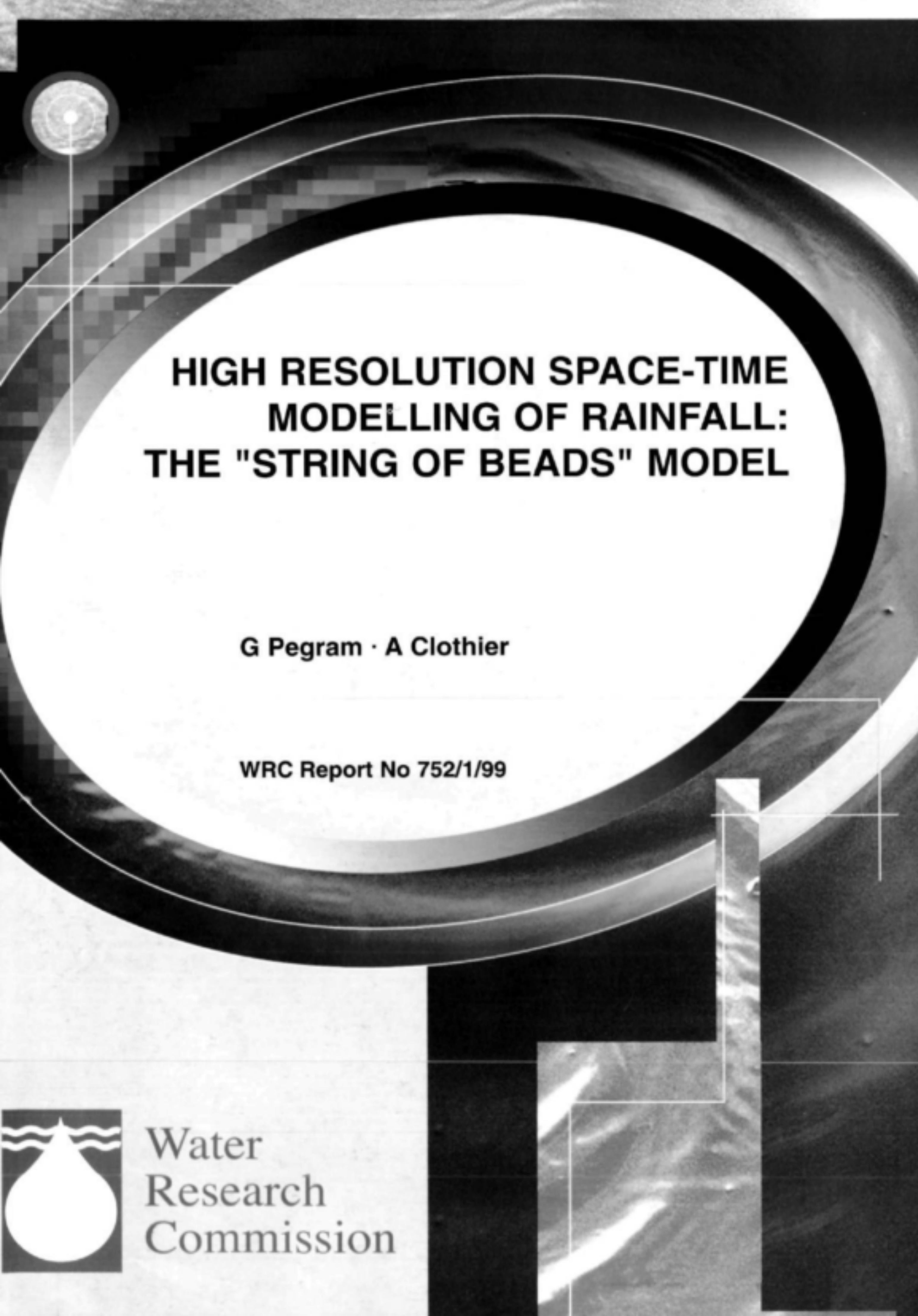


# HIGH RESOLUTION SPACE-TIME MODELLING OF RAINFALL: THE "STRING OF BEADS" MODEL

G Pegram · A Clothier

WRC Report No 752/1/99

Water
Research
Commission

# High Resolution Space-Time Modelling of Rainfall: The "String of Beads" Model

## Report to the Water Research Commission

GEOFFREY PEGRAM AND ANTONY CLOTHIER

Dept. of Civil Engineering,
University of Natal, Durban, South Africa
email: pegram@eng.und.ac.za
email: clothiera@eng.und.ac.za

WRC Report No : 752/1/99
ISBN No : 1 86845 504 1

# Executive Summary

### What is the background to this research?

A valuable measure of the country's potentially available water resources is through rainfall measurements. These measurements help with the quantitative definition of drought and possible climatic change, besides providing information for the design of structures such as spillways, bridges, culverts, channels, storm-water drains etc. On-line measurements of rainfall can help reduce the potential damage of floods; the calculations of runoff from flood-producing storms can be projected into the future to forecast flows in sensitive areas in flood-plains near rivers and through water-bodies stored behind dams.

Traditionally, rainfall measurement has been made using 120mm and 200mm diameter rain-gauges scattered randomly over the country. In South Africa the largest proportion of these were (and still are) read by volunteer observers at 8:00am each day and the daily totals collected by the South African Weather Bureau (SAWB). There was a time (1950) when 4500 or more such gauges were "live" on any particular day, but the number of daily read raingauges has dropped to 1750 in 1999 with less than 600 of these reporting daily for the public good; the numbers are still diminishing.

There is thus a need for remote sensing of rainfall using radar and satellite imagery. These technically advanced, relatively recent innovations provide detailed information about rain-rates over large areas in sequences of images which are typically at 5 minute intervals from radar and 30 minute intervals from satellite platforms. They thus have the potential to provide detailed information (in space and time) of rain rates which can be used in many applications.

**What did the research programme originally set out to achieve?**

The contractual objectives were:

1. To provide a means of describing rainfields in both time and space by means of fractals.
2. To facilitate the modelling of rainfields with economical mathematical descriptors.
3. To use the project as a means of enhancing collaboration between institutions in the KwaZulu Natal region.

The first of these objectives was modified from using fractals to Gaussian Random Fields, with better effect, and has been achieved successfully with the alternative approach. The second objective has been achieved if one interprets the words "with economical" as "parsimoniously with a small number of", as was originally intended. The third objective was achieved in a different sense than was originally intended, because although little collaboration initially developed in KwaZulu Natal proper, other than through the participation of the University of Zululand in the steering committee, the collaboration between researchers in Civil Engineering at Natal University and in the Bethlehem Precipitation Research Programme has become strong.

**Why model rainfall measured by radar in space-time?**

Increasing demands are being made on our meagre water resources and because there is greater population pressure causing invasion of high-risk flood-prone areas, there is an increased need for more detailed information about the spatial and temporal distribution of rainfall to assist with flood warning and mitigation.

In general the purpose of modelling is two-fold: to summarise complex physical processes in a descriptive or mathematical way with as few parameters as possible (parsimony) and to gain as deep an understanding of the phenomena as is possible, relevant and cost effective under the circumstances (insight).

The reason for modelling the rainfall process is to provide as much parsimony and insight as is reasonable, so that possible future scenarios can be artificially generated to conduct "what-if" investigations. Because of the relatively recent advent of remote sensing of rainfall in South Africa, the radar-records are quite short (3 to 5 years) and do not provide, in themselves, an objective basis for some aspects of decision-making. However there is a relatively large archive of data collected from daily read rain gauges which are too sparse, in many instances to provide the spatial detail required in some analyses/studies but give the information needed to quantify the variability of the rainfall processes over a long time-base. It is therefore prudent and desirable to marry the long-term (but poor spatial) information in the archive of daily rain-gauge records with the finely detailed (but short) records obtained from radar.

## What happened to the Fractals?

Fractal models have been proposed as appropriate candidates for modelling rainfields because, by their construction, they preserve certain statistical properties, amongst them self-similarity, scaling etc. The mathematics is difficult and the models used (in particular the multiplicative models) have had the tendency to produce "blocky" rainfields which are distinguishable from typical radar measured rainfields and not easily manipulated to give a movement of the rainfield. By contrast, the approach adopted here was to use the KISS principle, employing standard statistical analysis methods (yet exploiting what had been discovered about scaling fields in the search for appropriate fractal models) to craft a parsimonious, robust, adaptable, feasible model which is statistically correct and visually realistic when compared to typical radar images of areal rainfall. The outcome is a model which is relatively easy to understand, fit and apply, without becoming entangled in the thicket of arcane terminology which tends to make the subject of fractals unapproachable to all but the intrepid.

## Why the "String-of Beads" model?

Rainfall events (storms, showers, light rain etc) are patchy in space and patchy in time but are (in a wide sense) "wet". The time between rainfall events (when it is not raining on a given area) is evidently "dry". Modelling the space-time process of rainfall over an area for a few days, months or years thus consists of defining the process of alternating wet and dry periods and modelling the time-series structure. This is a one-dimensional process in time. Once the rain starts falling on the area during a wet period, the spatial as well as the temporal behaviour of the rainfall phenomenon is of paramount interest:

- how does the rain cover the area?
- how quickly do storm cells grow and decay?
- what is the speed and direction of the storm?

are some of the questions that demand answers.

The wet process is thus a three-dimensional one in two space dimensions (over the area) and time. The one-dimensional wet/dry process can be thought of as the "string" on which the three-dimensional process of the evolution of the rainfall event, the "bead", is threaded.

## How is the model fitted to the data?

The one-dimensional "string" describing the wet/dry process is defined by the sequence of wet and dry days on an area as specified by the rain gauge records. This process can be thought of as the alternation of three types of "climate" (or weather) occurring on the days of the observed record. The types are dry, scattered rain (usually caused by convective storms) and general rain (usually caused by stratiform, large systems). The distinction is a matter of convenience because there are days when the weather is a mixture of convective and stratiform clouds producing rain. This process is modelled by a three-state Markov chain which was thoroughly validated as a good model in WRC contract No. 550 conducted by Pegram and Seed (Report No. 550/1/98).

The three-dimensional space-time "bead" of a given rain event is defined by a sequence of images called Constant Altitude Plan Position Indicators (CAPPIs) which are derived from weather radar measurements. These images are sampled at approximately 5-minute intervals. Where more than 3% of the area covered by the CAPPI records rainfall above 1mm/hr, the image is classified as wet, otherwise the image is part of the "dry" time. Each CAPPI image is analysed to extract three parameters characterising its 9000 or more data points. This parsimonious use of parameterisation was one of the objectives of the original contract. These statistics ($\sigma$, $\beta$ and $\mu$) are stored for use in the simulation/generation process.

## How is a sequence of storms generated?

In generation, the process of analysis is reversed. First a plausible sequence of climate-type is generated. When (for example) a scattered rainday is encountered, a length of storm is randomly generated (usually between 2 and 4 hours in the late afternoon) and a reasonable set of values for the parameters $\sigma$, $\beta$ and $\mu$ chosen. The rainfall event is then produced by power-law filtering (in the complex domain) some Fourier transformed white noise, which is then reverse transformed, scaled and exponentiated to produce a sequence of CAPPI images. These are manipulated to provide realistic velocities of the storm-cells across the design area.

**Does the model work well?**

The answer is that it does with surprising facility and veracity. The sequence of images of CAPPIs generated for a wet period, not only have the right statistics and clustering behaviour, they are visually indistinguishable from the real images they mimic. This may seem like a quaint comment, but the production of images which look like the "real thing" is very important for the end user.

**How could the model be used in practice?**

In order for the model to realise its potential usefulness in application, a link must be formed between the parameterisation of the model and the statistics of raingauges recording one day totals in a region of interest. The proportion of raingauges recording rain on a particular day within a circle of diameter 130km (which should contain about 50 raingauges) is a good indicator of the Wetted Area Ratio, which is likely to have a close link with the parameters of the String of Beads model. The follow-on project (K5/1010) will investigate these links and allow the model to be widely applied.

# Acknowledgements

The authors gratefully acknowledge the following organisations and people:

**Water Research Commission of South Africa** for their generous financial support, encouragement and facilitation over the past two years,

**The WRC steering committee** members:

Dr George Green (Chairman) of the Water Research Commission,
Mr Hugo Maaren of the Water Research Commission,
Dr Deon Terblanche of the South African Weather Bureau,
Prof. Bruce Kelby of the University of Zululand,
Prof. Dennis Hughes of Rhodes University,
Dr Mark Dent of the Computing Centre for Water Research,
Mr Stefan van Biljoen of the Department of Water Affairs and Forestry,
Mr Steve Lynch of the University of Natal,
Mr Peter Visser of the South African Weather Bureau,

for their active interest in the development of the project and for the encouragement and valuable ideas that they offered during the steering committee meetings,

**South African Weather Bureau at Bethlehem** for providing excellent radar and raingauge data which were used in the analysis and for critically appraising the products of the research,

**Alan Seed** now of the Australian Weather Service in Melbourne for fuelling our interest in the topic and for sharing his contributions to the spatial and temporal modelling of rainfall in South Africa. His generous support regarding the radar database management and his critical appraisal of the model are also acknowledged with thanks,

# Acknowledgements...

**Merab Menabde** of the Physics Department of Auckland University for his guidance in the early stages of the research. The information he offered regarding the power-law filtering of white noise was crucial to the development of the model,

**Computing Centre for Water Research** for their assistance in the transfer of radar data between Bethlehem and Durban through the use of their FTP server,

**Duncan Forrest** for the insight he offered regarding image data storage and manipulation and for the CD Write facility that he offered on several occasions.

# Table of Contents

# Table of Contents...

# Table of Contents...

* a **Compact Disk** is available from the Water Research Commission at cost of reproduction.

# List of Figures

# List of Figures...

# List of Figures...

# List of Figures...

# List of Figures...

# List of Figures...

# List of Tables

# The String of Beads model

## 1. INTRODUCTION

Rainfall modelling has been a source of great interest in recent years and there have been a vast number of models proposed which vary in complexity from simple stochastic models through to highly complex, physically based models which attempt to model the micro-physical behaviour of clouds. The choice of the rainfall model will depend on its application. The model which is proposed in this study is designed to be used for real time rainfall simulation as seen by radar. It will find application in the simulation of rainfall for urban drainage control systems, large catchment flood control and water resources management or any other model that requires a detailed rainfall simulation in space-time. For reasons which will become apparent later in this report, it will be referred to as the *String of Beads model*.

### 1.1 Some relevant literature

The *String of Beads model* has been developed using ideas from a variety of existing models which are cited in the text of the report. It is perhaps appropriate at this point to present a brief review of some of the work which was particularly pertinent to the development of the *String of Beads model*.

A brief history of radar and radar meteorology was presented by Austin (1998) at an advanced study course on Radar Hydrology for Real Time Flood Forecasting held at the University of Bristol in 1998. Austin (1998) explains how radar was developed first by the allied forces as a military tool in World War 2 used to predict German air raids on Britain. Rainfall was observed as a nuisance on these early images and techniques were developed to distinguish aircraft from rainfall. Only after the war, it was realised that rainfall intensity could actually be measured by radar and so was born the science of radar meteorology. Early research in this field was focussed on the relationship between radar reflectivity and rainfall rate leading up to the well known Marshall-Palmer relationship published in 1948. Research has continued in this area and numerous improvements to the measurement of rainfall using radar have culminated in the modern weather radar.

Prior to the invention of radar in the mid 1930's, rainfall was mainly modelled as a point process based on raingauge data and various forms of interpolation between gauges were used to give an estimate of the rainfall over an area. With the advent of radar, the true space-time structure of the rainfall field was revealed and this sparked an interest in more realistic modelling techniques amongst both the hydrological and meteorological communities. Until recently, many of these techniques have been limited by the available computing power but as processors have been improved more elaborate space-time rainfall models have emerged. A general review of the more recent advances in rainfall modelling is given by Foufoula-Georgiou and Krajewski (1995) who consider the submissions to U.S. journals during the period 1991-1994. Foufoula-Georgiou and Krajewski (1995) divide rainfall models into two main categories, stochastic and dynamic models. They define dynamic models as being based on a set of partial differential equations describing conservation of mass, momentum and energy. This type of model belongs to the realm of cloud physicists and meteorologists and was not considered in this study since the String of Beads model is aimed more at the Hydrological community who are interested in the statistics of rainfields rather that the four dimensional (three space and one time) physical behaviour of clouds.

This leaves the stochastic models. Foufoula-Georgiou and Krajewski (1995) report how the trend in rainfall modelling has shifted towards models which are based on concepts of scale invariance which assumes that the stochastic structure of the rainfall field is independent of scale. This assumption has been shown to be true over a wide range of scales (Gupta and Waymire, 1990) and it lends itself to modelling through the use of multifractal algorithms such as the discrete random cascade and the multiplicative random cascade. Models of this type have been proposed by Schertzer and Lovejoy (1987), Over and Gupta (1994) and Menabde et al. (1997) amongst others, but although statistically sound, the rainfield images produced by these models have the unfortunate property of appearing blocky. In spite of this, a great deal of work has been done on the parameterisation and calibration of these models and to date, they are arguably the best statistical descriptors of the rainfall field.

Menabde et al. (1997) use a discrete random cascade to generate a field with the desired statistical structure which they then power law filter, thereby removing some of the blockiness and resulting in a more realistic looking rainfield. An alternative to this approach would be to simply generate a random field based on a suitable probability distribution and then power law filter the field to achieve the desired spatial correlation structure. This is the approach that was used by Bell (1987) for the simulation of satellite rainfall intensity data. He found that the lognormal distribution was a good descriptor of the marginal distribution of rainfall intensities on a field and this was confirmed by Crane (1990). The approach by Bell (1987) was simply to generate a Gaussian random field, power law filter it to achieve the desired spatial correlation structure and then exponentiate the field to obtain a spatially correlated, lognormally distributed field. These fields are convincing in appearance and possess a reasonable statistical structure.

In his work on fluid turbulence, Taylor (1938) hypothesised that the correlation of a point with itself in time (i.e. at a fixed point in space) was related to the correlation of that point with its neighbours in space (i.e. at a fixed time) through some velocity vector. This hypothesis is discussed with reference to the space-time structure of rainfall intensity fields by Gupta and Waymire (1987) and by Crane (1990). If Taylor's hypothesis is assumed to be true, the approach adopted by Bell (1987) for the simulation of two dimensional fields can be simply extended into three dimensional space-time by using a power law filter which defines the temporal correlation structure as a vector multiple of the average spatial correlation structure.

The algorithm used to generate these three dimensional fields makes use of the Fast Fourier Transform thereby reducing the computational effort and making it possible to generate a large number of fields in a short space of time. In his paper, Bell (1987) points out that *opposite edges of the Fourier filtered images behave as if they are joined* and this property can be easily exploited to introduce an overall advection of the rainfield. This gives the Fourier filtered images an upper hand when compared to generating correlated, lognormally distributed fields through the use of other algorithms such as the Turning Bands method (Matheron, 1973) which is described for example, by Bras and Rodriguez-Iturbe (1985). Bell's (1987) process produces a finite sequence of images which are correlated in both space and time and which incorporate an advection characteristic. However, it would not be feasible to simulate long periods of data through the use of this process alone, since the computational effort and data storage requirements would soon become unworkable.

An additional source of information with regards to the temporal structure of rainfall can be obtained through the use of raingauges and a variety of approaches have been proposed which mimic the behaviour of raingauge time series data. These include models such as the

Neymann-Scott shot noise model proposed by Cowpertwait (1994) and the daily climate model proposed by Pegram and Seed (1998). Pegram and Seed (1998) make use of daily raingauge data to classify the type of rainfall experienced during a day over a chosen area. They then use a three state Markov chain to simulate sequences of raindays throughout the year for that chosen area. Cowpertwait (1994) adopts an interesting approach to modelling a high resolution time series of rainfall intensity data by considering the duration of an event and the event arrival as two separate random processes. His concept can be used on a larger temporal scale by employing the daily climate model proposed by Pegram and Seed (1998) to model the type of rainfall and its arrival, whilst using power law filtering approach adopted by Bell (1987) to model the high resolution behaviour of the event. This is the essence of the String of Beads model which will be discussed in this report.

## 1.2 An introduction to the String of Beads model.

The *String of Beads model* is defined by a small number of parameters and is based on the combination of a modest data archive of rainfall intensity images measured by radar and a large archive of data collected from a daily raingauge network. It is a phenomenological model which is designed to run on a fast modern personal computer and provides a realistic simulated random rainfall scenario. It can be used for real-time simulation of a rainfall event or sequence of events over any number of years.

Radar and raingauge data for this study were supplied by members of the Bethlehem Precipitation Research Programme. The weather radar rainfall data were provided in a pre-processed, compressed file format and prior to any analysis being undertaken, it was necessary to convert it into an uncompressed format which is easy to manipulate and interpret. The process of extracting the raw data and converting it into a common image file format is discussed in detail in Appendix A - *Radar and raingauge data*. Other issues which are discussed in that appendix are data quality, data storage and data visualisation.

With the data in a common, uncompressed image file format, the radar rainfall images can be viewed in almost any Windows based graphics program. These images are referred to as **C**onstant **A**ltitude **P**lan **P**osition **I**ndicators (CAPPIs) and they are an instantaneous, two dimensional, best estimate of rainfall intensity at a constant altitude (2km) above the ground. The CAPPIs used in this study cover an area of 200km x 200km centred on the MRL5 radar near Bethlehem, at a resolution of 1km x 1km and are recorded at five minute intervals. Additional information regarding the configuration of the radar is given in Appendix A.

The raingauge network database is used to develop a model for the climatic behaviour which defines the type of rainfall on any day in the year. This *Daily Climate Model* was developed in a previous study by Pegram and Seed (1998) and is incorporated into the *String of Beads model* almost without modification. The definition of their daily climate model is paraphrased in the report and the full set of periodic transition probability curves are included in the Appendix B.

The weather radar database is used to develop a model for the high resolution space-time behaviour of the rainfall events by first analysing the CAPPI data and then using a technique known as *power-law filtering of Gaussian noise* to simulate sequences of CAPPIs. To improve the readability of the report, the theory used in the analysis and simulation processes is only briefly explained in the main report and more comprehensive discussions are provided in Appendices C and D. The daily climate model and the high resolution space-time model are combined to form the *String of Beads model*.

The two main modes of operation of the *String of Beads model* are described and a three-dimensional (two space and one time) random simulated rainfall event is generated and analysed as a means of model verification. The Generalised Structure Function, will be used to test the validity of the simulated CAPPIs and the results will be presented and discussed. It will be demonstrated that the String of Beads model simulates rainfall realistically and without obvious bias. Finally, animated images of the model output are included on the compact disk (refer to Appendix G) which accompanies this report.

## 2. A STOCHASTIC SPACE-TIME MODEL OF RAINFALL

The "String of Beads" model is a stochastic space-time model of rainfall. It does not attempt to explain or reproduce the physical behaviour of clouds or rainfall. It is a phenomenological model which is based on the statistics of data collected by raingauges and from radar rainfall images at a constant altitude.

Radar can provide good quantitative rainfall estimates over a large area and at a high spatial and temporal resolution and for this reason it is becoming increasingly popular in hydrological applications. The "String of Beads" model can provide a realistic simulated rainfall scenario which can be used for real-time simulation of a rainfall event or sequence of events. The output of the model is in the form of a sequence of animated images (each pixel a numerical value of rainfall intensity) that behaves like a visualisation of a storm of several hours (or days) duration as seen by radar in space and time.

## 3. THE STRING OF BEADS CONCEPT

The concept of the *String of Beads model* is similar to the Neymann-Scott shot noise model (Cowpertwait, 1994) in that it models each rainfall *event* and its *inter-arrival time* separately. The event *inter-arrival time* is a one dimensional process which is modelled as described in Section 6. Although the rainfall *event* is actually a four dimensional process (three space and one time), it has been simplified to a three dimensional process in the String of Beads model by analysing and simulating 2km CAPPIs rather than the sequence of full volume scans. The process of modelling the rainfall event is described in Section 9 and involves the power-law filtering of Gaussian noise. The one dimensional event *inter-arrival time* can be thought of as a *string* and the three dimensional rainfall *event* as a *bead* on the string as shown in Figure 1.

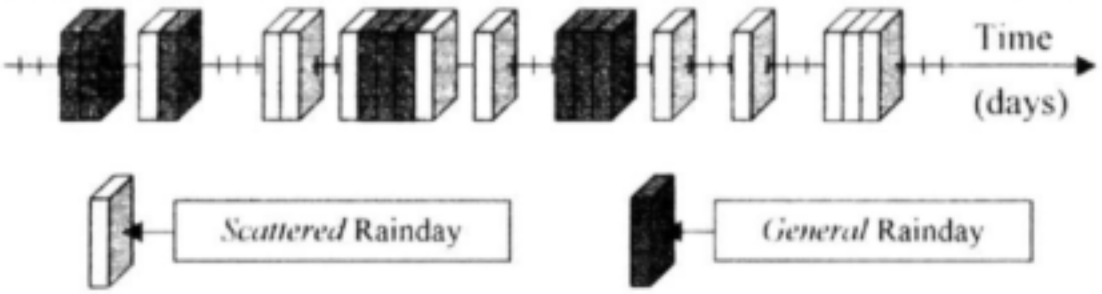


**Figure 1 - An illustration of the String of Beads concept**

Each *bead* in the model is a sequence in time of sequential CAPPIs at five minute intervals in the case of a *scattered* rainday, or at thirty minute intervals in the case of a *general* rainday. The time interval is adjusted in this way to account for the fact that a general rainfall event is less volatile than a scattered event and can therefore be modelled using a lower temporal resolution. In order to preserve disk storage space, it is important to exploit this lower volatility because general rainfall events can last for days, whereas scattered rainfall events usually only last a few hours. The number of CAPPIs in a bead therefore varies according to the duration and type of the rainfall event.

## 4. DATA PROVIDED BY THE SOUTH AFRICAN WEATHER BUREAU

An area of great strategic interest in terms of water resources and flood control in South Africa is the catchment of the Vaal Dam (38 505 km$^2$) which provides most of the stored water for Johannesburg and Pretoria and the surrounding areas. The Vaal catchment is monitored by two radars shown in Figure 2, a C-band at Ermelo and an S-band 20km north west of Bethlehem. The rainfall in the catchment is also monitored by a network of daily raingauges maintained by local volunteers, and a good network of tipping bucket raingauges on a sub catchment of 4 500km$^2$ which is maintained by the South African Weather Bureau (SAWB).

**Figure 2 - Coverage of the Bethlehem and Ermelo Radars**

The daily raingauge data for this region extends over approximately 80 years and the network of tipping bucket raingauges has recorded data for approximately 6 years. The MRL-5 S-band radar was installed near Bethlehem by the Water Research Commission in 1994 and continuous rainfall data during the wet summer months (November to April) are available from this radar from January 1995. Data from the C-band radar at Ermelo are available from

January 1998. Both radars are configured to do a full volume scan every 5 minutes starting with a base scan at an elevation of 1.5 degrees and finishing with an elevation of 55 degrees as discussed in Appendix A.

A basic assumption made in the development of the model is that the data received from the radar provides a reasonably accurate representation of what actually reaches the ground. In general, rainfall data measured by radar are subject to a wide variety of systematic errors such as bright band, attenuation, ground clutter and anomalous propagation. These effects are briefly discussed in Appendix A, although they are only rarely observed in the data received from the SAWB at Bethlehem. Apart from the Maluti Mountains in Lesotho, the terrain surrounding the Bethlehem radar is quite flat and there is therefore very little ground clutter and the rainfields tend to be reasonably homogeneous. The effect of attenuation is kept to a minimum by using an S-Band radar. Due to the high summer temperatures (wet season), the bright band is usually at or above the 2km level. A large proportion of the rainfall experienced in Bethlehem is convective in which case the bright band is not well defined and does not present a problem. The Z-R relationship $\boldsymbol{Z = 200R^{1.6}}$ given by Marshall and Palmer (1948) is used to convert reflectivity to rainfall rate. As methods are developed to improve the measurement of rainfall using radar providing improved data sets, the model will be better able to simulate real rainfields.

## 5. QUANTITATIVE LIMITS OF THE RADAR DATA

The data received are in the form of 200 x 200km images of instantaneous rain rate at 8bit precision (one byte per pixel) which gives a possible range of any integer value of rainfall intensity between 0 and 255mm/h. The base scan of the Bethlehem radar is at an elevation of 1.5° above the horizontal so that the range at which the base scan exceeds an altitude of 2km above ground level is 67km. Data recorded beyond this range are useful only for qualitative analysis of the weather system since these are usually within or above the melting layer. It is convenient in the analysis to use a radius of 64km as it is an integer power of two, a requirement for the Fast Fourier Transform whose use will be discussed in Section 7. CAPPIs are plotted as an average rainfall intensity at the centre of a pixel which represents an area which is 1km square in a horizontal plane (usually at an altitude of 2km above the ground).

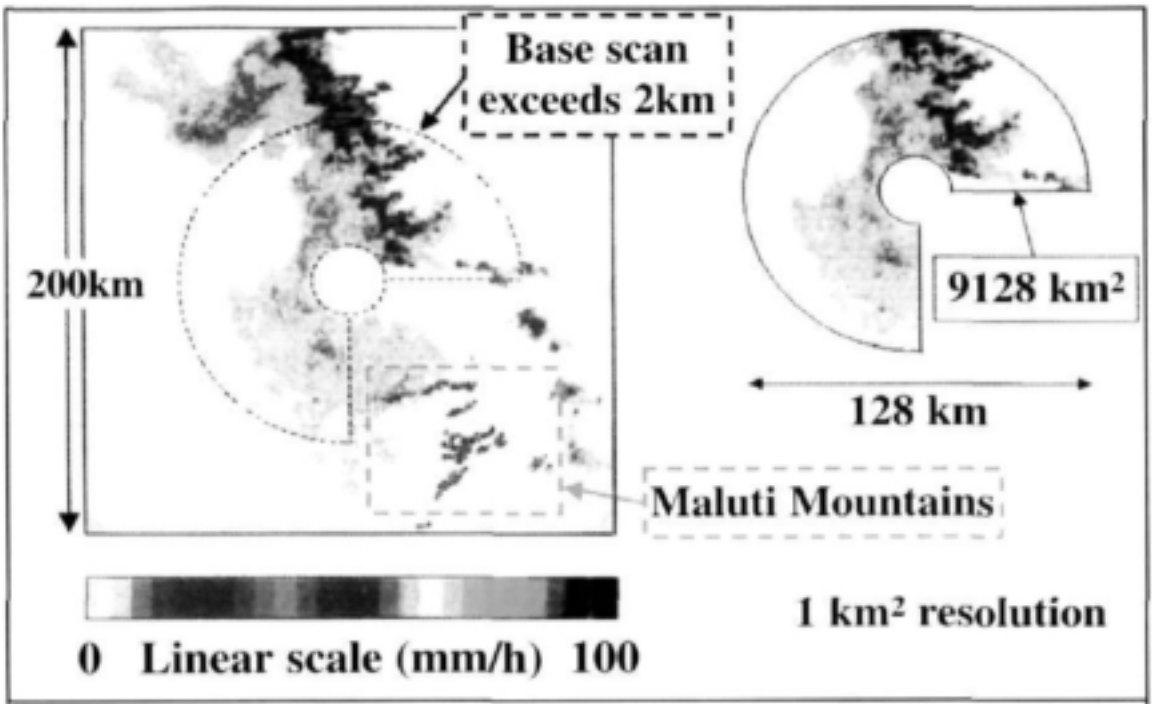


**Figure 3 - Masking the raw CAPPI data to eliminate ground clutter and data recorded at an altitude in excess of 2km.**

In the data-sets available to date, the radar did not record any reflections received within a 14km radius of itself although this "hole" will be eliminated in future recordings. For the purposes of the analysis, CAPPIs are masked as shown in Figure 3 so that only data within the $^{3}/_{4}$ doughnut shaped sample area of 9128 km$^2$, are considered.

## 6. MODELLING THE SEASONAL DAILY RAINFALL DISTRIBUTION

One of the benefits of radar data is that, when it comes to local application, the image gives very good spatial information that cannot be achieved using raingauges. However, the MRL5 S-Band radar has only been operational in Bethlehem for the past four years compared to nearly a century of daily raingauge recording. In order to include the daily climate variation in the model, it is necessary to make use of the information that is available from the network of raingauges recording daily totals in the area.

In the context of the *String of Beads model*, the daily climate is defined as the type of rain experienced on a particular day. Three *states* are considered in the model and they are *dry*, *scattered* and *general* raindays. Pegram and Seed (1998) analysed a 30 year collection of up to 330 daily read raingauges in the 200 kilometre square centred on Bethlehem. The minimum number of reporting gauges within the square at any time was 150. They used a three state

Markov Chain with parameters varying periodically over a year to model the temporal behaviour of the daily climate in the Bethlehem area. In this section, the parameters of their model will be redefined in terms of radar data and their methodology will be explained.

### 6.1. Definition of the three states of rainday

To define the Markov chain climate model, a definition of the states of the model is first required. Pegram et al. (1997) and Pegram and Seed (1998) used the three states defined in Table 1 in their analysis of the daily raingauge data.

**Table 1 - Daily weather classification for raingauge network (Pegram and Seed, 1998)**

| State | Description | Definition |
|---|---|---|
| 1 | Dry | < 3% gauges report rain |
| 2 | Scattered | > 3% gauges report rain but<br>< 50% gauges report rain in excess of 5mm |
| 3 | General | > 50% gauges report rain in excess of 5mm |

The definition of these states was based on a classification given by Court (1979) who attempted to find a correlation between seedable cloud formations (cumulus development, cloud tops colder that $-5^{\circ}$C, with no hail) and the fraction of raingauges in the area reporting rain. Scattered raindays are considered seedable. This classification was modified for CAPPI data and adopted for the String of Beads model to give the three states shown in Table 2. The mean wet area of Table 2 is the ensemble average of the Wetted Area Ratios (WAR, defined as the fraction of the CAPPI which experiences a rainfall intensity in excess of 1mm/h) of all of the CAPPIs recorded on the day.

**Table 2 - Daily weather classification for radar CAPPIs**

| State | Description | Definition |
|---|---|---|
| 0 | Dry | mean wet area < 3% |
| 1 | Scattered | 3% < mean wet area < 50% |
| 2 | General | 50% < mean wet area |

### 6.2. The three state Markov chain

The Markov chain is a statistical means of switching between the three states of rainday defined in Table 2. It works on the assumption that the type of rainfall experienced on a particular day is dependent on both the season and the type of rainfall experienced for some time prior to that day. For instance, general rainfall in South Africa is usually caused by a large weather system which moves across the country and dissipates over a period of one or two days. For this reason, given that the state of rainday in a sequence is *general*, the probability of the following day being *general* or *scattered* is relatively high when compared to the probability of the following day being *dry*. These are known as the *transition probabilities*.

It is conceivable that the state of a particular day is dependent on the state of two or more days prior as experienced in other daily rainfall models (e.g. Woolhiser and Pegram (1979), and Zucchini and Adamson (1984)). Pegram and Seed (1998) made use of the Akaike Information Criterion (as defined for a Markov chain and used by Pegram (1980)) to test this possibility, but found that the simple lag-one Markov chain was the most suitable descriptor of the observed data.

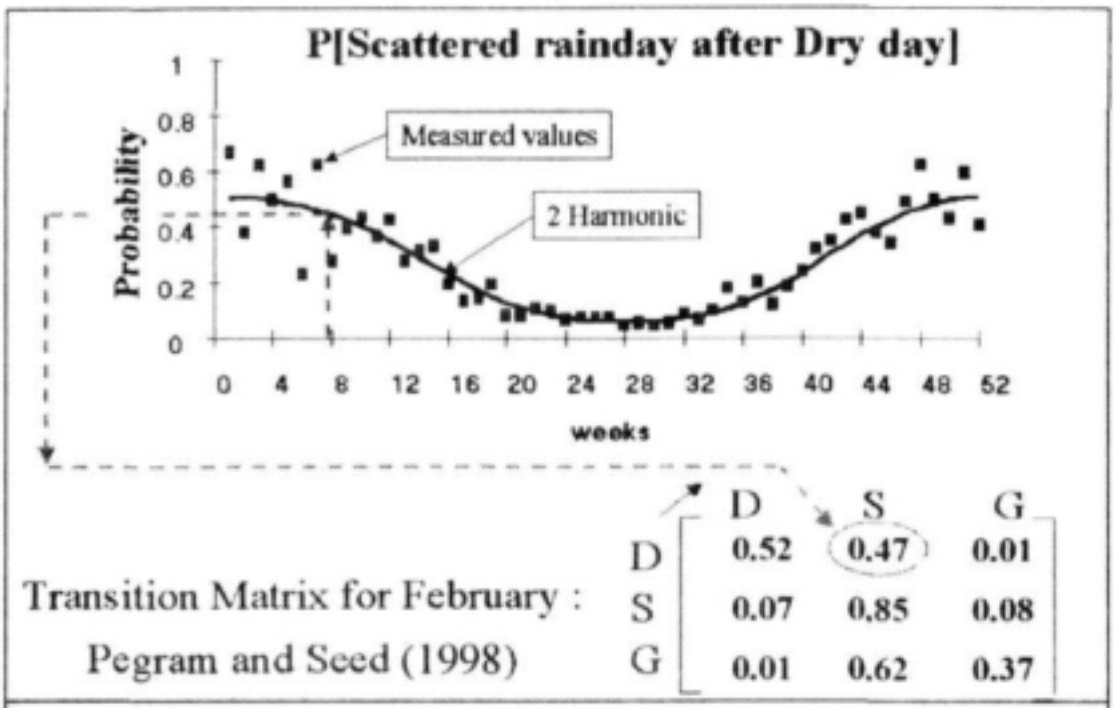
If the states of the lag-one Markov Chain are numbered 0, 1, 2 as shown in Table 2, the transition probabilities become:

$p_{ij}(t)$ = probability [given that the state is $i$, the state of the next day will be $j$ ] $(i, j = 0,1,2)$

These transition probabilities can be assembled into a *transition probability matrix* of the form:

$$\mathbf{P(t)} = \begin{bmatrix} p_{00}(t) & p_{01}(t) & p_{02}(t) \\ p_{10}(t) & p_{11}(t) & p_{12}(t) \\ p_{20}(t) & p_{21}(t) & p_{22}(t) \end{bmatrix}$$

Since the total of each row is constrained to unity, six transition probabilities suffice to define the Markov chain. The transition probabilities were estimated for each week in the year and were fitted with a two harmonic Fourier series to give a periodic relationship for the transition probabilities. An example of the Fourier fit to one of the nine transition probabilities is given in Figure 4.



Transition Matrix for February :
Pegram and Seed (1998)

$$\begin{array}{c|ccc} & D & S & G \\ \hline D & 0.52 & 0.47 & 0.01 \\ S & 0.07 & 0.85 & 0.08 \\ G & 0.01 & 0.62 & 0.37 \end{array}$$

**Figure 4 - The development of the transition matrix for the three-state Markov chain by fitting a Fourier series to 30 years of raingauge data.**

The full set of *periodic transition probability curves* for the Bethlehem area, calculated by Pegram and Seed (1998), are included in Appendix B. The points represent the weekly sampled transition probabilities from the 30 year collection of up to 330 daily read raingauges in the 200 kilometre square centred on Bethlehem. The line which has been fitted to these points is the two-harmonic Fourier fit. Pegram and Seed (1998) investigated the possibility of using higher order Fourier approximations, but found that in most cases this did not significantly improve the fit.

The rainfall experienced in Bethlehem in February is mainly comprised of afternoon convective rainfall caused by high summer temperatures on the ground. These convective events would usually be classified as *scattered* in terms of the three states defined in Table 2. The marginal distribution of the states changes *slowly* with time, so is *approximately* given by solving the homogeneous equation

$$\pi^{T} P = \pi^{T}$$

yielding

$$\pi^T = [0.12 \quad 0.78 \quad 0.10]$$

This means that the probability of a randomly chosen day in February being a *scattered* rainday is 78%. The probability of a randomly chosen day in February being a *general* rainday is 10%. Being the peak of the rainy season in Bethlehem, it is unlikely (12%) to observe a *dry* day in February.

The *monthly transition probability matrices* used in the String of Beads model are constructed from the curves using the method illustrated in Figure 4. A summary is given in Table 3 of the transition probabilities which make up the transition probability matrices for each month. As shown in Appendix B, the *general to anything* transition probabilities are independent of the season.

**Table 3 - Summary of monthly transition probabilities which make up the monthly transition probability matrices used in the String of Beads model.**

| **Month** | **Monthly Transition Probabilities** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | ***Dry* to...** | | | ***Scattered* to ...** | | | ***General* to ...** | | |
| | ***Dry*** | ***Scat*** | ***Gen*** | ***Dry*** | ***Scat*** | ***Gen*** | ***Dry*** | ***Scat*** | ***Gen*** |
| **Jan** | 0.49 | 0.49 | 0.02 | 0.06 | 0.86 | 0.08 | 0.01 | 0.62 | 0.37 |
| **Feb** | 0.52 | 0.47 | 0.01 | 0.07 | 0.85 | 0.08 | 0.01 | 0.62 | 0.37 |
| **Mar** | 0.61 | 0.38 | 0.01 | 0.10 | 0.82 | 0.08 | 0.01 | 0.62 | 0.37 |
| **Apr** | 0.74 | 0.26 | 0.00 | 0.21 | 0.72 | 0.07 | 0.01 | 0.62 | 0.37 |
| **May** | 0.85 | 0.15 | 0.00 | 0.33 | 0.61 | 0.06 | 0.01 | 0.62 | 0.37 |
| **Jun** | 0.92 | 0.08 | 0.00 | 0.43 | 0.52 | 0.05 | 0.01 | 0.62 | 0.37 |
| **Jul** | 0.94 | 0.06 | 0.00 | 0.44 | 0.52 | 0.04 | 0.01 | 0.62 | 0.37 |
| **Aug** | 0.92 | 0.08 | 0.00 | 0.38 | 0.57 | 0.05 | 0.01 | 0.62 | 0.37 |
| **Sep** | 0.86 | 0.14 | 0.00 | 0.29 | 0.65 | 0.06 | 0.01 | 0.62 | 0.37 |
| **Oct** | 0.73 | 0.26 | 0.01 | 0.19 | 0.73 | 0.08 | 0.01 | 0.62 | 0.37 |
| **Nov** | 0.59 | 0.40 | 0.01 | 0.12 | 0.79 | 0.09 | 0.01 | 0.62 | 0.37 |
| **Dec** | 0.52 | 0.48 | 0.00 | 0.07 | 0.83 | 0.10 | 0.01 | 0.62 | 0.37 |

The String of Beads model makes direct use of these relationships to generate a sequence of rainfall events. The probabilities in the rows of the transition matrix can be arranged in cumulative form and must obviously sum to unity. Given that the current state is *dry* for example, a uniformly distributed pseudo-random number ***R(0,1)*** is generated using the pseudo random number generator of Appendix D. For a uniformly distributed number ***R(0,1)*** less than that shown in row **D**, column **D** as 0.52 (Figure 4), will result in the state of the next day being *dry*. Similarly, if 0.52< ***R(0,1)*** < 0.99 (the last number being the sum of the transition probabilities in columns **D** and **S**) a *scattered* rainday will follow and ***R(0,1)*** > 0.99 would give a *general* rainday. The row of the transition matrix which is used for the calculation of the following rainday will then depend on the new state. This is computationally very fast and very simple to program.

### 6.3. Daily climate model validation

The climate model described by this three state Markov chain was carefully validated using statistical tests described in Pegram and Seed (1998). Briefly, they generated 100 30-year simulations and compared the statistics of those simulations to the 30-year historical sequence obtained from the raingauge network. Amongst the statistics that were compared were:

- The mean and standard deviation of the number of each rainday type observed in each month
- Frequency distributions of run-lengths of each of the three rainday types observed in each month.

The results of these two tests are particularly pertinent to the String of Beads model as they measure the ability of the model to replicate the sequence of rainday types. They indicated that:

- The mean number of each rainday type observed in each month was closely simulated by the model.
- During the dry months, the standard deviations of the number of *dry* days in a month of the simulated sequence were smaller than observed in the historical sequence.
- During the wet months, the standard deviations of the number of *scattered* days in a month of the simulated sequence were smaller than observed in the historical sequence.
- The standard deviation of the number of *general* raindays in a month was reproduced well throughout the year.
- The run lengths (number of consecutive days all of a given type) were very well mimicked by the model for the dry, scattered *and* general states.

Although this model does have its limitations in terms of reproducing the standard deviations of the number of *dry* and *scattered* raindays observed in some months, this could possibly be rectified by introducing a noise term into the appropriate transition probabilities. In its current state of development, the String of Beads model uses the transition probabilities of Table 3 *without* any random variation.

