only for design flood purposes, and was thus modified in order to achieve the objectives of this study. The modifications made to Boughton's (2000b) methodology in this study are related the disaggregation of all daily rainfalls, as opposed to only disaggregating the larger events, and the distribution of the time when the hour of maximum rainfall occurred. It was evident that the rainfall data at different stations displayed different distributions for the hour maximum rainfall. The distribution of the hour of maximum rainfall at each station was computed, and random sampling along the respective distributions was performed. This differs from the original Boughton (2000b) model, where the hour of maximum rainfall was determined by random sampling from a uniform distribution.

The resulting model uses the distribution of  $R$  and the distribution of the hour of maximum rainfall for each station in the selection of a temporal pattern. There are 20 ranges of the ratio  $R$  and these are combined with 24 patterns that allow the hour of maximum rainfall to occur during any of the 24 hours of the day. This combination gives  $20 \times 24 = 480$  different temporal patterns, which allows for variation between uniform and non-uniform rainfall.

This methodology, however, is designed to be applied at locations where short duration rainfall data are available. As mentioned in Chapter 1, there is relative paucity of short duration rainfall data compared to the more abundant daily rainfall data, both in the number of gauges and in length of the recorded series. This highlights the need for regionalisation. A regional approach attempts to supplement the limited information available with regional information from surrounding stations (Smithers and Schulze, 2000a). The following chapter is a description of the process of regionalising the disaggregation methodology described in this chapter.

## 5. REGIONALISATION

In order for the methodology, developed in the previous chapter, to be applied at a national scale, particularly at sites where only daily rainfall data are available, the need for regionalisation of the distributions of R becomes apparent. The following is a description of how the distributions of R were regionalised in South Africa and how the methodology can be applied at sites where only daily rainfall data are available.

As shown in Figure 4-2 it was found that the mean value of R for a particular site had a strong influence on the distribution of R for that site. It was thus decided that instead of creating one average distribution of R for the entire country, several distributions should be created to account for the variation in temporal rainfall distribution in South Africa.

It was found that the mean values of R ( $R_{\text{mean}}$ ) for each of the 157 stations used in this study fell between 0.385 at Jonkershoek (Station Jnk19a) and 0.639 at Pilanesberg (Station 0548290). Collating these 157 values and using the same ranges of R shown in Table 4-1, it was found that all but twelve stations had values of  $R_{\text{mean}}$  that fell within Ranges 9 to 12. Stations Jnk19a, Moko3a, Cp6 and C172 located as shown in Figure 5-1 with  $R_{\text{mean}} = 0.385$ , 0.406, 0.422 and 0.423 respectively, all fell into Range 8, but were included with the stations in Range 9 as there were no other stations with  $R_{\text{mean}}$  in Range 8. Before including these 4 stations in Range 9, the distributions of R for these stations were compared to the average distribution of all those stations that had an  $R_{\text{mean}}$  that fell in Range 9. Similar trends were observed, as shown in Figure 5-2, thus justifying their inclusion in Range 9.

Eight stations, viz., 0317476, 0092229, 0508047, 0631791, 0593489, 0809706, 0677802 and 0548290, located as shown in Figure 5-3, had  $R_{\text{mean}}$  values slightly larger than those in Range 12. As before, the distributions of R for these stations were compared with the average distribution of R for all the stations with an  $R_{\text{mean}}$  in Range 12 and are shown in Figure 5-4. Owing to the similarity it was decided that these stations could be included with the stations in Range 12 rather than creating another average distribution based on significantly fewer stations. Therefore, the ranges used for collating the  $R_{\text{mean}}$  values in South Africa needed to be changed accordingly and are shown in Table 5-1.



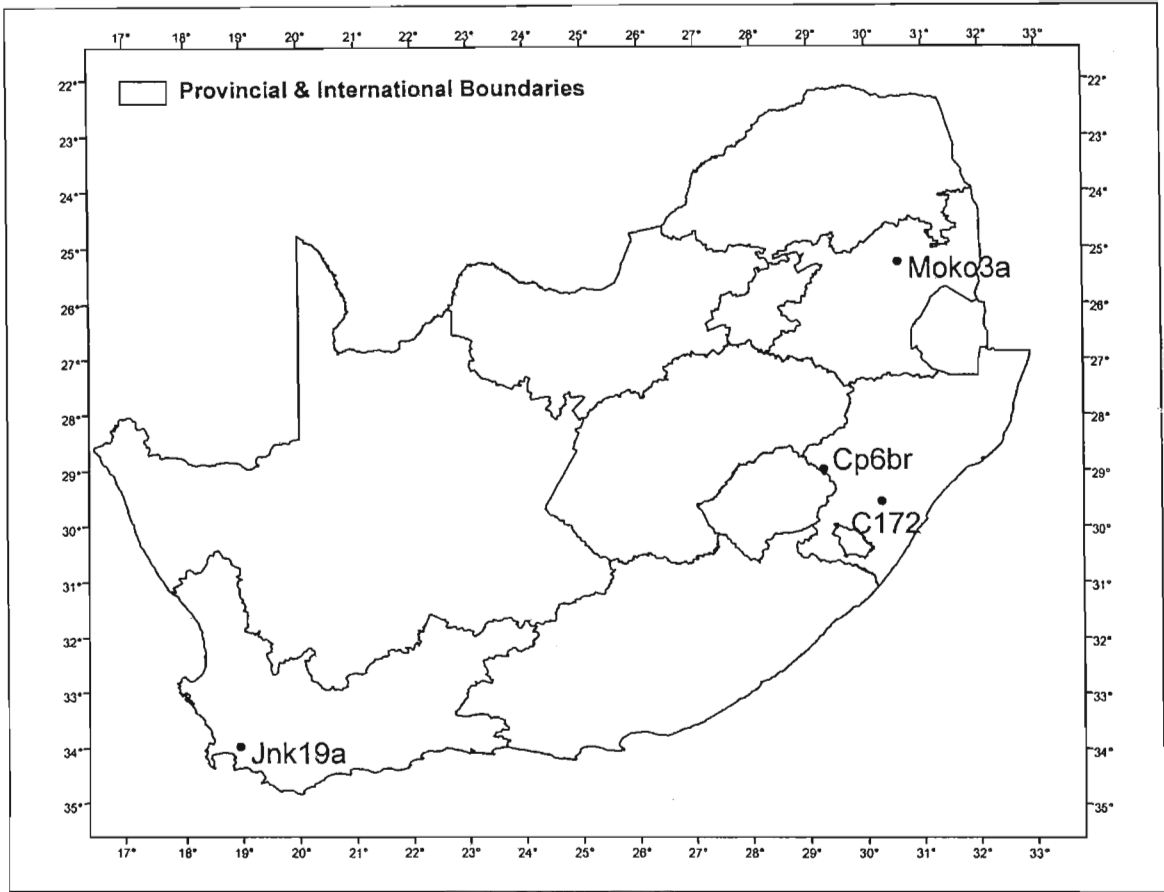


Figure 5-1 Stations that fell within Range 8 but were included in Range 9

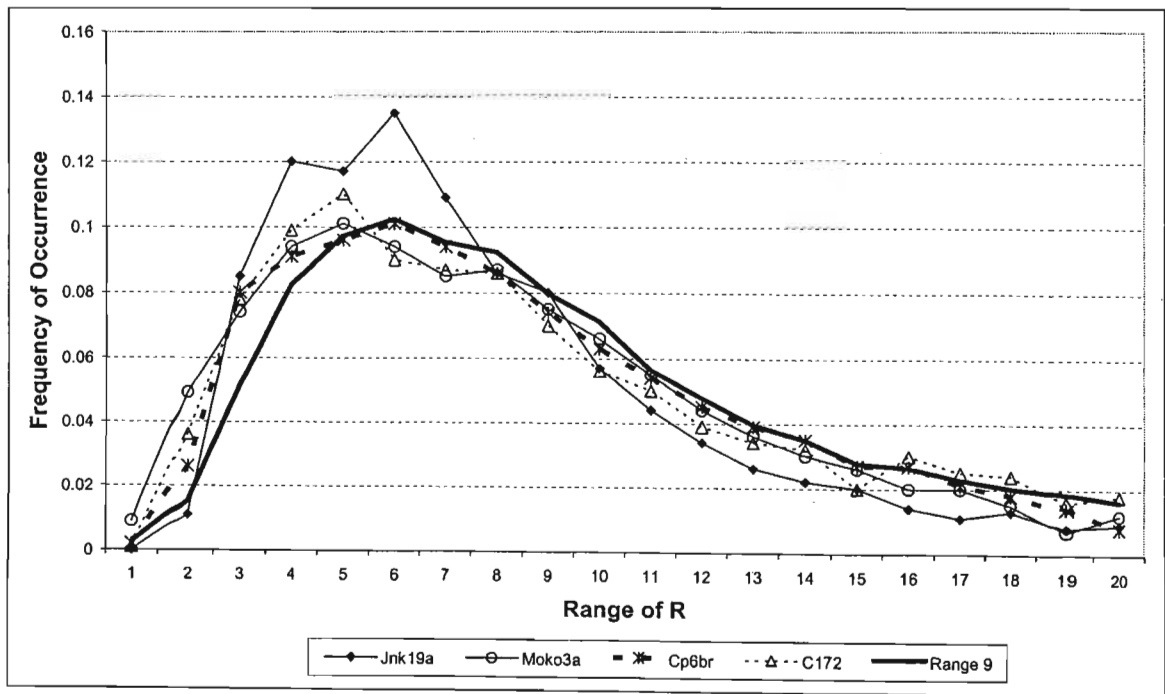


Figure 5-2 Distributions of R for sites included with those in Range 9

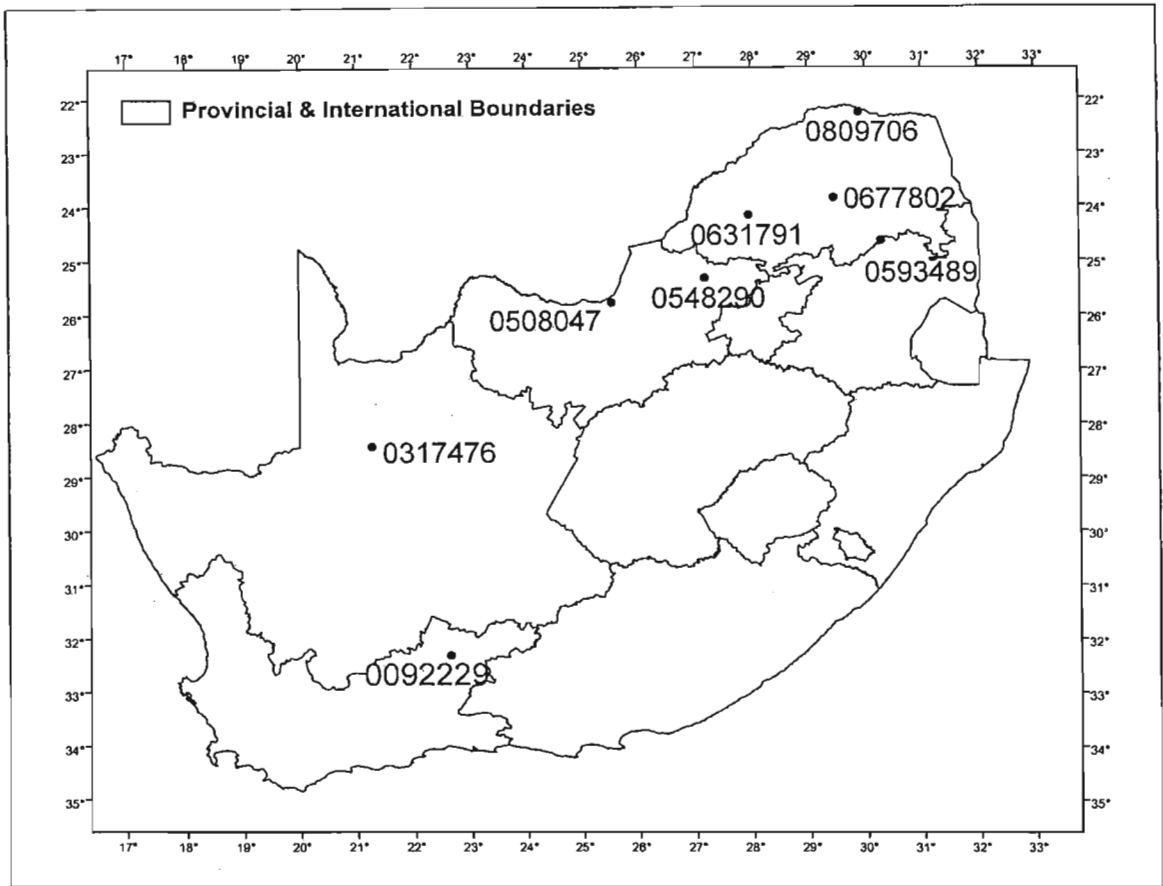


Figure 5-3 Stations that fell within Range 13 but were included in Range 12

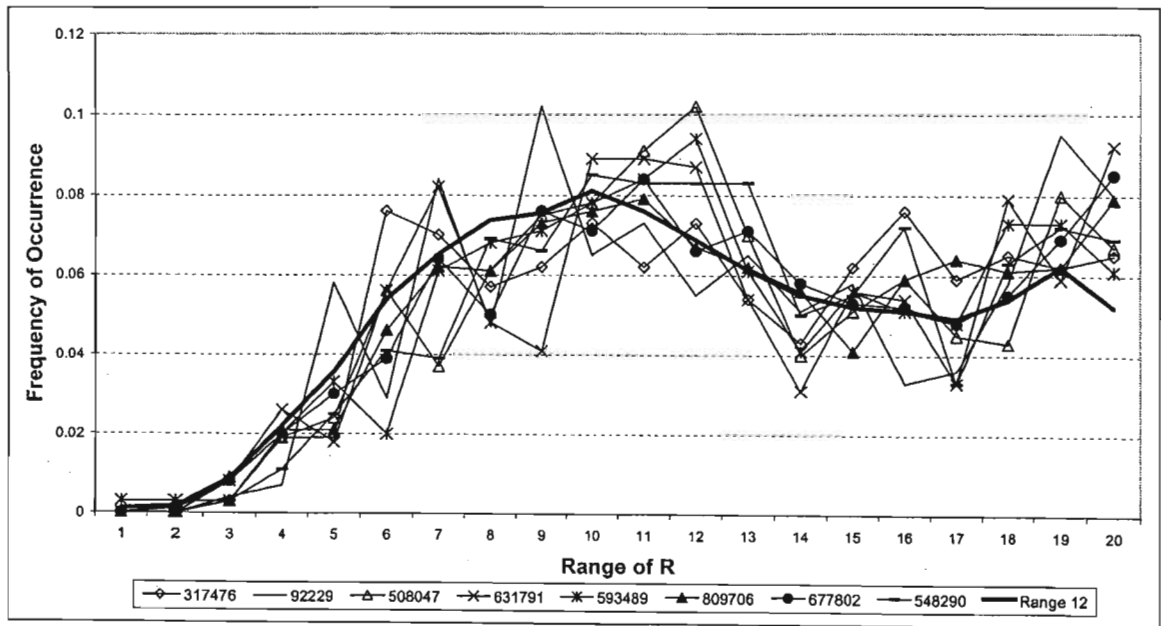


Figure 5-4 Distributions of R for sites included with those in Range 12

Table 5-1 Revised ranges used when collating  $R_{\text{mean}}$  values

No.	Range	No.	Range	No.	Range	No.	Range
I	0.375-0.475	II	0.475-0.525	III	0.525-0.575	IV	0.575-0.675

Once all 157 stations had been divided according to their respective  $R_{\text{mean}}$  values using Table 5-1, an average distribution of R was calculated using all the stations in each  $R_{\text{mean}}$  range. This resulted in 4 average distributions of R, which are shown in Figure 5-5. Similarly, 4 average distributions for the hour of maximum rainfall were calculated and are shown in Figure 5-6.

