THE FEASIBILITY OF STOCHASTICALLY MODELLING THE SPATIAL AND TEMPORAL DISTRIBUTION OF RAINFIELDS

by

GGS Pegram and AW Seed
Department of Civil Engineering
University of Natal
Durban

Report to the Water Research Commission on the Project

"Development of Models to Stochastically Generate Spatially Distributed Rainfields"

WRC Report No. 550/1/98
ISBN 1 86845 376 6

ACKNOWLEDGEMENTS

This report resulted from a research project funded by the Water Research Commission. The original title of the contract was "Development of Models to Stochastically Generate Spatially Distributed Rainfields"

One of us (Alan Seed) emigrated to New Zealand in 1993 and the other one (Geoff Pegram) was tempted to cancel the contract and return the funds, seeing only an impossible task ahead in conducting an approximation to antipodean research. It was only due to the patience and encouragement of the Director of the project, Dr George Green, that this task was completed. He persuaded and enabled us to get a paper into the European Geophysical Society meeting in Grenoble in April 1994 and made it possible for us to get together at a crucial time in September 1995 which also coincided with the WRC workshops "Characterizing spatially distributed rainfall" and "Hydrological uses of weather radar". We wish to acknowledge, with gratitude, the beneficial influence that Dr George Green had on this project.

Without question, we wish to thank the Water Research Commission for the funding which made the realization of this project possible. Water Science in South Africa is indebted to the Commission for its collective leadership and enablement through funding and guidance of us researchers. We are pleased to have been recipients of the Commission's benificence.

EXECUTIVE SUMMARY

How can rainfall be measured accurately?

Rainfall has traditionally been measured by raingauge. These are good instruments for understanding rainfall processes in time at various points, but are poor (even in fairly dense networks) at giving reliable information about the spatial distribution of rainfall in small time intervals. This is because of the great variability of rainfall in space as well as in time. In the recent past, radar has been used, with increasing success, in the measurement of rainfall. Not only does the technology give rainfall amounts at any specific location to a high degree of accuracy with fine time-resolution, it gives an accurate (and to those accustomed to raingauge data, a revolutionary) estimate of the spatial distribution of rainfall over large areas. This is something that raingauges cannot provide, except when they are used to give average values over relatively long intervals of time like months or years. More worrying is the reality that the number of daily reporting raingauges in South Africa is declining rapidly.

Why model rainfall?

We model the measurements of complex physical processes (in particular climate and rainfall) in order to understand what they mean. With understanding comes:

- an ability to justify simplifying assumptions about the structures
- an ability to mimic the behaviour of the processes for simulation purposes (for example in generating possible future storm patterns on a computer to train flood forecasters)
- an ability to use the structure of the processes to predict behaviour in a general way and to forecast in the short term (for example to increase the lead time in anticipating the arrival of severe storms).

What were the aims of the project?

As set out in the contract document, we wished to continue with the development of a rainday model (suggested for an earlier WRC contract) into a more complete model able to preserve the temporally varying rainfall distribution throughout the year. A second aim was to investigate methods to stochastically generate rainfields to fit a random set of point measurements in space.

As it turned out, the second aim was modified during the project to the fitting of a spatial model to the images of rainfall measured by radar rather than by networks of raingauges.

How was the research undertaken?

The task was split into achieving two objectives in series:

- modelling daily rainfall amounts over a large area using a climate model to describe occurrences of dry days, scattered raindays (due mainly to thunderstorm type rainfall) and general raindays
- modelling the spatial distribution of rainfall over a large area at given time-slices to replicate the variability and clustering nature of the rainfall process as observed by radar.

The first of these objectives was successfully achieved; the second was achieved with partial success, enough to satisfy the secondary aim of the project and to give good insight into the prerequisites for spatial modelling to be undertaken in a follow-up contract.

Conclusions and recommendations

We conclude that, in general, the temporal aspect of a space-time modelling of rainfall using a climate component has been successful. It has been shown that it is possible to model the temporal processes well. The rainfall amounts modelled by the space-time model are by and large fairly well modelled with minor exceptions. The combined model, as it stands, could be applied, fitted and used with some extra polishing. What this report shows is that the relatively simple approach that we adopted pays off.

When it comes to modelling the true spatial character of rainfall, we now thank that we are very unlikely to be able to use raingauge data as a basis for determining the spatial structure of rainfall - the raingauge networks are simply too sparse, so recourse has to be made to radar imagery. Preliminary attempts at modelling the space component of the rainfall process are very encouraging and these results will underpin the new work that is aimed at solving the space-time modelling problem.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

In applications of information derived from rainfall data in the fields of hydrology, engineering and agriculture, it is becoming increasingly important to know (or at least have a reasonable estimate of) the rainfall in space as well as time, in more detail than it is possible to deduce from the data collected at raingauges in a sparse network.

There is a large library of daily readings of rainfall totals that have been collected over many years at many gauges in Southern Africa. It is of concern that the number of operating gauges has diminished in the recent past and that the trend is likely to continue. To be more precise, the number of daily read raingauges reporting to the Weather Bureau in the 1970s was 1700, which reduced to 500 in the mid-1990s. Fortunately there is a relatively new methodology available for measuring rainfall spatially as well as temporally - radar.

There has been a period of overlap in radar and raingauge measurements for some time which is now reaching the stage where meaningful links can be found between the two sets. The advantage of this is that where there are sparse networks of gauges it should be possible to interpolate rainfall spatially, even when there is (was) no radar coverage. This means that the historical raingauge data can be used to infill spatially, if the links can be satisfactorily formed. Traditional mathematical methods of interpolation are not useful for this purpose. They give too smooth a surface over rainfields accumulated in periods shorter than a month in the case of convective rain and in periods shorter than a few days in the case of stratiform rain. Some other technique must be found.

Where there is a good rain measuring radar (as there is in the guise of the MRL5, S-band radar sited near Bethlehem in RSA) it has been shown that there is very good correspondence between the rainfall estimated by the radar and that collected from a dense network of tipping-bucket raingauges. The inference is that where there is such a radar, precipitation measurement is accurate and gives good information spatially, provided the topography is suitable. The problem is to estimate the spatial distribution of rainfall where there is (or was) no radar.

The purpose of this report is to set the stage for forging meaningful links between the rainfall measured by radar and by raingauges in order to build a good stochastic rainfall generator operating both in time and space. In the words (slightly paraphrased) of the contract document defining the research, we wished to continue with the "development of a rainday model (suggested for an earlier WRC contract) into a more complete model able to preserve the temporally varying rainfall distribution throughout the year. A second aim was to investigate methods to stochastically generate rainfields to fit a random set of point measurements in space. As it turned out, the second aim was modified during the project to the fitting of a spatial model to the images of rainfall measured by radar rather than by networks of raingauges."

Although it is desirable to work in finer time-scales than one day, that effort will be deferred to the future. Here it is the intention to outline the strategy for forming the links between gauged and radar-measured daily rainfall totals with the purpose of producing a working stochastic generating model that will produce daily rainfall maps that mimic the historically collected data from gauge and radar. Attention will be confined to the Bethlehem weather radar and the (between 158 and 330) daily read raingauges it covers in a 150 km sided square.

The intention is to model the rainfall spatially on a daily basis over several years, incorporating seasonality and spatial variation due to topography. To put this in other words, the model must replicate at each point, in the region covered by this study, a daily rainfall record that has the same *characteristics* as that measured by the radar or by a raingauge that may be at that position.

The steps covered so far in the journey to achieve the objective of modelling the full correspondence between measurements of rain by gauge and radar are recorded in the remainder of this report. In Chapter 2 the model used by Seed (1992) is described. It is known that the temporal and spatial distribution of rainfall depends on the type of causative weather system; to pool all the data would be to destroy important structure. Therefore, before working with the relatively short set of radar data, the set of raingauge data (consisting of a set which numbered between 158 and 330 raingauges) was examined to determine the frequency of the different weather types. As a first step, these were classified by Seed (1992) as Dry days, Scattered raindays and General raindays after Court (1979). It is possible that in future work a finer classification might be used, but for the purposes of this first exploratory model, a three-state classification was thought to be adequate. The efficacy of the model is proved in validation tests, and it is found that although the character of the rainfall has been captured there is a need to model the seasonal variation.

Chapter 3 describes the extension of the model to allow for time-varying parameters using finite Fourier series with a small number of harmonics. The model exploits the same strategy as the earlier one, namely the generation of a sequence of daily weather types. For each weather type, the model randomly selects (with replacement) an appropriate historical day's pattern of rainfalls. Statistical comparisons were made between the average rainfalls over the set of gauges as recorded and as generated and were found to be "close" to each other, even when using such tools as the sampling distributions of annual maxima over runs of various numbers of days.

The conclusion drawn from Chapter 3 is that the daily weather model is useful as a pilot model for simulation of daily rainfall over a large area (22 500 square kilometres) and will faithfully reproduce the temporal distribution of rainfall at each of the raingauges. The surprise of the modelling procedure was that, unlike the models for individual raingauges introduced by Woolhiser and Pegram (1979) and further developed to operational status by Zucchini and Adamson (1984), the three-state climate model used here is described by only one lag in the Markov chain. It is thus easy to use.

Chapter 4 introduces the idea of random (fractal) cascades to model rainfall spatially and also temporally. The information that is modelled is the rainfall measured by the radar at Bethlehem. The measurements of reflectivity from the precipitation are converted to rain-rate in minute-squares (approximately 1 km square) on the 200 kilometre square grid centred on the radar, computed at a constant altitude above the radar - typically 3 km. These CAPPIs (Constant Altitude Plan Position Indicators) are collected every 4 to 5 minutes using the MRL5 radar, while rain is falling, giving very detailed information in time and space. Rather than integrate these up to daily totals and model them, it was decided to model the instantaneous CAPPIs taken at four minute intervals, then if that is successful, aggregate each in the future, to test the correspondence.

The full fractal description is the work being done in a new contract with WRC, so it is not the purpose of this report to attempt to pre-empt that. The initial work on fractals reported here was done by Seed and colleagues at the University of Auckland and demonstrated by Seed and Pegram at a workshop on spatial rainfall characterization held in September 1995 at the WRC in Pretoria, so forms part of this exploratory study.

In summary then, this report examines the feasibility of usefully modelling rainfall in time and space, provides some partial answers and confirms that some methodologies such as climate modelling of raingauge data and the spatial description of CAPPIs by fractals are not only feasible but practically realizable.

It remains for the researchers involved with the next contract to polish the methodologies and come up with the final working results. Nevertheless, a brief outline of the (possible) linkages between the climate model and the fractal description of CAPPIs will now be given. This plan, it is thought, will enable the objective of a fully operational seasonal spatial daily rainfall model to be achieved. Once a good relationship between the fractals and the CAPPIs has been formed, the link will be sought between the daily rainfall pattern (as a start described by a few parameters - more than the three describing dry, scattered and general raindays) and the fractal pattern (also described by a small set of parameters). This will only be possible using the relatively short overlapping radar and gauged records.

The technique that is likely to be used is a form of classification using the covariance biplot that has been successfully applied to rainfall classification in the past (Pegram and Pegram, 1993, Basson et al., 1994 and Pegram, 1997). Some of what needs to be done here is to derive a meaningful spatial interpolation algorithm to compute plausible rainfalls at ungauged locations based on rain measured at gauges in a network. It is not sufficient to interpolate between gauges using the traditional techniques of polynomial, Fourier, Spline or multi-quadrics in two dimensions - they are too *smooth*. A form of fractal interpolation is sought which has the same character as the variability seen in the CAPPIs. Once meaningful interpolation is achieved, raingauge data might be filled in spatially where it has not been measured by gauge or radar thus achieving the objective of this sequence of studies.

Chapter 2

THE FIRST CLIMATE MODEL

2.0 INTRODUCTION.

The model described here was devised by us and used by Seed (1992) in a desk-top study to test the efficacy of rainfall augmentation by seeding.

The basis of the model is the classification adopted by Seed (1992) which is formed by lumping two states of the four state classification suggested by Court (1979) using raingauge data based on the classification of weather type proposed for BEWMEX by Hudak and Steyn (1978). Seed's explanation is repeated in the following indented paragraphs for completeness and ease of explanation, where the table and section numbering has been changed to fit in with this report.

2.1 CLASSIFICATION OF SEEDABLE AND NON-SEEDABLE RAINDAYS

> An early classification proposed by Hudak and Steyn (1978) for the BEWMEX project was based on cloud formations associated with convection. The classification scheme is listed in Table 2.1. Clearly, this classification is not helpful when attempting to classify raindays on the basis of raingauge data. To address this problem, Court (1979) attempted to find a correlation between the weather type as defined in Table 2.1 and the fraction of raingauges in the area reporting rain. Raindays were classified into dry, scattered, isolated, and general raindays after the scheme shown in Table 2.2.
>
> The main conclusions from the study were:
> a) most scattered and isolated days were type 5 days
> b) most type 5 and 6 days were scattered days
> c) most general raindays were either type 6 or type 4.

Table 2.1 Classification scheme proposed for BEWMEX by Hudak and Steyn (1978)

TYPE	DESCRIPTION
1	Blue skies
2	Cumulus Mediocris with tops warmer than -5° C
3	Night-time line storms tracking eastwards
4	General rain
5*	Cumulus development, with tops colder than -5° C but no hail
6	Cumulus development, tops cooler than -5° C with hail

* Type 5 weather is considered to be seedable.

Table 2.2 Classification of raindays using raingauge data only, after Court (1979)

Dry	Less than 3% stations report precipitation
Isolated	Between 3% and 15% of the stations report precipitation
Scattered	More than 15% stations report rain but less than 50% report 5 mm
General	More than 50% of the stations report at least 5 mm

This implies that as a first estimate, the rainfall pattern classification scheme can be used to differentiate seedable (type 5 or isolated and scattered raindays) from non-seedable raindays (dry or general raindays). A problem with this scheme is that the days on which large storms develop, the days with the highest potential for weather modification, could be miss-classified as general raindays, and thereby be excluded from the set of seedable raindays. Since both scattered and isolated raindays are considered to be seedable, a decision was made to combine these raindays into a single class, called "scattered rain" for this project. The classification into dry, scattered, and general raindays will be used to partition rainfall into seedable and non-seedable classes since:

a) it is based on raingauge data which is readily available,
b) it is objective, and
c) the connection to the meteorological situation has already been established.

Although the three-way classification was originally devised by Seed (1992) to enable a desk-top analysis of the effects of weather modification to be performed, we felt that it was a useful first step in creating a simulation model. Future refinements may be found to be desirable, but are not addressed in this report.

The frequency of scattered and general raindays as measured by a gauge network is evidently dependent on the density of the gauge network and the size of the study area. A sparse gauge network will tend to over-estimate the frequency of isolated raindays. The frequency of general raindays is dependent on the size of the study area and the gauge network density, with frequency of general raindays decreasing as the size of the study area is increased. The thirty years of available daily raingauge data for the up to 330 gauges in the study area appearing in Figure 2.1 were examined.

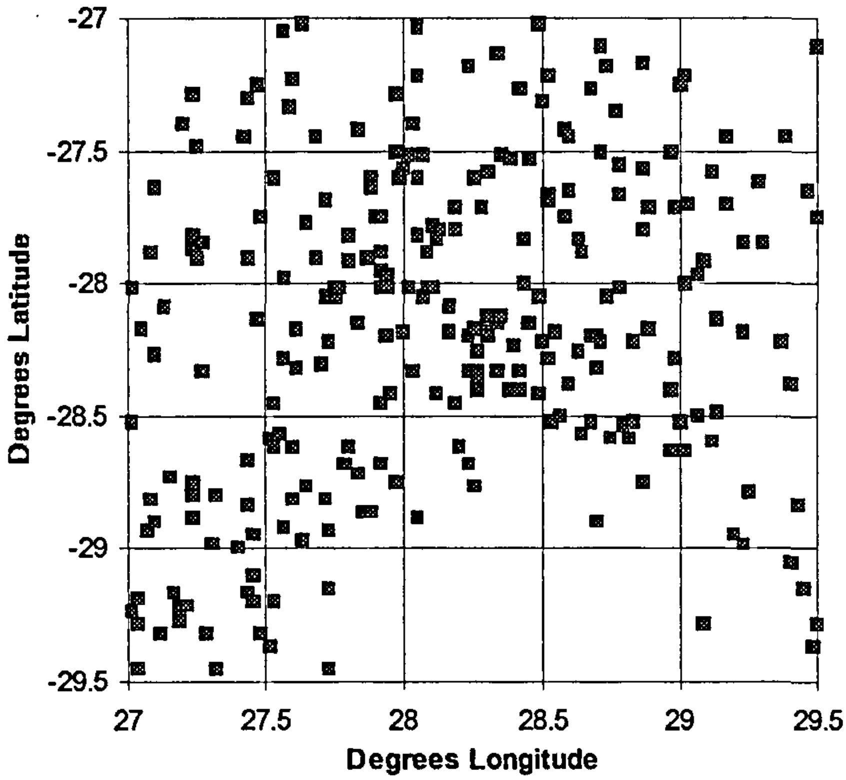


Figure 2.1 The locations of the 330 raingauges used in this study. Only gauges with 10 or more years of data in the 1962-1991 period were selected.

Not all of them were operating all the time as can be seen from Figure 2.2, however the minimum number did not drop below 150. (The sudden increase by 180 gauges in September 1977 was at the commencement of the Bethlehem Weather Modification experiment.)

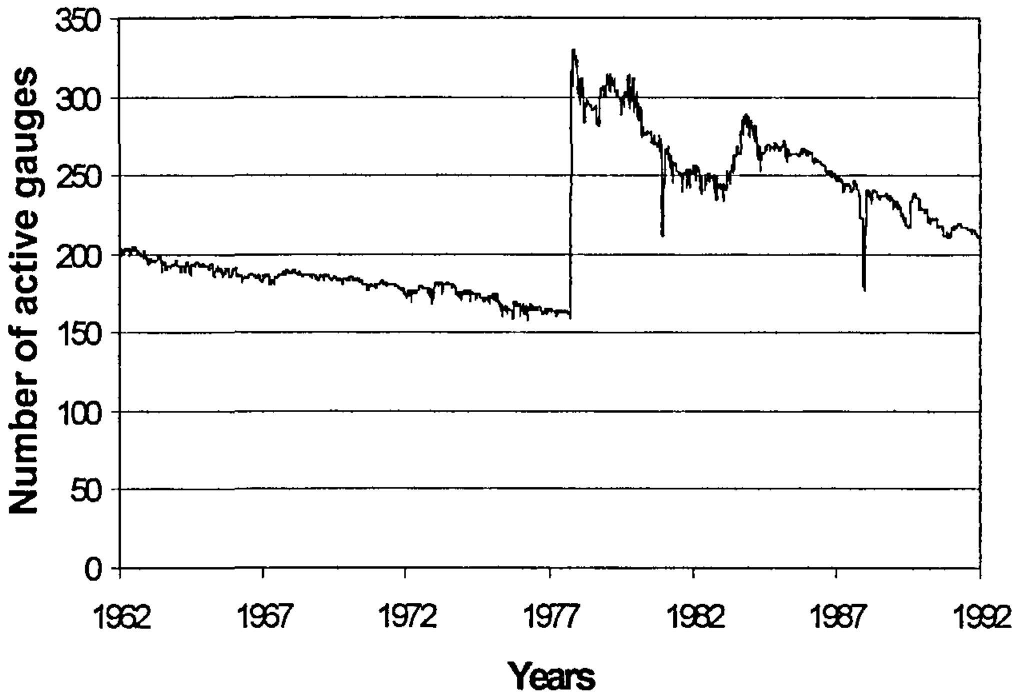


Figure 2.2 The number of active gauges at any time during the study period from the beginning of 1962 to the end of 1991

This number was enough to achieve the classification of daily weather types following the criteria of Table 2.3, which is the contraction of Court's scheme as suggested by Seed (1992).

Table 2.3 Daily weather classification based on Bethlehem raingauge network

dry	< 3% gauges report rain
scattered	> 3% gauges report rain but < 50% gauges report > 5 mm rain
general	> 50% gauges report > 5 mm rain

2.2 MODEL DESCRIPTION

In this chapter, a relatively simple hierarchical model is suggested. It comprises two levels:

- a daily climate model, which in turn drives
- a disaggregation model, describing the daily rainfall on a network of raingauges.

The climate model is a straightforward three-state lag-one Markov chain (not hidden as in the work of Zucchini and Guttorp, 1991) and the raingauge network disaggregation model is based on historical daily rainfalls on the network. The reason that this is appropriate is that, in the study area concerned, it was found that the difference between wet and dry years depends primarily on the number of general raindays that occur in the year. A lag-one climate model was found to be best when compared to three different models of multi-lag Markov chains.

In the following sections we discuss the differences between wet and dry years and describe the differences in the probability distribution of mean areal daily rainfall for sequences of general raindays before we describe and validate the model. In the modelling, a classification scheme based on daily raingauge data is selected, a Markov chain climate model is fitted and finally a number of 30-year sequences of daily rainfall are generated and compared with the historical 30-year record.

2.2.1 The space - time daily rainfall model

Mesoscale climate as measured from a network of raingauges is assumed to be fundamentally a switching type process which enables shifts between distinctive rainday states. This is in contrast to the continuous type of variable which could be modelled by a multivariate autoregressive moving average (ARMA) model. An obvious model for such states is a Markov chain, possibly incorporating a multiple lag structure. This is easily modelled once the states have been defined and sampled. The dependence within the climate model alone is assumed to be sufficient to maintain the general dependence structure of the rainfall process over the study area.

If it is assumed (as has been suggested by some preliminary investigations reported in Section 2.2.1) that the fundamental difference between a wet year and a dry year is the increased frequency of general raindays, and not any characteristic of the raindays themselves (for example higher mean areal rainfall per rainday during a wet year) then it should be possible to reconstruct a rainfall record by sampling at random from a library of historical raindays of the appropriate type.

The first suggested model was constituted as follows:
Model the year in two distinct seasons, wet and dry. (The more refined periodically varying model will be discussed in Chapter 3).

For the summer (wet) period:-

1. Classify a set of daily raingauge data into dry, scattered and general raindays according to the classification scheme of Table 2.3.
2. Use a Markov chain to model the transition between the three rainday states.
3. Generate a new sequence of rainday states from the Markov chain.
4. Build up a new sequence of daily rainfall amounts by selecting a rainday (or set of raindays) at random, with replacement, from an appropriate library of historical raindays according to type and thereafter insert the historical rainfall recorded at each gauge on that day into the synthetic record for each gauge.

For the winter (dry) period:-

Select at random a winter season as a block from the historical record and insert the season into the synthetic data set.

The characteristics of general raindays are assumed to be dependent on the number of consecutive general raindays. This appears to be a reasonable assumption (see Section 2.2.2) since very large scale events cause wide-spread rain over a number of days, and tend to produce more rainfall per day than the one-day events. A scattered rain field does not appear to have a carry-over effect to the following day, and the amounts of rain falling on scattered raindays are therefore assumed to be independent of each other. For general raindays, the resampling is conditional on the generated length of contiguous raindays.

2.2.2 Differences between wet years and dry years

It is of considerable interest to understand the difference between a "wet" and a "dry" summer since fundamental differences have a bearing on the methodology used to generate new sequences of daily rain fields. Garstang and Emmitt (1986) examined the differences between wet and dry years using 31 raingauges in the eastern Transvaal region of South Africa.

Their major findings were that *regardless of the annual rainfall*:

1. The top 12% of the raindays produce 50% of the annual rainfall.
2. The lowest 50% of the raindays produce 8% of the annual rainfall
3. The total number of raindays does not vary significantly between wet and dry years.
4. The major difference lies in the number and sequences of days reporting a mean areal rainfall greater than 7.5 mm.

Court (1979) analysed the contribution of scattered and general raindays to the annual rainfall using 15 years of daily rainfall data taken from 69 raingauges in the Bethlehem area of South Africa. Her conclusions were that dry years had approximately the same total number of raindays as the wet years, with the annual rainfall increasing as a function of the number of general raindays. These findings are corroborated by those of Garstang and Emmitt (1986) emphasizing that the differences in annual rainfall totals are due to the frequency of general raindays.

2.2.3 Dependence of the characteristics of general raindays on the number of consecutive general raindays

In the study area near Bethlehem (which is typical of the inland region of Southern Africa in that it experiences summer rainfall, rarely snow and often hail) Court's (1979) findings referred to in Section 2.2.1 demonstrate the importance of general rainday patterns in relation to the total amount of rainfall experienced in a year. An important first step in the analysis of the data is to find if there is dependence of the average rainfall on the number of contiguous raindays when general rain is experienced.

To achieve this, the classification scheme of Table 2.3 was used to classify a 30-year sequence of daily rainfall over a 200 km-sided square near Nelspruit in Mpumalanga, South Africa. The mean areal rainfall, based on over 200 raingauges in the study area, was calculated for each day for sequences of between one and four days of general rain. Table 2.4 shows the mean, standard deviation and number of days in each of the four data sets. From Table 2.4 it can be seen that the mean areal rainfall increases with duration, while the sample standard deviation does not exhibit a marked trend. From Figure 2.3, it is seen that there is a trend where the increased probability of extreme events increases with the duration of the events. It is evident that lengths of sequences of general rainfall need to be preserved in the re-randomization process.

Table 2.4

Average daily areal rainfall represented by: mean, standard deviation and number of raindays for one to four-day general rainday sequences.

duration (days)	mean (mm)	standard deviation (mm)	number of days
1	11.3	4.1	236
2	12.7	7.1	196
3	12.9	5.0	84
4	14.0	7.1	80

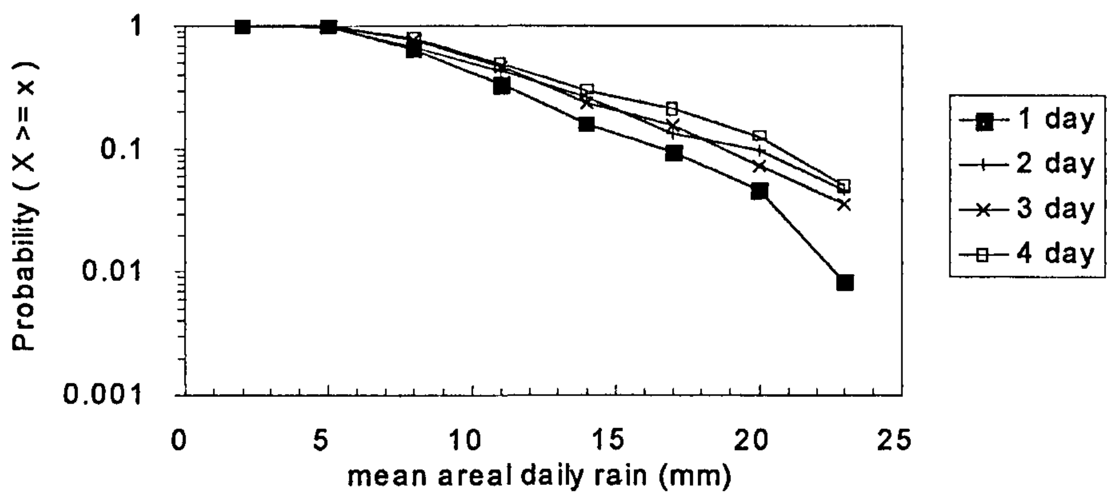


Figure 2.3 Exceedence probability distributions for the mean areal daily rainfall in sequences of consecutive general rainfall

2.2.4 Defining the 3-state Markov chain climate model

The model suggested to describe the rainday climate as classified in Table 2.3 is a 3-state Markov chain. Because of the possibility of there being a significant multiple lag (perhaps of two or three days as experienced in other daily rainfall models e.g. Woolhiser and Pegram (1979), and Zucchini and Adamson (1984)), three multiple lag models were tested using the Akaike Information Criterion (AIC) to choose between models. The models were: (i) the conventional multiple lag specification, (ii) the autoregressive model suggested by Pegram (1980) and (iii) the extension to the latter suggested by Raftery (1985). The lag which minimised the AIC on the 30 years of daily summer climate data for the study area for each of the three models was one day, all of which collapsed to the same simple lag-one Markov chain. The transition probabilities are given in Table 2.5.

Table 2.5

Transition probabilities for the three-state rainday model for the wet season assuming constant parameters

	dry day	scattered rainday	general rainday
dry day followed by a	0.583	0.407	0.010
scattered rainday followed by a	0.103	0.801	0.096
general rainday followed by a	0.007	0.624	0.369

In Table 2.5 it can be seen that dry and scattered raindays are most likely to be followed by their own type, while a general rainday is most likely followed by a scattered rainday. This applies during the wet season assuming constant parameters.

2.2.5 Model validation

Twelve sequences of daily rainfall of 30-years duration were generated using the climate model. Tables 2.6, 2.7 and 2.8 compare some of the characteristics of the historical data with the synthetic data sets verifying the model's performance.

Table 2.6

A comparison between the accumulated mean summer rainfall (mm) for a 30-year measured rainfall sequence and twelve 30-year synthetic rainfall sequences

	25 percentile	median	75 percentile
synthetic	465	519	540
measured	478	540	586

Table 2.7

A comparison between the mean annual number of scattered raindays for a 30-year measured rainfall sequence and twelve 30-year synthetic rainfall sequences.

	mean	standard deviation
synthetic	128	9
measured	130	10

Table 2.8

A comparison between the mean annual number of general raindays for a 30-year measured rainfall sequence and twelve 30-year synthetic rainfall sequences

	mean	standard deviation
synthetic	20	6
measured	20	5

Turning to the figures which follow, Figure 2.4 shows the mean daily rainfall for the measured and synthetic data sets. Figure 2.5 compares the relative frequency of the total summer rainfall for the historical record with the summer rainfall totals in the synthetic rainfall.