## 7. RAINFALL EVENT ANALYSIS

Having established an effective means of analysing and modelling the daily climate the following approach was applied to the rainfall events measured by radar. The first step was to devise a way of measuring and classifying a masked CAPPI (see Figure 3) using as few statistics as possible. The two parameter lognormal distribution has been used effectively in describing the marginal distribution of rainfall rates over an area as estimated by satellite (Bell, 1987) and by radar (Crane, 1990) and following their example, this distribution was applied to the radar CAPPI data obtained from the South African Weather Bureau.

### 7.1. The lognormal distribution applied to CAPPI data

Since the radar data received from the SAWB is in integer format in the range 0 to 255mm/h, it was necessary to use an algorithm for fitting the lognormal distribution for grouped data which is described by Aitchison and Brown (1957). The formal definition and the distribution properties are discussed in Appendix C. The multi-dimensional downhill simplex method of Nelder and Mead (1965) as given by Press et al. (1992) was used to maximise the log likelihood function ***L*** shown in equation (1) with respect to the parameters $\mu$ and $\sigma$ for each CAPPI.

$$L = \sum_i n_i \log\left\{ N\left( \frac{y_i - \mu}{\sigma} \middle| (0,1) \right) - N\left( \frac{y_{i-1} - \mu}{\sigma} \middle| (0,1) \right) \right\} \qquad (1)$$

where $\boldsymbol{i} = 1, 2, \ldots, 255$,

$\boldsymbol{y_i}$ = log (rainfall rate $\boldsymbol{i}$ mm/h )

$\boldsymbol{n_i}$ = number of pixels in the masked CAPPI recording a rainfall rate of $\boldsymbol{i}$ mm/h

$\boldsymbol{\mu}$ = the mean of the logs of the rainfall rates in the CAPPI

$\boldsymbol{\sigma}$ = the standard deviation of the logs of the rainfall rates in the CAPPI.

To illustrate the effectiveness of the lognormal distribution in describing the marginal distribution of the rainfall intensities on a CAPPI, five examples are given in Figure 5, each with a different type of weather system. The parameters ***m*** and ***s*** given in Figure 5 are the most likely estimators of $\mu$ and $\sigma$ of equation (1).

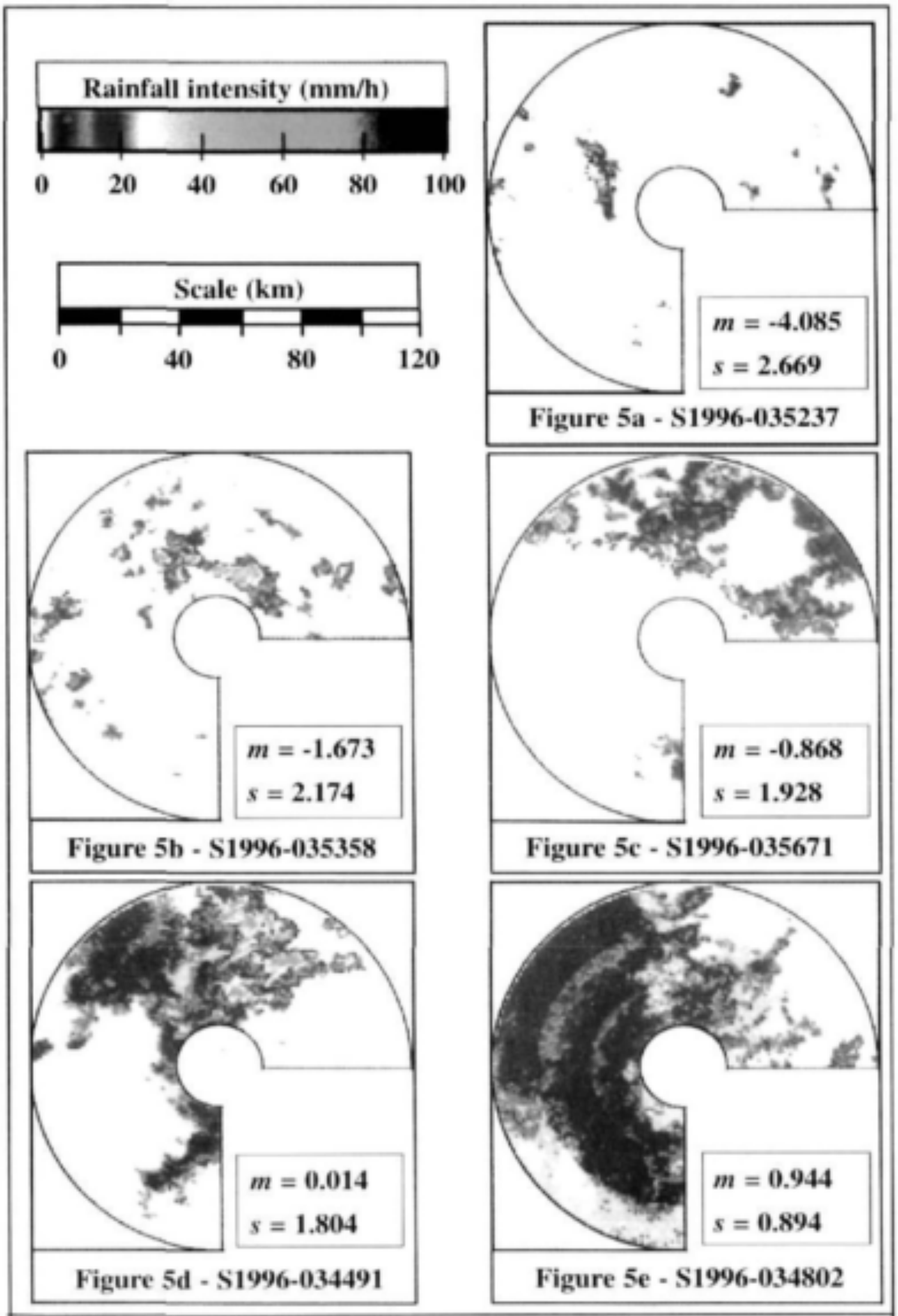


Figure 5a - S1996-035237

Figure 5b - S1996-035358

Figure 5c - S1996-035671

Figure 5d - S1996-034491

Figure 5e - S1996-034802

Figure 5 - Illustration of the types of rainfall observed in Bethlehem

The observed histogram of rainfall intensity and the expected histogram which corresponds to the maximum likelihood approximation of the lognormal distribution for each of the five CAPPIs of Figure 5, are plotted in Figures 6 to 10. To enable the critical examination of the fit, the histograms are plotted with a logarithmic abscissa and the final bin is lumped for the most intense 0.25% of the number of pixels on the CAPPI. The labels on the ordinate axis represent the lower limit of the bin (for grouped data) so that the first bin which is labelled 0mm/h includes all rainfall rates between 0 and 1mm/h.

Figure 5a shows a *free convective system* with isolated convective cells which are less than 15km in diameter. Due to the large proportion of the CAPPI with rainfall between 0 and 1mm/h (bin 0), a low ***m*** is observed, however the high rainfall intensities experienced within the convective cells results in a high ***s*** parameter. The expected and observed histograms for this CAPPI are plotted in Figure 6.

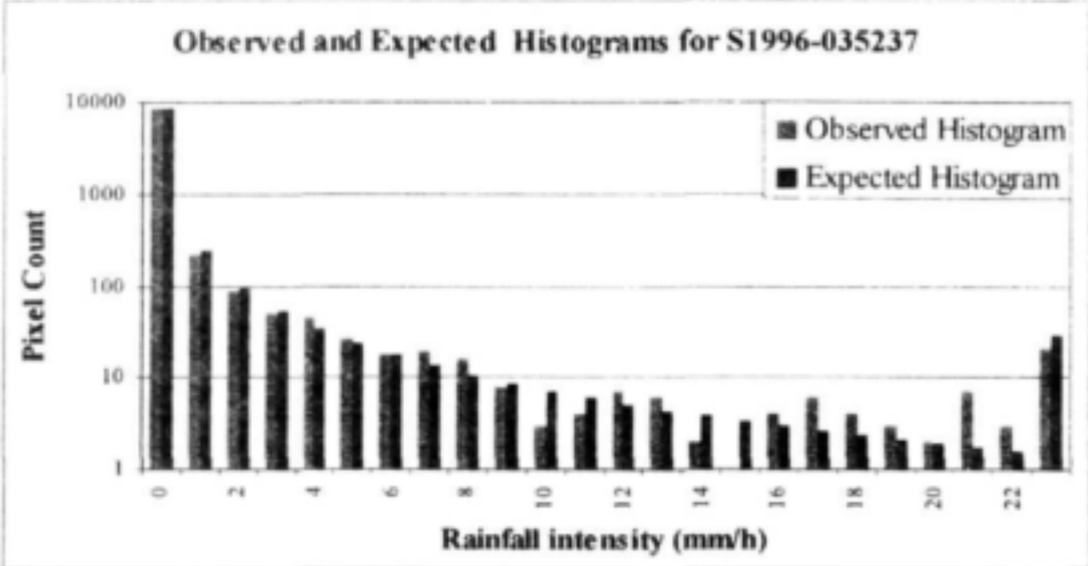


**Figure 6 - Expected and observed histograms for CAPPI S1996-035237 depicted in Figure 5a**

Figure 5b shows a *free convective system* with large isolated convective stormcells which are of the order of 20km in diameter. The result is an intermediate ***m*** parameter combined with a high ***s*** parameter. The histograms for this CAPPI are plotted in Figure 7.

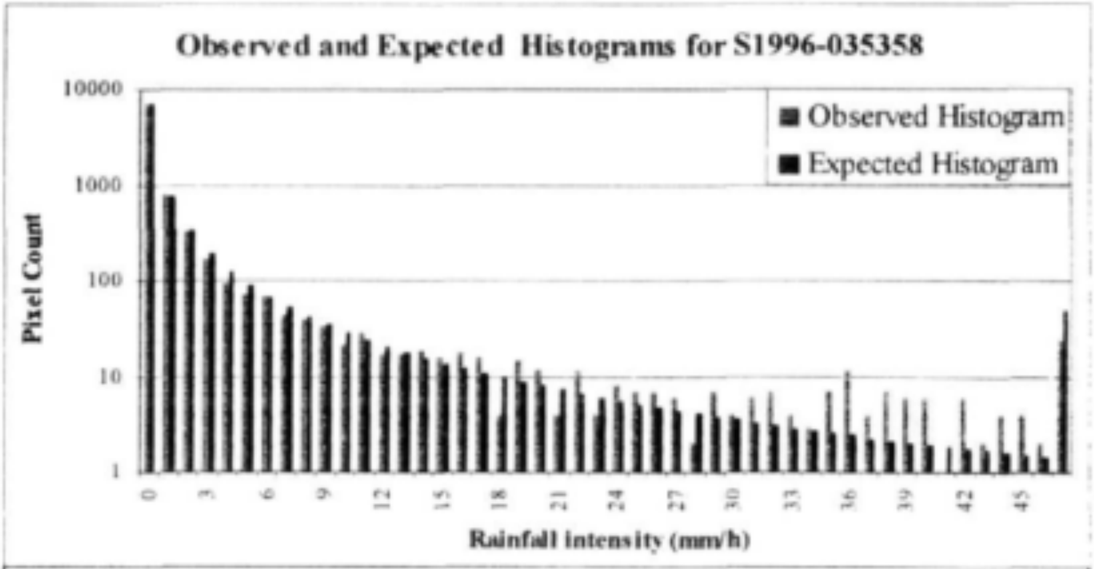


**Figure 7 - Expected and observed histograms for CAPPI S1996-035358 depicted in Figure 5b**

Figure 5c shows a weather system which is between a *free convective* and a series of *forced convective systems* with a characteristic high **s** parameter and a reasonably high ***m*** parameter. The most intense rainfall rate on this CAPPI is 85mm/h. The histograms for this CAPPI are plotted in Figure 8.

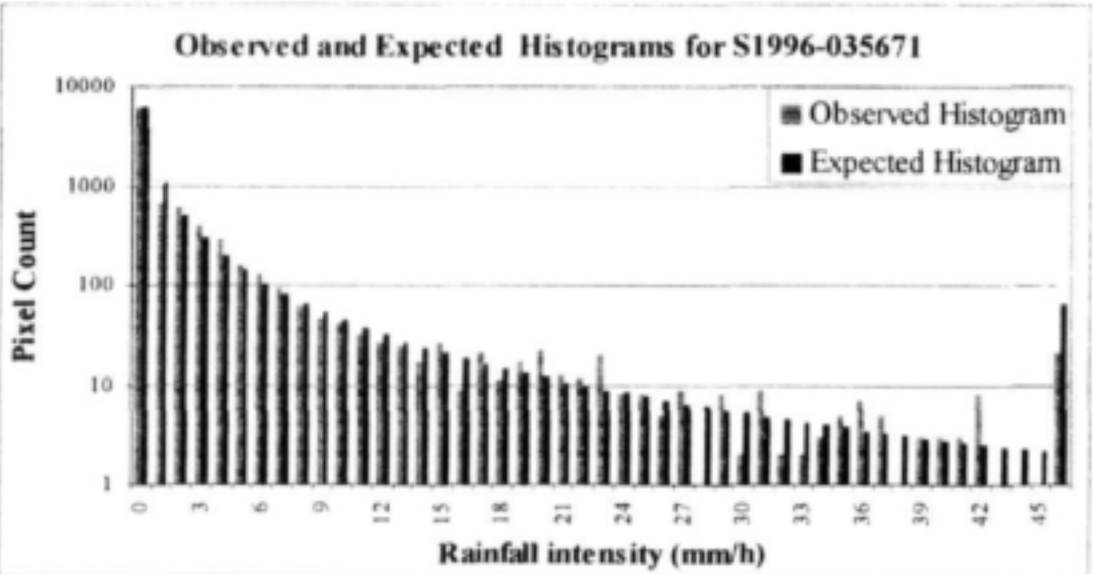


**Figure 8 - Expected and observed histograms for CAPPI S1996-035671 depicted in Figure 5c**

Figure 5d shows a *forced convective weather system*. The main convective cell is approximately 50km in diameter and the most intense rainfall experienced on the CAPPI is 75mm/h. The statistics reveal high ***m*** and low ***s*** parameters when compared to the other types of weather system depicted in Figure 5. The expected and observed histograms for this CAPPI are plotted in Figure 9.

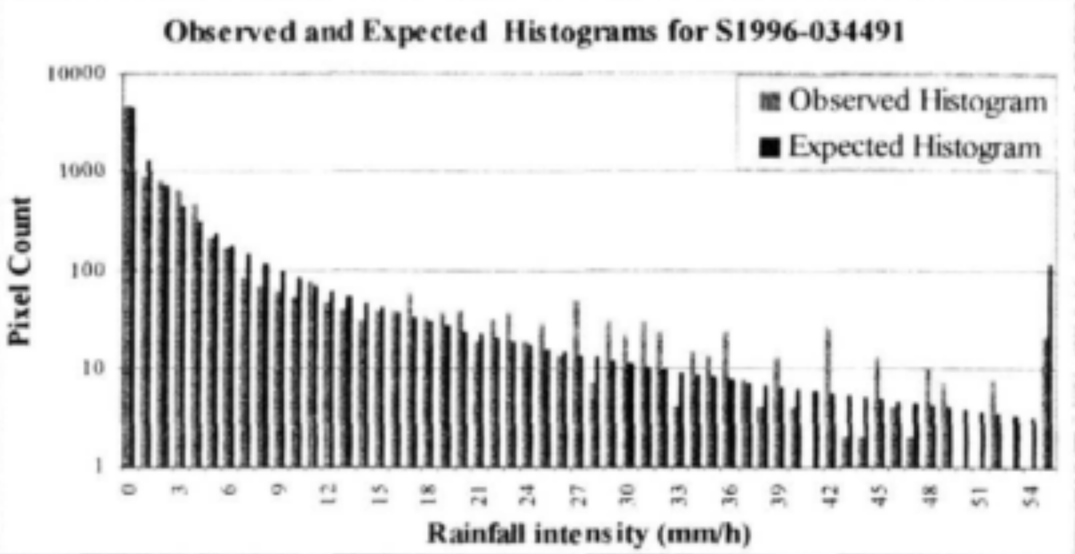


**Figure 9 - Expected and observed histograms for CAPPI S1996-034491 depicted in Figure 5d**

Figure 5e shows a typical *stratiform weather system* with widespread rainfall of low intensity. Due to almost complete coverage of the sample area with an almost constant rainfall rate, this type of rainfall gives rise to very high ***m*** and very low ***s*** values. The maximum intensity experienced on this CAPPI is 30mm/h, however less than 0.25% of the sample area experiences intensities above 15mm/h. The expected and observed histograms for this CAPPI are plotted in Figure 10.

The concentric rings which can be clearly observed in Figure 5e are due to the projection technique used to extract the CAPPI data from the volume scan data as discussed in Appendix A. The fact that the rainfall intensity appears to increase with radius (and therefore altitude) over the width of each ring also suggests the presence of the melting layer (discussed in Appendix A) at an altitude of approximately 2km.

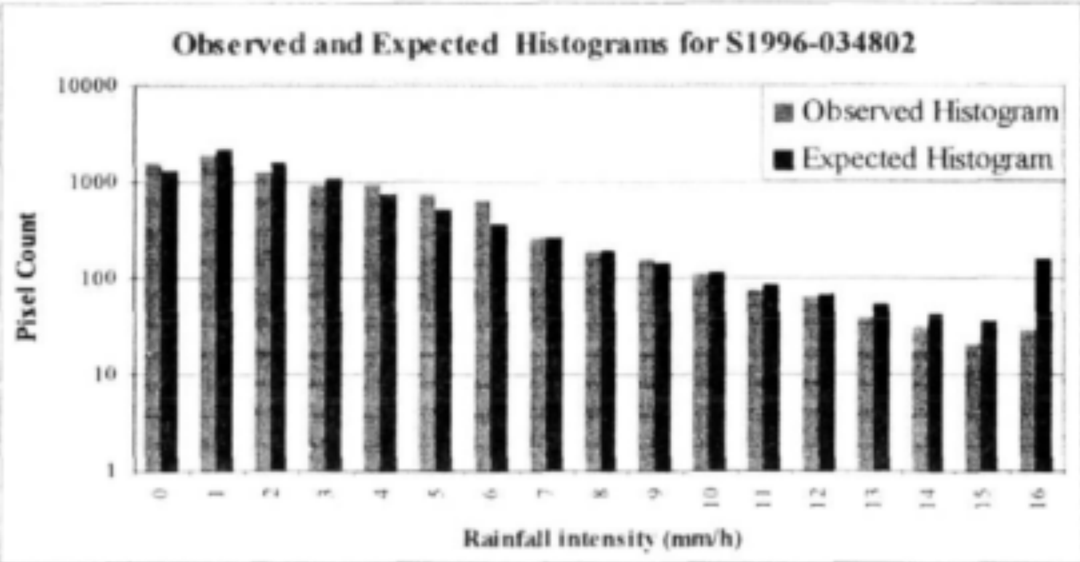


**Figure 10 - Expected and observed histograms for CAPPI S1996-034802 depicted in Figure 5e**

In all of the cases considered in Figures 6 to 10, the lognormal distribution appears to approximate the observed distribution well. Although the observed and expected histograms sometimes differ markedly for a particular intensity of rainfall at the tail end (higher rainfall intensities) it is important to note that these differences are of the order of 10 pixels out of a total of 9128. Similar fits are obtained by extracting the combined histogram of an entire sequence of CAPPIs.

### 7.2. The spatial correlation structure of CAPPI data

While the marginal distribution of the rainfall rates is described by two parameters, $\mu$ and $\sigma$ of the lognormal distribution, a third parameter $\beta_{space}$ describes the spatial correlation structure and hence the clustering behaviour of the pixels. For this the two dimensional power spectrum is used. After $\mu$ and $\sigma$ have been estimated the image is normalised using the log transform, standardised (shifted by $\mu$ and scaled by $\sigma$) and then transformed into Fourier space using the Fast Fourier Transform routine as described by Press et al (1992). The Fourier Transform (defined in Appendix C) of a real image is a complex valued image. The modulus of the complex number on each pixel is then squared to give the real valued, two dimensional power spectrum. This is illustrated in Appendix C for the one-dimensional case. For a masked CAPPI, the two dimensional power spectrum takes on the general shape shown in Figure 11.

The left hand image of Figure 11 shows a three dimensional oblique view of the power spectrum illustrating its draped nature. The right hand image serves to illustrate the skew symmetry of the spectrum which is exploited in order to find an average gradient.

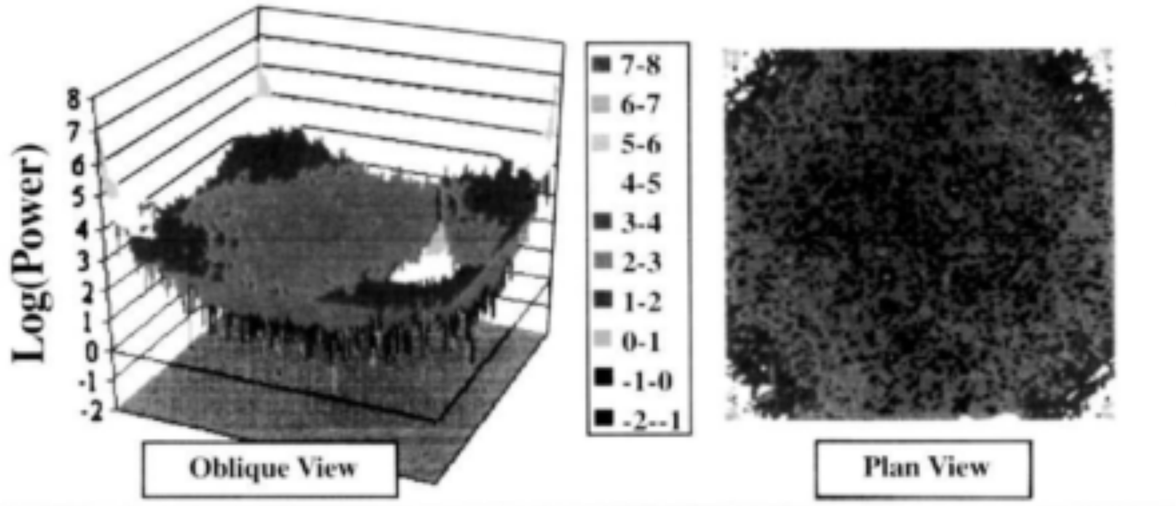


**Figure 11 - Typical two dimensional power spectrum for a masked CAPPI**

Due to this symmetry, half of the spectrum can be discarded. The remaining half is folded in half and then averaged to achieve a quarter conical shaped surface in three dimensional log space. The surface is then averaged radially, pivoting about the remaining corner, to obtain the *radially averaged, two dimensional power spectrum* an example of which is given in Figure 12. The straight line illustrates the power law relationship between the spectral density **P** and the wavenumber **w** for some constant **c**.

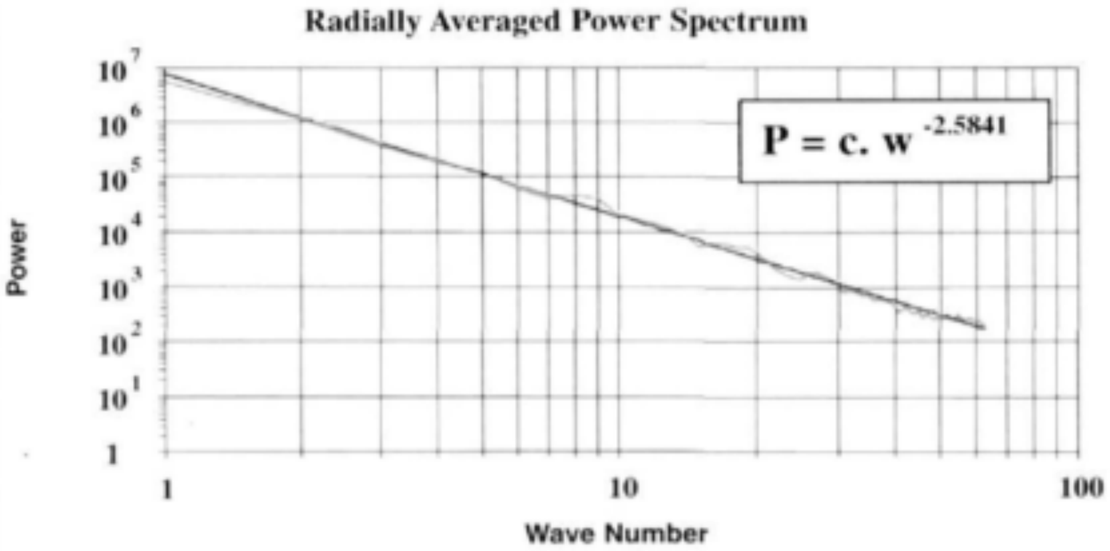


**Figure 12 - Radially averaged two dimensional power spectrum**

The gradient ($\beta_{space}$) of the radially averaged, two-dimensional power spectrum is estimated by least squares and defines the averaged spatial correlation of the CAPPI in all directions - the field is assumed to be homogeneous in this sense. This methodology was adapted from the work of Menabde et al. (1997). Typical values of $\beta_{space}$ for CAPPI data vary between 2.30 and 2.90 indicating non-stationarity in the statistics of the rainfields. As is the case in Figure 12, almost all of the CAPPIs encountered in the SAWB records are scaling in the sense that their radially averaged log-spectra are well approximated by a straight line.

With reference to Figure 5, a *forced convective weather system* as shown in Figure 5d would tend to have a relatively steep gradient (high $\beta_{space}$) due to its highly variant nature concentrating more power in the lower wave numbers. By contrast, a *stratiform* event (Figure 5e) is far less variable and will tend to have a shallow gradient (low $\beta_{space}$), that describes a more even distribution of power over the range of wave numbers. *Free convective events* (Figure 5a) have a large proportion of the CAPPI with zero rainfall making the field relatively smooth with steep "spikes" of highly variant *convective* rainfall which results in a power spectrum whose gradient is between that of the *forced convective* and *stratiform* (intermediate $\beta_{space}$ ) cases. Although some correlation between the $\beta_{space}$ parameter and the type of weather can be recognised in the data, the range of observed $\beta_{space}$ values is small and it is difficult to visually distinguish two images with the same $\mu$ and $\sigma$ that have $\beta_{space}$ values at either end of the typical range. It could therefore be argued that the average $\beta_{space}$ should be kept at a constant value for all types of CAPPI.

### 7.3. The temporal correlation structure of CAPPI data

The measurement of the exponent of the power spectrum in time ($\beta_{time}$) is not a trivial exercise as weather systems are usually moving and this has the effect of reducing the correlation in time when sampling a single pixel. The $\beta_{time}$ exponent is therefore dependent on the velocity of the weather system *and* $\beta_{space}$. Furthermore, the $\beta_{time}$ exponent is dependent on the resolution of the CAPPI, because the correlation in time for a large sample area is considerably stronger than that for a small sample area. It is therefore difficult to relate the $\beta_{time}$ parameter for a radar data set to that of a raingauge data set.

In the CAPPI data considered so far, it appears that for a spatial resolution of 1km$^2$ and at a temporal resolution of one CAPPI every 5 minutes, $\beta_{space} \cong \beta_{time}$. In other words, for a typical rainfall event sampled at this spatial and temporal resolution, a pixel appears to have a similar correlation with its neighbours in *space*, to the correlation with itself in *time*. This was hypothesised by Taylor (1938) and is discussed with reference to the space-time structure of rainfall intensity fields by Gupta and Waymire (1987) and by Crane (1990). Another qualitative observation that was made is that with a spatial resolution of 1km$^2$ and a sampling interval beyond 120 minutes two consecutive CAPPIs appear to be almost uncorrelated.

A somewhat naive, pragmatic approach to this problem was adopted for the *String of Beads* model by assuming a linear relationship which decays from $\beta_{time} = \beta_{space}$ for a temporal resolution of one CAPPI every 5 minutes, down to $\beta_{time} = 0$ for a temporal resolution of one CAPPI every 120 minutes (only applies to CAPPI data with a spatial resolution of 1km$^2$). The decision to model the time dependence in this way was an operational one based on qualitative observation of the data. Further investigation is required in order to quantify this parameter and it will be adjusted in a subsequent study once the process of extensive validation has begun. A possibly fruitful approach is that suggested by Bell (1987) who uses a Fourier series to approximate the time dependent structure by allowing the lagged correlation of the Fourier coefficients to decay exponentially in time. He achieves this by letting the coefficients satisfy Markov equations in time (Jenkins and Watts, 1968).

## 8. ANALYSIS OF A REAL EVENT

The event analysed in this section was nearly two days long and occurred in February 1996. It contributed to formidable flood event at the Vaal Dam. It was chosen for analysis due to the wide variety of rainfall types which were experienced within the 42 hour period, ranging from scattered thundershowers through to large scale stratiform rainfall with intense convective cells. Figure 13 shows the development of the estimates of the statistics of each CAPPI, $\sigma$, $\beta_{space}$ and $\mu$, in time. The fourth statistic, the Wetted Area Ratio (**WAR**), is plotted on a separate axis and represents the percentage of the sample area which received rainfall in excess of the wet/dry threshold of 1mm/h.

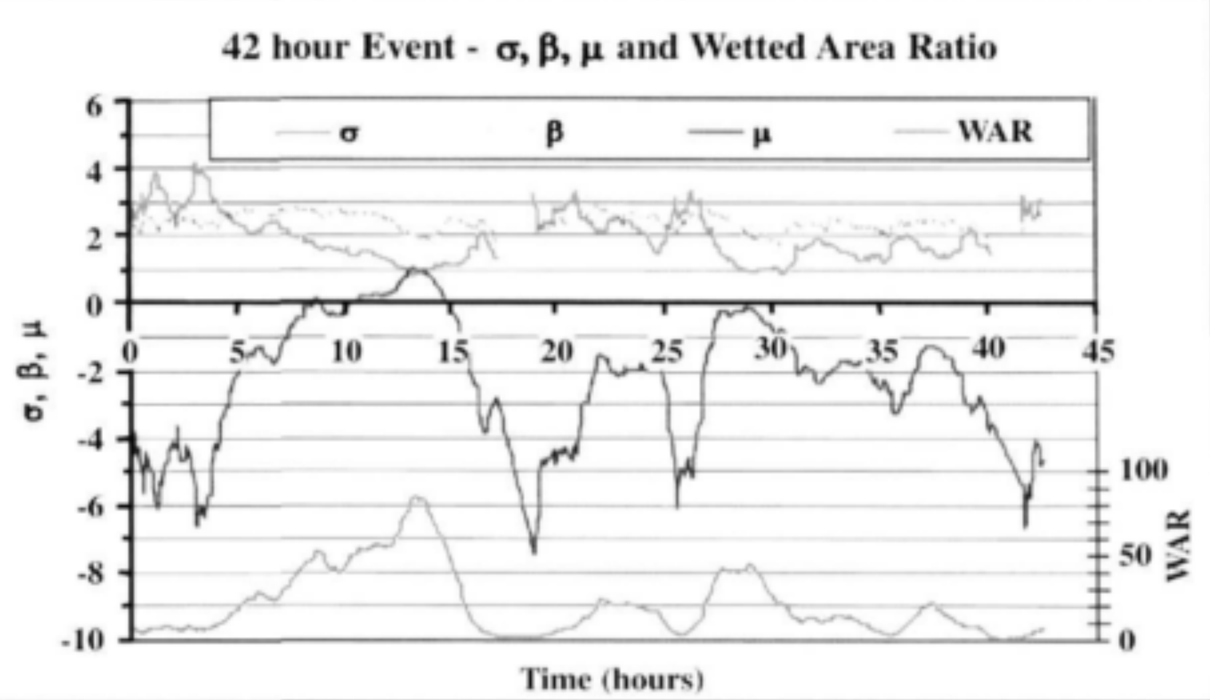


**Figure 13 - The behaviour of $\sigma$, $\beta$ and $\mu$ in time over a real 42 hour event.**

Three main points are evident in Figure 13. First, the gradient of the two dimensional power spectrum $\beta_{space}$ remains reasonably constant between 2 and 3 following the extent and type of rainfall on the CAPPI as discussed in the previous section. Second, the mean of the pixel logs ($\mu$) is very strongly correlated with the **WAR** since the logarithm of any value between 0 and 1 is negative and a CAPPI which is mostly dry (rainfall rates between 0 and 1mm/h) will have a low **WAR** and a correspondingly more negative $\mu$. Third, $\mu$ and $\sigma$ (the mean and standard deviation of the logs) are negatively correlated. This makes sense in terms of the observed behaviour of rainfall since the more general type of rainfall (*stratiform* rainfall) tends to be of a reasonably constant (low) intensity over a large area and will therefore have a low $\sigma$ and relatively high $\mu$. In contrast to this, the more scattered types of rainfall (*free convective* rainfall) tend to be highly variable in intensity (large $\sigma$) but tend to cover a small proportion of the total area (low $\mu$). The two exceptions to this behaviour are encountered with either a large *forced convective* event in which both the $\mu$ and $\sigma$ parameters are high, or alternatively with very light scattered showers (usually in the early or late stages of a *free convective* event) in which both $\mu$ and $\sigma$ are low. The negatively correlated relationship is confirmed in Figure 14 in which the sampled $\mu$ values are plotted against $\sigma$ for the 512 CAPPIs in the 42 hour event.

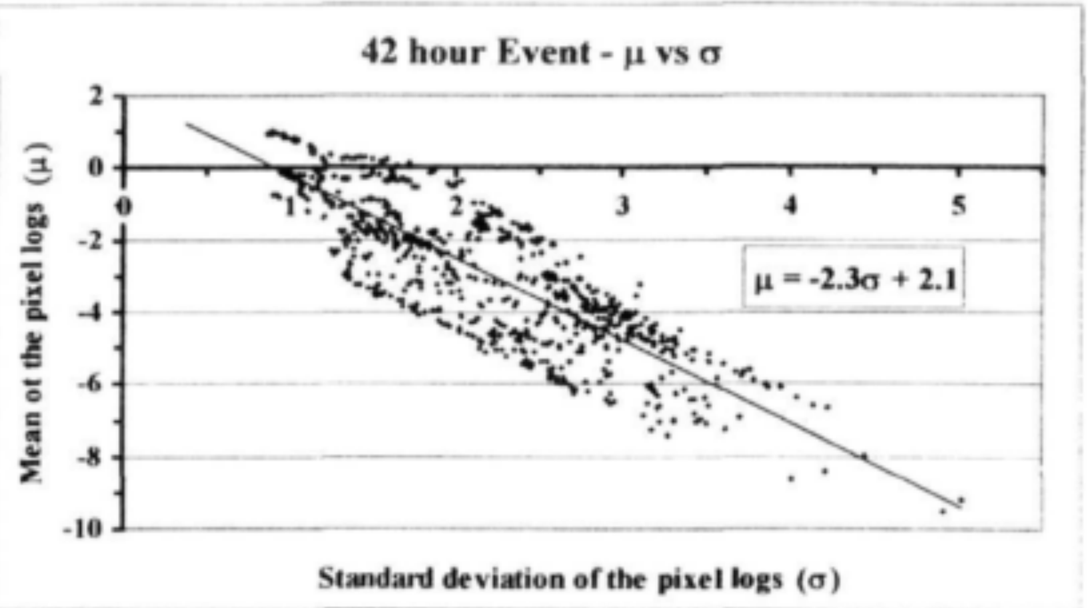


**Figure 14 - $\mu$ vs $\sigma$ for CAPPIs in a 42 hour period in February 1996**

The best fit regression line is plotted in Figure 14 and it is clear that $\mu$ and $\sigma$ fall within well defined boundaries on either side of this line. Three other major rainfall events (a total of 2500 CAPPIs) were analysed in a similar way and they all produced very similar results and characteristics to those presented for the 42 hour event analysed here. Graphs for the entire month's data, similar to that presented in Figures 13 and 14, are included in Appendix F.

## 9. MODELLING A SIMPLE RAINFALL EVENT

The process of analysis outlined in the Section 7 described a rainfall event (bead) by first breaking it down into individual masked CAPPIs and then extracting the histogram of the logs of the pixel values in order to find the best possible lognormal distribution which would describe the marginal distribution of the pixels. Each masked CAPPI was then normalised and the gradient of the radially averaged power spectrum was measured. In this way it is possible to describe a CAPPI in terms of three statistics. A simple rainfall event, that is one with a reasonably constant $\sigma$, $\beta_{space}$ and $\mu$, can be modelled by simply reversing this process starting with the desired $\sigma$, $\beta_{space}$ and $\mu$ and using a technique similar to that described by Wilson, Schertzer and Lovejoy (1991). The technique involves the power law filtering of a Gaussian noise field and is described in detail in Appendix E. For a single isolated CAPPI, the statistics can be reproduced very well in a simulated image using this technique.

Given the set of parameters $\sigma$, $\beta_{space}$ and $\mu$, the first step in the process is to generate a three dimensional field of normally distributed, real valued, random noise in a parallelepiped of size $(\mathbf{2^n \times 2^n \times 2^t})$ pixels where **n** and **t** are integer values. The integer **n** will determine the size of the rainfield and **t** will determine the number of CAPPIs that will be produced. The field is then transformed into Fourier space, power-law filtered with exponent $\beta_{space}$ in space and $\beta_{time}$ (= $c$. $\beta_{space}$ for some suitably chosen constant $c$) in time and back transformed to give a real valued three-dimensional field of correlated, normally distributed noise. The field is then scaled and shifted to achieve the required mean $\mu$ and standard deviation $\sigma$ and finally exponentiated to give a log-normally distributed field of correlated Gaussian noise with statistical descriptors $\sigma$, $\beta_{space}$, $\beta_{time}$ and $\mu$. The three-dimensional field is then "sliced up" in time into its individual CAPPIs which will each have a *unique* distribution of rainfall. These can then be assembled into an animated image after the work of Brenier (1990) and as such they simulate a *simple rainfall event* of $\mathbf{2^t}$ CAPPIs each $\mathbf{2^n} \times \mathbf{2^n}$ in size.

For general rainfall it is often not necessary to model the event to temporal resolution as high as 5 minutes, in which case it is possible to increase the time interval between CAPPIs to 15, 30 or even 60 minute intervals and to reduce the correlation in time ($\beta_{time}$ parameter) accordingly. This saves on computer memory requirements and hard disk storage space, two precious resources when working on a personal computer. A three-dimensional field of size 128 x 128 x 256 (1 day event at 5 min resolution) requires approximately 70MB of RAM to compute the double precision (8 byte) complex array.

A well known property of Fourier filtered images is their wrapped nature. This can be exploited in order to build in the advection of the field by cutting off the one edge of the image and pasting it onto its opposite edge as show in Figure 15. In this way the storm can be made to drift with any velocity provided that the distance moved by the storm between time steps is small compared to the size of the CAPPI.

The final stage of the modelling process for a *simple rainfall event* is to discard the edges so that the wrapped nature of the CAPPIs is no longer obvious. The approach suggested by Bell (1987) was to use only a quarter of the artificial image for simulation, thereby minimising the correlation between the opposite edges. However, since the filter decays exponentially from a point, it is very difficult to detect any relationship between the opposite edges of a simulated

CAPPI after discarding a border as small as $^{1}/_{16}{}^{\text{th}}$ of its width. This is illustrated in the right hand image of Figure 15.

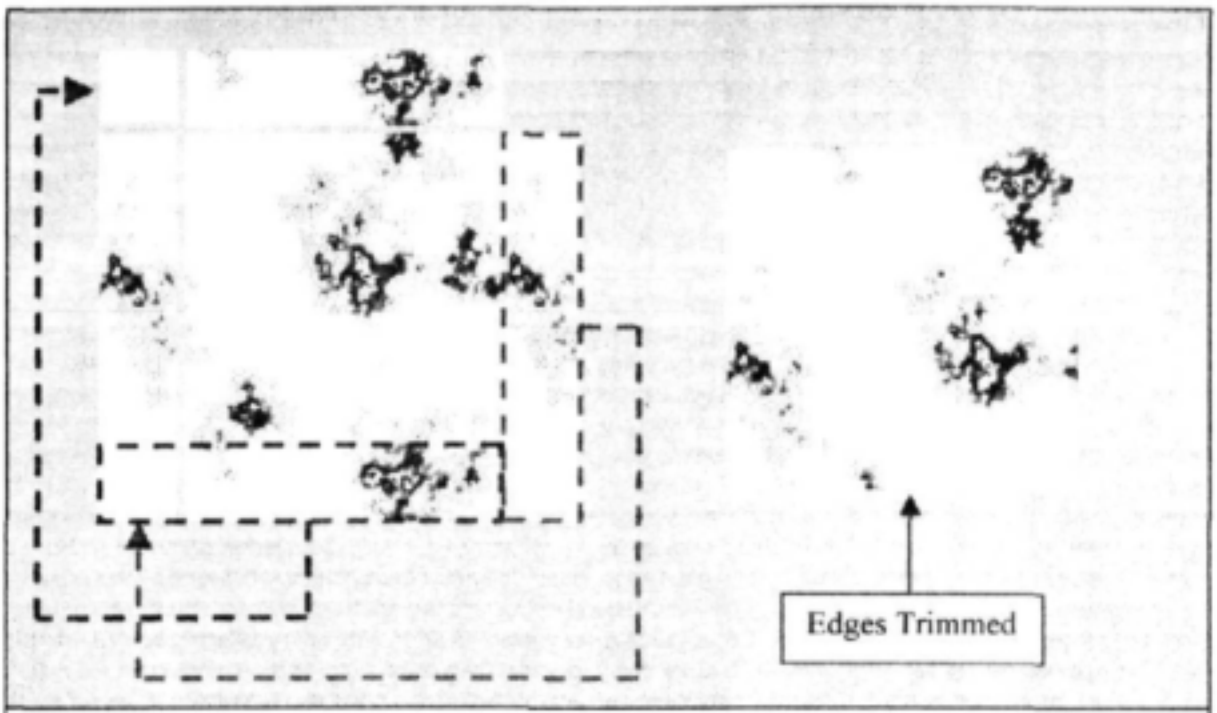


**Figure 15 - The method used to simulate storm advection. The trimmed image shows that Fourier wrapping is not obvious if the edges are discarded.**

Since the simple Fourier filtered event will also wrap in time, it is also necessary to discard the first or last few CAPPIs in the generated sequence, in order to avoid the scenario in which the first and last images of the event are too similar.

## 10. MODELLING A COMPLEX RAINFALL EVENT

In many cases the *simple rainfall event model* presented in Section 9 is all that is required when modelling monthly rainfall scenarios, but the String of Beads Model can be used to simulate far more complex events, such as the one whose sampled parameters are presented in Figure 13, where the parameters are varying over a wide range. Such an event may be modelled as a series of simple events with shifting means, which merge into each other to form one large, *complex rainfall event*. The event to be modelled is divided into discrete time blocks $\boldsymbol{i}$ of duration $\boldsymbol{T_i}$ containing values of $\mu$ which are randomly scattered about a constant value $\mu_i$ which is calculated over each time block. The corresponding standard deviation ($\sigma_i$) is then calculated from the regression line shown in Figure 14. A series of sub-events are then generated, each of duration ($\boldsymbol{T_i} + \delta$), where $\delta$ is the duration of the overlap of a sub-event with the next consecutive sub-event. The transition from one sub-event to the next (duration $\delta$) is achieved by linearly fading out the old sub-event whilst fading in the new one. This has the effect of smoothing the transition between sub-events and eliminating the abrupt change that would be observed if two independent sub-events of length $T_i$ and $T_{i-1}$ were simply placed one after the other.

The computational effort required to model a *complex rainfall event* is less than that for a large simple event of the same $n$ and $t$, because the complex event has the advantage of using a series of short Fourier transforms which can be computed very quickly. This is a direct result of the fact that the computational effort of the Fast Fourier Transform is proportional to $(n^2t).\log_2(n^2t)$, so that if a three dimensional field is sliced into $m$ blocks of length $s$ (so that $t = ms$), the effort is proportional to $(n^2t).\log_2(n^2s)$. A simulated complex event based on the statistics of the one presented in Figure 13 will be generated and analysed in Section 12.4.

## 11. MONTHLY SIMULATED RAINFALL

Combining the methodologies for modelling the seasonal daily rainfall distribution and a simple rainfall event presented in Sections 6 and 9, it is possible to simulate monthly rainfall in three dimensions. This is the essence of the String of Beads model.