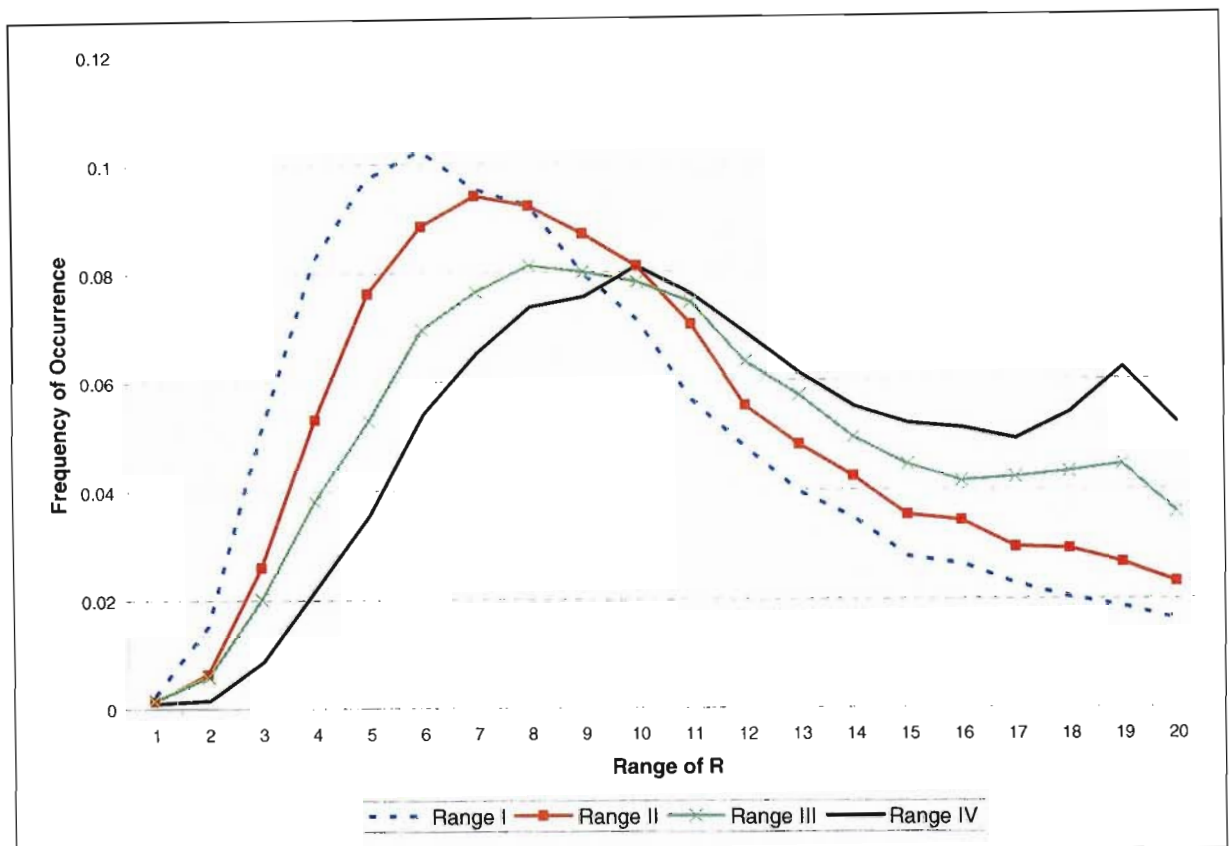


Figure 5-5 Regionalised distributions of R

In order to establish which distribution of R to use for a site of interest anywhere within South Africa, it was necessary to develop a regionalised map of the mean value of R. The mean values of R from the 157 used stations were regionalised using inverse distance weighting. The resulting spatial pattern is displayed in Figure 5-7, and shows that the smallest  $R_{\text{mean}}$  values, 0.325-0.375, occur in the south western part of South Africa as well as on the east coast, while the highest values occur in the northern and north eastern parts of South Africa.

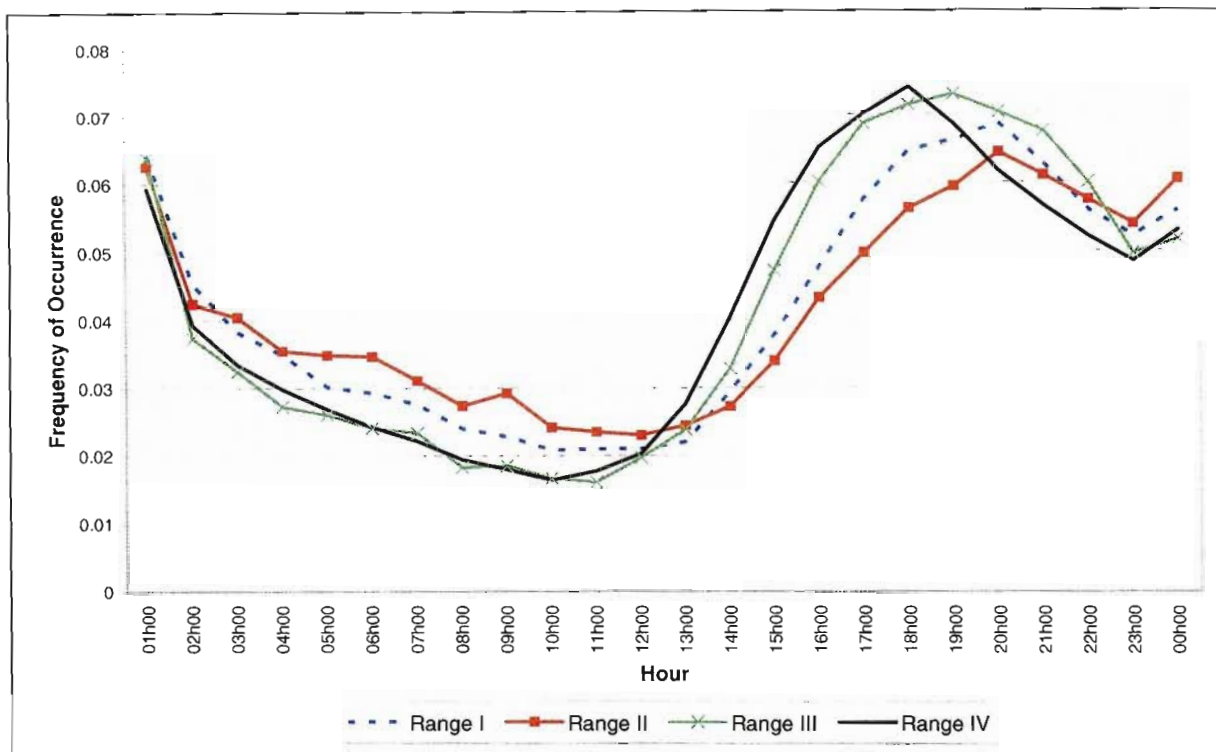


Figure 5-6 Regionalised distributions of the hour of maximum rainfall

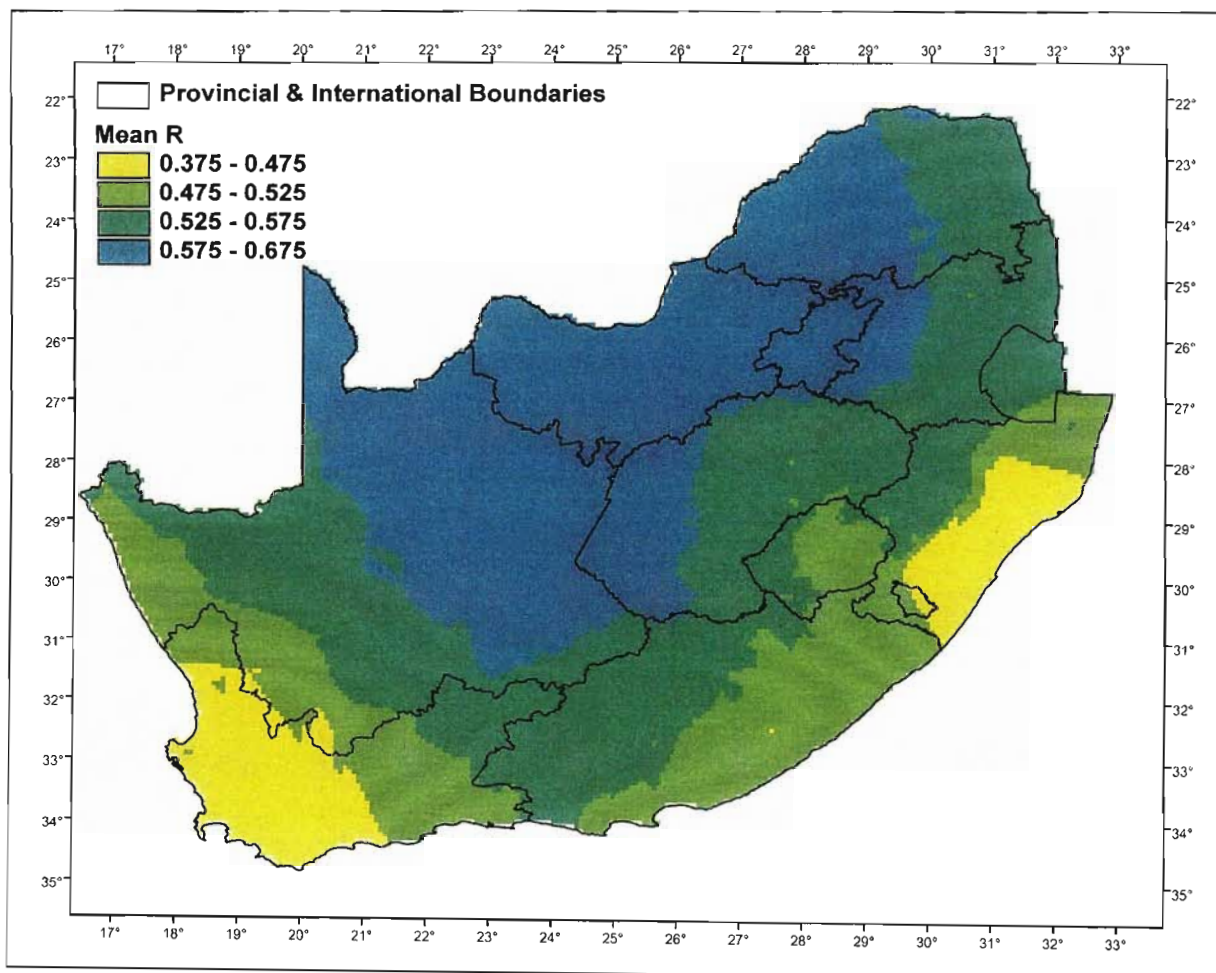


Figure 5-7 Regionalised map of  $R_{mean}$

In application, the range in which the mean value of R for the site of interest needs to be established in order to select the appropriate distribution of R. This is achieved with the use of Figure 5-7. Once the mean value of R has been determined for the site of interest, the disaggregation model operates as presented in Chapter 4. In this study the approach differs from the method developed by Boughton (2000b) by the use of regionalised distributions for R, and the distributions of the hour of maximum rainfall. However, a shortfall of the technique used to regionalise the methodology is that certain distributions of when the hour of maximum rainfall occurs are lost during regionalisation, which can be seen when comparing the regionalised distributions in Figure 5-6 to that of Station Jnk19a in Figure 4-4. It is yet to be discovered as what effect this will have on the resulting output from the model, which is presented and discussed in the following chapter.

## 6. MODEL TESTING AND RESULTS

In order to quantify the simulated performance of the disaggregation model, a similar approach to that used by Smithers and Schulze (2000a) is employed. Moments and other event characteristics computed from the disaggregated rainfall series are compared to the equivalent values computed from the observed data in Section 6.2. Similarly, design rainfall depths computed from the disaggregated rainfall series are compared to the equivalent values computed from the observed data in Section 6.3. The same analyses are performed using the output obtained from the model when regionalised input is used in Sections 6.4 and 6.5.

For each of the 15 test stations shown in Table 3-1, the observed hourly data are aggregated to give 24-hourly values. The disaggregation methodology is then applied to these data in order to attempt to simulate the hourly data for the respective sites. The performance of the model is assessed using two measures. Firstly, seventeen moments and statistics of the disaggregated series, *viz.*, mean, standard deviation, autocorrelations (lag1 – lag10), dry probability, skewness, inter-event duration, event duration and number of events, are compared to the corresponding characteristics of the observed data. The second measure of model performance is aimed at extreme values, where design rainfalls computed from the disaggregated series are compared to the design rainfalls computed from the historical data.

However, before performing the abovementioned analyses at all 15 test locations, five variations of the model were developed and tested at 5 of the test locations, shown in Figure 6-1, using the first measure of model performance mentioned above. The difference between the models lay in the selection of data used to:

- develop the 20 averaged ranked series, as described in Section 4.3, and
- calculate the average 1-hour, 2-hour, 3-hour and 6-hour fractions used in the clustering process, as described in Section 4.4.

The results from these analyses are shown in the following section.

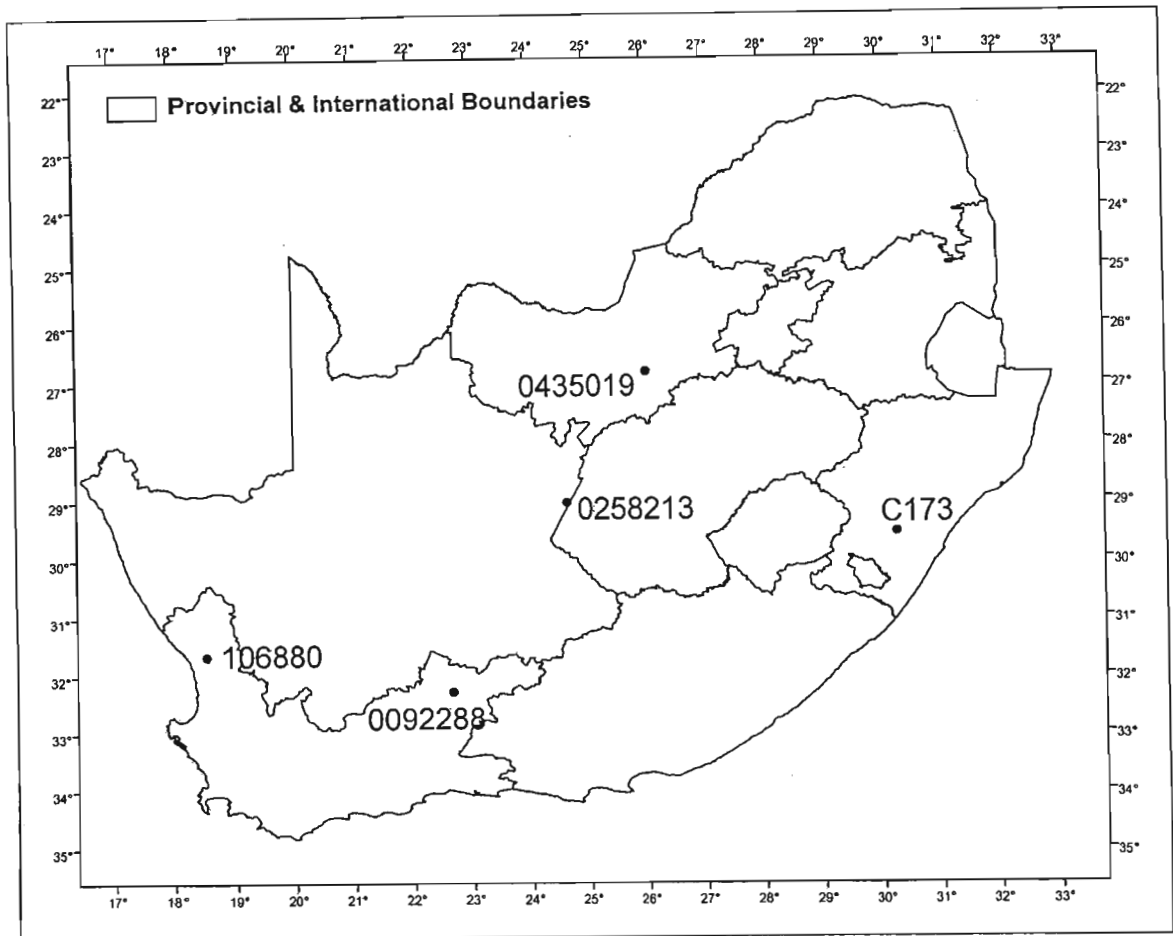


Figure 6-1 Locations of stations used for analysis of the 5 variations of the disaggregation model

### 6.1 Selecting Data Threshold for Model Development

When developing the model various data thresholds were tested to identify which would yield the best results. The five variations of the model that were analysed are based on clustered sequences developed using:

- (i) all days that rainfall occurred,
- (ii) all days for which the aggregated 24-hour total was greater than or equal to 1 mm,
- (iii) all days for which the aggregated 24-hour total was greater than or equal to 5 mm,
- (iv) all days for which the aggregated 24-hour total was greater than or equal to 10 mm, and
- (v) all days for which the aggregated 24-hour total was greater than or equal to 15 mm.

