Spatio-temporal Rainfall Estimation and Nowcasting for Flash Flood Forecasting

Scott Sinclair

In fulfilment of the requirements for the degree of Doctor of Philosophy in Engineering

School of Civil Engineering, Surveying and Construction Management

University of KwaZulu-Natal

March 31, 2007

Everything should be made as simple as possible, but not simpler.

–Albert Einstein

PREFACE

This Ph.D. thesis has resulted from work which took place over a period of around four and a half years. During this time I have been involved in several funded research projects where the general theme has been that of flood forecasting and spatial rainfall estimation.

Since this topic encompasses a broad range of tasks, techniques and algorithms, my time has been spent focussing on many different issues. Apart from reviewing and internalizing the extensive literature on the subjects presented herein (as well as finding ways to make my own contribution); I have had to gain expertise in computer programming using a number of languages (C, C++, Python, MATLAB) and related tools; code algorithms; develop and learn to use software for handling the large datasets associated with Radar and Satellite; assist other people in the group with manipulating these data; source, procure and install a Meteosat receiving station and data server; become proficient with both Linux and Windows operating systems and understand the organizational set up and relationships between the role players in the Flood Forecasting field within South Africa.

As a result this thesis does not attempt to document in fine detail everything that I have read, seen or worked on during this period, but rather to bring together a selection of these things that is aimed at providing solutions to real problems.

Scott Sinclair – December 2006

ACKNOWLEDGEMENTS

In producing this Thesis I'd like to acknowledge the support of my supervisor, Professor Geoff Pegram. Without his enthusiasm and knowledge of the subject matter, the opportunities for much of this research would not have arisen.

I also acknowledge the support of a number of organizations. The South African Water Research Commission for funding several projects related to improving rainfall estimation and flood forecasting, which I have been a part of. The South African Weather Service, in particular the team at Bethlehem for freely sharing data, software, enthusiasm and scientific input.

My fiance Angela for being quietly accepting of my frequent absences during the write-up period. Thank you.

Of course my family have been supportive throughout. Thanks Mom, Dad and Kirsty.

Finally to my friends and all of the people who keep life interesting by providing good memories of climbing trips and reminding me that sometimes it's just about having fun.

ABSTRACT

Floods cannot be prevented, but their devastating effects can be minimized if advance warning of the event is available. The South African Disaster Management Act (Act 57 of 2002) advocates a paradigm shift from the current "bucket and blanket brigade" response-based mindset to one where disaster prevention or mitigation are the preferred options. It is in the context of mitigating the effects of floods that the development and implementation of a reliable flood forecasting system has major significance. In the case of flash floods, a few hours lead time can afford disaster managers the opportunity to take steps which may significantly reduce loss of life and damage to property.

The engineering challenges in developing and implementing such a system are numerous. In this thesis, the design and implementation of a flash flood forecasting system in South Africa is critically examined. The technical aspects relating to spatio-temporal rainfall estimation and nowcasting are a key area in which new contributions are made. In particular, field and optical flow advection algorithms are adapted and refined to help predict future paths of storms; fast and pragmatic algorithms for combining rain gauge and remote sensing (radar and satellite) estimates are refined and validated; a two-dimensional adaptation of Empirical Mode Decomposition is devised to extract the temporally persistent structure embedded in rainfall fields. A second area of significant contribution relates to real-time forecast updates, made in response to the most recent observed information. A number of techniques embedded in the rich Kalman and adaptive filtering literature are adopted for this purpose.

The work captures the current "state of play" in the South African context and hopes to provide a blueprint for future development of an essential tool for disaster management. There are a number of natural spin-offs from this work for related fields in water resources management.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

API – Application Programming Interface
AR – AutoRegressive time series model
ARIMA – AutoRegressive Integrated Moving Average time series model
ARMA – AutoRegressive Moving Average time series model
ATS – Adaptive Time Series
CAPPI – Constant Altitude Plan Position Indicator
DISPLACE – algorithm
DM – Disaster Manager
DVB – Digital Video Broadcast
DWAF – Department of Water Affairs and Forestry
EMD – Empirical Mode Decompostion
EOF – Empirical Orthogonal Function
EUMETSAT – EUropean organization for the exploitation of METeorological SATellites
FFF – Flash Flood Forecasting
FFT – Fast Fourier Transform
GCM – Global Circulation Model
GRAHI – Grup de Recerca Aplicada en HIdrometeorologica
HTML – HyperText Markup Language
IMF – Intrinsic Mode Function
IMS – Intrinsic Mode Surface
IR – InfraRed channel
LAM – Limited Area Model
MA – Moving Average time series model
MAD – Mean Absolute Deviation
MDV – Meteorological Data Volume
MET-8 – Meteosat-8
MISO – Multiple Input Single Output
MISP – Mutually Interactive State and Parameter estimation algorithm
MSRR – MultiSpectral Rain Rate algorithm
MSVC – MicroSoft Visual C/C++ compiler
NWP – Numerical Weather Prediction

QPF – Quantitative Precipitation Forecast
RMSSE – Root Mean Sum of Squared Errors
S-PROG – Spectral PROGnosis model
SAR – Synthetic Aperture Radar
SAWS – South African Weather Service
SBM – "String of Beads" Model
SEVERI – Spinning Enhanced VisiblE and infraRed Imager
SIMAR – Spatial Interpolation and MApping of Rainfall
SISO – Single Input Single Output
SMF – Spatial Mean Flux
SVD – Singular Value Decomposition
TITAN – Thunderstorm Identification, Tracking, Analysis and Nowcasting software system
UH – Unit Hydrograph
UKZN – University of KwaZulu-Natal
UPS – Uninteruptable Power Supply
VIS – VISible channel
WAR – Wetted Area Ratio
WV – Water Vapour channel

CHAPTER 1

INTRODUCTION

There is a paucity of work in flood forecasting in South Africa and most concerning is the fact that functioning systems in operational use are virtually non-existent. Government has declared a "strategic intent" Buys (2005), to deliver on the contents of the Disaster Management Act (2002). In this Chapter, a background is given and the contributions made within the thesis are outlined.

1.1 Background

Severe floods cannot be prevented. Through appropriate planning and management strategies, the devastating effects of floods can be reduced. Loss of life and damage to infrastructure can be minimized but never completely eliminated. The white paper on disaster management (WPDM, 1998) presents the data in table 1.1, showing the estimated losses resulting from several disasters in South Africa. Of particular significance is the prominence of flood events amongst these disasters.

With large increases in population and increasing urbanization (largely driven by poverty) there are more people living in informal settlements (near cities). These settlements often encroach on flood plains as this is the only undeveloped land which remains available. The people living in these settlements are those who are most at risk, not only due to their geographical location in the floodplain but also because (without access to insurance policies) they do not have the financial resources to recover from damage caused by flooding. In addition many formal developments already exist (and continue to be developed) in areas that are susceptible to flooding. When the nature of the development is such that significant

Place	Disaster	Cost
Ladysmith	Floods, 1994	400 families evacuated, R50 million damages
Merriespruit	Slimes dam, 1994	17 lives lost, R45 million damages
Pietermaritzburg	Floods, 1995	173 lives lost, Emergency shelter needed for 5 500
Ladysmith	Floods, 1996	Damages to infrastructure: R25 million
South Africa	Drought, 1991-92	49 000 agricultural jobs lost, 20 000 non-agricultural jobs lost, Associated with 27% decline in agricultural gross domestic product
Northern Province	Floods, 1996	R105 million damages
Mpumalanga	Floods, 1996	R500 million damages

Tab. 1.1: Estimated damages for several floods and a drought in South Africa. [Adapted from WPDM, 1998]

damage would result from flooding, early warning systems provide a possibility of reducing the risk to acceptable levels.

The recently promulgated Disaster Management Act (Act 57 of 2002) advocates a paradigm shift from the current "bucket and blanket brigade" response-based mindset to one where disaster prevention or mitigation are preferred. It is in the context of mitigating the effects of flood events that the development and implementation of a reliable flood forecasting system has major significance. In the case of flash floods even a small amount of lead time, of the order of a few hours, can allow disaster managers to take steps which may significantly reduce loss of life and damage to property.

The main requirement of an effective flood forecasting system is the provision of reliable, intelligible forecasts of flood flows with a reasonable lead-time and explicit error bounds. The forecasts must be made available at frequent intervals to hydrological operators, decision makers and disaster managers, in a clearly understandable form. Unfortunately, South Africa is a long way off from realizing such a situation in practice.

Important key phrases for on-line operation are: long lead time, frequent up-

dates, estimates including an indication of uncertainty, confirmatory information from the catchment, redundancy and feedback to improve estimates as new information becomes available.

For off-line work, the key phrases are: evaluation of the system and instrumentation of catchments, improvement of models, commitment to recalibration, ability to simulate alternative scenarios, operator training and the development of a reliable intelligible alerting system for disaster managers.

In a typical flooding scenario, the alerting system should trigger the forecasting system to move from standby mode to yellow alert, upon which, the information stream goes on-line and the models are calibrated at operational, rather than at standby frequency. Once there is a perceived threat of flood damage as a result of no intervention, the system goes to orange alert and the flood mitigation strategy is employed. It is at this stage that the information stream arriving in the operations room is put to best use by disaster managers.

It is well known that in times of flood, things to do not work well: telemetering equipment gets damaged by water or lightning, phone lines go dead, radio links become erratic, computer links go down and people do unpredictable things. The important key to counteract these phenomena is redundancy in the information stream and well trained disaster management personnel. The system needs to have parallel measuring devices available, along with alternative communication links.

The tools available can be loosely divided into hardware, software, algorithms and the means of communication to the stakeholders. The appropriate hardware available for use in South Africa includes the following: meteorological satellites, weather radar, telemetering rain gauges, telemetering stream gauges, fast computers with large storage, telephones, cell phone networks, Internet, solar panels/batteries and UPS systems.

The software useful for flood forecasting, which is currently available and does not require a large amount of modification for incorporation into a system, includes: meteorological prediction of rainfall - location and probability, satellite estimation of rainfall, radar estimation of rainfall, Hydrologic catchment models

and Hydraulic river channel models.

There is a large number of conceptual or physically-based catchment models in use throughout the world. A few of them are fully distributed, the remainder are either lumped models or semi-distributed models; the latter are effectively aggregations of the lumped models.

In this context, lumping refers primarily to the spatial dimension, the vertical being compartmentalized into a varying number of storage elements, incorporating linear or non-linear models of the water transfer processes between the elements. A typical lumped model may have a number of adjustable parameters, which need calibration.

Physically based catchment models, which may mimic the physical processes on a small plot, need calibration when used in lumped form to model an area larger than a few square metres. The non-linear processes which are likely to vary a great deal in space and time, are often assumed to maintain their mathematical relationships when aggregated over tens or hundreds of square kilometres. Not surprisingly the numbers which come out of the calibration process (forcing the model to fit the calibration data set) in some cases bear no resemblance to their physical values in small areas.

There are other approaches to rainfall-runoff modelling. The first, discussed here, is the "difference equation" approach borrowed from time series analysis. These models have difficulty in coping with losses due to infiltration and evapotranspiration in the relatively dry periods but may be quite reasonable event models during very wet times where losses are lower and the rainfall to runoff conversion process is close to linear. Methods of estimating the parameters of these linear models include Least Squares in the time domain and spectral techniques using Fourier and Laplace transforms in the frequency domain.

Another type of model is related to the earliest known form of rainfall/runoff model used in Hydrology, namely the Unit Hydrograph (UH) and its successors. If the UH is approximated by a non-negative transfer function (like the impulse response function of a linear reservoir), modified by incorporating a loss function, it can conveniently be fitted to rainfall-runoff data. Its advantage is the small num-

ber of parameters defining the UH and the maintenance of positive flows due to the functional form of the impulse response function. It is no wonder that engineering hydrologists have for many years trustingly used the Unit Hydrograph in various forms to estimate extreme floods or for relatively impervious catchments.

There are two approaches to producing model forecasts in real time. The first school chooses to update the model (state variables, input or parameters) on-line. The second school produces the forecasts using a model calibrated on historical data, without updating the model parameters. At first sight the parameter updating approach appears to make better sense; however, in some cases this may result in a large amount of computation to be done on-line, adjusting the models parameters to give the best fit to the most recently acquired (often noisy) data. The model updating tasks are often performed by filtering techniques (e.g Kalman filter), which are efficient and have the additional benefit of providing the means to give explicit error estimates for the forecasts.

1.2 A proposed framework for flash flood forecasting in South Africa

In this thesis an attempt is made to provide a blueprint for flash flood forecasting systems in South Africa. The focus is not only on the science required to overcome the various difficulties which flood forecasting presents, but also to provide a generic framework which can be adapted and updated in response to specific needs, as skills in this area are developed. Figure 1.1 shows a schematic overview of the key components required by a flood forecasting system, the remaining chapters in this thesis tackle each of these components in greater depth.