As an example, consider the month of February in Bethlehem, South Africa. The appropriate probability transition matrix is shown in Figure 4 and it can be used to generate a random sequence of daily rainfall types for Bethlehem during February as described in Section 6.2. With the 3 possible states defined in Table 5.2, there are $3^{28}$ possible rainday sequences for this 28 day month, one of which could be:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G | S | S | D | S | S | S | S | S | S | S | S | D | S | G | G | G | S | G | G | G | S | S | S | G | S | S | S |

where

**G** is a *general* rainday
**S** is a *scattered* rainday
**D** is a *dry* day.

Each *scattered* rainfall event is considered independent of the events either side of it, as these tend to be the isolated convective thundershowers that occur during the hot summer afternoons. The duration of these scattered events (in hours) is modelled as an exponentially distributed random number with a mean of 1.5 hours. *Scattered* rainfall events are modelled at a fine temporal resolution of 5 minutes.

Sequential *general* rainfall events, such as the two 3-day general events in the scenario shown above, are treated as a single event, as these are large systems which persist throughout the night and into the following day. Their duration is measured in days as run length of general raindays which is automatically accounted for by the Markov chain as discussed in Section 6.3. They are modelled at a coarse temporal resolution of 30 minutes.

Since the median of the lognormal distribution is at $x = e^{\mu}$ and the **WAR** is defined as the percentage of the catchment with rainfall intensity of 1mm/h or greater, a *general* event has $\mu > 0$ (**WAR** > 50%) and a *scattered* event has $\mu < 0$ (**WAR** < 50%). The analysis of over 2500 *wet* CAPPIs for this month has shown that the marginal distribution of $\mu$ can be approximated by the trapezoidal distribution of Figure 16. Using this distribution, a random value of $\mu_i$ (where $\mu_i$ is the average of the $\mu$ values for all the CAPPIs in the event, $\mu_i > 0$ for the *general* event or $\mu_i < 0$ for the *scattered* event) is generated for each event. Its corresponding $\sigma_i$ is calculated from the relationship $\mu_i = -2.2\sigma_i + 1.4$, defined by a regression line similar to that shown in Figure 14, but calculated for 2500 CAPPIs rather than only 512 CAPPIs. Based on the analysis of these 2500 CAPPIs (refer to Appendix F), the mean value of $\beta_{space}$ for each event is randomly chosen between 2.4 and 2.7. The value for $\beta_{time}$ is taken to be equal to $\beta_{space}$ for the five minute resolution events and as 0.8 x $\beta_{space}$ for the thirty minute resolution events as discussed in Section 7.3.

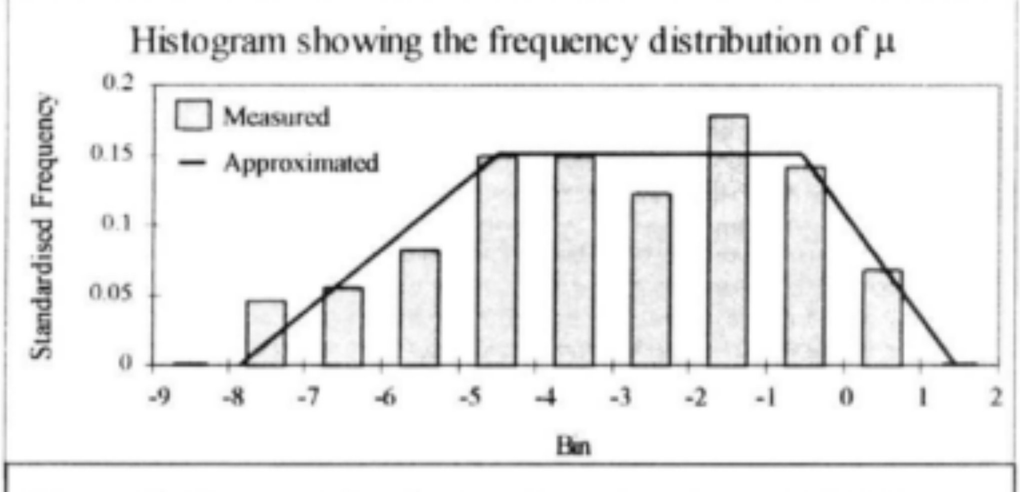


**Figure 16 - Measured distribution of statistic μ for real CAPPI data**

Once the duration and the parameters $\sigma_i$, $\beta_{space}$, $\beta_{time}$ and $\mu_i$ have been chosen for each independent event, sequences of artificial CAPPIs for each event are generated as described in Section 9. Since the number of CAPPIs in the sequence must be an integer power of 2 for the Fast Fourier Transform, the length of the bead is taken as the smallest integer power of 2 which is greater than the required event duration. Any excess CAPPIs are then discarded thereby ensuring that the sequence does not wrap in time. The first and last few CAPPIs of each independent event (approximately 10 CAPPIs for the scattered events and 15 CAPPIs for the general events) are then linearly faded to a dry state using a simple scalar multiplier in order to give the impression of a gradual change in weather. If a more refined monthly rainfall scenario is required the event *arrival* is modelled in the same way, but the *simple* rainfall events of Section 9 are replaced by the *complex* rainfall events of Section 10. The analysis of a simulated complex event is described in Section 12.4.

## 12. MODEL VERIFICATION

The purpose of model *verification* is to confirm that the statistics of the model output match those used as input. It will be shown in this section that the String of Beads model is capable of reproducing *statistically unbiased* simulations of CAPPI data which look and behave like real data. Due consideration is given to the effects of data precision and CAPPI masking on the measured statistics. Finally, a *complex simulated event* based on the event analysed in Section 8 is generated and then analysed as if it were real data. The analysis is presented in the same format as that of Section 8 to facilitate easy comparison between the real and simulated events.

### 12.1 The effect of data precision on the power spectrum

Simulated CAPPIs are generated using double precision (8 byte) pixels and then re-sampled as integer precision (1 byte) pixels to save computer hard disk space as discussed in Appendix A. The effect of this truncation of the continuous variable of rainfall intensity is to flatten the gradient of the power spectrum. This is due to the fact that when using integer data the Fourier coefficients are fitted to a stepped function rather than a smooth function and this shifts more of the total power into the high frequency coefficients.

The effect is illustrated in Figures 17 and 18 in which six CAPPIs were generated at double precision (32 bit) using the *String of Beads* model (all with $\sigma = 1$ and $\mu = 0$) with different $\beta_{space}$ varying from 0.5 to 3.0. The power spectra of these six CAPPIs were measured, first using the full 32 bit precision and then using the truncated 8 bit precision pixels. Figure 17 compares the input exponent $\beta_{in}$ to the output exponent $\beta_{out}$ for the 32 bit fields and Figure 18 compares the input exponent $\beta_{in}$ to the output exponent $\beta_{out}$ for the 8 bit fields.

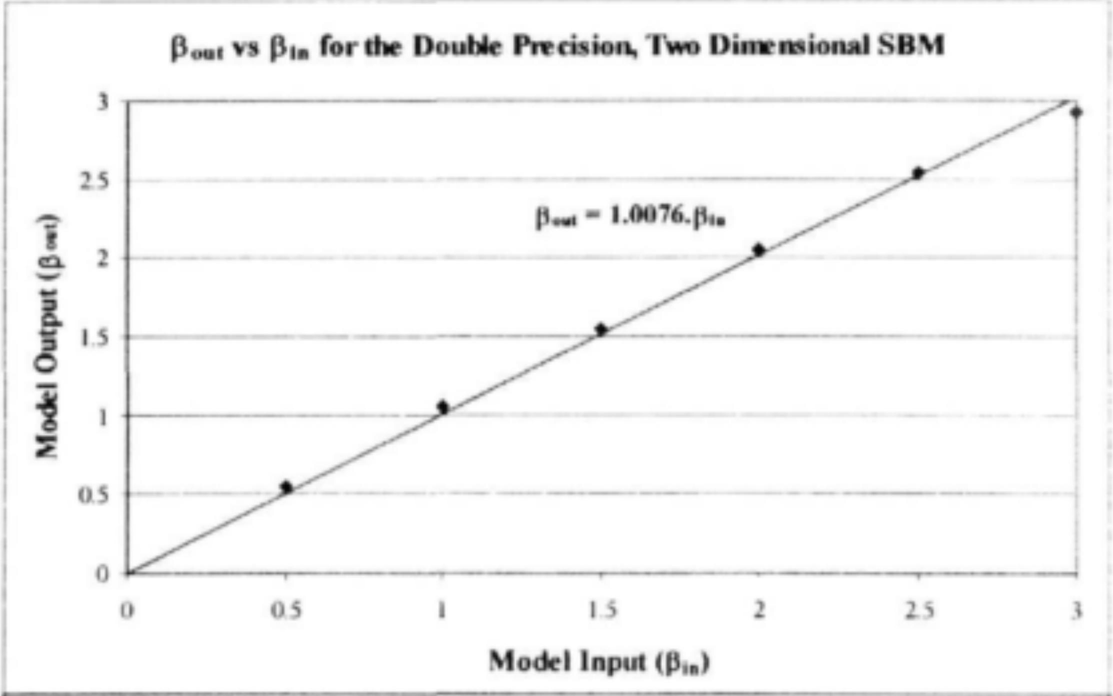


**Figure 17 - Comparison of input $\beta_{space}$ to the output $\beta_{space}$ for 32 bit fields**

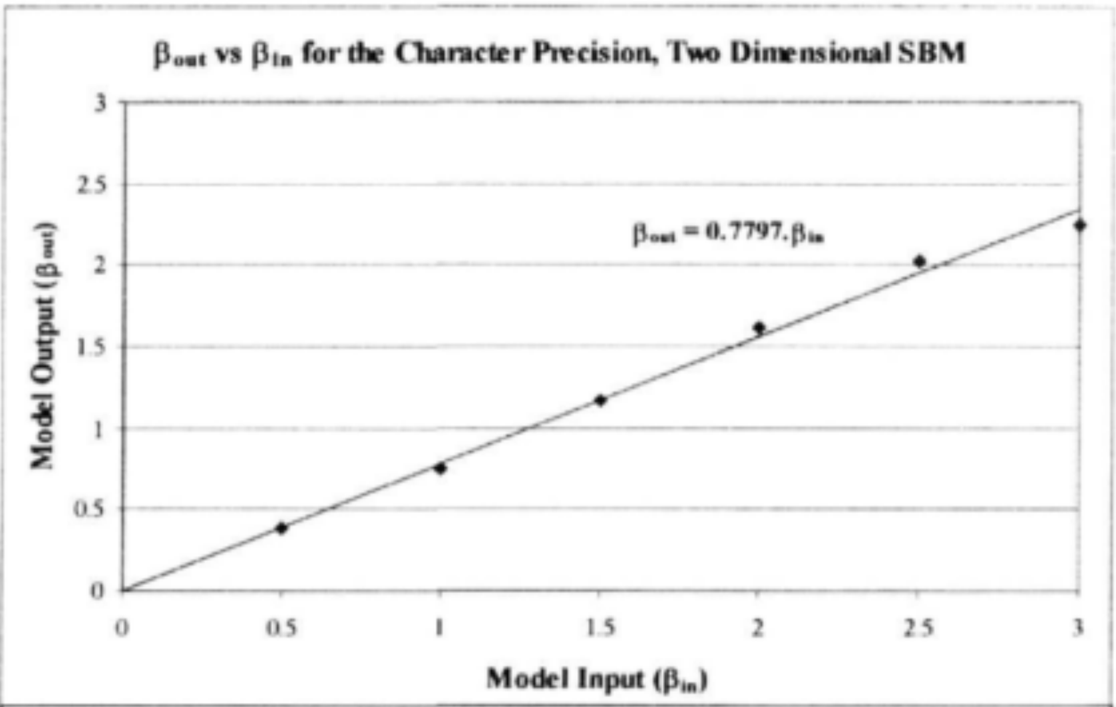


**Figure 18 - Comparison of input $\beta_{space}$ to the output $\beta_{space}$ for 8 bit fields**

The regression line of Figure 17 has a gradient very close to unity, clearly indicating that the input exponent is reproduced well for the entire range of $\beta$ in the case of the double precision simulation and there is no obvious bias when sampling the power spectrum in this case.

Figure 18 shows that when sampling the power spectrum from this simulation of an 8-bit CAPPI, the exponent of the power spectrum is biased by a factor of 0.8. Therefore it can be deduced that a $\beta_{space}$ exponent measured from an 8 bit CAPPI should be increased by 25% to obtain its equivalent 32 bit value to be used as input to the *String of Beads* model.

### 12.2 The effect of the $^{3}/_{4}$ doughnut mask on the power spectrum

In the case of real data, not all of the pixels within the 128x128 pixel square represent the rainfall rate at that pixel. Some of the image is masked out (set to zero) in order to eliminate ground clutter and data recorded above an altitude in excess of 2km above ground level as discussed in Section 5. The effect of this mask on the gradient of the power spectrum appears to be very small when considering a $\beta_{space}$ value which varies over a large range. To illustrate this point, the six CAPPIs (8 bit) used in Section 12.1 to demonstrate the effect of truncating the data to 8 bit precision, were re-analysed using the $^{3}/_{4}$ *doughnut mask*. The results of this analysis are presented in Figure 19.

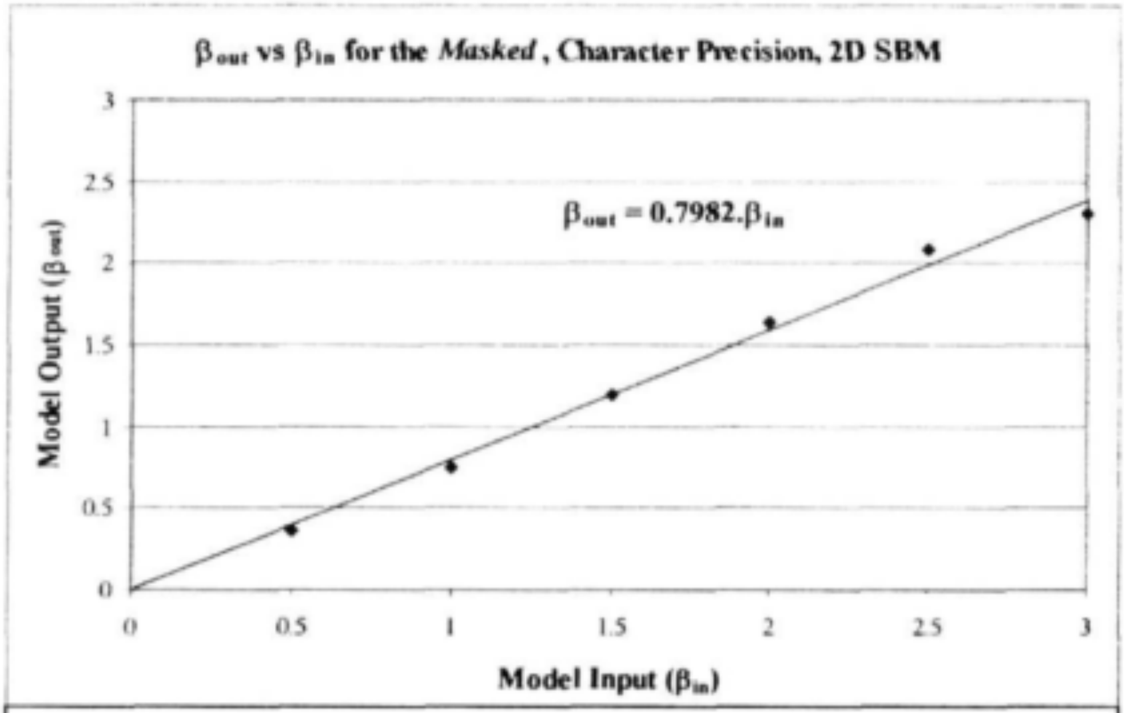


**Figure 19 - Comparison of input $\beta_{space}$ to output $\beta_{space}$ for masked 8 bit fields**

Comparing the slopes of the lines in Figures 18 and 19 there is very little difference between these two graphs. This suggests that the mask has little effect on $\beta_{space}$. In the case of measured (integer) data the $\beta_{space}$ (masked) value was found to vary between 2.0 and 3.0. Over this short range it is difficult to confirm the conclusions drawn from a comparison of Figures 18 and 19. 200 simulated CAPPIs were generated and analysed first with and then without a mask and the comparison of the estimates of the $\beta$ values is presented in Figure 20. In most cases the values of $\beta_{unmasked}$ are closely approximated by $\beta_{masked}$. The points appear to cluster evenly about the line $\beta_{masked} = \beta_{unmasked}$, thereby indicating that no obvious bias is introduced by the process of sampling the masked field.

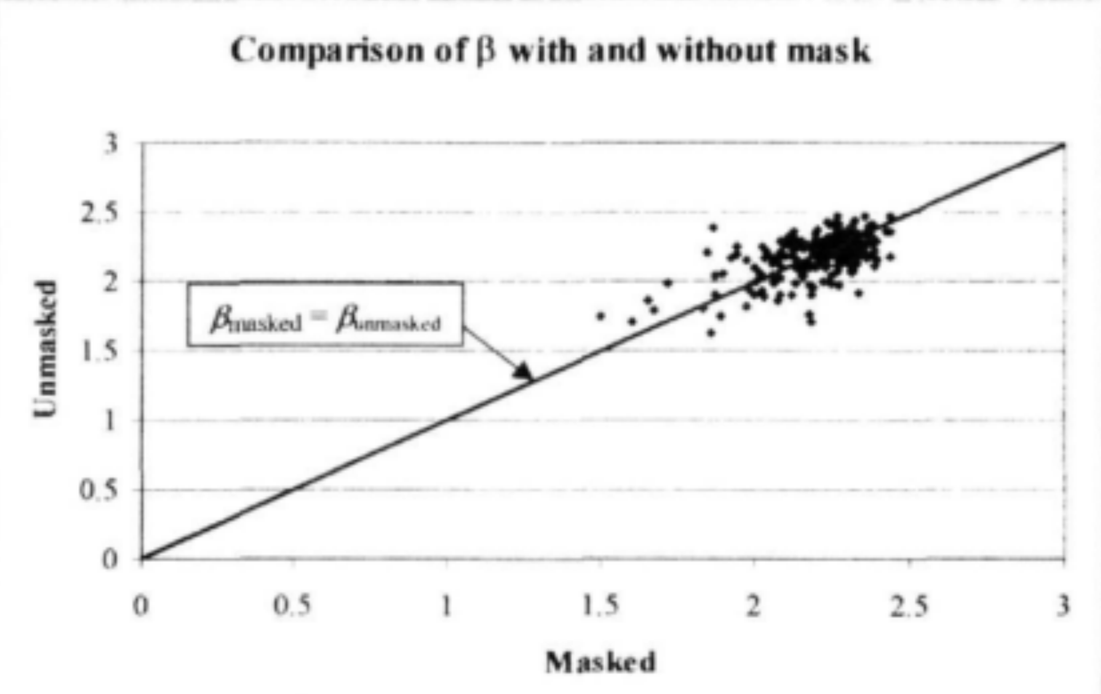


Figure 20 - $\beta_{masked}$ versus $\beta_{unmasked}$ for 200 simulated (8-bit) CAPPIs

The results presented in Sections 12.1 and this one, clearly show that although the spatial structure of the pixels (defined by $\beta$), cannot be reproduced *exactly* for an *individual* CAPPI, when considering the statistics of a *sequence* of CAPPIs, it is approximated without statistical bias by the model.

### 12.3 Effect of data precision on the verification of $\sigma$ and $\mu$

When operating the String of Beads model in double precision, the statistics of any real CAPPI can be reproduced *exactly* in a double precision simulated CAPPI. The modification of the CAPPI data to integer precision alters the marginal distribution of the pixels slightly and this causes difficulties when attempting to verify the performance of the model using the integer data. Furthermore, the parameters $\mu$ and $\sigma$ cannot be measured exactly and are only *estimates* of the parameters of the underlying continuous distribution of rainfall intensities as discussed in Section 7. In spite of this, the String of Beads model does produce CAPPIs in integer precision which can be verified to have good, statistically unbiased estimates of these parameters.

To demonstrate the unbiased nature of the model output, five individual CAPPIs were simulated using the measured estimates of $\sigma$, $\beta_{space}$ and $\mu$ obtained from the analysis of the five real CAPPIs which were presented in Figure 5 of Section 7.1. Since the original data are in masked integer format, the measured $\beta_{space}$ parameter was increased by 25% to account for the flattening of the power spectrum gradient as discussed in Sections 12.1 and 12.2. The simulated CAPPIs are shown in Figure 21 for qualitative comparison with those of Figure 5.

Qualitatively, the simulated CAPPIs of Figure 21 all appear to have a similar nature to their corresponding real CAPPIs of Figure 5. Real and Simulated CAPPIs (integer precision) were analysed using the same method and a comparison of the parameters of the two sets is given in Table 4.

**Table 4 - Verification of $\sigma$, $\beta_{space}$ and $\mu$ parameters estimated from the simulated integer precision CAPPIs of Figure 21.**

| Simulated CAPPI | $\sigma$ | | $\beta_{space}$ | | $\mu$ | |
|---|---|---|---|---|---|---|
| | Input | Output | Input | Output | Input | Output |
| Figure 21a | 2.669 | 2.633 | 2.073 | 2.146 | -4.085 | -3.939 |
| Figure 21b | 2.174 | 2.466 | 2.361 | 2.336 | -1.673 | -2.009 |
| Figure 21c | 1.928 | 2.305 | 2.567 | 2.467 | -0.868 | -1.249 |
| Figure 21d | 1.804 | 1.934 | 2.618 | 2.819 | 0.014 | -0.046 |
| Figure 21e | 0.894 | 0.885 | 2.122 | 2.117 | 0.944 | 0.941 |

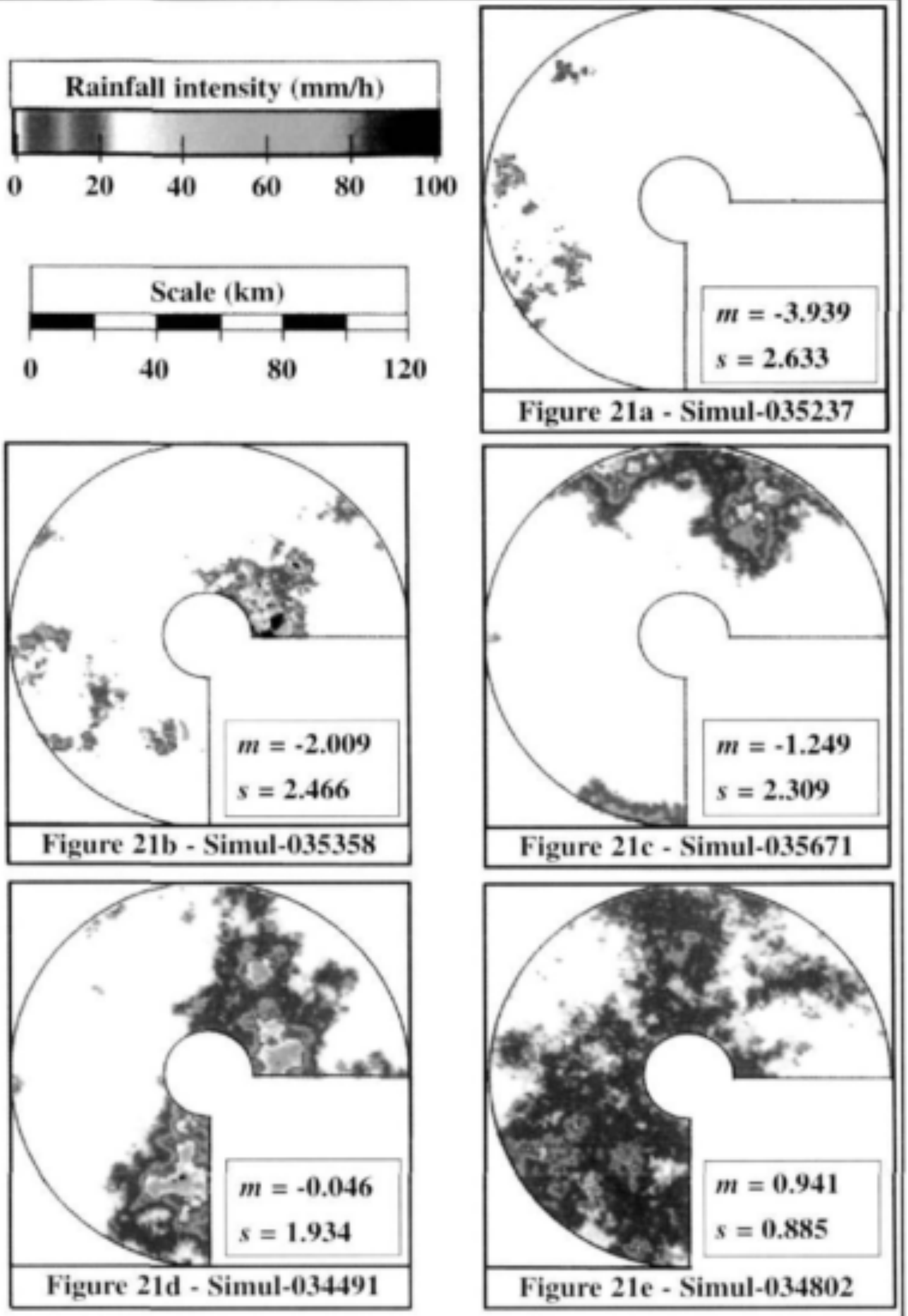


Figure 21a - Simul-035237

Figure 21b - Simul-035358

Figure 21c - Simul-035671

Figure 21d - Simul-034491

Figure 21e - Simul-034802

**Figure 21 - Simulated CAPPIs generated using the statistics estimated from the real CAPPIs presented in Figure 5**

A point of interest in Table 4 is the accuracy with which the String of Beads model reproduces the $\beta_{space}$ parameter of integer data when the input is increased by 25% as discussed previously. In most cases the $\mu$ and $\sigma$ parameters are reproduced reasonably well when using integer precision and this is confirmed in Figure 22 in which the relationship between the model input is compared to its output and a least squares approximation is given for each of the three parameters.

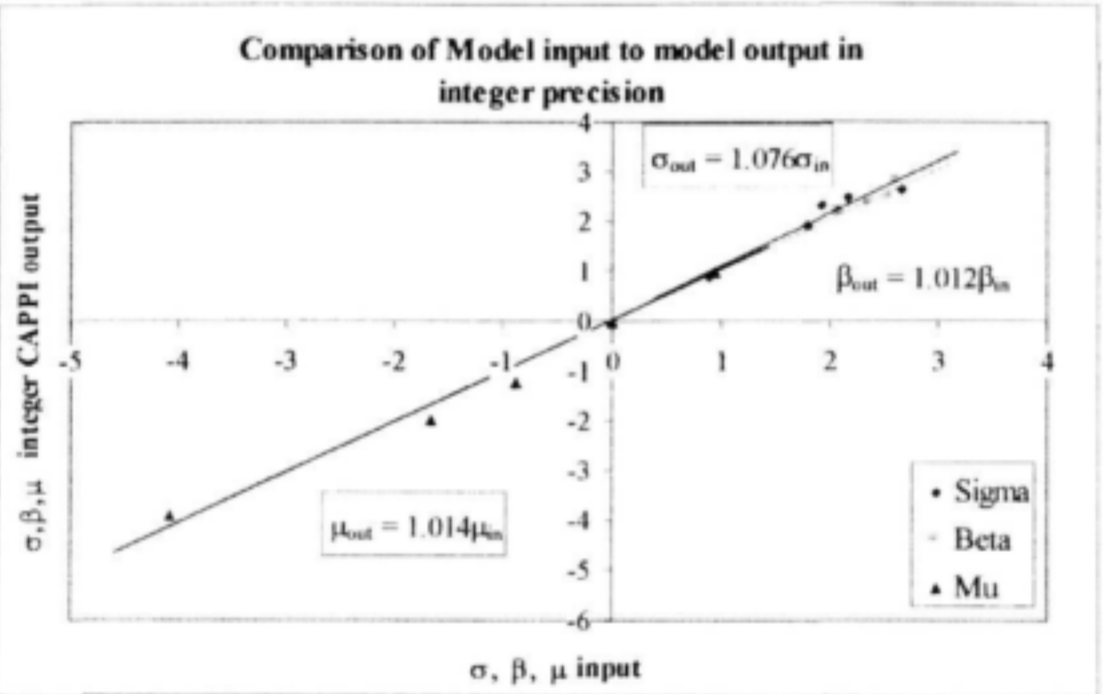


**Figure 22 - Verification of the parameters of the String of Beads model as applied to integer precision CAPPI data**

In all three cases the regression lines plotted in Figure 22 have nearly unit gradient thereby demonstrating the unbiased nature of the output of the String of Beads model. Figure 22 also serves to illustrate the typical range of values that is assumed by each parameter.

It is perhaps worth repeating at this point that the original five CAPPIs of Figure 5 were selected from the radar data set as representative of the wide variety of rainfall types observed by the MRL5 radar at Bethlehem. The quality of the images of the five CAPPIs modelled in Figure 21 is a typical representation of what can be expected from the String of Beads model.

Sections 12.1 through 12.3 have addressed the main problems encountered during the verification of the String of Beads model. The causes of these problems have been identified and wherever possible a solution has been proposed and tested. With these points in mind, a 42 hour simulated complex event was generated and the results of the analysis are presented in Section 12.4.

### 12.4 Analysis of a simulated complex event

Based on the analysis of the 42 hour rainfall event observed in February 1996 and presented in Section 8, a simulated *complex* event was generated as described in the Section 10 and then analysed to extract its statistics as a means of model verification. For the simulation, a constant discrete time block $T = 2$ hours 40 minutes (32 consecutive CAPPIs) was adopted for each sub-event with an overlap of $\delta = 1$ hour 20 minutes (16 CAPPIs). The average mean for each sub-event $\mu_i$ was calculated over time intervals of 1 hour 20 minutes from Figure 13, and the corresponding standard deviation $\sigma_i$ for each sub-event was calculated from the regression line given in Figure 14.

The resulting sequence of simulated CAPPIs is included with this document as an animated image on compact disk (refer to Appendix G) which can be compared to a similar animated image of the real event as seen by the MRL5 radar in Bethlehem. Some CAPPIs have been taken directly from the two sequences at time intervals of 4 hours and these are presented in Figures 23 (a), (b) and (c).

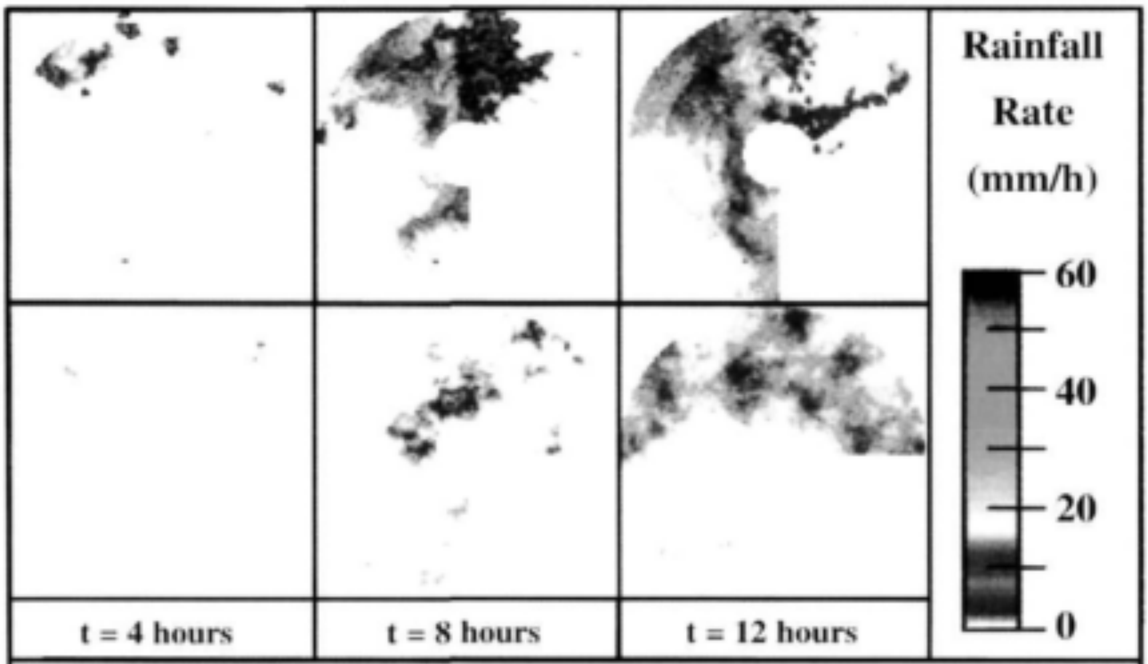


**Figure 23(a) - Selected CAPPIs from the real sequence (top) analysed in Section 8 and the simulated sequence (bottom), at 4 hour intervals**

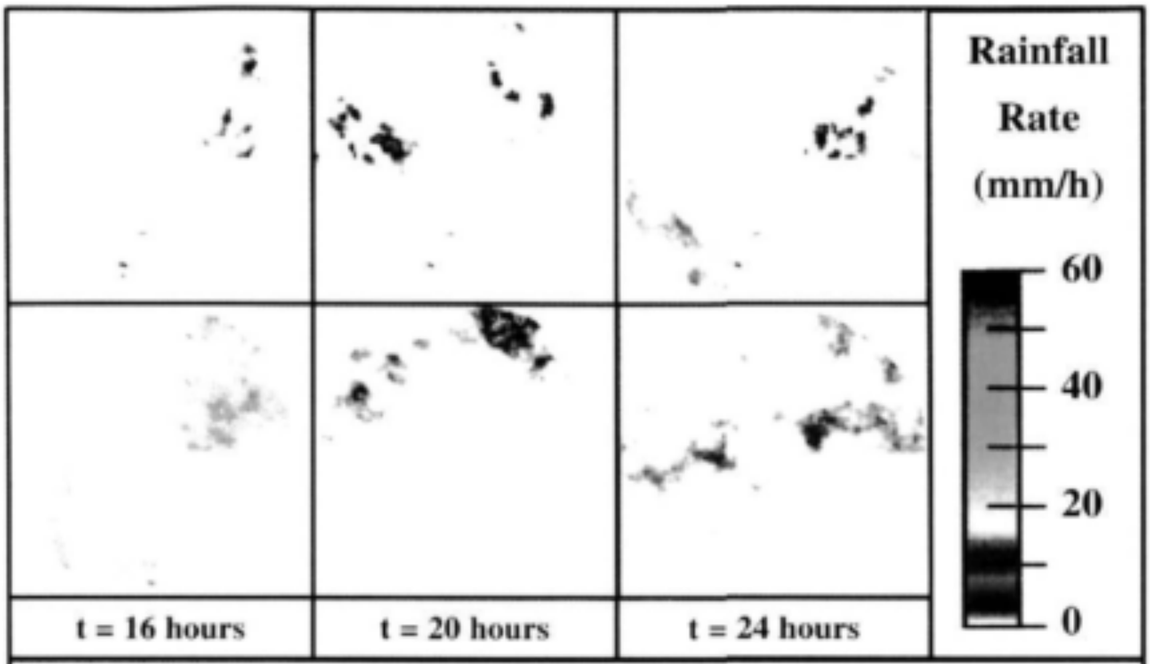


**Figure 23(b) - Selected CAPPIs from the real sequence (top) analysed in Section 8 and the simulated sequence (bottom), at 4 hour intervals**

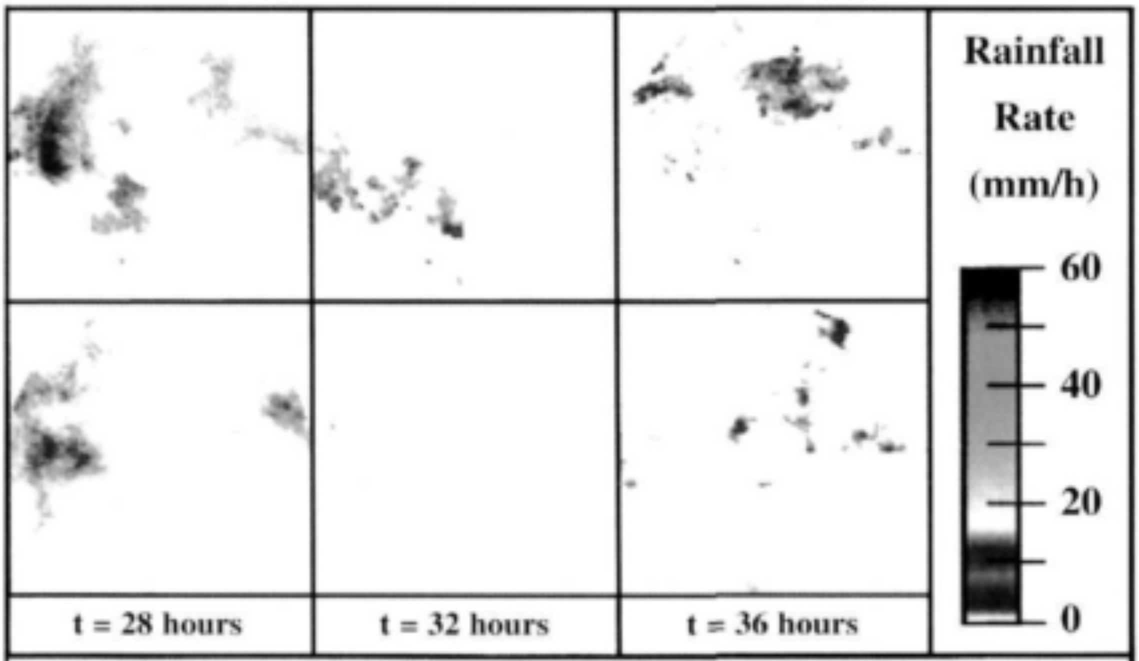


**Figure 23(c) - Selected CAPPIs from the real sequence (top) analysed in Section 8 and the simulated sequence (bottom), at 4 hour intervals**

Figures 23(a), (b) and (c) show how the storm builds from nothing at t = 0 (not shown) through scattered showers and peaks at t = 12h with general, light rainfall. This decays into scattered, isolated rainfall within 4 hours (t = 16h) and then gradually builds to a more general event at t = 28h before decaying again to nothing at t = 40h (again not shown).

This sequence of rainfall events also illustrates the ability of the String of Beads Model to model different kinds of rainfall. For both scattered and general rainfall the simulated images have a realistic distribution of rain. Storm cells are of reasonable size and random shape and the wet/dry interface has the right "roughness". When seen as an animated sequence, the storms move, grow and decay in a realistic manner. The most noticeable difference in the characteristics of the real and simulated images are the "flaws" observed in the real data such as the concentric rings which can be observed at time steps t = 12 and t = 28. This is an artefact of using the projection technique of CAPPI calculation as discussed in Appendix A.

Figure 24 shows a typical example of a single real, double precision CAPPI and its simulated partner. The real CAPPI was analysed to obtain the three statistics $\sigma$, $\beta_{space}$ and $\mu$ which were used as input for the String of Beads model to produce the simulated image which was then analysed in the same manner.

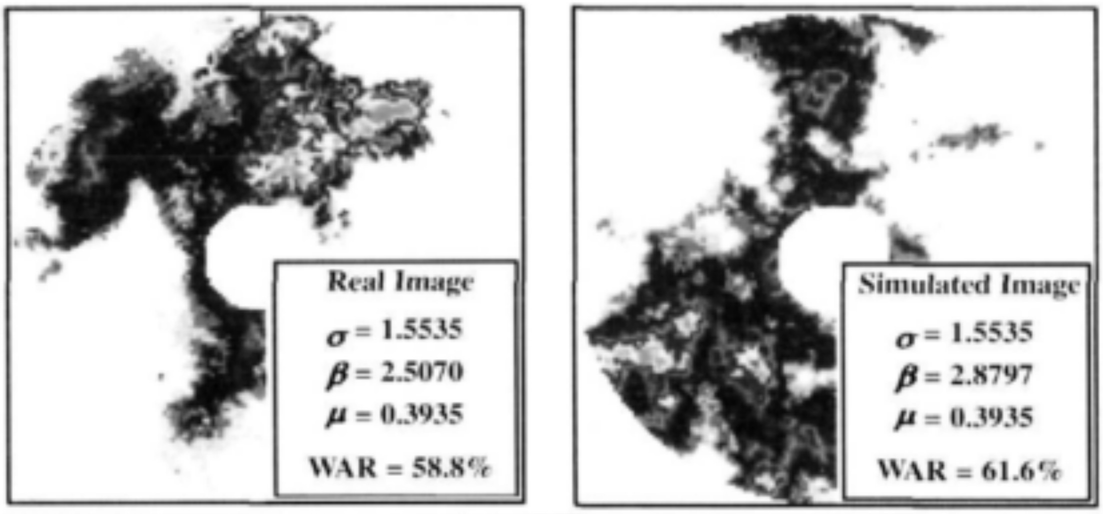


**Figure 24 - Comparison of a real CAPPI to its simulated equivalent**

The marginal distributions of the two images are identical through the reverse of the process of standardisation as is the case with all CAPPIs simulated in double precision. The wetted area ratios of the two images are very similar, as would be expected for two images with such similar marginal distributions. There is a small difference in the gradients $\beta_{space}$ of the radially averaged power spectra. This gradient cannot be matched exactly for simulated CAPPIs due to the effects of data precision and CAPPI masking as discussed in Sections 12.1 and 12.2 respectively.

When compared to the statistical signature of the real event of Figure 13 and 14, similar trends can be observed in the signature of the simulated event shown in Figures 25 and 26. Apart from a few isolated CAPPIs, the statistics $\sigma$, $\beta_{space}$ and $\mu$ all have a similar range to the real event and their trends and relationships to each other are reproduced reasonably well. Due to the fact that each sub-event has statistics chosen independently of its neighbours, discontinuities are clearly evident in the plot of the Wetted Area Ratio for the simulated event. This could be improved by more smoothing of the short duration (1 hour 20 minute) sub-events and in the extreme case, the statistics of each CAPPI could be reproduced individually and smoothed to give an almost identical set of curves to those produced in Figure 13.

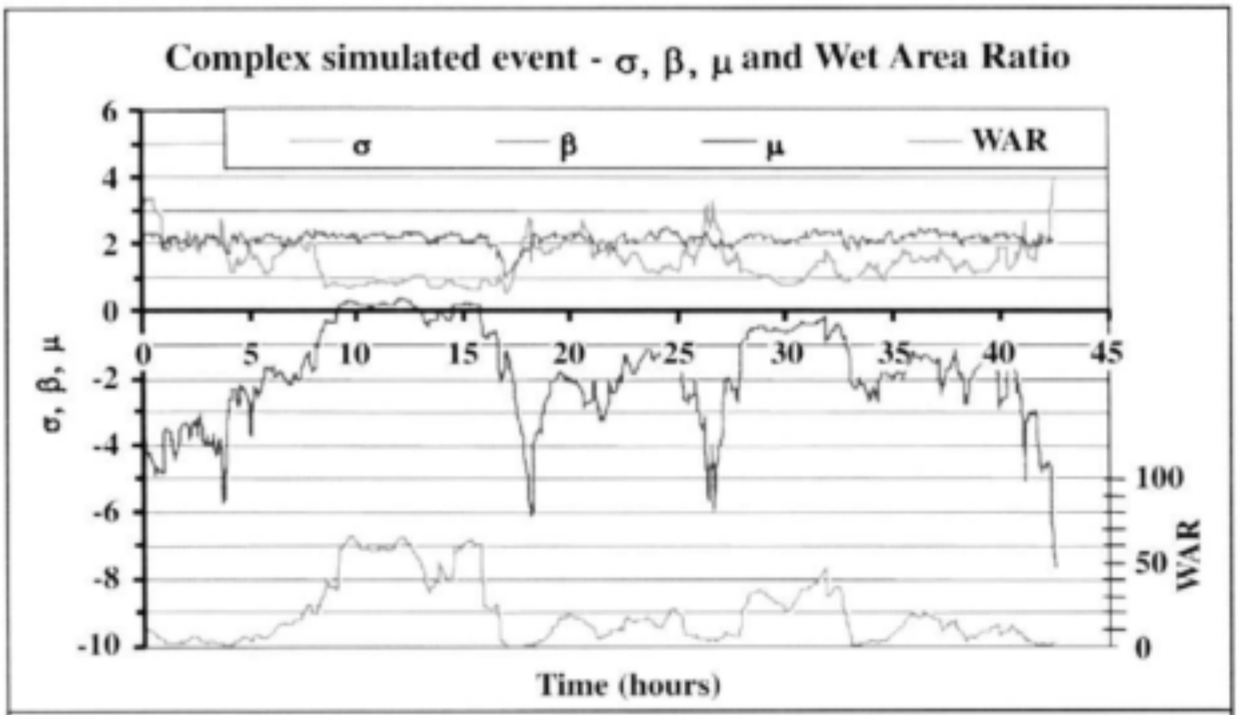


**Figure 25 - The behaviour of $\sigma$, $\beta_{space}$ and $\mu$ in time over a simulated 42 hour event. To be compared to that of the real event of Figure 13**

The line plotted on Figure 26 is the regression line of Figure 14. Comparing the two series of Figure 26, $\mu$ and $\sigma$ of the simulated event are more tightly grouped around the regression line than is observed in the real event. There was no noise added to $\sigma$ when computing it from the regression line, any variability is the direct consequence of the randomisation procedure used in generating white noise in the bead. There are more CAPPIs with relatively low $\sigma$ and high $\mu$ in the simulated event. These differences are a result of taking an average $\mu$ for each sub-event used in the simulation and thereby averaging out CAPPIs of extreme $\mu$. By reducing the duration of the sub-events the simulated event could be forced to resemble the real event more closely. An alternative would be to fit a time series to the means process although this might be an over elaborate route to follow.