In order to objectively assess the overall performance of the models, the Mean Absolute Relative Error of hourly rainfall (*MARE*), calculated as shown in Equation 6.1, are contained in Table 6-1.

$$MARE = \frac{100}{N_S \times N_M} \times \sum_{j=1}^{N_S} \sum_{i=1}^{N_M} \left( \frac{|S_{(i,j)} - O_{(i,j)}|}{O_{(i,j)}} \right) \quad \dots 6.1$$

where

- MARE* = mean absolute relative error of hourly rainfall (%)
- S<sub>(i,j)</sub>* = *j*-th statistic computed from hourly values, mean (*j*=1), standard deviation (*j*=2), lag-1 autocorrelation (*j*=3), lag-2 autocorrelation (*j*=4), lag-3 autocorrelation (*j*=5), lag-4 autocorrelation (*j*=6), lag-5 autocorrelation (*j*=7), lag-6 autocorrelation (*j*=8), lag-7 autocorrelation (*j*=9), lag-8 autocorrelation (*j*=10), lag-9 autocorrelation (*j*=11), lag-10 autocorrelation (*j*=12), dry probability (*j*=13), duration of wet periods (*j*=14), duration of dry periods (*j*=15), number of wet periods (*j*=16) and skewness (*j*=17), computed from the 100 disaggregated rainfall series for month *i*,
- O<sub>(i,j)</sub>* = *j*-th statistic computed from observed hourly data for month *i*,
- N<sub>M</sub>* = number of months of the year available for statistical analysis, and
- N<sub>S</sub>* = number of statistics and event characteristics calculated (=17).

Table 6-1 Mean absolute relative error computed for hourly rainfall values for the 5 variations of the disaggregation model tested at 5 different locations

Station #	MARE (%)				
	>0	≥1	≥5	≥10	≥15
0435019	593.0	258.3	313.3	363.2	499.8
C173	56.7	56.9	53.1	52.3	51.5
0092288	101.2	86.5	109.9	126.7	157.0
0258213	104.6	79.1	87.0	121.7	163.7
0106880	54.1	46.8	70.4	97.8	128.7

The relatively high *MARE* for all versions of the model at Station 0435019 can be attributed to long periods of missing data, and the way the statistics program used operates when missing data is encountered. When the various statistics are computed, months containing even a single missing hour of data are omitted from analysis. Then, a monthly statistical analysis is carried out on all the remaining months, yielding statistics for each month of the year. Thus, for a complete 20 year record, the statistics program would output the mean,



standard deviation, 25<sup>th</sup> percentile of non-exceedance, the 50<sup>th</sup> percentile of non-exceedance and 75<sup>th</sup> percentile of non-exceedance for each statistic for each month of the year, based on all 20 years of data. If, however, there is a particular month of the year that has less than 3 years of complete data, the statistics for that month of the year will not be calculated. This is the case with regard to Station 0435019. Although this station has a record length of 20 years, the record has much missing data, resulting in only nine months of the year available for analysis. Furthermore, the nine months of the year that were analysed are also plagued with missing data. The problem of missing data, particularly within the digitised SAWS data, was also identified by Smithers and Schulze (2000a). An example of the poor quality data is shown in Table 6-2, where the month of the year with the highest frequency of complete data for the twenty year record at Station 0435019 is April, where only seven months, out of a possible 20, had complete data.

Table 6-2 Number of months with complete data in a 20 year record at Station 0435019

Month of the Year	Number of Complete Months	Month of the Year	Number of Complete Months
January	4	July	1
February	3	August	3
March	5	September	2
April	7	October	3
May	5	November	5
June	2	December	3

As can be seen in Table 6-1, with the exception of Station C173, the model based on all days for which the aggregated 24-hour total was greater than or equal to 1 mm yielded the lowest *MARE*. Hence, this version of the model was selected to be used for the rest of the study.

## 6.2 Moments and Statistics Using At-site Short Duration Data

The two random processes that occur within the disaggregation model, *viz.*, the selection of the value of R and the timing of the hour of maximum rainfall, introduce stochastic variability. At each of the selected test stations the stochastic variability was simulated by generating one hundred disaggregated series. A frequency analysis was performed on the 100 sets of disaggregated values for each statistic and duration. High-Low bar graphs depicting the observed moments and the 25<sup>th</sup> and 75<sup>th</sup> non-exceedance percentiles of the 100 sets of disaggregated values are used to graphically depict the performance of the model.

In order to objectively assess the performance of the model at the 15 test stations, located as shown in Figure 3-1, the *MARE* of hourly rainfall, as calculated in Equation 6.1, and the Mean Absolute Relative Error for all durations (*MARE\_AD*), as calculated in Equation 6.2, are calculated. The number of aggregation levels in Equation 6.2 ( $N_L$ ) was set to 11 and the durations used were 1, 2, 3, 4, 5, 6, 9, 12, 15, 18, and 24-hour.

$$MARE_{AD} = \frac{100}{N_s \times N_M \times N_L} \times \sum_{k=1}^{N_L} \sum_{j=1}^{N_s} \sum_{i=1}^{N_M} \left( \frac{|S_{(i,j,k)} - O_{(i,j,k)}|}{O_{(i,j,k)}} \right) \quad \dots 6.2$$

where

- $MARE_{AD}$  = mean absolute relative rainfall error for all durations (%),
- $S_{(i,j,k)}$  = mean  $j$ -th statistic for aggregation level  $k$  computed from the 100 disaggregated rainfall series for month  $i$ ,
- $O_{(i,j,k)}$  =  $j$ -th statistic computed from observed data for aggregation level  $k$  for month  $i$ ,
- $N_M$  = number of months of the year available for statistical analysis,
- $N_L$  = number of aggregation levels used (=11, for 1, 2, 3, 4, 5, 6, 9, 12, 15, 18, and 24-hour durations), and
- $N_s$  = number of statistics and event characteristics calculated (=17, for mean, standard deviation, lag-1 autocorrelation, lag-2 autocorrelation, lag-3 autocorrelation, lag-4 autocorrelation, lag-5 autocorrelation, lag-6 autocorrelation, lag-7 autocorrelation, lag-8 autocorrelation, lag-9 autocorrelation, lag-10 autocorrelation, dry probability, duration of wet periods, duration of dry periods, number of wet periods and skewness)

The poorest results, i.e. the highest *MARE* and *MARE\_AD*, were achieved at Station 0435019, which had a *MARE* of 258.3% and a *MARE\_AD* of 138.9%. The best results, i.e. the lowest *MARE* and *MARE\_AD*, were achieved at Station Sacfs, which had a *MARE* of 37.8% and a *MARE\_AD* of 67.2%. The results from disaggregating the 24-hour data at Stations 0435019 and Sacfs are shown in Figure 6-2 and 6-3 respectively. The *MARE* and *MARE\_AD* for the 15 test station are presented in Table 6-3.

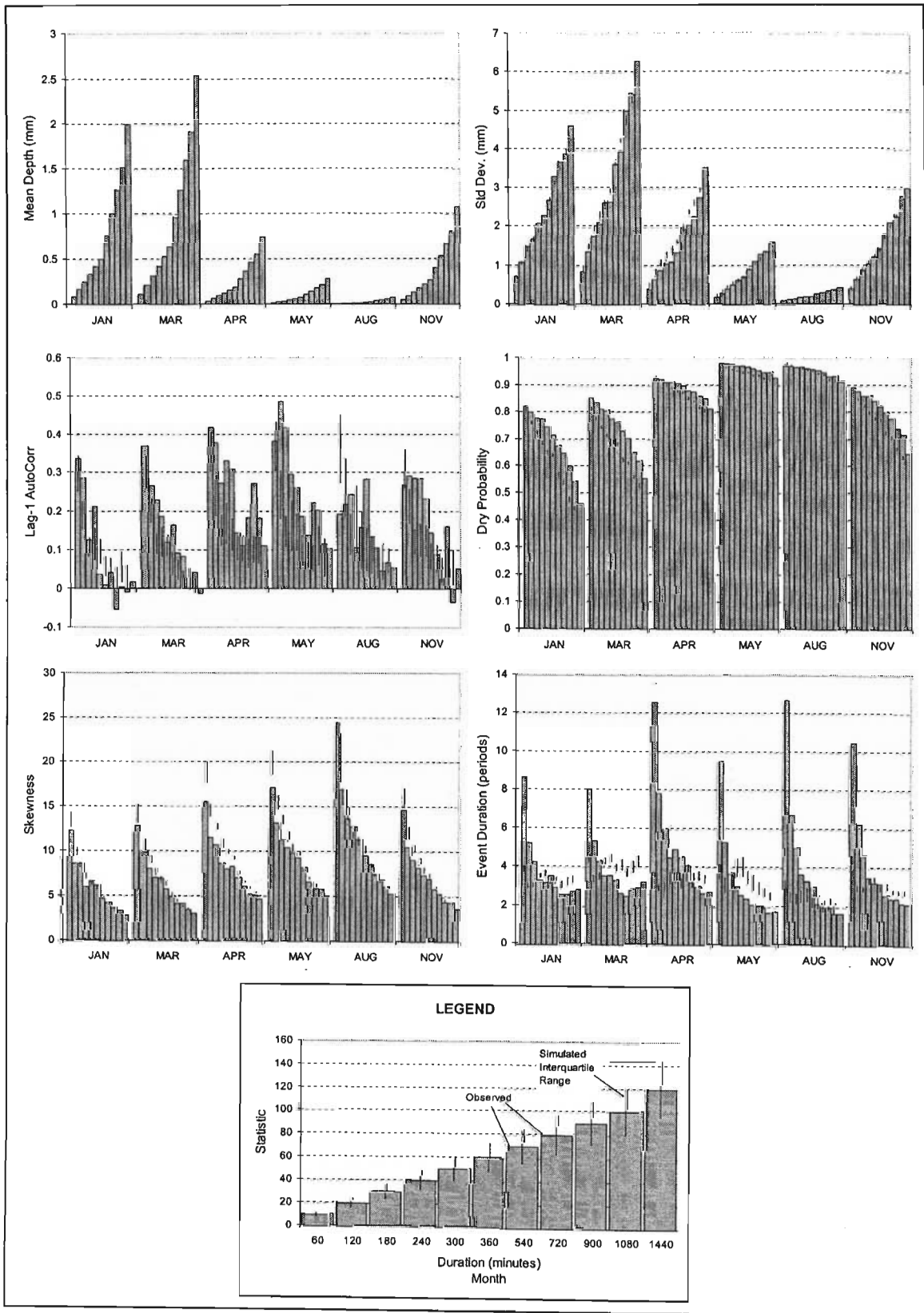


Figure 6-2 Simulated performance of the disaggregation model Station 0435019

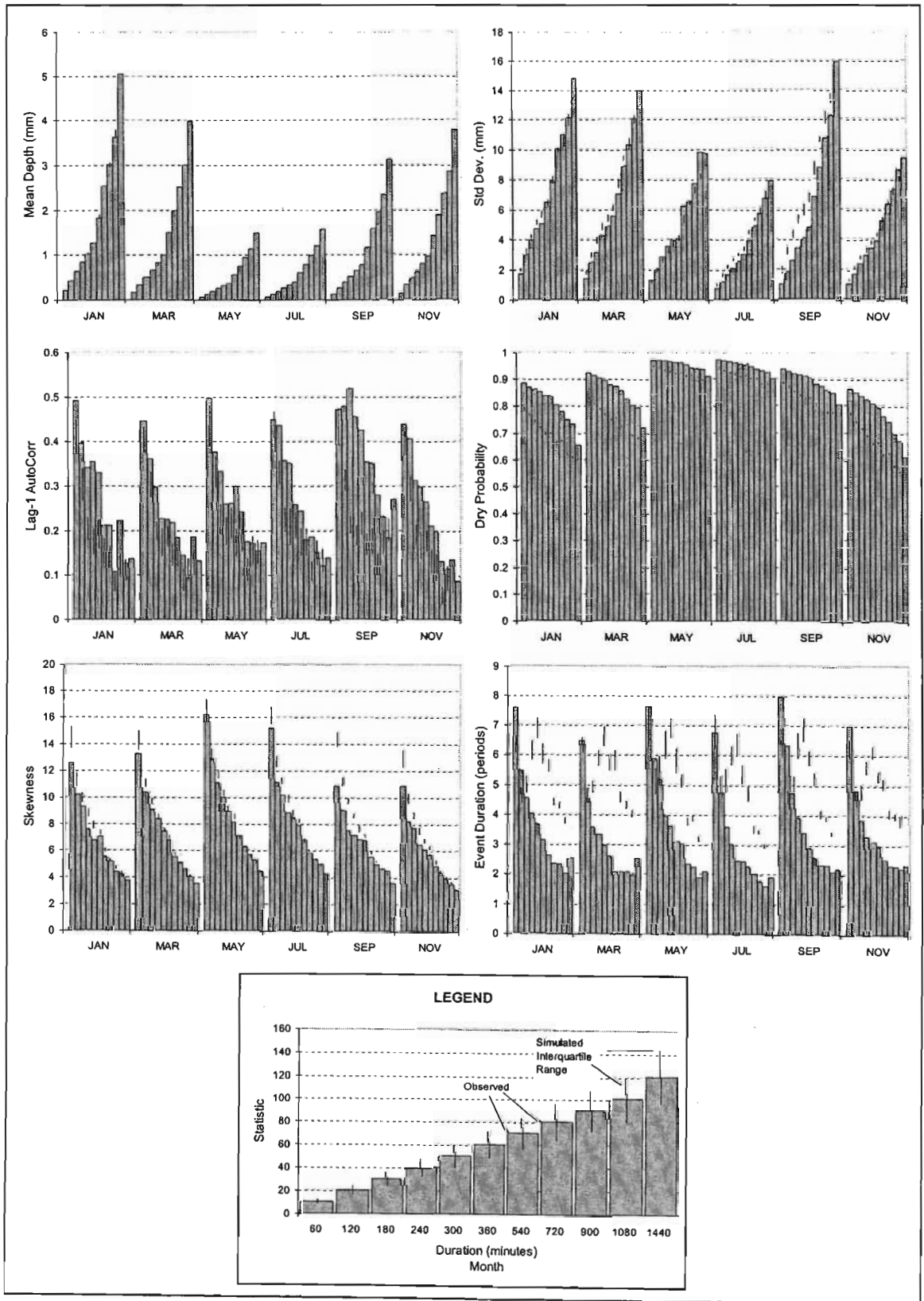


Figure 6-3 Simulated performance of the disaggregation model at Station Sacfs

Table 6-3 *MARE* and *MARE\_AD* for the 15 test stations

Station Number	<i>MARE</i> (%)	<i>MARE_AD</i> (%)
0435019	258.3	138.9
0552581	83.1	105.1
C173	56.9	137.5
0317474	55.8	96.5
0719370	237.7	135.4
0023710	45.0	76.8
304474	43.7	100.1
Sacfs	37.8	67.2
0028748	48.2	111.5
0092288	86.6	87.9
0474680	165.8	116.7
0261516	52.9	97.2
0127272	118.2	145.7
0258213	79.3	148.1
0106880	46.8	87.0

It can be seen in Figure 6-2 and Figure 6-3 that the disaggregation model performs similarly well in simulating the mean, standard deviation and skewness at Station 0435019 and Station Sacfs. The inter-quartile range for the mean cannot be seen owing to the near perfect simulation of the mean. It is expected that the mean rainfall for all levels of aggregation is simulated extremely well owing to the method of disaggregation, as explained in Chapter 4. The model tends to be less capable of simulating certain event characteristics and statistics such as event duration and dry probability. This is a weakness in the disaggregation model and suggests that more work needs to be done on refining the sequencing of the hourly rainfalls. The distinguishing factors between the best and worst simulations, according to the *MARE* and *MARE\_AD*, seem to be the autocorrelation lags 1 – 10. This can be verified by calculating the mean absolute relative error for each autocorrelation for all levels of aggregation (*MARE\_LAG*). The *MARE\_LAG* for the two stations, as calculated in Equation 6.3, are shown in Table 6-4.

$$MARE\_LAG = \frac{100}{N_M \times N_L} \times \sum_{j=1}^{N_L} \sum_{i=1}^{N_M} \left( \frac{|S_{(i,j)} - O_{(i,j)}|}{O_{(i,j)}} \right) \quad \dots 6.3$$

where

*MARE\_LAG* = mean absolute relative error lag autocorrelation for all durations (%),

$S_{(i,j)}$  = lag autocorrelation for aggregation level  $j$  computed from the 100 disaggregated rainfall series for month  $i$ ,

- $O_{(i,j)}$  = lag autocorrelation computed from observed data for aggregation level  $j$  for month  $i$ ,
- $N_M$  = number of months of the year available for statistical analysis, and
- $N_L$  = number of aggregation levels used (=11).