Arguably the most important input to any flash flood forecasting system is precipitation. In most of Southern Africa the influence of snow can safely be ignored and the measurement of rainfall becomes the most important factor in determining the system input. As described in detail in chapter 3, there are three measurement devices in South Africa which provide estimates of rainfall at suitable spatial and temporal resolution for flash flood forecasting. The rainfall estimates produced

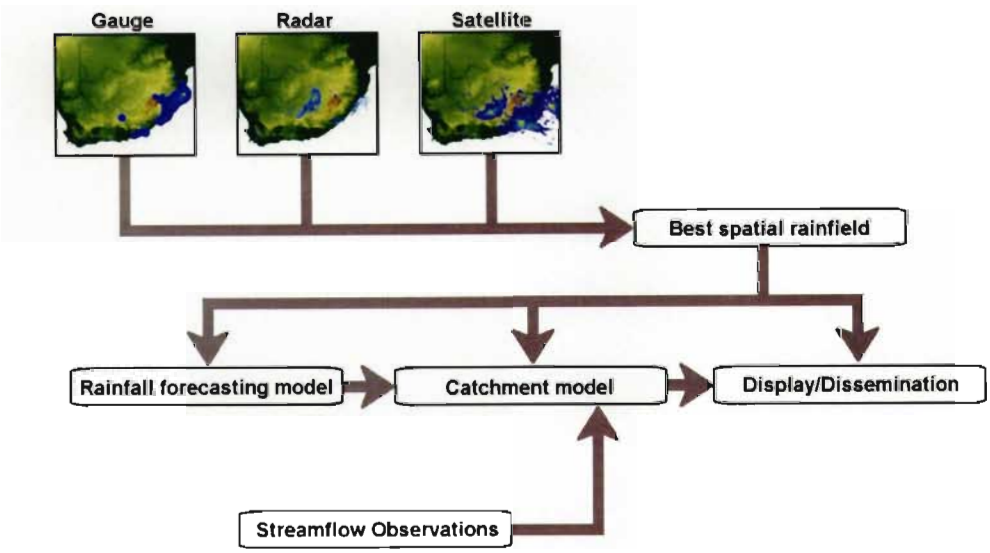


Fig. 1.1: A schematic overview of the main components required for a successful flood forecasting system.

from rain gauges, weather radar and meteorological satellite may be combined in an optimal way to produce the best spatial estimate of rainfall for the region and, in particular, the catchment of interest. The resulting combined rainfall estimate can be fed into display systems for direct visualization of instantaneous and accumulated rainfall. The estimates are also used as input to catchment models and converted to streamflow, the core output of the flood forecasting system.

The nature of the catchment model is not dictated here as its implementation and effective use are dependent on the expertise and data which is available. In South Africa the necessary hydrological expertise does not exist in many of the municipal structures which are responsible (by law) for ensuring appropriate flood mitigation strategies, including flood forecasting systems. For useful real-time operation, the catchment models need to be informed by real-time streamflow observations which provide a means of updating the models performance and improving forecasts. The concepts related to model updating are discussed

in chapter 2. To improve the lead-time of streamflow forecasts using current information, rainfall forecasts can prove useful. Short term rainfall nowcasting is dealt with in chapter 4. Stochastic nowcasting models can provide the means for extending the information from current best spatial rainfall fields into the future.

1.3 Summary of publications arising from this study

Some of the work presented in this thesis has appeared in the publications listed in this section. In a few cases the text and figures appear here in an almost unaltered form, while in others, additions, refinements and improvements have been made. Copies of the peer-reviewed journal articles are included in the Appendix of this document.

Peer reviewed journal articles

Sinclair, S. and Pegram, G. (2005a). Combining radar and rain gauge estimates using conditional merging. *Atmospheric Science Letters*, 6:19–22.

Sinclair, S. and Pegram, G. (2005b). Empirical Mode Decomposition in 2-D space and time: a tool for space-time rainfall analysis and nowcasting. *Hydrology and Earth System Sciences*, 9:127–137.

Technical reports

Pegram, G., Sinclair, S., Parak, M., Sakulski, D., and Nxumalo, N. (2006a). National Flood Nowcasting System: Towards an integrated mitigation strategy. Technical Report 1429/1/06, Water Research Commission, Pretoria, South Africa.

Pegram, G., Sinclair, S., and Wesson, S. (2006b). Daily Rainfall Mapping over South Africa: Mapping. Technical Report 1425/1/06, Water Research Commission, Pretoria, South Africa.

Sinclair, S. and Pegram, G. (2004a). A Flood Nowcasting System for the eThekwini Metro. Volume 1: Umgeni Nowcasting using Radar - An integrated pilot study. Technical Report 1217/1/04, Water Research Commission, Pretoria, South Africa.

Conference proceedings and presentations

Mkwananzi, N., Pegram, G., and Sinclair, S. (2003). Modelling flood inundation in the Mlazi river under uncertainty. In *Proceedings of the 11th South African National Hydrology Symposium*, Port Elizabeth, South Africa.

Pegram, G., Seed, A., and Sinclair, D. (2002). Comparison of Methods of Short-term Rainfield Nowcasting. In *Proceedings of the 27th EGS general assembly*, Nice, France.

Pegram, G., Deyzel, I., Sinclair, S., Visser, P., Terblanche, D., and Green, G. (2004). Daily mapping of 24 hr rainfall at pixel scale over South Africa using satellite, radar and raingauge data. In *Proceedings of the 2nd IPWG Workshop*, Monterey, CA, USA.

Pegram, G. and Sinclair, S. (2004). National Flood Nowcasting System towards an integrated mitigation strategy in South Africa. In *Proceedings of the 6th International Symposium on Hydrological Applications of Weather Radar*, Melbourne, Australia.

Sinclair, S., Ehret, U., Bardossy, A., and Pegram, G. (2003). Comparison of Conditional and Bayesian Methods of Merging Radar and Rain gauge Estimates of Rainfields. In *EGS - AGU - EUG joint assembly*, Nice, France.

Sinclair, S. and Pegram, G. (2003b). The Design and Implementation of a Real-Time Flood Forecasting System in Durban, South Africa. In *EGS - AGU - EUG joint assembly*, Nice, France.

Sinclair, S. and Pegram, G. (2003a). Combining Traditional and remote sensing techniques as a tool for Hydrology, Agriculture and Water Resources Management. In *Proceedings of the 11th South African National Hydrology Symposium*, Port Elizabeth, South Africa.

Sinclair, S. and Pegram, G. (2004b). Combining radar and rain gauge rainfall estimates for flood forecasting in South Africa. In *Proceedings of the 6th International Symposium on Hydrological Applications of Weather Radar*, Melbourne, Australia.

Sinclair, S. and Pegram, G. (2005c). Empirical Mode Decomposition in 2-D space and time: A tool for space-time rainfall analysis and nowcasting. In *Proceedings of the 12th South African National Hydrology Symposium*, Pretoria, South Africa.

Sinclair, S. and Pegram, G. (2005d). Space-time rainfall analysis and nowcasting using Empirical Mode Decomposition in 2D. In *EGU general assembly*, Vienna, Austria.

1.4 Technical and scientific contributions

The main technical and scientific contributions in this thesis are outlined in this section. Pointers are given to the relevant sections within the document and further discussion of each contribution is presented in the concluding chapter (Section 6.1).

1. A two-dimensional extension of Empirical Mode Decomposition (EMD) is presented and applied to the analysis of spatial rainfall data. EMD analysis explicitly accepts non-stationarity in the data and does not rely on the predefined basis functions common in other (e.g. Fourier, Wavelet) methods. The work is a new contribution, published in Sinclair and Pegram (2005b) and discussed in detail in section 4.2 of the thesis.

2. A novel technique for the accumulation of spatial rainfall fields is presented. It uses advection vectors computed between successive scans to generate path integrals. The approach presented here refines those reported by Anagnostou and Krajewski (1999a) and Hannesen (2002). The accumulation scheme is presented in section 3.3.

3. The String of Beads Model (SBM) presented in Pegram and Clothier (1999) and Clothier and Pegram (2002), is adapted for short term nowcasting and compared with the Spectral PROGnosis (S-PROG) model (Seed, 2001). SBM in nowcasting mode is a new contribution, described in detail in section 4.1.2.

4. The conditional merging technique of Ehret (2002) has been adopted and validated using cross-validation techniques for synthetic as well as observed rainfall data. The technique has been compared with an implementation of the Bayesian merging technique (Todini, 2001) and found to perform competitively. The conditional merging technique has been extended to provide a more formal discussion of its error structure. The conditional merging algorithm is discussed in more detail in section 3.5.

5. Although adaptive time series forecasting is certainly not new (Haykin, 2001; Ljung, 1987), the application of these techniques is useful in the Hydrometrological context where traditional time series methods (Box and Jenkins, 1970) and filters (Kalman filters, Extended Kalman filter) are quite common (e.g. Seed, 2003; Szollosi-Nagy and Mekis, 1987; Todini 1978). Adaptive filtering techniques are used to forecast non-stationary time series data. The adaptive filtering techniques and model fitting procedures adopted in this study are presented in chapter 2.

6. Implementation of data transfer mechanisms. A prototype system was set up in the eThekwini disaster management centre (Durban, South Africa). This system is a first in South Africa and has laid the ground work for the eventual implementation of a more robust countrywide flood warning system. This work is expanded upon in chapter 5.

CHAPTER 2

TIME SERIES FORECASTING AND ADAPTIVE FILTERS

Time series forecasting refers to the prediction of future values of a time series at a specified lead-time, or range of lead-times. The predictions are typically made using a model which has outputs based on previous values of the data that we wish to predict and/or additional time series that are strongly correlated with the data we wish to forecast. The first section describes the general concepts and the well-known ARMA models for stationary time series (Box and Jenkins, 1970). The second section describes the ARIMA models of Box and Jenkins (1970) used for forecasting non-stationary time series. Section 2.3 provides a discussion of Adaptive time series models (e.g. Ljung, 1987; Haykin, 2001) and the Kalman filter (Kalman, 1960). Recursive models are particularly suited to online computations, as they do not require all of the data history to be retained. The strong link between Kalman filters and adaptive time series models is examined. Finally, some applications of the techniques described are presented which relate to the forecasting of streamflow and important image scale statistics measured from images of observed radar rainfall fields.

2.1 Stationary models

A time series is a sequence of the measured values of a process. A formal definition of stationarity is given by Box and Jenkins (1970, pp 26 ff.) and paraphrased here. Let z_t be the observed value of a process at time t and $\mathbf{z}_t = (z_t, z_{t+1}, \cdots, z_{t+n-1})$ be a sample of the process values at n equally spaced temporal intervals. Further, let $\mathbf{z}_{t+k} = (z_{t+k}, z_{t+k+1}, \cdots, z_{t+k+n-1})$ be a different set

of n sampled process values with times of observation which have all been shifted forward or backward by an integer amount k. In order for the process to satisfy the conditions of *strict stationarity*, the joint probability distribution of the n observations $f(z_t, z_{t+1}, \cdots, z_{t+n-1})$ must remain unchanged for any integer shift k,

$$f(z_t, z_{t+1}, \cdots, z_{t+n-1}) = f(z_{t+k}, z_{t+k+1}, \cdots, z_{t+k+n-1}). \qquad \text{for all } k \in \mathbb{I}$$

Weak stationarity of order m is satisfied if the moments of the distribution up to order m are independent of the sampling position. As a special case; for a Gaussian process to be strictly stationary, it is sufficient that the process have a constant mean and covariance structure, since the multivariate Normal distribution is completely characterized by its first and second order moments.