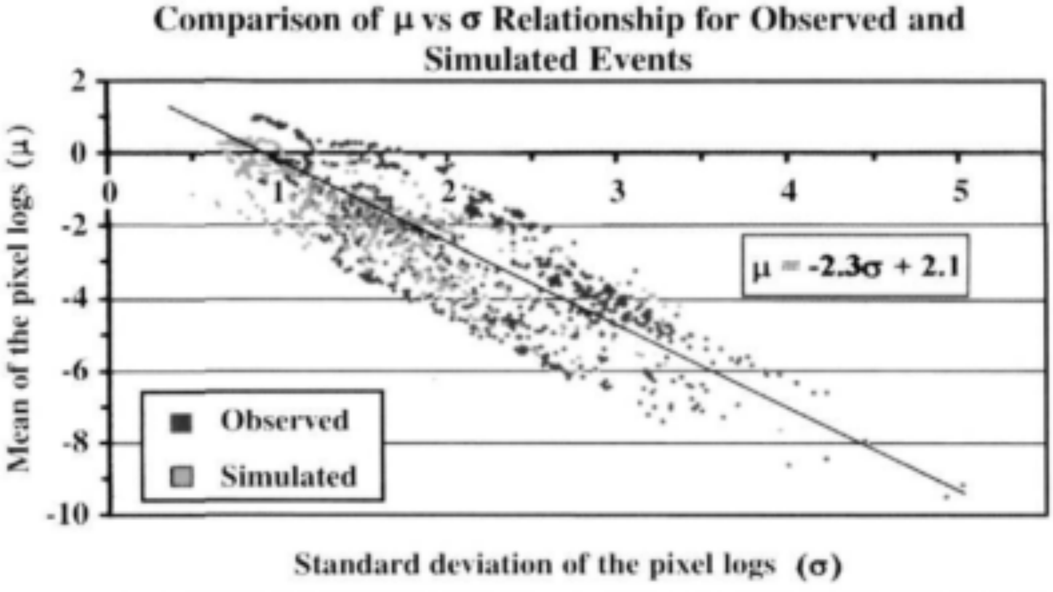


**Figure 26 - μ vs σ for CAPPIs in a simulated 42 hour event. To be compared to the real event of Figure 14**

The analyses presented in Section 12.4 verify the ability of the String of Beads model to reproduce statistically unbiased CAPPIs with any required $\sigma$, $\beta_{space}$ and $\mu$. The simulated CAPPIs have been demonstrated to be realistic in appearance and when considered as a sequence in time their statistics behave in a similar manner to that of real CAPPIs. What remains is to validate the output of the model using an independent test and this process is described in Section 13.

## 13. SIMULATED CAPPI VALIDATION

Unlike the two dimensional power spectrum and the lognormal distributions, the Generalised Structure Function was not used to determine the input parameters of the String of Beads Model, and therefore serves as a useful validation test of the simulated CAPPIs. It was developed for use in the field of fluid turbulence and has become a widely used technique for multifractal fields since the work of Anselmet et al. (1984). Its use is appropriate in the analysis of a non-stationary field of dimension $D$ where $\beta > D$. When $D = 2$ dimensions (as is the case in individual CAPPIs), $\beta > 2$ indicates that ordinary correlation analysis is inappropriate. Almost without exception, the values of $\beta$ estimated from real data were greater than 2 as seen in Figures 13 and 25.

Since the *Generalised Structure Function* only makes use of the data sampled from within the $^3/_4$ doughnut shaped sample area for its calculation (as explained in Appendix C) in both the real and simulated cases, the mask has no effect when comparing the GSF of two CAPPIs. The precision of the data however, does effect the shape of the GSF and this will be addressed in Section 13.1 before using it to validate the spatial structure of the CAPPI data in Section 13.2.

## 13.1 The effect of data precision on the Generalised Structure Function

Most of the CAPPI data analysed in this study were supplied at 8 bit precision and could assume any integer value of rainfall intensity between 0 and 255 mm/h. This presents a problem when calculating the Generalised Structure Function. The Generalised Structure Function $G_q(l)$ is defined in Equation 2 on a non-negative scalar random field R(x) as being the ensemble average of the $q^{th}$ power of the absolute differences of the $R$ values a distance $l$ apart. That is

$$G_q(l) = \left\langle \left| R(\vec{x} + \vec{l}) - R(\vec{x}) \right|^q \right\rangle \tag{2}$$

Equation 2 suggests that for a continuous variable $\boldsymbol{R(x)}$, the absolute difference between any two pixel intensities can assume any real, positive value $\mathbf{z}$. That is

$$\left| R(x + l) - R(x) \right| = \mathbf{z}, \qquad \mathbf{z} \in \Re,\ \mathbf{z} \geq 0. \tag{3}$$

In the case of *double precision data*, the $\mathbf{z}$ values of Equation 3 assume small *non-zero* values for small values of $\boldsymbol{l}$, and this leads to a small $\boldsymbol{G_q(l)}$. As $\boldsymbol{q}$ tends to zero, $\boldsymbol{G_q(l)}$ tends to unity and the function gradient $\boldsymbol{\zeta(q)}$ tends gradually to zero. That is to say that

$$\text{if } z \neq 0,\ \lim_{q \to 0} z^q = 1,$$

$$\therefore \lim_{q \to 0} \left\langle z^q \right\rangle = \lim_{q \to 0} G_q(l) = 1 \tag{4}$$

$$\therefore \lim_{q \to 0} \varsigma(q) = 0 \tag{5}$$

Conversely, as $\boldsymbol{l}$ increases, larger absolute differences will be encountered and the $\boldsymbol{G_q(l)}$ will gradually increase for any given $\boldsymbol{q} \neq 0$. A higher power of $\boldsymbol{q}$ will lead to a greater $\boldsymbol{G_q(l)}$ and therefore a steeper function gradient $\boldsymbol{\zeta(q)}$.

This is *not* the case when $R(x)$ can only assume *integer* values. The absolute difference between two integers must also be an integer value. This introduces zero values of **z** and the shape of the structure function is different from when it is calculated with full precision. Figures 27 and 28 show the gradients of the structure function for a CAPPI. Figure 27 was calculated for a *double precision* CAPPI and Figure 28 for the same CAPPI stored in *integer precision*.

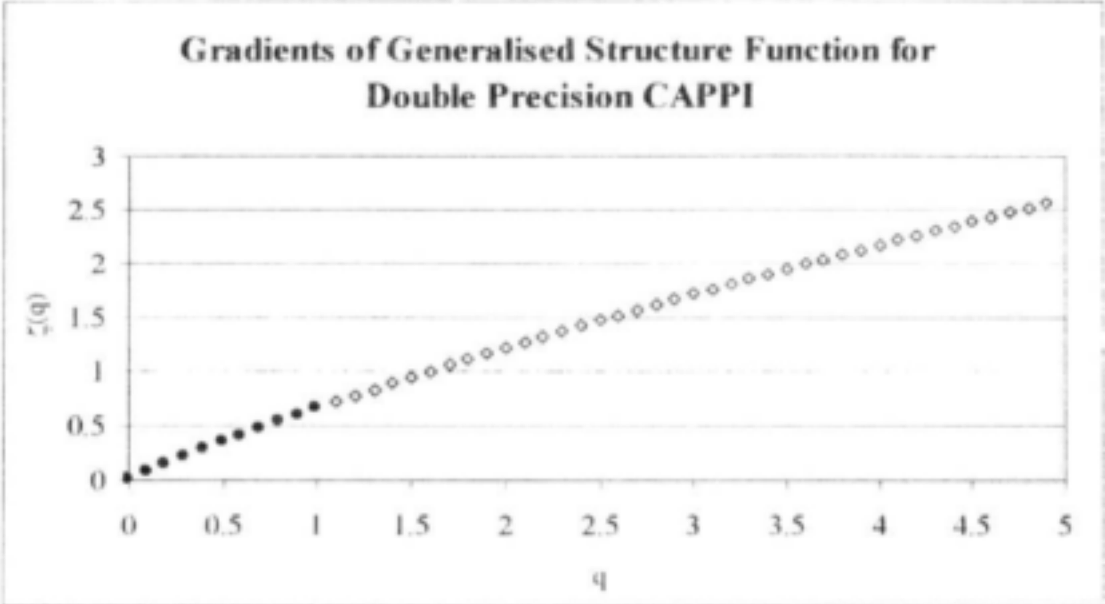


**Figure 27 - Gradients of the GSF for a double precision CAPPI**

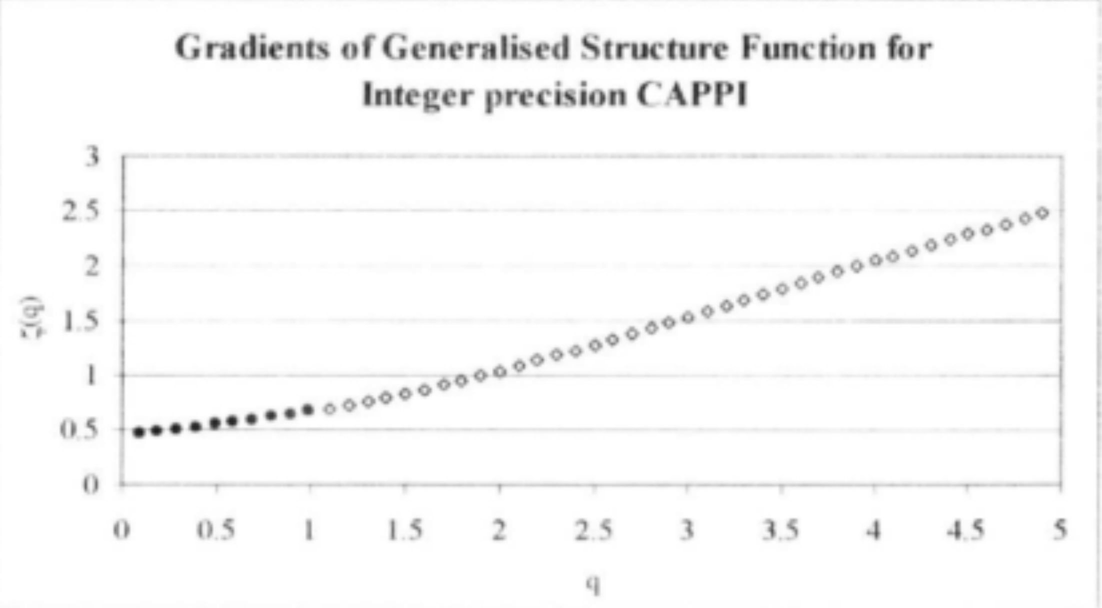


**Figure 28 - Gradients of the GSF for an integer precision CAPPI to be compared with Figure 27**

The explanation for the difference lies in the fact that the distribution of rainfall on a CAPPI is severely skewed towards the lower values (mostly below 2mm/h as shown in 7.1) and it stands to reason that the distribution of the *absolute differences* of rainfall intensities on the CAPPI will tend to be even more skewed towards the lower values. In addition, the sampled correlation of the intensity of a pixel with its closest neighbour is very high. The combination of these two conditions leads to the fact that for low values of ***l*** it is rare to observe an *absolute difference* of rainfall intensity in excess of 1mm/h. However, as ***l*** increases, the correlation between the two pixel intensities is reduced and there is a better chance of encountering a larger absolute difference between them. This is illustrated in Figure 29.

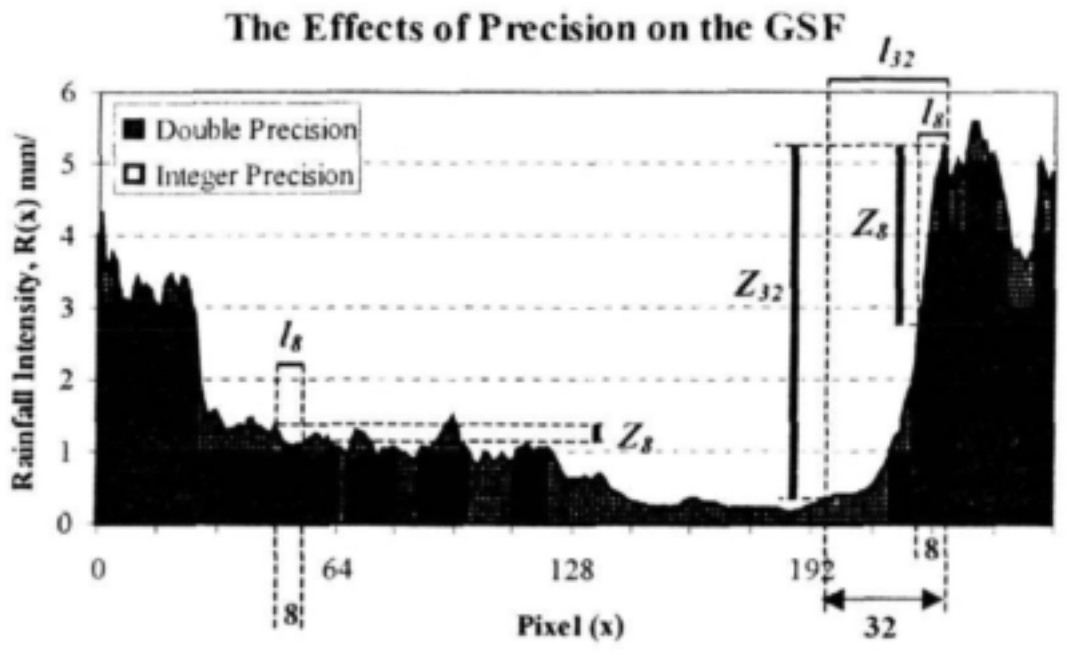


**Figure 29 - Choice of larger *l* corresponds to a greater chance of observing a large absolute difference (z)**

The horizontal bars in Figure 29 represent ***l*** values of 8 and 32 used to calculate the GSF in the simple one-dimensional example plotted in double precision as the dark grey series, and in integer precision as the light grey series. The vertical bars represent the corresponding $z$ values for the double precision case. Even on the steepest part of the curve, the largest value of $z$ that is experienced for an ***l*** value of 8 is approximately 2.5. On the same section of the curve a $z$ value of 4.9 can be obtained when using an ***l*** value of 32.

When working with integer values of rainfall intensity on the same series of data (plotted as the light grey series in Figure 29), a large proportion of the observed absolute differences (**z**) are zero. For a sample of ***N*** points containing ***m*** zero values

$$\lim_{q \to 0} \left\langle z^q \right\rangle = \lim_{q \to 0} G_q(l) = \frac{(N - m)}{N} \qquad (6)$$

since $0^q = 0$ and $1^q = 1$ for all ***q*** (defining $0^0 = 0$), and the ratio ***( N - m ) / N*** is strictly less than one. In the case where ***m*** = 0, Equation 6 reduces to Equation 4.

With reference to Equation 3, the smaller the value of ***l***, the more **z** = 0 will be observed and consequently, the smaller will be the limit of Equation 6. Conversely, a larger value of ***l*** will produce fewer **z** = 0 and therefore a higher ***$G_q(l)$*** will be calculated. This shows that when analysing data in which the absolute differences **z** *can* assume a value of zero, ***$G_q(l)$*** will increase with increasing ***l*** even when ***q*** = 0. The only cases in which ***ζ(0)*** can be zero are if the rainfield is either entirely zero or entirely non-zero. The result of this is that when calculating the structure function for integer precision CAPPI data, the gradient ***ζ(0)*** will be finite and positive as shown in Figure 28.

### 13.2 Comparison of the GSF for real and simulated CAPPIs

The *ζ (q) ~ q* functions of the real and simulated *double precision* CAPPIs which were shown in Figure 24, are given in Figure 30.

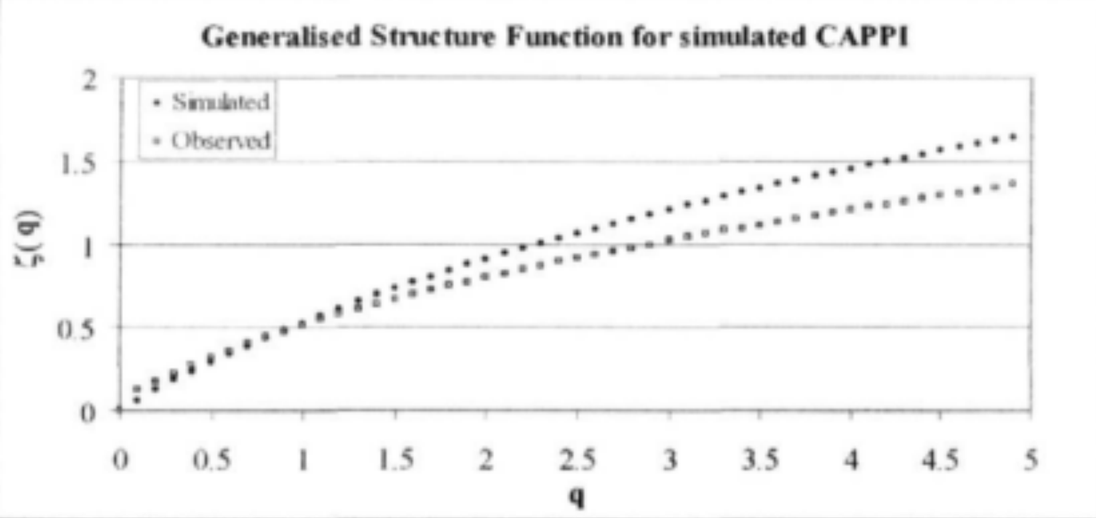


**Figure 30 - GSF Exponent for Real and Simulated CAPPIs of Figure 24**

The two sampled $\zeta(q)$ functions compare reasonably well out to a range of $q = 5$. A difference appears at the lower end of the curve near the origin. In the simulated case the function plots nicely through the origin as expected, however in the real case the function appears to be discontinuous near the origin. This is an artefact of the radar data received from Bethlehem which is recorded to a precision of 0.1 dBZ. The truncation of the data in this way causes the introduction of some zero values and consequently the $\zeta(q)$ function to bend upwards at low values of $q$, with a discontinuity at the $\zeta(q)$ axis as explained in Section 13.1. This effect becomes more severe as the precision of the data is reduced.

Menabde et al (1999) show that for a two dimensional non-stationary random field, the gradient $\beta_{space}$, of the radially averaged, two dimensional power spectrum (one of the three parameters of the String of Beads model) is related to $G_2(l)$ by Equation 7.

$$G_2(l) \propto l^{\beta - D} \tag{7}$$

Which implies that when $D = 2$ (as is the case for CAPPI data):

$$\beta = 2 + \zeta(2)$$

Figure 30 gives values of $\zeta(2)$ equal to 0.80 and 0.91 respectively, implying a $\beta_{real}$ of 2.80 and a $\beta_{simulated}$ of 2.91. These compare quite well to the measured values of 2.51 and 2.88 presented in Figure 24.

The validity of comparing the spatial structure of two *integer precision* CAPPIs using the Generalised Structure Function is questionable but is done here nevertheless for the five CAPPIs of Figure 5 and their simulated equivalents of Figure 21. Plots of the gradients of the Generalised Structure Functions for these 10 CAPPIs are presented in pairs of real and simulated CAPPIs in Figures 31 through to 35.

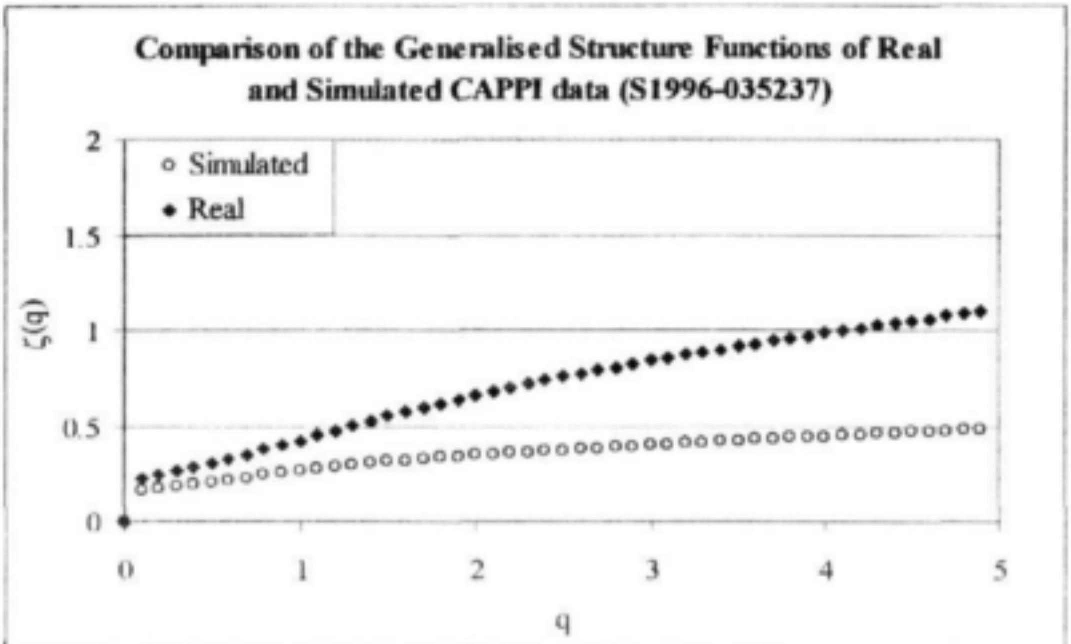


**Figure 31 - Comparison of the GSF gradients of the real and simulated integer precision CAPPIs**

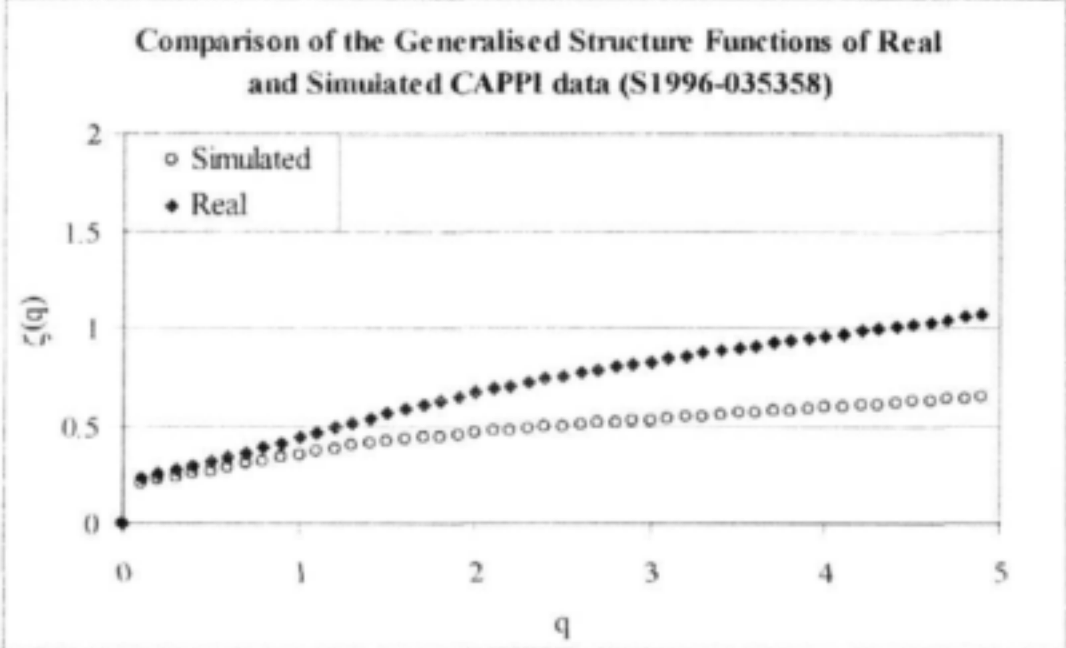


**Figure 32 - Comparison of the GSF gradients of the real and simulated integer precision CAPPIs**

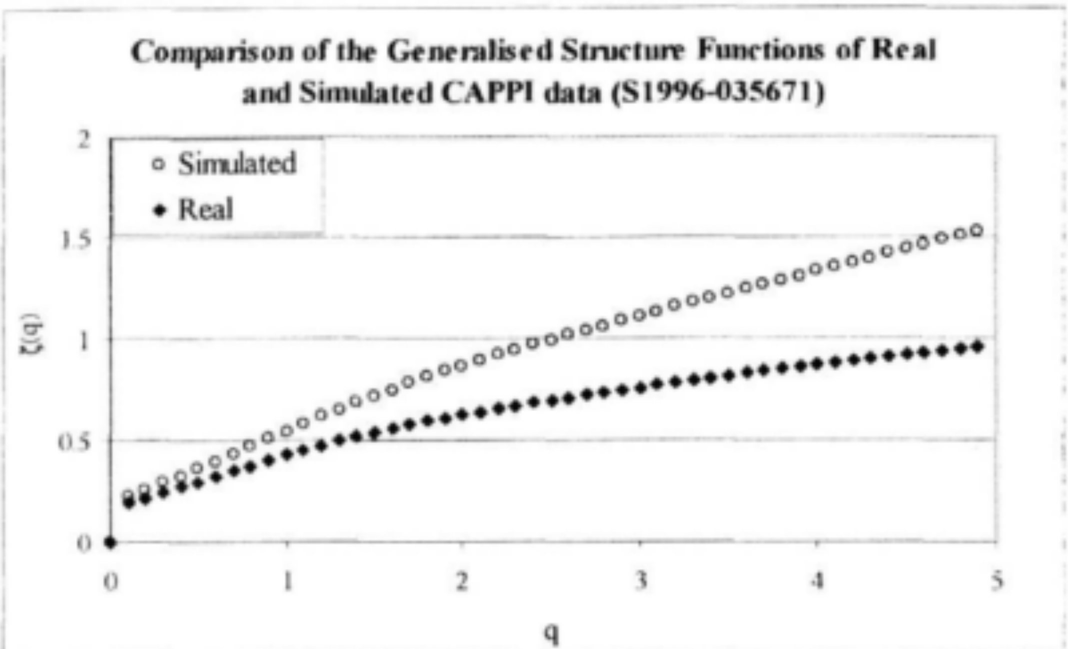


**Figure 33 - Comparison of the GSF gradients of the real and simulated integer precision CAPPIs**

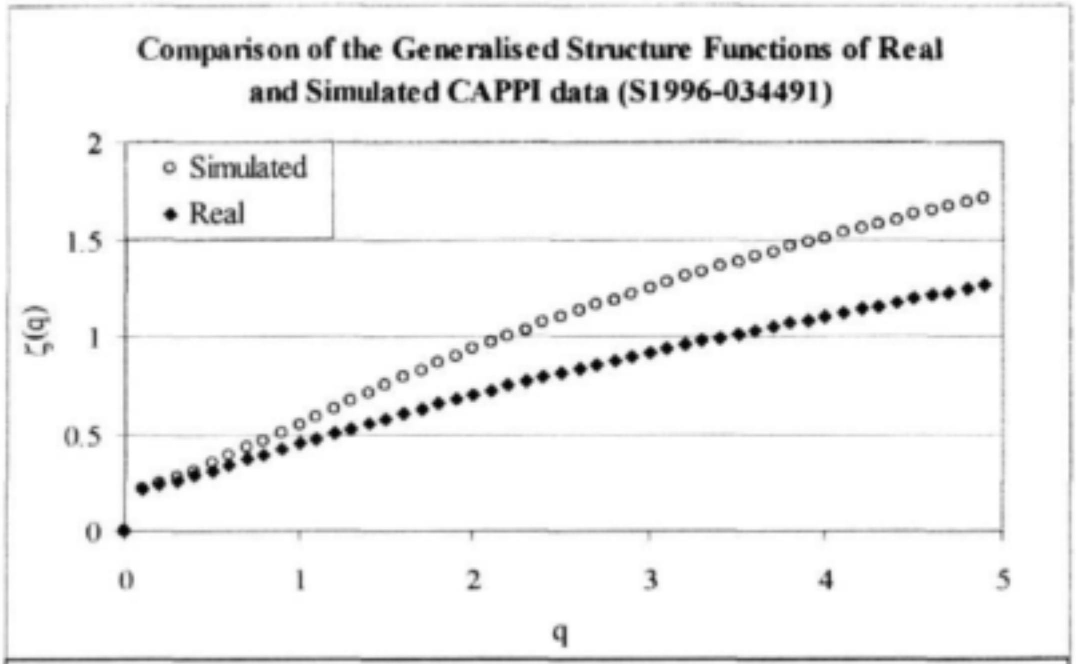


**Figure 34 - Comparison of the GSF gradients of the real and simulated integer precision CAPPIs**

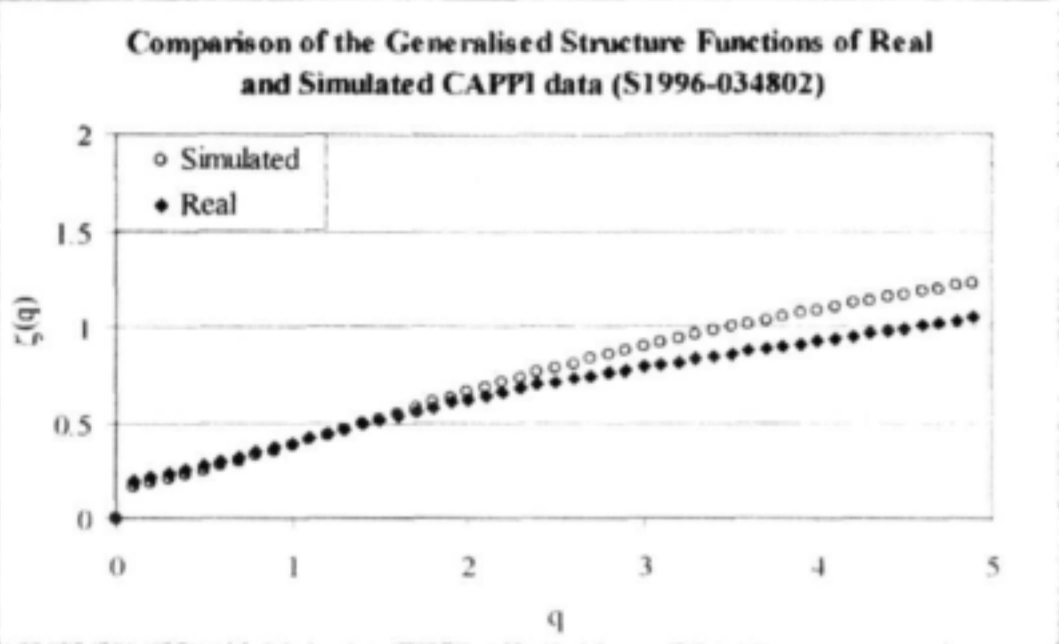


**Figure 35 - Comparison of the GSF gradients of the real and simulated integer precision CAPPIs**

The ζ(q) functions plotted in Figures 31 through to 35 suggest that the Generalised Structure Function is better approximated for a higher Wetted Area Ratio or for a lower number of zero values. The worst case is presented in Figure 31 for the CAPPI 035237 which has the lowest WAR and the best case is that of Figure 35 for CAPPI 034802 which has the highest. Apart from Figure 31, the ζ(q) functions are reasonably well approximated up to a ***q*** value of 1.0 but get progressively worse for the higher order moments.

When considering the *integer precision* radar data that has been analysed in this study, it is not possible to draw any firm conclusions regarding the spatial structure of the simulated (integer) CAPPIs by the use of the Generalised Structure Function. With the limited amount of *double precision* radar data analysed thus far, the String of Beads model appears to perform well when simulating *double precision* CAPPIs, but this needs to be confirmed by additional analysis. With the introduction of the MDV data storage format early in 1999, all radar data obtained from the Bethlehem Weather Bureau will be stored to a precision of 0.1dBZ and it is intended to use this data to further validate and improve the *String of Beads model* in a follow-on study, under WRC contract K5/1010.

## 14. LIMITATIONS OF THE STRING OF BEADS MODEL

A limitation of the String of Beads Model is that it produces rainfields which are statistically homogeneous in space-time. Simulating orographic effects on the rain field has not yet been explored in this model. A second limitation is that the mechanism used to drift the storm across the image, Fourier wrapping, demands that the entire weather system in the image move with the same velocity (not necessarily constant). This does not always happen in real rainfall, particularly when considering convective activity over a large area. Although the String of Beads Model has been designed to run on a personal computer, the routines used for analysis are RAM and CPU intensive. This is not a serious limitation for any good modern computer and will become even less of an issue as more powerful computers are made available. Typically, on a Pentium II 266 with 128MB RAM, simulation of a month's rainfall takes approximately 30 minutes. Analysis of the month's rainfall (if required) will take approximately 2 hours. This could be improved by using more efficient optimisation routines or bigger computing power.

## 15. CONCLUSIONS

In the context of other rainfall models, the String of Beads Model is a simple, effective means of simulating two dimensional rainfall fields as measured by radar. It can provide a credible sequence of artificial CAPPIs of instantaneous rainfall intensity which have the correct marginal distribution and clustering behaviour. When viewed as an animated image, the storm cells in the simulated sequence of CAPPIs grow, decay and translate realistically. The model is specified by few parameters and can be used to model simple rainfall events of reasonably constant $\sigma$, $\beta$ and $\mu$ or it can be used to create detailed *design storms*. It can provide a possible random rainfall scenario for any time period based on the analysed data. Due to the scaling nature of rainfall over a wide range of space and time, large or small areas can be modelled equally well. Finally, the model is designed to run on a personal computer in order to avoid the need for specialised computing equipment.

## 16. RECOMMENDATIONS

In its current state the String of Beads model is in an early state of development. Additional work is required in the validation of the model and in the linking of the radar data to raingauge data for use in areas not covered by radar. If the model is to be useful in the management of water resources outside of the study area, additional radar data will need to be obtained from other radars in the country. These data would then be analysed in order to calibrate the model for use in any region of South Africa. With the standardisation of the radar data storage format into MDV file format, the String of Beads model will also need to be adapted to perform analysis on these data. Lastly, although much of the operational core of the String of Beads model is complete, a considerable amount of work is required in the development of the user interface if it is to be used outside of an academic environment.

# Appendix A - Radar and Raingauge Data

## A.1 INTRODUCTION

The beauty of radar data is that it lends itself to the communication of rainfall intensity over a large area through the use of images thereby exploiting the ability of the human mind to qualitatively assess the data much faster, and often with far greater accuracy, than can be done by computer program. If poorly managed, the large volumes of data can be extremely cumbersome and lead to problems in computer disk storage space. It is therefore essential to briefly outline the processes used in this study to store, manipulate and display the radar data. In this appendix, the configuration of the MRL-5 weather radar is discussed and the conversion of its raw data to Cartesian co-ordinates at a constant altitude, that is the generation of CAPPIs, is explained. The quality of the radar data and some of the difficulties experienced when measuring rainfall using radar are considered. Finally, the methods used to store and to display the CAPPI data in an informative format, are presented.

## A.2 THE MRL-5 WEATHER RADAR AT BETHLEHEM

A photograph of the MRL-5 Radar is given in Figure A.1. The radar dome is mounted on top of a semi-mobile control centre. The data are transmitted via radio link to the offices of the Weather Bureau in Bethlehem which are roughly 20 km south east of the radar site. The radar is powered by an uninterruptable power supply (UPS) to ensure that the data set is not corrupted in the event of an electricity failure which is most likely to occur during a thunder storm. It is a combined S-band (10 centimetre wavelength) and X-band (3 centimetre wavelength) weather radar. The S-band radar requires a larger, more expensive antenna but is less prone to errors due to attenuation and is therefore a better choice when measuring rainfall intensity at long range. The X-band radar data is not considered in this study. In S-band mode, the MRL-5 is configured to perform a full volume scan starting with a base scan at an elevation of 1.5° and incrementing in elevation in 18 increasing steps to its final elevation of 55°.

**Figure A.1 - Photograph of the radar**

With an antenna 4,5 meters in diameter, the beam width of the S-band radar is 1.5°. Data is collected in 600m bins along the beam. Data analysed to date records its first 600m bin echo at a range of 14km from the radar, resulting in a characteristic hole at the centre of the CAPPIs. The MRL-5 has recently been reconfigured so that the first bin is at a range of 600m and future CAPPIs will record rainfall everywhere within a 150km radius of the radar. The MRL-5 completes one volume scan every 4.5 minutes.

### A.3 SOURCES OF ERROR IN RAINFALL MEASUREMENT USING RADAR

A basic assumption made in the development of the model is that the data received from the radar is a reasonably accurate representation of what actually reaches the ground. Radar measures rainfall by sending out an intense electromagnetic pulse, and then listening to the echoes of the pulse as minute parts of it are reflected back to the antenna off the raindrops. Being an electromagnetic wave, the pulse propagates in a reasonably straight line from the antenna and knowing the speed of light in air, the range of the raindrops can be computed by measuring the time between the pulse and the echo. This process is subject to a wide variety of systematic errors and these will be discussed briefly with reference to the conditions in Bethlehem.

### A.3.1 Ground clutter and beam blocking

Ground clutter is caused by the radar beam or the side lobes of the radar beam colliding with the ground at some point and the result is a strong echo which could be interpreted as an intense rainfall cell on a CAPPI. In a sequence of CAPPIs this is easily visually distinguished from rainfall due to the fact that it is stationary. The use of Doppler radar is one of the methods of eliminating the effects of ground clutter but this results in a much slower scan speed and consequently a much lower temporal resolution. Where the core of the radar beam strikes a large fixed object, the beam is said to be *blocked* and rainfall beyond that point will not be recorded by the radar. In mountainous areas ground clutter presents a major problem in the measurement of rainfall using radar. Apart from the Maluti Mountains in Lesotho, the terrain surrounding the Bethlehem radar is quite flat and there is therefore very little ground clutter and the rainfields tend to be reasonably homogeneous. A CAPPI altitude of 2km above ground level ensures that the effects of ground clutter are minimal, whilst the rainfall passing through that level is thought to be a reasonable representation of what actually reaches the ground (ignoring the effects of evaporation and updraughts).

### A.3.2 Bright band

The so called *bright band* is a band of high reflectivity which is usually at a reasonably constant altitude and corresponds to the melting layer in a cloud. When viewed by radar, snow and ice crystals have a reflectivity which is approximately 4 times lower than that of water. Consequently, the radar will tend to under estimate the rainfall on the ground when looking above the bright band. As snow and ice crystals pass through the melting layer en route to the ground, they begin to melt and become coated in a thin layer of water. This results in a very high reflectivity as the radar *sees* very large drops of rainfall. On a CAPPI this is seen as a ring of high reflectivity or rainfall intensity centred on the radar. When viewed with a vertically pointing radar this is seen as a band of high reflectivity. Research is underway at the University of Natal (Pegram and Mittermaier, 1998) in an attempt to predict the rainfall reaching ground level using the vertical reflectivity profile.

Due to the high summer temperatures experienced during the wet season in South Africa, the bright band is usually at or above the 2km level. A large proportion of the rainfall experienced in Bethlehem is convective in which case the bright band is not well defined and therefore does not present a problem.

### A.3.3 Beam Attenuation

As the radar beam strikes any object, a small part of the beam is reflected back to the receiver, some is scattered and the remaining (slightly weaker) part continues in its original direction. The radar beam becomes weaker as its range from the radar increases. When passing through an intense storm cell, it is possible in extreme cases for the beam to become so weak that rainfall falling beyond the cell is significantly under estimated. This is known as beam attenuation. Since the longer wavelength radars are more powerful and suffer from less Raleigh scattering, the effects of attenuation are kept to a minimum by using an S-Band radar.

### A.3.4 Anomalous propagation

When calculating the three dimensional position corresponding to the reflectivity recorded by the radar, it is assumed that the radar beam has propagated in a straight line from the radar to the target and back again. This is not always the case. Being an electromagnetic wave, the radar beam is bent by a change in density of the medium through which it is travelling. A simple analogy can be drawn to a straight stick when it is partly immersed in water. To the person holding the stick, it appears to be bent at the air-water interface. Changes in air density are experienced with the approach of a weather front and in extreme cases, a temperature inversion can cause the radar beam to be bent to such an extent that it strikes the ground. Part of the reflected beam will follow the same path back to the receiver and will be observed as a point of high reflectivity. This is difficult to identify in a CAPPI as it is not observed in a fixed location. In these extreme cases the beam will be blocked at the point of intersection with the ground and will be difficult to distinguish from beam attenuation. More information is required in this case, either from ground based stations or from satellite images. These extreme cases of anomalous propagation are very rare in the data collected at Bethlehem.

### A.3.5 The Z-R relationship

Much research has been done on the relationship between the observed reflectivity (***Z***) and the rainfall rate (***R***) recorded on the ground. The Z-R relationship $Z = 200R^{1.6}$ given by Marshall and Palmer (1948) is used to convert reflectivity to rainfall rate for the Bethlehem data. Various refinements of this relationship have been proposed, 69 of which are quoted by Battan (1973). It has been shown (Uijlenhoet, 1998) that the Marshall-Palmer relationship is very close to the mean of the 69 relationships referred to by Battan (1973). An important point to note is that the measurement of rainfall intensity by *any* of the known methods is by no means infallible and at *best* a reasonable estimate will be achieved. The use of the Marshall-Palmer relationship seems to be a reasonable one.

In spite of the possible sources of error, the weather radar data received from the Bethlehem Weather Bureau is of excellent quality and will only improve as better methods of rainfall estimation are developed.

### A.4 METHODS OF CAPPI EXTRACTION FROM VOLUME SCAN DATA

The CAPPIs considered in this study have been extracted from the volume scan data in one of two ways. The first is that used by Seed (1994) which involves simply projecting the nearest bin of radar data (shown as bold lines in Figure A.2) onto a horizontal plane (shown as a dashed line) at a chosen altitude which was taken as 2 kilometres for this study.

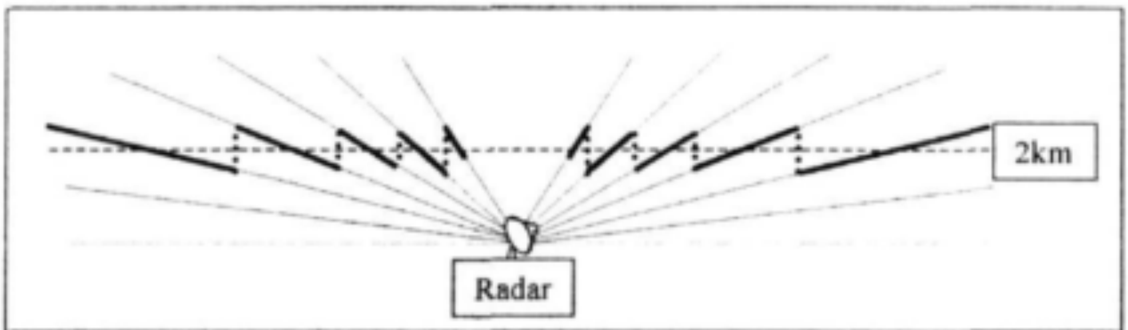


**Figure A.2 - Creating CAPPIs from volume scan data using the projection technique of Seed (1992)**

This method is a very fast and simple way to generate CAPPIs, however it has the disadvantage of having discontinuities in rainfall rate at the jump from one beam to the next. This effect is observed in the CAPPI of Figure A.3. Most of the CAPPIs analysed in this study were extracted from volume scan data in this way.