Table 6-4 *MARE\_LAG* for autocorrelation lags 1-10 at Station 0435019 and Station Sacfs

	<i>MARE_LAG</i> (%)									
	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
Station 0435019	171.8	339.6	157.3	280.1	535.0	169.4	207.7	112.7	170.7	96.3
Station Sacfs	29.9	138.9	93.7	92.4	114.6	117.9	95.7	131.2	103.3	76.7

It is clear from Table 6-4 that the autocorrelations for lags 1 – 10 are simulated with significantly higher accuracy at Station Sacfs than at Station 0435019. This can be attributed the quality of the available data, as mentioned in Section 6.1. When comparing the number of months with complete data for Station Sacfs (Table 6-5), also with a 20 year record, with that shown in Table 6-2 for Station 0435019, the importance of good quality data becomes apparent.

Table 6-5 Number of months with complete data in a 20 year record at Station Sacfs

Month of the Year	Number of Complete Months	Month of the Year	Number of Complete Months
January	12	July	17
February	15	August	17
March	15	September	16
April	15	October	14
May	16	November	11
June	16	December	14

Furthermore, it becomes apparent, from Figure 6-2 and Figure 6-3, that the dry probabilities, and hence the event durations, are better simulated at Station 0435019 than Station Sacfs. It is postulated that this may be attributed to the distribution of  $R$  at Station 0435019, which has a mean value of  $R$  ( $R_{\text{mean}}$ ) of 0.607, whereas Station Sacfs has an  $R_{\text{mean}}$  of 0.499. This appears to

be a shortcoming of using a single distribution of R to represent rainfall of all magnitudes. Although the disaggregated data will have the correct overall distribution of R, it is likely that the tails of the rainfall distribution will not receive the correct values of R. It is yet to be shown how this will effect the estimation of design rainfalls.

With the focus of this study not only being the disaggregation of 24-hour data to yield day to day short-duration rainfall values, but also on the estimation of design rainfall values, the second measure of assessing the performance of the model is an assessment of how well extreme events are modelled when using the disaggregated data.

### 6.3 Extreme Rainfall Events Using At-site Short Duration Data

Similar to the procedures used by Smithers and Schulze (2000a), design rainfall depths were calculated using the General Extreme Value (GEV) distribution fitted to the Annual Maximum Series (AMS) by L-moments, for the observed data and for each of the 100 disaggregated series generated from the disaggregation model. Design values for the 2, 5, 10, 20, 50, and 100-year return periods were computed for durations of 1, 2, 3, 4, 6, 8, 10, 12, 16, 20, and 24-hour. For each duration and return period, a frequency analysis was performed on the 100 values computed from the disaggregated rainfall series generated by the disaggregation model. High-Low bar graphs depicting the observed design rainfall computed from the observed data and the 25<sup>th</sup> and 75<sup>th</sup> non-exceedance percentiles of the design rainfall computed from the 100 disaggregated rainfall series were used to evaluate the performance of the model.

In order to objectively assess the performance of the model at the 15 test stations, with respect to the estimation of design rainfalls, the Mean Absolute Relative Error for the 1-hour duration (*MARE*), as calculated in Equation 6.4, and the Mean Absolute Relative Error for all durations (*MARE\_AD*), as calculated in Equation 6.5, are calculated.

$$MARE = \frac{100}{N_{RP}} \times \sum_{i=1}^{N_{RP}} \left( \frac{|S_{(i)} - O_{(i)}|}{O_{(i)}} \right) \quad \dots 6.4$$

where

*MARE* = mean absolute relative error for 1-hour design rainfall (%),

- $S_{(i)}$  = mean  $i$ -th return period, 1-hour design rainfall computed from the 100 disaggregated rainfall series,  
 $O_{(i)}$  =  $i$ -th return period, 1-hour design rainfall computed from observed data, and  
 $N_{RP}$  = number of return periods (= 6 for 2, 5, 10, 20, 50, and 100-year return periods).

$$MARE_{AD} = \frac{100}{N_L \times N_{RP}} \times \sum_{i=1}^{N_L} \sum_{j=1}^{N_{RP}} \left( \frac{|S_{(i,j)} - O_{(i,j)}|}{O_{(i,j)}} \right) \quad \dots 6.5$$

where

- $MARE_{AD}$  = mean absolute relative error of design rainfall of all durations (%),  
 $S_{(i,j)}$  = mean  $j$ -th return period,  $i$ -th hour design rainfall computed from the 100 disaggregated rainfall series,  
 $O_{(i,j)}$  =  $j$ -th return period,  $i$ -th hour design rainfall computed from observed data, and  
 $N_L$  = number of aggregation levels (= 11)  
 $N_{RP}$  = number of return periods (= 6 for 2, 5, 10, 20, 50, and 100-year return periods).

Examples of model performance, with respect to design rainfall estimation, are shown in Figure 6-4, which depict the worst (Station 0028748) and best (Station 0474680) simulations according to the  $MARE_{AD}$  values. Station 0028748 has an  $MARE$  of 163.1% and an  $MARE_{AD}$  of 39.1%, whereas Station 0474680 has an  $MARE$  of 14.4% and an  $MARE_{AD}$  of 4.2%. The  $MARE$  and  $MARE_{AD}$  for all test stations are shown in Table 6-6.

When comparing Table 6-6 with Table 6-3, there appears to be a direct contradiction. Sites that performed well in the statistical measures of model performance generally perform relatively poorer when estimating design floods and vice versa. The reason for some stations yielding poor results in the statistical measures of model performance, however, is postulated to be a consequence of poor data quality. The poor performances observed when estimating design floods seem to be related to the distribution  $R$ , i.e. those stations with higher  $R_{\text{mean}}$  values had the best results, and those with the lower  $R_{\text{mean}}$  values had the worst results (as shown in Table 6-7). This is because on those days when smaller rainfall events ( $\pm 1$  mm)



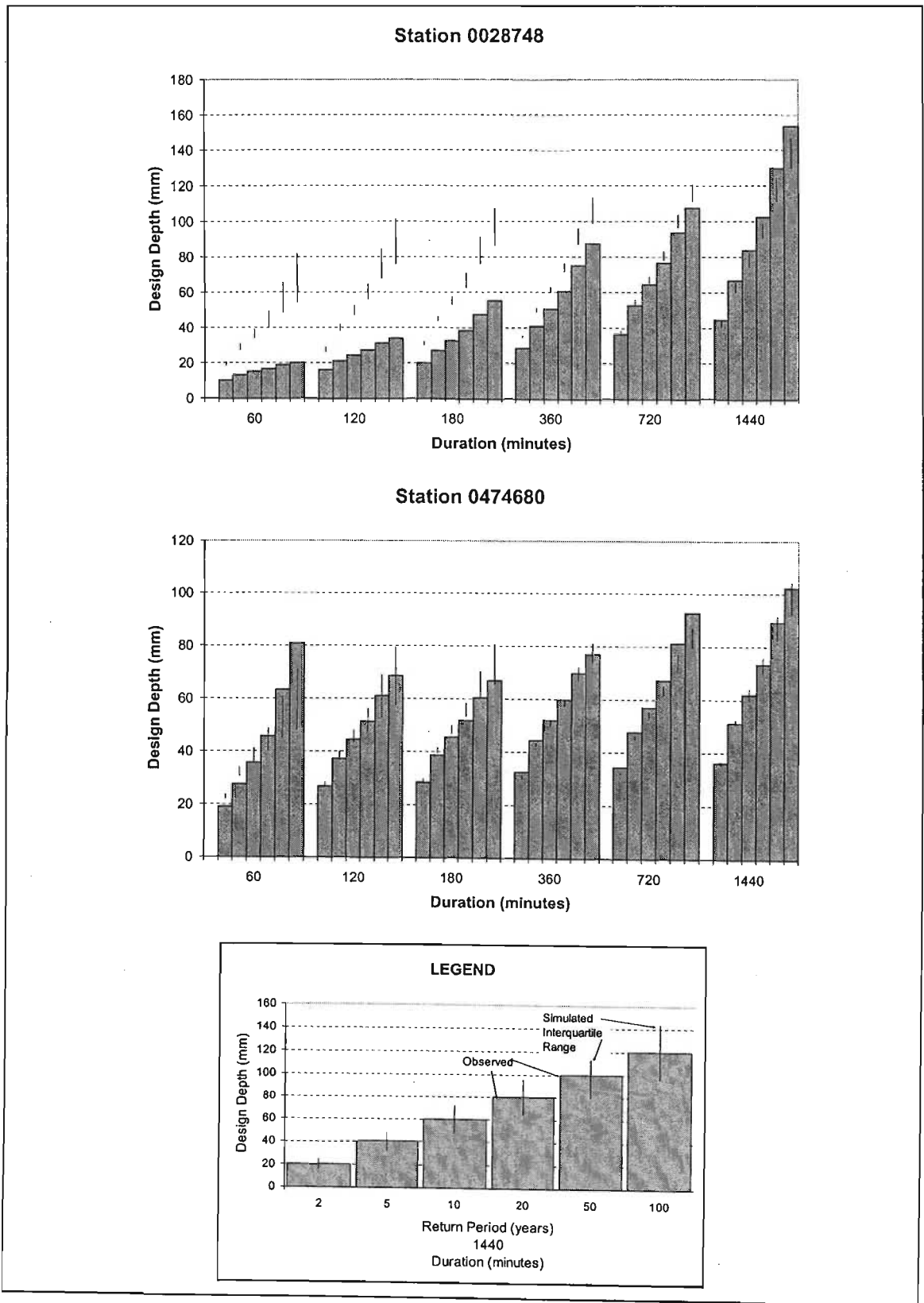


Figure 6-4 Design rainfall estimated using disaggregated data for Station 0028748 and Station 0474680

occurred it is likely that the all the day's rainfall fell within a few hours, thus unduly influencing the distribution of R. Although this will affect all the stations used, it appears that the error is exacerbated for those stations with lower mean R values. It is postulated that the use of different distributions of R to represent rainfalls of differing magnitudes will improve the performance of the rainfall disaggregation model.

Table 6-6 *MARE* and *MARE\_AD* for design rainfalls at the 15 test stations

Station Number	<i>MARE</i> (%)	<i>MARE_AD</i> (%)
0435019	18.4	7.4
0552581	21.6	6.1
C173	27.2	11.5
0317474	27.2	14.6
0719370	31.5	7.4
0023710	57.3	26.8
304474	55.8	24.0
Sacfs	74.3	24.2
0028748	163.1	39.1
0092288	31.9	9.1
0474680	14.4	4.2
0261516	38.9	11.8
0127272	11.2	8.3
0258213	30.3	11.6
0106880	53.7	14.8

Table 6-7 Mean values of R and *MARE* for the 15 test stations

Station Number	<i>MARE</i> (%)	Mean R	Station Number	<i>MARE</i> (%)	Mean R	Station Number	<i>MARE</i> (%)	Mean R
0435019	18.4	0.607	0023710	57.3	0.479	0474680	14.4	0.596
0552581	21.6	0.631	304474	55.8	0.458	0261516	38.9	0.571
C173	27.2	0.468	Sacfs	74.3	0.499	0127272	11.2	0.503
0317474	27.2	0.654	0028748	163.1	0.447	0258213	30.3	0.599
0719370	31.5	0.626	0092288	31.9	0.551	0106880	53.7	0.505

All of the above results are from applying the disaggregation model at locations where there are some hourly data available in order to calculate the distributions of R, and the hour of maximum rainfall, for the respective location. In order to estimate hourly rainfalls at a location where no hourly data are available, but daily rainfall data are available, it is necessary to utilise regionalised information in the disaggregation model. The following sections assess the performance of the disaggregation model when regionalised input is used.

#### 6.4 Moments and Statistics When Using Regionalised Information

It should be noted that the results from the disaggregation model shown in Section 6.2 and Section 6.3 used distributions of R, and distributions of the hour of maximum rainfall that were calculated from the hourly data at the location of interest. The results contained in Sections 6.4 and 6.5 were achieved using regionalised distributions, as detailed in Chapter 5, which do not include any of the data from the 15 test stations.

The performance of the model at the “ungauged” test stations was assessed in exactly the same manner as outlined in Section 6.2. The values of the Mean Absolute Relative Error for hourly rainfall (*MARE*) and the Mean Absolute Relative Error for all durations (*MARE\_AD*), as calculated in Equation 6.1 and Equation 6.2 respectively, are shown in Table 6-8.

Table 6-8 *MARE* and *MARE\_AD* for the 15 test stations using the regionalised methodology

Station Number	<i>MARE</i> (%)	<i>MARE_AD</i> (%)
0435019	243.4	129.7
0552581	85.1	105.7
C173	61.7	151.9
0317474	63.9	99.8
0719370	254.7	135.0
0023710	44.2	73.4
304474	42.7	99.7
Sacfs	40.5	67.1
0028748	45.8	114.7
0092288	89.5	89.2
0474680	159.7	111.5
0261516	55.1	101.0
0127272	104.8	131.9
0258213	163.7	204.6
0106880	128.7	132.7

Assessing the values in Table 6-8 and comparing these to the values in Table 6-3, it can be seen that when regionalised input is used, the model is able to produce results that are very similar to those produced by the model when at-site short duration information are available. This is depicted graphically in Figure 6-5 and Figure 6-6, where the results from the same stations analysed in Section 6.2, i.e. Station 0435019 and Station Sacfs respectively, are presented.