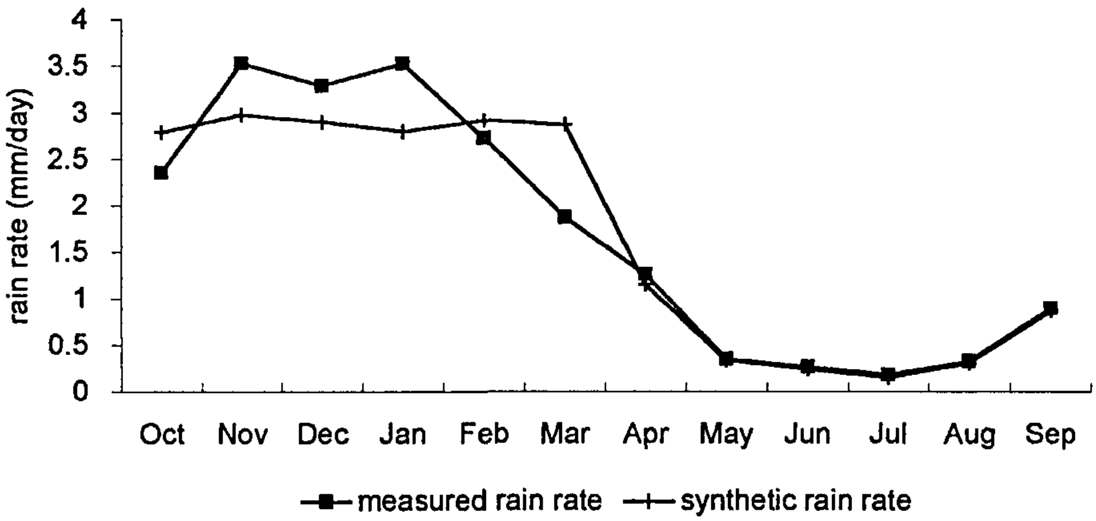


Figure 2.4 Mean monthly rain rate for synthetic and measured rainfall series

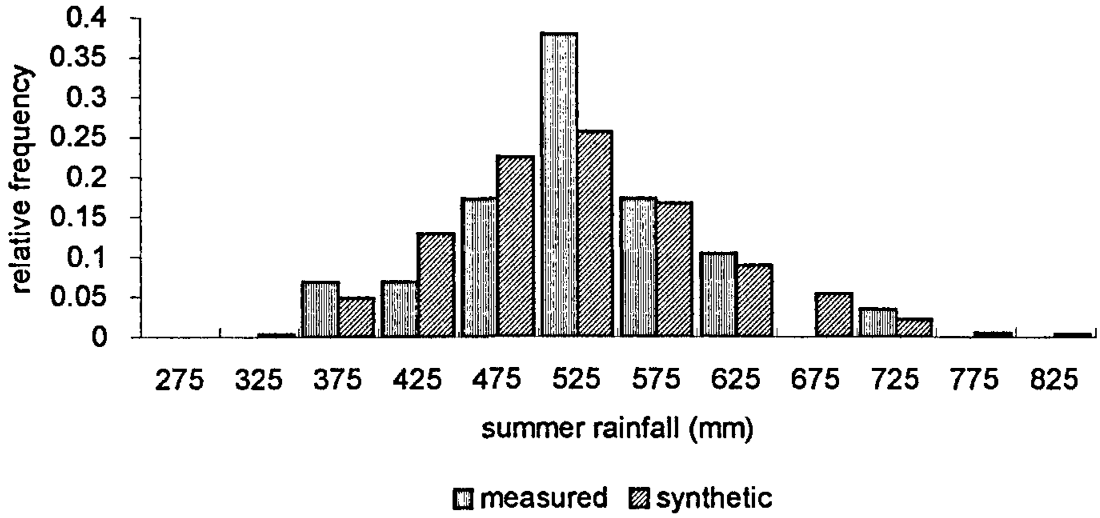


Figure 2.5 Relative frequency of total summer rainfall for the measured and synthetic data

From Table 2.6 and Figure 2.5 it can be seen that the model is able to reproduce the probability distributions of the summer rainfall totals although the inter-seasonal variability is not satisfactorily reproduced. This results from the assumption that the climate was statistically homogeneous during the summer rather than explicitly allowing for seasonal variation which is clearly violated as demonstrated in Figure 2.4.

Figures 2.6 and 2.7 show the relative frequency of the number of scattered and general raindays in the synthetic and historical data sets respectively, but only in the summer months when in this first model we assume constant parameters.

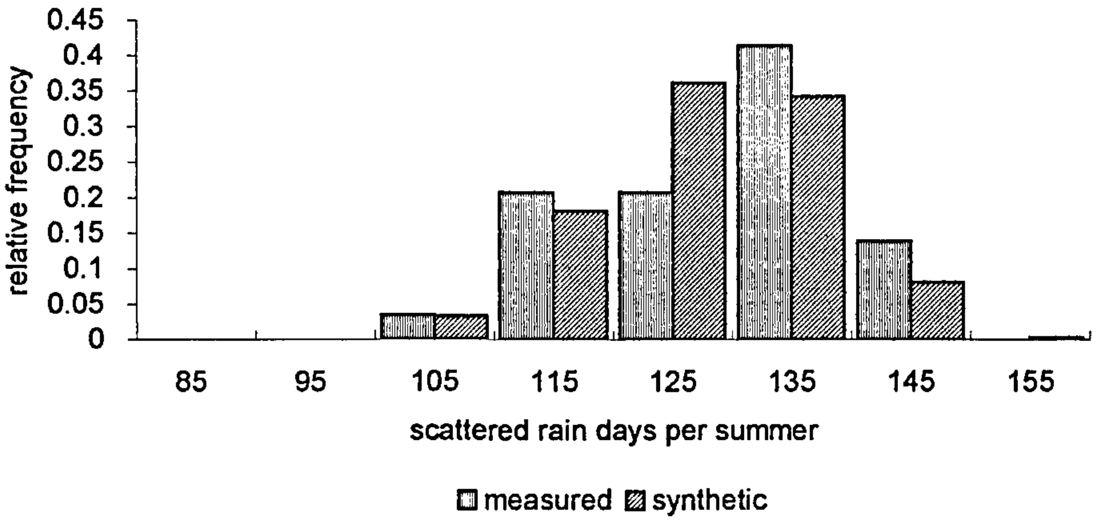


Figure 2.6 Relative frequency of the scattered raindays per summer for the measured and synthetic data

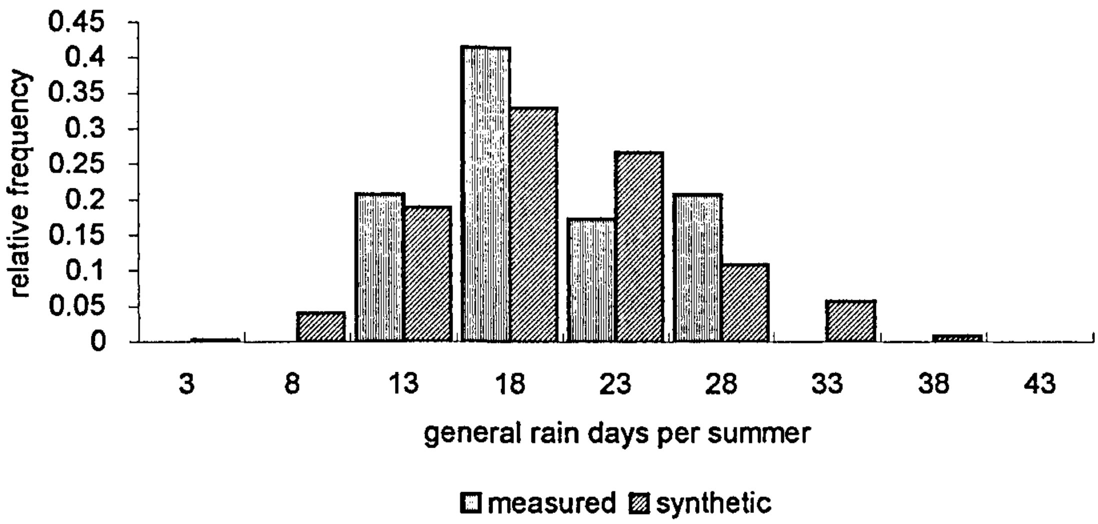


Figure 2.7 Relative frequency of the general raindays per summer for the measured and synthetic data

The frequency distributions for the number of scattered and general raindays per summer are reasonably well preserved as shown in Figures 2.6 and 2.7. The flatter distribution of simulated general raindays as portrayed in Figure 2.7 is probably due to the implicit assumption of homogeneity referred to above.

2.2.6 Summary of simple climate model

A spatial daily rainfall model was suggested which consists of a climate model and a dependant resampling scheme for the rainfall amounts. Its naive simplicity might tend to mask its efficacy as a serious contender in the spatial rainfall modelling stakes, but the results of a simulation experiment to verify and validate the model were particularly encouraging.

It was clear from these preliminary results that some attention needed to be given to describing seasonality of the climate process which might be addressed by using a Fourier series approach as suggested by Woolhiser and Pegram (1979), which will be addressed in Chapter 3.

In summary, the initial model is simple to define for a given target region, is simple to program and operate and yields encouraging results.

Chapter 3

A SEASONALLY VARYING SPACE-TIME MODEL OF DAILY CLIMATE AND RAINFALL

3.0 INTRODUCTION

In this Chapter, a model is described which is the direct out-growth of the model introduced in Chapter 2. There, a 3-state Markov chain was used to describe the transitions between the three climate types: dry, scattered rain and general rain. The parameters were kept constant during the wet period, and this resulted in some shortcomings when observed and simulated sequences were compared.

The obvious extension to the climate model is done here, in that the 6 parameters defining the Markov chain transition probability matrix are allowed to vary periodically over the calendar year on a daily basis. This makes for a great improvement to the model performance as anticipated. The method of achieving this will be outlined in some detail.

Rather surprisingly, the initial fit of the model in the wet months of January and February were not good - the mean monthly rainfall in those months was under estimated. We found that the fault lay in assuming that the amount of rain that falls on a scattered rainday is constant throughout the year. It is not. Once this was rectified, the model produced very good validation statistics which are presented in graphical form.

A summary of the work contained in this Chapter was presented as a paper at the European Geophysical Society meeting at Grenoble in April 1994.

3.1 THE DAILY RAINFALL DATA

30 years of daily rainfall at some of the 255 gauges whose locations are shown in Figure 2.1 were used in this analysis, in the same way as they were in Chapter 2. The gauges were selected which had 10 or more years of data during the period. The number of gauges recording rain on a particular day as a percentage of those working on that day were used to define the climate, or rainday type according to the classification adopted in Table 2.3.

To compute the average rainfall over the area, the simple average rainfall of the working raingauges on a particular day was computed. This was typically from 100+ raingauges and was felt to be as good as any method available. The difference from using Thiessen polygons, isohyets computed by surface fitting or by straight averaging is usually fairly close (Chow et al., 1988) especially when a large number of gauges is involved. The drawback from using a sample whose membership changes from time to time, as raingauges come and go, is that there is a likelihood of lack of homogeneity in the sample with time. A more ambitious program might use cross-correlation to infill and/or extend some of the sequences, however the methodology has yet to be developed which will enable the simultaneous patching of more than 5 to 10 gauges on a daily basis (Pegram, 1997).

Our opinion is that there is high enough cross-correlation between the surviving gauges to ensure that spatial homogeneity, when averaged, will be better maintained than any changes inducing temporal inhomogeneity that are due to climate change or the like. This strategy was used for extracting spatial averages of rainfall over the study area. Evidently, it is a small additional step to select the actual rainfall record for a day selected at random and to reconstruct the spatial distribution of that day from the raingauge readings. We feel that although this is reasonable, it does not go far enough to describing the small-scale spatial variability of rainfall, so the matter will be deferred until the treatment in Chapter 4.

3.2 MANAGING THE DATA

To fit a 3-state Markov chain model of daily climate with 6 free transition probabilities, each of which may vary periodically, involves a large amount of computation if it is done on a daily basis. We examined models involving up to 3 harmonics in the Finite Fourier series. Each harmonic requires 5 parameters, a constant term and two coefficients, thus there would be 30 or more parameters defining the full model. The transitions were assumed to change on a daily basis over the 30 years so the computation becomes lengthy when many iterations are needed to find the optimal set of 30 parameters.

In the event, the adopted strategy was to search for the optimum first for the model with 2 harmonics and then for the model with 3 harmonics and after each optimum was found, to perturb the parameters randomly in the vicinity of the achieved optimum and start the search again for a better optimum set. On a 80 MHz 486 PC, the program took about 3000 iterations to converge in each search and about 7 hours to complete the calculation. Some coding improvements could speed things up, but it was not the intention to go into production. It is true that the statistical estimation of each of the transition probabilities for a time - homogeneous Markov chain can be performed independently of the rest, but this is not necessarily true of a seasonally varying Markov chain. Hence the parameters are estimated simultaneously by minimizing an objective function chosen as the Akaike Information Criterion (AIC), which is defined as the sum of the negative of the log likelihood and a penalty function involving the number of parameters. The function minimization algorithm chosen was the simplex algorithm of Nelder and Mead (1965).

To make for pleasing presentations of the transition probabilities without sacrificing much detail, each year was defined in 52 weeks and the parameters were assumed constant for the week. Thus the data were analysed week-by-week for the 30 years and the number of transitions of each type in each week were recorded and it is these which are compared to the fitted Fourier series models presented in Section 3.9

3.3 THE PERIODICALLY VARYING DAILY CLIMATE MODEL

The climate states were given symbols:

- 0 for a dry day
- 1 for a scattered rainday
- 2 for a general rainday.

The transition probabilities then become: $p_{ij}(t)$ = probability [rainday on day i is of type j given that the climate on the previous day was of type i], $(i, j = 0, 1, 2)$. Assembled in a transition probability matrix, the relationships are:

$$P(t) = \begin{bmatrix} p_{00}(t) & p_{01}(t) & 1-p_{00}(t)-p_{01}(t) \\ p_{10}(t) & p_{11}(t) & 1-p_{10}(t)-p_{11}(t) \\ p_{20}(t) & p_{21}(t) & 1-p_{20}(t)-p_{21}(t) \end{bmatrix}$$

and because the total of each row is constrained to unity, only six transition probabilities suffice to define the Markov chain.

To define the periodically varying transition probabilities in terms of a finite Fourier series we adopt the following notation:

$$p_{ij}(t) = a_{ij0} + \sum_{k=1}^{m} a_{ijk} \sin\left[\frac{2\pi\, m\, t}{L} + b_{ijk}\right] \tag{3.1}$$

for $i, j = 0, 1, 2$, $m = 2$ or 3 or more (harmonics) and L is the length of the period of the first harmonic. L=365.25 for fitting and generating daily probability transitions.

3.4 Fitting The 3-State Markov Chain With Periodically Varying Parameters

Zucchini and Adamson (1984) used a clever idea in fitting the periodically varying parameters of a Markov chain for describing alternating wet and dry days. We use the same idea here.

Transition probabilities (like all probabilities) are constrained to be in the interval (0, 1). When fitting an equation like (3.1) to a time-varying parameter, using an unconstrained function minimization routine, such as the simplex algorithm of Nelder and Mead (1965), it is quite likely that in exploring the parameter space the trial parameter may wander outside the interval (0, 1). The workaround suggested by Zucchini and Adamson, was to transform the transition probabilities into logits of probability and to fit Fourier series to the transformed variables. The transform is simple: if $a < x < b$ then $y = \ln[(x-a)/(b-x)]$ lies on the real line i.e., $-\infty < y < \infty$. The reverse transform is

$$x = [a + b\exp(y)]/[1 + \exp(y)] \tag{3.2a}$$

Thus if $a = 0$ and $b = 1$, the logit transform pair is:

$$y = \text{logit}(x) = \ln[x/(1 - x)]$$
$$x = \exp(y)/[1 + \exp(y)] \tag{3.2b}$$

The Fourier series of the probabilities fitted directly (if we could so without violating the (0, 1) bounds) and by fitting the logits of the probabilities will not result in the same shaped periodically varying probability estimates, but they are likely to be quite similar.

To be more precise, instead of working with equation (3.1), a similar equation describes the logits of the transition probabilities, whose amplitude coefficients A_{ijk} are unconstrained, but whose phase coefficients are limited to the range $(0, 2\pi)$ for neatness:

$$\text{logit}[p_{ij}(t)] = l_{ij}(t) = A_{ij0} + \sum_{k=1}^{m} A_{ijk} \sin\left[\frac{2\pi\, m\, t}{L} + B_{ijk}\right] \tag{3.3}$$

where $L = 365.25$ and $m = 0, 1, 2$.

The operation of the function minimizing algorithm is to:

- alter the A_{ijk} and B_{ijk} by small amounts, then for all $t = 1, 2, \ldots, 10956$
- determine i and j, the current and following rainday types
- calculate the appropriate $l_{ij}(t)$ only for the current i and $j = 0, 1$, using (3.3)
- reverse the transform using (3.2b) to calculate the transition probabilities for the calendar week (t) as $p_{i0}(t)$ in the range (0, 1), use (3.2b) to calculate $p_{i1}(t)$ in the range (0, $p_{i0}(t)$) and then obtain $p_{i2}(t) = 1 - p_{i0}(t) - p_{i1}(t)$
- form the sum of $-\log[p_{ij}(t)]$ over the observed transitions to define the negative of the loglikelihood which is the variable part of the objective function being minimized. This is equivalent to computing the logarithm of the likelihood function which we wish to maximize which is:

$$\text{loglikelihood} = \sum_{t=1}^{10956} \sum_{i=0}^{2} \sum_{j=0}^{2} n_{ij}(t) \ln[p_{ij}(t)] \tag{3.4}$$

  in which n_{ij} is an indicator variable taking on the value 1 if the transition from day t to the next is actually ij, and 0 otherwise.

Thus the minimization of the function takes place in probability space, not in logit space. The effect of using the logits is to flatten the sine and cosine curves near the boundaries of the interval (0,1); there is less distortion as the midpoint of the interval (0.5) is approached. Nevertheless, good fits can be achieved by this means using any conventional, computationally efficient, unconstrained non-linear function minimization/maximization routine.

3.4.1 The Model Fitting Criterion and the Number of Harmonics.

A given model (the number of harmonics describing each transition probability) will be fitted by maximizing the likelihood function, or equivalently minimizing its negative logarithm. This is the objective function used by the simplex algorithm of Nelder and Mead (1965).

When a harmonic is added to the model, one is likely to experience better fit, because more parameters allow more flexibility since there are more degrees of freedom. To counteract this effect, the objective function should embody a penalty term dependent on the number of parameters. An appropriate objective function in this context is the AIC which is

$$AIC = -\text{loglikelihood}|\underline{\theta} + 2p \qquad (3.5)$$

where the likelihood depends on the parameter set $\underline{\theta}$ and p is the number of parameters in the model. A better fit using more parameters will increase the likelihood, reducing the AIC, but at the cost of an increase of p. One chooses the model between parameter choices which minimizes the AIC overall.

3.6 THE VARIATION OF THE DAILY RAINFALL PER RAINDAY TYPE OVER THE YEAR.

Initial fits of the models came up with some disappointments, among which was that the January and February rainfall totals were underestimated by the simulations using the model when compared with the measured data. This lead us to re-examine the data. One of us (Seed) remembered that January storms were larger than the rest when looking at Bethlehem radar data. Extracting the data by rainday type, the mean and standard deviation of the average depth of rain over the study area per scattered rainday by month are shown in Table 3.1.

Table 3.1

Variation of Mean and Standard deviation of depth of rain per scattered rainday over the year (mm)

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	2.76	2.55	2.16	1.88	1.19	1.15	1.13	1.24	1.42	1.88	2.19	2.32
Stddev	2.36	2.45	2.10	2.05	1.69	1.58	1.50	1.66	1.81	2.05	2.16	2.08

It is clear that January experiences average rainfall on the gauge network which is wetter than the rest and that July's figure is less than half of it, in fact it is 41% of January's average. This becomes an important difference when it is noted that the average number of scattered raindays in January is 25 which has the consequence that, to ignore the variation, penalizes the modelling of the January rainfall by about 20mm. The model thus needs to take this variation into account - it is not acceptable to ignore the systematic change in the mean and variance of the average amount of rain on scattered raindays throughout the year. The effect of the variation of general rainday intensities is not as marked. Some systematic variation is evident in the figures reported in Table 3.2.

Table 3.2

Variation of Mean and Standard deviation of depth of rain per general rainday over the year (mm)

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	14.3	13.4	13.5	12.3	11.5	10.1	9.1	17.1	15.3	14.1	13.3	12.9
Stddev	6.5	5.1	8.1	5.5	5.7	2.4	3.8	7.3	7.3	6.5	6.2	4.5

The most marked is the months of June to August when the range is 8mm. However, coupled with a mean number of raindays of 0.3, 0.3 and 0.6 in each of the 3 months with deviations from the mean (13.07) of -3, -4 and +4, the error in the monthly totals incurred by ignoring this variability is -0.9, 0.8 and 2.4mm in June to August respectively.

The effect on January and February is to reduce the estimated monthly average by 3.7 and 0.8 mm respectively. This does not show up in the validation tests to be described in Section 3.9 so the amount of error is not great and to attempt to remove it by allowing for within-year variation of the general rainday intensity would mean sampling by month (or season) from a very small pool of general rainday data. It therefore does not seem worth the effort to get a marginal increase in accuracy. In the event, variation of general raindays within the year was ignored for this model but could be incorporated in future if desired; but it should be remembered that the general rainday set is divided by runs of days into 1, 2, 3, 4 and 5 day bins which will complicate things considerably.

3.7 THE ADOPTED SPACE-TIME MODEL.

The two components of the model are:

- the climate generator in the guise of a 3-state Markov chain with periodically varying parameters and
- the bins of rainday data i.e. collections of historical dates on which the various types of rain occurred.

An economy of 8 (or 12) parameters for the two (or 3) harmonic model was achieved by estimating the general-to-anything transition probabilities as constant. This was because there was no detectable systematic variation over the year in those probabilities when fitted by Fourier series. The result was that the 3-state climate model is specified by 22 parameters with an AIC of 5886 for 2 harmonics or 30 parameters with an AIC of 5890 for 3 harmonics. The Fourier coefficients for the logits of the probabilities for the 2 models follow:

Table 3.3

The 2-harmonic model - Fourier coefficients of the logits of the transition probabilities - in all, 22 parameters - AIC = 5886.13 in 7445 iterations in 2 hours.

	a_0	a_1	a_2	b_1	b_2
p_{00}	1.1927	1.3973	0.1812	4.6095	1.1184
p_{01}	5.1275	-1.1123	-0.3994	5.8646	2.6946
p_{10}	-1.4910	-1.3142	-0.1284	1.4545	0.3569
p_{11}	2.2804	-0.0367	0.0842	1.8098	0.1960
p_{20}	-4.6553				
p_{21}	0.7892				

Table 3.4

The 3-harmonic model - Fourier coefficients of the logits of the transition probabilities - in all, 30 parameters - AIC = 5895.72 in 12 236 iterations in 5 hours

	a_0	a_1	a_2	a_3	b_1	b_2	b_3
p_{00}	1.1884	1.3767	0.1696	0.0313	4.6090	1.3795	2.2207
p_{01}	7.4091	-0.9662	1.5517	1.7585	-0.8043	-0.1001	-0.6870
p_{10}	-1.5016	-1.2947	0.1130	-0.0711	1.4398	3.7168	5.4468
p_{11}	2.3226	-0.0585	0.1569	0.1352	2.1229	0.4796	2.4165
p_{20}	-4.6272						
p_{21}	0.7998						

The 2-harmonic model seems (by the AIC) to be more appropriate at this stage. Incidentally, for exposition purposes, the probabilities p_{20} calculated by the two variations of the model are different, but quite close, indicating that estimating the 22 to 30 parameters of the transition probabilities simultaneously is likely to lead to imprecise estimates; it is arguable that to initially estimate them independently will give a quicker and more precise convergence. The values of these constant probabilities is easily computed from equation (3.2b) as 0.00942 and 0.00969 for the 2- and 3-harmonic models respectively. Note that to compute the probabilities p_{21}, use must be made of equation (3.2a) with a and b equal to 0 and $1 - p_{20}$ respectively.

Turning to the rainfall amounts on the two types of rainday, the scattered rainday data were sorted and stored by 12 calendar months while the general rainday data were sorted and stored by run lengths of 1 to 5 days, because the average intensity of general rain per day increases with the duration of the event as shown in Table 2.4.

In operation then, the model works like this:

- **starting**: the largest proportion of climate type is the scattered rainday so we start with that

- **continuing**: given a current rainday type (i = 0, 1 or 2), compute the 2 transition probabilities $p_{i0}(t)$ and $p_{i1}(t)$ for the given day (t) to the next and hence compute $p_{i2}(t) = 1 - p_{i0}(t) - p_{i1}(t)$. Generate a uniformly distributed random number and use it to select a rainday type for the next day.

  If the type is dry record it as 0 rainfall at all gauges.

  If the type is scattered, select (from the collection of scattered raindays in the appropriate month) an historical date at random. Look up the set of data for that date (which in the current model is the mean value of rainfall) and record it.

  If the type is general note how many general raindays are in the current sequence of general raindays and resample the climate type for the day after next. If the current rainday is general and the next one is other than general, select a general rainday sequence from the sets of runs of 1 to 5 days and record that sequence as it occurred historically.

The model will thus produce a sequence of daily averages of rainfall based on the historical record. It is a trivial extension to recover the actual rain that fell at all the active raingauges on that date and obtain the spatial variation true to the topography and distribution of the gauges.

3.8 VERIFYING THE MODEL - REPRODUCING THE TRANSITION PROBABILITIES.

The objective of verification tests is to determine whether what went in to a model is what comes out. For the climate component of the model, this means resampling the transition probabilities from many generated sequences of raindays and comparing them to the historical.

This was not done for two reasons:

- to extract the transition probabilities by function minimization over the historical sequence of 30 years was very laborious and took about 7 hours on a 486 computer as shown in the heading to Tables 3.4 and 3.5. To repeat the calculations for a large number of artificially generated sequences of the same length of the historical sequence presented a daunting computational task. There are possibilities of refining the code to increase the speed of computation but it was not felt it was worth it at this stage, since this is a feasibility study.

- The resampled probabilities will, in all likelihood, behave as they should which is to look like the ones put into the model, since they will come out without bias. However having said this, it should be noted that the rigorous validation tests which follow in Section 3.9 have elements of verification in them. In particular Section 3.9.1 compares the means of the number of the various climate days on a monthly rather than a weekly basis as was employed in the model fitting. True, these are not transition probabilities but are marginal probabilities which come from the transitions, however they give a useful, if indirect verification check.

One result is that in this section we display only the diagrams comparing the sampled (by week) transition probabilities and the fitted, smooth, 2 and 3 harmonic finite Fourier series of the model. There are 9 transition probabilities in all, defined by 6 independent ones as described in Section 3.3. These nine transition probabilities are displayed in Figures 3.1 to 3.9 which follow.

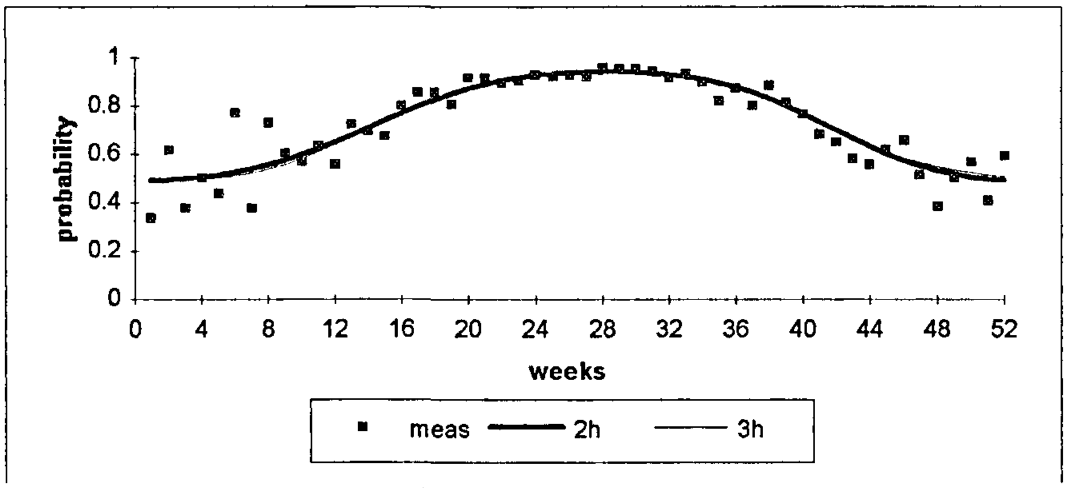


Figure 3.1 Probability of a dry day following a dry day

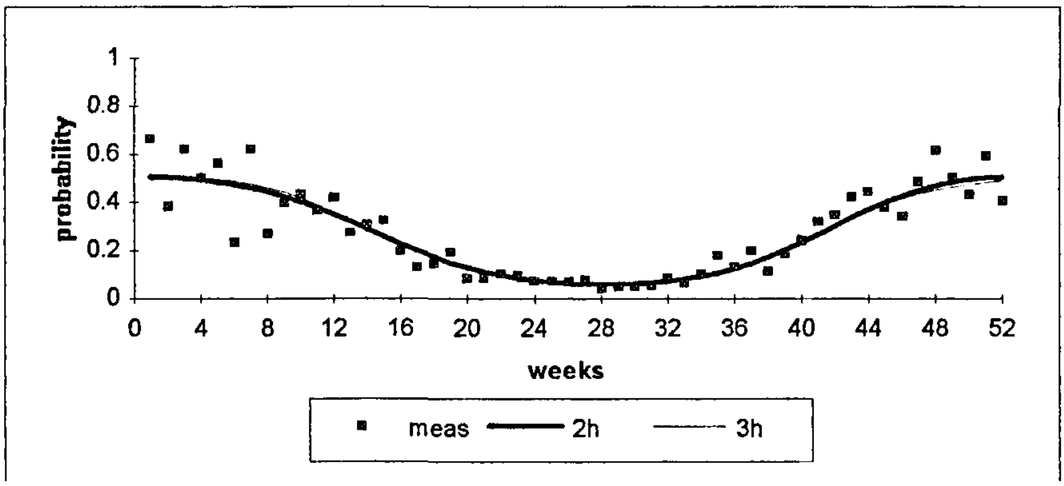


Figure 3.2 Probability of a scattered rainday following a dry day.

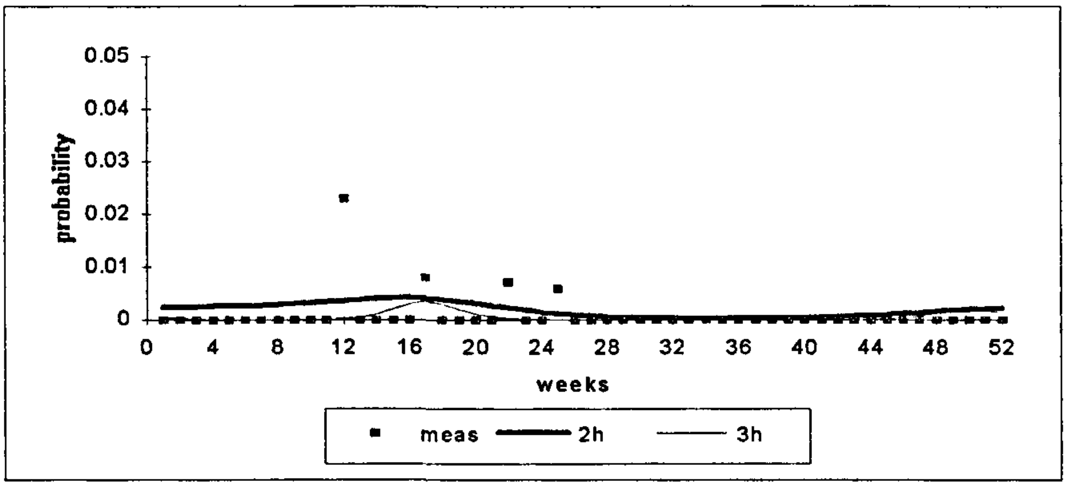


Figure 3.3 Probability of a general rainday following a dry day - note the change in vertical scale.