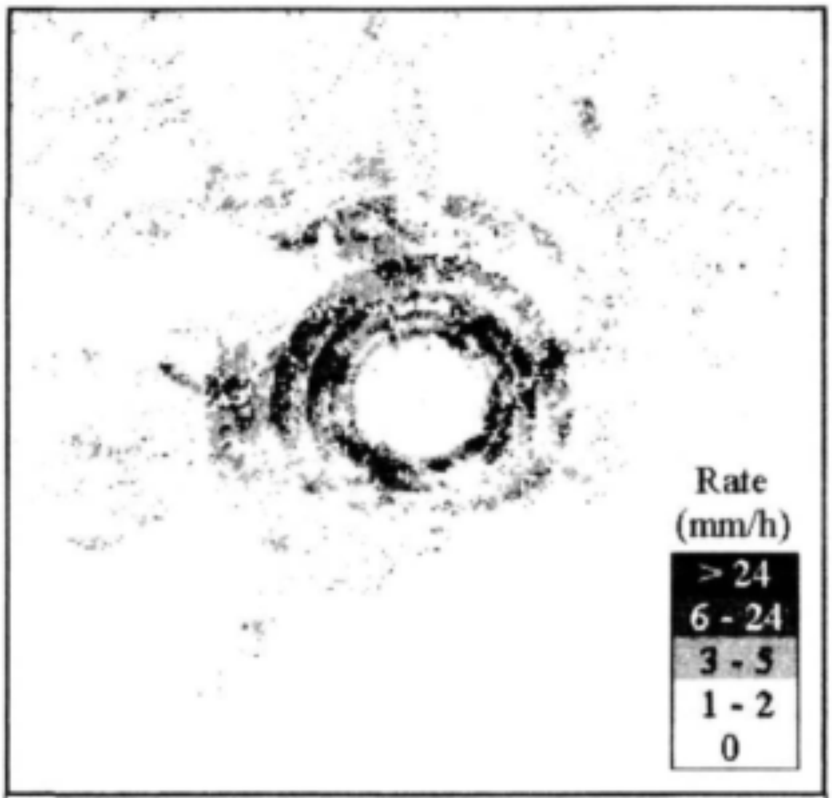


**Figure A.3 - 200km x 200km CAPPI generated from volume scan data using the projection technique (Mittermaier and Terblanche, 1997).**

A better method of extracting CAPPIs from volume scan data is that suggested by Mittermaier and Terblanche (1997). The algorithm is incorporated into a more comprehensive system of radar signal processing known as the DISPLACE system (Terblanche, 1996). It involves the interpolation between eight radar bins at two levels to achieve a weighted average reflectivity at a point $(x, y, z)$ in Cartesian space. This point corresponds to a point $(r, \phi, \theta)$ in spherical co-ordinates and the relationships between the two frames of reference are defined by the four equations (A.1) through to (A.4).

$$r = \sqrt{(x^2 + y^2 + z^2)} \qquad \text{(A.1)}$$

$$\theta = \arctan\left(\frac{x}{y}\right) \qquad \text{(A.2)}$$

$$z_h = z - \left( \frac{x^2 + y^2}{aD} \right) \tag{A.3}$$

where $\boldsymbol{a}$ is the standard compensating tropospheric beam refraction factor of $^4/_3$ (Battan, 1973) and $\boldsymbol{D}$ is the approximate diameter of the Earth. Finally,

$$\phi = \arcsin\left( \frac{z_h}{r} \right) \tag{A.4}$$

For each chosen Cartesian grid point, the eight closest radar grid points are identified (four on the elevation *cone* above the point and four on the *cone* below). This is illustrated in Figure A.4. Mittermaier and Terblanche (1997) explain that since the radar grid points do not necessarily match the Cartesian grid points, weighting factors reflecting these differences are calculated to allow for linear interpolation from the eight surrounding spherical data grid points.

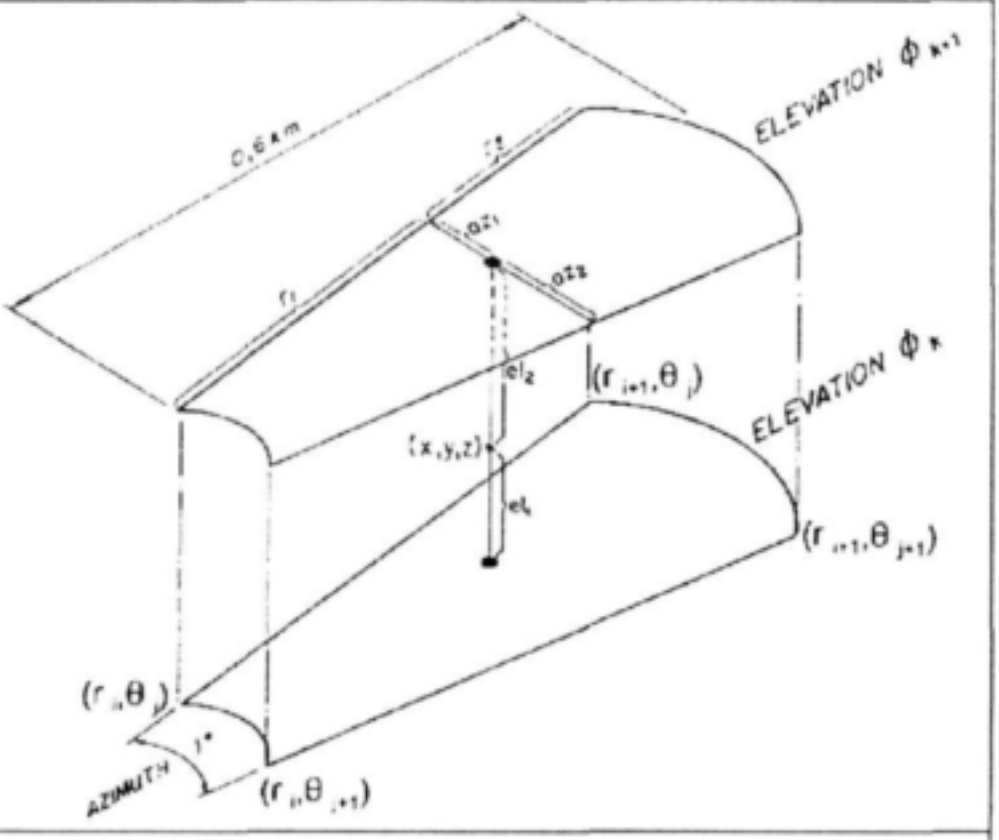


**Figure A.4 - Illustration of Cartesian point and the eight surrounding radar data points used in DISPLACE averaging to extract CAPPIs from volume scan data (Mittermaier and Terblanche, 1997).**

Figure A.4 illustrates that for each Cartesian point, six weighting factors have to be calculated, two for the slant range ($r_1$ and $r_2$), two for the azimuth ($az_1$ and $az_2$) and two for the elevation ($el_1$ and $el_2$). The reflectivity at the point $(x, y, z)$ is then calculated by taking the weighted average of consecutive bins ($r_i$ , $r_{i+1}$) in the ratio ($r_2 : r_1$), thus reducing eight points to four which surround $(x, y, z)$ in a vertical plane. These four points are then averaged in azimuth in the ratio ($az_1 : az_2$) to give an average reflectivity directly above and directly below the point in Cartesian space. Finally, the reflectivity of the point above and the point below the point $(x, y, z)$ are averaged in elevation ($el_1 : el_2$) to give the weighted average reflectivity at $(x, y, z)$. The result of this process is a far smoother CAPPI which does not exhibit the concentric rings observed in CAPPIs calculated using the projection technique. An example of the DISPLACE type of CAPPI, extracted from the same raw radar data as that used to generate the CAPPI of Figure (A.3), is given in Figure (A.5).

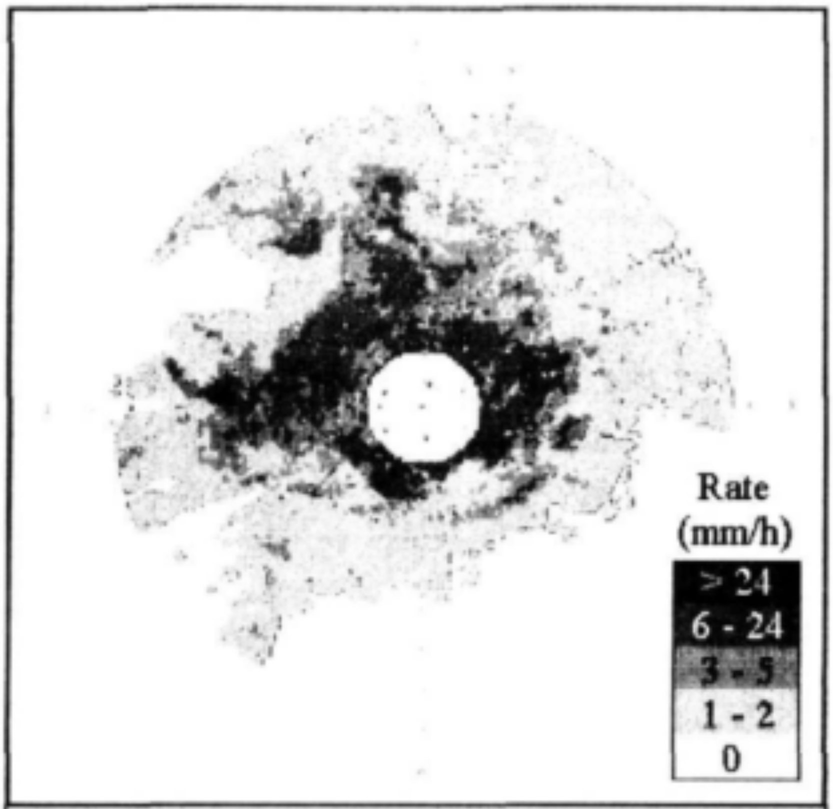


**Figure A.5 - 200km x 200km CAPPI generated from volume scan data using the DISPLACE averaging technique (Mittermaier and Terblanche, 1997).**

Data received beyond a range of 67 km from the radar is not used in the generation of 2km CAPPIs when using this technique since the altitude of the base scan exceeds 2km at that range. Although the DISPLACE method is clearly a more thorough approach to CAPPI generation, it is a relatively new technique in South Africa and has only recently been implemented in the radar data received from the Bethlehem Weather Bureau. Consequently only a small proportion of the analysis was performed on this type of CAPPI.

### A.5 DATA STORAGE CONSIDERATIONS

Raw radar reflectivity data in dBZ was processed by the Bethlehem Weather Bureau to a precision of 0.1dBZ and made available for analysis in one of two main formats. In the first, reflectivity data (dBZ) is stored uncompressed as a continuous array of 40000 integers per CAPPI, at 16 bit integer precision. To maintain the precision of 0.1dBZ, reflectivity was multiplied by 10 prior to being stored as an integer so that an integer of 253 would represent an average reflectivity for that pixel of 25.3dBZ, for example. Data are stored row-wise (West to East), top to bottom (North to South), so that every 200th integer in the array is at the beginning of the next row located 1km South of its predecessor. This method of data storage is extremely inefficient as it requires roughly 78kB (1kB = 1024 bytes, 1 byte = 8 bits) of disk storage space per CAPPI, or 22MB of disk storage space per day if a temporal resolution of 5 minutes is assumed. Due to limitations of disk storage space on the personal computer, very few CAPPIs of this nature were analysed in this study.

A far more efficient method of data storage is achieved by first converting the reflectivity data to rainfall intensity data and then truncating the data to an 8 bit integer. The result of this process is that the pixel intensities can only assume integer values between 0 and 255 inclusive. Most of the CAPPI data analysed in this study were supplied in this format. The conversion to 8 bit precision results in any rainfall rate below 1mm/h being truncated to 0mm/h. Although some detail is lost in the truncation process, a saving in disk space of 50% is achieved so that the data for one day can be stored *uncompressed* using 11MB.

Under prevailing South African conditions, a large proportion (usually more than 50%) of a typical CAPPI is either dry or experiencing a rainfall rate less than 1mm/h at any instant in time so that the array of integers will usually contain a large proportion of zero values. This is due to the fact that a lot of the rainfall experienced in South Africa is in the form of isolated thundershowers. The storage method adopted by Seed (1992) exploits this fact and stores

CAPPIs as a sequence of short *non-zero* arrays. Each non-zero array is preceded by a 32 bit integer *address* corresponding to the position of the first non-zero pixel in the 40000 pixel array, and a 16 bit integer *length* corresponding to the number of consecutive pixels before the next zero. This same principle is extended to a large file format (*.*ima* file) which stores all the CAPPIs for a day and uses an index file (*.*ind* file) to locate the start of each new CAPPI in the *.*ima* file. In this way, data storage is typically reduced by approximately 60% depending on the extent of the rainfall recorded on the CAPPI and the amount of disk storage space required for a day's radar data is approximately 4.5MB. All data analysed in this study is included in this format on compact disk (Refer to Appendix G).

Furthermore, when considering the quantitative limits of the data discussed in Section 5, there is no need to store data beyond a range of 64km from the radar. Since it is convenient to store data as a rectangular or square image, the stored data is reduced to an array of 16384 bytes per CAPPI or 4.5MB per day in an *uncompressed* format. In addition, Section 5 showed that only a sample area of 9128 pixels (out of the 16384 pixels in a 128 x 128 pixel CAPPI) is considered in this analysis. If no information is required outside of the $^3/_4$ doughnut mask (as is the case when analysing CAPPI data), pixels located outside of the mask can be set to zero and the Seed compression algorithm (conversion to *.*ima* and *.*ind* file) can be applied, thereby reducing the amount of disk storage space to a mere 1.5MB for a typical day's rainfall. This represents an overall saving of 93% in disk storage space which is an important consideration when operating on a personal computer.

These two methods of data storage have recently been replaced by the *MDV format* used in the TITAN software (Dixon and Wiener, 1993) and this is soon to become the standard *compressed* format used for storing radar data in the South African Weather Bureau. No data stored in this format has been included in this study. The MDV format is an extremely efficient method of storing radar data. The TITAN software has the option of converting the reflectivity data (in MDV file format) into an uncompressed ASCII or binary file which can then be analysed in the same manner as the other data used in this study.

### A.6 CONVERTING BINARY DATA INTO IMAGES

In order to view the CAPPI data as an image it was necessary to first select an image file format and for the purposes of this study the 256 colour Windows Bitmap was chosen as being the most suitable. There are several reasons for this choice the first being that it is an uncompressed, simple file format which makes it easy to manipulate during the analysis process. Since the 256 colour bitmap also uses one byte per pixel, the raw data can be stored as an image without any loss of precision. Almost any Windows based graphics software will display, edit and convert the bitmap into any other desired image format. This makes it easily portable and accessible to most computers. The indexed colour palette is also readily edited in most simple graphics programs to give a different viewpoint without any manipulation of the data. The binary structure of the bitmap will be explained by simple example in section A.6.1.

### A.6.1 The binary file structure of the Windows bitmap

In order to illustrate the binary structure of the Bitmap, it is convenient to discuss the structure of a small 16 colour bitmap, an example of which is given in Figure A.6 - each small square represents a single pixel on the image. The principles are the same for 16 and 256 colour bitmaps.

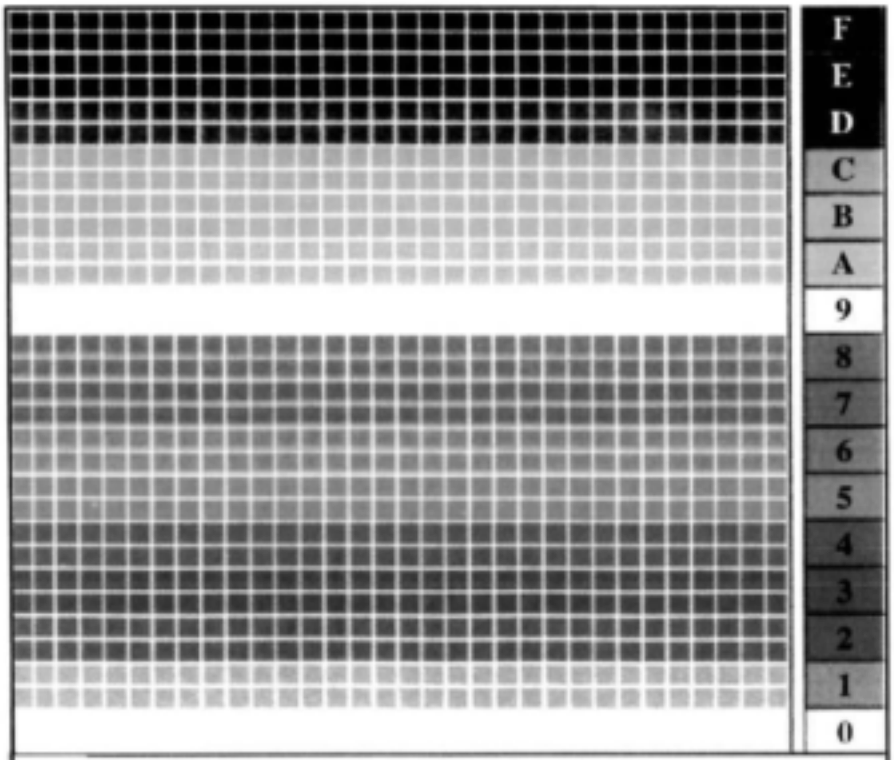


Figure A.6 - Example of 32 x 32 pixel, 16 colour bitmap

A bitmap file consists of an identifying string, a definition of rectangular size (image height and width in pixels), a colour palette and finally a continuous array of numbers which refer to the colour palette index. The colour palette defines and indexes each of the colours in the image in terms of the three primary colours of light (viz. Red, Green and Blue) at 24 bit precision or one byte per primary colour. The size of the 256 colour palette is 1078 bytes

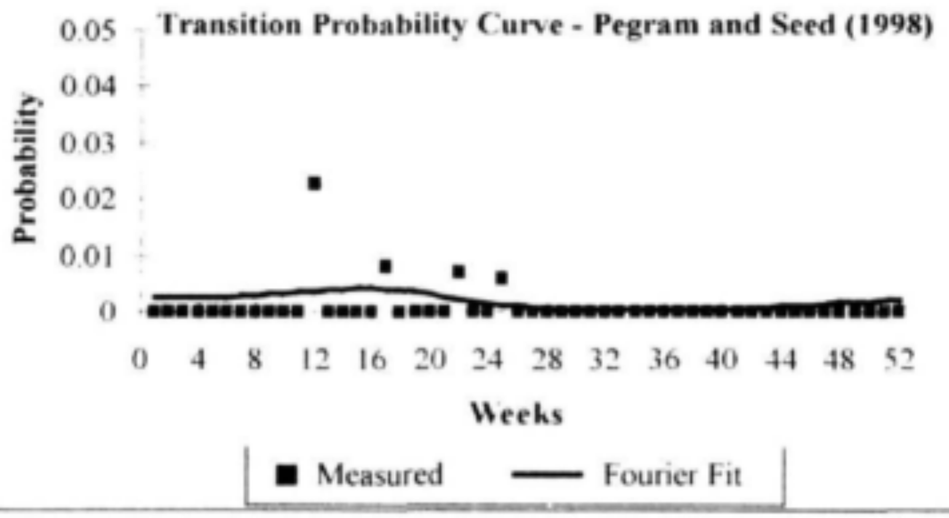


Figure B.3 - Probability of a general rainday following a dry day

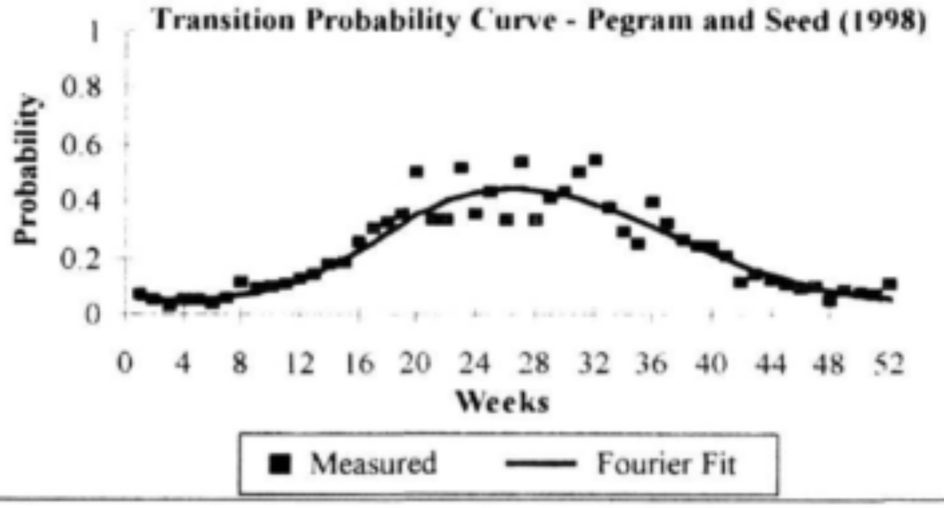


Figure B.4 - Probability of a dry day following a scattered rainday

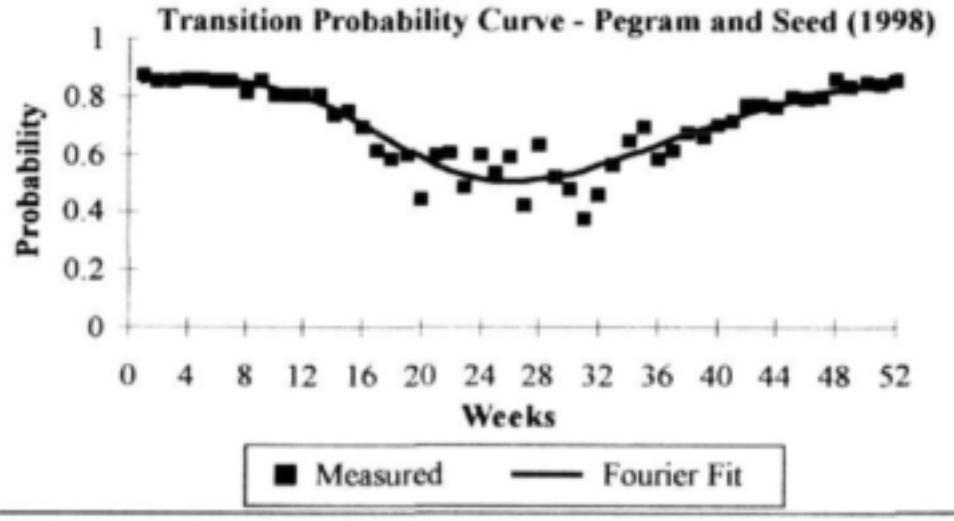


Figure B.5 - Prob. of a scattered rainday following a scattered rainday

# Appendix B - Transition Probability Curves

## B.1 INTRODUCTION

The full set of *periodic transition probability curves* for the Bethlehem area, calculated by Pegram and Seed (1998), are included in Figures 5.2 to 5.10. The points represent the weekly sampled transition probabilities from the 30 year collection of up to 330 daily read raingauges in the 200 kilometre square centred on Bethlehem. The line which has been fitted to these points is the two-harmonic Fourier fit.

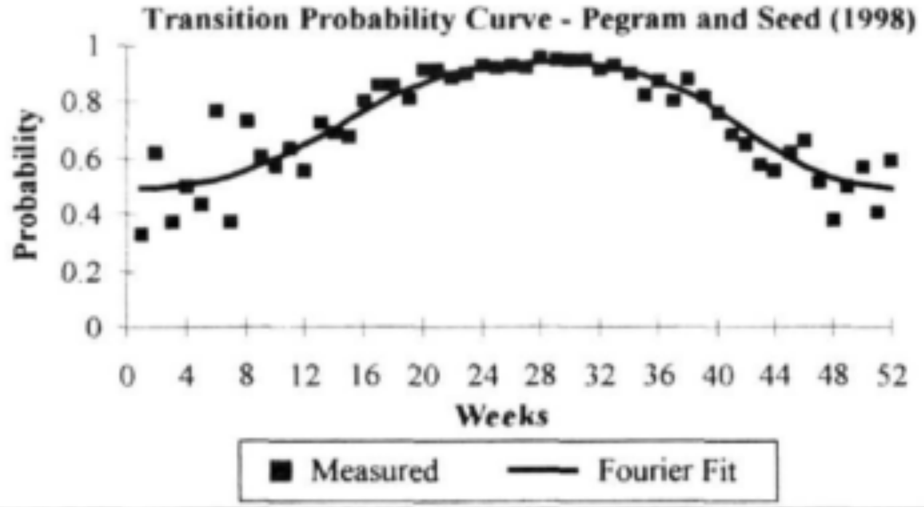


**Figure B.1 - Probability of a dry day following a dry day**

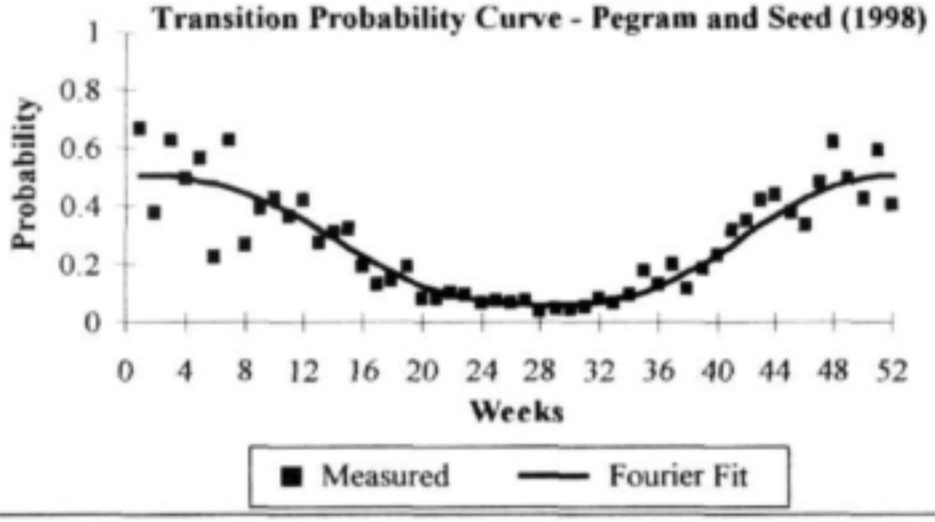


**Figure B.2 - Probability of a scattered rainday following a dry day**

## A.8 SUMMARY

This appendix has served to explain the process involved in radar data collection, storage, interpretation and manipulation. It is acknowledged that rainfall measurement by radar is by no means perfect, however it is able to give a good representation of the rainfall in real time over a large area and at a high spatial and temporal resolution. The techniques that have been outlined are easily taken for granted, but they are essential to the data visualisation, modelling and analysis processes.

The use of colour proved to be far superior to greyscale in the communication of CAPPI structure. The linear Red-Green-Blue (RGB) palette of Figure A.8c provides an image which gives a clear idea of the distribution of rainfall intensity on a CAPPI and is particularly useful when considering the structure of less intense storm cells in the CAPPI. The logarithmic RGB palette of Figure A.8d better describes the structure of the more intense storm cells on the CAPPI although intuitively it gives the impression of a very intense weather system. When qualitatively comparing CAPPIs, they should be viewed using both these palettes.

## A.7 CAPPI IMAGE FILE NOMENCLATURE

A point of great importance when dealing with a large number of image files is that they should have sensible nomenclature. During analysis it is necessary to open and close each file in turn so that if this process is to be automated the number of characters in the file name should be constant. Another point to note is that it is far easier to reference a file by number than by letter. The file nomenclature that was adopted for the data was:

First character - **S, C or X** (the type of radar)
Second to fifth characters - the **year** (four characters for 2000 compliance)
Sixth character - the **dash** character
Seventh to twelfth characters - **number of minutes from the start of the year**
(Six numerical characters are needed as
there are 527040 minutes in a *leap* year.)
Followed by the four character extension ***.bmp***

An example of an image filename could be:

**S1996-034591.bmp**

which would be a bitmap image of a CAPPI, recorded with an S-band radar in the 34591st minute of the year 1996.

Since the numbers 0 to 9 are ASCII characters 48 to 57 respectively, the next image in the sequence can be automatically referenced by simply incrementing the right-most character in the file pointer (before the *.bmp* extension) and testing to see if it is greater than 57. If it is greater than 57, set it to character 48 (zero) and increment the character to its left.

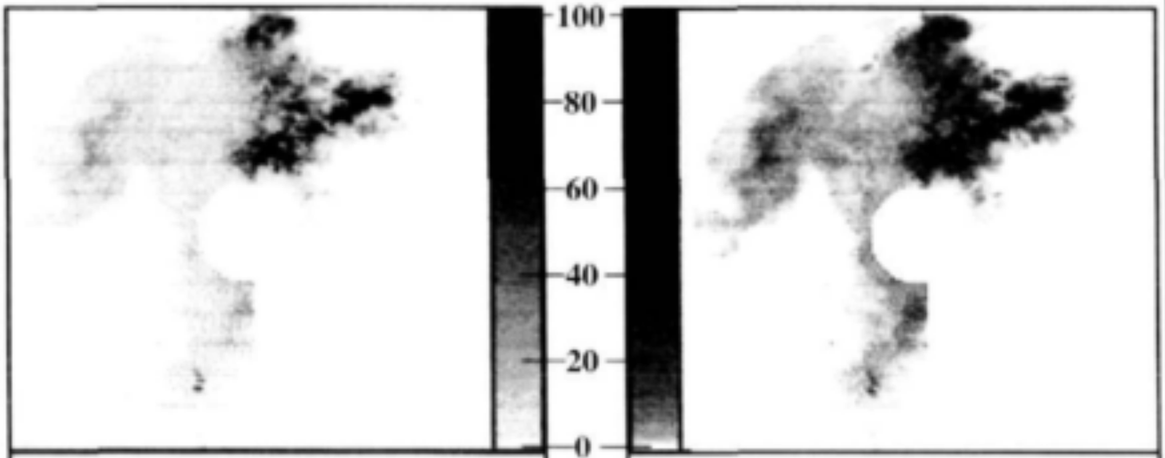


Figure A.8a - Linear grey palette

Figure A.8b - $\log_2$ grey palette

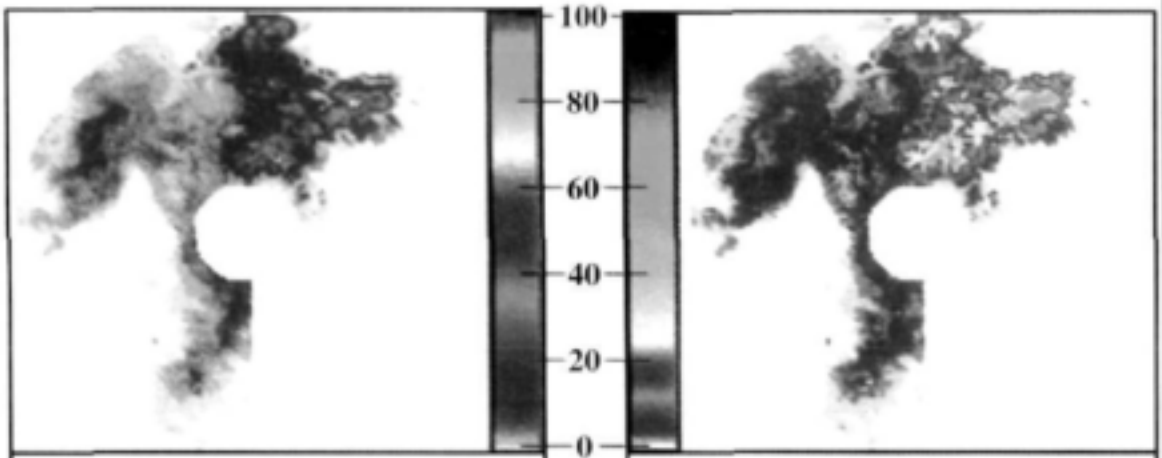


Figure A.8c - Linear RGB palette

Figure A.8d - $\log_2$ RGB palette

Figure A.8 - The use of four different colour palettes to display a CAPPI

All palettes in Figure A.8 have a range of 0-100mm/h, but the left hand images are drawn to a linear scale and the right hand images are drawn to a logarithmic scale.

In linear greyscale (Figure A.8a) it is very difficult to visualise the structure of the CAPPI because the human eye is not able to distinguish between the neighbouring intensities in the palette. Figure A.8a serves to illustrate a poor definition of colour palette. The logarithmic greyscale (Figure A.8b) provides a better picture of the CAPPI structure, but is still not able to convey the detail of the CAPPI structure, particularly at the higher rainfall intensities. Figure A.8b was found to be the best choice of palette when limited to greyscale.

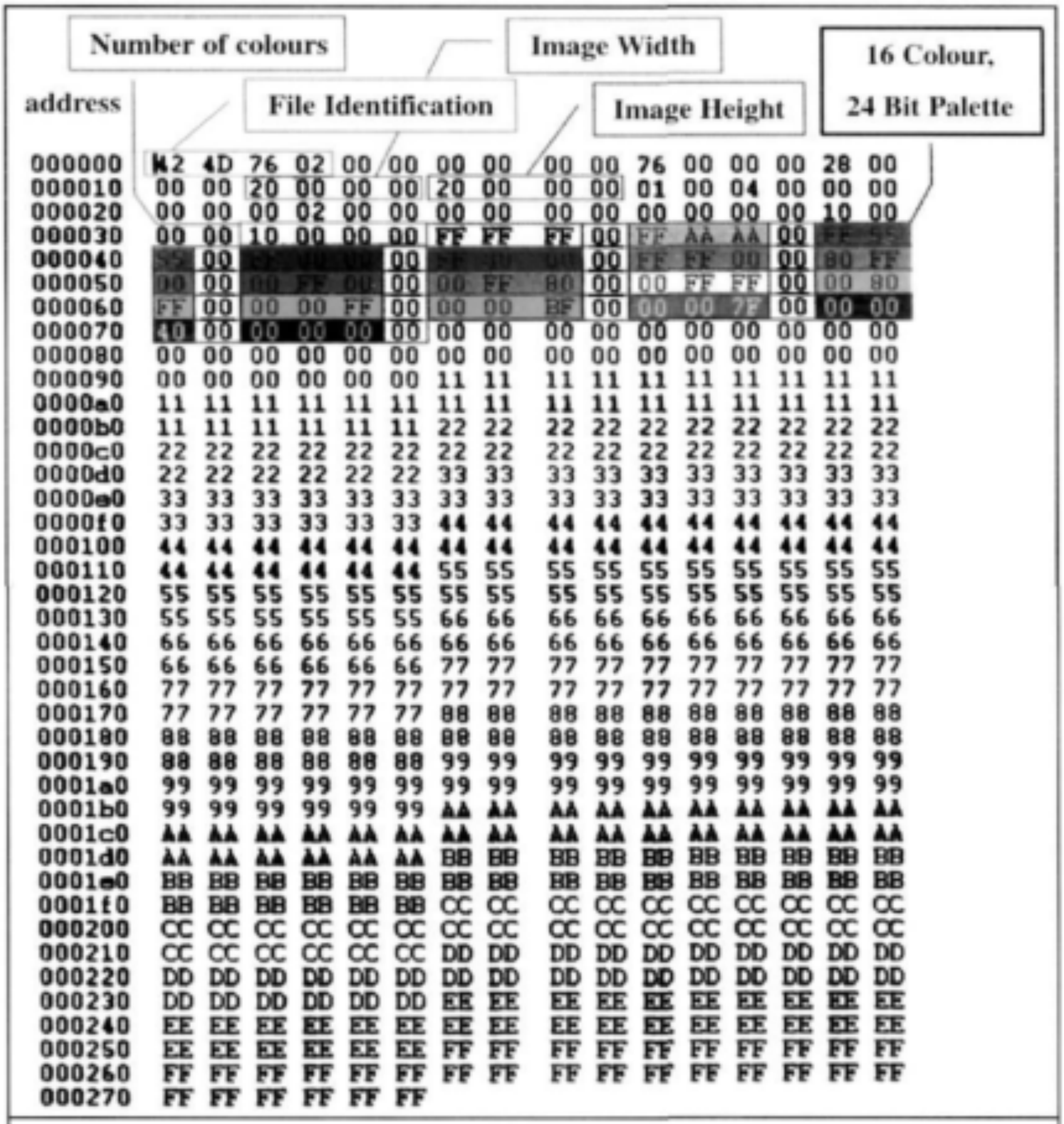
After the number of colours is declared, the next byte in the file is the NULL character, followed by the 24 bit Blue-Green-Red definition of the first colour (Colour 0) which in this case is FF FF FF or White. The 24 bit definition of each colour in the palette is separated by the NULL character. The last character before the image data, character 000076 is the NULL character which is then followed by the index value of the colour of the pixel to be drawn in the bottom left hand corner of the rectangular image. In this example, since the image is a 16 colour bitmap, this index is the first 4 bits of the next byte - Colour 0 which is White in this case. The bitmap is drawn row-wise from the bottom left hand corner up to the top right. Since there are 32 pixels in each row of the bitmap of Figure A.7, there are 64 consecutive 4 bit zeros before the start of the next colour (Colour 1).

### A.6.2 Selecting the colour palette for displaying a CAPPI

Since the colour palette is a fixed size and format for a 256 colour bitmap, it is possible to change the way in which the image is displayed by simply redefining the colour palette in the binary file. As explained in section A.6.1, this is done by altering the mixture of primary colours (RGB) in each of the 256 indices.

Manipulating the colour palette in binary is an extremely laborious process. A shortcut is to read a single CAPPI into any good graphics package and edit the palette using the tools provided in the graphics package. Once a satisfactory palette has been defined for the CAPPI, it should be saved as a new 256 colour bitmap. In order to change the palettes of a series of CAPPIs, a very simple program can be written which opens the image with the newly defined palette and copies the first 1078 bytes into a buffer. Each CAPPI in the series is then opened in turn and the first 1078 bytes of the file is over written with the contents of the buffer. Figure A.8 shows an image which is drawn using four different colour palettes.

compared to 118 bytes for the 16 colour palette and the array of data which follows the palette is at 8 bit precision for the 256 colour bitmap as opposed to 4 bit precision for the 16 colour bitmap. For the bitmap shown in Figure A.6, the binary file is given in Figure A.7.



```
000000  42 4D 76 02 00 00 00 00  00 00 76 00 00 00 28 00
000010  00 00 20 00 00 00 20 00  00 00 01 00 04 00 00 00
000020  00 00 00 02 00 00 00 00  00 00 00 00 00 00 10 00
000030  00 00 10 00 00 00 FF FF  FF 00 FF AA AA 00 FF 55
000040  55 00 FF 00 00 00 FF 80  00 00 FF FF 00 00 80 FF
000050  00 00 00 FF 00 00 00 FF  80 00 00 FF FF 00 00 80
000060  FF 00 00 00 FF 00 00 00  BF 00 00 00 7F 00 00 00
000070  40 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00
000080  00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00
000090  00 00 00 00 00 00 11 11  11 11 11 11 11 11 11 11
0000a0  11 11 11 11 11 11 11 11  11 11 11 11 11 11 11 11
0000b0  11 11 11 11 11 11 22 22  22 22 22 22 22 22 22 22
0000c0  22 22 22 22 22 22 22 22  22 22 22 22 22 22 22 22
0000d0  22 22 22 22 22 22 33 33  33 33 33 33 33 33 33 33
0000e0  33 33 33 33 33 33 33 33  33 33 33 33 33 33 33 33
0000f0  33 33 33 33 33 33 44 44  44 44 44 44 44 44 44 44
000100  44 44 44 44 44 44 44 44  44 44 44 44 44 44 44 44
000110  44 44 44 44 44 44 55 55  55 55 55 55 55 55 55 55
000120  55 55 55 55 55 55 55 55  55 55 55 55 55 55 55 55
000130  55 55 55 55 55 55 66 66  66 66 66 66 66 66 66 66
000140  66 66 66 66 66 66 66 66  66 66 66 66 66 66 66 66
000150  66 66 66 66 66 66 77 77  77 77 77 77 77 77 77 77
000160  77 77 77 77 77 77 77 77  77 77 77 77 77 77 77 77
000170  77 77 77 77 77 77 88 88  88 88 88 88 88 88 88 88
000180  88 88 88 88 88 88 88 88  88 88 88 88 88 88 88 88
000190  88 88 88 88 88 88 99 99  99 99 99 99 99 99 99 99
0001a0  99 99 99 99 99 99 99 99  99 99 99 99 99 99 99 99
0001b0  99 99 99 99 99 99 AA AA  AA AA AA AA AA AA AA AA
0001c0  AA AA AA AA AA AA AA AA  AA AA AA AA AA AA AA AA
0001d0  AA AA AA AA AA AA BB BB  BB BB BB BB BB BB BB BB
0001e0  BB BB BB BB BB BB BB BB  BB BB BB BB BB BB BB BB
0001f0  BB BB BB BB BB BB CC CC  CC CC CC CC CC CC CC CC
000200  CC CC CC CC CC CC CC CC  CC CC CC CC CC CC CC CC
000210  CC CC CC CC CC CC DD DD  DD DD DD DD DD DD DD DD
000220  DD DD DD DD DD DD DD DD  DD DD DD DD DD DD DD DD
000230  DD DD DD DD DD DD EE EE  EE EE EE EE EE EE EE EE
000240  EE EE EE EE EE EE EE EE  EE EE EE EE EE EE EE EE
000250  EE EE EE EE EE EE FF FF  FF FF FF FF FF FF FF FF
000260  FF FF FF FF FF FF FF FF  FF FF FF FF FF FF FF FF
000270  FF FF FF FF FF FF
```

Figure A.7 - Binary structure of 16 colour bitmap example of Figure A.6

Each pair of hexadecimal numbers is an 8 bit BYTE. The first column represents the byte address in the bitmap file. The file is identified by four characters 42 4D 76 02, the first three of which represent the ASCII characters BMv. The image width and height are defined by two 32 bit integers as is the number of colours in the colour palette. This structure is very similar for a 256 colour bitmap, the main difference being that the number of colours would be shown as HEX 100 which is the decimal equivalent of 256.