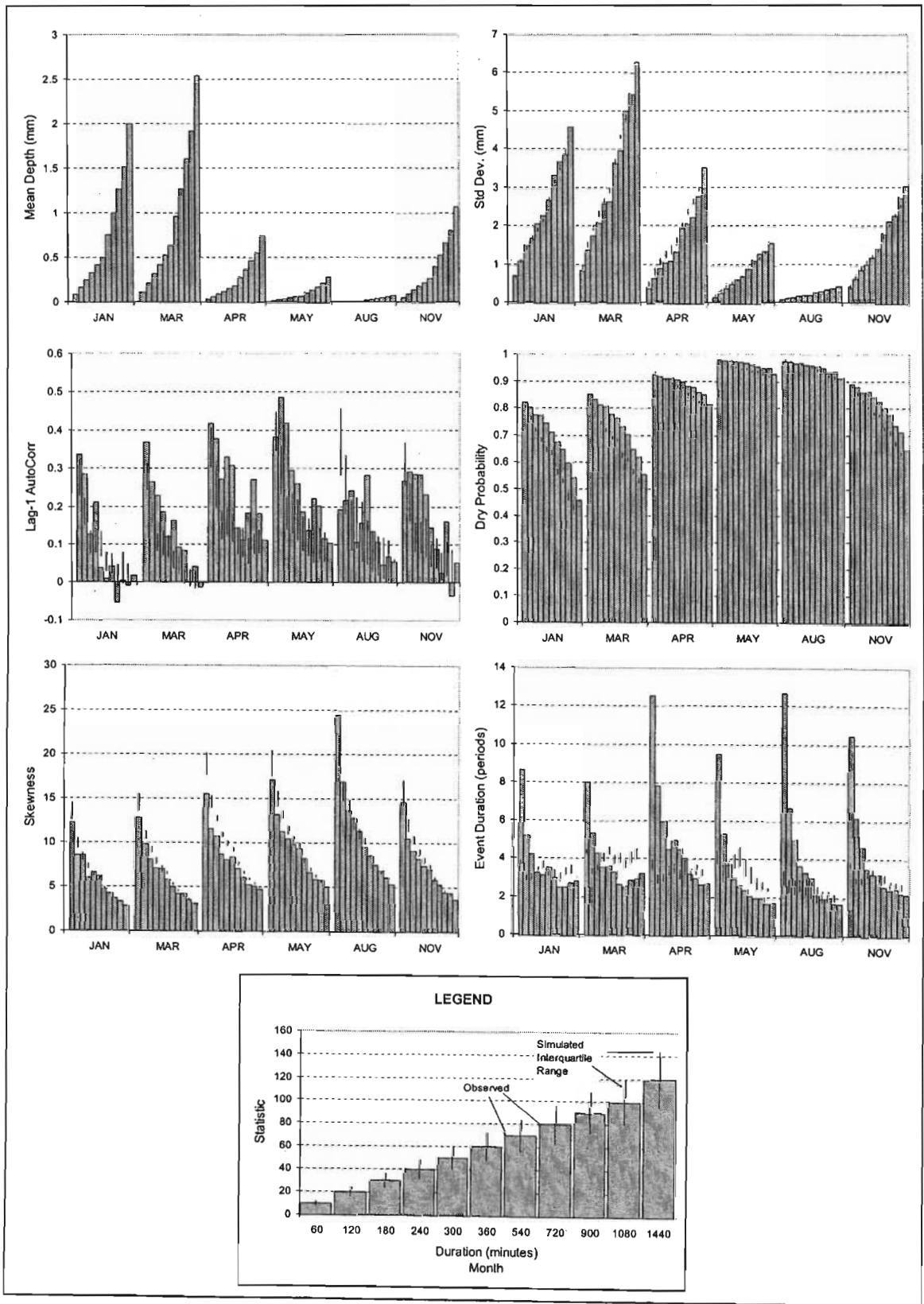


Figure 6-5 Simulated performance of the disaggregation model at Station 0435019, when regionalised input is used

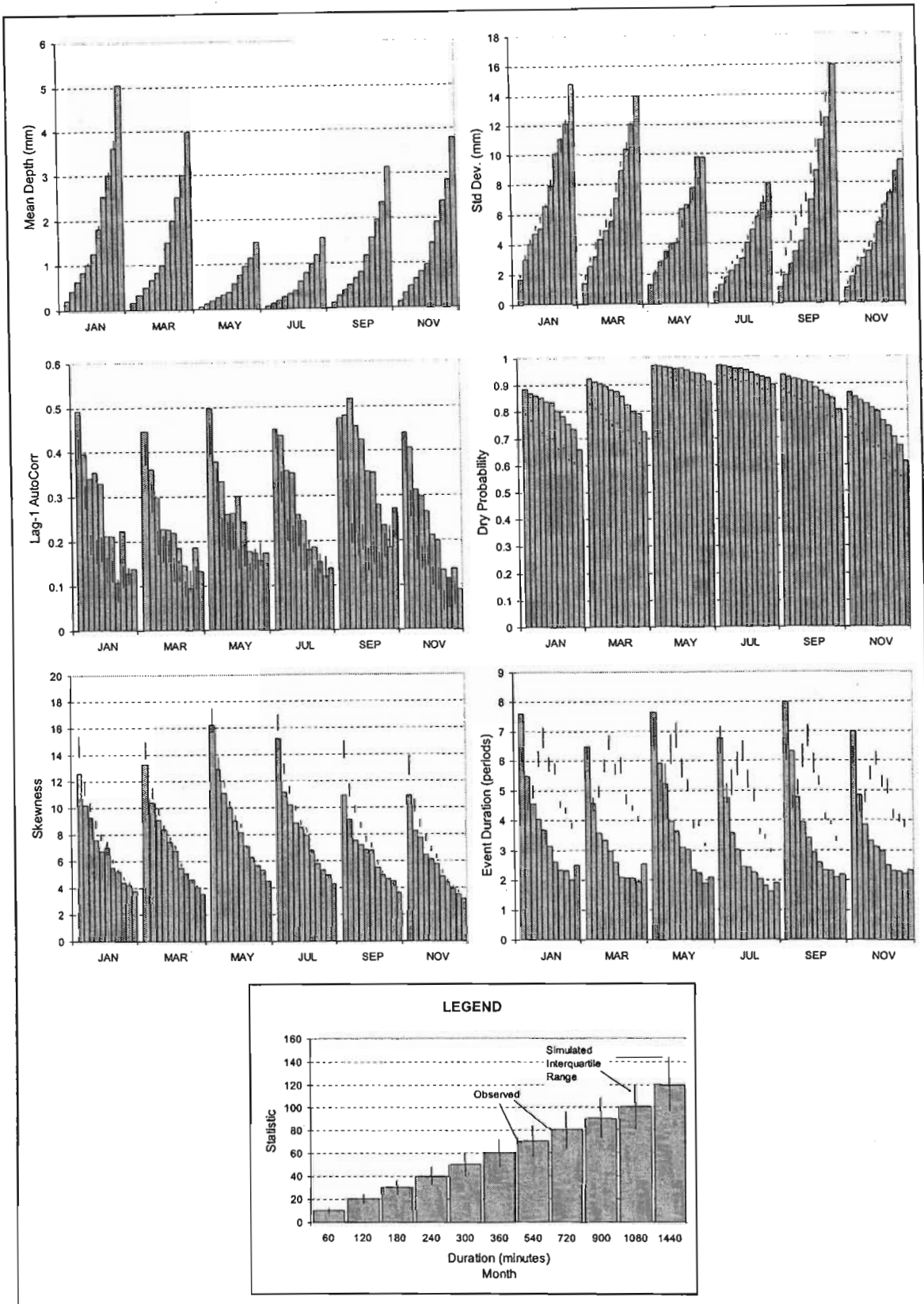


Figure 6-6 Simulated performance of the disaggregation model at Station Sacfs, when regionalised input is used

In the case of Station 0106880, however, the use of regionalised input results in a poorer simulation compared to when at-site information is used. This is due to the methodology used in the regionalisation. Although the actual distribution of R for this site seems to be well represented by the appropriate regionalised distribution of R, the regionalised distribution of the hour of maximum rainfall is not representative to that observed at the site, as shown in Figure 6-7.

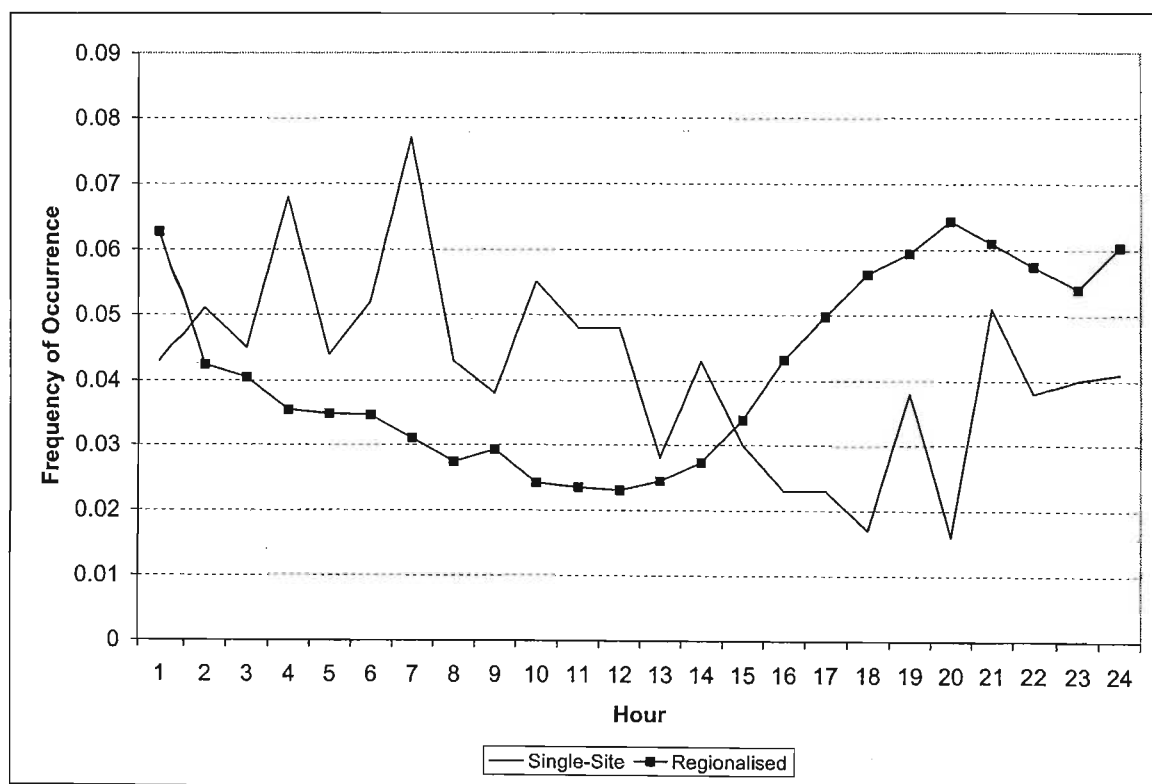


Figure 6-7 Single-site and regionalised distributions for the hour of maximum rainfall applied at Station 0106880

The very poor *MARE* and *MARE\_AD* values for Stations 0435019 and 0719370 are attributed to the poor observed records at these sites. Although the methodology would normally be used to disaggregate observed daily rainfall data, which are observed at 08:00 every day in South Africa, for the purpose of this investigation the same daily rainfall totals aggregated from the hourly data as used in Section 6.2 were disaggregated back to hourly values. This facilitated the comparison of the results from the disaggregation model when at-site short duration information was available and when regionalised input was used. As a result of the way the statistics are calculated (explained in Section 6.2), the large errors indicated by the *MARE* and *MARE\_AD* were to be expected for the abovementioned stations owing to the

large percentage of missing data within the rainfall records. The following section assesses the performance of the regionalised methodology when estimating extreme events.

**6.5 Extreme Rainfall Events When Using Regionalised Information**

The performance of the regionalised methodology with respect to design rainfall was assessed using the same methodology detailed in Section 6.3. The values of the Mean Absolute Relative Error for the 1-hour duration (*MARE*) and the Mean Absolute Relative Error for all durations (*MARE\_AD*), as calculated in Equation 6.4 and Equation 6.5 respectively, are summarised in Table 6-9.

Table 6-9 *MARE* and *MARE\_AD* for design rainfalls at the 15 test stations using the regionalised methodology

Station Number	<i>MARE</i> (%)	<i>MARE_AD</i> (%)
0435019	18.4	7.1
0552581	19.8	5.6
C173	23.6	11.3
0317474	21.8	13.7
0719370	25.9	6.6
0023710	69.4	27.8
304474	55.2	23.1
Sacfs	77.3	25.4
0028748	195.4	42.8
0092288	35.7	9.5
0474680	13.9	4.0
0261516	40.5	11.7
0127272	13.7	9.0
0258213	32.8	12.0
0106880	60.8	15.3

From the results for the *MARE\_AD* values presented in Table 6-9, it is evident that the disaggregating procedure performed the most poorly at Station 0028748 and the best results were obtained at Station 0474680, as was the case when short duration data obtained from the site of interest were used as input into the disaggregation model. As was observed in Section 6.4, the design rainfall *MARE* and the *MARE\_AD* values when using regionalised input into the disaggregation model are very similar to the values calculated when site-specific data were used. As noted in Section 6.3, it is postulated that the simulated performance of design rainfall is related to the  $R_{mean}$  value, and hence the distribution R that is used.

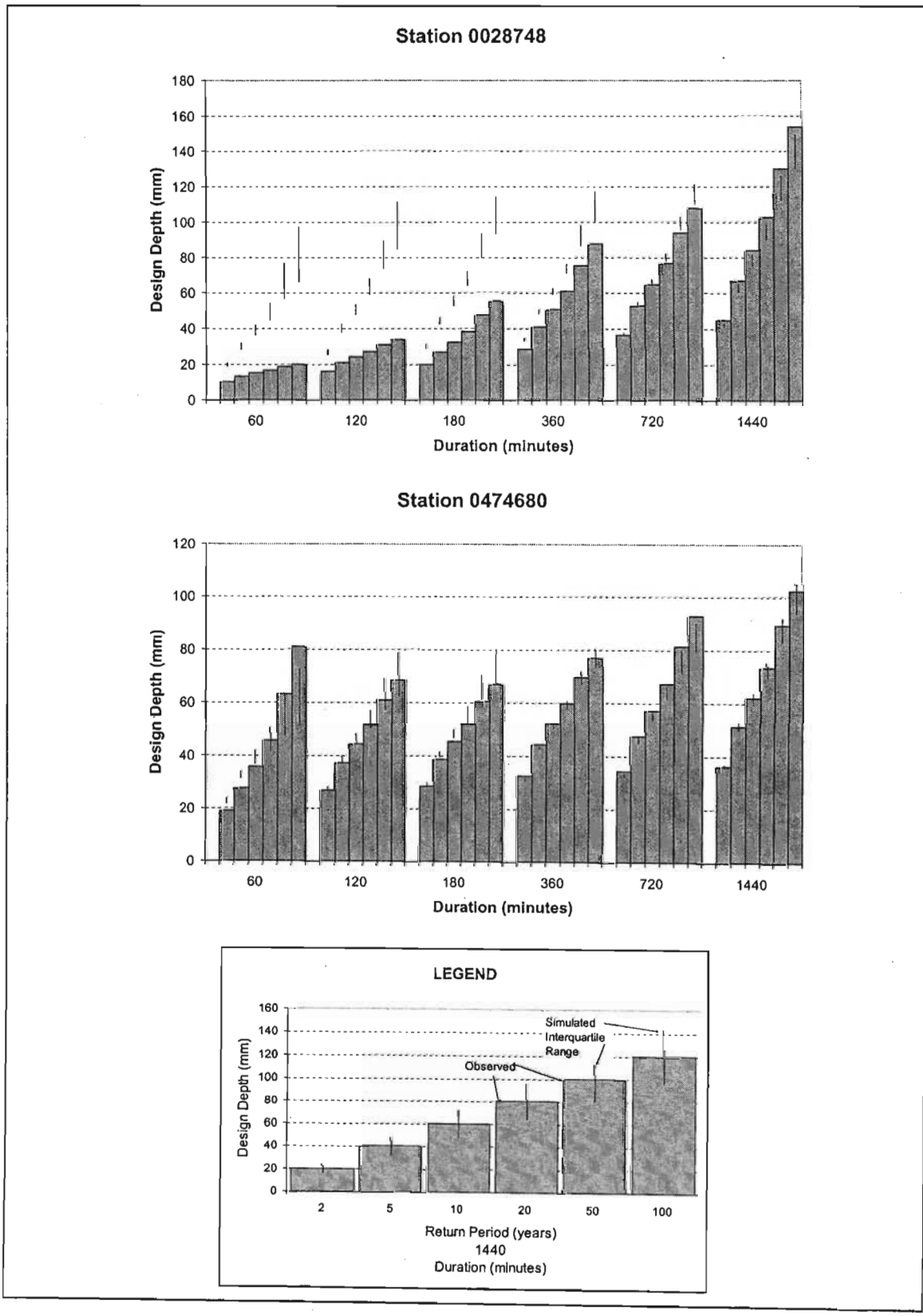


Figure 6-8 Design rainfall estimated using disaggregated data for Station 0028748 and Station 0474680, when regionalised input is used



## 6.6 Chapter Conclusions

The performance of the disaggregation model was analysed firstly by comparing moments and event characteristics computed for the observed and disaggregated data, and secondly by comparing design rainfall values computed from the observed and disaggregated data. This was done for sites where short duration data were available, as well as for simulations where it was assumed that no short duration data were available, in order to assess the performance of the rainfall disaggregation model.