2.1.1 Autoregressive models

Autoregressive time series models use a linear combination of past values of the process to be modelled, as a means of predicting future values. Following the treatment of Box and Jenkins (1970), let y_t be the deviations of the measured process z_t from its mean. The process may then be modelled as an Autoregressive (AR) process of order p (where p is to be determined by appropriate model identification procedures). The AR process model has the following form

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + a_t$$

where the ϕ_j are autoregressive weights and a_t is a temporally uncorrelated, random error term. If a_t is Gaussian, then y_t is a linear combination of Gaussian variables and is therefore also Gaussian. It is thus conventional to standardize the sample $\mathbf{z}_i$ by its mean μ and standard deviation σ such that

$$y_t = \frac{z_t - \mu}{\sigma}$$

and $\mathbf{y}_i = (y_t, y_{t+1}, \cdots, y_{t+n-1})$ is Gaussian. In this case a_t has a zero mean and a variance to be estimated from past values of the process y_t and the model parameters. The terms μ and σ are the sample mean and standard deviation, estimated as

$$\mu = \frac{1}{n}\sum_{k=1}^{n} z_{t+k-1}$$

and

$$\sigma = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(z_{t+k-1} - \mu)^2}$$

The introduction of the backward shift operator $\mathbf{B}$, for which $\mathbf{B}y_t = y_{t-1}$ and $\mathbf{B}^i y_t = y_{t-i}$, allows the AR model to be written in a more condensed form as

$$\phi(\mathbf{B})y_t = a_t$$

where the AR operator, of order p is in polynomial form:

$$\phi(\mathbf{B}) = 1 - \phi_1\mathbf{B} - \phi_2\mathbf{B}^2 - \cdots - \phi_p\mathbf{B}^p$$

2.1.2 Moving average models

A moving average model defines the current value of a process as a linear combination of the values of the random process a_t. Thus a Moving Average (MA) model of order q is given by

$$y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

where the moving average weights θ_i are not constrained to be positive nor sum to unity, but do have restrictions related to the stationarity conditions of the process. As was the case for the AR model, we can define a moving average operator of order q (again based on the backward shift operator)

$$\theta(\mathbf{B}) = 1 - \theta_1\mathbf{B} - \theta_2\mathbf{B}^2 - \cdots - \theta_q\mathbf{B}^q$$

so that the MA model may be written in a condensed notation as

$$y_t = \theta(\mathrm{B})a_t$$

2.1.3 Autoregressive moving average models

The specification of a mixed Autoregressive Moving Average (ARMA) model is given as follows

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

This is an ARMA(p, q) process. It follows easily that this equation can also be expressed in terms of the AR and MA operators, defined previously, giving the condensed form for a general ARMA model as

$$\phi(\mathrm{B})y_t = \theta(\mathrm{B})a_t$$

2.2 Non-stationary (fixed parameter) models

For certain classes of non-stationary processes the ARIMA models and their counterparts can be used to model the time series. Autoregressive Integrated Moving Average (ARIMA) models (Box and Jenkins, 1970; pp 85) basically operate on differences of the original time series on the assumption that the process tends to exhibit locally stationary behaviour but that the process mean is non-stationary. The idea is to average (integrate) out the effect of the mean and produce a stationary process, which may then be efficiently modelled using the stationary models described in section 2.1.

Consider the process z_t described earlier. If the process z_t is non-stationary it is sometimes possible to define a stationary process

$$\begin{aligned} w_t &= \nabla^d z_t \\ &= z_t - z_{t-d} \end{aligned} \tag{2.1}$$

where d is an integer, so that w_t can now be modelled as a stationary ARMA process

$$\phi(\mathbf{B})w_t = \theta(\mathbf{B})a_t$$

Equation 2.1 can also be written as

$$z_t = \mathbf{S}^d w_t \tag{2.2}$$

where S is the summation operator. Equation 2.2 shows how the non-stationary process z_t can be obtained by summing the stationary process w_t, d times, hence the name of the model: Autoregressive Integrated Moving Average (ARIMA). Box and Jenkins (1970, pp. 90 ff.) show that the ARIMA model is useful for a number of non-stationary time series typically encountered in time series modelling applications.

Other alternatives include the ARMAX and ARIMAX models, where the X stands for eXogenous variables. This class of model is equivalent to the transfer function models described by Dooge (2003) and have been used often in the Hydrological literature (e.g. Szollosi-Nagy and Mekis, 1987).

2.3 Adaptive models and the Kalman filter

The models described in sections 2.1 and 2.2 are only applicable to time series which are stationary or can be transformed to stationarity using the methods described in section 2.2. The parameters describing these models are assumed to remain constant in time. Many naturally occurring time series do not conform to these notions of stationarity but may, nevertheless, be described by similar models if it is accepted that the parameters will vary in time. The solution is to re-estimate the parameters as new data becomes available.

This section begins by describing the Kalman filter (Kalman, 1960), and then moves through recursive least squares parameter estimation and finally ends with univariate and multivariate adaptive filters. There is a close relationship between the adaptive least squares algorithms and the Kalman filter (although the Kalman

filter is only relevant to stationary time series), these links will be emphasized where appropriate, while retaining a consistent notation throughout.

2.3.1 The Kalman filter

The Kalman filter (Kalman, 1960) is a recursive estimation scheme. It produces optimal estimates (in a linear, least squares sense) of a dynamic system's state vector $\mathbf{x}$. The system may be described by the following linear stochastic difference equation

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B}\mathbf{w}_k \tag{2.3}$$

at discrete times $k = 0, 1, 2, \cdots$, where $\mathbf{A}_k$ is the state transition matrix defining the transition of the states from one time-step to the next, $\mathbf{B}$ is a dimensional conversion matrix and $\mathbf{x}_k$ is the state vector at step k. The states are not directly observable but measurements of the system $\mathbf{y}$ are related to the states by the system measurement equation

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k. \qquad k = 0, 1, 2, \cdots \tag{2.4}$$

where $\mathbf{y}_k$ is the system measurement vector and $\mathbf{H}_k$ is the state to measurement transfer matrix.

The state noise $\mathbf{w}_k$ and the measurement noise $\mathbf{v}_k$ are assumed to be white noise described by their mean and covariance

$$E[\mathbf{w}_k] = \bar{\mathbf{w}};\ E[\mathbf{v}_k] = \bar{\mathbf{v}}, \qquad k = 0, 1, 2, \cdots$$
$$E[\mathbf{w}_k \mathbf{w}_j{}^T] = \mathbf{Q}_k \delta_{kj};\ E[\mathbf{v}_k \mathbf{v}_j{}^T] = \mathbf{R}_k \delta_{kj}$$

with δ_{kj} the Kronecker delta, ensuring that both $\mathbf{w}_k$ and $\mathbf{v}_k$ are serially uncorrelated. Although not strictly necessary (Sorenson, 1985, pp 13) it is also usual to require that

$$E[\mathbf{w}_k \mathbf{v}_j{}^T] = 0. \qquad \text{for all } j, k$$

The derivation of the filter equations can be found in various texts (e.g. Gelb,

1974; Young, 1984) and is omitted here. An excellent review and collection of key papers on Kalman filtering and its practical applications is given by Sorenson (1985).

The filter equations fall into two categories: prediction (equations 2.5 and 2.6) and correction (equations 2.7-2.9). The prediction equations produce *a priori* estimates of the system states and state error covariance matrix, while the correction equations compute the *a posteriori* estimates of the system states on the basis of a new observation vector. The measurements are given by equation 2.4. The matrices $\mathbf{A}, \mathbf{B}, \mathbf{Q}, \mathbf{R}, \mathbf{H}$ and the measurements $\mathbf{y}_k$ are all assumed to be known up to time k; the unknowns are the values of the noise terms $\mathbf{w}_k$ and $\mathbf{v}_k$, which are replaced by their expected values ($\bar{\mathbf{w}}$ and $\bar{\mathbf{v}}$). A suitable, application dependent, estimate is made for the initial value of the state vector $\hat{\mathbf{x}}_{0|0}$.

Prediction Equations: The *a priori* estimate $\hat{\mathbf{x}}_{k|k-1}$ of the system states at time k is given by

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{A}_{k-1}\hat{\mathbf{x}}_{k-1|k-1} + \bar{\mathbf{w}} \tag{2.5}$$

where $\hat{\mathbf{x}}_{k-1|k-1}$ is the *a posteriori* estimate of the system states at time $k-1$. The *a priori* estimate $\mathbf{P}_{k|k-1}$ of the state estimation error covariance matrix at time k is obtained from

$$\mathbf{P}_{k|k-1} = \mathbf{A}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{A}_{k-1}^T + \mathbf{B}\mathbf{Q}_k\mathbf{B}^T \tag{2.6}$$

with $\mathbf{P}_{k-1|k-1}$ the *a posteriori* estimate of the state estimation error covariance matrix at time $k-1$.

Correction Equations: $\mathbf{K}_k$ is the gain of the filter, computed from

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1} \tag{2.7}$$

$\hat{\mathbf{x}}_{k|k}$ the *a posteriori* estimate of the system states at time k is found from

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} - \bar{\mathbf{v}}\right) \tag{2.8}$$

and $\mathbf{P}_{k|k}$ the *a posteriori* estimate of the error covariance matrix at time k from

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\,\mathbf{P}_{k|k-1}\,(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_t{}^T \tag{2.9}$$

The filter, summarized by equations 2.5–2.9, is applied recursively with initial estimates of the system states and expected values of the error terms being made at the first step. The prediction equations are used to extrapolate information to the next time step and the correction are equations applied to update the estimates, once a measurement becomes available.

It should be noted that the *a posteriori* update of the error covariance matrix $\mathbf{P}_{k|k}$ (equation 2.9) is often given in the following simplified form

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\,\mathbf{P}_{k|k\ \ 1} \tag{2.10}$$

Althought this form is theoretically equivalent (if derived by substituting equation 2.7 into equation 2.9) better numerical stability is achieved by computing the update on the basis of equation 2.9, which is valid for *any* gain $\mathbf{K}_k$. Equation 2.10 is only valid for a symmetrical $\mathbf{K}_k$ since the error covariance matrix $\mathbf{P}_{k|k}$ must be symmetrical by definition.

2.3.2 Recursive least squares algorithm

Consider a general univariate linear process model given by

$$y_k = \mathbf{z}_k{}^T\mathbf{x} + v_k \tag{2.11}$$

where y_k is the variable of interest at time-step k, $\mathbf{z}_k$ is an m-vector of regression variables at time-step k, $\mathbf{x}$ is the m-vector of model parameters and v_k is a zero mean, Gaussian, random error term. To simplify the discussion, the process y_k and regression variables $\mathbf{z}_k$ are both assumed to have zero mean.

In the context of time series forecasting, the goal is to produce an estimate $\hat{y}_{k+n}$ of the process value y_{k+n} using $\hat{\mathbf{x}}_k$, the best estimate of the parameters $\mathbf{x}$ given the information up to time k. The lead time of the forecast is n. Values of

$\hat{y}_{k+n}$ are computed using equation 2.11, by replacing $\mathbf{x}$ with $\hat{\mathbf{x}}_k$.

The best estimate of $\hat{\mathbf{x}}_k$ is still to be found. Defining the prediction error at time k as

$$\begin{aligned} e_k &= y_{k+n} - \hat{y}_{k+n} \\ &= d_k - \mathbf{z}_k{}^T \hat{\mathbf{x}}_k \\ &= d_k - \hat{d}_k \end{aligned}$$

where d_k and $\hat{d}_k$ can be generalized to represent (for example), a sum of forecast values if this is required as the model fitting criterion, i.e. $d_k = y_{k+1} + y_{k+2}$ and $\hat{d}_k = \hat{y}_{k+1} + \hat{y}_{k+2}$.

If the prediction errors are required to have minimum variance, then the following objective function

$$U = \sum_{i=1}^{k} e_i^2 \tag{2.12}$$

must be minimized. This is the well known Least Squares criterion and the estimate of $\hat{\mathbf{x}}_k$ which minimizes U is given by the solution of the Normal equations:

$$\boldsymbol{\Phi}_k \hat{\mathbf{x}}_k = \boldsymbol{\nu}_k \tag{2.13}$$

where

$$\boldsymbol{\Phi}_k = \sum_{i=0}^{k} \mathbf{z}_i \mathbf{z}_i{}^T$$

and

$$\boldsymbol{\nu}_k = \sum_{i=0}^{k} \mathbf{z}_i d_i$$

The solution of the Normal equations (the linear system of equation 2.13) requires the computation of $\boldsymbol{\Phi}_k{}^{-1}$. However, this matrix inversion is computationally wasteful for large m and, as it turns out, there is an elegant solution which leads naturally to the Recursive Least Squares algorithm (e.g. Young, 1984).