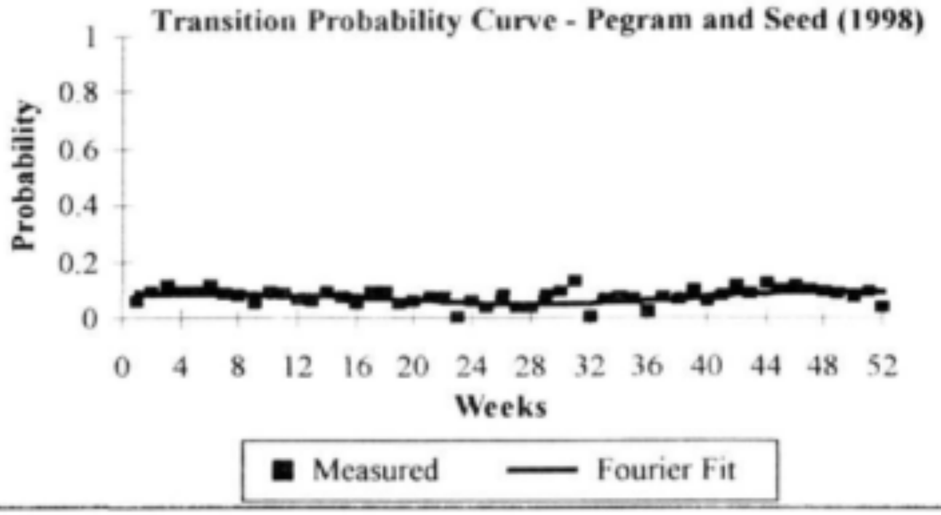


Figure B.6 - Prob. of a general rainday following a scattered rainday

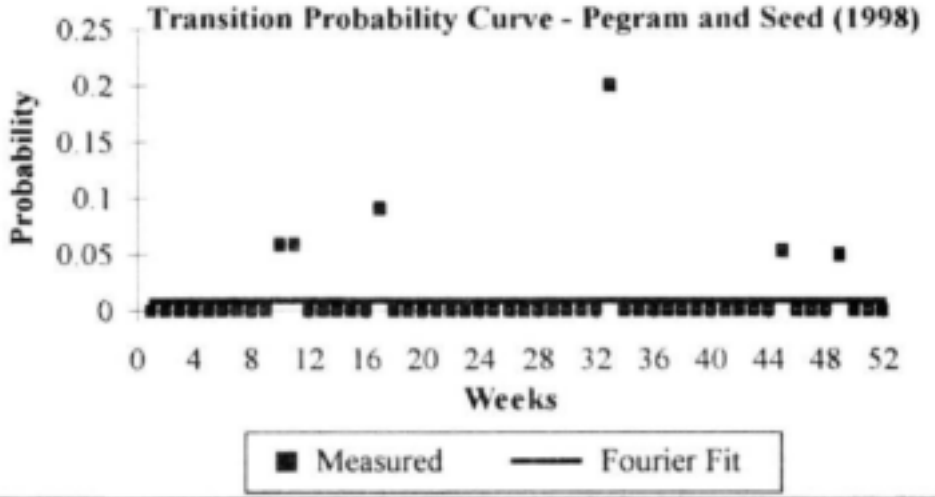


Figure B.7 - Probability of a dry day following a general rainday

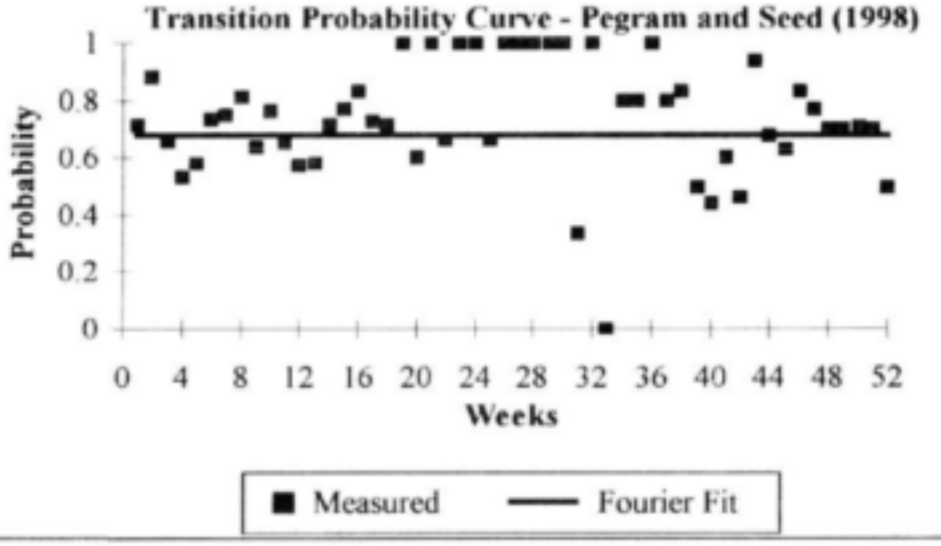


Figure B.8 - Prob. of a scattered rainday following a general rainday

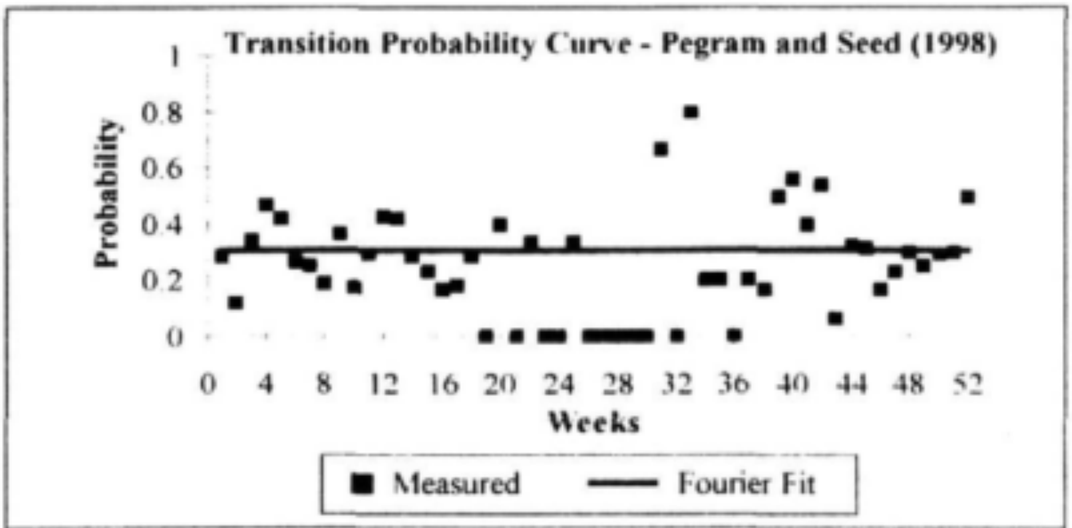


**Figure B.9 - Prob. of a general rainday following a general rainday**

In their discussion, Pegram and Seed (1998) point out that Figures B.7, B.8 and B.9 exhibit no recognisable trend and it is therefore reasonable to keep the general-to-anything transition probabilities constant throughout the year. Note the change in vertical scale on Figures B3 and B7 illustrating the rarity of a *General* rainday following a *Dry* day and vice-versa.

# Appendix C - Tools for Data Analysis

## C.1 INTRODUCTION

The Lognormal distribution, the Fourier Transform and the Generalised Structure Function are defined and the method of calculation of their parameters is discussed. These are the tools used in the development and *verification* of the output of the *String of Beads model* and in the case of the Generalised Structure Function, the *validation* of the spatial correlation structure of the simulated CAPPIs..

## C.2 THE TWO PARAMETER LOGNORMAL DISTRIBUTION

A variate is considered to be *lognormally distributed* if the logarithms of the variate are described by the Normal Distribution. The formal definition and properties of the two parameter lognormal distribution are described by Aitchison and Brown (1957) and paraphrased here for completeness and clarity of exposition.

Using their notation, consider a positive variate $\boldsymbol{X}$ $(0<x<\infty)$ such that $\boldsymbol{Y} = \log\boldsymbol{X}$ is normally distributed with mean $\mu_y$ and variance $\sigma_y^{\,2}$. To simplify the notation, the $\boldsymbol{y}$ subscript will be dropped and the parameters $\mu_y$ and $\sigma_y^{\,2}$ will be written as $\mu$ and $\sigma^2$ respectively. $\boldsymbol{X}$ is then said to be lognormally distributed, or $\boldsymbol{X}$ is a $\Lambda$-variate written as $\boldsymbol{X}$ is $\Lambda(\mu, \ \sigma^2)$ and correspondingly $\boldsymbol{Y}$ is $N(\mu, \ \sigma^2)$. The distribution of $\boldsymbol{X}$ is therefore specified by the two parameters $\mu$ and $\sigma^2$. Clearly $\boldsymbol{X}$ must be positive and non-zero as the transformation $\boldsymbol{Y} = \log\boldsymbol{X}$ is not defined for zero or negative $\boldsymbol{X}$. The distribution functions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are then given by $\Lambda(x \mid \mu, \sigma^2)$ and $N(y \mid \mu, \sigma^2)$ respectively so that

$$\Lambda(x \mid \mu, \sigma^2) \;=\; P\{X \le x\} \qquad \text{(C.1)}$$

and

$$N(y \mid \mu, \sigma^2) \;=\; P\{Y \le y\} \qquad \text{(C.2)}$$

where $P\{X \le x\}$ and $P\{Y \le y\}$ represent the probabilities that $X \le x$ and $Y \le y$.

The two parameters of the lognormal distribution $\mu$ and $\sigma^2$ are related to the mean $\mu_x$ and standard deviation $\sigma_x$ of the untransformed $X$-variate by the equations:

$$\mu_x = \exp(\mu + {}^{1}/_{2}\,\sigma^2) \qquad \text{(C.3)}$$

$$\sigma_x = \exp(2\mu + \sigma^2)\,(\exp(\sigma^2)\ - 1) \qquad \text{(C.4)}$$

and the median is at $x = \exp(\mu)$.

### C.2.1 Estimation of the parameters $\mu$ and $\sigma^2$

Given a sample $S_n$ of data consisting of $n$ observations described by the real numbers $\{x_1, x_2, \ldots x_n\}$ sampled from $\Lambda(\mu, \sigma^2)$ a method is required for measuring the parameters $m$ and $s^2$, the *most likely* estimators of the parameters $\mu$ and $\sigma^2$. This is achieved by finding the maximum value of the Likelihood function of the sample given as:

$$\frac{1}{\sigma^n.(2\pi)^{n/2}.\prod_{i=1}^{n} x_i} \exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(\log x_i - \mu)^2\right\} \qquad \text{(C.5)}$$

from which the maximum likelihood estimators $m$ and $s^2$ are found to be:

$$m = \frac{1}{n}\sum_{i=1}^{n}\log x_i \qquad \text{(C.6)}$$

and

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(\log x_i - m)^2 \qquad \text{(C.7)}$$

Equations C.6 and C.7 can be used explicitly to find the maximum likelihood estimation of the parameters for continuous data sets (i.e. data sets in which each $x_n$ is given to its full precision). Radar data can be recorded to any precision, but at the expense of processing time and data storage space. The rainfall data analysed in this study were mostly at 8 bit precision (i.e. any integer value between 0 and 255 mm/h) and it was therefore necessary to adopt a slightly different approach when estimating the parameters of the continuous lognormal distribution of rainfall intensity on a CAPPI. In this case, the method of maximum likelihood is applied for *grouped data*, where the observations are considered to be grouped into 256 bins of even interval 1mm/h. Supposing that the interval $i$ is $(x_{i-1}, x_i)$ for $i = 1, 2, \ldots, 256$ and that $n_i$ of the total $n$ observations fall within the $i^{th}$ interval, then the likelihood of the sample is proportional to

$$\prod_{i} \left\{ \Lambda(x_i \mid \mu, \sigma^2) - \Lambda(x_{i-1} \mid \mu, \sigma^2) \right\}^{n_i} \qquad \text{(C.8)}$$

and the *log likelihood* function **L** is thus

$$L = \sum_{i} n_i \log\left\{ N\left( \frac{y_i - \mu}{\sigma} \middle| (0,1) \right) - N\left( \frac{y_{i-1} - \mu}{\sigma} \middle| (0,1) \right) \right\} \qquad \text{(C.9)}$$

for $i = 1, 2, 3, \ldots, 256$ where $y_i = \log x_i$ and for CAPPI data,

$y_i$ = log (rainfall rate $i$ mm/h )

$n_i$ = number of pixels in the masked CAPPI recording a rainfall rate of $i$ mm/h

The maximum value of Equation C.9 can be found changing its sign and using a function minimisation routine such as the *Downhill Simplex* method devised by Nelder and Mead (1965) the algorithm for which is given by Press et al. (1992).

### C.2.2 The nature of the lognormal distribution

The lognormal distribution is positively skew and it has been demonstrated in Section 7.1 that the marginal distribution of CAPPI data is also positively skew. That is to say that the *skewness* given by Equation C.10 is positive valued and therefore the distribution has an asymmetric tail which extends towards the more positive $x$.

$$Skewness = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{x_i - m_x}{s_x} \right]^3 \qquad \text{(C.10)}$$

where $m_x$ and $s_x$ are given by Equation C.3 and C.4 for the continuous case.

Figure (C.1) compares the frequency curves of two distributions, $N(0,1)$ and $\Lambda(0,1)$ illustrating the positive skewness of the latter and the positions of its mean, median and mode. In the case of the standard normal distribution $N(0,1)$ the mean, median and mode coincide at zero and the skewness is zero.

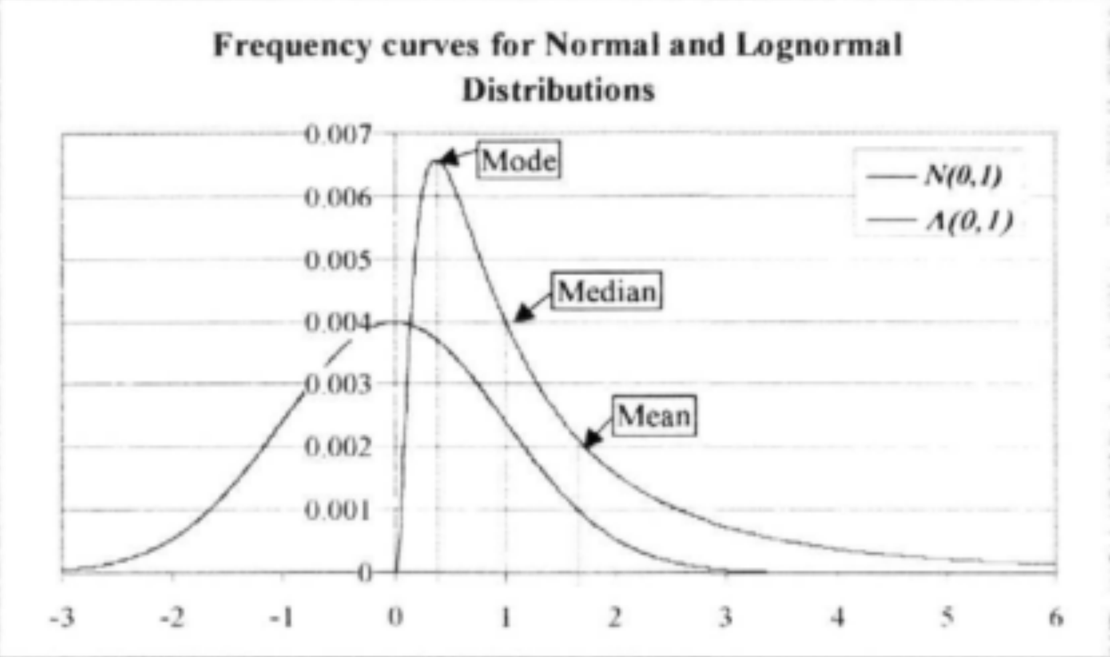


**Figure C.1 - Comparison of the normal and lognormal distributions**

The lognormal distribution is very flexible in so far as it is able to assume a wide variety of shapes depending on the ratio of its parameters $\mu$ and $\sigma$. This is illustrated in Figures (C.2) and (C.3) in which the effects of varying $\sigma$ whilst maintaining a constant $\mu$, and vice-versa, are depicted.

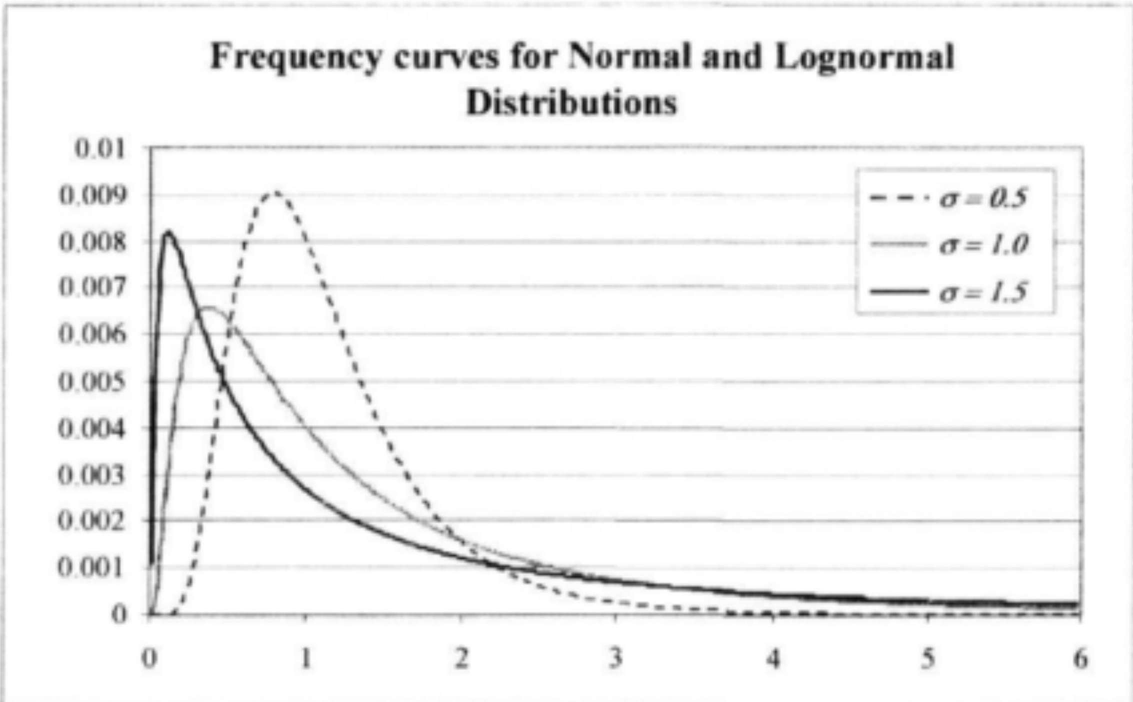


**Figure C.2 - The effect on the shape of the lognormal distribution of increasing $\sigma$ whilst maintaining a constant $\mu$**

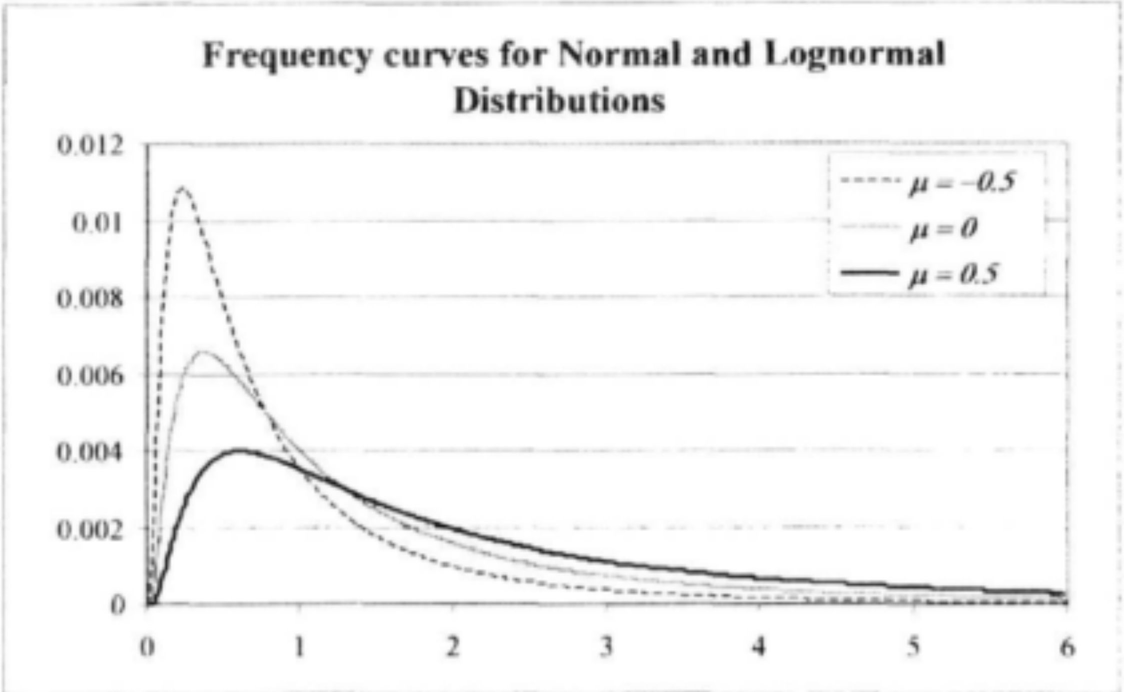


**Figure C.3 - The effect on the shape of the lognormal distribution of increasing $\mu$ whilst maintaining a constant $\sigma$**

The higher the value of the parameter $\sigma$ in relation to $\mu$, the more skew is the distribution. This kind of flexibility is essential when fitting a distribution to rainfall intensity data on a CAPPI. The distribution of rainfall intensities for a *stratiform event* (general rain of low intensity) will be very different from that of an intense *forced convective system* (large convective storm cells which are of the order of 50km in diameter) which in turn will be very different from that of a *free convective system* (isolated thunderstorms of the order of 10km in diameter).

### C.2.3 Goodness of fit

In spite of its good appearance, the two parameter lognormal fit is *not* perfect and a problem lies not in the tail of the fit, but in the first bin of the data (0 - 1mm/h). By fitting a simple two parameter lognormal distribution, it is assumed that the entire CAPPI is experiencing rainfall although in some parts of the CAPPI it is *extremely* light. This is not in fact the case, as parts of the CAPPI are actually dry. Owing to the integer nature of the data, these dry pixels are included in the first bin of the histogram. The result is that the number of pixels in the first bin is erroneously high and the skewness of the distribution of the rainfall intensities appears greater than it really is. The effect of this phenomenon is most clearly seen when examining the differences between the expected and observed histograms as shown in Figure C.4 in which the number of pixels in the second bin (1 - 2mm/h) is clearly over-estimated.

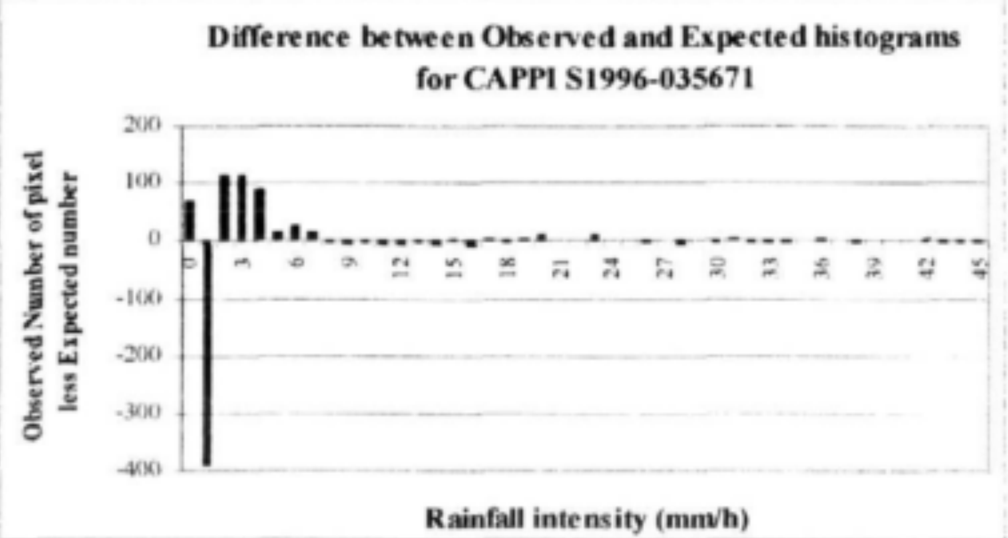


**Figure C.4 - Errors of fit of the lognormal distribution**

Critical examination of Figures 6 to 10 (Section 7.1) will also show this trend. For this reason, many of the fitted distributions will fail the well known Chi-square test. This problem was discovered too late in the study to repeat all of the analysis, however a possible solution (untested to date) is proposed here. At this stage the modelling process has been shown to be valid, and refinements will be dependent on a follow-on study.

A small modification to the methods presented thus far should improve the estimation of the parameters of the distribution although it has not yet been tested. CAPPI data are provided in integer format between 0 and 255mm/h. This does not give any information regarding the distribution of rainfall intensity between 0 and 1mm/h, therefore there is no way to distinguish between areas of light rainfall and areas that are dry. The methods given thus far assume that the entire CAPPI is experiencing rainfall, although it is *very* light in some parts of the CAPPI. Consider the complete set of $n$ pixels arranged into 256 groups (or bins). The first group of the CAPPI data comprises all rainrates $0 \leq i < 1$ including $n_0$ points which are dry ($i = 0$) and $n_1$ points which are experiencing very light rainfall ($0 < i < 1$mm/h). The remaining points $n_2$ all experience rainfall rates $1 \leq i \leq 255$mm/h. The points $n_0$ are not defined on the lognormal distribution and are shown as a vertical pole at 0mm/h. Figure C.5 depicts a mixed frequency diagram with a pole of values at zero, the remaining $n_1+n_2$ points being distributed continuously on the positive part of the real line.

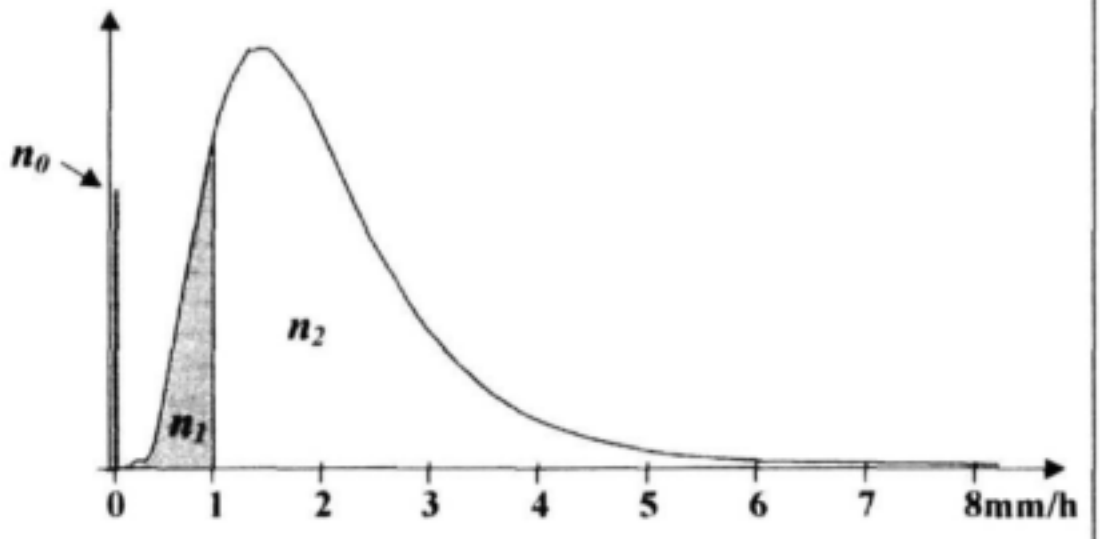


**Figure C.5 - Mixed frequency diagram of rainfall intensity on a CAPPI**

Clearly,

$$n = n_0 + n_1 + n_2 \tag{C.11}$$

therefore, $n_1$ must lie in the interval

$$0 < n_1 < n - n_2 \tag{C.12}$$

Since the function maximisation routine used in the parameter estimation algorithm is unconstrained, it is necessary to make use of the *logit transform* in order to express Equation C.12 in the unconstrained form of C.13.

$$q = \log\left(\frac{n_1}{n_0}\right) = \log\left(\frac{n_1}{n - nl - n_2}\right) \tag{C.13}$$

where $q$ is a dummy variable $(-\infty < q < \infty)$ such that, once it has been defined, $n_1$ can be found from it by reversing the transform as shown in Equation C.14:

$$n_1 = \frac{(n - n_2)\exp(q)}{1 + \exp(q)} \qquad \text{(C.14)}$$

Thus $q$ can be added as an extra variable in the log likelihood function which is passed to the function maximisation routine, thereby improving the estimation of the parameters $\mu$ and $\sigma^2$ and statistically defining the boundary between areas of the CAPPI which are receiving rainfall and those that are not.

## C.3 THE FOURIER TRANSFORM AND POWER SPECTRUM

A mathematical transform is simply a procedure whereby a function is *transformed* into another *domain* which may have some convenient properties that can simplify the manipulation of the function. A common example is the *Log Transform* where any positive, non-zero number can be expressed in terms of its logarithm. In the log domain, multiplication of positive numbers is replaced by the addition of their logarithms. Addition is a much simpler operation than multiplication and this transformation can save a great deal of time when multiplying large numbers by hand. Of course the transformation has no real value without an *inverse* to transform the logarithm back into the natural domain and this role is played by the *Exponential Transform*.

The Fourier transform is an essential ingredient in both the analysis and the modelling processes due to the fact that it can be calculated quickly for a large field of data through the use of the Fast Fourier Transform.

### C.3.1 The Continuous Fourier Transform in one dimension

Although the transform is not used in this form in the *String of Beads model*, this is the simplest case of the Fourier Transform and will serve to illustrate some of the properties of the Fourier Transform and how it can be applied to one dimensional data. A concise general development of this theory is given by Press et al. (1992) and the relevant sections are paraphrased in this section.

Analogous to the Log Transform, the *Fourier Transform* transforms a function, or a sequence of numbers in the case of the Discrete Fourier Transform, from the *time domain* into the *frequency domain*. The procedure which defines the Fourier Transform from a function of a complex variable $t$ in the *time domain* **$h(t)$** to its corresponding function in the *frequency domain* **$H(f)$** is defined by Equation C.15.

$$H(f) = \int_{-\infty}^{+\infty} h(t).e^{2\pi i.f.t}\,dt \qquad \text{(C.15)}$$

The term $e^{2\pi i.f.t}$ is a complex multiplier which can be represented as

$$e^{i\theta} = \cos\theta + i\sin\theta \qquad \text{where } \theta = 2\pi ft$$

The inverse transform, from the *frequency domain* into the *time domain* is given by Equation C.16.

$$h(t) = \int_{-\infty}^{+\infty} H(f).e^{-2\pi i.f.t}\,df \qquad \text{(C.16)}$$

### C.3.2 The Discrete Fourier Transform in one dimension

Rainfall data measured by radar are obtained in discrete space-time intervals and it is therefore necessary to extend this theory to its discrete form. Following Press et al. (1992), consider the case of a sample set containing $N$ consecutive values in the *time domain* $h_k$ ( = $h(t_k)$, $k = 0, 1, 2, \ldots, N-1$ , where $t_k = k\Delta$ and $\Delta$ is the sampling interval). Assuming that $N$ is even as will be the case in CAPPI data, with $N$ numbers of input, no more than $N$ independent numbers of output will be produced so that the Fourier Transform **$H(f)$** need only be estimated for the discrete values of $f_n$ where

$$f_n = \frac{n}{N\Delta}, \qquad n = -\frac{N}{2}, \ldots, \frac{N}{2} \qquad \text{(C.17)}$$

The extreme values of $n$ in Equation C.17 correspond to the lower and upper limits of the Nyquist critical frequency ( $f_c$ ) range. The Fourier Transform for frequencies beyond these limits cannot be calculated due to the fact that critical sampling of a sine wave is two sample points per cycle. It can be shown that the discrete form of Equation C.15 is given by

$$H(f_n) = H_n = \sum_{k=0}^{N-1} h_k.e^{2\pi i.k.n/N} \qquad (C.18)$$

Where $\boldsymbol{H_n}$ is the *Discrete Fourier Transform* of the $\boldsymbol{N}$ points $\boldsymbol{h_k}$. The Discrete Fourier Transform maps $\boldsymbol{N}$ complex numbers ($\boldsymbol{h_k}$'s) in the time domain onto $\boldsymbol{N}$ complex numbers ($\boldsymbol{H_n}$'s) in the frequency domain independently of the time interval between the sampled points. The inverse of the Discrete Fourier Transform is calculated using Equation C.19 which will recover *exactly* the $\boldsymbol{h_k}$'s from the $\boldsymbol{H_n}$'s.

$$h_k = \frac{1}{N}\sum_{n=0}^{N-1} H_n.e^{-2\pi i.k.n/N} \qquad (C.19)$$

### C.3.3 The Fast Fourier Transform

The Fast Fourier Transform is a method of computing the Discrete Fourier Transform using fewer calculations. The data sets used in the development of the *String of Beads model* are extremely large and it therefore becomes important to consider the computational efficiency of the algorithm used to extract the Fourier Transform. It has been shown that for a sample of $\boldsymbol{N}$ points, the computational effort required for the Fourier Transform is proportional to $\boldsymbol{N^2}$ calculations. The computational effort for the Fast Fourier Transform algorithm is proportional to $\boldsymbol{N.\log_2 N}$ calculations. To illustrate the significance of this saving, consider the case in which the simulation of a 40 hour rainfall event is required at 5 minute intervals and 128x128 spatial resolution. The saving in computational effort is of the order of 320 000 times which means the difference between 1 minute and 220 days on a modern computer.

The derivation of the *Decimation in Time Fast Fourier Transform algorithm* is covered in many texts (Kuc (1988), for example) and will not be repeated here. It takes advantage of certain symmetries in the transform by splitting the sequence $\boldsymbol{h_k}$ into two half sequences, even indexed $\boldsymbol{h_{k(even)}}$ and odd indexed $\boldsymbol{h_{k(odd)}}$, which are independently transformed and added together. Since $\boldsymbol{4N^2}$ real valued multiplications are required to compute an $\boldsymbol{N}$ point Discrete Fourier Transform, $\boldsymbol{4(N/2)^2}$ are required for each of the half sequences $\boldsymbol{h_{k(even)}}$ and $\boldsymbol{h_{k(odd)}}$, or a total of $\boldsymbol{2N^2}$ for both. The two half sequences can themselves be split into two half sequences again reducing the multiplication count. Decimation can continue until $N = 1$. It is for this reason that the sequence to be transformed must be an integer power of 2 in length. A sequence of arbitrary length should be padded with zeros to achieve this condition.

### C.3.4 The Fast Fourier Transform in two or more dimensions

Consider a complex valued function $\boldsymbol{h(k_1, k_2)}$ defined on a two dimensional space where $\boldsymbol{k_1} = 0, 1, \ldots, N_1 - 1$ and $\boldsymbol{k_2} = 0, 1, \ldots, N_2 - 1$, its complex valued two dimensional discrete Fourier Transform $\boldsymbol{H(n_1, n_2)}$ is given by Equation (C.20).

$$H(n_1,n_2) = \sum_{k_2=0}^{N_2-1} \sum_{k_1=0}^{N_1-1} \exp\left(\frac{2\pi.i.k_2.n_2}{N_2}\right).\exp\left(\frac{2\pi.i.k_1.n_1}{N_1}\right).h(k_1,k_2) \qquad \text{(C.20)}$$

This can be rearranged in the form of Equation (C.21).

$$H(n_1,n_2) = \sum_{k_2=0}^{N_2-1} \exp\left(\frac{2\pi.i.k_2.n_2}{N_2}\right).\sum_{k_1=0}^{N_1-1} \exp\left(\frac{2\pi.i.k_1.n_1}{N_1}\right).h(k_1,k_2) \qquad \text{(C.21)}$$

In this way the two dimensional transform can be computed by taking sequential one dimensional Fast Fourier Transforms on each index of the original function (in this case 1 and 2). The inverse transform $\boldsymbol{h(k_1, k_2)}$ for the two dimensional case is given by:

$$\frac{1}{N_2.N_1} \sum_{k_2=0}^{N_2-1} \sum_{k_1=0}^{N_1-1} \exp\left(-\frac{2\pi.i.k_2.n_2}{N_2}\right).\exp\left(-\frac{2\pi.i.k_1.n_1}{N_1}\right).H(n_1,n_2) \qquad \text{(C.22)}$$

The theory presented here for the two dimensional case can be shown to extend to multi-dimensional space. Efficient programming of this algorithm is not a trivial exercise. The routine given by Press et al. (1992) was used for this study.

### C.3.5 Estimating the power spectrum using the FFT

Given a one dimensional sample of $\boldsymbol{N}$ complex valued points (where $\boldsymbol{N}$ is an integer power of two) in the time domain $\boldsymbol{h_k}$ (where $\boldsymbol{k} = 0, 1, \ldots, N-1$), the one dimensional power spectrum is proportional to $| \boldsymbol{H_n} |^2$, where the $\boldsymbol{H_n}$ ($\boldsymbol{n} = -{}^{N}/_{2}, \ldots, {}^{N}/_{2}$) are defined by Equation C.18 and represent the corresponding sequence in the frequency domain. That is to say that in order to *estimate the power spectrum of a sequence of $N$ complex valued points* $\boldsymbol{h_k}$, *the sequence is* transformed into Fourier space via the FFT to get the transformed complex valued sequence $\boldsymbol{H_n}$ and the power spectral density $\boldsymbol{P(f_n)}$ corresponding to each frequency $(f_n)$ is obtained by squaring the $\boldsymbol{H_n}$ term.

### C.3.6 Some properties of power spectra

An example of a one-dimensional *real valued* sequence and its power spectrum is given in Figure C.6. The Fourier Transform of a *real valued sequence* $\boldsymbol{h_k}$ is a complex valued sequence in which $\boldsymbol{H_n}$ $(\boldsymbol{n} = -{}^{N}/_{2}, \ldots, {}^{N}/_{2})$ is the complex conjugate of $\boldsymbol{H_{-n}}$. Since the absolute values of a complex number and its complex conjugate are equal, only half of the power spectrum is plotted (i.e. $|\boldsymbol{H_n}|^2$, where $(\boldsymbol{n} = 0, \ldots, {}^{N}/_{2})$). Although it is only a one-dimensional example, the sequence has some properties similar to those observed in CAPPI data. Firstly, all the points are positive, real numbers and the sequence appears to have a strong auto-correlation. This is also the case in CAPPI data and the result of this property is that the pixels on the CAPPI have a definite clustered appearance. The second point is the power-law relationship between the frequency and the power spectral density - the logarithm of the power is proportional to the logarithm of the frequency as shown in Equation C.23.

$$\log P(f_n) \sim \log f_n \qquad \text{(C.23)}$$

which implies that

$$P(f_n) \sim f_n^{\;-\beta} \qquad \text{(C.24)}$$

for some exponent $\beta$. A random field will exhibit *scaling* properties in its statistical moments if the power spectrum of the field takes the form of Equation C.24. This in turn means that the spatial correlation on the image is uniquely defined by a single parameter. In the extreme case where $\beta = 0$, the data in the sequence are completely uncorrelated with each other and this would indicate a pure noise process. Conversely a high value of $\beta$ would indicate a highly correlated sequence. Finally, a $\beta$ value greater than the dimension of the sequence indicates a non-stationary underlying process in the sense that it does not have a fixed mean value. The $\beta$ exponent of the example presented in Figure C.6 has a value of 1.93 and since the original process is one-dimensional it is likely to have been sampled from a non-stationary process.

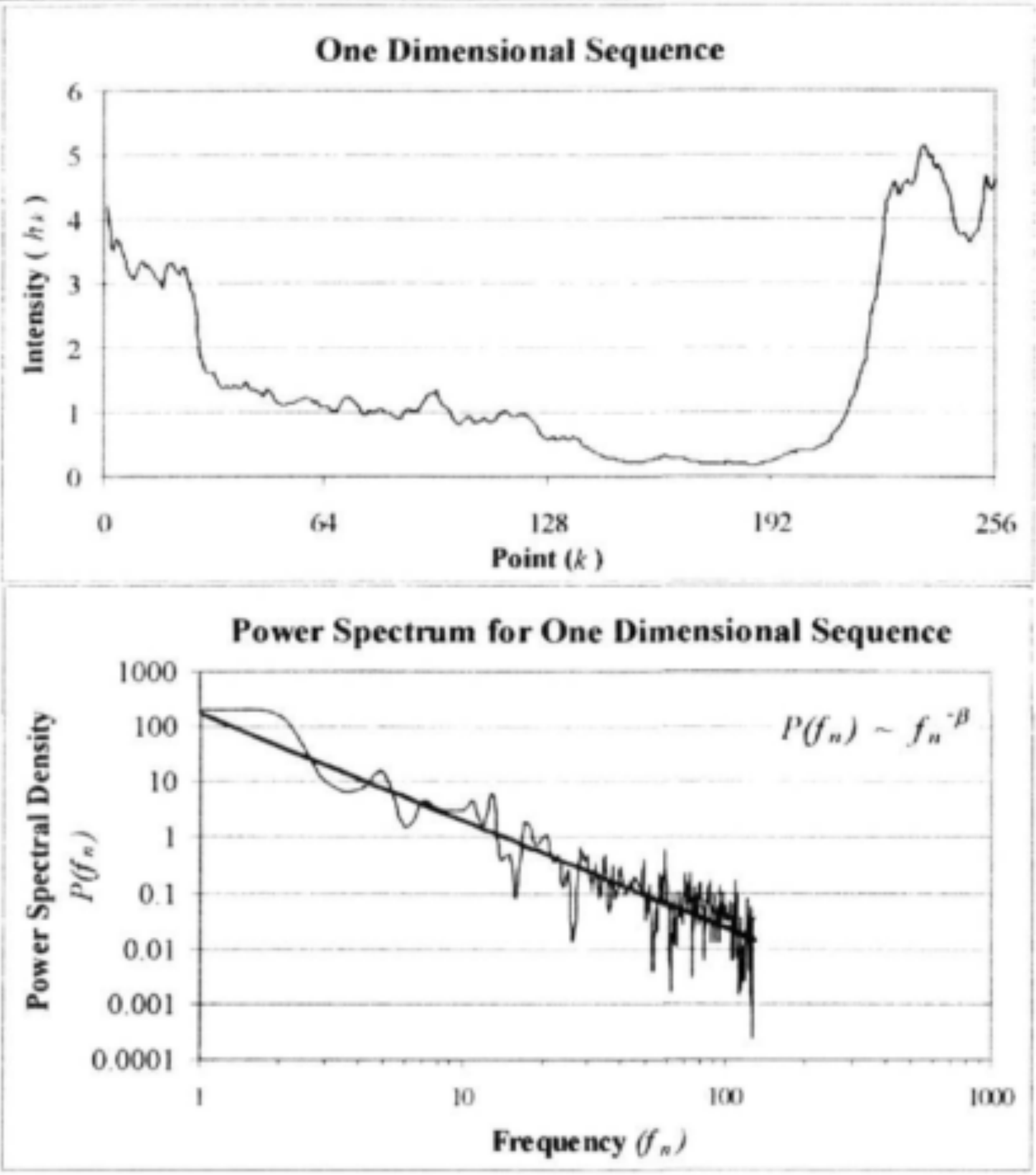


**Figure C.6 - Example of a 1-D real sequence and its power spectrum**

The ideas presented for the one dimensional case in Sections C.3.5 and C.3.6 can easily be extended to multi-dimensional space, the only difficulty being that n-dimensional space produces a corresponding n-dimensional power spectrum which becomes difficult to visualise and interpret.

## C.4 THE GENERALISED STRUCTURE FUNCTION

Although the gradient of the radially averaged two dimensional power spectrum *($\beta$)* is a sufficient descriptor of the spatial correlation structure of the rainfall intensities on a CAPPI, an important tool for validating the modelling procedure is through the use of the so called *Generalised Structure Function*. This has become a widely used technique in multifractal fields since the work of Anselmet et al. (1984). A concise description of this technique with specific reference to rainfall fields is given by Menabde et al. (1998). Paraphrasing their work, the Generalised Structure Function $G_q(l)$ is defined on a non-negative scalar random field R(x) as being the ensemble average of the $q^{th}$ power of the absolute differences of the $R$ values a distance $l$ apart. That is

$$G_q(l) = \left\langle \left| R(\vec{x} + \vec{l}\,) - R(\vec{x}) \right|^q \right\rangle \qquad \text{(C.25)}$$

In addition, the non-stationary random field is called ***multiaffine*** (Benzi et al., 1993) if $\boldsymbol{G_q(l)}$ satisfies the condition:

$$\boldsymbol{G_q(l) \propto l^{\zeta(q)}} \qquad \text{(C.26)}$$

where $\boldsymbol{l} = |\,\vec{l}\,|$ and $\boldsymbol{\zeta(q)}$ is some non-linear function of $\boldsymbol{q}$. For the case in which $\boldsymbol{q} = 2$, $\boldsymbol{\zeta(2)}$ is related to $\boldsymbol{\beta}$, the gradient of power spectrum, by Equation C.27.

$$\boldsymbol{\beta = D + \zeta(2)} \qquad \text{(C.27)}$$

where $\boldsymbol{D}$ is the dimension of the data ($\boldsymbol{D} = 2$ for CAPPI data).

For the purposes of this study, the generalised structure function of Equation C.25 was approximated for CAPPI data in the two principle directions (north-south and east-west) and for 6 discrete values of $\boldsymbol{l}$ chosen to be 1, 2, 4, 8, 16, 32 and 64. Since both points $\boldsymbol{R(x)}$ and $\boldsymbol{R(x + l)}$ of Equation C.25 are required to be within the unmasked region of the CAPPI, significant areas of the CAPPI do not contribute to the calculation of the structure function for the larger values of $\boldsymbol{l}$. This is illustrated in Figure C.7. The area shown in yellow represents the area of the CAPPI which contributes to the calculation of the structure function in both the east-west and the north-south directions. Areas shown in red and green represent regions of the CAPPI which contribute to the structure function in only the east-west and the north-south directions respectively. Regions shown in black do not contribute to the calculation of the structure function. The image on the left shows the case in which $\boldsymbol{l} = 64$ and the image on the right when $\boldsymbol{l} = 16$.