From the results obtained at the 15 independent test sites located in differing climatic regions of South Africa, it is evident that when short duration data are available, the disaggregation model is able to produce short duration rainfall where the mean, standard deviation and skewness are very similar to that of the observed data. However, statistics and event characteristics related to the structure of the rainfall are not simulated as well, and it is postulated that this is the result of the methodology used to sequence the disaggregated hourly rainfall values. Furthermore, it is postulated that the use of different distributions to represent rainfalls of different magnitudes will improve the simulation of dry probabilities and design rainfall depths.

Comparing the *MARE* and *MARE\_AD* values from the disaggregation model when at-site information is available to those yielded when regionalised input is used, it can be seen that the results for a particular location are very similar. This indicates that, with the exception of the distribution of the hour of maximum rainfall and other event characteristics related to the sequencing of the hourly rainfalls, the disaggregation model, when using regionalised information, is able to produce short duration rainfall with similar characteristics to the actual rainfall observed at the respective location.

## 7. DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

The temporal distribution of rainfall is an important factor affecting the timing and magnitude of peak discharge from a catchment. The ability to represent the temporal distribution of rainfall is thus an essential prerequisite for the design of hydraulic structures and other water resource studies. The methods reviewed in Chapter 2 were developed to fulfil a common function, which is to account for the variability of rainfall intensities within a given event.

The objectives of this study were:

- to identify techniques for the disaggregation of daily rainfall to produce hourly increments which aggregate to equal the observed daily values; and
- to select and apply a disaggregation technique and assess the applicability of the method in South Africa.

A comprehensive description of techniques to account for varying rainfall intensities within rainfall events, as an aid for design flood estimation, has been provided in Chapter 2. These included methods for developing temporal rainfall distribution curves and methods for the disaggregation of coarser-scale rainfall data.

Of the techniques reviewed, Boughton's (2000b) methodology for disaggregating daily rainfall to hourly values was selected for use and assessment in this study. The selection was largely based on the promising results reported by Boughton (2000b) when estimating design storms, and the simplicity of the method. However, the methodology developed by Boughton (2000b) was to disaggregate only those daily rainfalls that were greater than or equal to 15 mm, as these events were viewed to be the most important when considering flood studies (Boughton, 2000b). Hence, it was decided that the methodology developed by Boughton (2000b) would need to be modified in order to fulfil the objectives of this study.

### 7.1 Short Duration Data Used

The short duration rainfall database compiled by Smithers and Schulze (2000a) was used in this study. Only stations with record lengths greater than 10 years were used. Of the 172 stations identified, 15 stations were excluded from model development in order to

independently evaluate the model. All the data were made available as 24 hourly blocks, extracted for 0:00 to 0:00 periods.

## 7.2 Methodology

The disaggregation model developed by Boughton (2000b) was developed only for design flood purposes, and hence only focussed on disaggregating larger rainfalls. In order to achieve the objectives of this study, the methodology was modified. An analysis of the Mean Absolute Relative Error (*MARE*) for various data thresholds revealed that using all days for which the aggregated 24-hour rainfall total was greater than or equal to 1 mm yielded the best results when disaggregating all non-zero 24-hour rainfalls.

Further modifications were made to the methodology regarding the distribution of the time when the hour of maximum rainfall occurred. It was evident that the rainfall data at different stations displayed different distributions for the hour maximum rainfall. The distribution of the hour of maximum rainfall at each station was computed, and random sampling along the respective distributions was performed. This differs from the original Boughton (2000b) model where the hour of maximum rainfall was determined by random sampling along a uniform distribution.

The resulting model is capable of producing 480 different temporal patterns with ranging levels of uniformity. The distribution of  $R$  and the distribution of the hour of maximum rainfall for each station determine which of the 480 possible temporal patterns to select for a particular day's rainfall.

## 7.3 Regionalising the Methodology

Owing to the relative paucity of available short duration rainfall data in South Africa, and relative abundance of daily rainfall data compared to continuously recorded rainfall data, it was necessary to regionalise the methodology. This enables the disaggregation of daily rainfall for a particular location where daily rainfall data are available and where short duration rainfall data are not available. A regionalised map of the mean value of  $R$  ( $R_{\text{mean}}$ ) was developed for South Africa. This enables a user to identify the range of the  $R_{\text{mean}}$  value for the site of interest, hence determining which regionalised distributions to use in the

disaggregation process. The spatial pattern displayed by the regionalised map shows that the smallest  $R_{\text{mean}}$  values, indicative of less intense, frontal or general rainfall, occur in the south western part of South Africa as well as on the east coast, while the highest values, indicative of higher intensity convective rainfall, occur in the northern and north eastern parts of South Africa.

#### **7.4 Application of the Disaggregation Model**

Two measures were employed in order to quantify the performance of the disaggregation model. Firstly, moments and other event characteristics were computed from the disaggregated data and compared to the equivalent values computed from the observed data. Secondly, design rainfall depths were computed from the disaggregated data and compared to the equivalent values computed from the observed data. These two measures were applied to the disaggregation model for when both single-site and regionalised information was used.

Owing to the stochastic nature of the disaggregation models, 100 disaggregated series were generated for each test location and a frequency analysis performed. High-Low bar graphs depicting the observed moments and the 25<sup>th</sup> and 75<sup>th</sup> non-exceedance percentiles of the 100 sets of disaggregated data were used to graphically depict the adequacy of the disaggregation model.

##### **7.4.1 Application of the Model Using At-Site Short Duration Data**

The results from the model where at-site short duration data are available indicate that the model is able to produce synthetic hourly data which resembles the general distribution of the observed hourly data for a particular site. However, the results also indicate that the model is less capable of simulating some of the statistics considered i.e. the probability of dry periods and design rainfalls for selected return periods. It was shown in Chapter 4, that the  $R_{\text{mean}}$  value is directly related to the shape of the distribution of  $R$ . In Section 6.3, it was noted that there appeared to be a correlation between the  $R_{\text{mean}}$  values and the ability to simulate extreme design rainfalls, which were presented in Table 6-7. Station 0028748 had an  $R_{\text{mean}}$  of 0.447 and a *MARE* of 163.08%, compared to the *MARE* of 14.36% observed at Station 0474680, which had an  $R_{\text{mean}}$  of 0.596. Hence, it is postulated that this is owing to the manner in which the distributions of  $R$  are derived.

The relatively poor results obtained for the various lag autocorrelations from the disaggregated rainfall is postulated to be the result of the way that the hourly rainfalls are sequenced. Although it was shown that the lag autocorrelations were better simulated for longer and better quality data, it is suggested that a different method of sequencing the disaggregated hourly rainfalls may improve the model in this area.

#### **7.4.2 Application of the Model Using Regionalised Input**

The results obtained when using the regionalised input in the model tended to be similar to those obtained when at-site information were available, for both statistical and extreme value measures of model performance. This is a positive result as it implies that the model can be used to reasonably disaggregate daily rainfall at locations where no short duration data are available.

Although the use of regionalised input is able to adequately represent the distribution of R at the various test sites, the method of regionalisation fails to adequately capture the distribution of the hour of maximum rainfall. This was observed for Station 0106880, shown in Figure 6-5, where the use of at-site information in the disaggregation model produced a *MARE*, for the 1-hourly duration, over 2.5 times less than that produced by the disaggregation model when regionalised input was used.

#### **7.5 Summary**

From this study it can be concluded that:

- The temporal distribution of rainfall is an important factor affecting the timing and magnitude of peak discharge from a catchment.
- Owing to the relative paucity of short duration rainfall data, a methodology for the disaggregation of daily rainfall to produce hourly increments which aggregate to equal the observed daily values has been identified and applied in South Africa.
- The methodology is shown to reproduce the general distribution of rainfall relatively well, both when observed short duration data are available as well as in the absence of such information.

- Refinements need to be made to the methodology in order to improve upon the shortfalls which have been highlighted.

## 7.6 Recommendations for Future Research

Owing to the model being developed using data extracted for 0:00 to 0:00 24-hour periods, as opposed to 8:00 to 8:00 periods, it is recommended that the model should be tested on daily rainfall before being applied in practice. This is because totals of 24-hour rainfall, i.e. any 24-hour period, are not necessarily the same as the daily rainfall, which conventionally in South Africa is recorded for the 24-hour period ranging from 8:00 to 8:00. Hence, the direct application of the model developed in this study may not yield satisfactory results when applied to daily rainfall totals for 08:00 to 08:00 periods. If testing shows the direct application of the model to be unsatisfactory, it is recommended that the model be redeveloped using 8:00 to 8:00 periods.

It is postulated that the weakness of the model in simulating both extremes of the rainfall spectrum (dry probabilities and design rainfall) is a result of the use of a single distribution of  $R$  to represent an entire range of magnitudes within a rainfall record. It is recommended that, for a particular rainfall record, the data be collated according to the daily rainfall total, using pre-determined ranges, and a distribution of  $R$  be calculated for each of these ranges. It is further recommended that more research be done on how to sequence the disaggregated hourly rainfalls, in order to improve the simulation of the structure of the rainfall, as measured by the lag autocorrelations, number of events and event durations.

As noted in Chapter 5, the ranges used when regionalising the  $R_{\text{mean}}$  values were the same as the ranges used to collate the values of  $R$  in Chapter 4. It is recommended that this not be done in the future as it becomes confusing when trying to differentiate between the collation  $R$  and the collation of  $R_{\text{mean}}$ . Furthermore, it was noted in Section 7.4.2, that the distribution of the hour of maximum rainfall for certain stations is lost using this method of regionalisation. In this study it was assumed that the distribution was related to the  $R_{\text{mean}}$  value. In hindsight, it is now postulated that this distribution is probably more related to geographic location than to the mean value of  $R$ . It is thus suggested that a different approach to regionalising the methodology should be attempted. A possible starting point could be to regionalise the

distributions of R and the hour of maximum rainfall according to the 15 relatively homogeneous short duration rainfall clusters identified by Smithers and Schulze (2000a).

It is further recommended that the disaggregation model be linked with a reliable daily rainfall generator. This would facilitate the generation of long sequences of hourly data for any location in South Africa in an attempt to improve water resources modelling and design flood estimation.

## 8. REFERENCES

- Arnbjerg Nielsen, K., Madsen, H. and Harremoes, P. 1998. Formulating and testing a rain series generator based on tipping bucket gauges. *Water Science and Technology*, 37(11): 47-55.
- Bonta, J.V. and Rao, A.R. 1987. Factors affecting development of Huff curves. *Transactions of the American Society of Agricultural Engineers*, 31(3): 756-760.
- Bonta, J.V. and Rao, A.R. 1988. Comparison of four design-storm hyetographs. *Transactions of the American Society of Agricultural Engineers*, 31(1): 102-106.
- Bonta, J.V. and Rao, A.R. 1989. Regionalisation of storm hyetographs. *Water Resources Bulletin*, 25(1): 211-217.
- Boughton, W. 2000a. *Disaggregation of daily rainfall to hourly rainfall for flood studies*. Cooperative Research Centre for Catchment Hydrology, Monash University, Clayton, Victoria, Australia. Working Document 00/2, 25 pp.
- Boughton, W. 2000b. *A model for disaggregating daily to hourly rainfalls for design flood estimation*. Cooperative Research Centre for Catchment Hydrology, Monash University, Clayton, Victoria, Australia. Report 00/15, 36 pp.
- Boughton, W. 2005. Generation of hourly rainfalls for design flood estimation. *29th Hydrology & Water Resources Symposium*, Institution of Engineers. Canberra, Australia.
- Burian, S.J. and Durrans, S.R. 2002. Evaluation of an artificial neural network rainfall disaggregation model. *Water Science and Technology*, 45(2): 99-104.
- Burian, S.J., Durrans, S.R., Nix, S.J. and Pitt, R.E. 2001. Training artificial neural networks to perform rainfall disaggregation. *Journal of Hydrologic Engineering*, 6(1): 43-51.



- Burian, S.J., Durrans, S.R., Tomic, S., Pimmel, R.L. and Wai, C.N. 2000. Rainfall disaggregation using artificial neural networks. *Journal of Hydrologic Engineering*, 5(3): 299-307.
- Campolo, M., Andreussi, P. and Soldati, A. 1999. River flood forecasting with a neural network model. *Water Resources Research*, 35: 1191-1197.
- Chow, V.T., Maidment, D.R. and Mays, L.W. 1988. *Applied hydrology*. McGraw-Hill, New York, USA.
- Connolly, R.D., Schirmer, J. and Dunn, P.K. 1998. A daily rainfall disaggregation model. *Agricultural and Forest Meteorology*, 92(2): 105-117.
- Cordery, I., Pilgrim, D.H. and Rowbottom, I.A. 1984. Time patterns for rainfall for estimating design floods on a frequency basis. *Water Science and Technology*, 16(8-9): 155-165.
- Cowpertwait, P.S.P. 1991. Further developments of the Neyman-Scott clustered point process for modeling rainfall. *Water Resources Research*, 27(7): 1431-1438.
- Cowpertwait, P.S.P. 1998. A Poisson-cluster model of rainfall: High-order moments and extreme values. *Proceedings of the Royal Society, London, A*, 454: 885-898.
- Cronshey, R.G. 1982. Synthetic regional rainfall time distributions. In: V.P. Singh (Editor), *Statistical analysis of rainfall and runoff*. Water Resources Publications, Littleton, Colorado, USA. 55-66 pp.
- Dawson, C.W. and Wilby, R.L. 2001. Hydrological modelling using artificial neural networks. *Progress in Physical Geography*, 25(1): 80-108.
- Entekhabi, D., Rodriguez-Iturbe, I. and Eagleson, P.S. 1989. Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott rectangular pulses model: Parameter estimation and validation. *Water Resources Research*, 25(2): 295-302.
- Frost, A., Srikanthan, R. and Cowpertwait, P.S.P. 2004. *Stochastic generation of point rainfall data at subdaily timescales: a comparison of DRIP and NSRP*. Cooperative

- Research Centre for Catchment Hydrology, Monash University, Clayton, Victoria, Australia. Draft Technical Report, 31 pp.
- Garcia-Guzman, A. and Aranda-Oliver, E. 1993. A stochastic model of dimensionless hyetograph. *Water Resources Research*, 29(7): 2363-2370.
- Glasbey, C., Cooper, G. and McGechan, M. 1995. Disaggregation of daily rainfall by conditional simulation from a point process model. *Journal of Hydrology*, 165: 1-9.
- Görgens, A.H.M. 2002. Design Flood Hydrology. *Design and rehabilitation of dams*, Institute for Water and Environmental Engineering, Department of Civil Engineering, University of Stellenbosch, RSA.
- Gyasi-Agyei, Y. 1999. Identification of regional parameters of a stochastic model for rainfall disaggregation. *Journal of Hydrology*, 223: 148-163.
- Gyasi-Agyei, Y. 2005. Stochastic disaggregation of daily rainfall into one-hour time scale. *Journal of Hydrology*, 309: 178-190.
- Heneker, T.M., Lambert, M.F. and Kuczera, G. 2001. A point rainfall model for risk-based design. *Journal of Hydrology*, 247: 54-71.
- Hershendorff, J. and Woolhiser, D.A. 1987. Disaggregation of daily rainfall. *Journal of Hydrology*, 95(3/4): 299-322.
- Hingray, B., Monbaron, E., Jarrar, I., Favre, A.C., Consuegra, D. and Musy, A. 2002. Stochastic generation and disaggregation of hourly rainfall series for continuous hydrological modelling and flood control reservoir design. *Water Science and Technology*, 45(2): 113-119.
- Hoang, T.M.T. 2001. Joint probability approach to design flood estimation. Unpublished PhD Thesis, Monash University, Australia, 278 pp.