It is possible to recursively compute

$$\begin{aligned}\mathbf{\Phi}_k &= \sum_{i=1}^{k-1} \mathbf{z}_i \mathbf{z}_i^T + \mathbf{z}_k \mathbf{z}_k^T \\ &= \mathbf{\Phi}_{k-1} + \mathbf{z}_k \mathbf{z}_k^T \end{aligned} \tag{2.14}$$

and as a result of similar reasoning

$$\boldsymbol{\nu}_k = \boldsymbol{\nu}_{k-1} + \mathbf{z}_k d_k$$

What remains is to produce a recursive estimate of $\mathbf{\Phi}_k^{-1}$ in order to solve

$$\hat{\mathbf{x}}_k = \mathbf{\Phi}_k^{-1} \boldsymbol{\nu}_k$$

A neat solution is provided by the Matrix Inversion Lemma (e.g. Haykin, 2001, pp 440), which can be stated as follows. If matrices $\mathbf{A}$ and $\mathbf{B}$ are positive-definite $m \times m$, $\mathbf{D}$ is a positive-definite $n \times n$ matrix, $\mathbf{C}$ is an $m \times n$ matrix and they are related by

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^T \tag{2.15}$$

then

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}\left[\mathbf{D} + \mathbf{C}^T\mathbf{B}\mathbf{C}\right]^{-1}\mathbf{C}^T\mathbf{B} \tag{2.16}$$

Now, let $\mathbf{A} = \mathbf{\Phi}_k$, $\mathbf{B}^{-1} = \mathbf{\Phi}_{k-1}$, $\mathbf{C} = \mathbf{z}_k$ and $\mathbf{D} = 1$. Combining equations 2.14, 2.15 and 2.16, it follows that

$$\mathbf{\Phi}_k^{-1} = \mathbf{\Phi}_{k-1}^{-1} - \mathbf{\Phi}_{k-1}^{-1}\mathbf{z}_k\left[1 + \mathbf{z}_k^T\mathbf{\Phi}_{k-1}^{-1}\mathbf{z}_k\right]^{-1}\mathbf{z}_k^T\mathbf{\Phi}_{k-1}^{-1} \tag{2.17}$$

Simplifying the notation, let

$$\mathbf{F}_k = \mathbf{\Phi}_k^{-1}$$

and

$$\mathbf{K}_k = \frac{\mathbf{F}_{k-1}\mathbf{z}_k}{1 + \mathbf{z}_k^T\mathbf{F}_{k-1}\mathbf{z}_k}$$

it then follows from equation 2.17 that

$$\mathbf{F}_k = \mathbf{F}_{k-1} - \mathbf{K}_k \mathbf{z}_k{}^T \mathbf{F}_{k-1}$$

which is a recursive update of $\mathbf{F}_k$, as required. Finally, it remains to recursively compute $\hat{\mathbf{x}}_k$

$$\begin{aligned}
\hat{\mathbf{x}}_k &= \mathbf{F}_k \boldsymbol{\nu}_k \\
&= \mathbf{F}_k \left(\boldsymbol{\nu}_{k-1} + \mathbf{z}_k d_k \right) \\
&= \mathbf{F}_k \boldsymbol{\nu}_{k-1} + \mathbf{F}_k \mathbf{z}_k d_k \\
&= \left(\mathbf{F}_{k-1} - \mathbf{K}_k \mathbf{z}_k{}^T \mathbf{F}_{k-1} \right) \boldsymbol{\nu}_{k-1} + \mathbf{F}_k \mathbf{z}_k d_k \\
&= \mathbf{F}_{k-1} \boldsymbol{\nu}_{k-1} - \mathbf{K}_k \mathbf{z}_k{}^T \mathbf{F}_{k-1} \boldsymbol{\nu}_{k-1} + \mathbf{F}_k \mathbf{z}_k d_k \\
&= \hat{\mathbf{x}}_{k-1} - \mathbf{K}_k \mathbf{z}_k{}^T \hat{\mathbf{x}}_{k-1} + \mathbf{F}_k \mathbf{z}_k d_k
\end{aligned}$$

and since it can be shown that $\mathbf{K}_k = \mathbf{F}_k \mathbf{z}_k$, it follows that

$$\begin{aligned}
\hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k \left[d_k - \mathbf{z}_k{}^T \hat{\mathbf{x}}_{k-1} \right] \\
&= \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k \xi_k
\end{aligned}$$

where ξ_k is the *a priori* measurement estimation error, which (in general) is different from e_k, the *a posteriori* estimation error.

The univariate recursive least squares algorithm for updating the parameter estimate $\hat{\mathbf{x}}_k$, can now be summarized in equations 2.18 to 2.20.

$$\mathbf{K}_k = \frac{\mathbf{F}_{k-1} \mathbf{z}_k}{1 + \mathbf{z}_k{}^T \mathbf{F}_{k-1} \mathbf{z}_k} \tag{2.18}$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k \xi_k \tag{2.19}$$

$$\mathbf{F}_k = \mathbf{F}_{k-1} - \mathbf{K}_k \mathbf{z}_k{}^T \mathbf{F}_{k-1} \tag{2.20}$$

In equations 2.18 - 2.20, $\mathbf{K}_k$ is the $m \times 1$ gain vector, $\mathbf{F}_k$ is the $m \times m$ inverse of $\boldsymbol{\Phi}_k$, $\hat{\mathbf{x}}_k$ is the *a posteriori* estimate of the parameters and ξ_k is the *a priori*

measurement estimation error

$$d_k - \mathbf{z}_k{}^T \hat{\mathbf{x}}_{k-1}$$

2.3.3 Univariate adaptive time series algorithm

For time series which exhibit non-stationarity, the parameters of the model described by equation 2.11 will be time varying. Adaptive time series models allow for the possibility of time varying parameters by modifying the objective function U (Equation 2.12) in the following way (e.g. Haykin, 2001)

$$U = \sum_{i=1}^{k} e_i^2 \lambda^{k-i}$$

λ is a factor ($0 < \lambda \leq 1$) controlling the "memory" of the filter by ensuring that the most recent values of e_k contribute most towards the objective function. This factor explicitly allows for the possibility of model parameters that change with time. If $\lambda = 1$ then the algorithm is exactly equivalent to the recursive least squares algorithm (Equations 2.18 to 2.20) presented in section 2.3.2.

The derivation of the adaptive filter equations (omitted for brevity) closely follows that presented in section 2.3.2 and results in a very similar sequence of equations.

$$\mathbf{K}_k = \frac{\mathbf{F}_{k-1}\mathbf{z}_k}{\lambda + \mathbf{z}_k{}^T \mathbf{F}_{k-1}\mathbf{z}_k}$$
$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k \xi_k$$
$$\mathbf{F}_k = \lambda^{-1}\left[\mathbf{F}_{k-1} - \mathbf{K}_k \mathbf{z}_k{}^T \mathbf{F}_{k-1}\right]$$

In the preceding three equations, the variables are the same as those discussed in section 2.3.2. The equations are applied in the order they are presented above (as new data become available) and forecast values of the time series may be obtained using

$$y_{k+n} = \mathbf{z}_{k+n}{}^T \hat{\mathbf{x}}_k$$

where n is the lead time of the forecast, y_{k+n} is the forecast estimate at time $k+n$ given all of the information up to time k and $\mathbf{z}_{k+n} = [y_{k+n-1}, y_{k+n-2}, \cdots, y_{k+n-m}]^T$ for an autoregressive model. Initial conditions must be specified for some of the terms, the initial conditions suggested by Haykin (2001, pp 620) are $\hat{\mathbf{x}}_0 = 0$ and $\mathbf{P}_0 = \mathbf{I}\delta^{-1}$ with δ a very small number and $\mathbf{I}$ the $m \times m$ identity matrix.

2.3.4 Multivariate adaptive time series algorithm

The univariate algorithm (Section 2.3.3) is easily extended to the multivariate case when the fading memories are all the same. The model retains a similar form

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k$$

where $\mathbf{y}_k$ is a p-vector containing the variables of interest at time-step k, $\mathbf{H}_k$ is a $p \times d$ matrix of regression variables at time-step k, $\mathbf{x}_k$ is the d-vector of (possibly) time varying parameters, $\mathbf{v}_k$ is a p-vector of error terms and $d = mp^2$ (if the model is of an autoregressive type). The algorithm for a recursive update of the parameter estimate $\hat{\mathbf{x}}_k$ is

$$\begin{aligned}
\mathbf{K}_k &= \mathbf{F}_{k-1}\mathbf{H}_k^T[\lambda\mathbf{I} + \mathbf{H}_k\mathbf{F}_{k-1}\mathbf{H}_k^T]^{-1} \\
\hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k[\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k-1}] \\
\mathbf{F}_k &= \lambda^{-1}[\mathbf{F}_{k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{F}_{k-1}]
\end{aligned}$$

Now, $\mathbf{K}_k$ is the $d \times p$ system gain matrix, $\mathbf{F}_k$ is the $d \times d$ inverse of the covariance matrix describing the regression variables $\mathbf{H}_k$, $\hat{\mathbf{x}}_k$ is the estimate of the parameter vector and $\lambda\mathbf{I}$ is a $p \times p$ matrix, weighting the contributions of each component in $\mathbf{v}_k$ to the prediction error criterion (e.g. the sum of squared, step-ahead, prediction errors). Forecast values of the multivariate time series may be produced using

$$\mathbf{y}_{k+n} = \mathbf{H}_{k+n}\hat{\mathbf{x}}_k$$

where $\mathbf{y}_{k+n}$ is the $p \times 1$ forecast vector at time $k+n$, given all of the information up to time k, and $\mathbf{H}_{k+n} = [\mathbf{y}_{k+n-1}, \mathbf{y}_{k+n-2}, \cdots, \mathbf{y}_{k+n-m}]^T$ for a multivariate autoregressive model. The initial conditions are consistent with the univariate case described in section 2.3.3.

2.4 Application of recursive filters to the forecasting of streamflow and spatial rainfall data

2.4.1 Case study - Stream flow forecasting using Kalman filters

Linear reservoir model

In order to demonstrate a simple use for Kalman filters in the context of flood forecasting, a case study is presented. A linear catchment model is used to forecast streamflow on the Liebenbergsvlei catchment in South Africa. Similar techniques have been used successfully in the past (Szollosi-Nagy and Mekis, 1987).

Figure 2.1 shows (schematically) the structure of the linear model used. The model consists of an arrangement of three interlinked linear reservoirs. The variable S_i represents the storage in reservoir i while the k_i are parameters which determine the reservoirs response to a given storage. Each reservoir has a loss parameter associated with it and the first reservoir is defined as the one that accepts the rainfall input. The model is conveniently represented in state-space form as (Pegram and Sinclair, 2002)

$$\begin{aligned} \mathbf{S}_t &= \mathbf{A}\mathbf{S}_{t-1} + \beta \mathbf{u}_{t-1} \\ \mathbf{y}_t &= \mathbf{C}\mathbf{S}_t \end{aligned}$$

with $\mathbf{S}_t$ the vector of storages at time t, while

$$\mathbf{A} = \begin{bmatrix} 1 - \left(\frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_5}\right) & 0 & 0 \\ \frac{1}{k_1} & 1 - \left(\frac{1}{k_3} + \frac{1}{k_6}\right) & 0 \\ \frac{1}{k_2} & \frac{1}{k_3} & 1 - \left(\frac{1}{k_4} + \frac{1}{k_7}\right) \end{bmatrix} \text{ and}$$

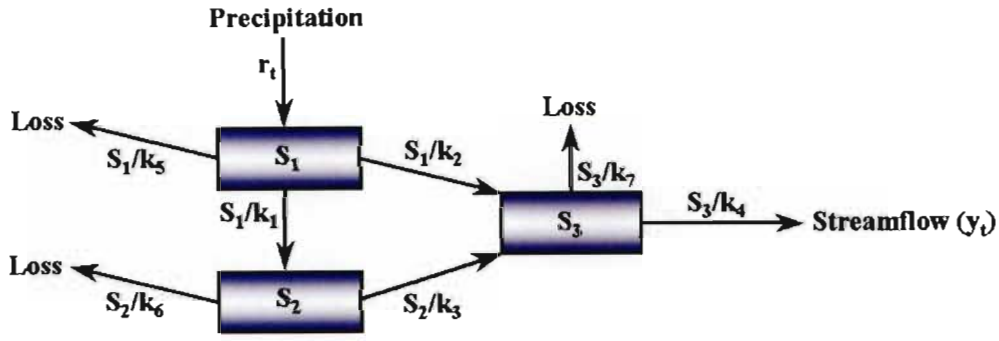


Fig. 2.1: A linear reservoir model for flood forecasting.

$$\beta = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \mathbf{C} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{k_4} \end{bmatrix}$$

The state-space formulation of the model may be directly interpreted in terms of equations 2.3 and 2.4. The Kalman filter (equations 2.5 to 2.9) may therefore be readily applied in order to update the state estimates $\hat{\mathbf{S}}_t$.

Figure 2.2 shows schematically how the triple reservoir model described above has been used in a semi-distributed sense to produce forecast flows based on the application of a Kalman filter for state updates. The catchment has been sub-divided into twelve smaller homogeneous units. Each sub-catchment is then modelled in a spatially lumped sense using average rainfall input and the linear model shown in figure 2.1. The model parameters were fitted using a different flood event of similar magnitude to those presented here.