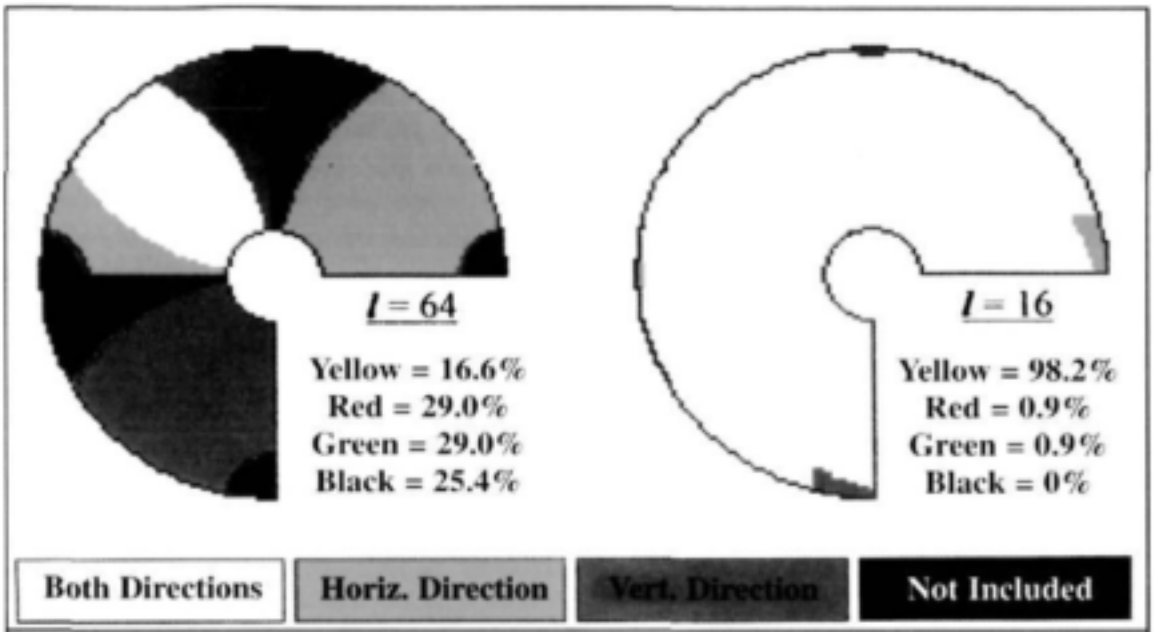


**Figure C.7 - Areas contributing to GSF calculation for different values of *l***

Less than 75% of the CAPPI is used to calculate the structure function in the case where *l* = 64. In comparison, the entire CAPPI is used for the case when *l* = 16 and more than 98% of the CAPPI is used in both the north-south and east-west directions. Consequently the behaviour of the structure function at high values of *l* is dependent on the distribution of rainfall on the CAPPI and is somewhat erratic. An example of a structure function is plotted in Figure C.8.

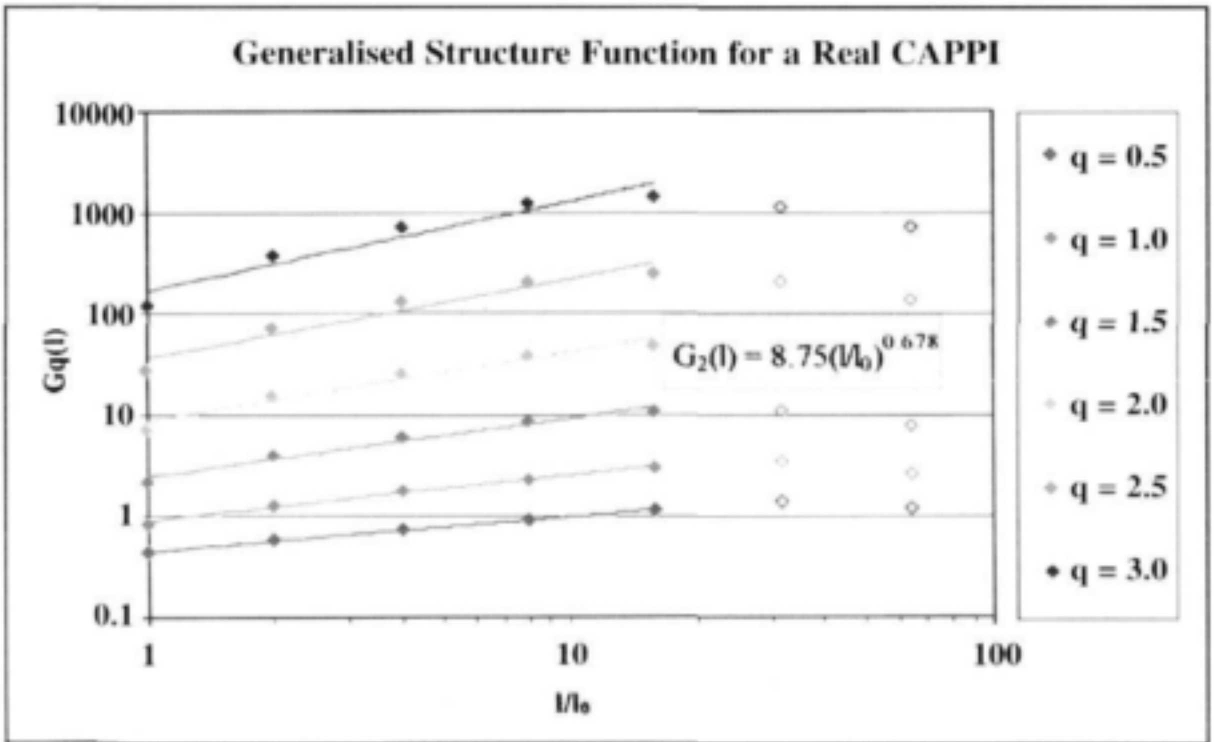


**Figure C.8 - Typical example of Structure Function for a real CAPPI**

For a chosen $l$, beginning in the north-west corner of the CAPPI, the position of the first point $R(x, y)$ $(x, y = 0,1,\dots,127)$ was first checked to ensure that it fell within the limits of the $^3/_4$ *doughnut sample area*. Next the position of the point $R((x + l), y)$ was checked in a similar fashion. If both $R(x, y)$ and $R((x + l), y)$ were found to be within the limits of the sample area, their absolute difference was calculated which was then raised to various powers $q$ and stored in a table. This process was then repeated in the north-south direction. The ensemble average of the absolute differences in intensity raised to a power $q$ for all pairs of pixels a distance $l$ apart was then calculated for selected values of $q$.

The Generalised Structure Function is defined for all $q \geq 0$ and has been plotted with logarithmic axes in Figure C.8 for discrete values of $l/l_0$ up to 64, and discrete values of $q$ up to 3.0. For values of $q$ less than 3.0 and $l/l_0$ less than 16 the function, $\log G_q(l)$ versus $\log (l/l_0)$ is approximately linear for a given $q$. In addition, the gradients $\zeta(q)$ of the linear approximations appear to be a function of increasing $q$. These observations in real rainfields suggest the *multiaffine* behaviour defined in Equation C.26 and can be expressed in the form of Equation C.28.

$$\log G_q(l) \propto \zeta(q) \log l \qquad \text{(C.28)}$$

The gradients $\zeta(q)$ of Equation C.28 are obtained by fitting power law relationships by least squares in the linear space of $G_q(l) \sim l/l_0$. These can be plotted for discrete values of $q$ and an example of such a plot is given in Figure C.9 for the six values of $q$ shown in Figure C.8. According to the relationship defined in Equation C.27 for a two dimensional field, the $\zeta(2) = 0.678$ shown in Figure C.9 should correspond to a $\beta$ of 2.678. The measured $\beta$ for this CAPPI was found to be 2.692.

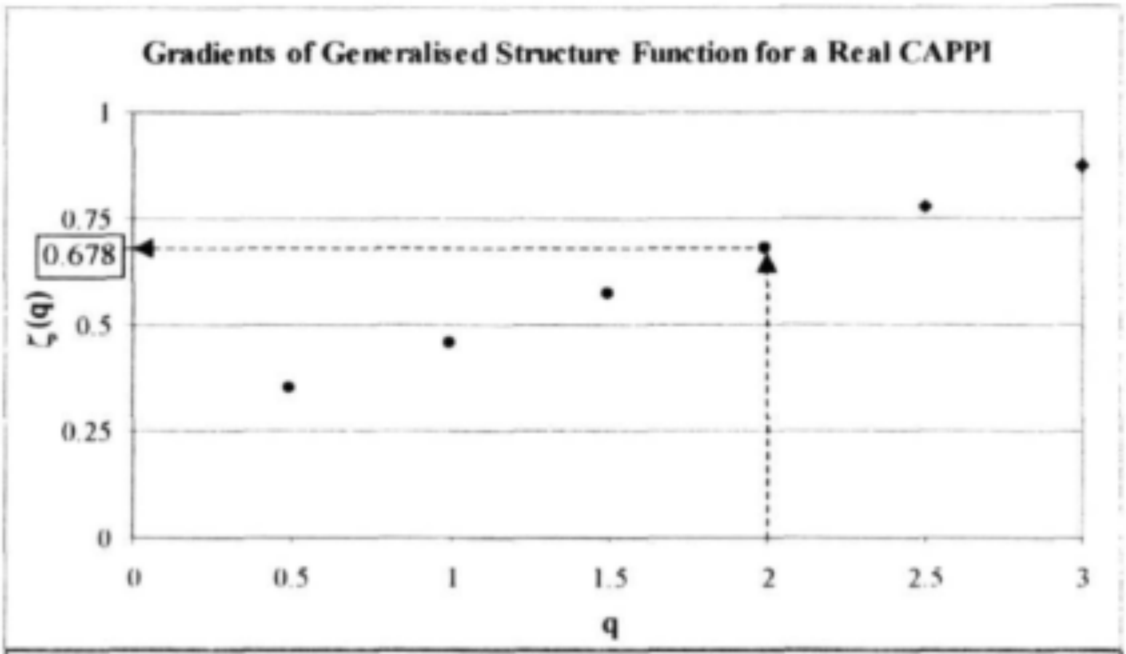


**Figure C.9 - Gradients of the Generalised Structure Function presented for a real CAPPI**

Unlike the power spectrum and the marginal distribution, the *Generalised Structure Function* is not used to determine the input parameters ($\sigma$, $\beta$ and $\mu$) of the String of Beads model and it therefore serves as a useful *validation test* for the model output.

# Appendix D - Random Number Generation

## D.1 INTRODUCTION

The object of the *String of Beads model* is to generate a realistic pseudo-random rainfall scenario in space-time. One of the key ingredients in the model is a good *uniformly distributed pseudo-random number generator*. Such generators commonly available range from reasonable to awful. The purpose of this appendix is to revisit this well worn topic and indicate the care taken to ensure that the simulations of the String of Beads model do not suffer from a poor generator. Two algorithms are presented in this appendix; the first is *an efficient and portable pseudo-random number generator* (Wichmann and Hill, 1982); the second is to transform a sequence of uniformly distributed random numbers into a white noise sequence using, *the percentage points of the normal distribution* (Beasley and Springer, 1985). A quantitative measure of the quality of output for both algorithms is given.

## D.2 PSEUDO-RANDOM NUMBER GENERATION

Rainfall, like many other natural phenomena, can be effectively modelled as a random process. The randomness of the sequence of pseudo-random numbers produced by the random number generator is of great importance in the simulation process. Algorithms which are built into the compiler software cannot always be relied upon to produce sufficiently random numbers and for this reason one devised by Wichmann and Hill (1982), was adopted to generate *uniformly distributed* random noise.

Strictly speaking, the sequences of uniformly distributed numbers used are *pseudo-random* and make use of an algorithm which calculates a number based on the previous number in the sequence. This ensures that any generated sequence of random numbers can be repeated when necessary by initialising the algorithm with the same seed. Obviously the random number generator should not refer to the time or date or any other volatile variable for its initial seed if the sequence is to be reproducible.

Several factors were considered in the selection of the pseudo-random number generator. The algorithm needed to be *portable* in the sense that it should be machine independent. Routines which make use of low-level commands such as bit shifts and other binary arithmetic, rely on a format of data storage which is specific to the computer operating system, the programming language and the brand of compiler. If the routine is to be portable these low-level commands should be avoided. The *randomness* of the original uniformly distributed sequence is of great significance in the model. Simulation of the rainfall for a single day at 128x128 pixel, five minute resolution requires 5 million random numbers. The consequences of using a poor random number generator would be that patterns would soon become apparent in the simulated rainfields. Another important factor considered in the selection of the algorithm was the speed or efficiency. An algorithm which involves too many arithmetic operations would result in a very slow program when generating long sequences of numbers.

### D.2.1 The multiplicative congruential generator

The method chosen was that suggested by Wichmann and Hill (1982) who make use of a combination of three *multiplicative congruential* generators. The algorithm AS 183 is published in Applied Statistics Algorithms (Griffiths and Hill, 1995). Uniformly distributed random number generation is discussed at length by Knuth (1969) and his description of the *linear congruential* generator is outlined here. The *multiplicative congruential* sequence is a special case of the *linear congruential* sequence ($X_n$) defined by Equation D.1.

$$X_{n+1} = (a.X_n + c) \bmod m, \qquad n \geq 0. \qquad \text{(D.1)}$$

where

$X_0$ is the starting value or *seed*; $X_0 \geq 0$

$a$ is the *multiplier*; $a \geq 0$

$c$ is the *increment*; $c \geq 0$

$m$ is the *modulus*; $m > X_0,\ m > a,\ m > c$

and the "mod" function is the remainder when a number is divided by a divisor.

It can be shown that all such sequences will eventually repeat themselves for a large enough $n$ known as the *period*. A larger period implies a ***more random*** sequence of numbers. The choice of these four parameters is critical to the period of the sequence and an example given by Knuth (1969) of a *poor* choice of parameters is when $X_0 = a = c = 7$, $m = 10$. In this case the second value of the sequence ($X_1$) would be calculated as

$$\begin{aligned} X_1 &= (7.7 + 7) \bmod 10 \\ &= 56 \bmod 10 \\ &= 6 \end{aligned}$$

and the entire sequence $X_n$ would be

$$X_n = 7, 6, 9, 0, 7, 6, 9, 0, 7, 6, 9, 0 \ldots \qquad (n \geq 0)$$

This sequence has a period of length 4 and is therefore not useful as a random number generator.

For the special case in which $c = 0$, the number generation is slightly faster with a small sacrifice in the length of the period. This is often referred to as the *multiplicative congruential* method and is the only case that will be considered in this document. Equation D.1 then reduces to Equation D.2.

$$\boldsymbol{X_{n+1} = a.X_n \text{ mod } m, \qquad n \geq 0} \qquad \text{(D.2)}$$

The modulus $\boldsymbol{m}$ should be chosen as a large number since the period can never have a value larger than $\boldsymbol{m}$. Furthermore, if $\boldsymbol{m}$ is a prime number the maximum period possible is $\boldsymbol{(m - 1)}$ and this can be achieved provided that two conditions are satisfied. These are

(1) $\boldsymbol{X_0}$ is relatively prime to $\boldsymbol{m}$

(2) $\boldsymbol{a}$ is a *primitive root* of $\boldsymbol{m}$

If $\boldsymbol{m}$ is chosen as a prime number then $\boldsymbol{X_0}$ will automatically be prime relative to $\boldsymbol{m}$.

The number $x$ is said to be a *primitive root* of $\boldsymbol{m}$ if there exists an integer value $\boldsymbol{k}$ which satisfies the relationship given in Equation D.3.

$$\boldsymbol{x^k \text{ mod } m = 1} \qquad \text{(D.3)}$$

At least one solution ($\boldsymbol{x}$) to Equation D.3 exists for all prime numbers $\boldsymbol{m}$.

As previously mentioned, Wichmann and Hill (1982) make use of three multiplicative congruential generators each of which uses a prime number for its modulus and a corresponding primitive root for its multiplier. The modulus and its corresponding multiplier for each of the generators are listed as

| | *modulus* | *multiplier* |
|---|---|---|
| (1) | 30269 | 171 |
| (2) | 30307 | 172 |
| (3) | 30323 | 170 |

Although the sum of $n$ independent rectangular distributions will tend to a normality as $n$ tends to infinity, the *fractional part* of a sum of rectangular distributions will remain rectangular for all $n$. This property is exploited by Wichmann and Hill (1982) in order to compensate for imperfections in the randomness of the individual generators. Each generator is seeded with an initial value between 1 and 30000, $X_{10}$ , $X_{20}$ , $X_{30}$ respectively. The results of the three generators $X_{1n}$ , $X_{2n}$ , $X_{3n}$ are then added and the fractional part of the sum is taken as the uniformly distributed pseudo-random number $R(0,1)$.

Wichmann and Hill (1982) claim that since each generator has a period of 1 less than its prime modulus, the periods are all of even length and therefore have a common factor of 2 but no other common factors exist for the values chosen. For this reason the three generators are not completely independent but the results have been well tested and perform very well. Due to the common factor of 2 in the periods of the generators, they have a combined period of one quarter of the product of the individual periods. That is a period exceeding $6.95 \times 10^{12}$. Considering that $5 \times 10^{6}$ random numbers are required for a single day's simulation at 128 x128 pixel, 5 minute resolution, about 3800 years of simulated rainfall data of a continuous wet period could be generated before the pseudo random sequence starts to repeat itself and a cyclic behaviour might possibly be observed in the CAPPIs.

### D.2.2 Percentage points of the normal distribution

The *String of Beads model* makes use of *normally distributed* random noise in the generation of artificial CAPPIs. Having established an effective means of generating uniformly distributed random noise, the next step is to transform it into normally distributed noise by means of an inverse normal transformation.

The normal distribution for given mean $\mu$ and standard deviation $\sigma$ is defined by Equation D.4.

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \qquad -\infty < x < \infty \qquad \text{(D.4)}$$

The variable $x$ is normally distributed, with mean $\mu$ and variance $\sigma^2$ and this is written as $x \sim N(\mu, \sigma^2)$. In the standard normal case $\mu = 0$ and $\sigma = 1$ and this equation reduces to (D.5).

$$f(x;0,1) = \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{x^2}{2}\right], \qquad -\infty < x < \infty \qquad \text{(D.5)}$$

Equation D.5 describes the well known bell-shaped probability density function which can be represented in the cumulative form of Figure D.1 which is described by Equation D.6.

$$p = P(x < x_p) = \int_{-\infty}^{x_p}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{x^2}{2}\right].dx \qquad \text{(D.6)}$$

where $p$ is the probability that a randomly chosen number $x$ is less than a chosen $x_p$.

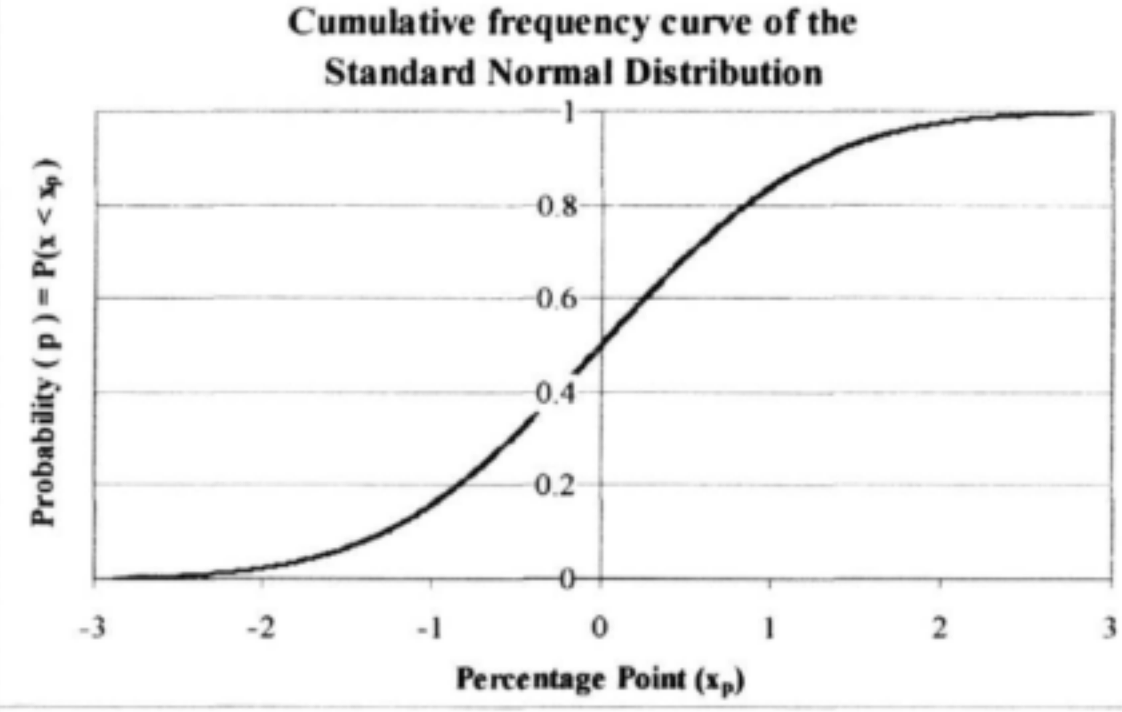


**Figure D.1 - Cumulative frequency curve for the Standard Normal distribution**

A uniformly distributed random variate $Y \sim U(0,1)$ can be converted to a standard normally distributed random variate $X \sim N(0,1)$ by setting $Y = p$ in Equation D.6 and solving for $X = x_p$. Equation D.6 has no known solution and must therefore be approximated.

A good approximation is given by Beasley and Springer (1977) and published as AS111 in Applied Statistics Algorithms (Griffiths and Hill, 1995). Paraphrasing from that text, the routine replaces the $p$ of Equation D.6 by a dummy variable $q = p - 0.5$ and then compares $|q|$ with 0.42. If $|q| \leq 0.42$, $x_p$ is obtained through the rational approximation of Equation D.7.

$$x_p = q.A(q^2)\ B(q^2) \qquad \text{(D.7)}$$

where $A$ and $B$ are polynomials of degrees 3 and 4 respectively. If $|q| > 0.42$, another dummy variable $r = \sqrt{\{\ln(0.5-|q|)\}}$ is introduced and $x_p$ is calculated from a different rational approximation given by Equation D.8.

$$x_p = \pm C(r)\ D(r) \qquad \text{(D.8)}$$

where $C$ and $D$ are polynomials of degrees 3 and 2 respectively, and the sign is taken to be that of $q$. They claim that in the absence of rounding error, the $x_p$ calculated for a given $p$ will correspond to a true value $p'$ which will satisfy the relationship

$$|p' - p| < 2^{-31}$$

In the presence of rounding error, they claim an accuracy which will satisfy the relationship

$$|p' - p| < \varepsilon$$

Where $\varepsilon$ is of the order of 20 times the value of the smallest digit in the mantissa. The routine was re-written in the C programming language using 64 bit double precision with an 11 bit exponent and a 52 bit mantissa, the remaining bit used for sign. In this case the value of $\varepsilon$ would be $20 \times 2^{-52}$ which is much less than $2^{-31}$ so rounding error can be neglected in this case and the algorithm is used to its full potential.

# Appendix E - Power-law Filtering

## E.1 INTRODUCTION

The method of simulating sequences of CAPPIs presented in this section are based on the work of Bell (1987), Brenier (1990) and Wilson et al. (1991). Bell (1987) used the technique of power-law filtering of a Gaussian noise field to simulate rainfields in three dimensional space-time (two space and one time) as viewed by satellite. This idea was used by Brenier (1990) and Wilson et al. (1991) to produce a sequence of CAPPIs in time which were assembled into an animated image and published on video. CAPPI sequence simulation in the String of Beads model is done using a modified version of their techniques and this will be illustrated first in the simple one-dimensional case and then extended into two and three dimensions.

## E.2 ONE-DIMENSIONAL POWER-LAW FILTERING

It was shown in Section 7 that three parameters $\sigma$, $\beta_{space}$ and $\mu$ are required to describe the spatial distribution of rainfall on a CAPPI. $\mu$ and $\sigma$ are the mean and standard deviation of the pixel logs respectively, whilst $\beta_{space}$ is the gradient of the radially averaged two-dimensional power spectrum. Any *pure* noise process is uncorrelated and therefore has a $\beta$ of zero.

By power-law filtering a field of random data it is possible to increase the correlation of the data from zero to any desired level $\beta$. This is done by transforming the sequence into Fourier space via the Fast Fourier Transform, multiplying (pointwise) the transformed sequence by a filter function, and then back transforming into natural space. In the one-dimensional case, the power-law filter is defined in one-dimensional, *complex* valued Fourier space, as a distance function which decays exponentially (with exponent $^{\beta}/_{2}$ ) from the first frequency to the Nyquist critical frequency and is then reflected about that point. The *scalar* coefficients of a typical filter are plotted in Figure E.1.

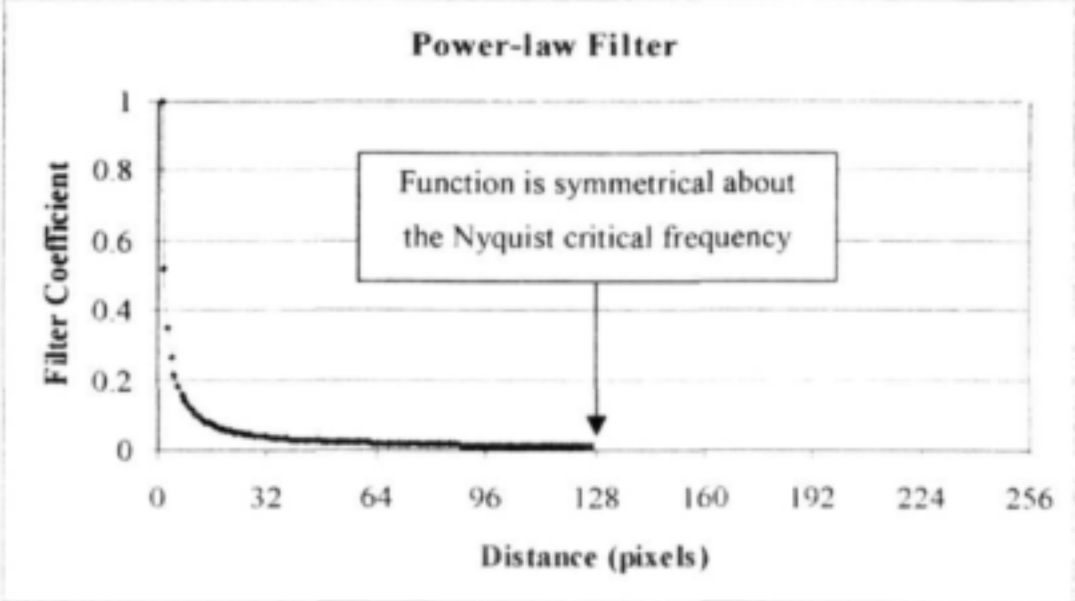


**Figure E.1 - Plot of scalar coefficients for a typical power-law filter**

As an illustration of the power-law filtering process, consider the one-dimensional sequence of 256 standard, normally distributed, real valued random numbers plotted in Figure E.2.

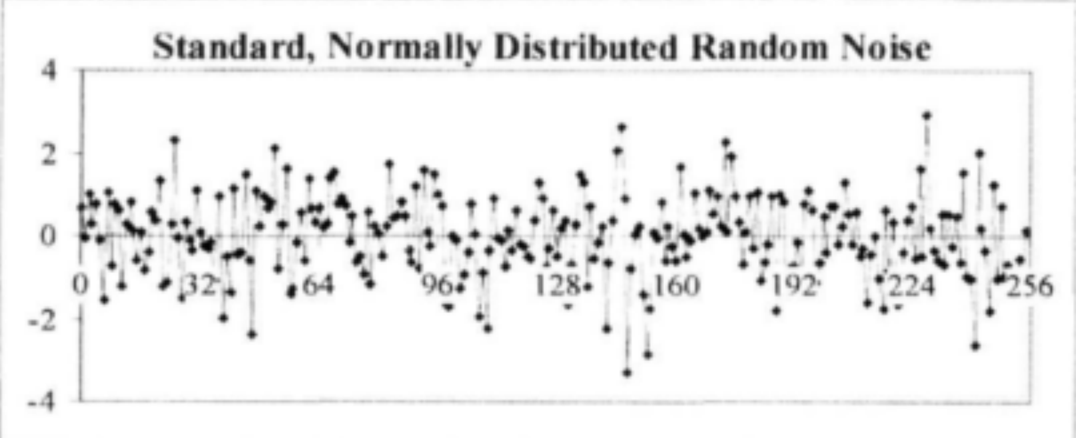


**Figure E.2- Sequence of standard, normally distributed random numbers**

This *real valued* sequence can be transformed into Fourier space via the FFT to produce a sequence of 256 *complex valued* numbers comprised of 128 pairs of random complex conjugates. Next, each *complex valued* member of the transformed sequence is multiplied by its corresponding *scalar* filter function coefficient at that point. The *real* and *imaginary* components of the product of the transformed sequence and its scalar multiplier, are plotted separately in Figure E.3. The conjugate symmetry of this power-law filtered sequence is illustrated by the fact that the real component of the sequence is symmetrical about the

Nyquist frequency (128), whilst the imaginary component exhibits skew symmetry about the Nyquist frequency.

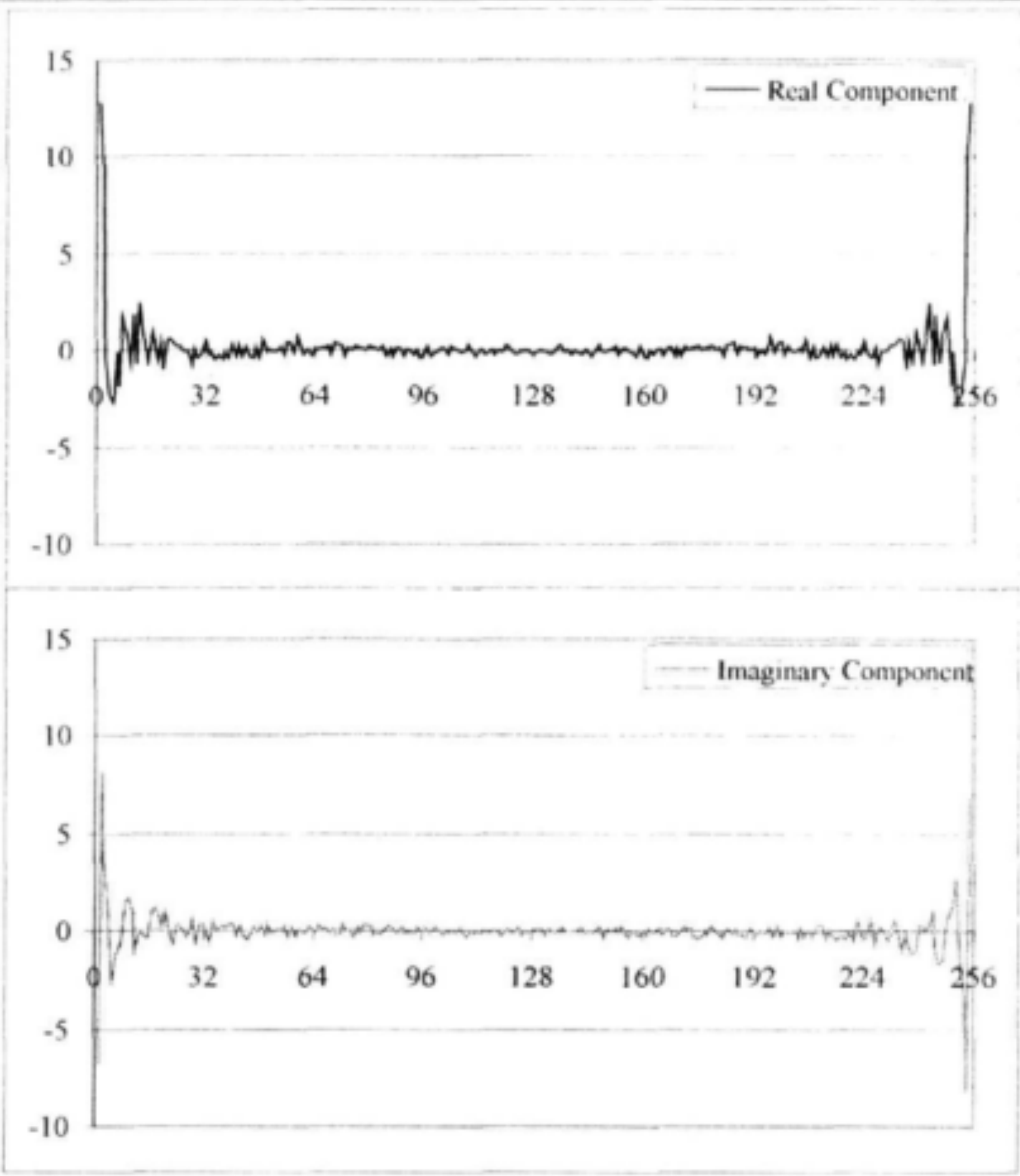


**Figure E.3 - Real and imaginary components of the product of the scalar filter of Figure E.1 and the complex valued, Fourier transformed sequence of random noise**

Using the *Inverse Fast Fourier Transform*, the sequence of complex numbers in Fourier space shown in Figure E.3 can then be back-transformed into natural space to yield the corresponding sequence of real valued, normally distributed random numbers which now have a correlation structure defined by $\beta$. The sequence can then be scaled and shifted to achieve the required marginal distribution with parameters $\mu$ and $\sigma$. For the purposes of this example, $\mu$ and $\sigma$ values of 0 and 1 respectively were chosen and the resulting sequence is plotted in Figure E.4.

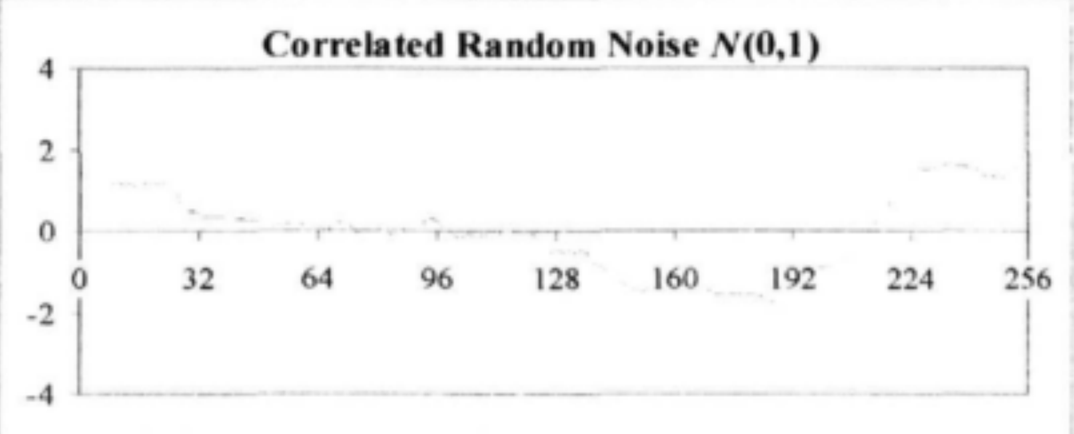


**Figure E.4 - Power-law filtered (Exponent β = 1.9) sequence of the standard, normally distributed random noise given in Figure E.2**

The sequence of numbers plotted in Figure E.4 has the same marginal distribution as that of Figure E.2, however the spatial correlation has been introduced by power-law filtering the sequence and thereby increasing the $\beta$ parameter from 0 in Figure E.2, to 1.9 in Figure E.4. Since this sequence is normally distributed $N(\mu,\sigma)$, the conversion to a lognormal distribution $\Lambda(\mu,\sigma)$ is achieved by exponentiating each element in the sequence. The result is a *real* valued sequence of *positive* numbers, plotted in Figure E.5, with marginal distribution $\Lambda(\mu,\sigma)$ and a spatial correlation defined by $\beta$.

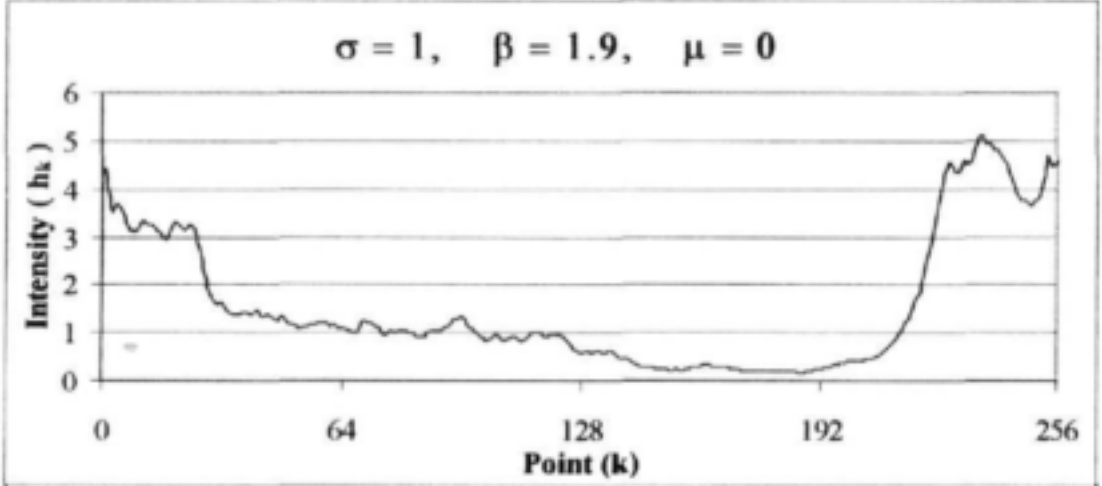


**Figure E.5 - Lognormally distributed, power-law filtered white noise**

By varying the $\beta$ parameter, it is possible to alter the spatial correlation in the filtered sequence. A higher exponent in the filter results in a smoother sequence. This is due to the fact that there is more power in the lower frequency Fourier coefficients and therefore a much higher correlation between any point in the sequence and its immediate neighbours. Figure E.6 is a plot of a sequence which was generated using the set of white noise plotted in Figure E.2, but with a $\beta$ parameter of 3.0 as opposed to the value of 1.9 used to produce the sequence plotted in Figure E.5. The two sequences have been scaled and shifted so that they have identical marginal distributions.

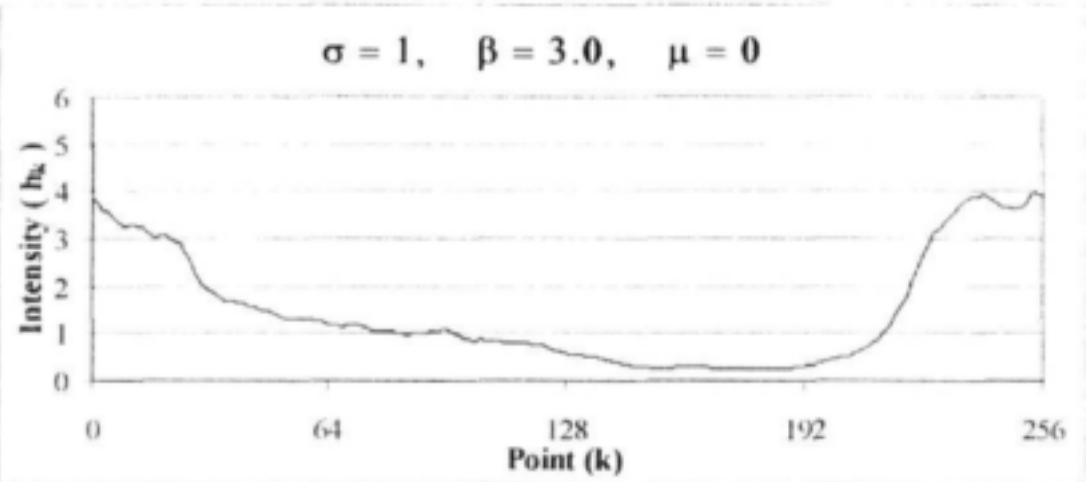


**Figure E.6 - Lognormally distributed, power-law filtered white noise, the greater β parameter produces a smoother sequence**

Figures E.5 and E.6 are plotted to the same scale and illustrate how the higher exponent power-law filter smoothes out the extreme points in the sequence. The *wrapped nature* of the Fourier filtered sequence can be seen by the fact that both generated sequences finish with the intensity with which they began. This property of Fourier filtered sequences is exploited in Section 9.

The power spectra of these two sequences are plotted in Figures E.7 and E.8 and least squares approximations of their gradients have been calculated. The gradients of 1.928 and 3.028 are very close to the intended values of 1.900 and 3.000. The additional 0.028 is due to the fact that the original noise process is not *absolutely pure* and prior to any filtering had a spectral gradient of 0.028.

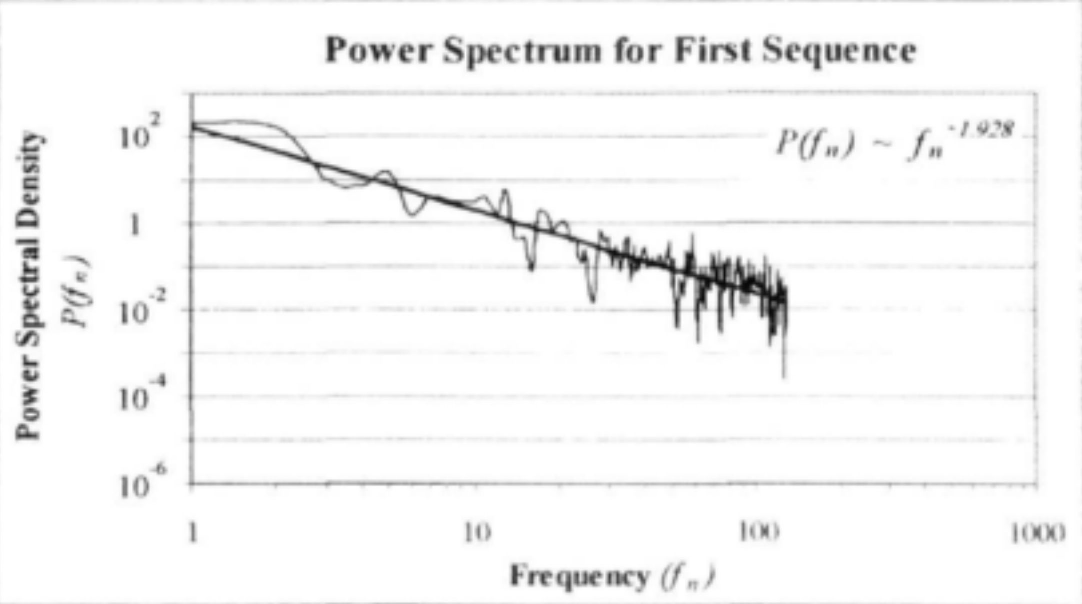


**Figure E.7 - Power spectrum for sequence plotted in Figure E.4**

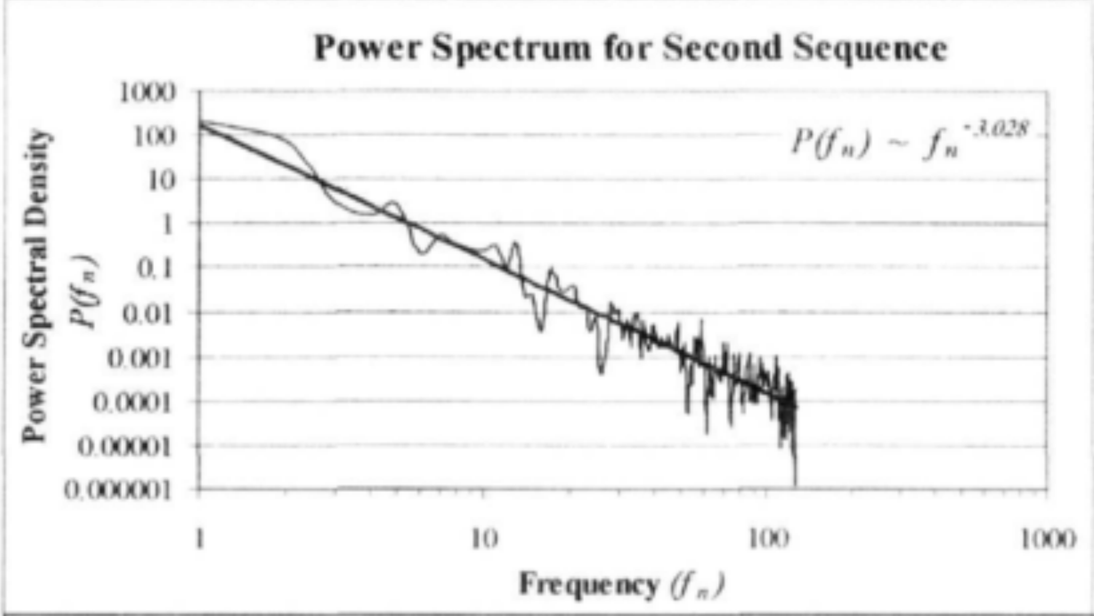


**Figure E.8 - Power spectrum for sequence plotted in Figure E.5**

All of the properties and methods presented here for the one-dimensional case, readily extend into two and three space, however it becomes increasingly difficult to visualise the shape of the power spectrum, filter function and random field as the dimension of the problem increases.

## E.3 TWO DIMENSIONAL POWER-LAW FILTERING

Following the format adopted for the one dimensional case in Section E.2, consider two two-dimensional fields generated using the same technique presented for the one-dimensional case, only using the two-dimensional Fast Fourier Transform on a two-dimensional field. With reference to Figure E.9, the field on the left was generated using a $\beta_{space}$ exponent of 2.0 in both x and y directions and the field on the right was generated using a $\beta_{space}$ exponent of 3.0 in both directions. The two fields were generated using the same set of white noise and have the same marginal distribution.