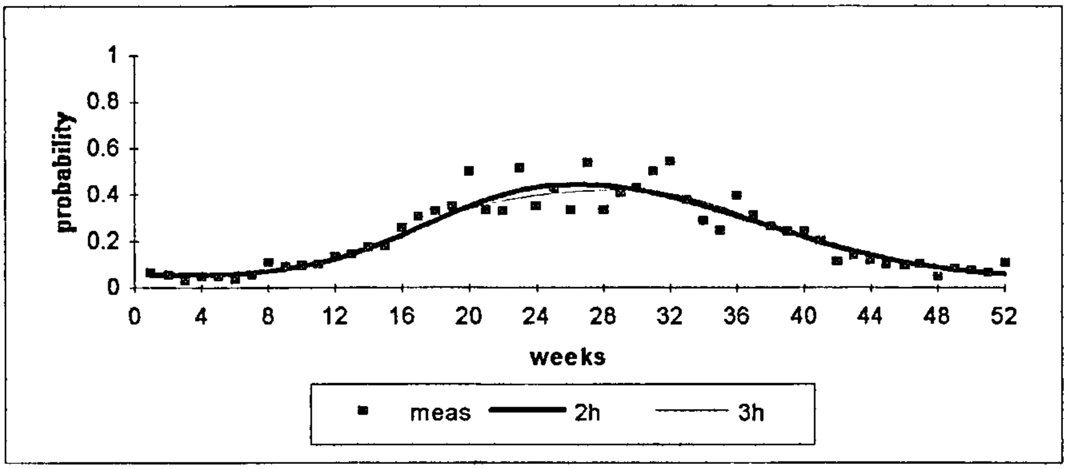


Figure 3.4 Probability of a dry day following a scattered rainday

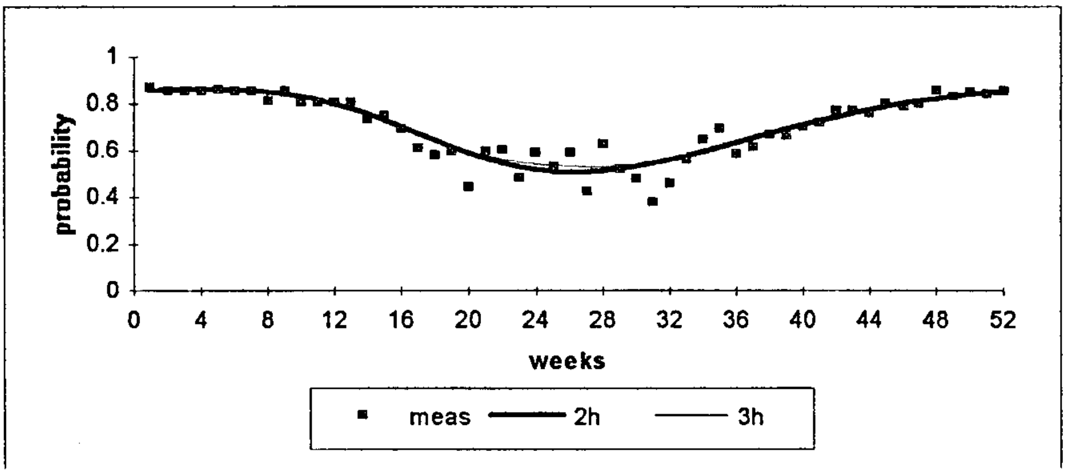


Figure 3.5 Probability of a scattered rainday following a scattered rainday

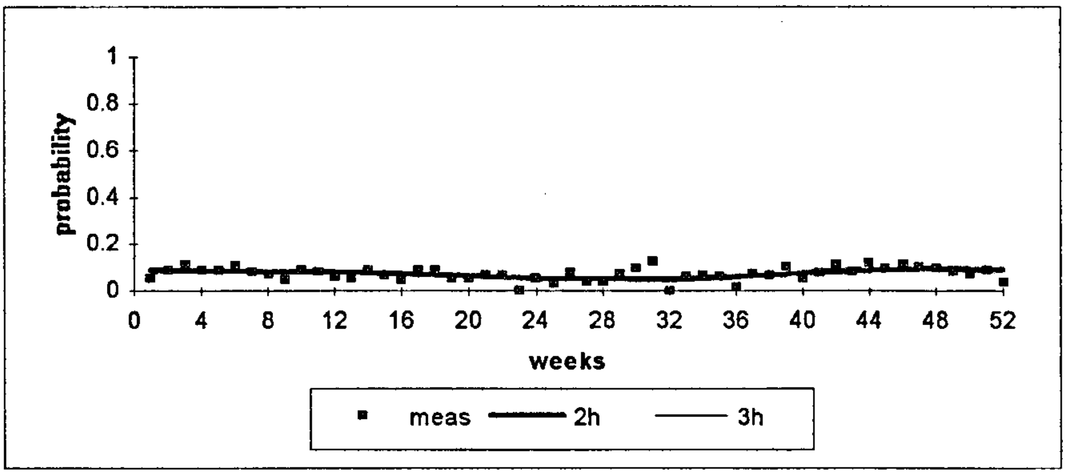


Figure 3.6 Probability of a general rainday following a scattered rainday

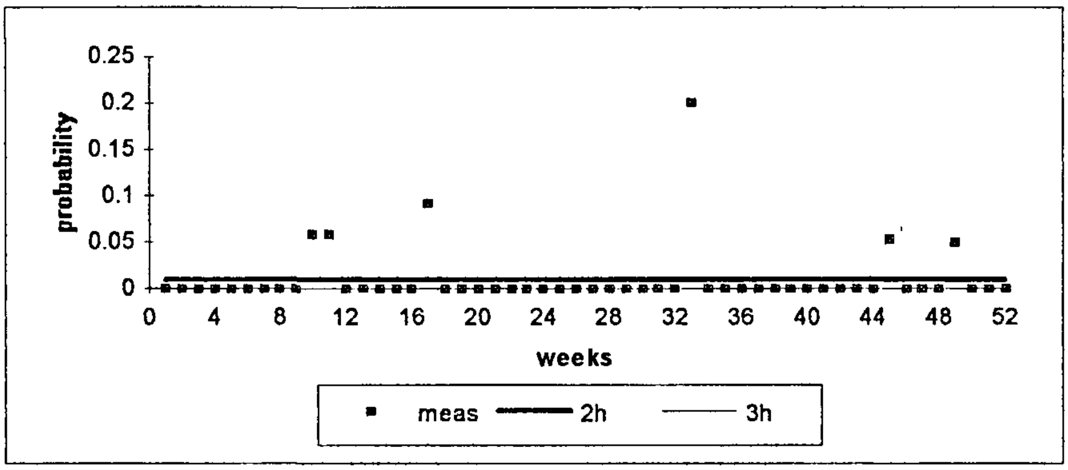


Figure 3.7 Probability of a dry day following a general rainday - note the change in the vertical scale

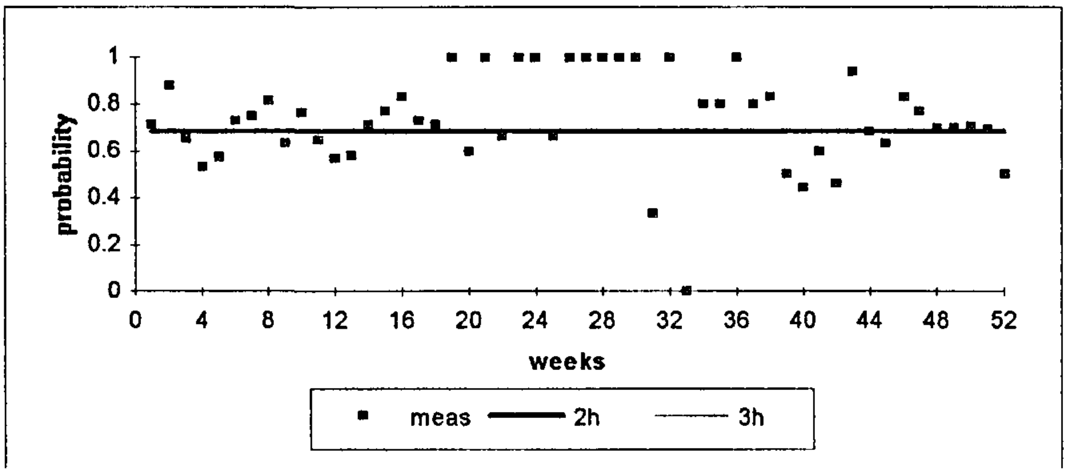


Figure 3.8 Probability of a scattered rainday following a general rainday

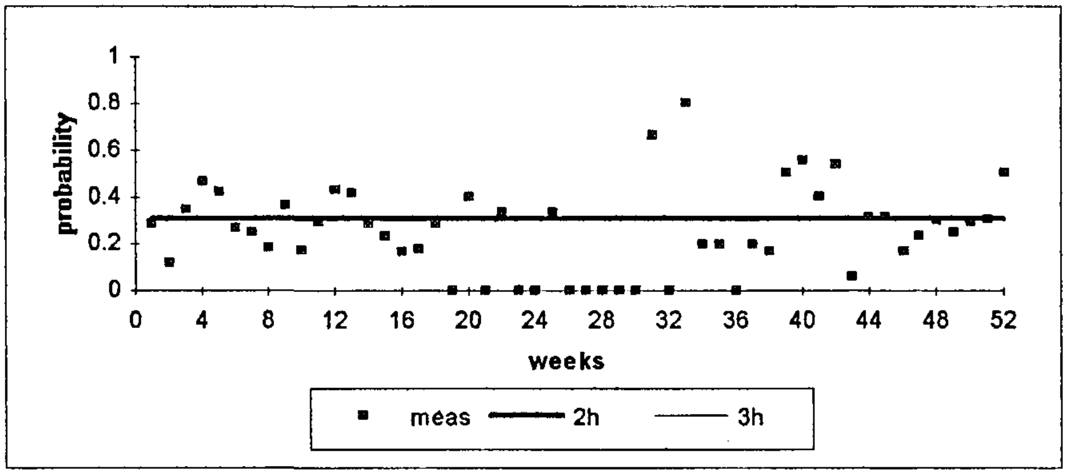


Figure 3.9 Probability of a general rainday following a general rainday

Although it was included merely for comparison purposes, the 3rd harmonic affects only the scattered-to-anything probabilities to any extent, as shown in Figures 3.4 to 3.6, where the maximum deviation from the 2-harmonic fit is of the order of 0.02 during the dry months. These are the months in which there is more scatter (and therefore less precision in definition) in the measured transition probabilities.

Note the flatness of the curves in Figures 3.1 and 3.2 near the boundaries. This is a combination of having 2 harmonics and working in logit space. There is no doubt of the goodness of fit in the constrained regions.

It is evident from Figures 3.7, 3.8 and 3.9 that a reasonable choice of parameterization of the general-to-anything transition probabilities was to keep them constant throughout the year. The probability of a dry day following a general rainday is zero for all but six weeks in the year, so that Figures 3.8 and 3.9 are all but complements of each other.

Again, the marginal probability of getting a general rainday on any day around the 30th week (July/August) is very low indeed. This will be confirmed in Figures 3.52 and 3.53 (to be discussed in Section 3.9.3) where it is seen that the total number of general raindays observed in August in 30 years (a total of 930 days) is only 7. This converts to a probability of observing a general rainday on any given day in August equal to 1/132 = 0.0075. (This checks with Figure 3.15 to be discussed in Section 3.9.1).

Figures 3.7 and 3.8 show that it is very rare for a dry day to follow a general rainday and in fact it is almost sure that a general rainday sequence ends with a scattered rainday, indicating that the general rainday system slowly decays over a couple of days in the wetter months, which makes sense meteorologically.

Overall the transition probabilities are well modelled by the 2-harmonic Fourier series model which, as was shown in Section 3.7, has a lower AIC than the 3-harmonic model, so statistically explains the data "better".

3.9 VALIDATION OF THE MODEL BY VARIOUS STATISTICAL CHECKS.

In contrast to verification, validation is a process of examining the effectiveness of a model in capturing those aspects of the behaviour of the process being modelled which were not explicitly written into the model specification and estimation procedure. This section contains 4 sub-sections in which various validation experiments are reported and discussed. They contain comparisons between the statistics of the 30-year historical sequence and 100 simulated 30-year sequences and comprise the following:

- the mean and standard deviation of the 3 rainday types by month, modelled by the 2- and 3-harmonic Fourier series models.
- the run-lengths of the 3 rainday types by month modelled by only the 2-harmonic model
- the means and standard deviations of the monthly rainfall totals using the 2- and 3-harmonic models
- the recurrence intervals of annual maxima of 1, 2, 3 and 7 day rainfall totals using the 2- and 3-harmonic Fourier series models.

It will been seen from these validation statistical comparisons that the 2- and 3-harmonic models produce very similar results. Those comparisons where they are both used are for exposition purposes. The hope was that the extra harmonic would improve the behaviour of the model where the fit was not as good as one would like. The marginal improvement obtained by using the extra harmonic does not seem to justify the more complicated model (an increase of 8 parameters with an increase, not a reduction, in AIC).

It will be seen that, by and large, the model works well at reproducing those things that are of interest. More refinement is possible but that will be deferred to the future. The objective here is to demonstrate that the basic strategy is good, feasible and worth pursuing.

3.9.1 Means and Standard Deviations of the Number of Raindays per Month of the 3 Climate Types Using 2- and 3-Harmonic Fourier Series Models.

The comparisons between the historical and the simulated results are presented in Figures 3.10 to 3.21. It will be seen that there is very little difference between the results of the simulations produced by the 2- and 3-harmonic models.

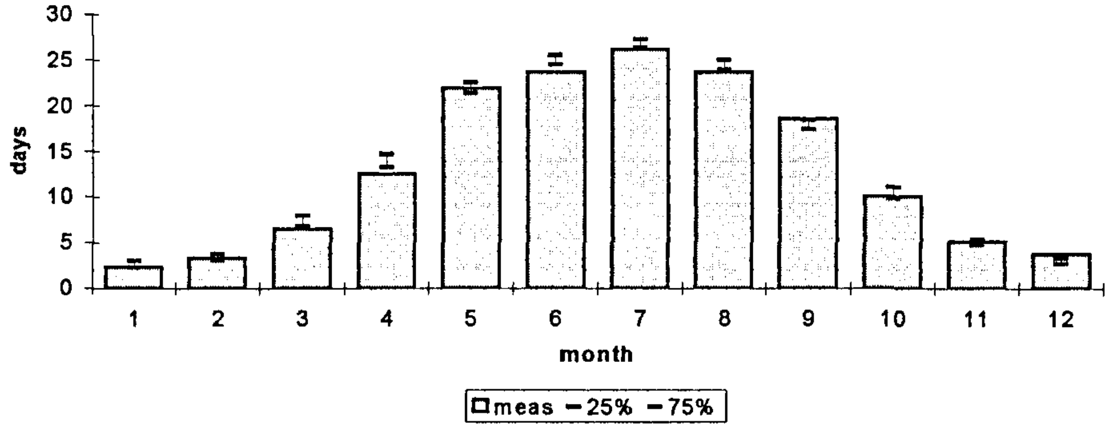


Figure 3.10. Mean number of dry days per month, using 100 30-year simulations of the 2-harmonic model.

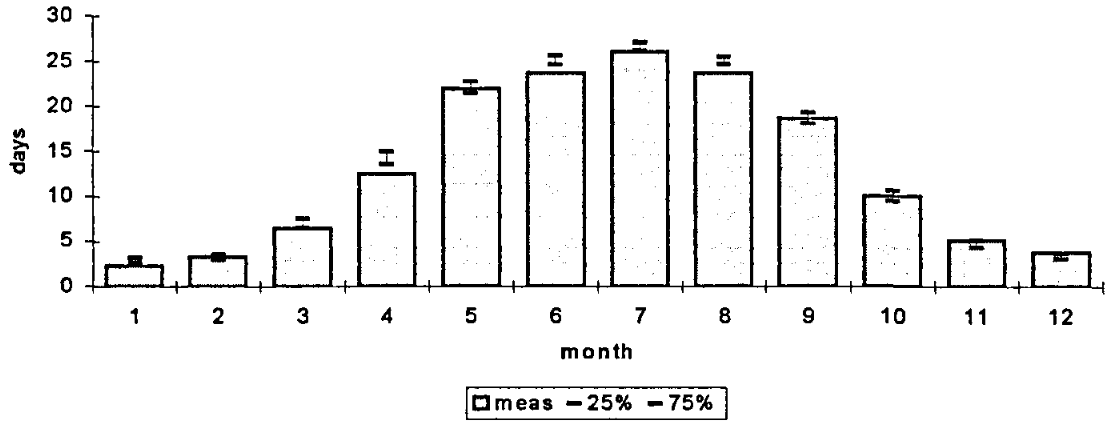


Figure 3.11. Mean number of dry days per month, using 100 30-year simulations of the 3-harmonic model.

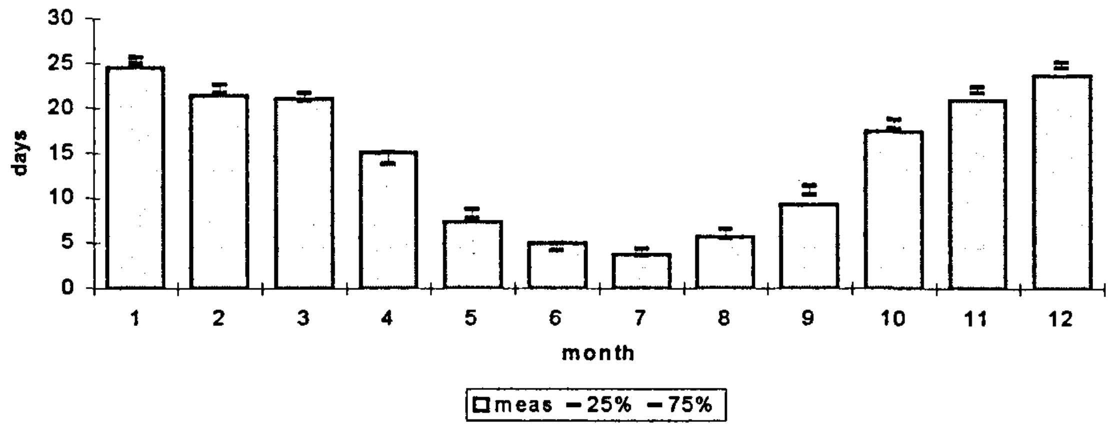


Figure 3.12. Mean number of scattered rain days per month, using 100 30-year simulations of the 2-harmonic model.

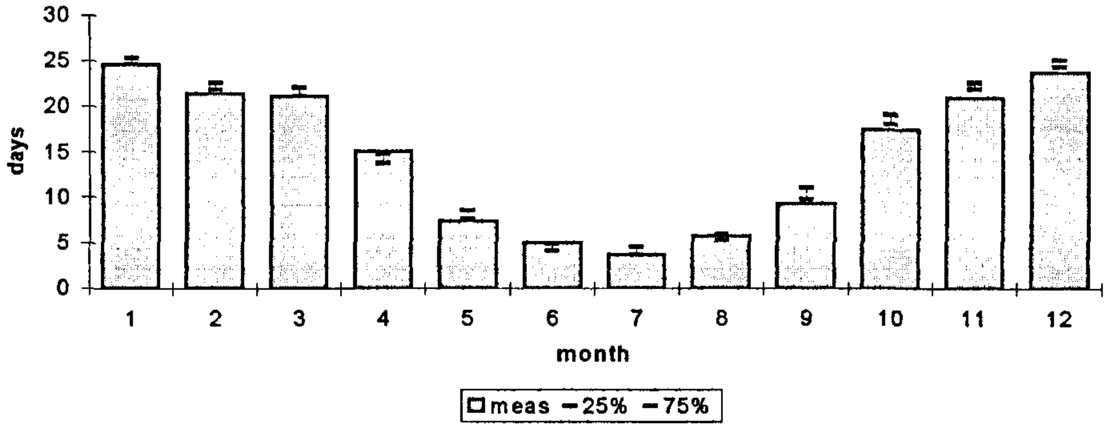


Figure 3.13. Mean number of scattered rain days per month, using 100 30-year simulations of the 3-harmonic model.

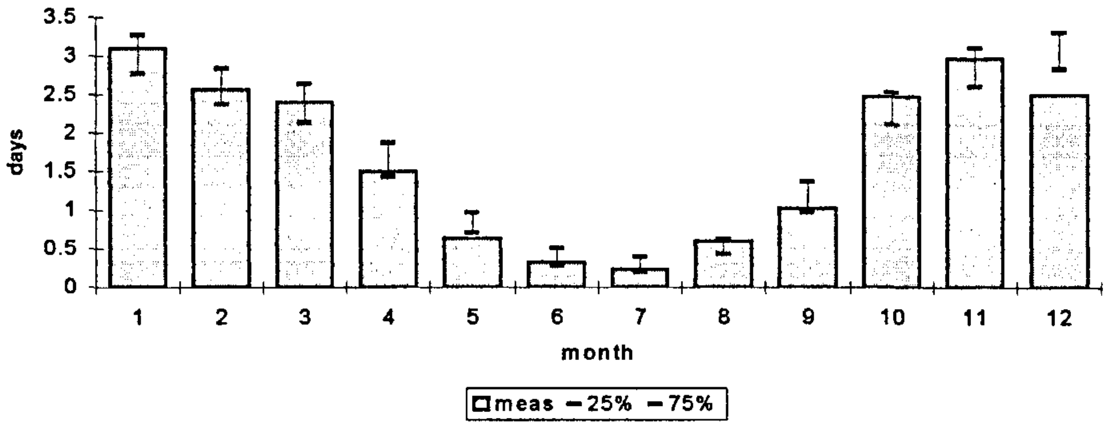


Figure 3.14. Mean number of general rain days per month, using 100 30-year simulations of the 2-harmonic model.

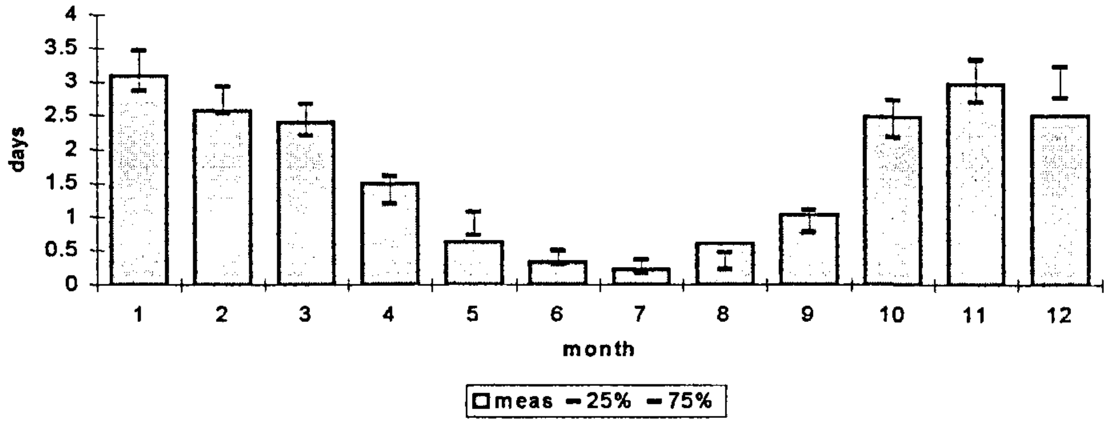


Figure 3.15. Mean number of general rain days per month, using 100 30-year simulations of the 3-harmonic model.

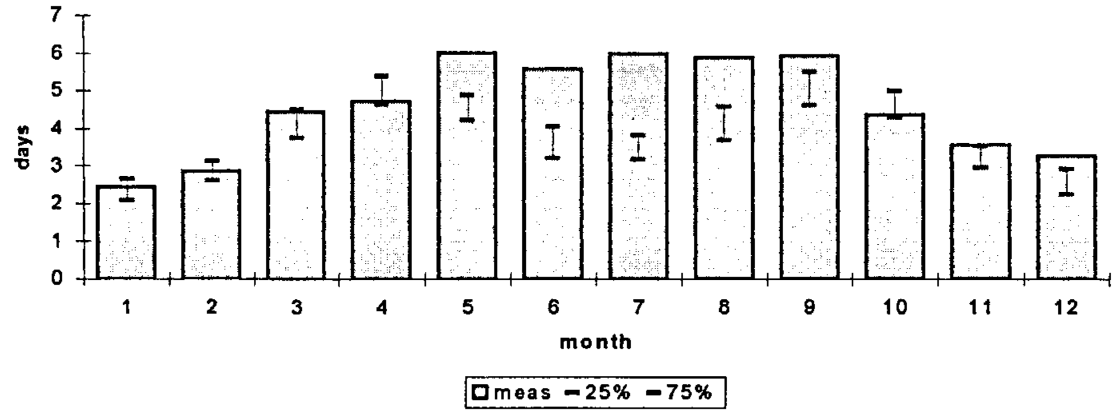


Figure 3.16. Standard deviation of number of dry days per month, using 100 30-year simulations of the 2-harmonic model.

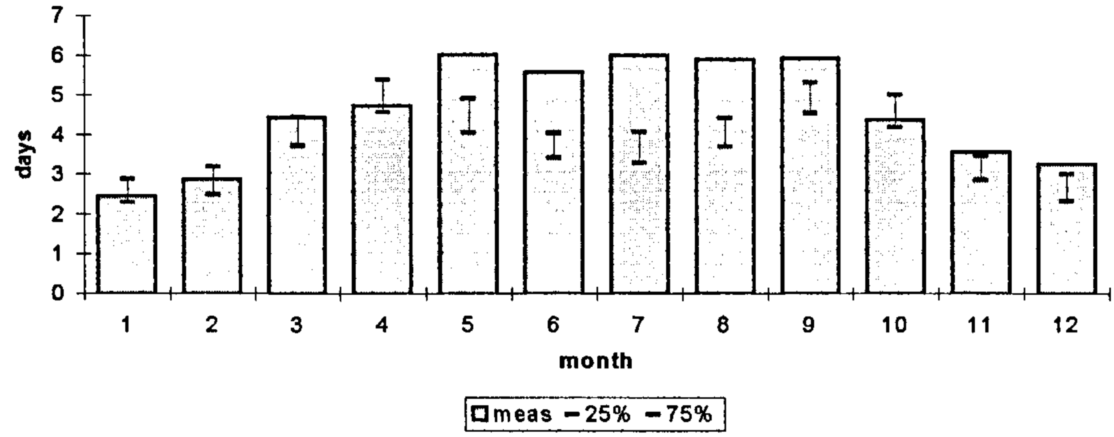


Figure 3.17. Standard deviation of number of dry days per month, using 100 30-year simulations of the 3-harmonic model.

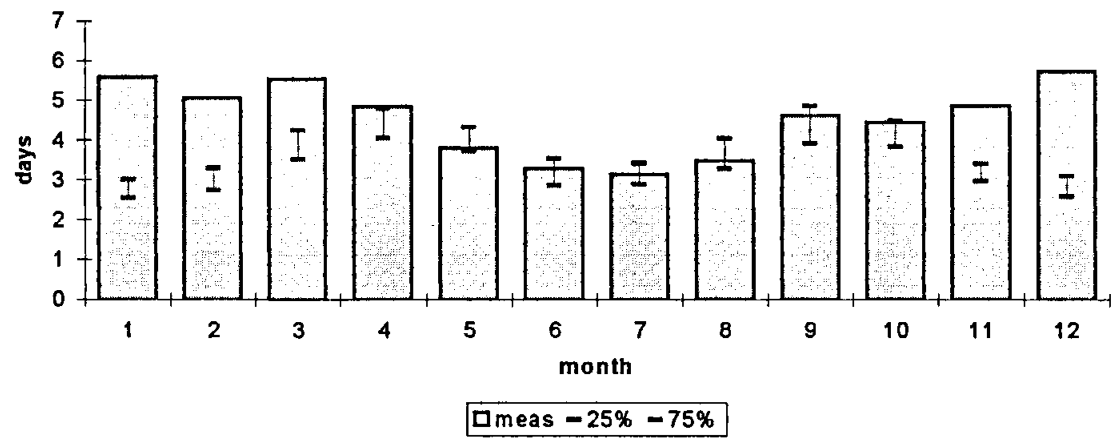


Figure 3.18. Standard deviation of number of scattered days per month, using 100 30-year simulations of the 2-harmonic model.

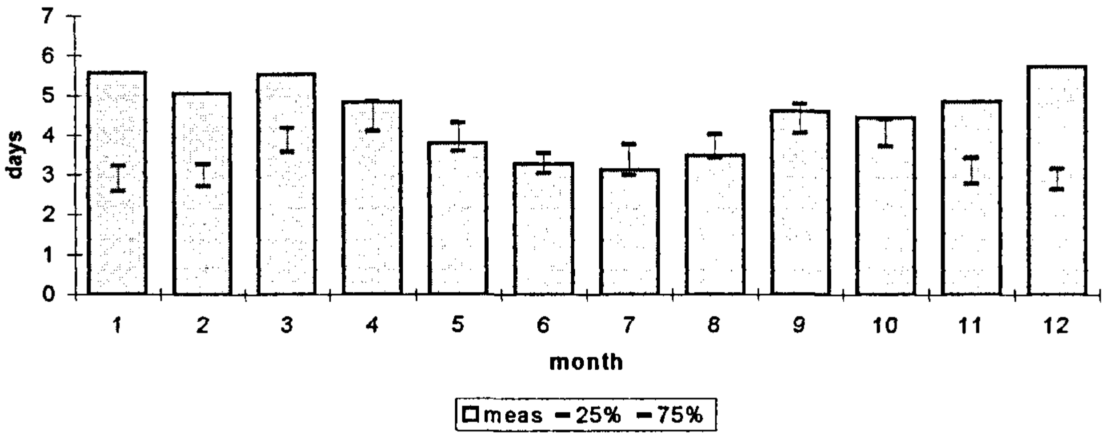


Figure 3.19. Standard deviation of number of scattered days per month, using 100 30-year simulations of the 3-harmonic model.

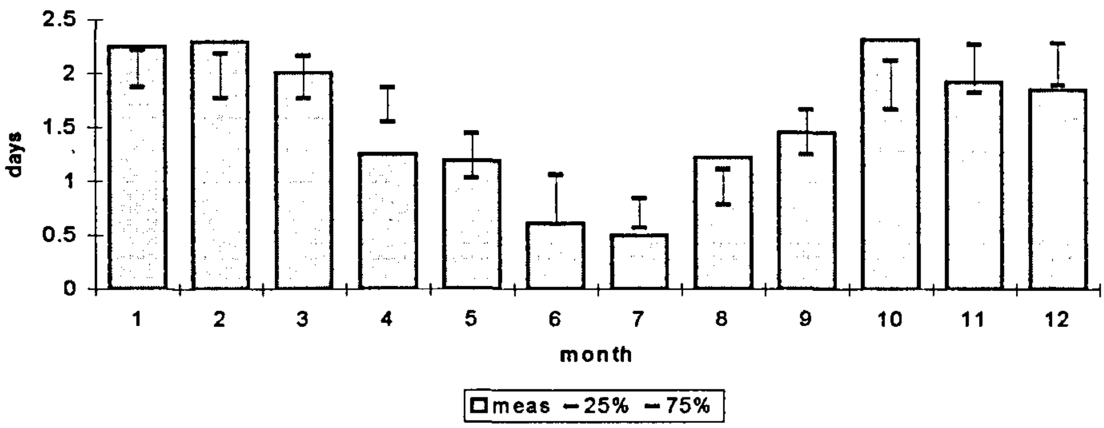


Figure 3.20. Standard deviation of number of general rain days per month, using 100 30-year simulations of the

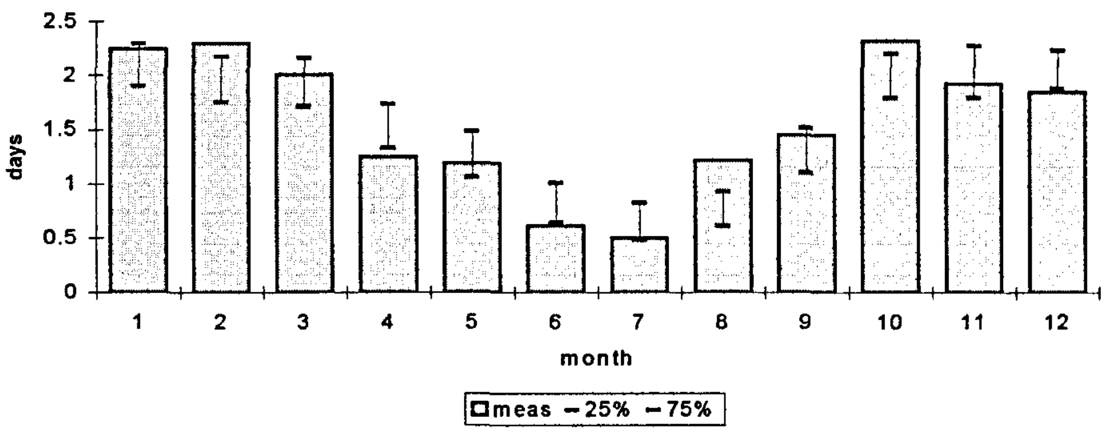


Figure 3.21. Standard deviation of number of general rain days per month, using 100 30-year simulations of the 3-harmonic model.