Results

Figures 2.3 and 2.4 show forecasts for a number of lead-times of up to twelve hours ahead. The forecasts were produced for flood events on the Liebenbergsvlei catchment in the Free State province, South Africa. It is clear from these figures that the accuracy of the forecasts reduces with lead-time, as expected. This is inescapable as the confidence with which forecasts can be made reduces with lead-time even for a perfect catchment model, if the rainfall input is not known

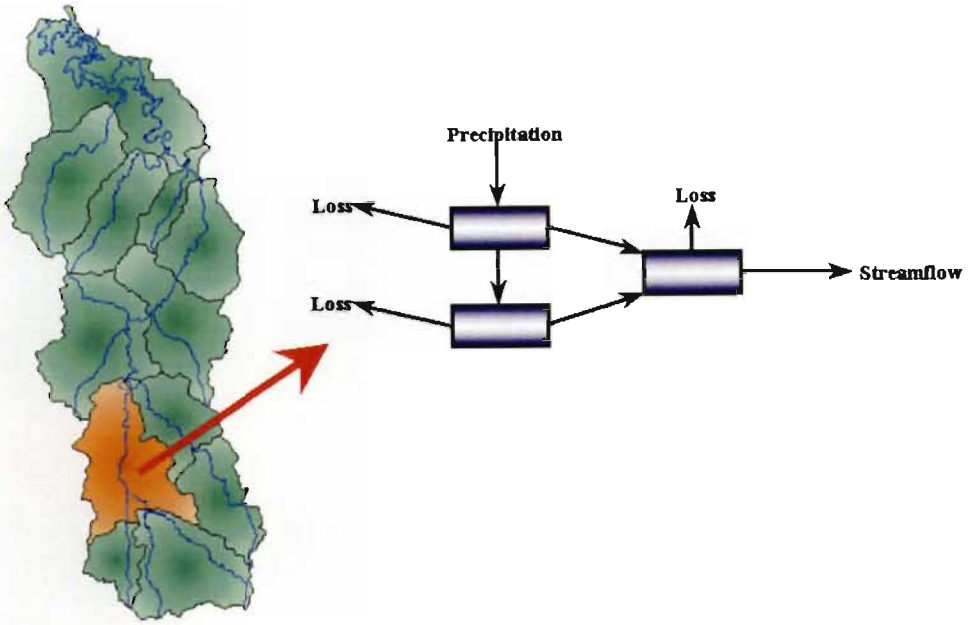


Fig. 2.2: An illustration showing how a linear reservoir model was used to model a sub-catchment area. The Liebenbergsvlei was sub-divided into 12 relatively homogeneous areas, each of which is modelled in a spatially lumped manner.

with certainty in advance. It is pleasing to note that the general character of the events is retained even by the twelve hour ahead forecasts, as this indicates a reasonable model response to the observed rainfall. In fact, it is quite remarkable that such a simple catchment model works so well. It is conjectured that the antecedent conditions of the catchment were such that a fairly linear rainfall-runoff relationship existed in both cases. No investigation was made as to whether this is the case when the catchment is relatively dry at the onset of a rainfall event.

Figures 2.5 and 2.6 show the square root of the mean sum of squared forecast errors (Root Mean Sum of Squared Errors - RMSSE) for forecasts made using the linear reservoir model with Kalman filtering compared to the RMSSE from simple persistence forecasts. A persistence forecast is one in which the current value of streamflow is used as the forecast value for all lead-times. The value is, of course, updated with each change in the forecast origin. This is equivalent to assuming that no change in streamflow is expected with time. Persistence forecasts take

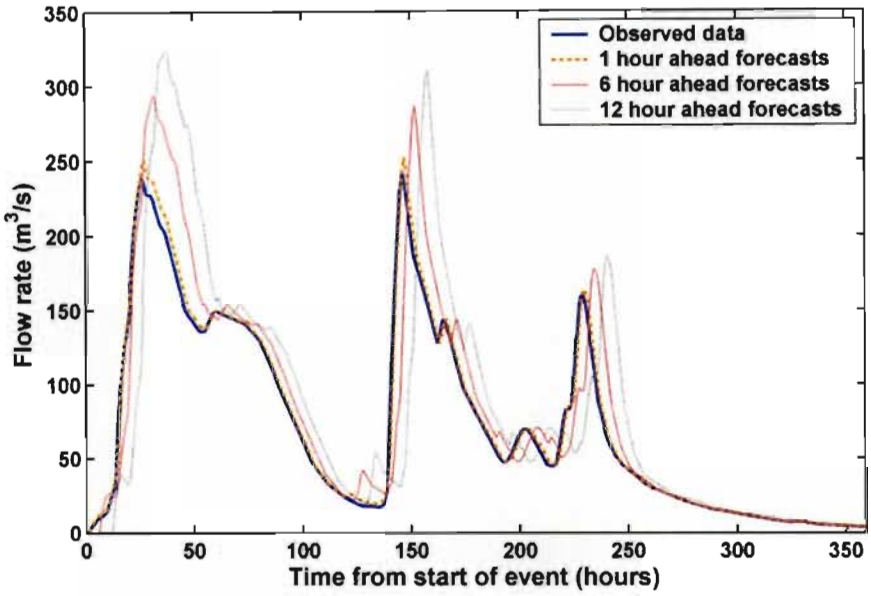


Fig. 2.3: Forecasts made using the Linear reservoir model and the Kalman filter (December 1995, Liebenbergsvlei).

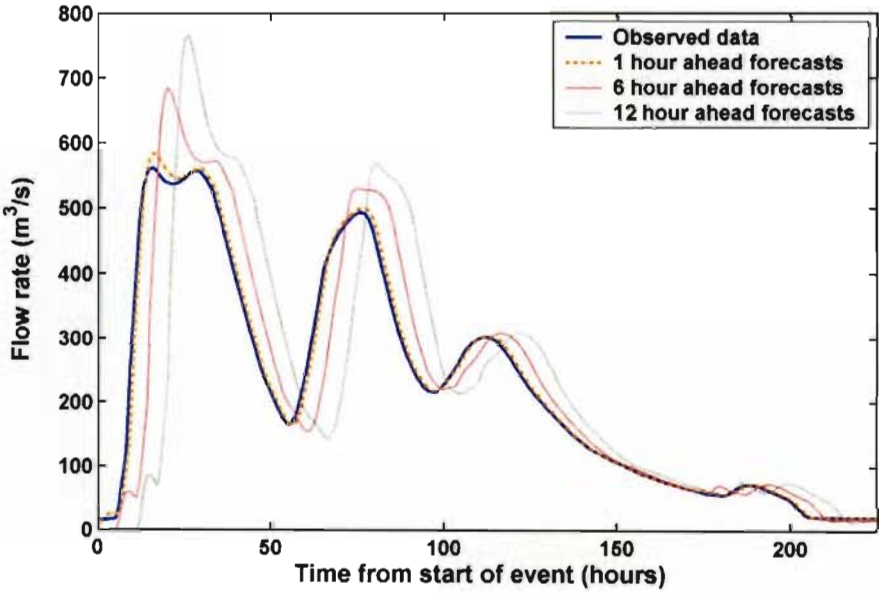


Fig. 2.4: Forecasts made using the Linear reservoir model and the Kalman filter (February 1996, Liebenbergsvlei).

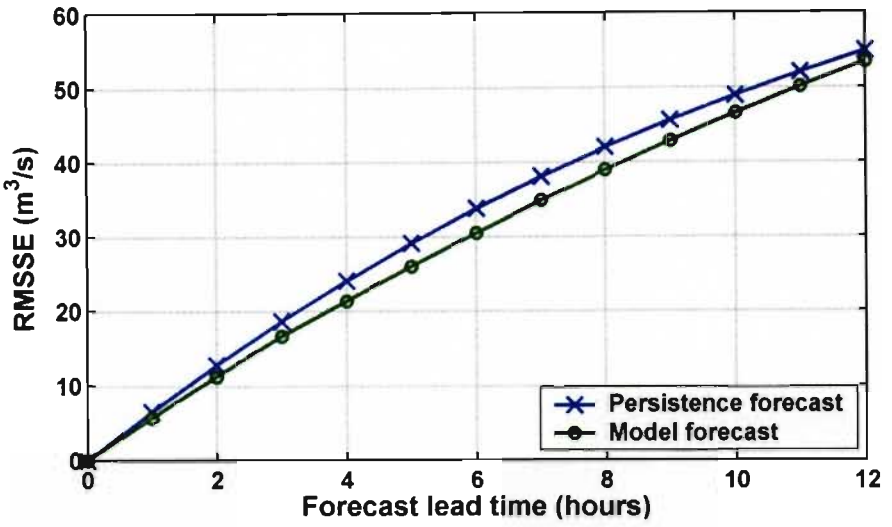


Fig. 2.5: Comparison of the root mean sum of squared forecast errors (December 1995, Liebenbergsvlei).

no cognisance of the streamflow history and figures 2.5 and 2.6 show that the Kalman filtered model forecasts outperform the persistence forecasts (despite a relatively simple rainfall-runoff model), due to their incorporation of the observed characteristics of the recent flows via the Kalman filter.

Figure 2.7 shows a set of scatter plots comparing observed and forecast stream-flows for the 1995 event of figure 2.3. The left hand column of the plots show the comparison for forecasts made using the linear reservoir model and Kalman filter while the right hand column shows the comparisons for persistence forecasts - note the larger R^2 values in the plots in the left hand column. Figure 2.8 shows the same comparisons for the 1996 event of figure 2.4. These figures provide an alternative way of showing the model forecasts to be an improvement over the persistence forecasts.

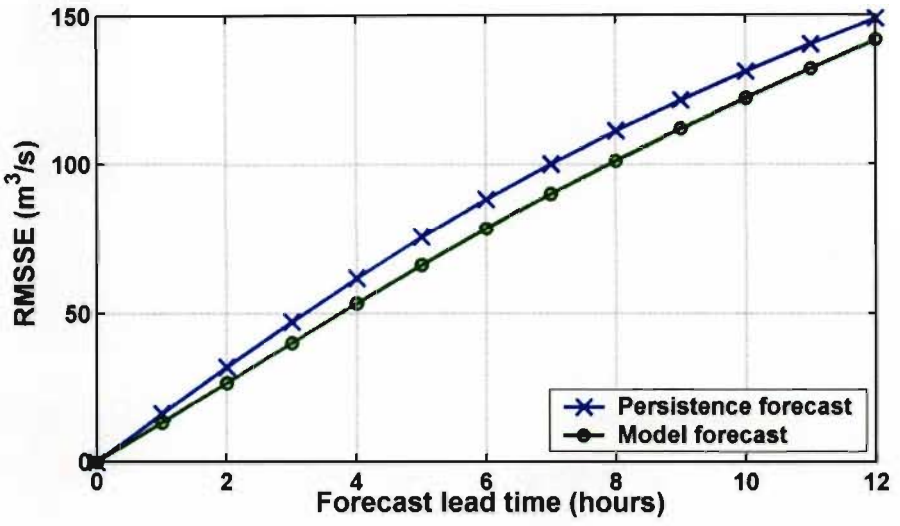


Fig. 2.6: Comparison of the root mean sum of squared forecast errors (February 1996, Liebenbergsvlei).

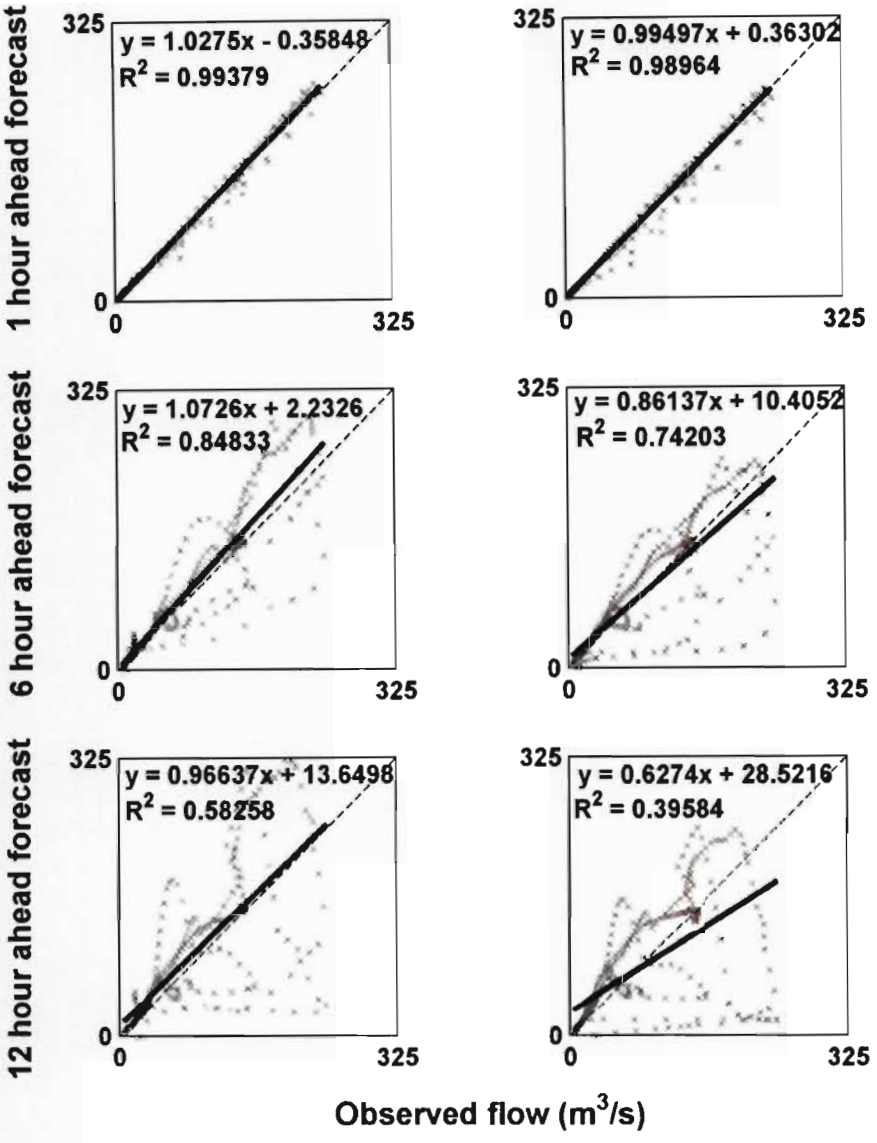


Fig. 2.7: Observed and forecast scatter-plots (December 1995, Liebenbergsvlei). Left-hand column model and Kalman filter, right-hand column persistence. Both axes of the plots have units of m^3/s.