- Hoang, T.M.T., Rahman, A., Weinmann, P.E., Laurenson, E.M. and Nathan, R.J. 1999. Joint probability description of design rainfall. *Conference on Water Resources and Environmental Research*, Brisbane, Australia.
- Huff, F.A. 1967. Time distribution of rainfall in heavy storms. *Water Resources Research*, 3(4): 1007-1019.
- Huff, F.A. 1990. *Time distributions of heavy rainstorms in Illinois*. Illinois State Water Survey, Champaign, Illinois, USA. Circular 173.
- Jennings, S., Lambert, M.F., Frost, A. and Kuczera, G. 2002. Regionalisation of a high resolution point rainfall model. *27th Hydrology and Water Resources Symposium*, Australia.
- Keifer, C.J. and Chu, H.H. 1957. Synthetic storm pattern for drainage design. *Journal of Hydraulic Division - ASCE*, 83(HY 4): 1-25.
- Kottegoda, N.T., Natale, L. and Raiteri, E. 2003. A parsimonious approach to stochastic multisite modelling and disaggregation of daily rainfall. *Journal of Hydrology*, 274: 47-61.
- Koutsoyiannis, D. 2003. Rainfall disaggregation methods: Theory and applications. *Workshop on Statistical and Mathematical Methods for Hydrological Analysis*, Rome, Italy.
- Koutsoyiannis, D. and Manetas, A. 1996. Simple disaggregation by accurate adjusting procedures. *Water Resources Research*, 32(7): 2105-2117.
- Koutsoyiannis, D. and Onof, C. 2000. A computer program for temporal rainfall disaggregation using adjusting procedures. *XXV General Assembly of European Geophysical Society*, Geophysical Research Abstracts, 2, European Geophysical Society. Nice, France.
- Koutsoyiannis, D. and Onof, C. 2001. Rainfall disaggregation using adjusting procedures on a Poisson cluster model. *Journal of Hydrology*, 246: 109-122.

- Koutsoyiannis, D. and Mamassis, N. 2001. On the representation of hyetograph characteristics by stochastic rainfall models. *Journal of Hydrology*, 251: 65-87.
- Koutsoyiannis, D., Onof, C. and Wheater, H.S. 2001. Stochastic disaggregation of spatial-temporal rainfall with limited data. *26th General Assembly of European Geophysical Society*, Geophysical Research Abstracts, 3, European Geophysical Society. Nice, France.
- Lambourne, J.J. and Stephenson, D. 1986. *Research in Urban Hydrology and Drainage: Factors Affecting Storm Runoff in South Africa 1. Short Duration Rainfall*. Water Research Commission, Pretoria, RSA. WRC Report No. 115/7/86, 1-284 pp.
- Loukas, A. 2002. Flood frequency estimation by a derived distribution procedure. *Journal of Hydrology*, 255: 69-89.
- Maier, H.R. and Dandy, G.C. 1996. The use of artificial neural networks for the prediction of water quality parameters. *Water Resources Research*, 32: 1013-1022.
- Midgley, D.C. and Pitman, W.V. 1978. *A depth-duration-frequency diagram for point rainfall in Southern Africa*. University of Witwatersrand, Johannesburg, RSA. HRU Report 2/78, 57 pp.
- Mikkelsen, P.S., Madsen, H., Arnbjerg Nielsen, K., Jorgensen, H.K., Rosbjerg, D. and Harremoes, P. 1998. A rationale for using local and regional point rainfall data for design and analysis of urban storm drainage systems. *Water Science and Technology*, 37(11): 7-14.
- Nagesh Kumar, D. 2003. Strengths, weaknesses of ANN and its potential applications in hydrology. *Proceedings of Workshop on ANN Applications in Hydraulic Engineering*, 6 July, 2003, Pune, India.
- Nagesh Kumar, D. and Srinivasa Raju, K. 2000. Temporal disaggregation of rainfall data using Artificial Neural Networks. *Proceedings of International Conference on*

*Integrated Water Resources Management for Sustainable Development*, New Delhi, India.

Olsson, J. and Burlando, P. 2002. Reproduction of temporal scaling by a rectangular pulse rainfall model. *Hydrological Processes*, 16: 611-630.

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

Onof, C. and Wheater, H.S. 1994. Improvements to the modelling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model. *Journal of Hydrology*, 157: 177-195.

Ormsbee, L.E. 1989. Rainfall disaggregation model for continuous hydrological modeling. *Journal of Hydrologic Engineering*, 115(4): 507-525.

Pilgrim, D.H. and Cordery, I. 1975. Rainfall temporal pattern for design floods. *Journal of Hydraulic Division - ASCE*, 100(HY1): 81-95.

Rahman, A., Weinmann, P.E., Hoang, T.M.T., Laurenson, E.M. and Nathan, R.J. 2001. *Monte Carlo simulation of flood frequency curves from rainfall*. Cooperative Research Centre for Catchment Hydrology, Monash University, Clayton, Victoria, Australia. Report 01/4, 24-26 pp.

Rodriguez-Iturbe, I., Cox, D.R. and Isham, V. 1987. Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society, London, A*, 410: 269-288.

Schmidt, E.J. and Schulze, R.E. 1987. *SCS-based design runoff*. Department of Agricultural Engineering, University of Natal, Pietermaritzburg, RSA. ACRU Report No. 24, 164 pp.

- Schulze, R.E. 1984. *Hydrological models for application to small rural catchments in Southern Africa : Refinements and developments*. University of Natal, Pietermaritzburg. ACRU Report No. 19, 248 pp.
- Schulze, R.E. and Arnold, H. 1979. *Estimation of volume and rate of runoff in small catchments in South Africa, based on the SCS technique*. University of Natal, Pietermaritzburg, RSA. ACRU Report No. 8, 79 pp.
- Schulze, R.E., Schmidt, E.J. and Smithers, J.C. 1992. *PC-Based SCS design flood estimates for small catchments in southern Africa*. Department of Agricultural Engineering, University of Natal, Pietermaritzburg, RSA. ACRU Report 40, 47 pp.
- Smithers, J.C. and Schulze, R.E. 2000a. *Development and evaluation of techniques for estimating short duration design rainfall in South Africa*. Water Research Commission, Pretoria, RSA. WRC Report No. 681/1/00, 356 pp.
- Smithers, J.C. and Schulze, R.E. 2000b. *Long duration design rainfall estimates for South Africa*. Water Research Commission, Pretoria, RSA. WRC Report No. 811/1/00, 69 pp.
- Smithers, J.C. and Schulze, R.E. 2001. Design runoff estimation: A review with references to practices in South Africa. *Tenth South African National Hydrology Symposium, SANCLAHS*, CSIR. Pietermaritzburg, RSA.
- Smithers, J.C. and Schulze, R.E. 2003. *Design rainfall and flood estimation in South Africa*. Water Research Commission, Pretoria, RSA. WRC Report No. 1060/1/03, 156 pp.
- Socolofsky, S., Adams, A. and Entekhabi, D. 2001. Disaggregation of daily rainfall for continuous watershed modeling. *Journal of Hydrologic Engineering*, 6(4): 300 - 309.
- Thirumalaiah, K. and Deo, M.C. 1998. Real-time flood forecasting using neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 13: 101-111.

Walker, S. and Tsubo, M. 2003. *Estimation of rainfall intensity for potential crop production on clay soil with in-field water harvesting practices in a semi-arid area*. Water Research Commission, Pretoria, RSA. WRC Report No. 1049/1/02, 113 pp.

Weddepohl, J.P. 1988. Design rainfall distributions for Southern Africa. Unpublished M.Sc. Thesis, Department of Agricultural Engineering, University of Natal, Pietermaritzburg, RSA, 162 pp.

Woolhiser, D.A. and Osborn, H.B. 1985. A Stochastic model of dimensionless thunderstorm rainfall. *Water Resources Research*, 21(4): 511-522.

Yen, B.C. and Chow, V.T. 1980. Design hyetographs for small drainage structures. *Journal of Hydraulic Division - ASCE*, 106(HY6): 1055-1076.

## APPENDIX A

### SITE CHARACTERISTICS OF STATIONS USED IN MODEL DEVELOPMENT

Organisation	Station No.	Location	Years Record	Latitude			Longitude			ALTITUDE	MAP
				Deg	Min	Sec	Deg	Min	Sec	(m)	(mm)
SAWS	0010425	RIVERSDALE	12	34	5	0	21	15	0	137	377
SAWS	0010456	RIVERSDALE	27	34	6	0	21	16	0	116	416
SAWS	0021054	CAPE TOWN- WINGFIELD	19	33	54	0	18	32	0	17	440
SAWS	0021178	CAPE TOWN D F MALAN	28	33	58	0	18	36	0	46	535
SAWS	0021179	CAPE TOWN D.F.MALAN	11	33	59	0	18	36	0	42	556
SAWS	0021591	ELSENBERG - AGR	32	33	51	0	18	50	0	181	658
SAWS	0023708	ROBERTSON	11	33	48	0	19	54	0	209	345
SAWS	0028428	ROOIHEUWEL - AGR	12	33	38	0	22	15	0	301	348
SAWS	0028690	GEORGE - P.W. BOTHA	15	34	0	0	22	23	0	193	581
SAWS	0034767	UITENHAGE - PUR	40	33	47	0	25	26	0	32	400
SAWS	0035179	PORT ELIZABETH - WK	55	33	59	0	25	36	0	60	611
SAWS	0037541	BATHURST - AGR	31	33	31	0	26	49	0	259	669
SAWS	0043566	MATROOSBURG	37	33	26	0	19	49	0	967	263
SAWS	0044081	TOUWSRIVIER	14	33	21	0	20	3	0	778	256
SAWS	0050887	WILLOWMORE - MUN	37	33	17	0	23	30	0	840	233
SAWS	0059572	EAST LONDON - WK	51	33	2	0	27	50	0	125	874
SAWS	0061298	LANGEBAAWEG WO	20	32	58	0	18	10	0	31	263
SAWS	0074296	JANSENVILLE	26	32	56	0	24	40	0	417	268
SAWS	0076134	SOMERSET EAST - HOSP	35	32	44	0	25	35	0	717	580
SAWS	0079712	KING WILLIAMS TOWN	17	32	52	0	27	24	0	400	594



Organisation	Station No.	Location	Years Record	Latitude			Longitude			ALTITUDE	MAP
				Deg	Min	Sec	Deg	Min	Sec	(m)	(mm)
SAWS	0079811	DOHNE - AGR	33	32	31	0	27	28	0	899	752
SAWS	0088293	SUTHERLAND	38	32	23	0	20	40	0	1459	339
SAWS	0092141	BEAUFORT WEST - TNK	16	32	21	0	22	35	0	857	238
SAWS	0092229	BEAUFORT WEST - WK	11	32	19	0	22	38	0	869	190
SAWS	0096045	GRAAFF-REINET - TNK	25	32	15	0	24	32	0	741	326
SAWS	0098190	CRADOCK-MUN	12	32	10	0	25	37	0	927	312
SAWS	0113025	FRASERBURG - POW	40	31	55	0	21	31	0	1264	181
SAWS	0123654	QUEENSTOWN - TNK	22	31	54	0	26	52	0	1066	520
SAWS	0125409	NCORA FLATS	19	31	49	0	27	44	0	990	648
SAWS	0127485	UMTATA	17	31	35	0	28	47	0	685	595
SAWS	0134478	CALVINIA - TNK	26	31	28	0	19	46	0	980	210
SAWS	0145059	GROOTFONTEIN - AGR	34	31	29	0	25	2	0	1263	354
SAWS	0165898	CARNARVON - AGR	24	30	58	0	22	0	0	1280	204
SAWS	0170009	DÉ AAR - WK	32	30	39	0	24	1	0	1243	303
SAWS	0175371	ALIWAL NORTH - TNK	14	30	41	0	26	43	0	1310	524
SAWS	0175373	ALIWAL NORTH WELVERD	16	30	43	0	26	43	0	1348	511
SAWS	0178689	SHEEPRUN	22	30	59	0	28	23	0	1213	813
SAWS	0180722	KOKSTAD - WILLOWS	21	30	32	0	29	25	0	1304	756
SAWS	0193561	VANWYKSVLEI	35	30	21	0	21	49	0	962	175
SAWS	0207531	MATATIELE	11	30	21	0	28	48	0	1490	838
SAWS	0214636	OKIEP	26	29	36	0	17	52	0	921	173
SAWS	0224430	PRIESKA	31	29	40	0	22	45	0	932	228
SAWS	0229556	FAURESMITH	32	29	46	0	25	19	0	1363	422
SAWS	0233044	WEPENER	36	29	44	0	27	2	0	1438	503

Organisation	Station No.	Location	Years Record	Latitude			Longitude			ALTITUDE	MAP
				Deg	Min	Sec	Deg	Min	Sec	(m)	(mm)
SAWS	0237591	WATERFORD	20	29	51	0	29	20	0	1643	975
SAWS	0237618	SHALEBURN	16	29	48	0	29	21	0	1614	977
SAWS	0239482	CEDARA	46	29	32	0	30	17	0	1076	876
SAWS	0239577	PIETERMARITZBURG-PUR	14	29	37	0	30	20	0	765	949
SAWS	0239756	PIETERMARITZBURG-PUR	19	29	36	0	30	26	0	613	817
SAWS	0240808	LOUIS BOTHA - WK	36	29	58	0	30	57	0	8	986
SAWS	0247668	POFADDER	34	29	8	0	19	23	0	989	130
SAWS	0256424	DOUGLAS	14	29	4	0	23	45	0	994	316
SAWS	0258157	RIETRIVIER - AGR	15	29	7	0	24	36	0	1140	385
SAWS	0261307	BLOEMFONTEIN - WK	24	29	7	0	26	11	0	1422	537
SAWS	0268631	ESTCOURT - AGR	15	29	1	0	29	52	0	1181	700
SAWS	0274034	ALEXANDER BAY - WK	38	28	34	0	16	32	0	21	43
SAWS	0290468	KIMBERLEY - WK	43	28	48	0	24	46	0	1198	414
SAWS	0296005	UINTJIESHOEK	11	28	35	0	27	31	0	1584	639
SAWS	0296583	GLEN MORGAN	11	28	43	0	27	50	0	1676	685
SAWS	0300423	LADYSMITH - MUN	13	28	33	0	29	45	0	1034	768
SAWS	0300454	LADYSMITH - WK	21	28	34	0	29	46	0	1079	734
SAWS	0300690	ESTCOURT - TNK	24	29	0	0	29	53	0	1148	731
UZ	304320	Kwa-Dlangzwa	12	28	50	0	31	41	0	378	1201
UZ	304353	Kwa-Dlangzwa	12	28	53	0	31	42	0	173	1325
UZ	304410	Kwa-Dlangzwa	12	28	50	0	31	44	0	331	1269
UZ	304412	Kwa-Dlangzwa	12	28	52	0	31	44	0	142	1310
UZ	304470	Kwa-Dlangzwa	11	28	50	0	31	46	0	252	1314
UZ	304473	Kwa-Dlangzwa	12	28	53	0	31	46	0	63	1310

Organisation	Station No.	Location	Years Record	Latitude			Longitude			ALTITUDE	MAP
				Deg	Min	Sec	Deg	Min	Sec	(m)	(mm)
UZ	304501	Kwa-Dlangzwa	12	28	51	0	31	47	0	142	1320
UZ	304530	Kwa-Dlangzwa	12	28	50	0	31	48	0	142	1243
UZ	304562	Kwa-Dlangzwa	12	28	52	0	31	49	0	95	1384
UZ	304593	Kwa-Dlangzwa	12	28	53	0	31	50	0	95	1476
UZ	304622	Kwa-Dlangzwa	12	28	52	0	31	51	0	95	1390
SAWS	0305168	RICHARDS BAY	13	28	47	30	32	6	0	47	1226
SAWS	0317476	UPINGTON - WK	18	28	26	0	21	16	0	814	180
SAWS	0323102	KOOPMANSFONTEIN II	39	28	12	0	24	4	0	1341	419
SAWS	0330421	ROODEPOORT	11	28	1	0	27	45	0	1569	672
SAWS	0330843	CHICAGO	11	28	3	0	27	59	0	1615	616
SAWS	0331520	LOCH LOMOND - AGR	27	28	10	0	28	18	0	1631	662
SAWS	0331585	BETHLEHEM - WO	13	28	15	0	28	20	0	1680	670
SAWS	0337143	BABANANGO	15	28	23	0	31	5	0	1288	883
SAWS	0360453	TAUNG	11	27	33	0	24	46	0	1124	453
SAWS	0362710	HOOPSTAD - TNK	13	27	50	0	25	54	0	1239	446
SAWS	0363239	PLESSISDRAAI	19	27	59	0	26	8	0	1249	479
SAWS	0365430	KROONSTAD - MUN	26	27	40	0	27	15	0	1348	593
SAWS	0370734	NEWCASTLE	11	27	44	0	29	55	0	1235	846
SAWS	0370765	NEWCASTLE - TNK	13	27	45	0	29	56	0	1197	818
SAWS	0393778	KURUMAN	26	27	28	0	23	26	0	1312	480
SAWS	0403537	CILLIERSRUS	11	27	27	0	28	18	0	1630	617
SAWS	0403886	FRANKFORT - TNK	37	27	16	0	28	30	0	1500	647
SAWS	0411323	MAKATINI	15	27	23	0	32	11	0	63	558
SAWS	0411324	MAKATINI - AGR	16	27	24	0	32	11	0	73	571