For example the changes do not capture the difference between the number of historical and simulated general raindays in December as shown in Figure 3.21, which differ by a small 0.50 of a day out of 3. Irritating as this difference may be, it is nevertheless not going to have a significant impact on the overall behaviour of the model. The means of the remainder of the rainday types are comfortably recaptured by the simulation.

Turning to the standard deviations of the dry days, here the model is not as successful. Historically the variability seems to be greater than that generated on average by the model. The scattered and general rainday statistics are better, as can been seen in Figures 3.18 to 3.21. The interpretation of Figures 3.10 and 3.11 which give the means and Figures 3.16 and 3.17 which give the standard deviations of the dry days per month, is that in the very dry months (6, 7 and 8) the model overestimates the number of dry days by about 1.5 to 2 days per month, giving over 25 dry days in a 30 to 31 day month. Because of the 30/31 day upper bound to the number of possible dry days per month, the standard deviation must be reduced to compensate and, in the case of the dry months, is reduced from about 6 days to about 3.5 to 4 days.

The same argument explains why the standard deviations of the number of scattered raindays occurring in wet months is depressed in the simulated compared to the historical sequences. Here the reduction is about 2.5 to 3 days from 5.5 to 6 as a result of the increase of about a day, to 25 wet days in a 31 day month when compared to the historical.

The behaviour of the general raindays is much more acceptable but, the mean number of raindays per month being small, makes this understandable in the light of the above arguments.

It is comforting to note that Figures 3.14 and 3.15 show that the decision to model the general-to-anything transition probabilities as constant does not damage the within-year distribution of the frequency of general raindays in any given month. Beside the slight anomaly of December the rest of the year is good.

Taken all in all the model yields acceptable probabilities of wet and dry days.

3.9.2 The Frequency Distributions of Run-Lengths of the 3 Rainday Types by Month, Using the 2-Harmonic Model.

The tests in this section are all done using the 2-harmonic model, which was shown in Section 3.9.1 to behave almost the same as the model with the extra harmonic and is, in addition, a better model statistically.

The figures are presented in 3 groups of 12. Each group records the results for a given climate - dry day, scattered rainday and general rainday - and in each group there are 12 figures, one for each month. The longest length of run sampled was 14. This is not because there are no runs of greater length but because runs up to 14 give a good idea of the general behaviour and because it is difficult to deal with overlaps. The reason for this is that what was sampled was the number of complete runs within a month.

If the run overlapped into (finished in) the following month it was ignored. This occurs particularly with run of dry days in the dry months and scattered days in the wet months. Therefore, attempting to compute the total number of historic days as 10957 from the figures will fail! The rationale behind this ploy was simplicity and efficacy. If a sequence overlapped 2 or more months, to which should it belong, the starting or finishing month or the one with a greater proportion of the run? What about ties? Keeping it simple made for ease of programming and also made the result unequivocal.

There are occasions when there are fluctuations between the simulations under- and over-estimating the frequency of recorded runs of various types but, by and large, what is observed in the following is entirely satisfactory. In particular, it is gratifying to note that the model, *in no case*, generated a run of general raindays longer than 5, the upper limit of what was observed historically.

Runs of dry days

Figures 3.22 through 3.33 record the behaviour of the runs of dry days. What each figures shows is the number of runs of dry days of a given length in each 30 year sequence, not the frequency of run of dry days occurring in a particular month. To get that frequency, one needs to divide the values in the figures by 30.

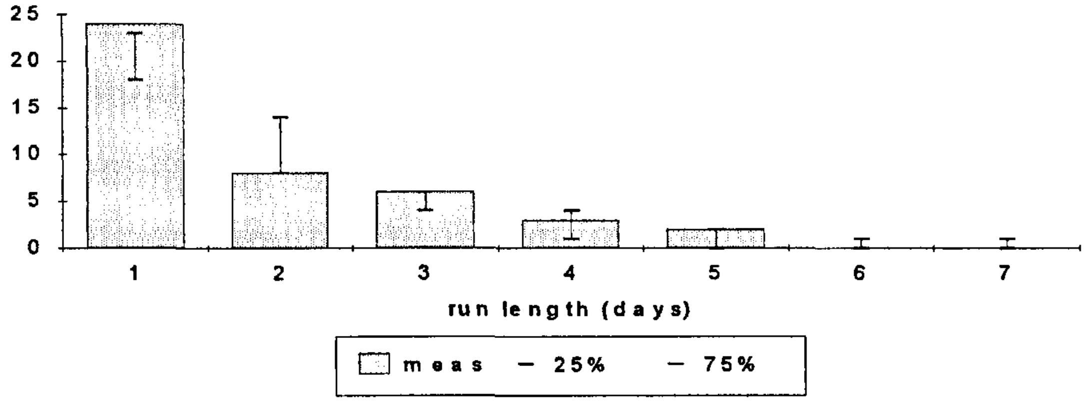


Figure 3.22. Histogram of frequency of run lengths of dry days in January using 100 30-year simulations of the 2-harmonic model.

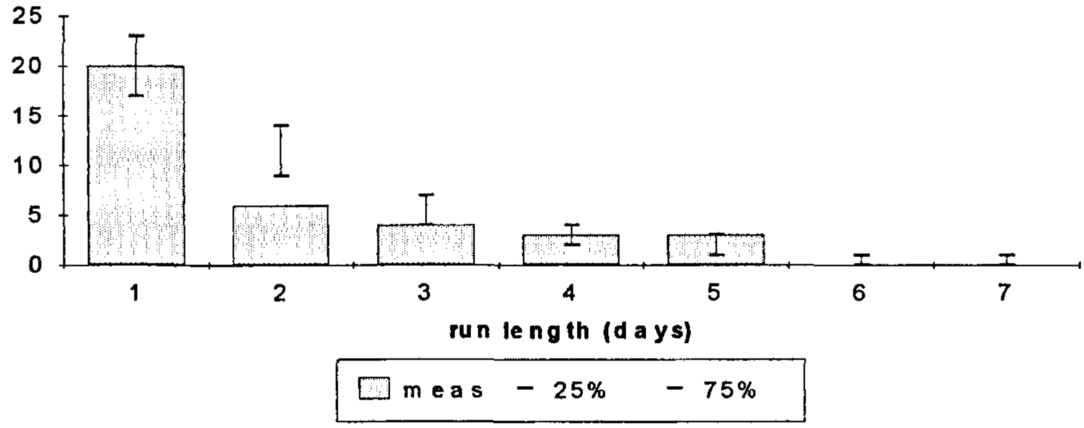


Figure 3.23. Histogram of frequency of run lengths of dry days in February using 100 30-year simulations of the 2-harmonic model.

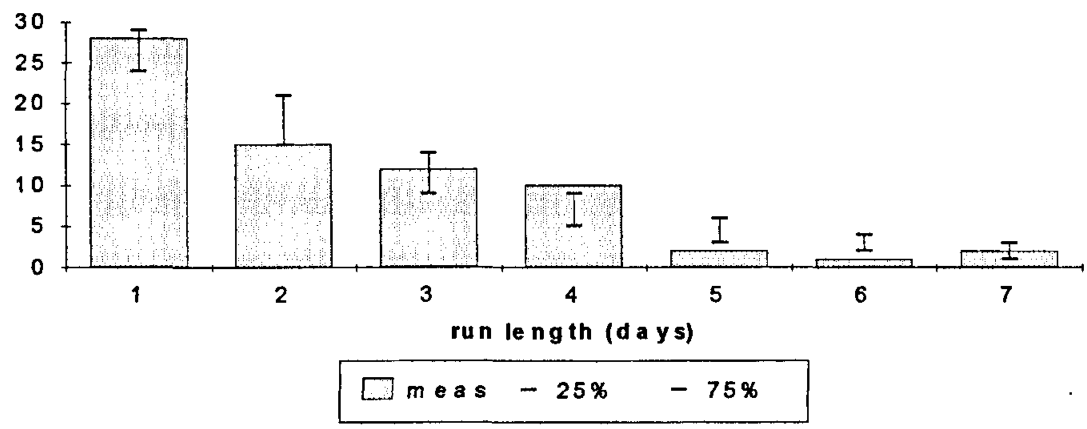


Figure 3.24. Histogram of frequency of run lengths of dry days in March using 100 30-year simulations of the 2-harmonic model.

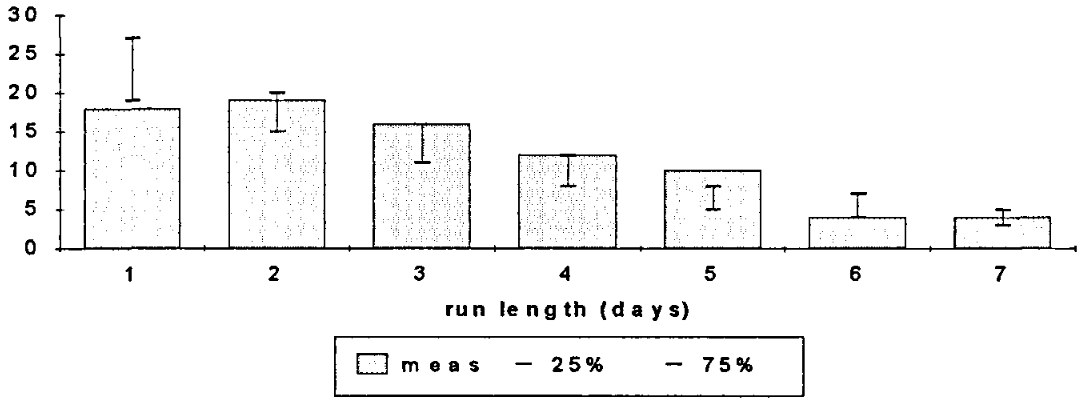


Figure 3.25. Histogram of frequency of run lengths of dry days in April using 100 30-year simulations of the 2-harmonic model.

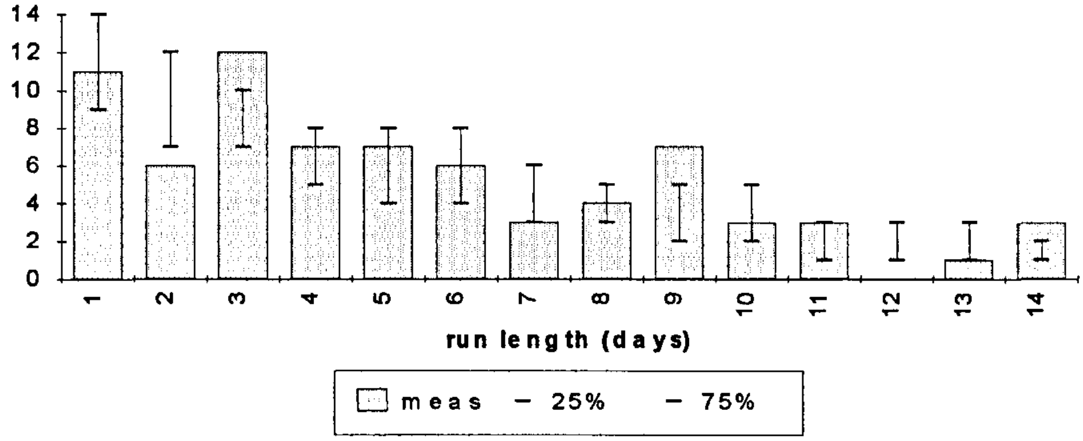


Figure 3.26. Histogram of frequency of run lengths of dry days in May using 100 30-year simulations of the 2-harmonic model.

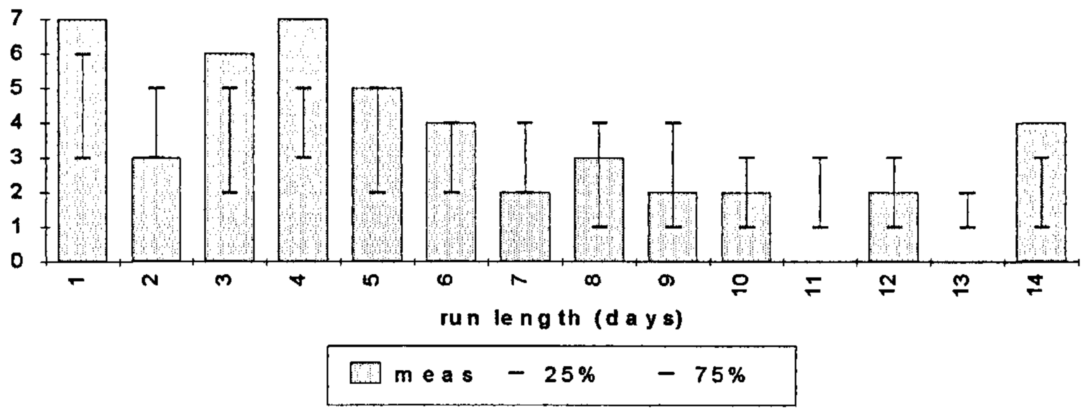


Figure 3.27. Histogram of frequency of run lengths of dry days in June using 100 30-year simulations of the 2-harmonic model.

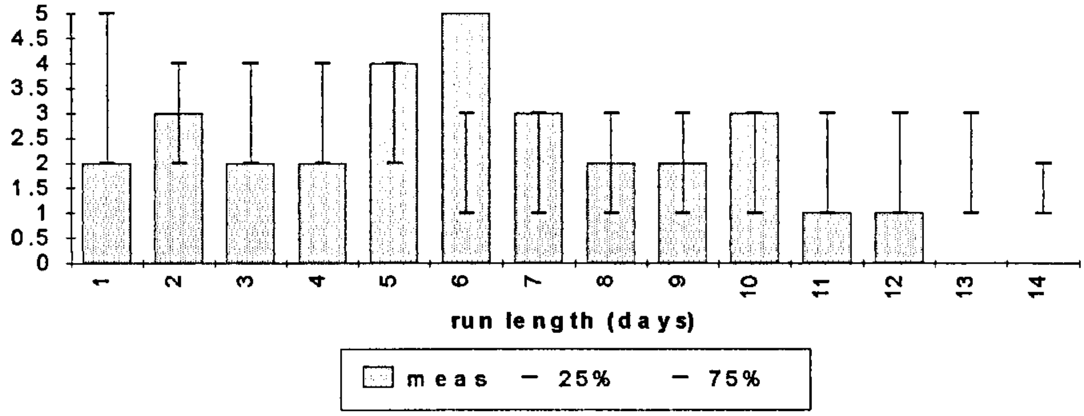


Figure 3.28. Histogram of frequency of run lengths of dry days in July using 100 30-year simulations of the 2-harmonic model.

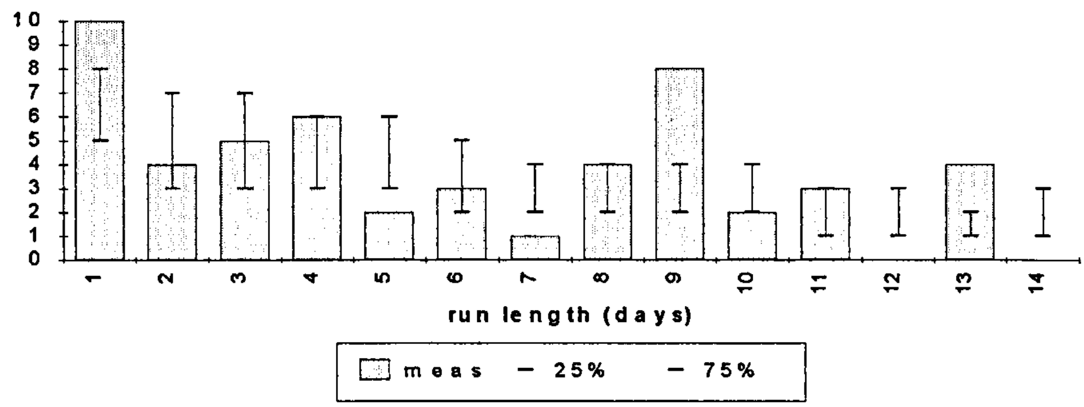


Figure 3.29. Histogram of frequency of run lengths of dry days in August using 100 30-year simulations of the 2-harmonic model.

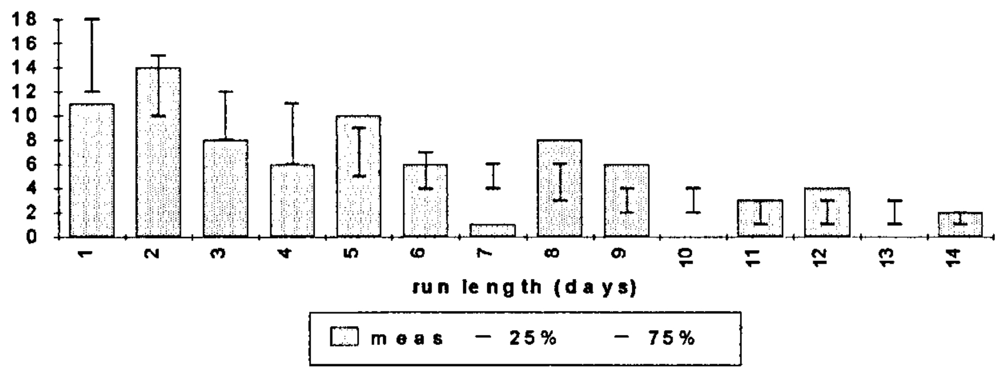


Figure 3.30. Histogram of frequency of run lengths of dry days in September using 100 30-year simulations of the 2-harmonic model.

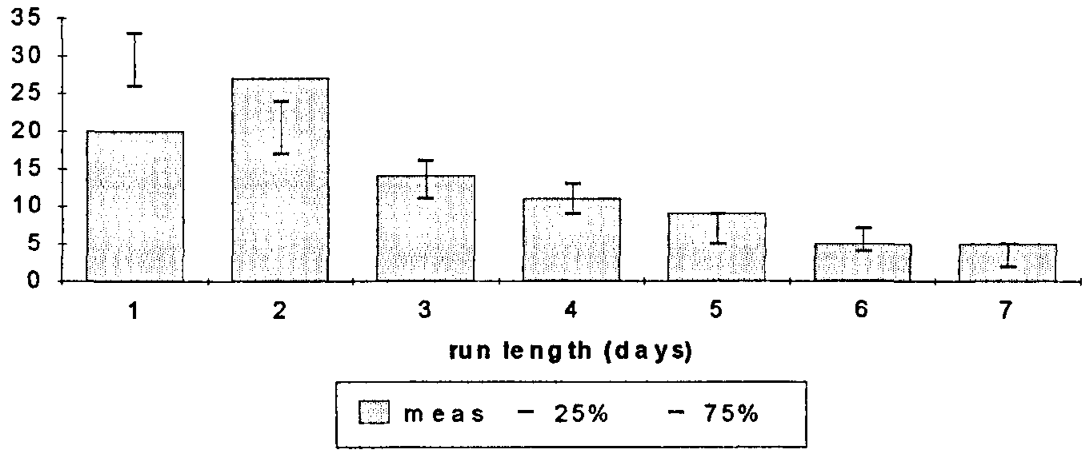


Figure 3.31. Histogram of frequency of run lengths of dry days in October using 100 30-year simulations of the 2-harmonic model.

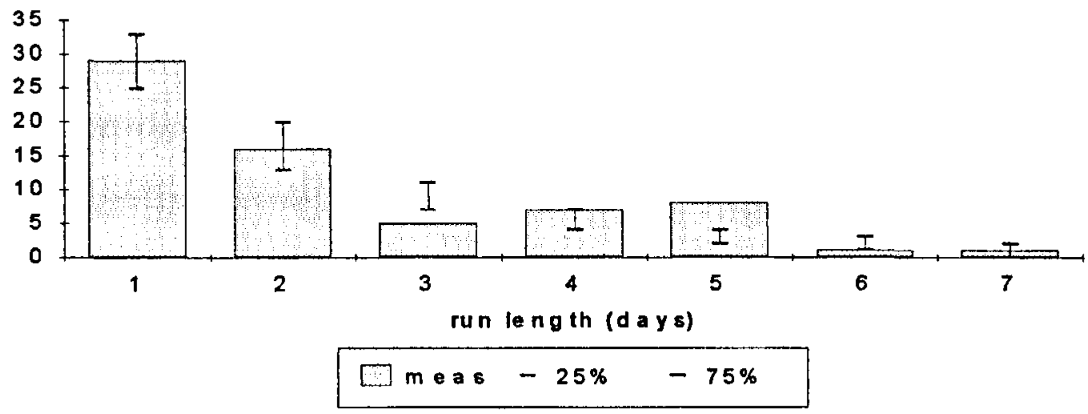


Figure 3.32. Histogram of frequency of run lengths of dry days in November using 100 30-year simulations of the 2-harmonic model.

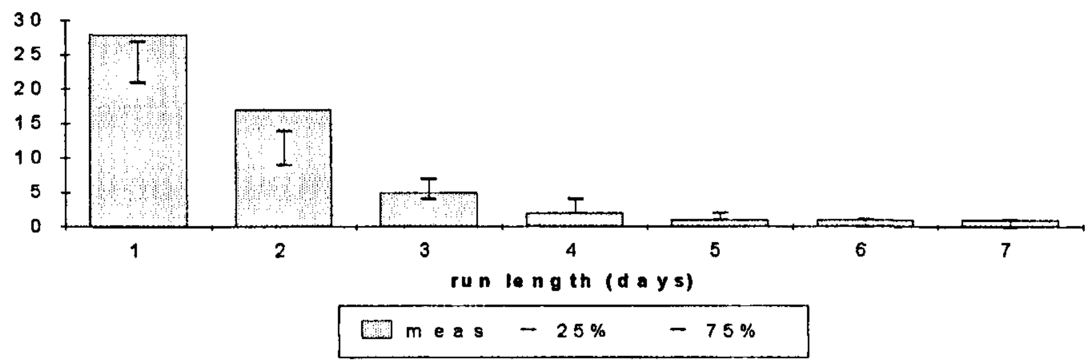


Figure 3.33. Histogram of frequency of run lengths of dry days in December using 100 30-year simulations of the 2-harmonic model.

Thus in Figure 3.22, which shows the historical and simulated runs for January, in all 30 historical years there were:

- 24 dry day runs of length 1 day
- 8 runs of length 2 days
- 6 runs of length 3 days
- 3 runs of length 4 days
- 2 runs of length 5 days

The dumbbells record the inter-quartile range of the values for the 100 sets of 30-year sequences which of necessity are computed in integers. Thus 75 of the simulated sequences recorded 1 run or no run of 7 days in January and 25 of the 30-year simulated sequences were found to have 8 or less complete runs of 2 dry days in January.

With these means of interpreting the figures kept in mind it appears that the simulations mimic the historical dry run behaviour very well. It is particularly interesting to note the change from a steep geometric (exponential) decay in the wet months to a flat decay in the dry months. Indeed in July (Figure 3.28) the median number of runs of 1 dry day is about 3.5 and this only drops to a median frequency of 2 for runs of 13 dry days; the long tail in the distribution is hinted at in this figure and in others in the dry months. By contrast in Figure 3.33 for December, the historical count of 28 and the simulated median of about 24 runs of 1 dry day decays quickly to 1 and (possibly) 0.5 respectively for the runs of 7 dry days.

This is a very good validation of the Markov chain climate model and confirms that the simple lag-one Markov chain is indeed appropriate. The reason that this is so is because any extra dependence (an extra positive lag) which might lie undetected in the historical sequence would change the run behaviour from what is observed to a much heavier tail. The test is a severe one and the model passes it well.

Runs of scattered raindays

Figures 3.34 to 3.45 record the behaviour of runs of scattered raindays by month. The interpretation of these figures is the same as those of the dry day runs.

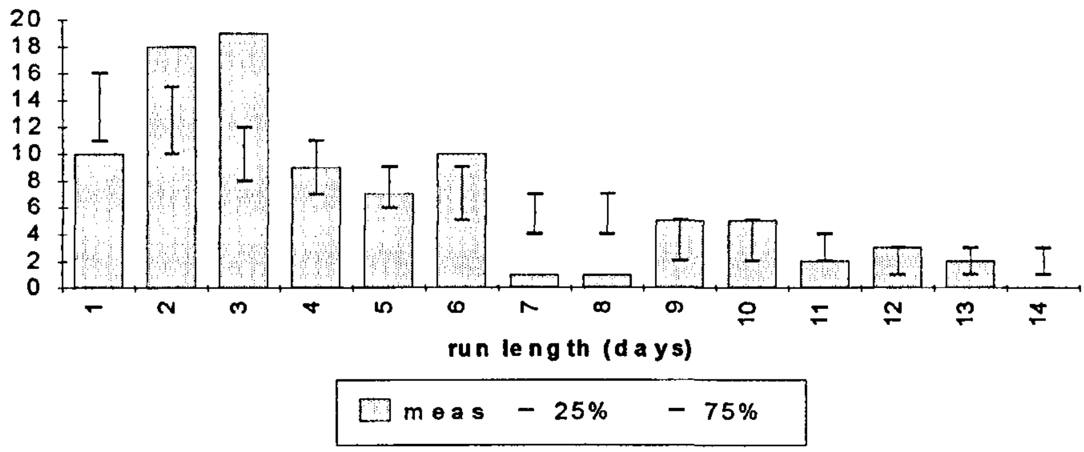


Figure 3.34. Histogram of frequency of run lengths of scattered rain days in January using 100 30-year simulations of the 2-harmonic model.

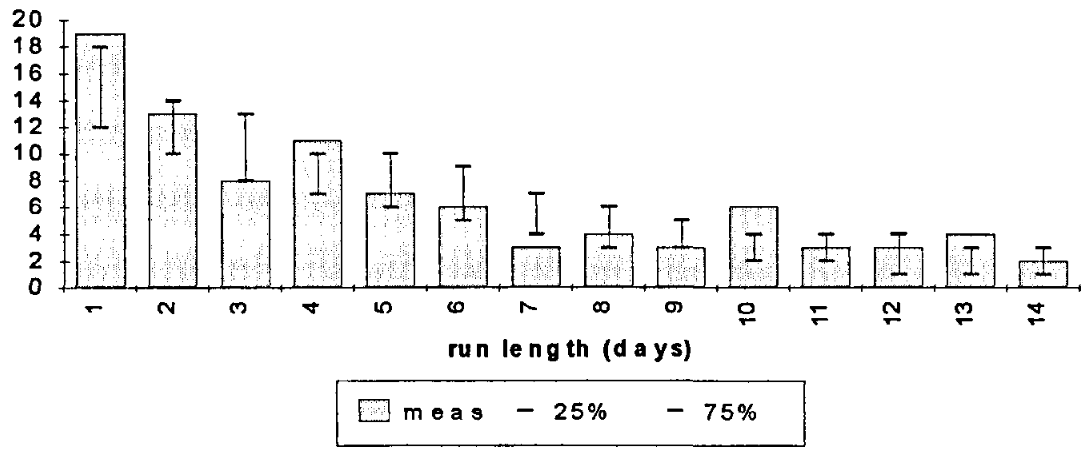


Figure 3.35. Histogram of frequency of run lengths of scattered rain days in February using 100 30-year simulations of the 2-harmonic model.

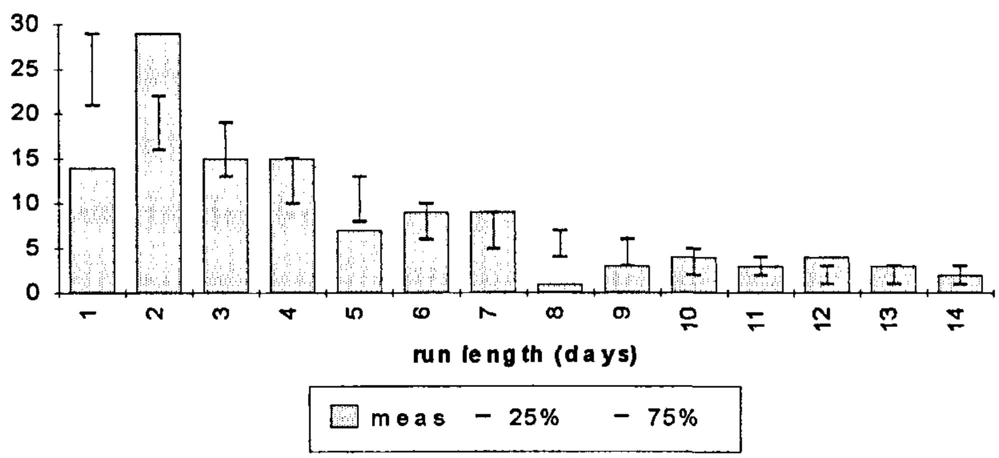


Figure 3.36. Histogram of frequency of run lengths of scattered rain days in March using 100 30-year simulations of the 2-harmonic model.

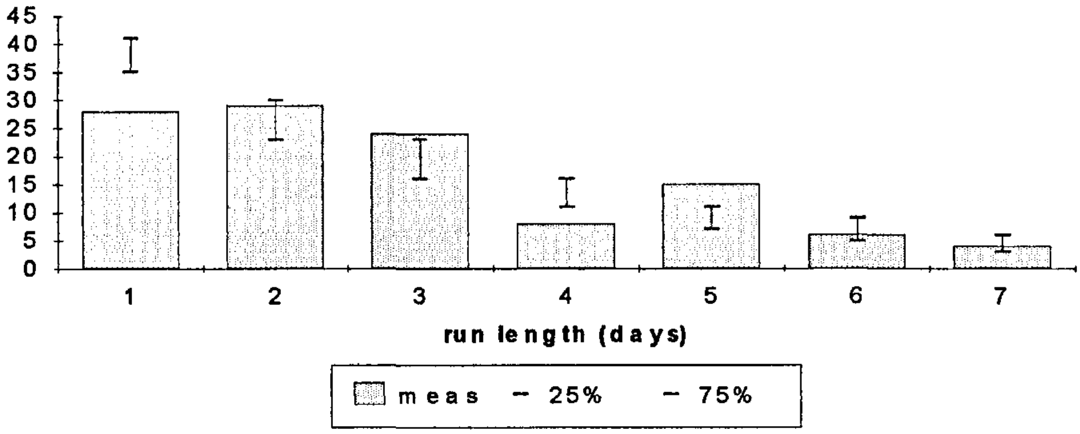


Figure 3.37. Histogram of frequency of run lengths of scattered rain days in April using 100 30-year simulations of the 2-harmonic model.