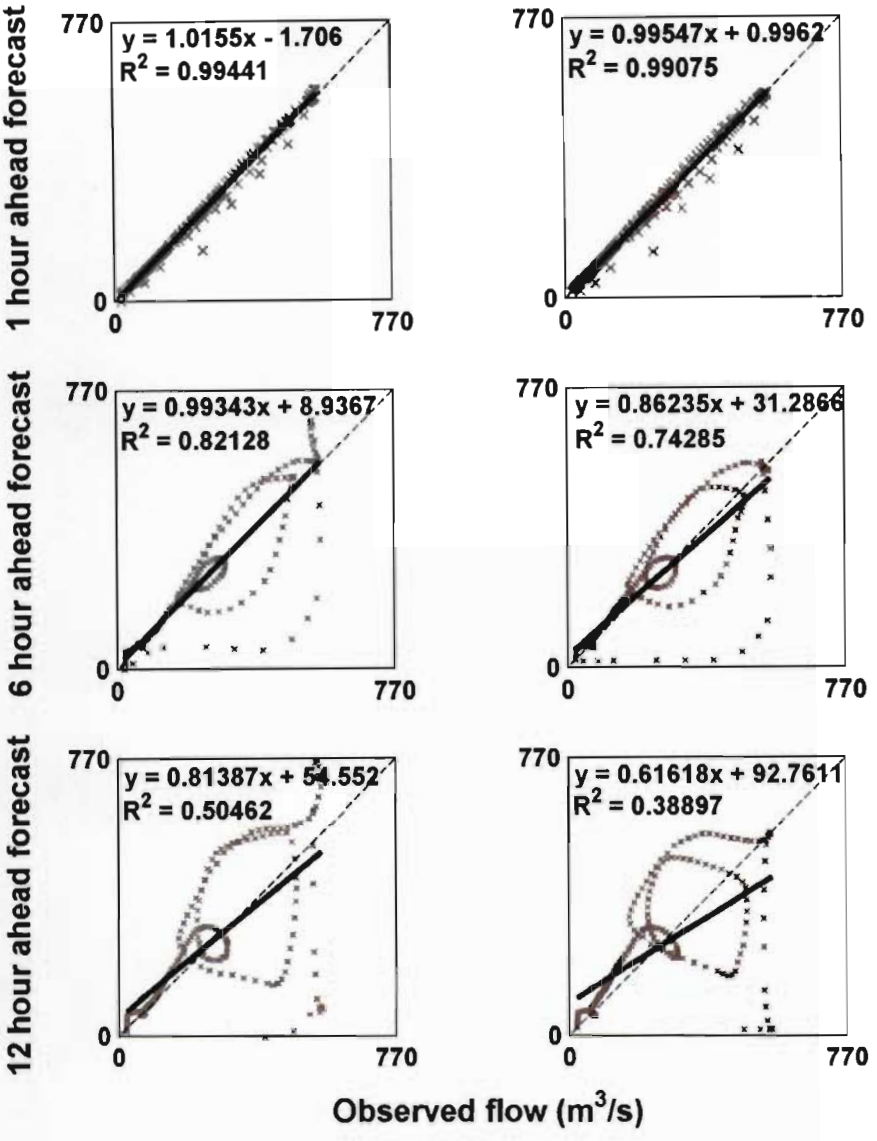


Fig. 2.8: Observed and forecast scatter-plots (February 1996, Liebenbergsvlei). Left-hand column model and Kalman filter, right-hand column persistence. Both axes of the plots have units of m^3/s.

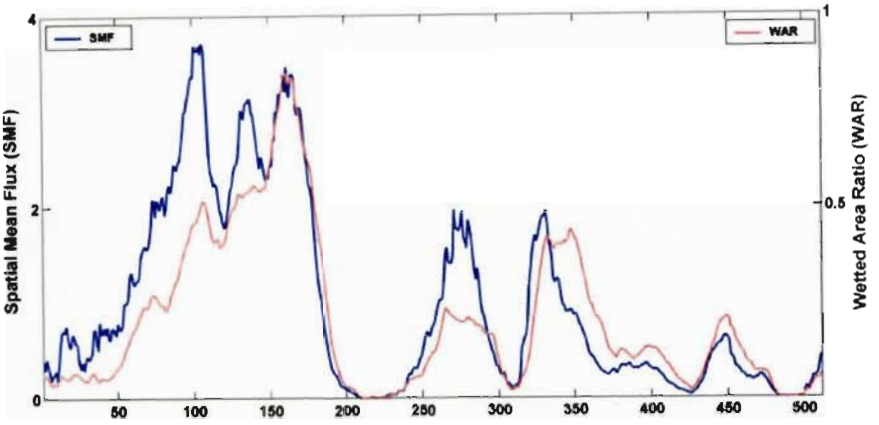


Fig. 2.9: A time series plot of WAR and SMF. The data has been sampled at 5 minute intervals over the duration of a rainfall event lasting 42 hours during the 24th and 25th of February 1996. The observations of the event were made with the SAWS MRL5 radar in Bethlehem, South Africa.

2.4.2 Case study - Adaptive time series forecasting of image scale statistics

Pegram and Clothier (2001) describe the image scale structure of instantaneous rainfall fields using two parameters, Wetted Area Ratio (WAR) and Spatial Mean Flux (SMF) – a detailed description of these parameters is given in section 4.1.2. As a useful test of adaptive forecasting algorithms for observed Hydrometeorological time series, the ATS models described in sections 2.3.3 and 2.3.4 are fitted to sample WAR and SMF time series estimated from South African radar data. Figure 2.9 shows the WAR and SMF time series for a rainfall event that lasted 42 hours on 24/25 February 1996.

This event, caused the largest flood on record entering the Vaal dam, with flows of up to 4700 m^3/s recorded (Pegram and Terblanche, 1998).

The Vaal dam catchment has an area of 38500 km^2 and experiences a mixture of stratiform and convective rainfall, most often in the summer months. In February 1996 the reservoir was full and experienced the largest inflow peak on record 4700 m^3. Co-operation between the DWAF Hydrology group and the SAWS

team in Bethlehem, resulted in controlled releases from the reservoir such that the peak outflow was attenuated to 2300 m^3. Rainfall accumulations from the MRL-5 weather radar near Bethlehem, South Africa were produced at hourly intervals for selected sub-catchments of the Vaal and relayed to the DWAF Hydrology group by telephone. The rainfall values were then manually input into the rainfall runoff model running in the DWAF offices and the resulting flow forecasts used to time releases from the reservoir. This practical solution to a large flood event (with prior warning due to the size of the catchment) relied on the skilled judgement of knowledgeable people. In the case of flash flood forecasting it is unlikely that these resources can be mobilized or that a large dam exists to allow for some attenuation of the flood. Techniques therefore require more automation and forecasting becomes critical.

Adaptive time series model fitting

One of the advantages of adaptive models is that the values of the parameters are estimated from recent observations in a recursive manner and the model is thus fitted "on the fly". However, it is necessary to select the number of parameters defining the model in an objective way (for autoregressive models, the model length fixes the number of parameters).

A generally accepted measure (Haykin, 2001), when using autoregressive models, is the Akaike information criterion, AIC (Akaike, 1973). The AIC balances the dimensionality m of the model relative to the sample size n by penalizing the sum of squares fit as m becomes larger. Increasing m increases the degrees of freedom in the model which will result in (spuriously) good fits to the data as m approaches n. However, such close fits will not necessarily contain much structural information and the AIC (a common formulation is given below) addresses this through a penalty term that enforces parsimony in the selected model.

$$\text{AIC} = n\left(\log \hat{\sigma}^2 + 1\right) + 2\left(m + 1\right)$$

Here $\hat{\sigma}^2$ is the sum of squared errors between the model forecasts and the data, m

is the model length and n the number of data. The corrected Akaike, AIC_C (Hurvich and Tsai, 1989) has been selected in this case for it's improved performance in model selection for non-stationary time series.

$$\text{AIC}_\text{C} = n\left(\log \hat{\sigma}^2\right) + n\frac{1+\frac{m}{n}}{1-\frac{(m+2)}{n}}$$

The choice of the decay factor λ, which controls the "memory" of the algorithm will also have an effect on the model fit and must therefore be taken into account. Therefore, the approach adopted here was to minimize the sum of squared errors with respect to λ for a number of different model lengths. The best model length was then selected on the basis of the AIC_C criterion.

As an example of a suitable methodology for the application of adaptive models to time series of the kind we wish to forecast, the following is offered. The decay factors λ of two univariate models were selected by minimizing the AIC_C with respect to λ and the model length for the first 200 data points of the WAR and SMF sequences shown in figure 2.9 (total sequence length 512 data points). In both cases it turned out that the optimum model length of each process fitted independently was 1. Figures 2.10 and 2.11 show the variation of the AIC_C and optimum $\lambda's$ with increasing model length m.

Figures 2.12 and 2.14 show comparisons between the observed WAR and SMF and the corresponding forecasts out to one hour ahead. The left hand panels show the forecasts produced using the adaptive time series model and it is evident for the longer lead times that the forecasts can occasionally be very unlikely. A pragmatic approach to detecting and removing these occasional "outliers" has been adopted. Ratios of successive values of both WAR and SMF were examined for the first 200 data points and a threshold ratio selected that is 2 – 3 times as high as the highest ratio observed during this period. This ratio is then applied to screen forecast values which are very different from the most recent observed value. Any forecast which doesn't pass the screening is simply replaced with the most recently observed value (i.e. a persistence forecast). The plots in the right hand panels show the forecasts produced (with the same adaptive model) using this "constrained"

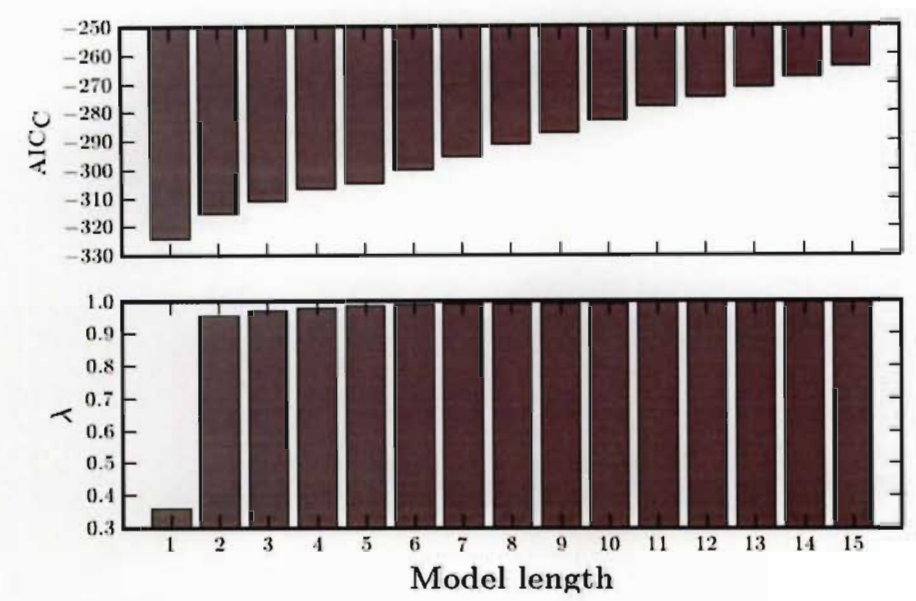


Fig. 2.10: Plots of the AIC_C and corresponding optimum λ value for a variety of model lengths. These results are for the WAR time series shown in figure 2.9.

version of the forecasting algorithm.