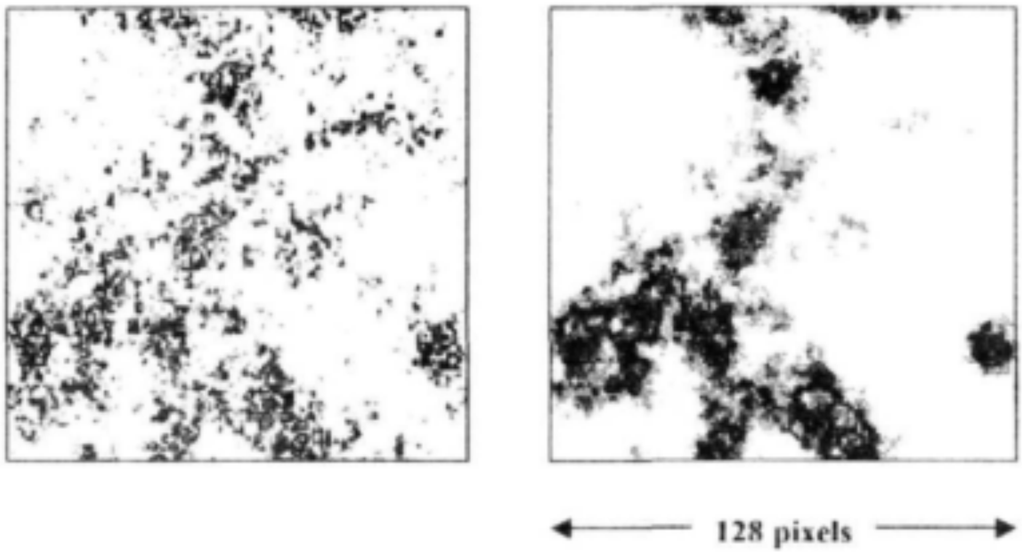


**Figure E.9 - Comparison of two two-dimensional power-law filtered fields**

As was observed in the one-dimensional case, the higher $\beta$ exponent in the filter produces a smoother image with a lower maximum intensity. The Fourier wrapping is evident in both the x and y directions of both images. The radially averaged two-dimensional power spectra of these two fields are plotted in Figures E.10 and E.11 and least squares approximations of their gradients have been calculated. The gradients of 2.049 and 2.939 are very close to the intended values of 2.000 and 3.000. The difference between the calculated values and the intended values are again due to the fact that the original noise process is imperfect and in addition, the data used to compute the power spectra are sampled from within the limits of the $^3/_4$ doughnut mask which was discussed in Section 5. The errors in estimation of the power spectrum due to the use of the mask and integer precision data are discussed in detail in Section 12.

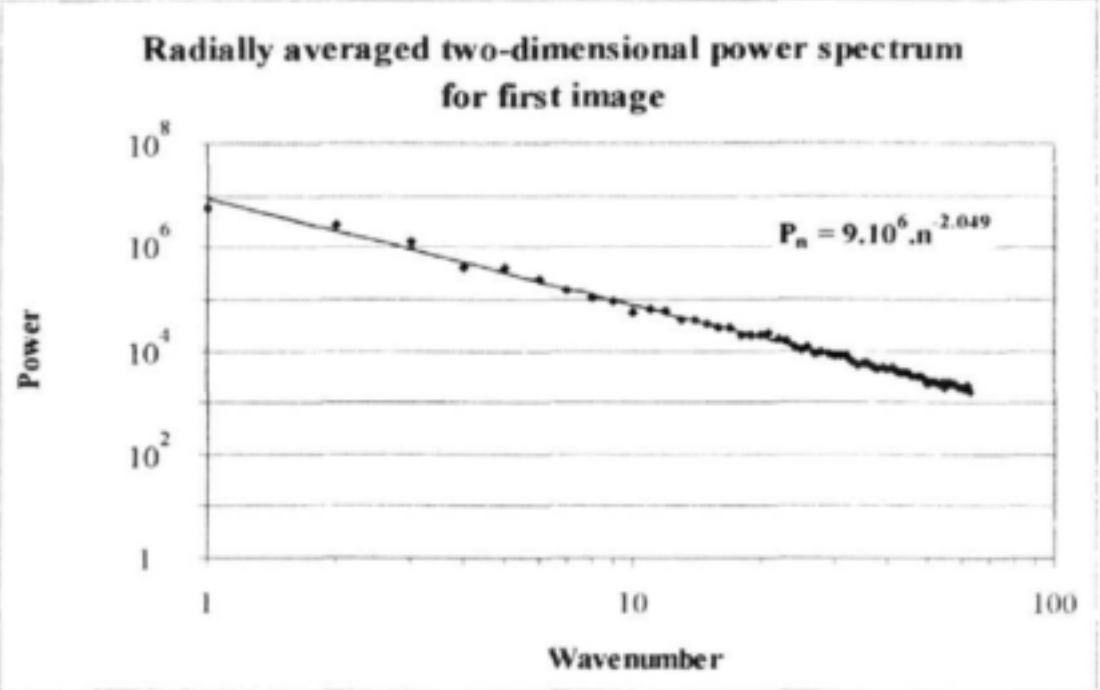


Figure E.10 - Averaged power spectrum for left hand image of Figure E.8

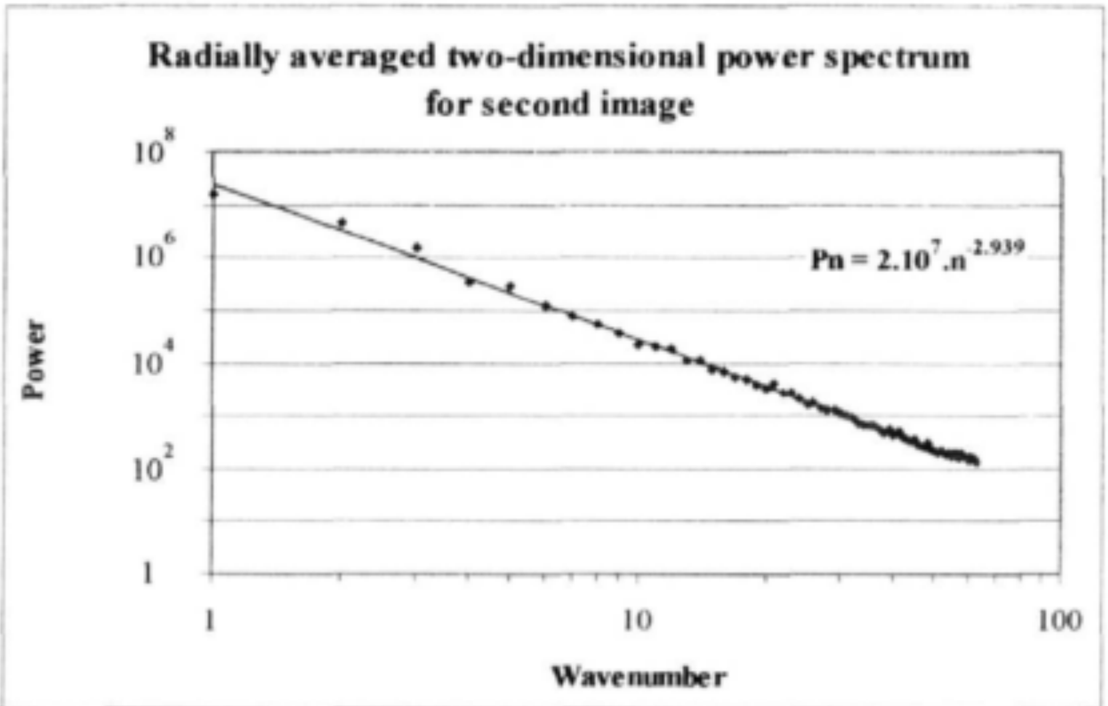


Figure E.11 -Averaged power spectrum for right hand image of Figure E.8

The two spectra shown in Figures E.10 and E.11 are plotted up to a wavenumber of 64 since the original images were a size of 128 x 128 pixels. Due to the averaging processes about the Nyquist frequency *and* about the apex, the power spectra are of a much smoother nature than those seen in the one dimensional cases of Section E.1, however, the basic properties observed in the one-dimensional case are preserved in the two-dimensional case and beyond. That is, the simulated fields are scaling, non-stationary random fields, with a lognormal distribution $\Lambda(\mu, \sigma)$ and a correlation structure defined by a single parameter $\beta$.

# Appendix F - σ, β, μ and WAR plots for February 1996

The Graphs which are presented in this appendix are all plotted to the same scale so that they can be easily compared. The software designed to analyse these data first tests to see if there is a Wetted Area Ratio in excess of 1% before attempting to fit a lognormal distribution and extract the power spectrum. Only if the WAR is found to be greater than 1% will the CAPPI be analysed for the three parameters σ, β and μ. This condition was deliberately built into the analysis software to avoid wasting unnecessary time searching for a maximum likelihood solution on a dry CAPPI. The function minimisation routine converges to a solution considerably quicker for a wet CAPPI than for a dry CAPPI. For WAR values under 1%, the routine can take several minutes to find a solution for a single CAPPI, whereas for CAPPIs with a WAR in excess of 10% it will usually converge to a solution within a second. The WAR is plotted for all 3000 CAPPIs in Figures F.1 to F.6 and σ, β and μ are plotted only for the 2500 CAPPIs with a WAR above 1%.

Additional plots in this appendix are the relationships between μ and σ parameters (Figure F.7) and between the β and σ parameters (Figure F.8). Clearly the μ versus σ, exhibits a relationship similar to that presented in Figure 14, and the regression line which is plotted on this figure is that used in the String of Beads model when calculating the σ from a given μ.

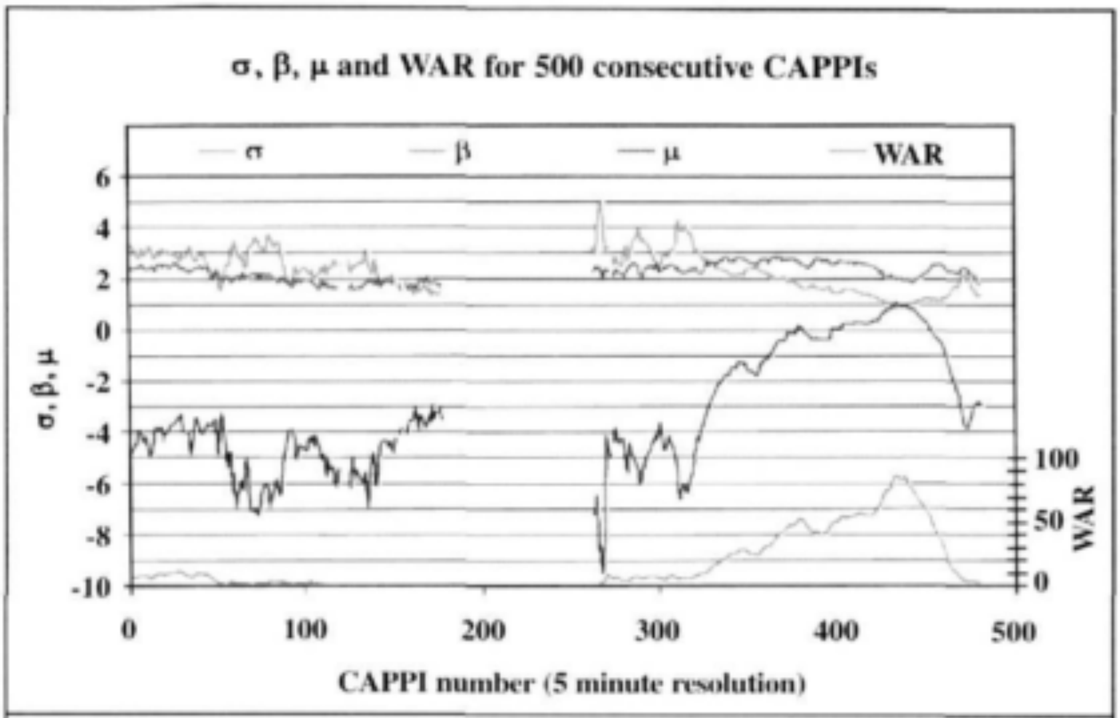


Figure F.1 - Analysis of CAPPIs 0 - 500 of February 1996

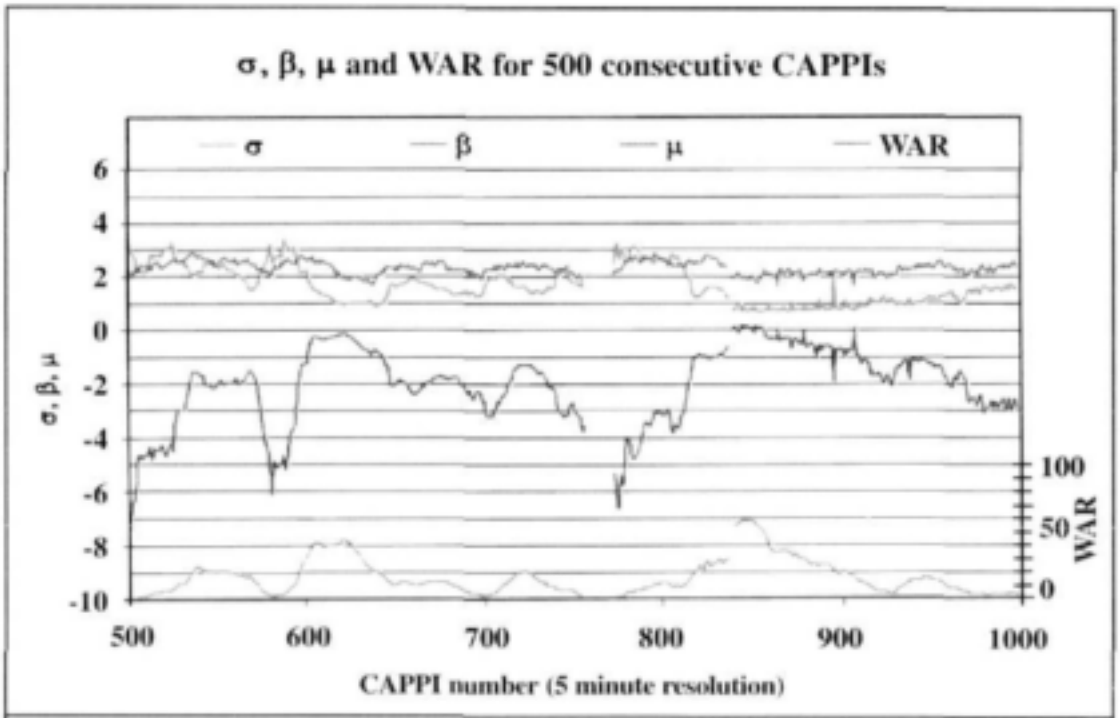


Figure F.2 - Analysis of CAPPIs 500 - 1000 of February 1996

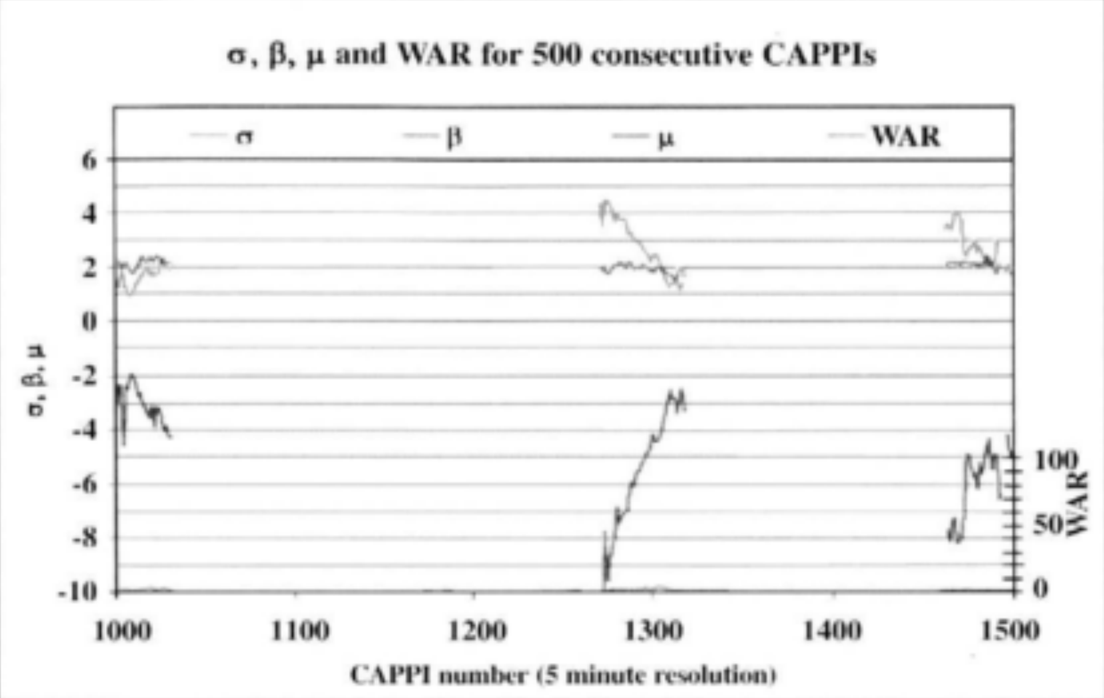


Figure F.3 - Analysis of CAPPIs 1000 - 1500 of February 1996

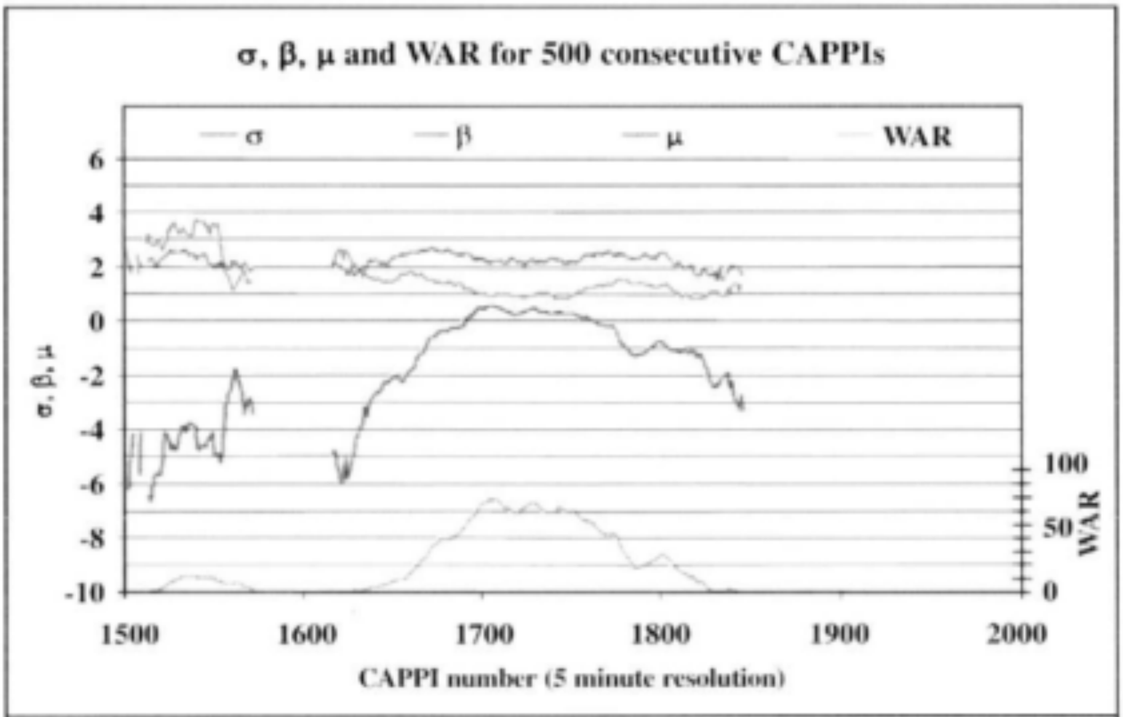


Figure F.4 - Analysis of CAPPIs 1500 - 2000 of February 1996

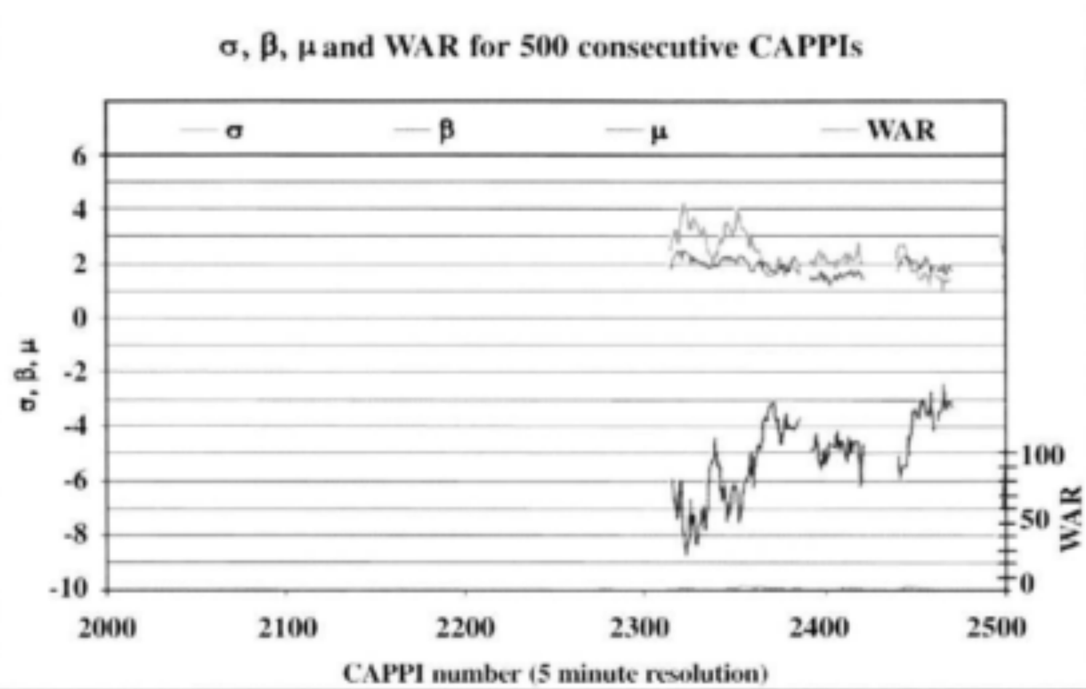


Figure F.5 - Analysis of CAPPIs 2000 - 2500 of February 1996

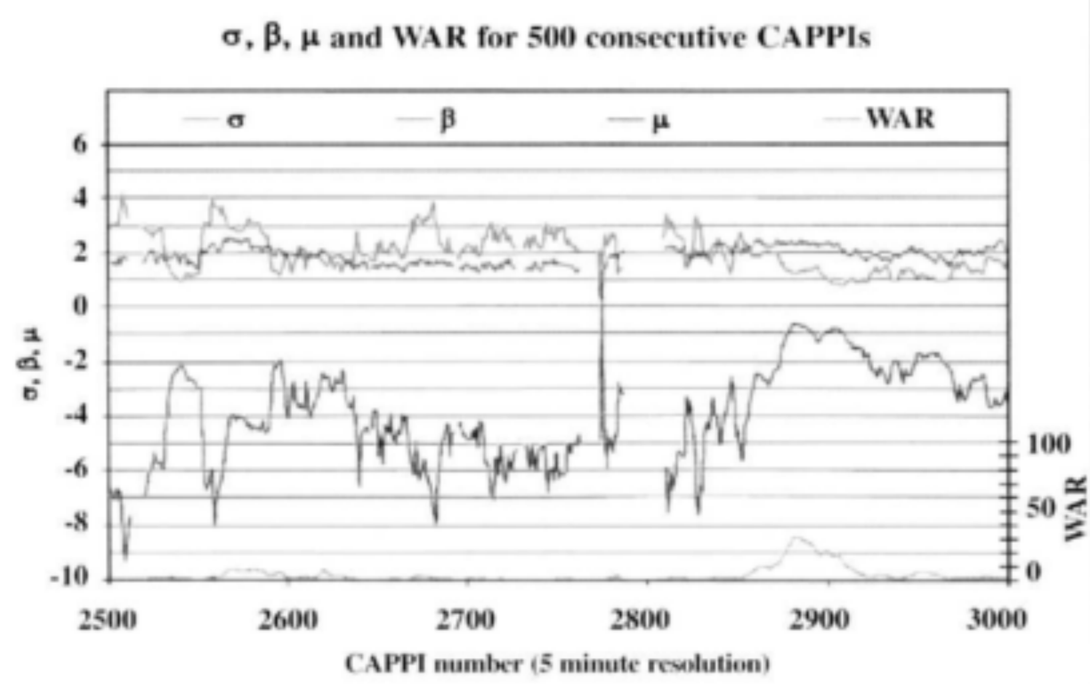


Figure F.6 - Analysis of CAPPIs 2500 - 3000 of February 1996

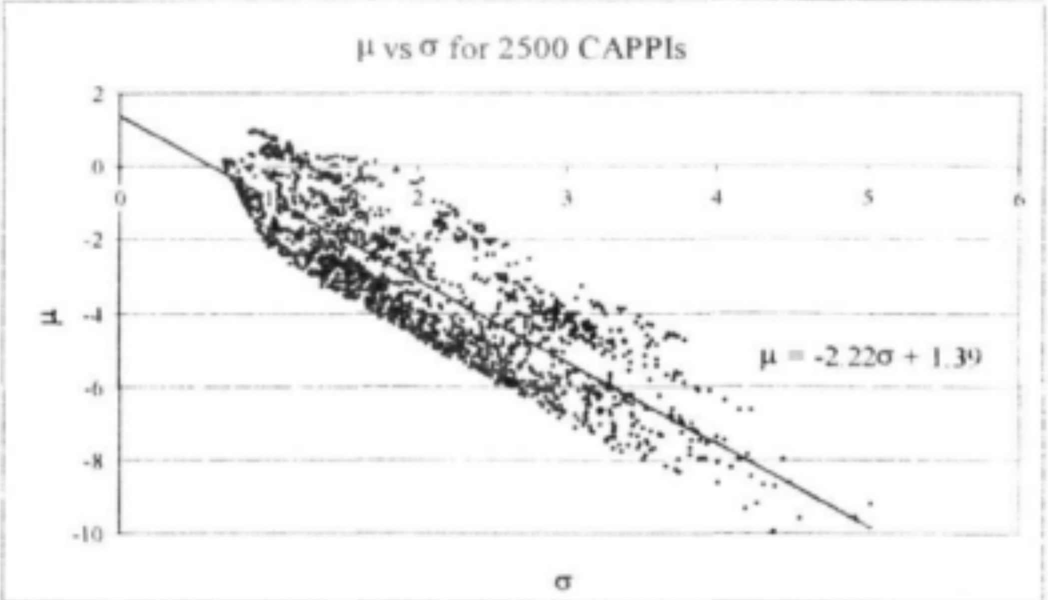


**Figure F.7 - μ versus σ plot for 2500 wet CAPPIs of February 1996**

The final plot (Figure F.8) in this appendix for the 2500 wet CAPPIs illustrates that there is no significant relationship between the β and σ statistics and consequently no relationship between the $\beta$ and $\mu$ statistics of a CAPPI. This shows that the clustering behaviour of the pixels as defined by $\beta$ is independent of the marginal distribution defined by $\mu$ and $\sigma$.

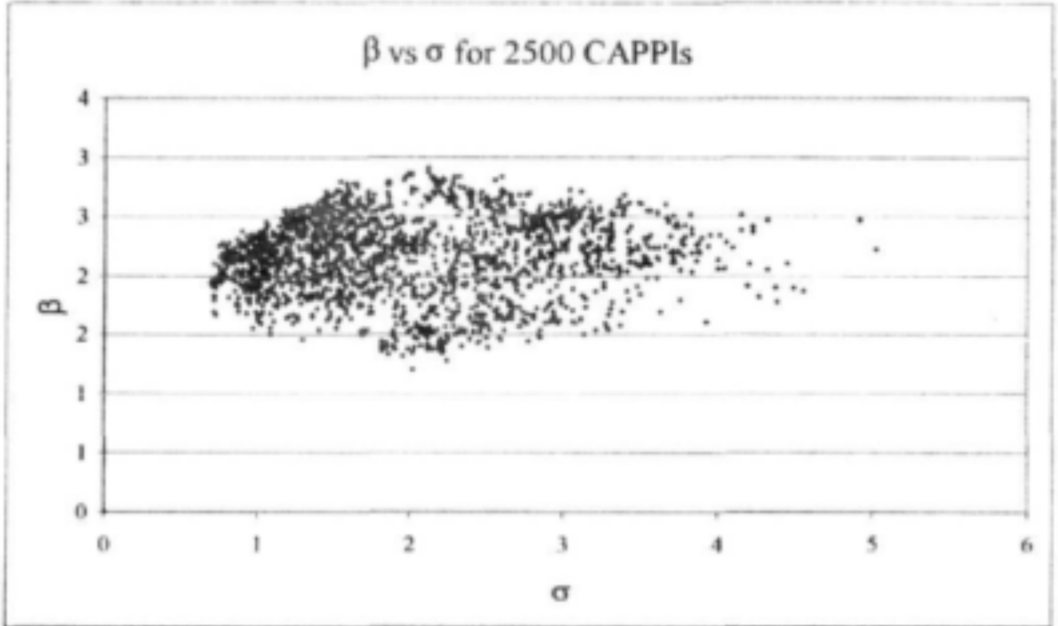


**Figure F.8 - β versus σ plot for 2500 wet CAPPIs of February 1996 illustrating that there is no obvious relationship between the marginal distribution and the clustering behaviour**

# Appendix G - Instructions for the use of the Compact Disk

The Compact Disk which accompanies this report is available on request at cost of reproduction (Approximately R20) from the Water Research Commission. It requires a Windows 95 or Windows NT (version 4.0 or later) operating system for viewing on a personal computer. The user should run the program using the built in *Windows Explorer*. There are several directories on the disk, each of which contains a brief text file explaining its contents.

## G.1 ANIMATED RAINFALL SEQUENCES

The only directory referred to in the text of the report is that containing the animated sequences of CAPPI images and has been named ***Animation***. This directory contains an executable file entitled *Setup.exe*. To execute the file, **double-click on it** with the mouse. In three interactive steps, the program will set up a MS PowerPoint presentation in the temporary directory of the computer hard drive. First, the program will prompt the user to select a file in which to extract the presentation. The directory **C:\Temp** is suggested as the most suitable. The program will take about 1 minute to install the presentation into the selected directory and once it has completed the installation, it will display a warning that the presentation will not run on Windows 3.1 - **select the OK button**. Finally, the program will prompt the user to choose whether or not to run the presentation immediately - **select the Yes button**. The presentation will begin with a brief slide instructing the user of three basic navigation commands, using the left and right mouse button.

The presentation consists of 6 brief slides, two of which contain animated images. The animation is started by moving the mouse pointer over the image, at which point it will convert to a Hypertext pointer (the Hand), and selecting the image with a **single click of the left mouse button**. In the case where there are two animated images on one slide (Slide number 4) both animated images can be run at the same time by starting one image and then immediately starting the second. This is a useful way to compare the real sequence to the simulated one. The third animation is of the entire sequence of CAPPI data for February 1996. This image runs for 8 minutes 30 seconds and contains a lot of dry time. There is no need to view the *entire* sequence, it is included for completeness of exposition.

## G.2 EXTRA INFORMATION

The compact disk contains five extra directories of information which contain files which may be of interest to the reader. The contents of each are briefly described to assist in the navigation of the Compact Disk.

**C Source Code** contains the source code which can be compiled and run on the Microsoft Visual C++ compiler. The program entitled *Analyser* is used to analyse a sequence of CAPPIs for their σ, β and μ and their WAR. The program entitled *Cropper* reads in a 200x200 pixel CAPPI, crops the inner 128x128 pixels and applies the 3/4 doughnut mask to the CAPPI. *Cubicbead* generates a single rainfall event of any required *average* σ, β and μ. *ExtractIMA* program used to extract the bitmaps from the original data, and an illustration of the binary file structure of the compressed *.ima format is included in this file. Finally, *StrFunction* reads a single 128x128 pixel CAPPI and extracts the data needed to plot the Generalised Structure Function.

**C++ Source Code** also contains source code which can be compiled and run on the Microsoft Visual C++ compiler. *SBM* is the simple String of Beads model configured to simulate the rainfall for a 28 day February in Bethlehem.

**CAPPIs** contains the data that was analysed in this study. There are several sub directories to this directory, the first of which contains the original 200x200 CAPPI data. The remaining subdirectories contain CAPPIs in a masked bitmap format. The data is divided into 4 sub-events and an *.avi video clip is included for each. This can be viewed in most modern computers by double-clicking on it. The sequences are the same as that presented in the MS PowerPoint Presentation.

**SB_Presentation** contains a PowerPoint presentation of a joint paper by GGS Pegram and AN Clothier. It was first presented at the 4th International Symposium on Hydrologic Applications of Weather Radar in San Diego, California, USA 5th - 9th April 1998 and again at the fall AGU meeting in San Francisco, December 1998. This presentation can be viewed by following the instructions given in Section G.1.

# References

---

**Aitchison, J. and JAC Brown.** (1957). *The lognormal distribution with special reference to its uses in economics*, University of Cambridge, Department of Applied Economics, Monograph: 5, Cambridge University Press.

**Anselmet F., Y Gagne, EJ Hopfinger, and R Antonia.** (1984). High-Order Velocity Structure Functions in Turbulent Shear Flows. *J. Fluid Mech.*, Vol. 140, p63.

**Austin, GL.** (1998). History of Radar and Radar Meteorology, *Proceedings of Advanced Study Course on Radar Hydrology for Real Time Flood Forecasting*, University of Bristol, 24th June to 3rd of July.

**Battan, LJ.** (1973). *Radar Observation of the Atmosphere*. University of Chicago Press, Chicago, USA.

**Beasley, JD. and SG Springer.** (1985). Algorithm AS111 - The Percentage Points of the Normal Distribution, *Applied Statistics Algorithms*, P Griffiths and I D Hill (eds). Royal Statistical Society.

**Bell, TL.** (1987). A Space-Time Stochastic Model of Rainfall for Satellite Remote Sensing Studies, *J. Geophys. Res.*, Vol. 92, No. D8, p 9631.

**Benzi, R., L Biferale, A Crisanti, G Paladin, M Vergassola and A Vulpiani.** (1993). A Random Process for the Construction of a Multiaffine Field, *Physica* D, Vol. 65, p353.

**Bras, RL. and I Rodriguez-Iturbe.** (1985). *Random Functions and Hydrology*. Addison-Wesley.

**Brenier, P.** (1990). *Simulations Dynamique Multifractale Des Nuages*. MS Thesis. Ecole Normale de Science et Technologie, Paris, France.

**Court, AP.** (1979). Rainfall Characteristics of Classification Systems used by the BEWMEX project. *Bethlehem Weather Modification Experiment*. Progress report no 14. Weather Bureau, Pretoria, South Africa.

# References...

**Cowpertwait, PSP.** (1994). A Generalised Point Process Model for Rainfall. *Proc. Royal Soc. London*, A, Vol. 447, pp 23-37.

**Crane, RK.** (1990). Space-Time Structure of Rain Rate Fields. *J. Geophys. Res.*, Vol. 95, No. D3, pp 2011-2020.

**Dixon, M. and G Wiener.** (1993). TITAN: Thunderstorm Identification, Tracking, Analysis and Nowcasting - a Radar Based Methodology. *J. Atmos. Ocean. Tech.*, Vol. 10, No.6, pp785-797.

**Foufoula-Georgiou, E., and Krajewski, W.** (1995). Recent Advances in Rainfall Modelling, Estimation and Forecasting. *Reviews of Geophysics*, Supplement, pp1125-1137, July 1995.

**Griffiths, P. and ID Hill.** (1995). (eds). *Applied Statistics Algorithms*. Royal Statistical Society.

**Gupta, VK and EC Waymire.** (1987). On Taylor's Hypothesis and Dissipation in Rainfall. *J. Geophys. Res.*, Vol. 92 , No. D8, pp 9657-9660.

**Gupta, VK and EC Waymire.** (1990). A Statistical Analysis of Mesoscale Rainfall as a Random Cascade. *J. App. Meteorology*, Vol. 32, pp251-267.

**Jenkins, GM and DG Watts.** (1967). *Spectral analysis and its applications*. Chap 5, Holden-day, Oakland, Calif.

**Knuth, DE.** (1969). *The Art of Computer Programming. Vol. 2 / Seminumerical Algorithms*. pp 9-34, Addison-Wesley.

**Kuc, R.** (1988). *Introduction to Digital Signal Processing*. McGraw Hill, New York.

**Marshall, JS. and WM Palmer.** (1948). The Distribution of Raindrop Size. *J. Meteor.* Vol. 5, p165.

# References...

**Matheron, G.** (1973). The Intrinsic Random Functions and Their Applications. *Adv. Appl. Prob.* 5: pp439-468.

**Menabde, M., D Harris, A Seed and G Austin.** (1997). Self-similar Random Fields and Rainfall Simulation. *J. Geophys. Res.* Vol. 102. D12. p13509.

**Menabde, M., A Seed, D Harris and G Austin.** (1999). Multiaffine Random Field Model of Rainfall. *Water Resour. Res.* Vol. 35, No. 2, pp509-514.

**Mittermaier, MP. and DE Terblanche.** (1997). Converting Weather Radar Data into Cartesian Space: A New Approach Using DISPLACE Averaging. *Water SA*, Vol. 23, No. 1, pp46-50.

**Nelder, JA. and R Mead.** (1965). A Simplex Method for Function Minimisation. *Computer Journal*, Vol. 7, pp. 308-313.

**Over, TM. and VK Gupta.** (1994). Statistical Analysis of Mesoscale Rainfall: Dependence of a Random Cascade Generator on Large Scale Forcing. *J. Appl. Meteorology*. Vol. 33. pp1526-1542.

**Press, W., SA Teukolsky, WT Vetterling and BP Flannery.** (1992). *Numerical Recipes in C - The art of scientific computing*. 2nd ed. Cambridge University Press.

**Pegram, GGS.** (1980). An Auto-regressive Model For Multi-lag Markov Chains. *J. Applied Probability*. Vol. 17. pp350-362.

**Pegram, GGS., AW Seed and AN Clothier.** (1997). Temporal and Spatial Modelling of Rainfields Using Fractals. *Proceedings of the 8th South African National Hydrology Symposium* (17th - 19th November 1997), Pretoria, South Africa.

**Pegram, GGS. and AN Clothier.** (1998). Space-Time Modelling of Rainfall in Fine Intervals: The "String of Beads" model. *Proceedings of 4th Int. Symp. on Hydrological Applications of Weather Radar*, San Diego, CA, 5th to 9th April.

# References...

**Pegram, GGS. and MP Mittermaier.** (1998). Estimating Rainfall at Low Altitude at Distance, From Radar CAPPI Data. *Proceedings of 4th Int. Symp. on Hydrological Applications of Weather Radar*, San Diego, CA, 5th to 9th April.

**Pegram, GGS. and AW Seed.** (1998). *The Feasibility of Stochastically Modelling the Spatial and Temporal Distribution of Rainfields.* WRC Report No. 550/1/98. Report to the Water Research Commission, Pretoria, South Africa.

**Schertzer, D. and S Lovejoy.** (1987). Physical Modelling and Analysis of Rain and Clouds by Anisotropic Scaling Multiplicative Processes. *J. Geophys. Res.*, Vol. 92 , No D8, pp 9693-9714.

**Seed, AW.** (1992). *Generation of a Spatially Distributed Daily Rainfall Database for Various Weather Modification Scenarios.* WRC 373/1/92. Report to Water Research Comm., Pretoria, South Africa.

**Taylor, GI.** (1938). The Spectrum of Turbulence. *Proc. R. Soc. London, Ser.* A, Vol. 164, pp476-490.

**Terblanche, DE.** (1996). A simple Digital Signal Processing Method to Simulate Linear Quadratic Responses from a Radar's Logarithmic Receiver. *J. Atmos. and Oceanic Tech*, Vol. 13, No. 2, pp533-538

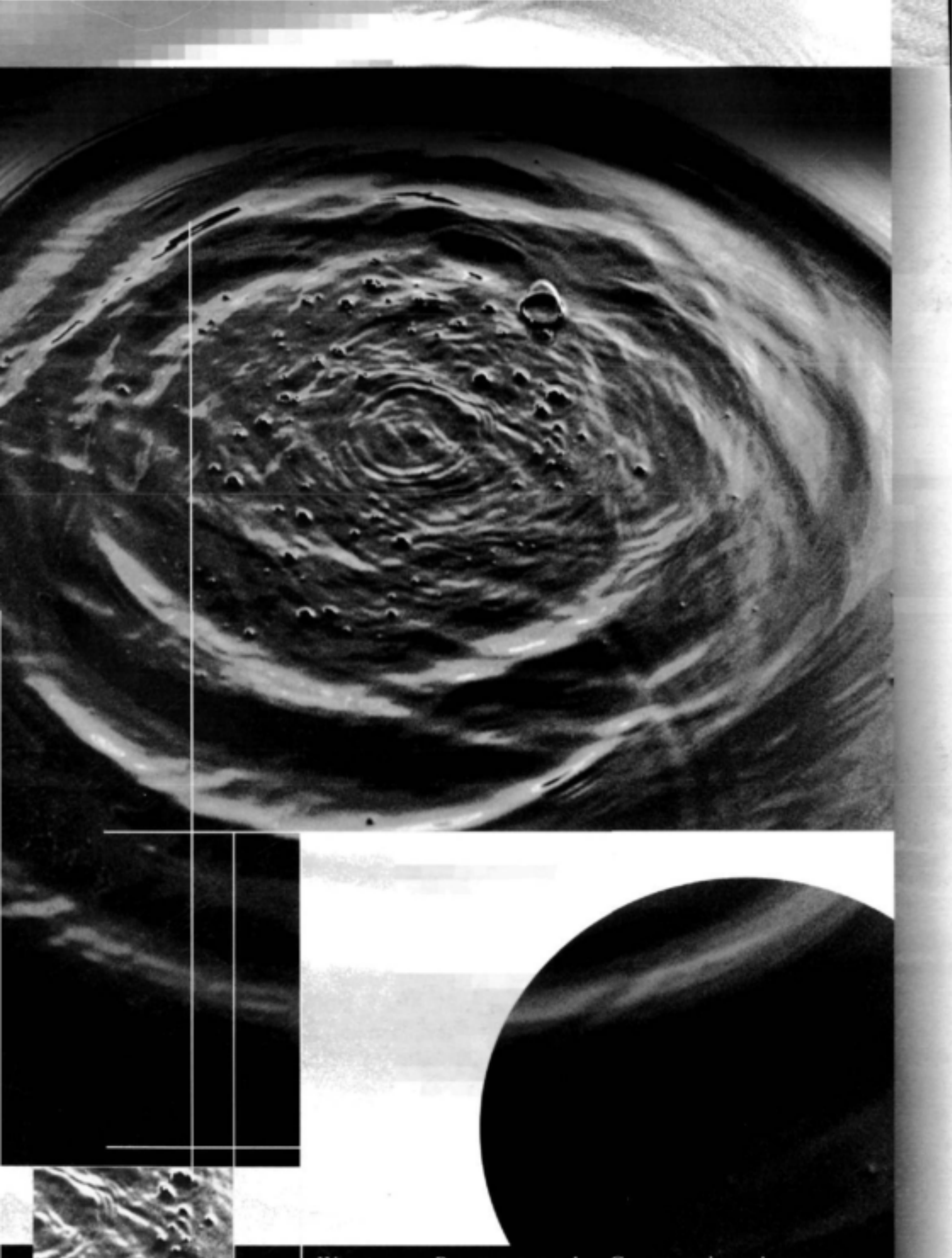
**Uijlenhoet, R.** (1998). Raindrop Size Distributions and the Z-R Relationship. *Proceedings of Advanced Study Course on Radar Hydrology for Real Time Flood Forecasting*, University of Bristol, England, 24th June to 3rd July.

**Wichmann, BA. and ID Hill.** (1982). A Pseudo-Random Number Generator. NPL Report DITC 6/82. Algorithm AS183 in *Applied Statistics Algorithms*, P Griffiths and I D Hill (eds). Royal Statistical Society.

**Wilson, J., D Schertzer and S Lovejoy.** (1991). Continuous Multiplicative Cascade Models of Rain and Clouds in *Non-Linear Variablility in Geophysics : Scaling and Fractals.* D Schertzer and S Lovejoy (eds). Kluwer Academic.

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

**Zucchini, WS. and PT Adamson.** (1984). *Assessing the risk of deficiencies in streamflow.* Report No. WRC 91/2/84. Report to the Water Research Commission, Pretoria, South Africa.



Water Research Commission

PO Box 824, Pretoria, 0001, South Africa

Tel: +27 12 330 0340, Fax: +27 12 331 2565

Web: http://www:wrc.org.za