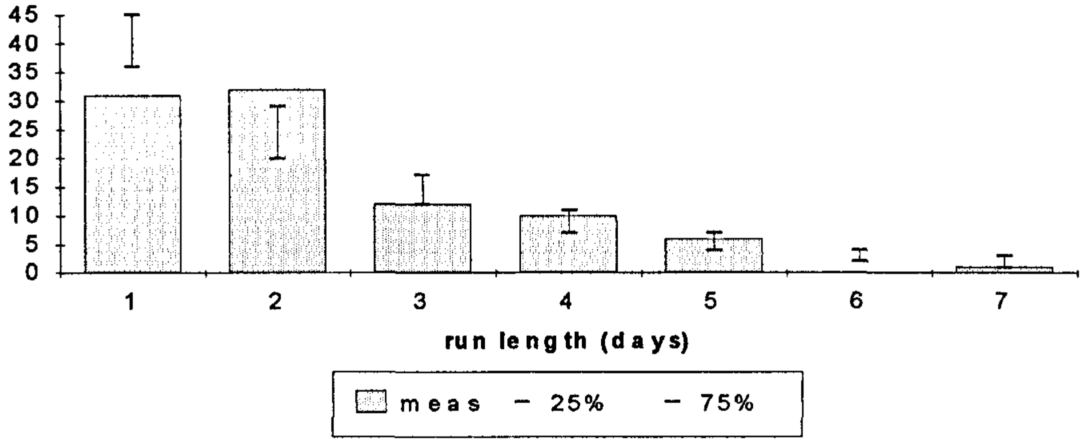


Figure 3.38. Histogram of frequency of run lengths of scattered rain days in May using 100 30-year simulations of the 2-harmonic model.

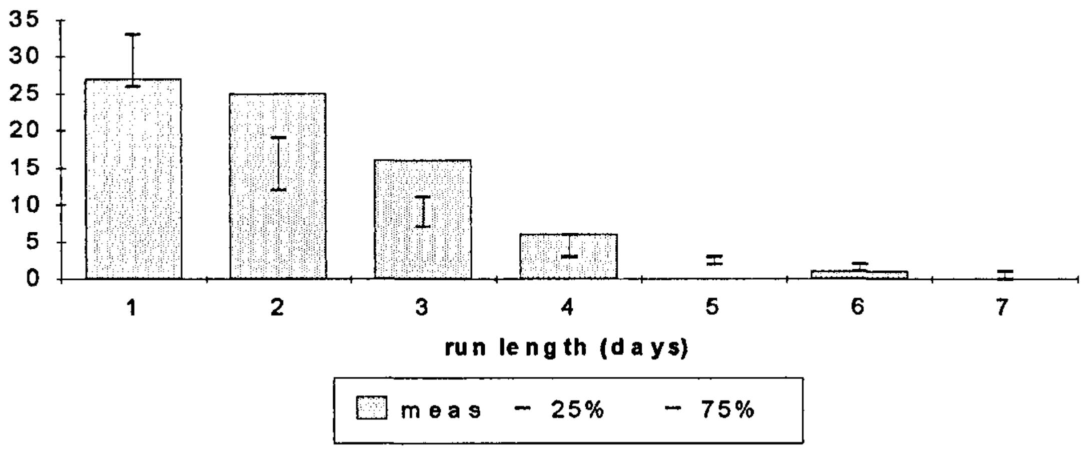


Figure 3.39. Histogram of frequency of run lengths of scattered rain days in June using 100 30-year simulations of the 2-harmonic model.

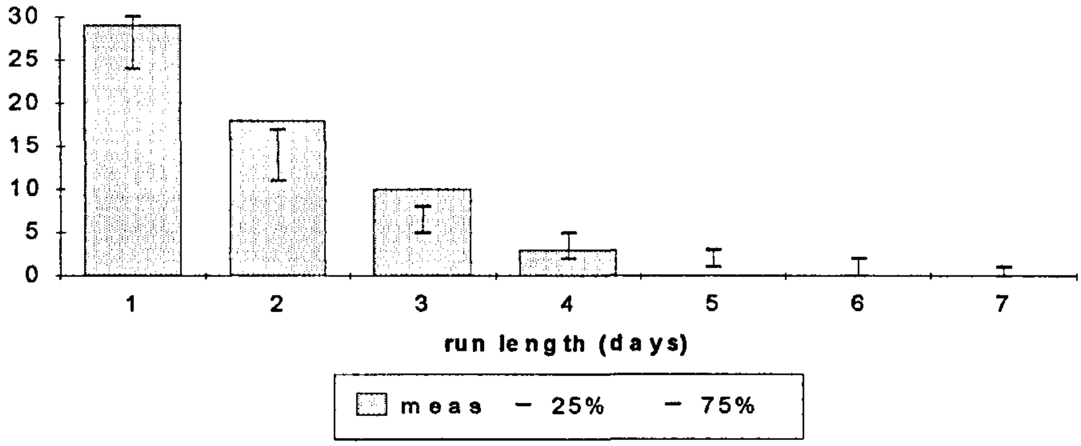


Figure 3.40. Histogram of frequency of run lengths of scattered rain days in July using 100 30-year simulations of the 2-harmonic model.

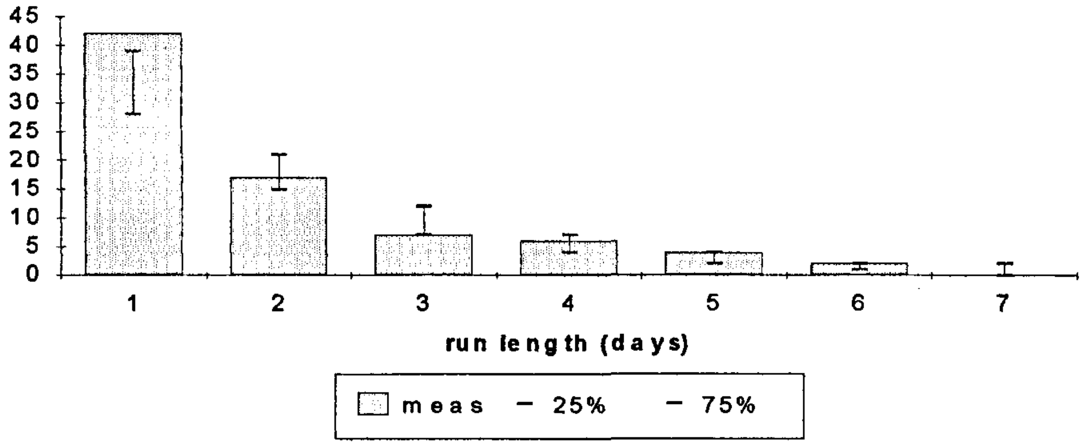


Figure 3.41. Histogram of frequency of run lengths of scattered rain days in August using 100 30-year simulations of the 2-harmonic model.

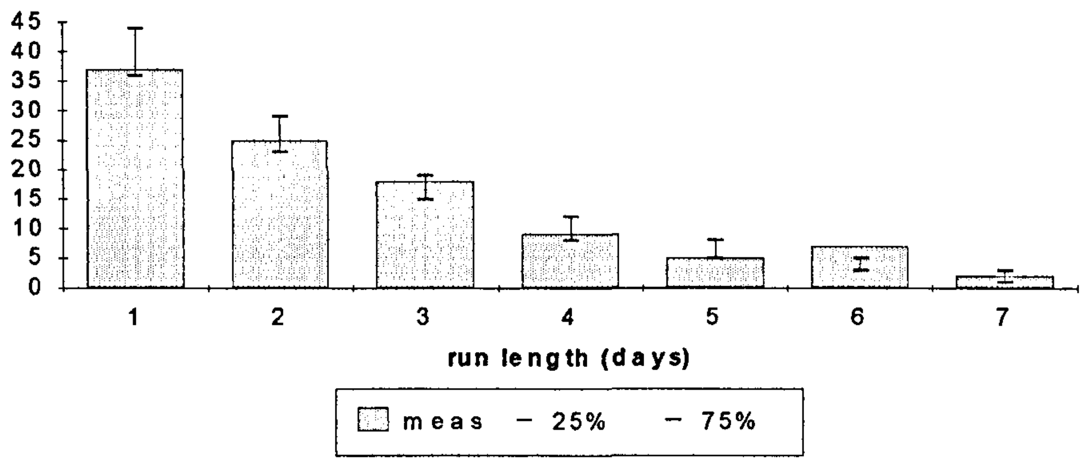


Figure 3.42. Histogram of frequency of run lengths of scattered rain days in September using 100 30-year simulations of the 2-harmonic model.

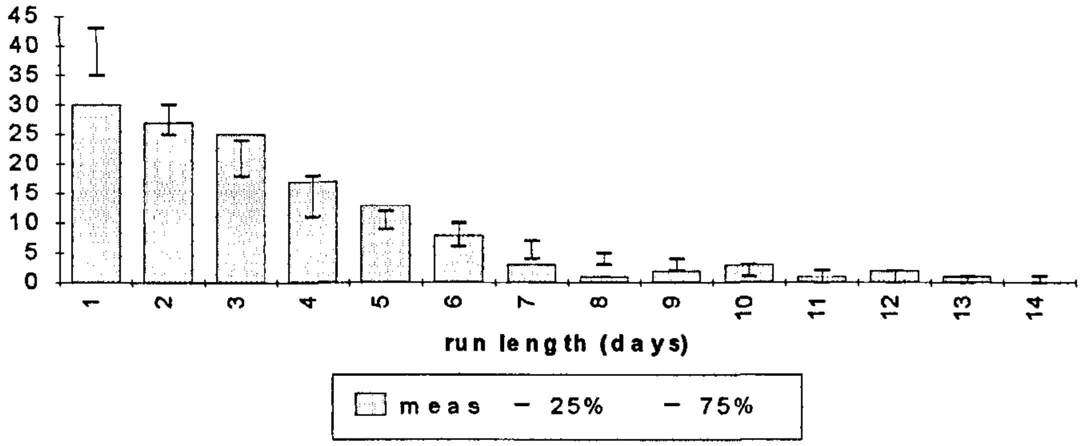


Figure 3.43. Histogram of frequency of run lengths of scattered rain days in October using 100 30-year simulations of the 2-harmonic model.

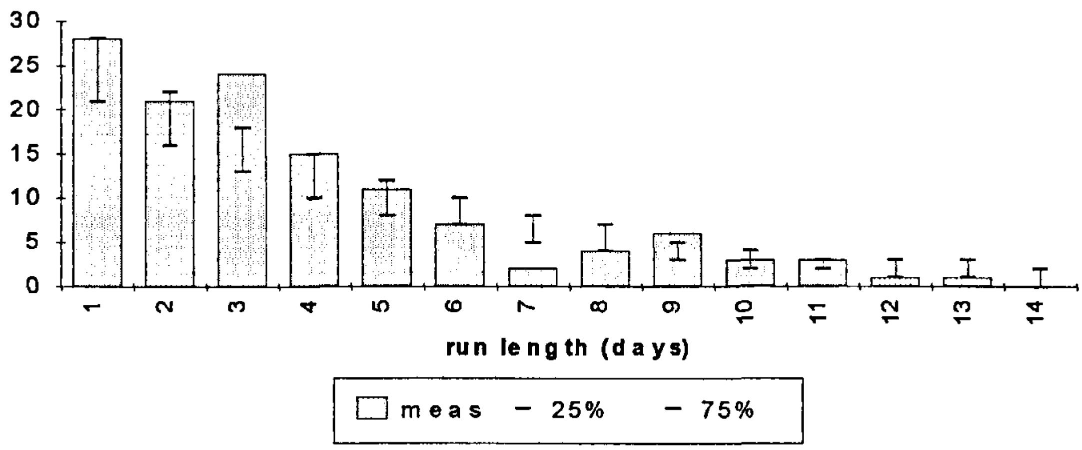


Figure 3.44. Histogram of frequency of run lengths of scattered rain days in November using 100 30-year simulations of the 2-harmonic model.

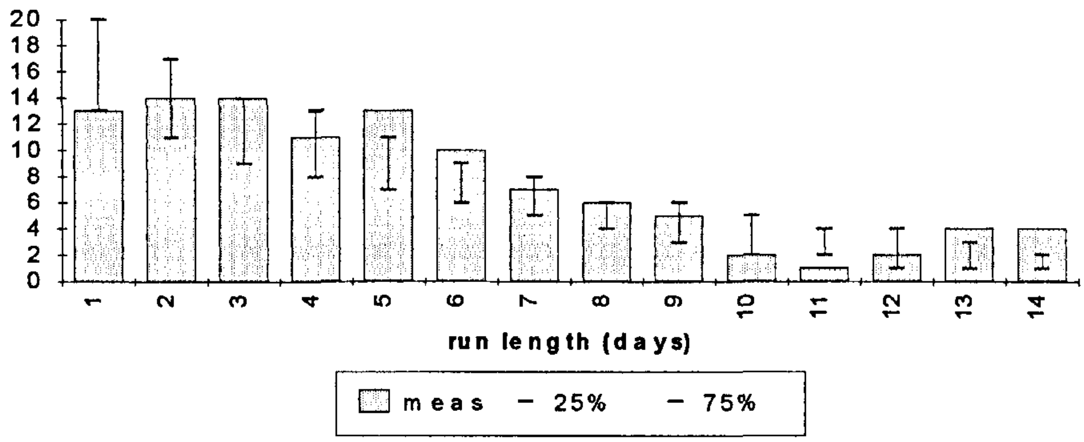


Figure 3.45. Histogram of frequency of run lengths of scattered rain days in December using 100 30-year simulations of the 2-harmonic model.

A feature of these diagrams is the variability of the historical sequences compared to the simulated ones, particularly between the various run lengths in a given diagram. Although there is a general geometric trend in each set, January and February in particular show this marked variation, which can be ascribed to sampling. The remainder of the figures exhibit much smoother behaviour of the historical sequence.

Another feature which draws comment is the almost alternating under- and over-estimation of the frequency of the observed runs of only one scattered rainday from one month to the next. Figures 3.34 to 3.36 exhibit this as do Figures 3.39 to 3.45. Some, though not all, months where there are a small number of historical one day runs seem to compensate by having a larger number of 2 day and even 3 day runs than would be expected from a geometric model.

This commentary is tending to be a little on the elaborate side but it does indicate that although the model has captured the broad-brush behaviour of the climate, there are details which may possibly be improved in future work. It is difficult to say if the departures of the historical runs of the scattered raindays are due to the method of sampling (which seems to be unlikely because the simulated sequences were treated the same way) or due to a more complex time series model being required. The matter will not be further addressed in this report.

Runs of general raindays

Figures 3.46 to 3.57 record the behaviour of the runs of general raindays by month. The interpretation of these figures is the same as the dry day and scattered rainday data. The correspondence between the historical and simulated sequence is excellent and justifies the decision to keep the general-to-anything transition probabilities constant.

Of particular significance is the observation that in January to April the maximum length, both observed and simulated, was 5 days. This result is a particularly gratifying validation of the time series model of the climate using a 3-state Markov chain.

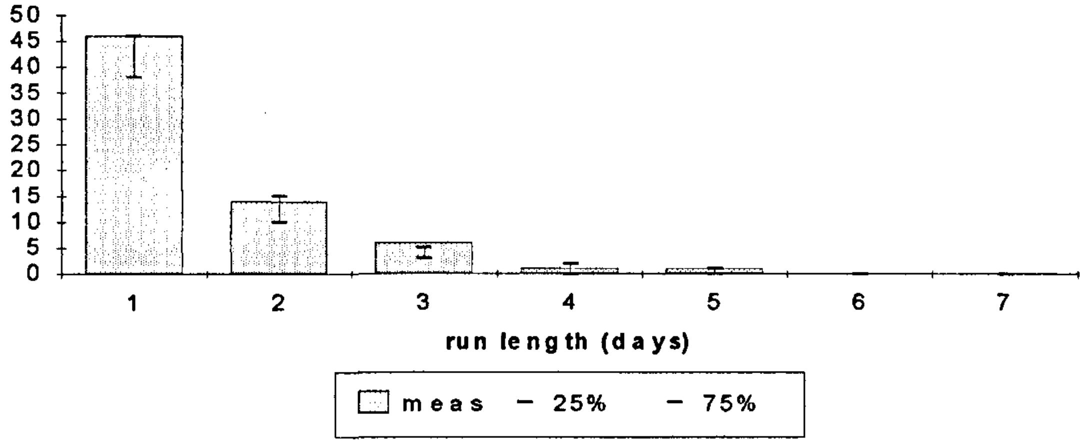


Figure 3.46. Histogram of frequency of run lengths of general rain days in January using 100 30-year simulations of the 2-harmonic model.

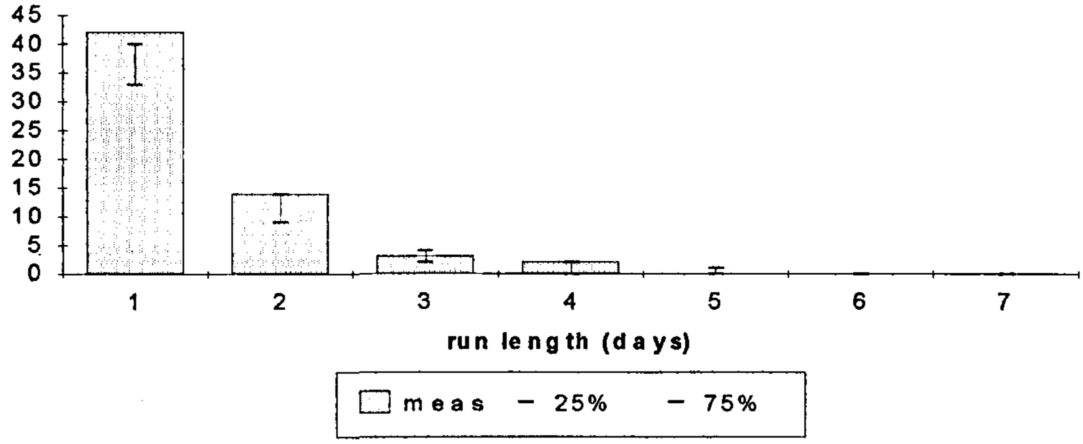


Figure 3.47. Histogram of frequency of run lengths of general rain days in February using 100 30-year simulations of the 2-harmonic model.

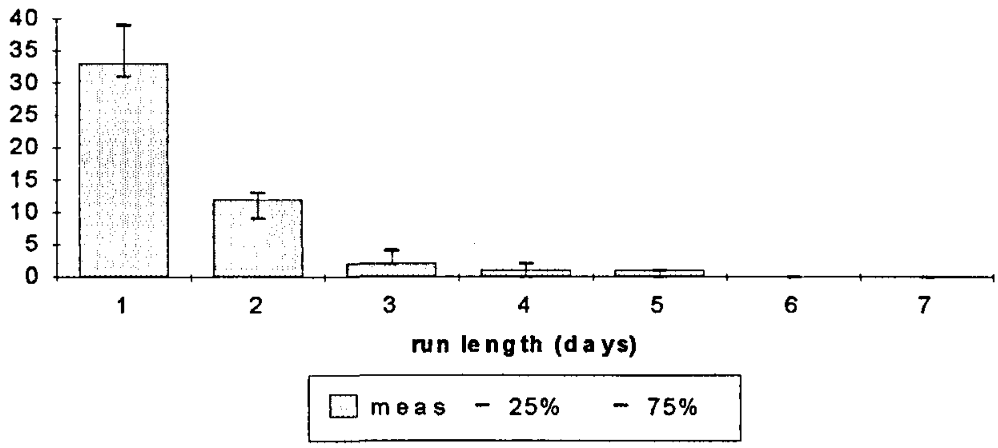


Figure 3.48. Histogram of frequency of run lengths of general rain days in March using 100 30-year simulations of the 2-harmonic model.

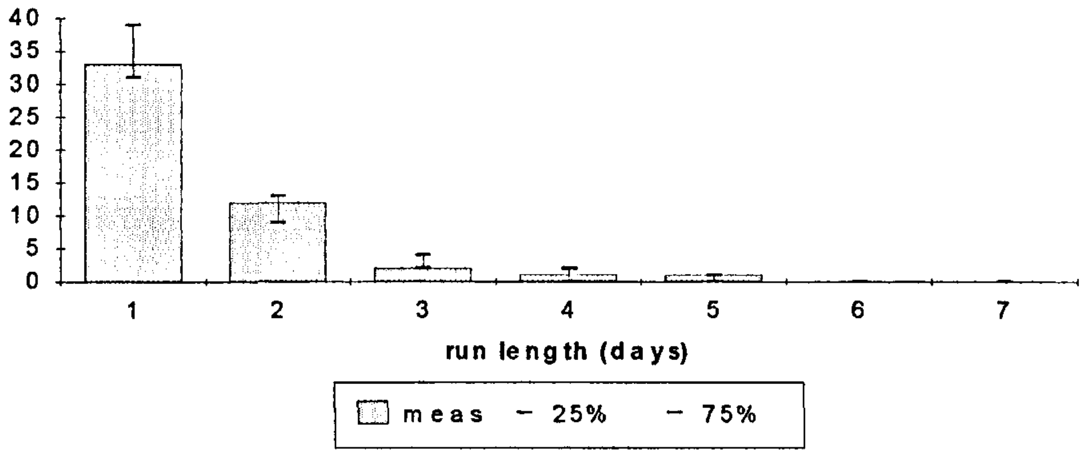


Figure 3.49. Histogram of frequency of run lengths of general rain days in April using 100 30-year simulations of the 2-harmonic model.

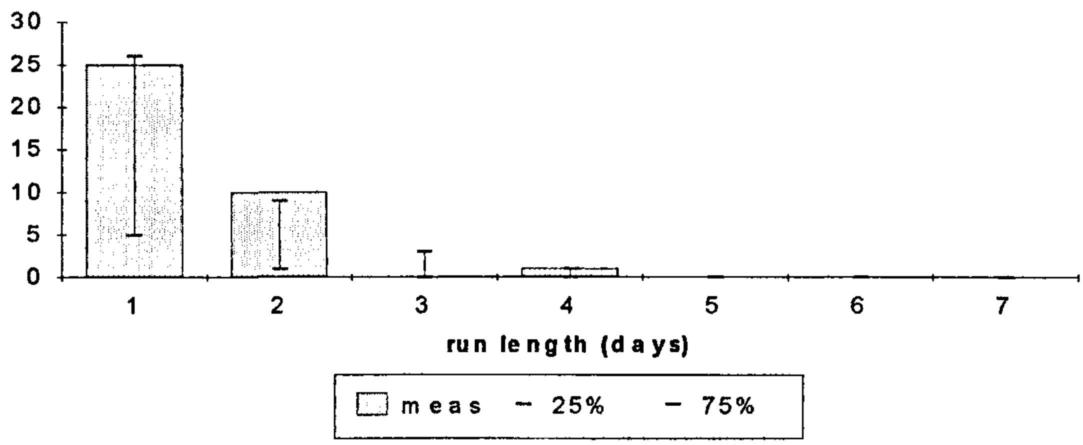


Figure 3.50. Histogram of frequency of run lengths of general rain days in May using 100 30-year simulations of the 2-harmonic model.

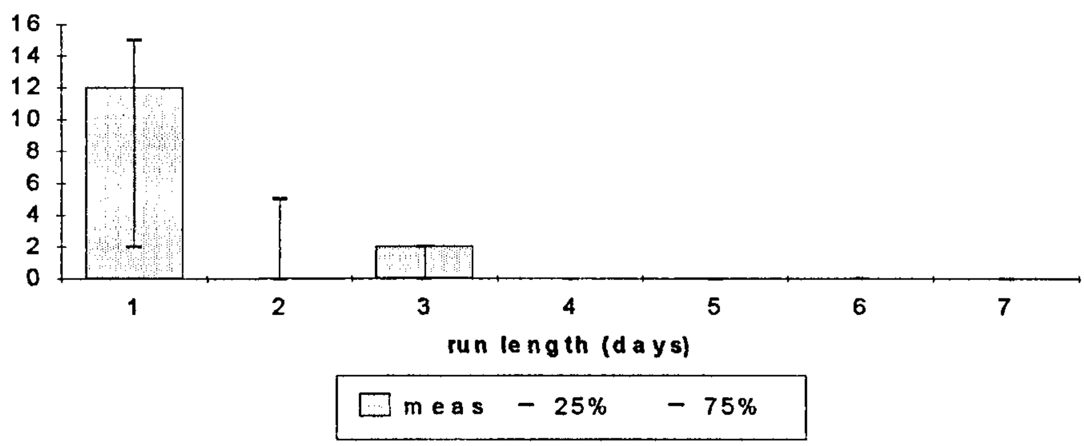


Figure 3.51. Histogram of frequency of run lengths of general rain days in June using 100 30-year simulations of the 2-harmonic model.

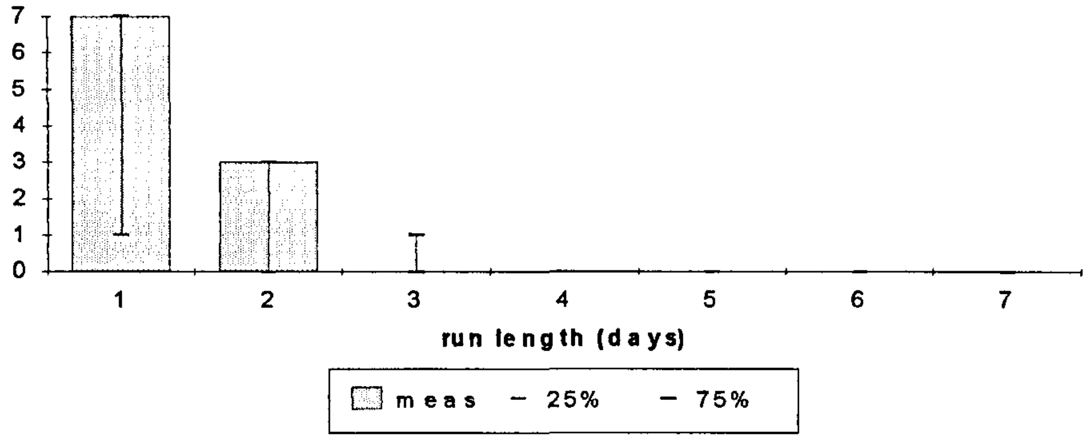


Figure 3.52. Histogram of frequency of run lengths of general rain days in July using 100 30-year simulations of the 2-harmonic model.

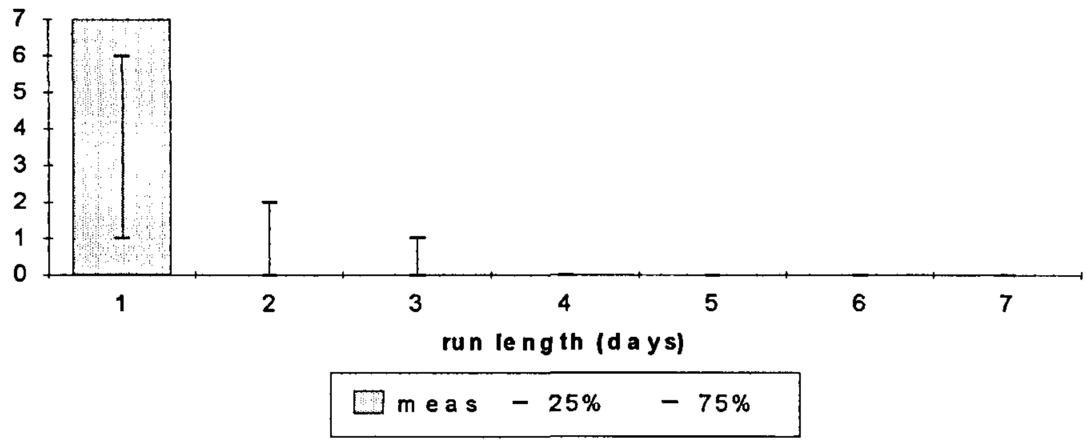


Figure 3.53. Histogram of frequency of run lengths of general rain days in August using 100 30-year simulations of the 2-harmonic model.

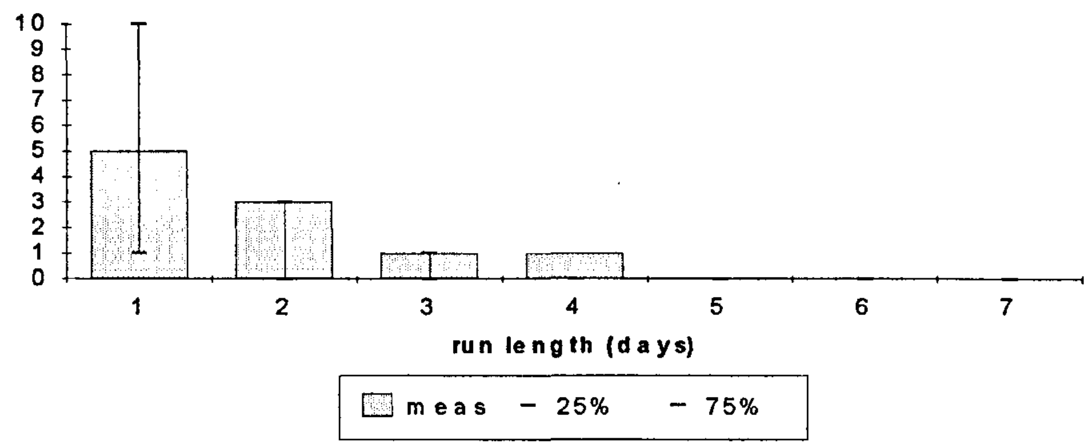


Figure 3.54. Histogram of frequency of run lengths of general rain days in September using 100 30-year simulations of the 2-harmonic model.

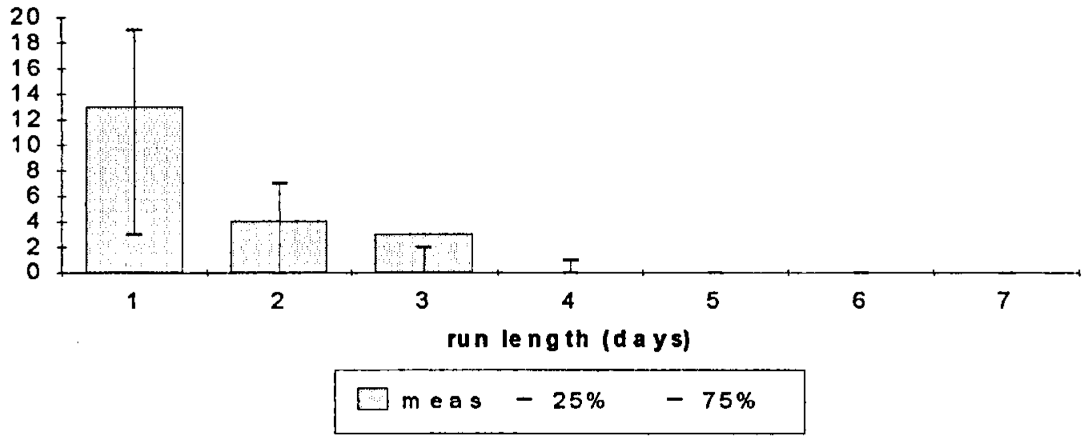


Figure 3.55. Histogram of frequency of run lengths of general rain days in October using 100 30-year simulations of the 2-harmonic model.