Scatter plots of the observed and forecast values are shown in figures 2.13 and 2.15. Again the forecasts are made as far as an hour ahead and the standard application of the adaptive algorithm is shown in the left panel while the constrained version is shown in the right hand panel. In each plot the linear regression R^2 value is also shown and this indicates a clear improvement for the constrained algorithm at the longer lead times. At shorter lead times the results of both approaches are virtually identical.

The improvement in the sum of squared errors (achieved by using the constrained algorithm) is indicated in figure 2.16 where the y-axis has been plotted with a logarithmic scale. Note that WAR is a ratio and therefore the sum of squared errors is a dimensionless quantity.

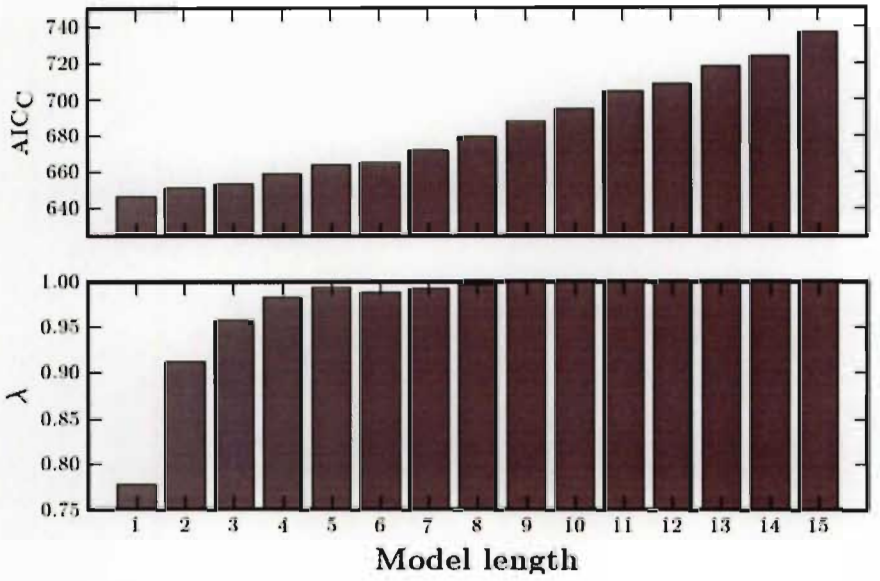


Fig. 2.11: Plots of the AIC_C and corresponding optimum λ value for a variety of model lengths. These results are for the SMF time series shown in figure 2.9.

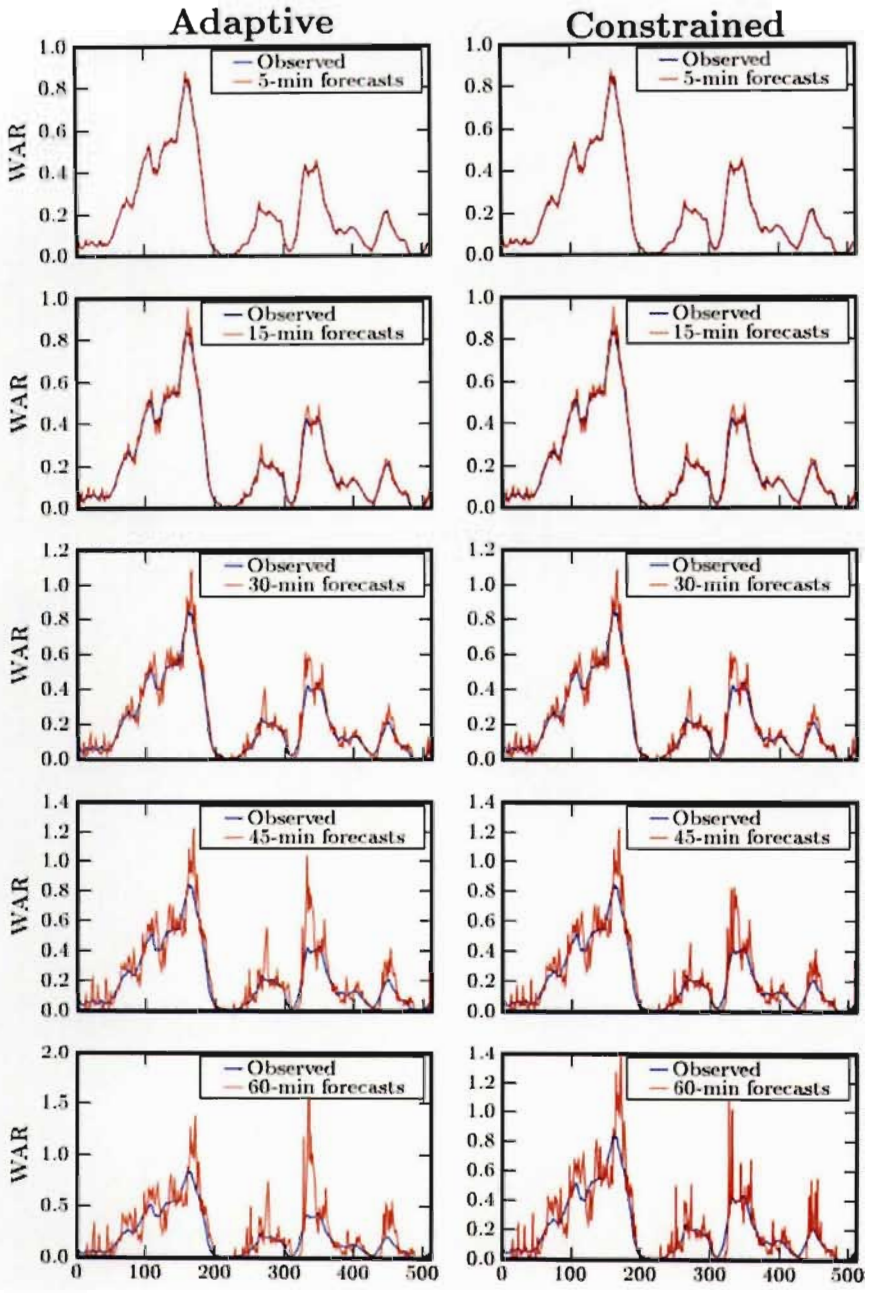


Fig. 2.12: Forecasts of WAR produced using an adaptive time series algorithm (left hand panels) and by constraining the ratios between forecasts and the most recent observation (right hand panels). Since WAR is only defined to lie between 0 and 1 the forecasts should also be appropriately thresholded.

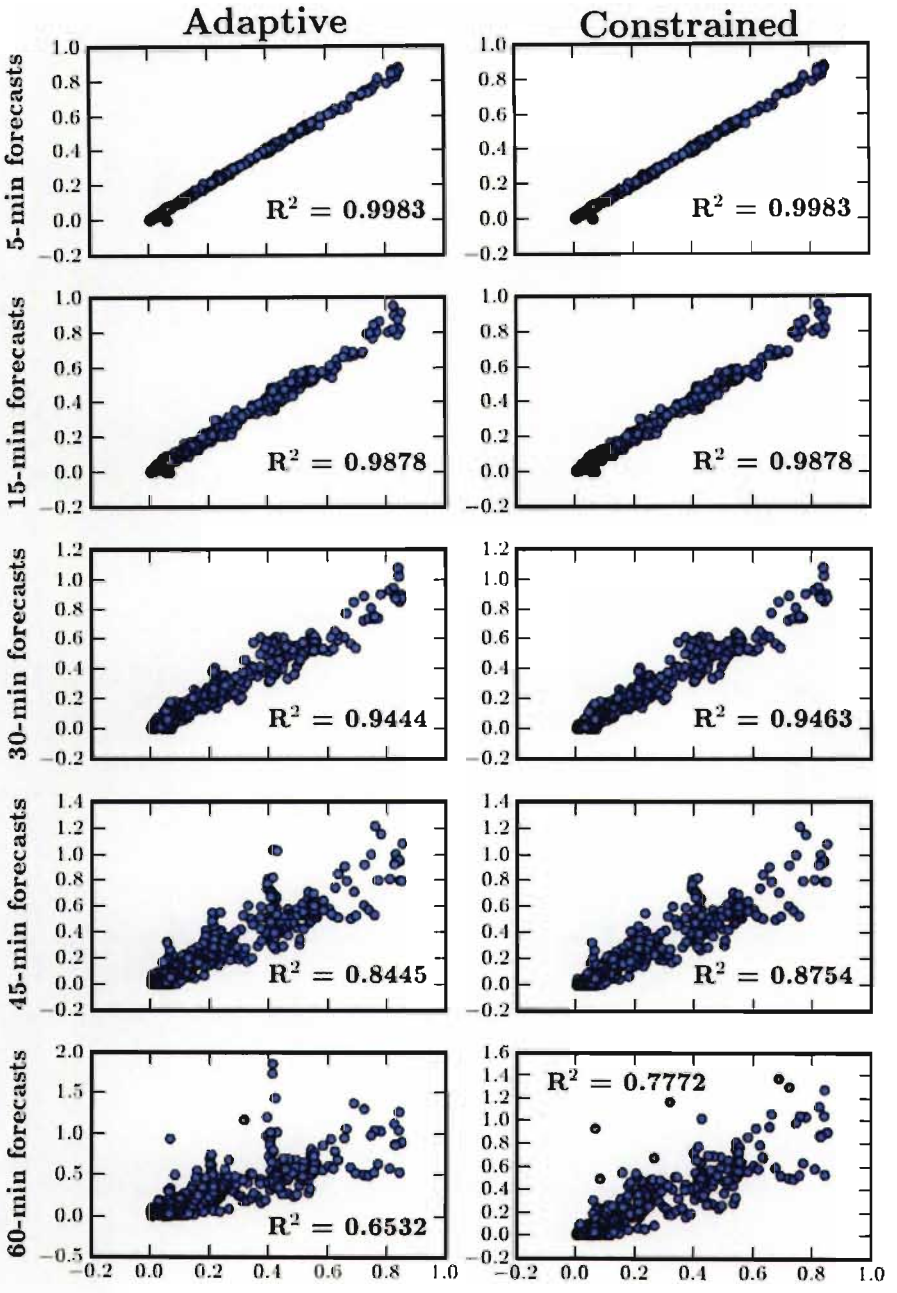


Fig. 2.13: Scatter plots of observed and forecast WAR produced using an adaptive time series algorithm (left hand panels) and by constraining the ratios between forecasts and the most recent observation (right hand panels). In each panel the R^2 value of a linear regression relating the observations and forecasts is given.