Organisation	Station No.	Location	Years Record	Latitude			Longitude			ALTITUDE	MAP
				Deg	Min	Sec	Deg	Min	Sec	(m)	(mm)
SAWS	0432237	ARMOEDSVLAKTE	36	26	57	0	24	38	0	1234	437
SAWS	0435157	DOORNLAAGTE	16	26	37	0	26	6	0	1473	574
SAWS	0437104	POTCHEFSTROOM	15	26	44	0	27	4	0	1350	618
SAWS	0437134	POTCHEFSTROOM - AGR	31	26	44	0	27	5	0	1345	618
SAWS	0438553	VANDEBYLPARK - PUR	12	26	43	0	27	49	0	1496	674
SAWS	0441416	STANDERTON - SKL	15	26	56	0	29	14	0	1570	610
SAWS	0442811	NOOITGEDACHT - AGR	28	26	31	0	29	58	0	1694	722
SAWS	0444540	PIET RETIEF - TNK	21	27	0	0	30	48	0	1235	887
SAWS	0475456	KRUGERSDORP KRONINGS	40	26	6	0	27	46	0	1699	798
SAWS	0476042	JHB - BURGERSDORP	16	26	12	0	28	2	0	1719	701
SAWS	0476131	JHB - BEZUIDENHOUT V	17	26	11	0	28	5	0	1700	784
SAWS	0476398	JAN SMUTS - WK	33	26	8	0	28	14	0	1692	696
SAWS	0478867	BETHAL	25	26	27	0	29	29	0	1663	689
SAWS	0480184	CAROLINA	32	26	4	0	30	7	0	1696	749
SAWS	0508047	MMABATHO - AER	13	25	47	0	25	32	0	1281	503
SAWS	0508261	MAFIKENG - TNK	11	25	51	0	25	39	0	1278	585
SAWS	0511523	RUSTENBERG - AGR	45	25	43	0	27	18	0	1157	639
SAWS	0513314	PRETORIA - FORUM	29	25	44	0	28	11	0	1330	674
SAWS	0513385	IRENE -WK	19	25	55	0	28	13	0	1524	667
SAWS	0513405	PRETORIA - BROOKLYN	37	25	45	0	28	14	0	1372	765
SAWS	0513465	PRETORIA - UNIV PROE	31	25	45	0	28	16	0	1372	687
SAWS	0513531	RIETVLEI - AGR	20	25	51	0	28	18	0	1524	743
SAWS	0513605	ROODEPLAAT - AGR	25	25	35	0	28	21	0	1164	653
SAWS	0548290	PILANESBERG	12	25	20	0	27	10	0	1043	611

Organisation	Station No.	Location	Years Record	Latitude			Longitude			ALTITUDE	MAP
				Deg	Min	Sec	Deg	Min	Sec	(m)	(mm)
SAWS	0554816	LYDENBERG - VIS	31	25	6	0	30	28	0	1439	670
SAWS	0555837	NELSPRUIT - AGR	14	25	27	0	30	58	0	660	750
SAWS	0555866	NELSPRUIT-FRIEDENHEI	20	25	26	0	30	59	0	671	752
SAWS	0589594	WARMBAD - TOWOOMBA -	51	24	54	0	28	20	0	1143	629
SAWS	0593489	TSWELOPELE	11	24	39	0	30	17	0	700	566
SAWS	0596179	SKUKUZA	38	24	59	0	31	36	0	263	526
SAWS	0631791	GROENDRAAI	11	24	11	0	27	57	0	1025	546
SAWS	0634011	POTGIETERUS-TABAK CO	33	24	11	0	29	1	0	1116	624
SAWS	0674311	ELLISRAS	11	23	41	0	27	41	0	849	471
SAWS	0677802	PIETERSBURG - WK	39	23	52	0	29	27	0	1230	458
SAWS	0677866	PIETERSBURG - WK	14	23	56	0	29	29	0	1294	446
SAWS	0679260	TZANEEN	13	23	50	0	30	9	0	716	972
SAWS	0679289	PUSELLA - GSH	14	23	49	0	30	10	0	749	1015
SAWS	0681266	PHALABORWA	24	23	56	0	31	9	0	407	531
SAWS	0719369	MARNITZ	14	23	9	0	28	13	0	946	388
SAWS	0722099	MARA	36	23	9	0	29	34	0	897	438
SAWS	0723485	LEVUBU	32	23	5	0	30	17	0	706	882
SAWS	0766898	THOHOYANDOU	15	22	58	0	30	30	0	600	812
SAWS	0767046	TSHANDAMA	12	22	46	0	30	32	0	600	555
SAWS	0809706	MESSINA - MACUVILLE	32	22	16	0	29	54	0	525	345
CTCE	athlone	Athlone	40	33	57	11	18	30	55	14	638
UKZN	c161	CEDARA	15	29	35	13	30	13	38	1340	974
UKZN	c162	CEDARA	20	29	34	40	30	13	53	1207	913
UKZN	c165	CEDARA	20	29	33	0	30	14	45	1130	848

Organisation	Station No.	Location	Years Record	Latitude			Longitude			ALTITUDE	MAP
				Deg	Min	Sec	Deg	Min	Sec	(m)	(mm)
UKZN	c172	CEDARA	20	29	34	10	30	15	50	1175	883
UKZN	c181	CEDARA	11	29	35	43	30	15	43	1445	906
UKZN	c182	CEDARA	20	29	35	18	30	14	50	1261	957
UKZN	c191	CEDARA	20	29	32	37	30	16	34	1058	873
UKZN	c201	CEDARA	20	29	32	40	30	16	57	1121	873
CSIR	cp6br	CathPeak CP6	32	28	59	15	29	15	7	1920	1046
UKZN	d1	DE HOEK	11	29	0	7	29	39	55	1201	925
CSIR	jnk19a	JnkBiesievlei	52	33	58	21	18	56	56	282	1095
CSIR	moko3a	Mokobulaan	29	25	16	15	30	34	0	1359	1004
UKZN	n11	NTABAMHLOPE	19	29	0	44	29	37	38	1529	851
UKZN	n18	NTABAMHLOPE	20	29	2	26	29	39	43	1448	1103
UKZN	n20	NTABAMHLOPE	11	29	1	10	29	40	21	1473	859
UKZN	n23	NTABAMHLOPE	31	29	3	29	29	39	23	1456	900
CTCE	newlands	Newlands	20	33	58	1	18	27	3	140	973
SASRI	sall0	La Mercy	20	29	36	0	31	1	0	81	937
SASRI	samte	Mt_Edgecombe	19	29	42	0	31	2	0	96	951
SASRI	samtz	Mtunzini	14	28	56	0	31	42	0	36	1338

## APPENDIX B

### AVERAGE RANKED SERIES OF HOURLY FRACTIONS

The following shows the average ranked series of hourly fractions from 157 stations for each of the 20 ranges of R. For a given value of R, the following series give the fractions of the daily total for the other 23 hours ranked in order of magnitude.

The results are displayed in the following format:

Range of R

Ranked values for 12 highest hours in order 1 to 12

Ranked values for other 12 hours in order 13 to 24

Range 1: 0.0417 – 0.075

0.066	0.064	0.063	0.062	0.061	0.060	0.058	0.056	0.054	0.051	0.049	0.048
0.047	0.044	0.042	0.038	0.027	0.024	0.021	0.019	0.016	0.014	0.011	0.007

Range 2: 0.075 – 0.125

0.107	0.097	0.089	0.082	0.076	0.069	0.064	0.058	0.053	0.048	0.043	0.038
0.033	0.028	0.024	0.021	0.017	0.014	0.011	0.009	0.007	0.005	0.003	0.002

Range 3: 0.125 – 0.175

0.153	0.131	0.114	0.100	0.087	0.075	0.064	0.055	0.046	0.037	0.031	0.025
0.020	0.016	0.012	0.010	0.007	0.006	0.004	0.003	0.002	0.002	0.001	0.001

Range 4: 0.175 – 0.225

0.201	0.164	0.135	0.111	0.091	0.072	0.056	0.043	0.032	0.024	0.018	0.013
0.010	0.008	0.006	0.004	0.003	0.002	0.002	0.001	0.001	0.001	0.000	0.000

Range 5: 0.225 – 0.275

0.250	0.194	0.151	0.115	0.084	0.060	0.042	0.030	0.021	0.015	0.011	0.008
0.005	0.004	0.003	0.002	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000

Range 6: 0.275 – 0.325

0.300	0.221	0.161	0.110	0.071	0.045	0.030	0.020	0.013	0.009	0.006	0.004
0.003	0.002	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000

Range 7: 0.325 – 0.375

0.350	0.244	0.159	0.095	0.056	0.034	0.021	0.013	0.009	0.006	0.004	0.003
0.002	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 8: 0.375 – 0.425

0.399	0.259	0.149	0.080	0.044	0.025	0.015	0.009	0.006	0.004	0.003	0.002
0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 9: 0.425 – 0.475

0.449	0.272	0.131	0.064	0.034	0.018	0.010	0.006	0.004	0.003	0.002	0.001
0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 10: 0.475 – 0.525

0.499	0.280	0.108	0.050	0.025	0.014	0.008	0.005	0.003	0.002	0.001	0.001
0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 11: 0.525 – 0.575

0.549	0.267	0.093	0.041	0.020	0.011	0.006	0.004	0.003	0.002	0.001	0.001
0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 12: 0.575 – 0.625

0.599	0.240	0.082	0.036	0.017	0.009	0.005	0.003	0.002	0.001	0.001	0.001
0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 13: 0.625 – 0.675

0.649	0.222	0.068	0.028	0.013	0.007	0.004	0.002	0.002	0.001	0.001	0.001
0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000



Range 14: 0.675 – 0.725

0.699	0.192	0.057	0.023	0.011	0.006	0.003	0.002	0.001	0.001	0.001	0.001
0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 15: 0.725 – 0.775

0.749	0.159	0.048	0.019	0.009	0.005	0.003	0.002	0.001	0.001	0.001	0.001
0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 16: 0.775 – 0.825

0.800	0.132	0.036	0.014	0.006	0.003	0.002	0.001	0.001	0.001	0.001	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 17: 0.825 – 0.875

0.849	0.102	0.027	0.010	0.004	0.002	0.002	0.001	0.001	0.001	0.001	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 18: 0.875 – 0.925

0.900	0.069	0.016	0.006	0.003	0.002	0.001	0.001	0.001	0.001	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 19: 0.925 – 0.975

0.952	0.031	0.008	0.003	0.002	0.001	0.001	0.001	0.001	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 20: 0.975 – 1.000

0.989	0.006	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

## APPENDIX C

### CLUSTERED SEQUENCES OF HOURLY FRACTIONS

Using all days where the 24-hour total was  $\geq 1$  mm from all 157 stations, the fractions of the daily totals in the maximum 2, 3, 6, and 12-hours were averaged in each range of R. The average ranked series, shown in Appendix B, were then rearranged such that the maximum 2, 3, 6, and 12-hour totals matched the average values calculated from the observed data.

The following clustered sequences are displayed in the following format:

Range of R

The maximum hourly fraction R is the first value on the first string of values, followed by the fraction which combines with R to form the maximum 2-hour total, then the value which forms the maximum 3-hour total with the first two values, etc.

Range 1: 0.0417 – 0.075

0.066 0.063 0.060 0.062 0.061 0.047 0.058 0.056 0.051 0.049 0.038 0.024  
0.064 0.054 0.048 0.044 0.042 0.027 0.021 0.019 0.016 0.014 0.011 0.007

Range 2: 0.075 – 0.125

0.107 0.082 0.076 0.069 0.058 0.053 0.097 0.048 0.038 0.033 0.024 0.021  
0.089 0.064 0.043 0.028 0.017 0.014 0.011 0.009 0.007 0.005 0.003 0.002

Range 3: 0.125 – 0.175

0.153 0.114 0.087 0.131 0.055 0.012 0.075 0.064 0.046 0.037 0.007 0.006  
0.100 0.031 0.025 0.020 0.016 0.010 0.004 0.003 0.002 0.002 0.001 0.001

Range 4: 0.175 – 0.225

0.201 0.135 0.091 0.164 0.056 0.003 0.111 0.043 0.032 0.008 0.004 0.002  
0.072 0.024 0.018 0.013 0.010 0.006 0.002 0.001 0.001 0.001 0.000 0.000

Range 5: 0.225 – 0.275

0.250	0.151	0.115	0.194	0.015	0.005	0.084	0.060	0.008	0.004	0.003	0.000
0.042	0.030	0.021	0.011	0.002	0.002	0.001	0.001	0.001	0.000	0.000	0.000

Range 6: 0.275 – 0.325

0.300	0.161	0.110	0.221	0.000	0.000	0.045	0.030	0.020	0.013	0.009	0.004
0.071	0.006	0.003	0.002	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000

Range 7: 0.325 – 0.375

0.350	0.159	0.095	0.244	0.000	0.000	0.056	0.021	0.002	0.000	0.000	0.000
0.034	0.013	0.009	0.006	0.004	0.003	0.001	0.001	0.001	0.001	0.000	0.000

Range 8: 0.375 – 0.425

0.399	0.149	0.080	0.259	0.000	0.000	0.025	0.015	0.004	0.002	0.001	0.000
0.044	0.009	0.006	0.003	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000

Range 9: 0.425 – 0.475

0.449	0.272	0.034	0.131	0.000	0.000	0.064	0.000	0.000	0.000	0.000	0.000
0.018	0.010	0.006	0.004	0.003	0.002	0.001	0.001	0.001	0.001	0.000	0.000

Range 10: 0.475 – 0.525

0.499	0.280	0.025	0.108	0.000	0.000	0.050	0.000	0.000	0.000	0.000	0.000
0.014	0.008	0.005	0.003	0.002	0.001	0.001	0.001	0.001	0.001	0.000	0.000

Range 11: 0.525 – 0.575

0.549	0.267	0.020	0.093	0.000	0.000	0.011	0.006	0.004	0.003	0.002	0.001
0.041	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 12: 0.575 – 0.625

0.599	0.240	0.017	0.082	0.000	0.000	0.009	0.005	0.003	0.002	0.001	0.001
0.036	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 13: 0.625 – 0.675

0.649	0.222	0.013	0.068	0.000	0.000	0.007	0.004	0.002	0.002	0.001	0.001
0.028	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 14: 0.675 – 0.725

0.699	0.192	0.006	0.057	0.000	0.000	0.023	0.001	0.000	0.000	0.000	0.000
0.011	0.003	0.002	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000

Range 15: 0.725 – 0.775

0.749	0.159	0.009	0.048	0.000	0.000	0.005	0.003	0.002	0.001	0.001	0.001
0.019	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 16: 0.775 – 0.825

0.800	0.132	0.006	0.036	0.000	0.000	0.003	0.002	0.001	0.001	0.001	0.001
0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 17: 0.825 – 0.875

0.849	0.102	0.001	0.010	0.004	0.002	0.027	0.000	0.000	0.000	0.000	0.000
0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 18: 0.875 – 0.925

0.900	0.069	0.001	0.006	0.003	0.001	0.016	0.000	0.000	0.000	0.000	0.000
0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 19: 0.925 – 0.975

0.952	0.031	0.001	0.003	0.002	0.000	0.008	0.000	0.000	0.000	0.000	0.000
0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Range 20: 0.975 – 1.000

0.989	0.006	0.001	0.001	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000