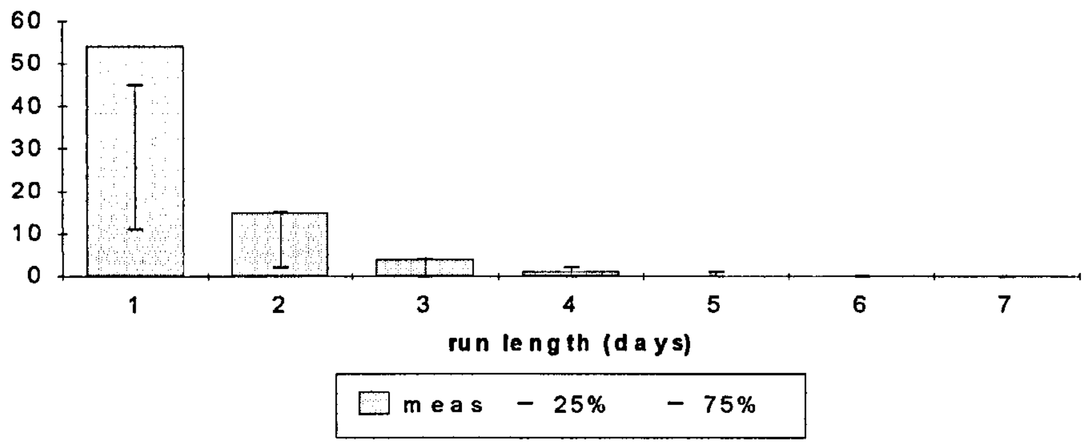


Figure 3.56. Histogram of frequency of run lengths of general rain days in November using 100 30-year simulations of the 2-harmonic model.

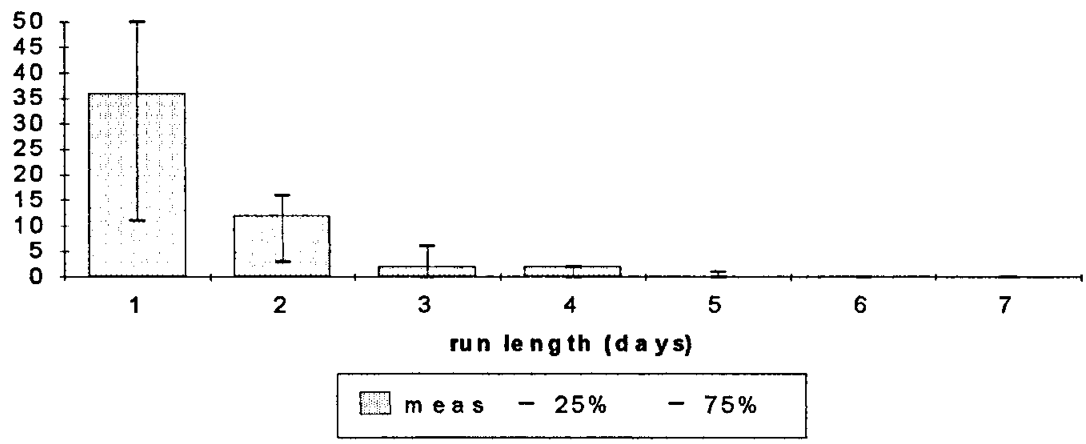


Figure 3.57. Histogram of frequency of run lengths of general rain days in December using 100 30-year simulations of the 2-harmonic model.

In May this maximum length, observed and simulated, reduces to 4 days. In June to September the simulated sequences yielded a maximum of 3 days which was exceeded by one 4-day event in September. (This must surely have been a misclassification of one of the historic days!) In October the simulated maximum increases to one run of 4 days and in November and December goes back to a maximum of 5.

This run behaviour is purely the artefact of the Markov chain climate model and it is significant that the number of 5 day runs of general raindays observed historically was only 3. Even more satisfying is the fact that the model did not simulate any runs of general raindays longer than 5. We suggest that the model is thus good at replicating the important rain-producing mechanism, namely the number of general raindays, which has been shown to be the most significant decider between what is a good and bad rainfall year.

3.9.3 Mean and Standard Deviations of Monthly Rainfall Totals using the 2 and 3-Harmonic Models.

This set of validation tests records the behaviour of the combined model, which consists of the climate component on the one hand and amount per rainday type on the other. It is a crucial test of the efficacy of the model because it tests the effective output of the quantity of interest which was not explicitly incorporated into the model, namely the amount of rain that falls in a given month. The Figures recording this set of validation tests is the sequence from 3.58 to 3.61

Figures 3.58 and 3.59 record the behaviour of historical and simulated means of the 30-year sequences by month using the 2- and 3-harmonic models respectively. The differences between the two figures are minor indicating that in this context the 2-harmonic model is adequate.

There are only two marked differences between the historical and simulated means. For January the median of the simulated sequences is 6mm below the historical observed mean of 117mm, while for December the median of the simulated amounts is 9mm above the observed mean of 91mm.

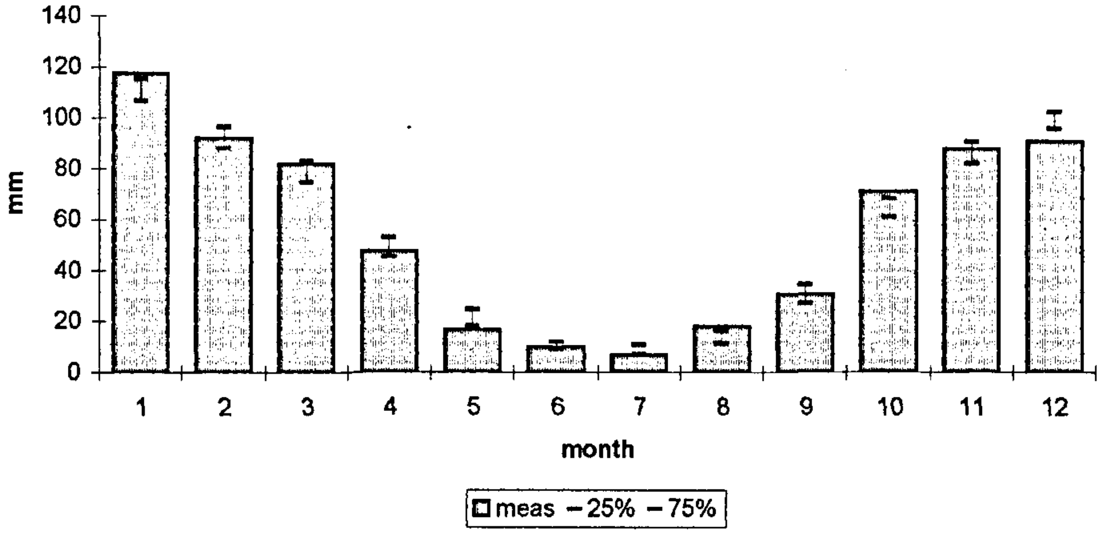


Figure 3.58. Mean monthly rainfall, using 100 30-year simulations of the 2-harmonic model.

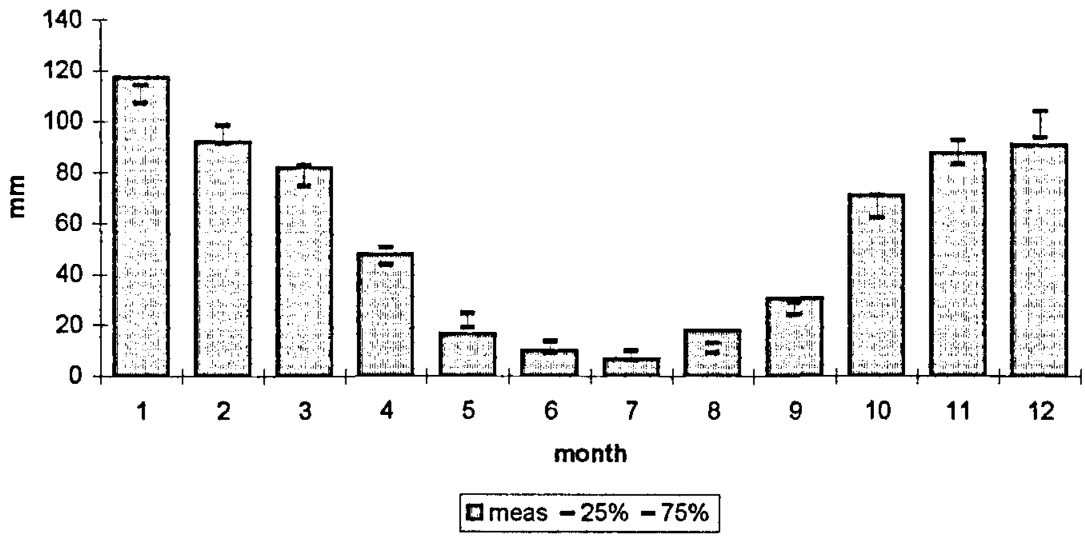


Figure 3.59. Mean monthly rainfall, using 100 30-year simulations of the 3-harmonic model.

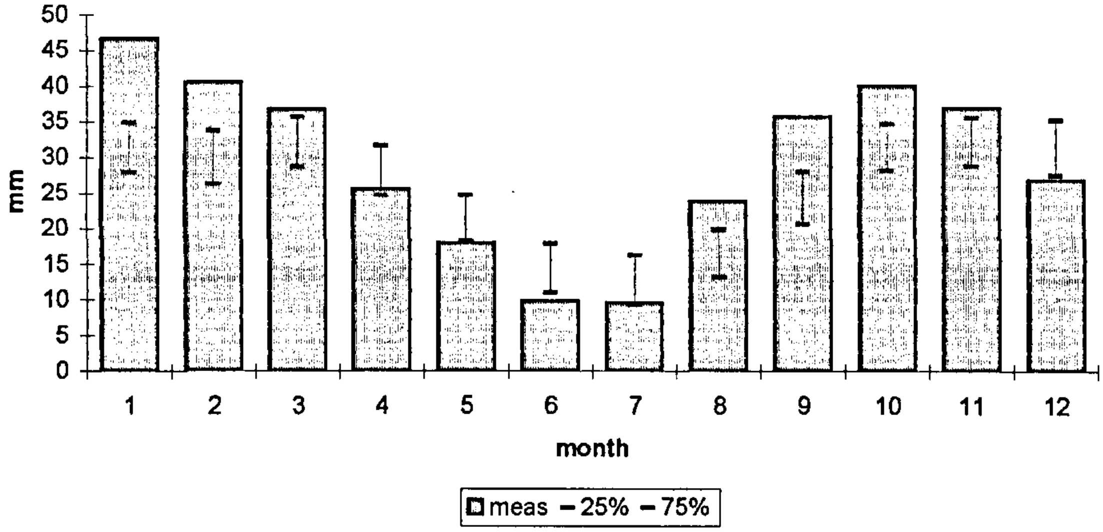


Figure 3.60. Standard deviation of monthly rainfall, using 100 30-year simulations of the 2-harmonic model.

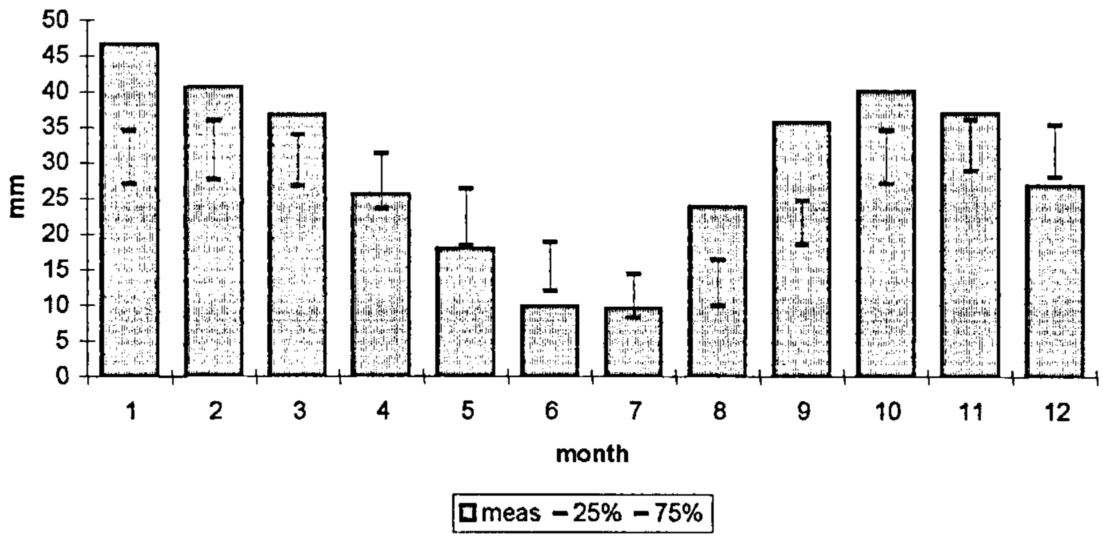


Figure 3.61. Standard deviation of monthly rainfall, using 100 30-year simulations of the 3-harmonic model.

These values are not excessively different, being within 10% of the historical means and the remainder of the months are very well replicated by the simulations. The disturbing aspect of January and December is that the historical means lie outside the interquartile range of the corresponding simulated values. This implies that there is some systematic bias in the model. We have not been able to put a finger on the reason so have left the small differences unexplained. The easy way out is to ascribe the differences to sampling error but that is not necessarily true and without much further work it would be wrong to suggest such a cause.

Figures 3.60 and 3.61 show the standard deviations of the amounts of rain falling in 30 years by month. It is disappointing to see that the model predicts less variability in the wet months and more variability in the dry months than exhibited by the historical sequences. Again, like the differences in the means, we have not been able to suggest a satisfactory explanation for this departure. We note however that they are in the same sense as those of the means, i.e. standard deviation and mean tend to be either under- estimated or over-estimated during the same months.

It is possible that these departures can be ascribed to our decision to resample the general raindays independently of the time of year. We noted in Section 3.6 that ignoring the variation of intensity of general rain per day over the year would induce a bias and this may be a contributory cause but is probably not the whole story.

The partial conclusion of these tests is that the model works reasonably well at capturing the total rainfall behaviour on an annual basis.

3.9.4 Recurrence Intervals of Annual Maxima of Rainfall Totals in 1, 2, 3 and 7 Day Runs, using the 2- and 3- Harmonic Models.

Figures 3.62 to 3.69 record the last of the validation tests on the model. Each successive pair is modelled by the 2- and 3-harmonic model in turn. There are 4 pairs of figures corresponding to the 1, 2, 3 and 7 day total rainfalls.

Each figure represents the exceedence probability distribution of the 30 historical annual maxima as estimated by fitting an EV1 (Gumbel) distribution to the data set. The percentage points, or the values exceeded for each of the recurrence intervals (RI) of exceedence probability given on the horizontal axis, are calculated and these historical values are given as shaded squares. Of course it is fanciful to estimate a 100-year storm from 30 years of data but this is commonly done in engineering hydrology so we make no apology!

The 100 sets of 30-year simulated sequences were each treated the same way as the historical; the data were scanned for annual maxima, an EV1 distribution fitted and percentage points calculated corresponding to the RIs, the 100 percentage points at each RI were ranked and the interquartile range calculated. The thin dashed lines on the figures are the loci of the interquartile range in each case.

This is a very severe test of model capability especially for the longer duration storms. To produce the correct behaviour for rare storms from a model of this nature is surely to validate its ability to model the climate/rainfall process.

Examining the 8 figures shows that the model works surprising well. One can discuss minor details but overall it is quite capable of producing extreme events of the correct exceedence probability. In this it is entirely successful.

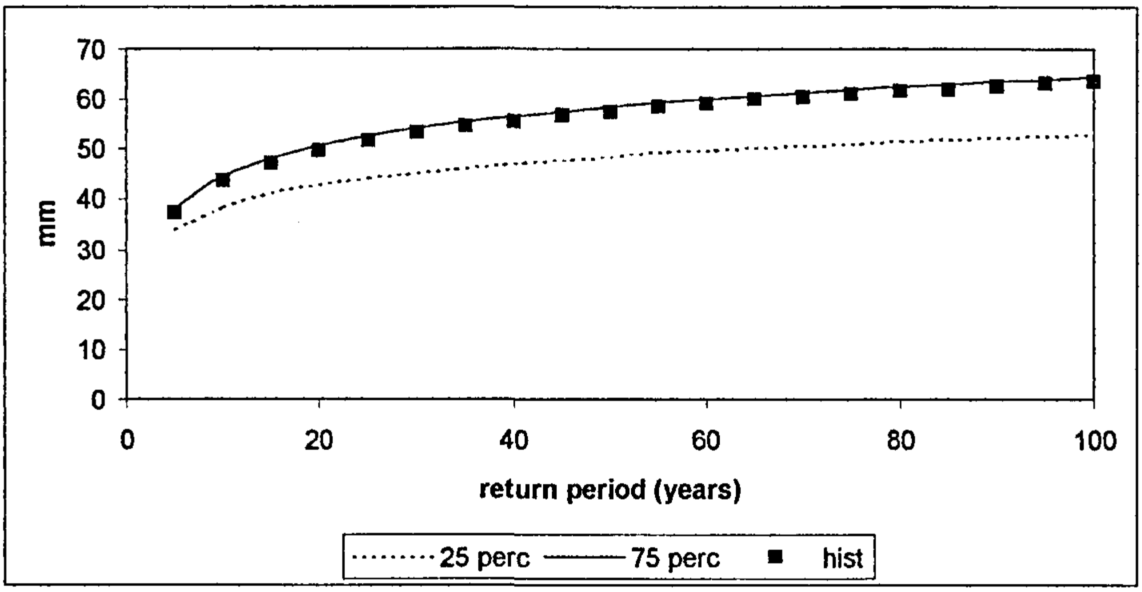


Figure 3.62. 1-day storm rainfall, using 100 30-year simulations of the 2-harmonic model.

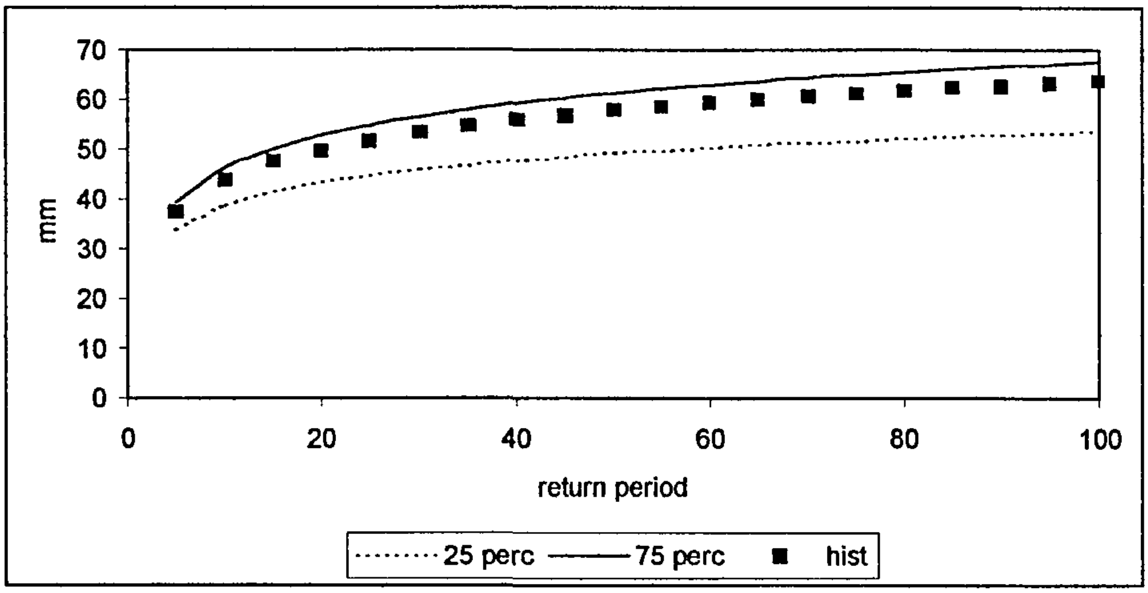


Figure 3.63. 1-day storm rainfall, using 100 30-year simulations of the 3-harmonic model.

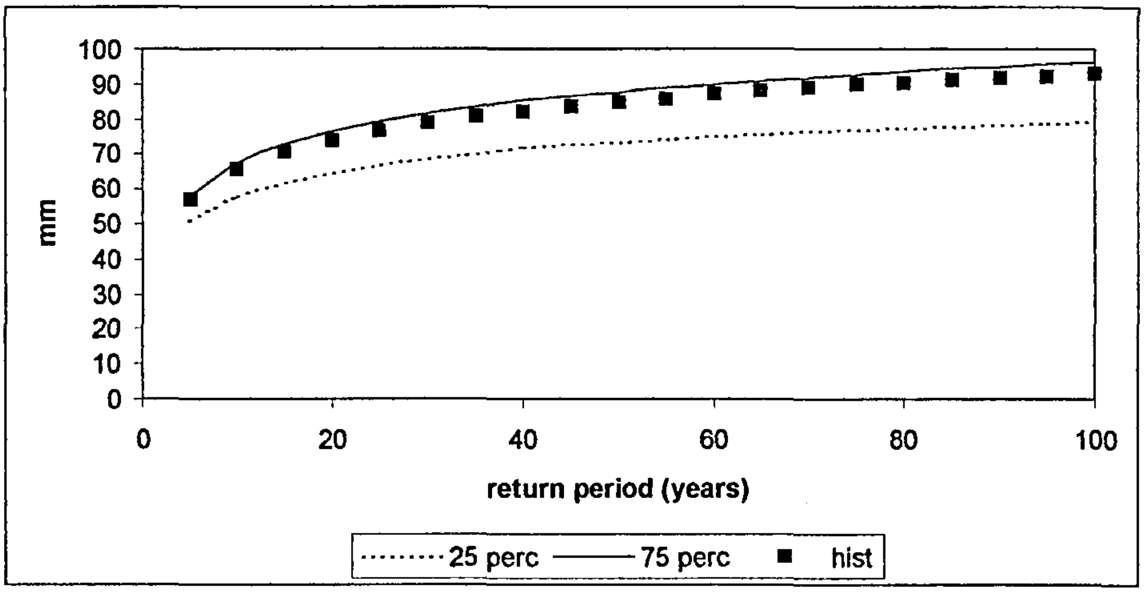


Figure 3.64. 2-day storm rainfall, using 100 30-year simulations of the 2-harmonic model.

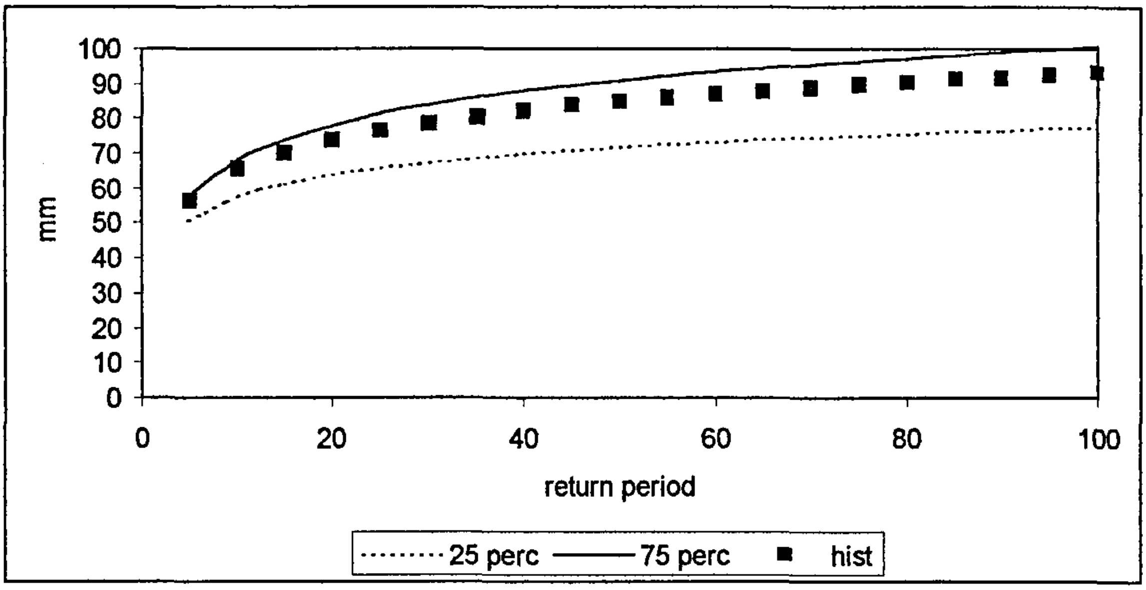


Figure 3.65. 2-day storm rainfall, using 100 30-year simulations of the 3-harmonic model.

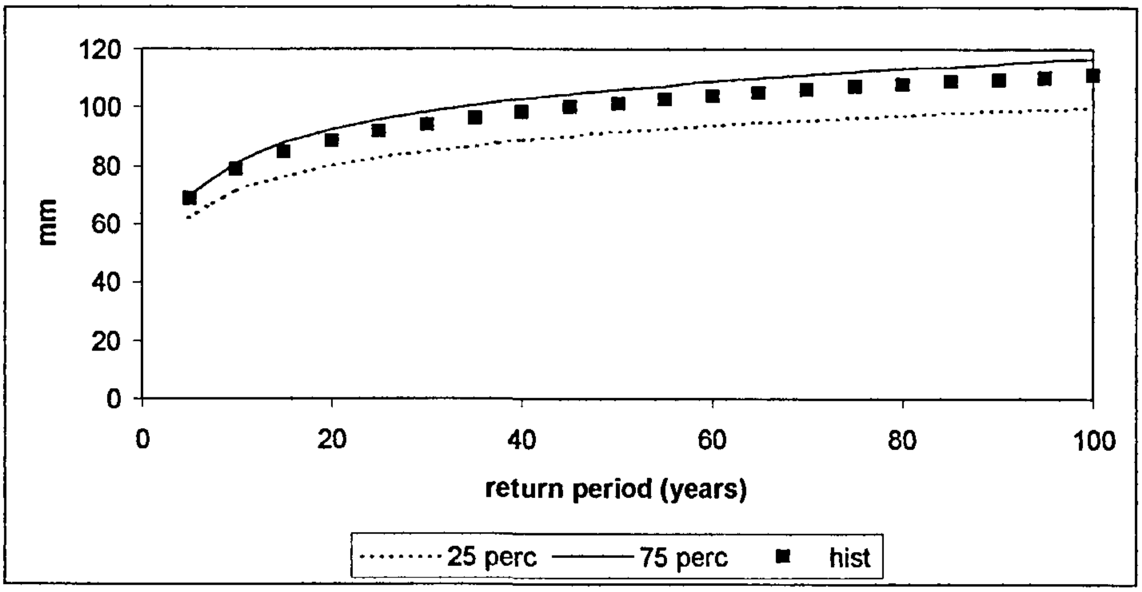


Figure 3.66. 3-day storm rainfall, using 100 30-year simulations of the 2-harmonic model.

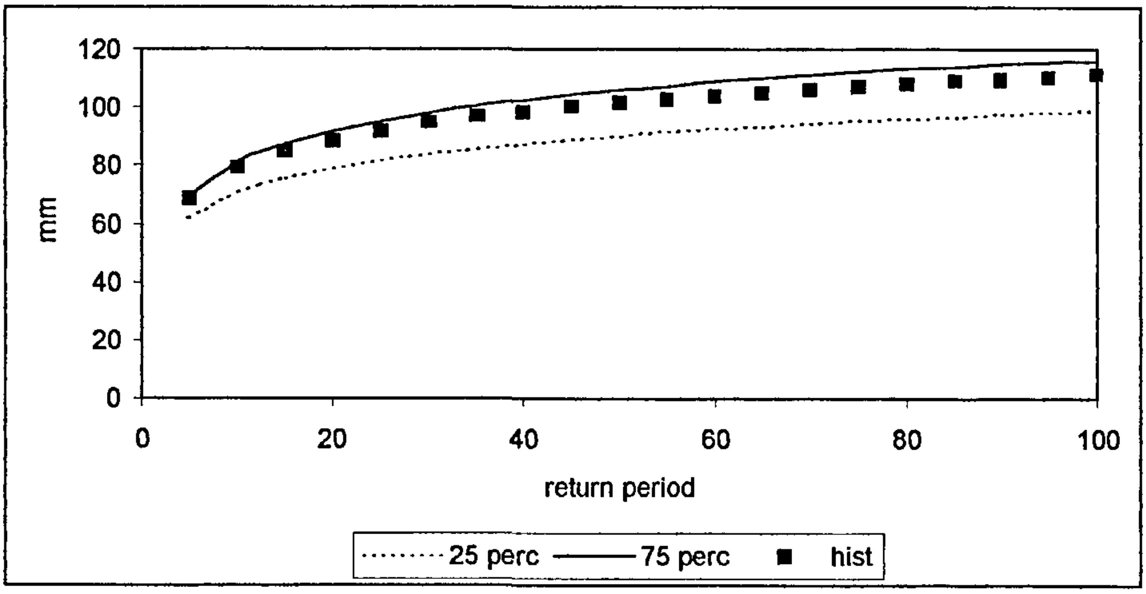


Figure 3.67. 3-day storm rainfall, using 100 30-year simulations of the 3-harmonic model.

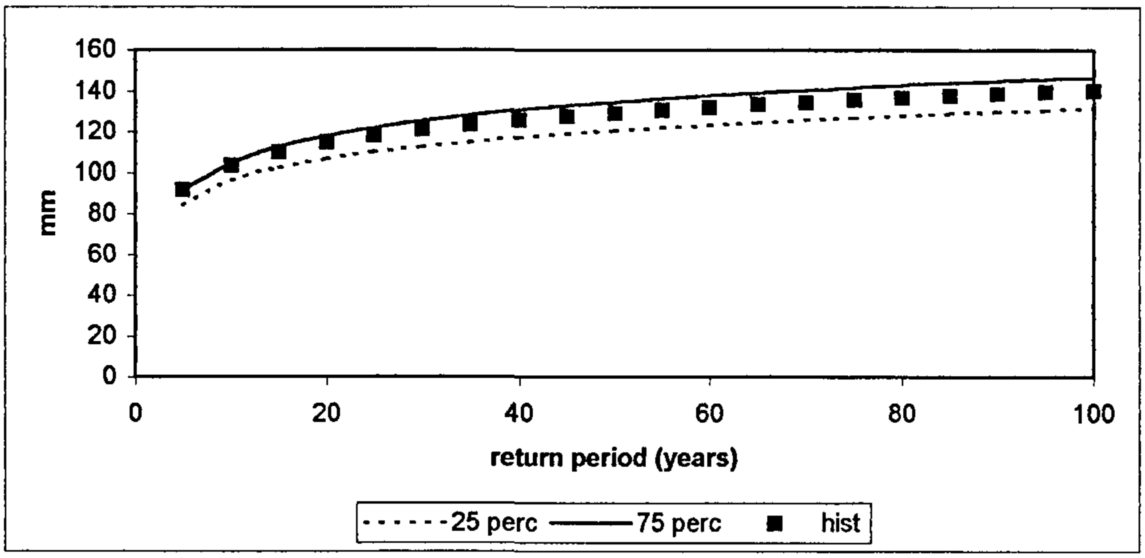


Figure 3.68. 7-day storm rainfall, using 100 30-year simulations of the 2-harmonic model.

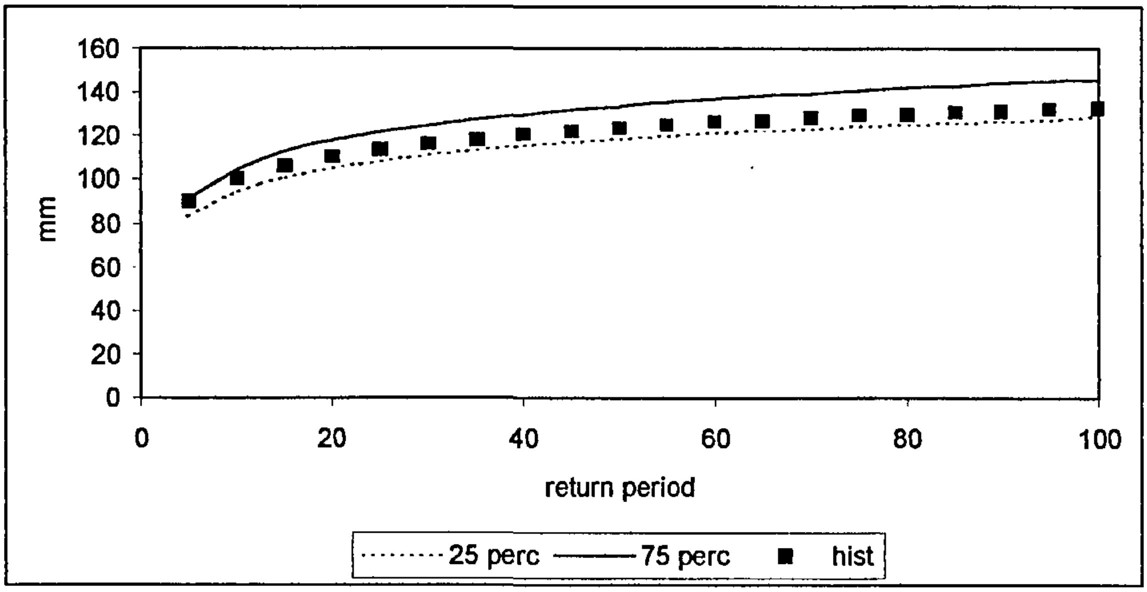


Figure 3.69. 7-day storm rainfall, using 100 30-year simulations of the 3-harmonic model.

To get down to detail, the difference between Figures 3.62 and 3.63 is that the 75th percentile of the simulated values moves up by about 2mm from the 2 harmonic model to the 3 harmonic model over much of the range of RIs and that they tend to bias the estimates downward by about 4 and 2 mm respectively. Both models yield simulation results which capture the historical values within the inter-quartile range, the 3-harmonic model the better of the two in this case.

The difference between Figures 3.64 and 3.65 for the 2-day storms is less marked, with the 3 harmonic model improving things but not by as much as in the 1 day storms. Figures 3.66 and 3.67 for the 3 day storms shows less change and less bias.

By the time we come to the 7 day storms whose behaviour is recorded in Figures 3.68 and 3.69, the improvement due to the 3 harmonic model has gone and the bias has all but disappeared completely in Figure 3.68 which shows a remarkably good fit of model to data.

In summary, the rare storms test shows that the model works extremely well. The slight downward bias of the 1 and 2 day storms, as represented by the interquartile range of simulated values in Figures 3.62 and 3.64, could be ascribed to the observation, drawn from Figure 3.58, that the model produces less rain in January than the historical sequence. Because we are modelling average rainfall over a large area (about 40 000km^2) the rare storms, even for 1 day duration, are likely to be produced by general raindays. It is just these storms that tend to be under-produced in December as is evidenced in Figure 3.14 which, combined with the minor tendency of December's general rainday intensity to be a little below the average as described in Section 3.6, will possibly bias 1 and 2 day rare storms downwards as observed above.

3.10 Assessment of the Climate/Rainfall Model.

Section 3.9 contains the results of a rigorous verification and validation program. Beside some minor differences between the behaviour of the 30-year historical sequence and the simulated sequences which have been explained (some only partially) at the end of each sub-section, the ability of the model to replicate or mimic the behaviour of the historical record in areas of importance is good.

The observations in the sub-sections attempt to provide an explanation which borders on being over refined. We do not believe that the good return we are getting from the model in its current simple state warrants much further fine tuning. It remains to test the model's ability to replicate long term drought patterns, but that test is beyond the scope of this study.

A comment is in order however; the model has a short memory climate component of a lag-one Markov chain. It has been shown that it is capable of replicating medium-term behaviour like runs of various types and rare storms of various durations, however it is unlikely to be able to mimic the tendency for wet and dry years to group into critical sequences, except by some stroke of good fortune. Short memory models are notoriously bad at replicating long term behaviour and to rectify this would possibly require that the Fourier parameters are allowed to vary randomly from year to year which would severely complicate the model.

The next stage in the modelling process is to refine the rainfall part of the climate/rainfall model, now that the climate part has been shown to work well. This next stage will concentrate on using radar images of rainfall and their models to describe the daily rainfall process in space and time. This is the subject of Chapter 4.

Chapter 4

MEASUREMENT OF RAINFALL USING RADAR AND ITS SPACE-TIME MODELLING USING FRACTAL CASCADES

4.0 INTRODUCTION.

In September 1995, one of us (Seed) travelled from New Zealand to spend 3 weeks in South Africa, so that we could work together on the space-time modelling of radar images of rainfields. During that time the WRC organized a pair of 1-day workshops:

- Characterizing spatially distributed rainfall and
- Hydrological uses of weather radar

The aims of the first workshop mentioned above were:

> ... to (i) establish, with regard to hydrology and water resources, the needs for spatial rainfall characterisation, (ii) assess the state of the art with respect to measuring systems, interpolation methods and stochastic modelling and (iii) make recommendations regarding future activities and research.

The need for spatial rainfall characterization was summarized in the following paragraphs which have been extracted from the recommendations which came from the minutes of the workshop:

> 1.2.1 <u>Needs related to the effective use of simulation models</u>
>
> A wide range of water resources planning and management decisions require the use of deterministic simulation models, which in turn require adequate and appropriate <u>spatially-related</u> rainfall information (gridded values, rainfall surfaces, rainfields) as input. This latter requirement gives rise to the following secondary needs:
>
> > (i) The need to distinguish between the different classes of spatially-related rainfall information required by different model applications, e.g. (a) historic (actual) information (for calibration/validation purposes, simulation of historic runoff time series), (b) real-time information (for real-time management decisions, forecasting) and (c) stochastically generated

information (scenario-based studies, risk analysis).

(ii) The need for methods and techniques of acquiring different classes of spatially-related rainfall information (see (i) above) at appropriate space and time scales. Excessive complexity of techniques (particularly in the case of stochastic modelling) would discourage the use of techniques and should thus be avoided.

(iii) The need to establish what the appropriate and/or cost-effective spatial and temporal scales are for various applications. Sensitivity analysis my prove of value.

(iv) The need, where derivation of spatially related information is based solely on raingauge data, to ensure an adequate number of gauges producing measurements coincident in time, especially in mountain areas (where most runoff is generated) and the more arid areas (where the spatial variation at shorter time scales can be extremely large).

1.2.2 Needs related to engineering hydrology (hydraulic design, erosion control, water quality management)

There is a need to both exploit all available data and enlarge the current data base to enable better depth-area-duration-frequency relationships to be established, especially at the smaller area/shorter duration/higher intensity range of the scale.

A better understanding of the mechanisms governing the higher intensity events is also required.

1.2.3 Needs related to widespread flood forecasting and civil engineering design

There is a need to develop a better understanding of the mechanisms and statistics of multiple-day, large-area rainfall occurrences. This implies the need for country-wide monitoring with a view to (a) generating a database of specific events to enable detailed study with a view to prediction and (b) building up time series for statistical analysis relating to the recurrence of events in space and time.

1.2.4 General needs (agricultural, hydrological, geohydrological, etc.) relating to areas unserved by rainfall measuring sites

There is an urgent need for the generation of daily rainfall time series for areas (as opposed to single sites) currently unserved by raingauges. This need is common to:

- agriculture (e.g. to enhance the spatial resolution of rainfall information to the point of comparability with that of other natural resource information for the assessment of crop production potential)

- hydrology (e.g. especially where runoff from ungauged mountain source areas needs to be quantified)

- geohydrology (for groundwater recharge assessment)
- disaster control (flood and drought risk assessment)

[This to some extent overlaps with the need expressed in 1.2.1(ii)]

1.2.5 Needs relating to research into rainfield structure, occurrence, modelling and prediction

These comprise primarily the need for sets of high resolution (space and time) observations of rainfields and for better understanding of appropriate mechanisms.

Some discussion was held on interpolation as a means of characterizing the spatial distribution of rainfall. For long time intervals, from a year possibly down to a month, smooth interpolation techniques (which use irregularly measured data from raingauge networks) such as multiquadric surfaces, are reasonable (Pegram and Pegram, 1993). However the workshop came to the following conclusion:

> The information content in current South African raingauge network data in general appears to be largely inadequate to justify generation of daily rainfall surfaces at a spatial scale as small as 1min x 1min, irrespective of the interpolation technique which may be selected from those which are currently available.

This is dramatically emphasized by Figure 4.1 which shows the total rainfall in 24 hours as measured by theMRL5 radar near Bethlehem on 11th January 1996. A raingauge network which has an average density of 100 per 40 000 square km (averaging 20km between gouges or ten gauges per side of the 200km square) would have a deal of difficulty obtaining a representative estimate of rainfall on this convective rainday. To use those few recordings to estimate the detailed spatial distribution of total rainfall on that day is farcical. (Incidentally, Figure 4.1 is one frame from a computer generated animation of the rainfall recorded by the MRL5 in January presented at the 8th SA National Hydrology Symposium held in Pretoria in November 1997, sponsored by SANCIAHS and the WRC.)

In considering the stochastic modelling of rainfields, the workshop recommendations continue:

> Stochastic modelling of daily rainfall at discrete, independent points is well-established in South Africa, with the Zucchini-Adamson models having proven their worth and utility and parameters needed to run the models being freely available for

4.1 An Outline Of How A Weather Radar Produces A Space-Time Record Of Rainfall.

Two types of weather radar are used in South Africa. The commonest are Enterprise C-band (5cm wavelengths) radars which are deployed at all the major airports and some minor ones in the country and of which there are 9 currently (1997) in operation. There is one MRL5 radar combining an S-band (10cm wavelength) and an X-band (3cm wavelength) deployed near Bethlehem in the Free State. The latter is a particularly useful tool for rainfall measurement, because it experiences less attenuation of power in its beam with distance than do the C-band radars.

Weather radars can be operated in a variety of modes. Among them are:

- vertically pointing mode, so that a radar acts as a raingauge, but gives additional information about the vertical profile of moisture in a 20km high column above the radar
- base scan mode, in which the radar scans a very flat conical surface at a fixed angle above the horizontal (typically 1° to 2° to avoid ground clutter). In this mode it completes a revolution every 10 seconds or so.
- volume scan mode, which is a generalization of base scan mode, in which the radar scans cones of successively steeper angles starting at 1.5°, incrementing in 18 increasing steps, until it reaches 55° with the last increase of 5°, when it starts again at the base scan. The extra information gained by going above the 55° angle is small compared to the time spent getting to the vertical so this is a compromise cut-off. A volume scan takes between 4 and 7 minutes to complete; the MRL5 has a cycle of 4.5 minutes.

The range of the MRL5 in base or volume scan is 144km radius, starting at 9km. The base scan at 150km will be about 5 km above the surface of the earth at the altitude of the radar because of refraction of the beam and the curvature of the earth.

In spherical co-ordinates centred on the radar the information is gathered in 600mm thick hemispherical shells (rather like an Eskimo igloo) made up of blocks, called bins, of small locations coinciding with map grid points at a spacing of 1min x 1min.

The transition from successful modelling at a point to the modelling of rainfields in both space and time is, however, a challenging undertaking which has yet to be successfully accomplished both here and overseas. The statistics which such models need to preserve at various scales are marginal distributions as well as temporal and spatial correlations.

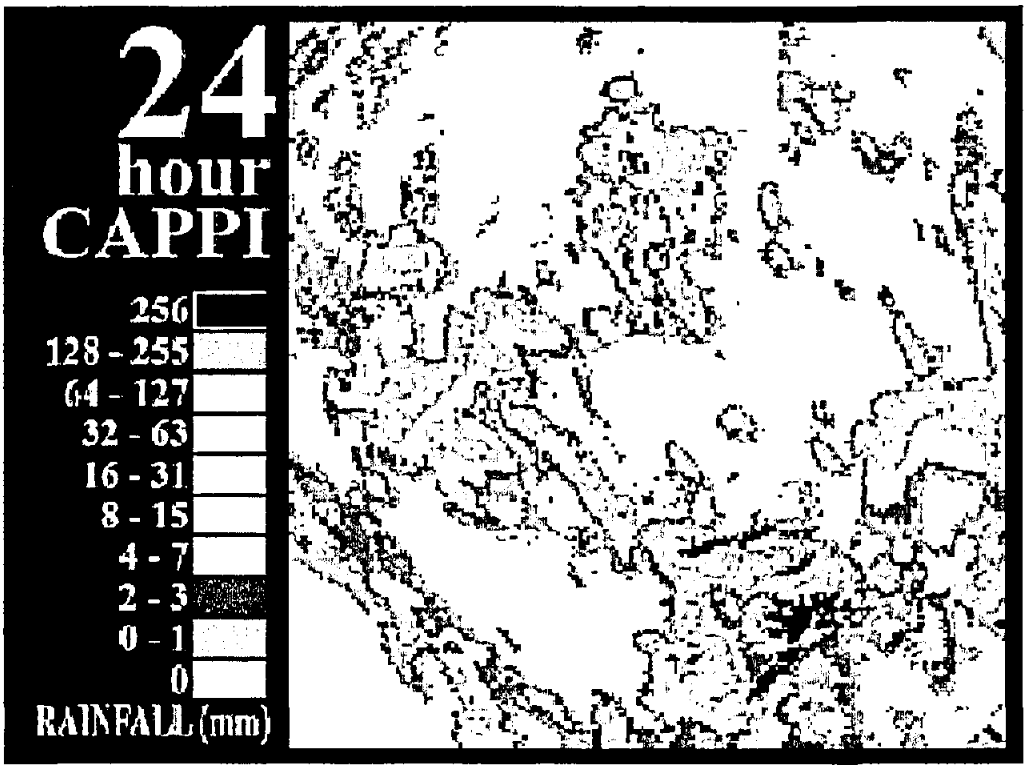


Figure 4.1 The MRL5 radar record of the total 24-hour rainfall near Bethlehem on 11th January 1997. Note the Lesotho mountains in the right hand bottom corner with high reflectivity – they are not raining at 130+ mm in a day!

We have reproduced the above extracts from the minutes of the workshop because they highlight the reasons reached by consensus of the participants as to the purposes to which any spatial rainfall model would be put, who would be likely to use one and what limitations of existing models need to be overcome in such a model.

The bulk of this present report contains discussion on deliberations attempting to address these very issues. It was with some pride that, at the workshop, we presented a summary of our work up to that time, including the 3 weeks we had spent together developing a stochastic space-time rainfield model. This chapter explains the background to that presentation.

4.1 An Outline Of How A Weather Radar Produces A Space-Time Record Of Rainfall.

Two types of weather radar are used in South Africa. The commonest are Enterprise C-band (5cm wavelengths) radars which are deployed at all the major airports and some minor ones in the country and of which there are 9 currently (1997) in operation. There is one MRL5 radar combining an S-band (10cm wavelength) and an X-band (3cm wavelength) deployed near Bethlehem in the Free State. The latter is a particularly useful tool for rainfall measurement, because it experiences less attenuation of power in its beam with distance than do the C-band radars.

Weather radars can be operated in a variety of modes. Among them are:

- vertically pointing mode, so that a radar acts as a raingauge, but gives additional information about the vertical profile of moisture in a 20km high column above the radar
- base scan mode, in which the radar scans a very flat conical surface at a fixed angle above the horizontal (typically 1° to 2° to avoid ground clutter). In this mode it completes a revolution every 10 seconds or so.
- volume scan mode, which is a generalization of base scan mode, in which the radar scans cones of successively steeper angles starting at 1.5°, incrementing in 18 increasing steps, until it reaches 55° with the last increase of 5°, when it starts again at the base scan. The extra information gained by going above the 55° angle is small compared to the time spent getting to the vertical so this is a compromise cut-off. A volume scan takes between 4 and 7 minutes to complete; the MRL5 has a cycle of 4.5 minutes.

The range of the MRL5 in base or volume scan is 144km radius, starting at 9km. The base scan at 150km will be about 5 km above the surface of the earth at the altitude of the radar because of refraction of the beam and the curvature of the earth.

In spherical co-ordinates centred on the radar the information is gathered in 600mm thick hemispherical shells (rather like an Eskimo igloo) made up of blocks, called bins, of small solid angle of 1.5°. These bins will be quite narrow near the radar but get to be about 3.75km across at 150km radius. The average reflectivity in each bin is then converted by mathematical transformation to a variable Z which is related to the rainfall rate R in the bin, which is itself related to the drop size distribution of the moisture in the bin.

The conversion from Z to R is via the Marshall-Palmer relationship $Z = 200R^{1.6}$. The basic variable worked with is dbZ (which is equal to 10 $\log_{10}Z$) and this will frequently be the variable used to map a radar's view of rainfall. To get a two-dimensional plan of rain falling on the earth's surface the dbZ in the spherically centred bins obtained in a volume scan are sampled at a constant height above ground - typically 2 km as a compromise - and presented as a Cartesian (rectangular) grid of 1km sided squares. This presentation is called a CAPPI (Constant Altitude Plan Position Indicator) and the variables in each square will be in steps of dbZ, or rain rate such as mm/hr.

A CAPPI of a large storm obtained from the MRL5 radar at Bethlehem is shown in grey scale in Figure 4.2.

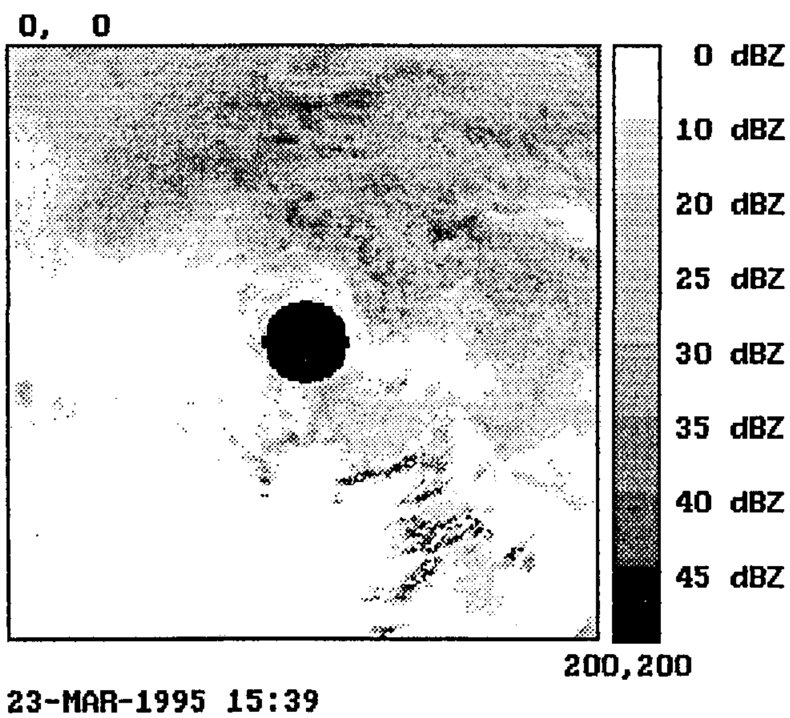


Figure 4.2 A cappi of a storm at Bethlehem on a 200 kilometre square. Ignore the ground clutter at the bottom right hand corner – it comes from the reflection of the Maluti mountains in Lesotho

The CAPPIs obtained from the MRL5 volume scan data are thus effectively snap shots of the rain falling through the 2km level above the ground and are separated in time by about 4.5 minutes, with one difference. The reflectivity of the inner ring of squares on one CAPPI is obtained a few seconds before the outer ring on the following CAPPI, so what is observed is a saw-tooth in time depending on radius from one CAPPI to the next. However in most storms (except for the fast moving ones) the changes over 4 minutes are reasonably gradual when measured to a 1km square resolution, so the CAPPIs obtained from volume scans are often more than adequate for hydrological purposes.

Figure 4.2 shows a CAPPI which measures 200km on each side. There are thus 40 000 pieces of data in the square. This amount of data arrives every 4.5 minutes, so handling it requires some special statistical techniques and fast computers.

The remainder of this chapter addresses the problem of representation and modelling of the large data sets that sequences of CAPPIs provide. The intention is to summarize the information in a model defined by a small number of descriptive statistics so that the CAPPIs can be reconstructed or mimicked with computational facility.

4.2 ONE DIMENSIONAL REPRESENTATION OF RAINFALL.

The record of rainfall at a point in space (like a raingauge or a CAPPI square) is very similar to the record at one point in time sampled across a CAPPI in one direction or another, as can be seen from Figure 4.3.

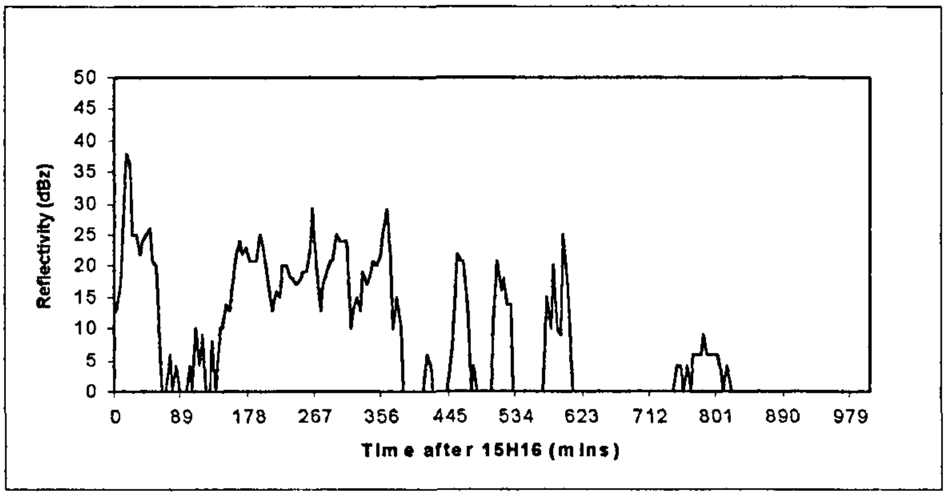


Figure 4.3a The 16hr record at a selected **pixel** as a function of time. The event was the storm of March 23 1995 at Bethlehem as recorded by the MRL5. The scale is dBZ, not rain-rate.
Rain-rate, R = [{10^(dBZ/10)}/200]^0.625, so 30dBZ = 2.73mm/h

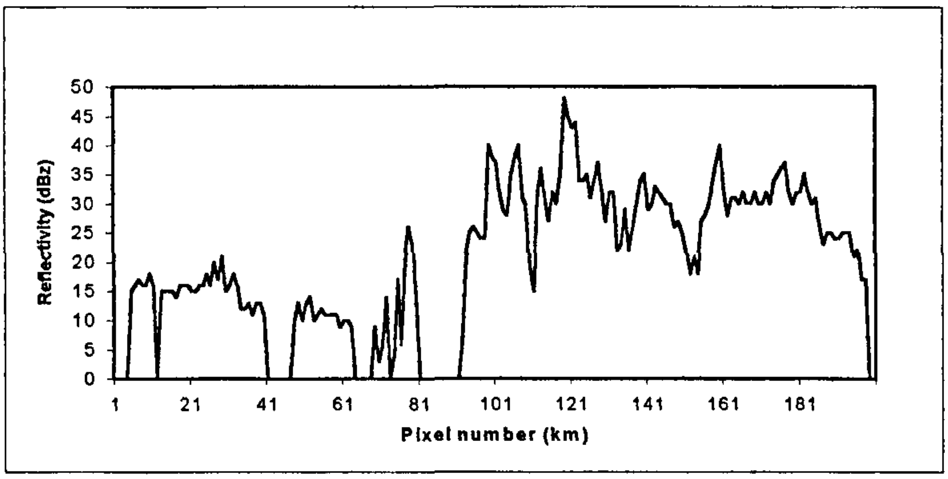


Figure 4.3b The 200km record across a selected **CAPPI** as a function of distance. The event was the storm of March 23 1995 at Bethlehem as recorded by the MRL5. The scale is dBZ, not rain-rate.
Rain-rate, R = [{10^(dBZ/10)}/200]^0.625, so 40dBZ = 11.53mm/h

These traces are not dissimilar; they were selected so that there was a reasonable correspondence between the amount of dry record and rain-rate. We are not suggesting that these records are mathematically similar – for that we have to be more precise in our definitions, which will follow.

So-called similarity depends on the size of the sampling interval relative to the total length of the sample. The concept is called self-similarity and means (by way of example) that if the rainfall at a point is accumulated in 10-second intervals for an hour giving 360 readings of rainfall depth, the graph of the time series presented in histogram form will look the same (and have the same statistics) as the graph of 3 hours of rainfall sampled in 360 half-minute totals. Of course this is only true if the 3 hour storm is statistically homogeneous in time and lasts that long, but this is approximately the situation in a large enough number of cases to make the idea useful. Fortunately the concepts carry over easily from 1 to 2 or more dimensions so it is convenient to remain in 1 dimension for descriptive and definition purposes.

There are two things which describe the one-dimensional sequences shown in Figure 4.3. The first is the (marginal) probability distribution of the amount of rain falling in any given interval. The second is the covariance structure relating the amounts of rain in adjacent and more distant intervals.

An equivalent technique for estimating the marginal distribution of the amounts is to evaluate the non-central moments of many orders which together uniquely define the marginal distribution. An equivalent way of interpreting the covariance structure is via its Fourier transform, the power spectrum, which is often used to decide whether a sampled stochastic process exhibits self-similarity. The moments and the spectrum are commonly used in this work because they are so convenient to graph.

A thorough description and derivation of the relationships we use here is given in a paper by Menabde et al (1997b) and we will make use of a sub-set of their notation, giving limited definitions of the variables to convey the ideas with the minimum of complexity, in order to make them practically useful.

4.2.1 The measurements - the sampled data

Assume that rainfall has been observed at a point over a period T, which may be from one to several hours. Following the ideas of Menabde et al. (1997b) the basic observations are conventionally made as accumulative volumes by a raingauge or as intensity measurements by a radar. These observations are made at intervals of fixed duration t_0, whose position from the beginning of the sample is given by the time co-ordinate τ, where $t_0 < \tau < T$ if we are recording at the epochs between intervals. There will thus be $N = T/t_0$ basic observations in the sample.

If we are only observing intensity (not accumulated volume) at the epochs between each interval, t_0 must be small enough to ensure that the intensity is representative of that in the interval (i.e. has an approximately linear variation from one epoch to the next) so that the intensity measurements and accumulated volume measurements would yield essentially the same information. If t_0 is too long, essential information will not be captured by an intensity measuring device (like a radar); less information will be lost by a mass accumulation instrument like a raingauge of the tipping-bucket or drop-counting type.

4.2.2 The spectrum

The power spectrum of the sample of data is the Fourier transform of the auto-correlation function and as such embodies all the time-series structure of the sample. Examining the power spectra of rainfall recorded in very short intervals (15 seconds or so) over some hours, it is evident that many of the homogeneous records exhibit spectra with a very particular structure.

This special structure is, that the average power at a given frequency is inversely proportional to a constant power of the frequency, showing an exponential die-off of power with an increase in frequency.

Mathematically,

$$p(f) \sim f^{-\beta} \tag{4.1}$$

where $p(f)$ is the power in the spectrum at a given frequency f and β is a constant exponent.

This behaviour has been observed in homogeneous fluid turbulence; fields exhibiting this property are said to be "scaling". Many observed rainfall events exhibit a value of β very nearly constant across the spectrum (as long as they do not contain too many breaks of appreciable length) and in addition, most observed values of β are greater than unity. A few exhibit β less than unity. (In D-dimensional space, the criterion is $\beta \gtrless$ D.)

It turns out that the value of β is crucial in deciding which model to adopt to explain the rainfall process in one or more dimensions as will be shown in the next few paragraphs in the discussion on moments.

4.2.3 The Moments

As explained in the earlier part of this section, the moments (either central about the mean, or non-central about the origin) of the values in the observed intervals are enough to characterize the marginal probability distribution of these values. It is necessary to introduce some notation here for precise development of the ideas.

Given that there are $N = T/t_0$ basic measurements in the sample of length T, we can aggregate adjacent observations into longer intervals t, which are multiples of t_0, such that $t/t_0 = n = 1, 2, \dots, N$, with n usually chosen to be incrementing in integer powers of 2 for convenience.

Thus we define the jth observation (for any reasonably large number of intervals n) in the interval of length t as $\mathrm{R}_t(j), j = 1, 2, 3, \dots, n$. The qth sample moment of the ensemble of n observations is then defined as:

$$< \mathrm{R}_t(j)^q > \; = n^{-1} \sum_{j=1}^{n} \mathrm{R}_t(j)^q$$

where it is convenient to use the notation <...> to denote ensemble average.

When β, the negative slope of the power spectrum introduced above, is less than unity, it has been observed that

$$< R_t(j)^q > \sim (t/T)^{-K(q)} \quad . \tag{4.2}$$

In words, the qth moment is inversely proportional to the size of the aggregation interval t raised to some constant value -K(q) which itself depends on q. Thus if we were to plot log$<R_t(j)^q>$ versus log(t), the slope of the line will appear as a constant equal to -K(q) if the basic process is what is called "multi-scaling". (The process would be called "scaling" if K was a linear function of q, but this is not observed in rainfall processes.) A nice way to represent the characteristics of multi-scaling (and to decide if a particular sample has this property) is to plot K(q) versus q. (We show examples of these calculations in Section 4.3.)

The idea of multi-scaling was predicated on the assumption that β < D. This is seldom observed in rainfall samples. Usually β > 1 (for 1-dimensional time-series), in which case the nice constancy of K(q) expressed by equation (4.2) does not hold. What has to be done in those circumstances is to form a new series $\Delta R(\tau)$ by taking the absolute values of the small scale differences (gradients) of the basic process, i.e.

$$\Delta R(\tau) = |R_{t_0}(\tau + t_0) - R_{t_0}(\tau)| \tag{4.3}$$

It turns out that for many samples which exhibit β > 1 the moments of $\Delta R(\tau)$ scale in the sense that

$$< |R_{t_0}(\tau + t_0) - R_{t_0}(\tau)|^q > \sim (t/T)^{-K(q)} \tag{4.4}$$

A field that exhibits this property is called quasi self-similar, while the field derived from it using equation 4.3 and those fields with β < 1 and satisfying equation (4.2) are called self-similar (Menabde et al, 1997b). It is a further property of self-similar random fields that β is related to K(2) for a D-dimensional field by the simple relationship

$$\beta = D - K(2) \tag{4.5}$$

K(2) can be shown to be positive for any self-similar random field, hence β is less than D in all such fields. Thus when a field is observed which has β > D, to use the theory of self-similar fields to analyse it, requires the pre-treatment indicated by equation (4.3) or some other transformation (such as power-law filtering) to convert the quasi self-similar field into a self-similar field. Thus the steps to take in the analysis are to compute β first and then (after possible transformation) to compute K(q).

4.2.4 A discrete random multiplicative cascade model - the α-model.

A useful model of a self-similar random field is the random multiplication cascade generating a sequence of random variables on discrete intervals. This is best introduced in 1 dimension (D = 1) and can easily be generalized to higher dimensions. The model is variously called: a simple cascade model, a canonical random cascade (Mandelbrot, 1974) or (when it takes on a special form) an α-model after Schertzer and Lovejoy (1987). The simplest formulation whereby to grasp the concepts is via the α-model and that is the one we will use to introduce the ideas and will also use in Section 4.4 where we perform computations and simulations.

The description we give is a gently edited passage taken from Harris et al. (1996) who give a clear and concise description and definition of the cascade.

> Consider a homogeneous distribution of rainfall in one-dimension (i.e., a time series; extensions to higher dimensions carry over in a simple way) with rain rate R_0 representing the average rain rate (and is thus, in effect, simply a normalization constant) over the time interval T. On the first step we divide the interval into two halves and assign each of them a rain rate $R_1=R_0W(1)$ and $R_2=R_0W(2)$, respectively. Note that for simplicity each interval here is split into two halves (i.e., branching number of 2); however, one could also divide each interval into quarters for a 2-dimensional cascade or eighths for a 3-dimensional cascade or any other fraction. The randomly chosen weights, $W(i)$, are independently and identically distributed (iid) random variables produced by a generator satisfying the condition that $<W> = 1$. This condition is often referred to as canonical conservation where the mass in each step is conserved only on the average. The more restrictive microcanonical cascade is the case when cascades are constructed with exact conservation of mass at each step. Such cascades lead to much calmer behavior *Mandelbrot* [1974]. We only consider canonical cascades in this report. This procedure is repeated n times, dividing each resulting interval into smaller and smaller increments, at the end of which there are 2^n subintervals with rain rate
>
> $$R_n(i_1, I_2, \ldots, i_n) = R_0 W(i_1) W(i_1, i_2) \quad W(i_1, i_2, \ldots, i_n) \qquad (4.6)$$
>
> where the set of binary indices, $i_1, \ldots, i_n$ indicates 2^n possible realizations of the random field. Raising equation (4.6) to power q and taking the ensemble average of both sides we get:
>
> $$\left\langle R_n^q \right\rangle = R_0^q (\left\langle W^q \right\rangle)^n = R_0^q 2^{n\log_2 \left\langle W^q \right\rangle}. \qquad (4.7)$$

Since $2^n=(t/T_0)^{-1}$, one can substitute this relation for 2^n into the left side of (4.7) to obtain the scaling moments relation:

$$\langle R_n^q \rangle \sim (t/T_0)^{-K(q)}, \qquad t << T_0$$

with

$$K(q) = log_2\langle W^q \rangle. \tag{4.8}$$

The last equation (4.8) gives a useful way of characterizing the cascade from real data, especially if one has a given distribution of W in mind described by a few parameters, an example of which is the α-model.

In the α-model, W can take on one of only two values at each step and in each sub-interval, with the following probabilities:

$$P[W=a]=p, \qquad P[W=b]=1-p$$

Because we want a canonical cascade,

$$E[W]=ap+b(1-p)=1$$

which constrains b to $(1-ap)/(1-p)$.

If we use α in place of a in the above development, we have the α-model, described by two parameters α and p.

The second moment of W is:

$$E[W^2]=p\,\alpha^2+(1-p)[(1-\alpha p)/(1-p)]^2$$
$$=p\,\alpha^2+(1-\alpha p)^2/(1-p)$$

and in general

$$E[W^q]=p^{q-1}\,\alpha^q+(1-\alpha p)^q/(1-p)]^{q-1} \tag{4.9}$$

which holds for all $q>0$.

Equations (4.8) and (4.9) provide a link between the observed exponents K(q) and α and p if we match population and sample moments, i.e.

$$K(q)=\log_2[p^{q-1}\alpha^q+(1-\alpha p)^q/(1-p)]^{q-1}] \tag{4.10}$$

Equation (4.10) is transcendental so one needs to resort to numerical methods to extract the values of α and p for a finite set of K(q) values. Because there are only two parameters in the alpha model we need only two values of K(q).

K(1) is identically zero always, so in practical computations it is common to evaluate K(2) and K(3), corresponding to the second and third moments. To use higher moments would be to invite numerical difficulties - moments higher than the third are notoriously imprecisely estimated except for a very large number of observations from a homogeneous sample.

4.2.5 Summary

Given a D-dimensional realization of a stochastic sequence of rainfall (in time or space) the steps in analysis before modelling the process with a cascade are:

- estimate the negative slope β of the log of the power spectrum of the data: if it is constant proceed because the sequence is scaling, if not, search for another sample
- if $\beta < D$ proceed directly to moment estimation;
  if $\beta > D$ form a new sequence of derived variables by taking the absolute local gradients of the original series, then proceed to estimating the moments
- estimate the moments $\langle R_t(j)^q \rangle$ from the data for various aggregations over intervals of length t; plot $\log\langle R_t(j)^q \rangle$ against $\log(t/T)$; the lines for each q will have constant slope if the process is self-similar (or multiscaling)
- extract the K(q) versus q function for the set of moments; in particular get K(2) and K(3)
- find α and p for the α-model of the canonical cascade to be used in simulations by a two-dimensional root-finding algorithm such as Newton-Raphson employing equation (4.10)

In the next section we report some experiments conducted on CAPPI data from the MRL5 radar at Bethlehem.

4.3 MODELLING RADAR CAPPI DATA

During an intense 3-week period of collaboration during September 1995, we examined some of the C-band radar data from Bethlehem and found that it was not good enough for analysis. The CAPPIs had many shadows and holes in them. With the help of the Weather Bureau personnel at Bethlehem (in particular Deon Terblanche and Karel de Waal) we obtained some (then) recent data from March 1995. In particular we started to analyse the data on March 23 which described a 7 hour storm, lasting from 15:16 until 1:00 the following morning, consisting of 79 CAPPIs. We appended a later storm on March 25 to bring the set up to 256 CAPPIs which we proceeded to analyse.

To obtain K(2) and K(3) from a 2-dimensional CAPPI is a direct extension of equation (4.2) where the ensemble average is taken over all the points. (In this preliminary analysis we ignored the step of calculating β, we just assumed that the field would be scaling. This was a deliberate omission to speed things up; in future work we would have to go the rigorous route summarized at the end of section (4.2).)

The plot of the 256 sampled K(q) versus q functions is shown in Figure 4.4

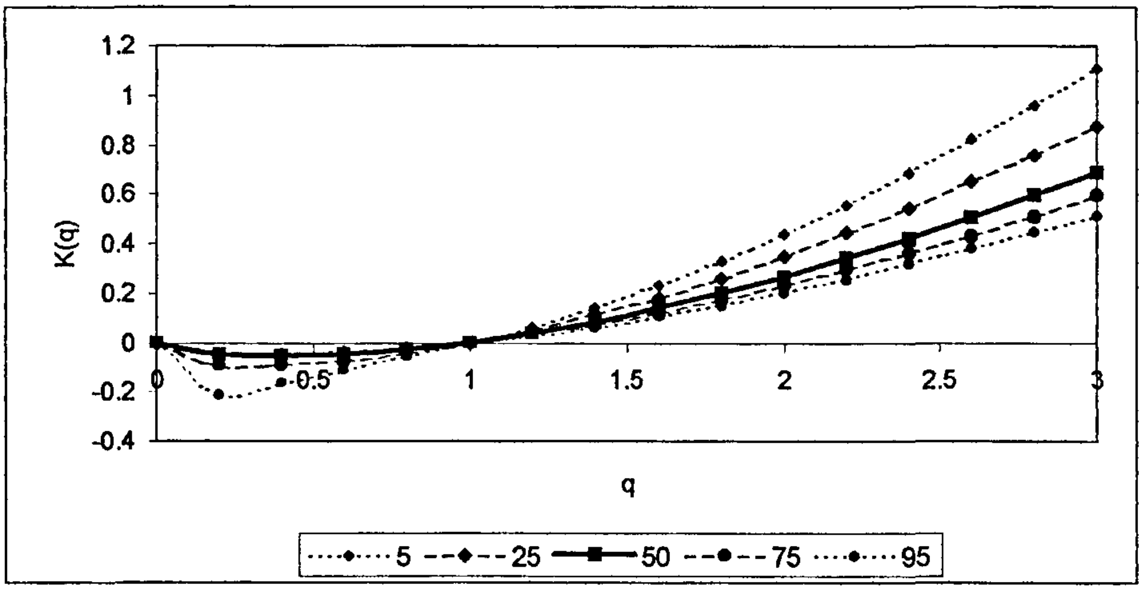


Figure 4.4 The $K(q)$ versus q functions of 256 CAPPIs from the storm of March 23 1995, plotted co-axially – only the 5, 25, 50, 75 & 95 percentiles are shown

The minimum, median and maximum of the values of K(2) and K(3) are given in Table (4.1) where, in addition, the corresponding values of the parameters of the alpha model are also given as derived by a Newton-Raphson gradient search routine based on equation (4.10).

Table 4.1

Values of K(2), K(3), α and p
derived from the CAPPI data of the storms of March 1995

	K(2)	K(3)	α	p
minimum	0.169	0.416	0.625	0.437
median	0.266	0.688	1.610	0.337
maximum	0.505	1.319	1.968	0.309

It turns out that there is an ambiguity in the values given by the search routine. In attempting to get the parameters α and p for a given K(2) and K(3) it is equally probable that one will derive the other pair $(1 - \alpha p)/(1 - p)$ and $(1 - p)$. It is important to constrain the variable α either in the range (0, 0.5) or in the range (0.5, 1) otherwise it hops from one interval to the other. That was done in Table 4.1. It was observed from examining the 256 (α, p) pairs that there is a very strong correlation between them.

We think that these pairs may be helpful in characterizing the type of a given storm - i.e. whether it is convective or stratiform or a mixture, helping to tie into the rainday classification introduced into Chapter 2 and developed in Chapter 3. Further development of this idea is deferred to the future.

It remains to give an example of a simulation which follows in Section 4.4.

4.4 SIMULATION OF A PART OF A CAPPI.

In September 1995 we did not have as much computer power available to us as we now have two years later, so we restricted ourselves to generating a sequence of 64 CAPPIs on a 64 by 64 square grid of pixels. This was done in a 3-D cube, using the parameters $(\alpha, p) = (0.5, 0.5)$ and smoothing the values a little using a Laplacian smoothing molecule. The details are really unimportant at this stage; suffice is to say that we thought that the simulations show considerable promise - these were the ones presented at the WRC workshop in September 1995.

In Figure 4.5 a, b, c and d are shown two sets of adjacent (in time) pairs of simulated CAPPIs selected from the set generated using the simple α-model in 3 dimensions (the third dimension being time) using the 3-D cube.

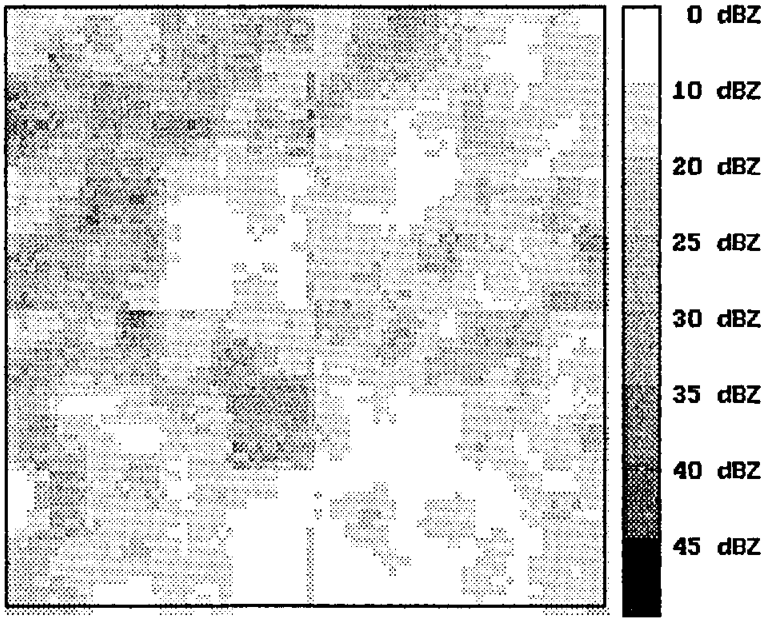


Figure 4.5a Frame 1 of simulation Figure 4.5 consists of 4 figures taken from a sequence of 64 frames of a cascade simulation on a 64 by 64 square using (0.5, 0.5) as the parameter set, 8 levels in the cascade and 2 smoothings. Frames 3 & 4 are contiguous time frames, 40 frames later than Frames 1 & 2

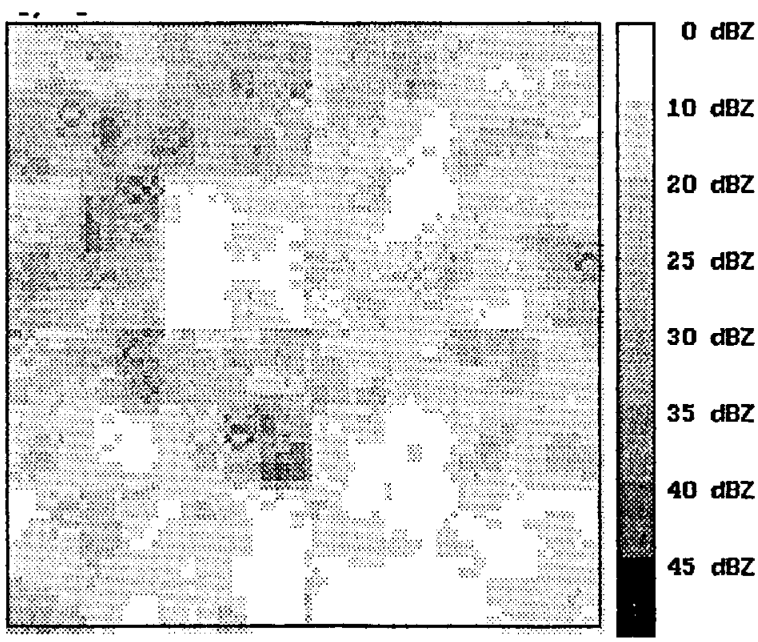


Figure 4.5b Frame 2 of simulation

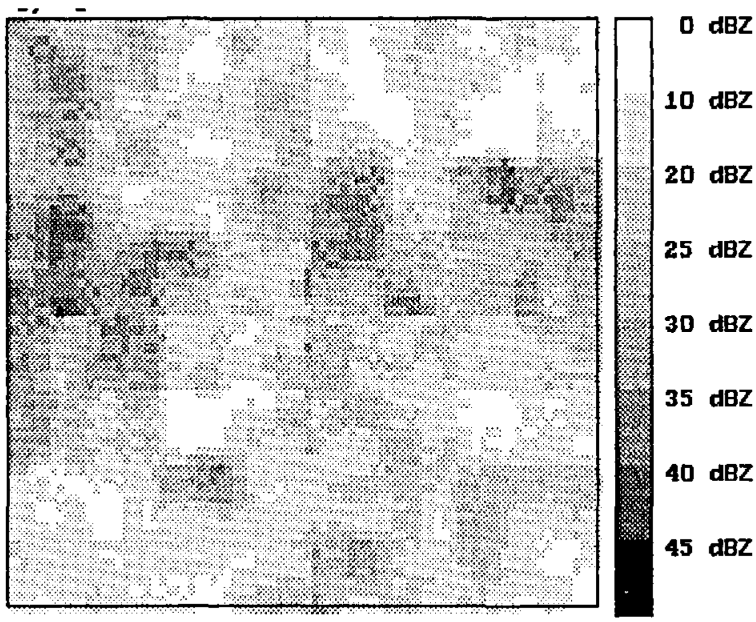


Figure 4.5c Frame 3 of simulation

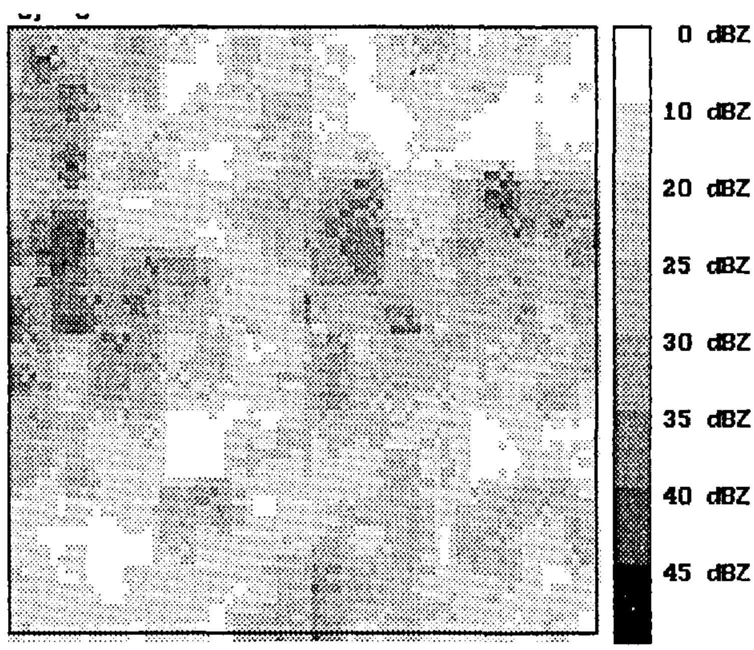


Figure 4.5d Frame 4 of simulation

Figure 4.6 shows a 64 by 64 square taken from the upper right hand corner of the full 200 by 200 CAPPI of March 23 1995 at 15:39 in Figure 4.2.

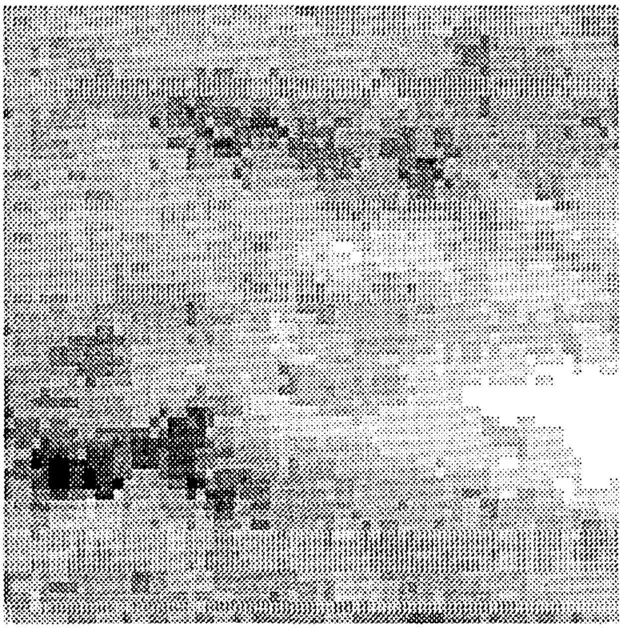


Figure 4.6 Frame 5 A 64 by 64 kilometre square cappi, being the portion of a CAPPI of real rain from the top right-hand corner of the full 200 by 200km CAPPI shown in Figure 4.2 of a storm over the Bethlehem region on 23rd March 1995.

The simulated CAPPIs of Figure 4.5 show some similarity to the measured one in Figure 4.6 in terms of the clustering behaviour and variability in intensity. What is a feature of the simulations is the "blockiness" or "squareness" of the simulated images. This artefact comes from the cascade and is something we wish to deal with, but as yet are not sure how.

However a pair of CAPPIs lifted (with permission) from the paper by Menabde et al. (1997a) using a new variation of the α-model, shows promise at removing this artefact. The CAPPIs shown in Figure 4.7 come from that paper.

The weather radar was configured so as to measure the rainfall out to a maximum range of 15 km. These data were mapped onto a 128 square Cartesian grid with 120 m resolution pixels. Figure 4.7a shows the measured data and Figure 4.7b shows the simulated CAPPI with the same statistics as the measured.

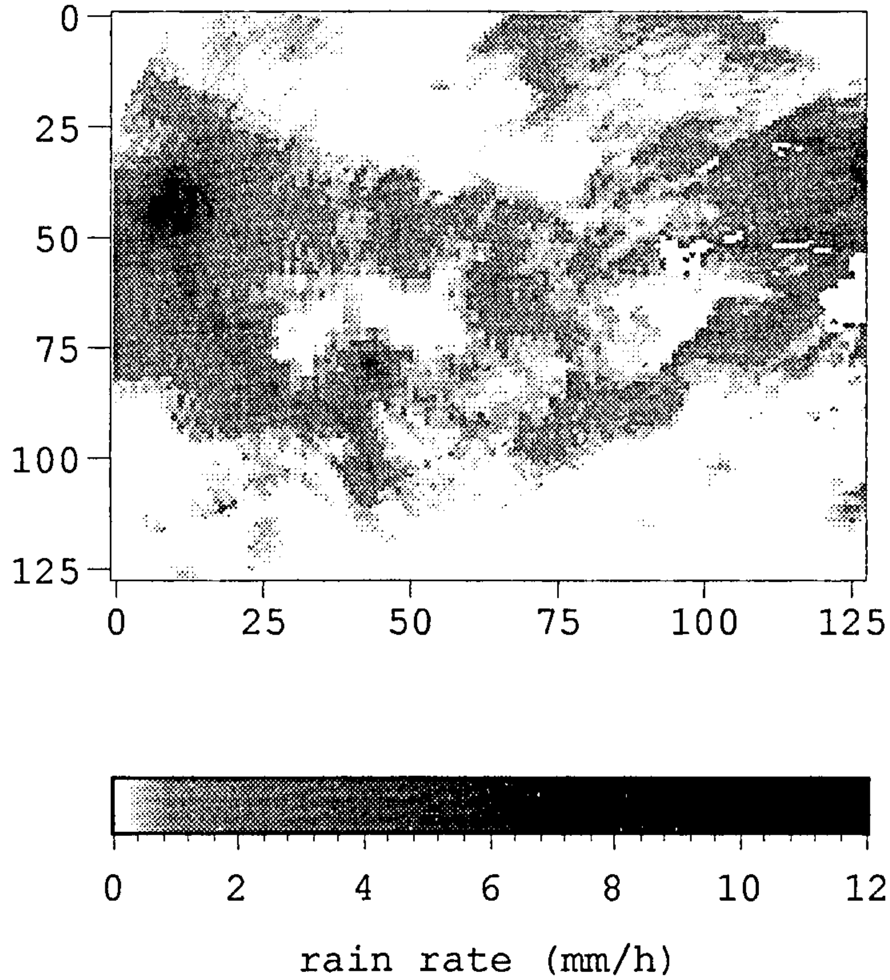
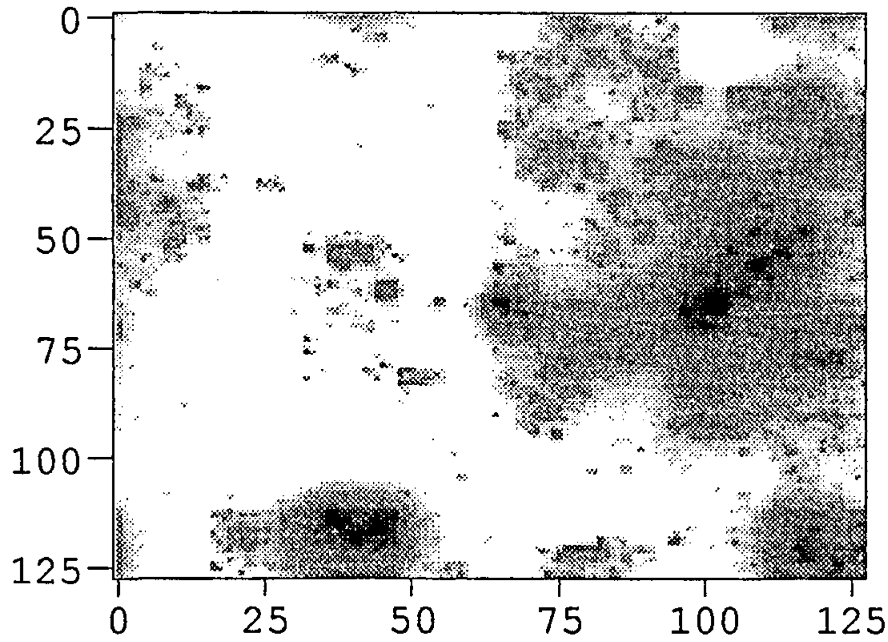


Figure 4.7a. Measured data at a site in South Island, New Zealand. The CAPPI shows a 128 square grid with resolution of 120 m on a pixel side. (from Menabde et al., 1996, with permission.)



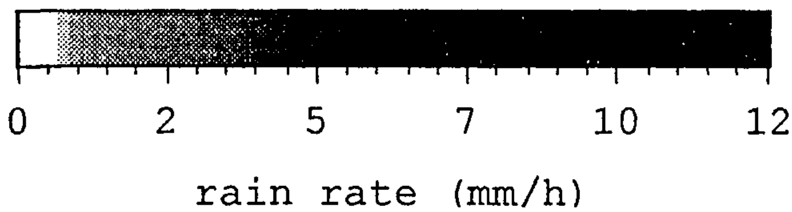


Figure 4.7b. Simulated data with the same statistics as the measured data in Figure 4.7a. The CAPPI shows a 128 square grid with resolution of 120 m on a pixel side. (from Menabde et al., 1997a, with permission.)

These two images give promise for the usage of fractal cascades in the modelling of spatial rainfall processes.

Chapter 5

CONCLUSIONS AND PROGNOSIS.

We conclude that, in general, the temporal aspect of a space-time modelling of rainfall using a climate component has been successful. There are minor things that may need attention, but it has been shown in Chapter 3 that it is possible to model the temporal processes well.

The rainfall amounts modelled by the space-time model introduced in Chapter 3 are by and large fairly well modelled with minor exceptions. The combined model, as it stands, could be applied, fitted and used with some extra polishing. What this report shows is that the relatively simple approach that we adopted pays off.

When it comes to modelling the true spatial character of rainfall, we now think that we are very unlikely to be able to use raingauge data as a basis for determining the scaling behaviour etc, but we have not yet checked this out properly. The raingauge networks are simply too sparse, so recourse has to be made to radar imagery which is getting increasingly more competent at measuring rainfall accurately. Even if it does not always match raingauge data exactly (some may argue the case as to which method yields a record which is more truly representative of the rainfall measured at a point!), enough has been done to show that when integrated over an area such as a quaternary catchment, the information yielded far exceeds that coming from any existing raingauge network for short time steps shorter than a day.

There are relatively few years of good radar data over-lapping the raingauge data collected from the network which lies under the gaze of the best weather radar in the country sited at Bethlehem. What has been shown, using modern analysis techniques for processing the very large streams of data coming from radars introduced in Chapter 4, is that it is feasible to continue the effort to model the spatial distribution of rainfall using models based on self-similarity or multifractal behaviour. From an operational point of view, the cascade modelling is more likely to start with the instantaneous rainfields and then move on through accumulations to daily totals; there may be problems with advection of the field – we are unsure what happens when we take out the mean advection and try the cascade modelling on the stationary fields.

We are mindful of the fact that the space/time link is still unresolved: when we "dimensionalise" the cascade to represent rainfall over say 1 km, what time step does the time dimension of the cascade represent? However, these are technical niceties and we are reasonably comfortable that the main task of getting a temporal and spatial model is not beyond our grasp.

The future research effort will be directed at :

- obtaining an improvement to the modelling of individual CAPPIs and removing the "squareness" evident in the simulations
- deriving an extension to the 3-D cascade cube to model CAPPIs in space and time; this will require matching the parameters to ensure consistency
- ascertaining what the outer and inner limits to the cascade process are. Based on new concurrent 100km and 10km radar data collected in New Zealand, we think that the likely limits are from 100km down to 2-5 km for quasi self-similarity at which stage the process becomes self-similar down to 100m scales. This will be important for urban, rather than rural catchments; the latter have much larger areas. Nevertheless, we think that there is a need for some high resolution work in South Africa to assist with these questions.
- tying in the characterization of rainfall measured in CAPPIs to the climate or rainday type
- combining the climate and cascade models to give a daily rainfall model in the first instance, followed by models with shorter time scales

Based on the observations and conclusions coming from this report, we think these are feasible and achievable objectives.

---oooOooo---

REFERENCES

Basson MS, Allen RB, Pegram GGS and van Rooyen JA, (1994).
Probabilistic Management of Water Resource and Hydropower Systems. Water Resources Publications, Colorado

Chow VT, Maidment LR and Mays LW (1988).
Applied Hydrology. McGraw Hill

Court AP, (1979).
Rainfall characteristics of classifications systems used by the BEWMEX project. *Bethlehem Weather Modification Experiment, Progress Report No 14.* Weather Bureau, Private Bag X97, Pretoria, 0001.

Garstang M and Emmitt GD, (1986).
A description of PAWS study area rainfall. In: in Programme for Atmospheric Water Supply Phase I 1983-1986, Volume 2: Natural clouds and rain in Nelspruit, Simpson Weather Associates and Cansas International Corporation. *Water Research Commission Report No 133/2/88,* P O Box 824 Pretoria, South Africa.

Harris D, Menabde M, Seed A and Austin G (1996).
Multifractal characterization of rain fields with a strong orographic influence. *J. Geophysical Research.* Vol 101, pp 26405-26414.

Hudak DR and Steyn PCL (1980)
Forecasting in the BEWMEX area. *Bethlehem Weather Modification Progress Report No 19.* Weather Bureau, Private Bag X97, Pretoria.

Mandelbrot B (1974).
Intermittent turbulence in self-similar cascades: Divergence of high moments and dimension of the carrier. *J. Fluid Mechanics.* Vol 62, No 2, pp 331-358.

Menabde M, Harris D, Seed A and Austin G, (1997a).
Self-similar random fields and rainfall simulation. *J. Geophysical Research.* Vol 102, (D12), p 13509.

Menabde M, Harris D, Seed A, Austin G and Stow D, (1997b).
Multiscaling properties of rainfall and bounded random cascades. *Water Resources Research.* (in press)

Nelder JA and Mead R, (1965).
A simplex method for function minimization. *Comp. J.* Vol 7, pp 308-313.

Pegram GC and Pegram GGS (1993).
Estimating areal rainfall by integration of multi-quadric surfaces over polygons. *J. Hydraulic Eng., American Soc Civil Eng.* Vol 119, pp 151-163.

Pegram GGS (1980).
An Autoregressive Model for Multilag Markov Chains. *Journal of Applied Probability,* Vol 17, pp 350-362.

Pegram GGS (1997).
Patching rainfall data using regression methods. 3. Grouping, patching and outlier detection. *J. Hydrology,* Elsevier, Vol 198, pp 319-334.

Raftery AE (1985).
A Model for High-order Markov Chains. *Journal of the Royal Statistical Society,* Series B, Vol 47, part 3, pp 528-539.

Schertzer D and Lovejoy S (1987).
Physical modeling and analysis of rain and clouds by anisotropic scaling of multiplicative processes. *J. Geophysical Res.* Vol 92 (D8), pp 9693-9714

Seed AW (1992)
The Generation of a Spatially Distributed Daily Rainfall Database for various Weather Modification Scenarios. WRC report 373/1/92, Water Research Commission, Pretoria.

Woolhiser DA and Pegram GGS (1979).
Maximum Likelihood of Fourier Coefficients to Describe Seasonal Variations of Parameters in Stochastic Daily Precipitation Models. *Journal of Applied Meteorology,* Vol 18, pp 34-42.

Zucchini WS and Adamson PT (1984).
Assessing the risk of deficiencies in streamflow. WRC Report 91/2/84, Water Research Commission, Pretoria.

Zucchini WS and Guttorp P (1991).
A hidden Markov model for space-time precipitation. *Water Resources Res.* Vol 28, No 8, pp 1917-1923.