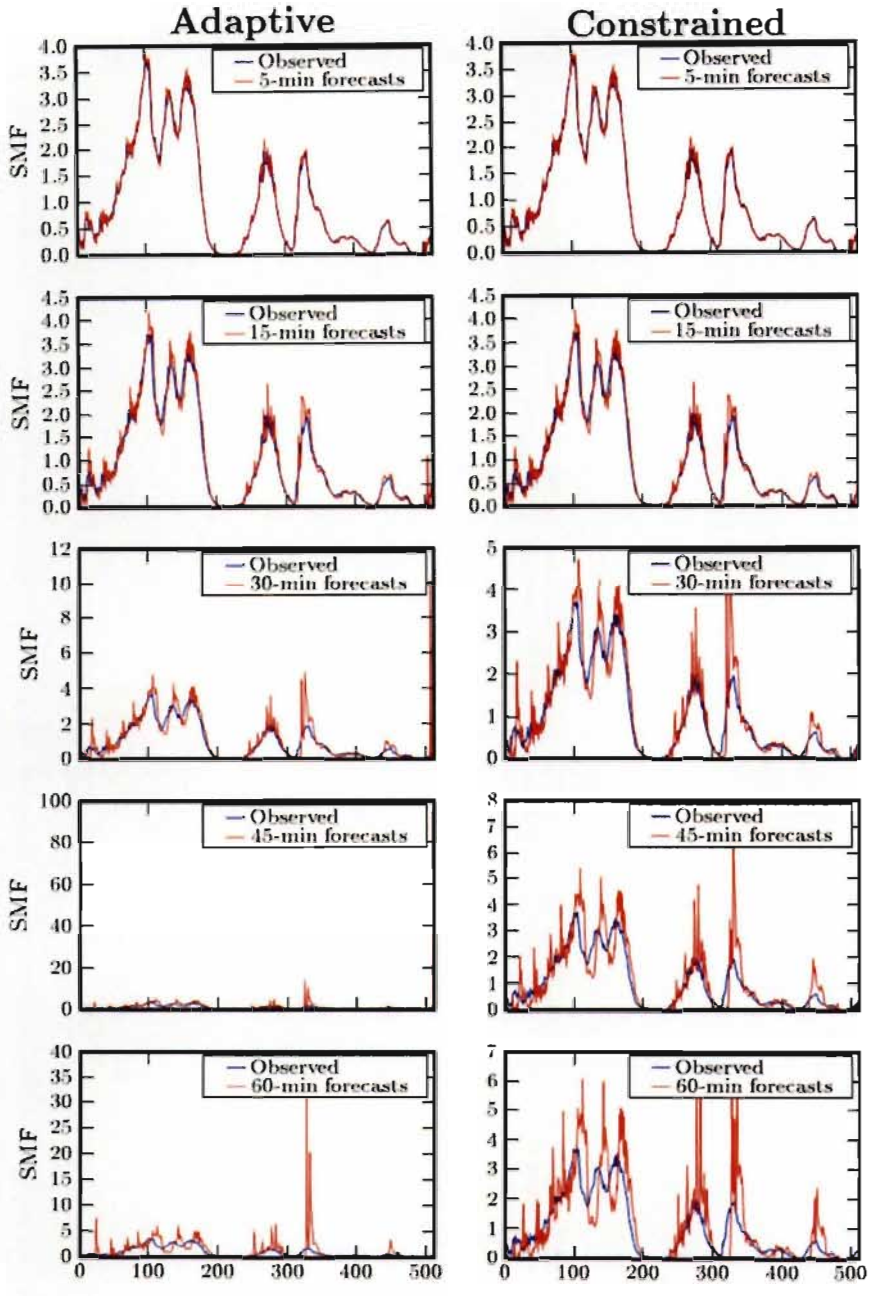


Fig. 2.14: Forecasts of SMF produced using an adaptive time series algorithm (left hand panels) and by constraining the ratios between forecasts and the most recent observation (right hand panels).

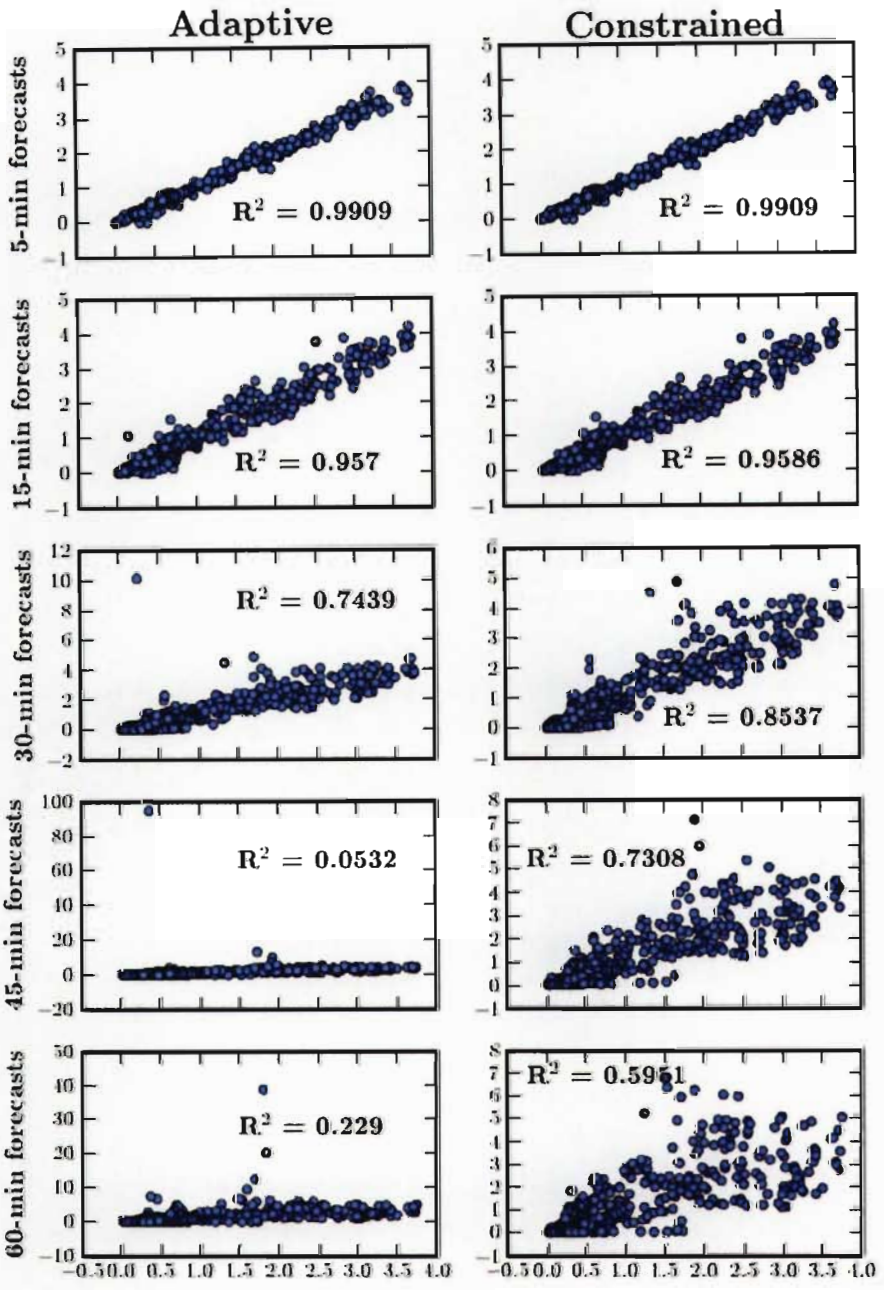


Fig. 2.15: Scatter plots of observed and forecast SMF produced using an adaptive time series algorithm (left hand panels) and by constraining the ratios between forecasts and the most recent observation (right hand panels). In each panel the R^2 value of a linear regression relating the observations and forecasts is given.

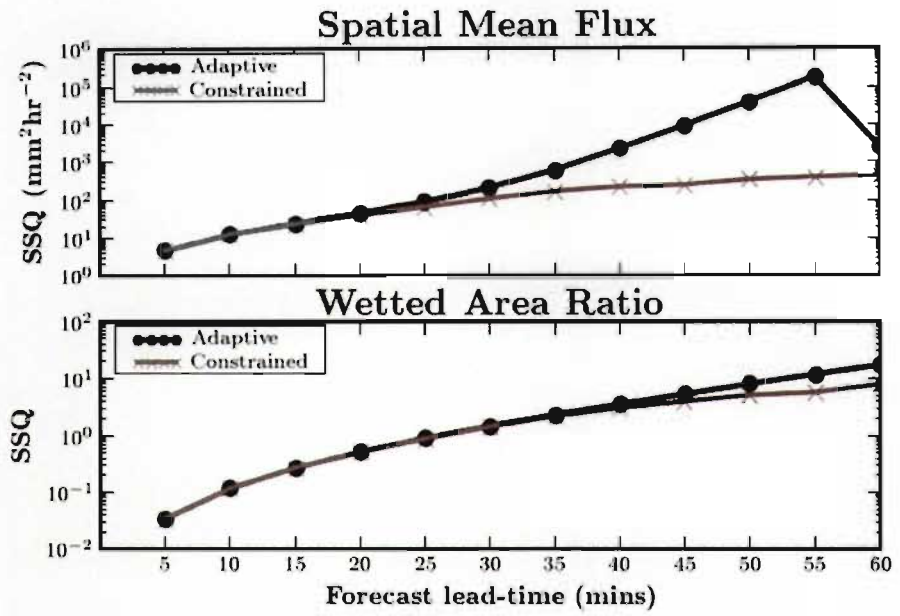


Fig. 2.16: Comparisons of the sum of squared errors for forecast lead times between 5 and 60 minutes. The adaptive and constrained algorithms are compared.

Chapter Summary: In this chapter two approaches to time series forecasting were examined as tools for dealing with the non-stationary time series frequently encountered in Hydrometeorology. The first approach is to make use of structural time series (ARMA) models (Box and Jenkins, 1970). These are well known techniques and the main focus of the chapter was therefore on adaptive methods in an attempt to handle non-stationarity in Hydrometeorological time series. The ARMA models were outlined for completeness and then the equations for the Kalman filter presented and discussed in some detail.

The set of equations for recursive estimation of the parameters of a linear regression model were developed. This was done for the univariate case in order to demonstrate the derivation of an adaptive model that allows for the possibility of time varying parameters. The model equations were then extended to the multivariate case. These adaptive filters are typically used in non-stationary signal processing applications for forecasting and control.

Finally, the Kalman filter and the adaptive filters were applied to the forecasting of real data. The Kalman filter was applied to update the states of a conceptual linear reservoir model using data from the Liebenbergsvlei catchment in South Africa. The model and Kalman filter combination was found to outperform persistence forecasts of streamflow for two flood events.

The adaptive filters were applied to the forecasting of WAR and SMF time series. The model length and decay factor λ were fitted by minimizing the AIC_C as a fitting criterion. Forecasts were produced for lead times of up to an hour ahead. The adaptive filters did a reasonable job of producing forecasts for these time series and the correspondence between observed and forecast values at longer lags was improved using a pragmatic technique of thresholding the ratio between forecasts and the most recent observation. It is interesting to note that the optimum model length was one in both cases. Also noteworthy is that the optimum values for λ are considerably less than one, suggesting that the model parameters vary throughout the event indicating non-stationary behaviour.

CHAPTER 3

RAINFALL ESTIMATION AND DATA MERGING ALGORITHMS

This chapter focuses on obtaining the optimum estimate of spatial rainfall for input to a catchment model. The challenges and pitfalls associated with combining estimates from different sensors are discussed. There is extensive literature on the subject indicating the significant research effort that has already been directed towards solving this difficult problem. Notable techniques and algorithms from the literature are highlighted and a fast, pragmatic merging algorithm suggested for use in South Africa. The chapter also deals with algorithms for the accumulation of spatial rainfall and presents a technique using advection based morphing of the rainfall field between successive observations.

3.1 Estimating rainfall

Streamflow is the direct response of a catchment to rainfall. Rainfall is a forcing mechanism which should be properly understood in order to credibly model and forecast streamflow. Understanding the dynamics of rainfall and how to measure it accurately at short space and time scales is therefore an important factor in flash flood forecasting (but also has obvious uses for rainfall-runoff model calibration). It is essential that real–time data are available since historical data are of no practical use in flood forecasting except for initial calibration of models. The instruments and data transfer mechanisms must be reliable and robust. Redundancy of the data transfer mechanisms and sources of rainfall measurement is important in this regard. The data should also be measured at, or converted

to, spatial and temporal scales which are relevant to the catchment model being used and the scale of the rainfall processes which most affect streamflow. This implies that accurate and validated accumulations are required in both space and time. The temporal accumulation of spatial rainfall fields is specifically dealt with in section 3.3.

It is beyond the scope of this thesis to focus in great detail on obtaining the best estimate of rainfall from each of the instruments which may be used to measure rainfall. This topic remains the focus of extensive international research efforts. Instead, sections 3.1.1, 3.1.2 and 3.1.3 provide a brief overview of the three principal sources of rainfall estimates used in South Africa. The main purpose of these three sections is to provide some background for the discussion that follows in the remaining sections of the chapter, which focus on the challenge of combining these estimates to produce a single "best estimate" of rainfall.

3.1.1 Rain gauges

The Hydrologist's traditional tool for measuring rainfall is the rain gauge. Rain gauges are relatively cheap, easy to maintain and provide a direct and suitably accurate estimate of rainfall at what is essentially point scale. What rain gauges fail to capture well is the spatial variability of rainfall, an important aspect for the credible modelling of a catchment's response to rainfall. This spatial variability is particularly evident at short accumulation timescales of up to several days. As the period of accumulation increases the decay of correlation (between nearby estimates) with distance is less rapid and rain gauges provide improved spatial rainfall estimates. Due to the highly variable nature of rainfall in space, simple interpolation between rain gauges does not necessarily provide an accurate estimate of the true spatial rainfall field, particularly at the short time scales relevant for flash flood forecasting.

There are several types of rain gauges in use throughout the world, the two types of gauge used operationally by the South African Weather Service (SAWS) are illustrated in figure 3.1. Climatological records are based mainly on the standard rain gauge which is read manually, on a daily basis. In most cases the gauges

Fig. 3.1: Rain gauges used by SAWS. The left-hand image shows a 127mm diameter standard rain gauge. The rainfall depth is manually read by an observer at 08:00 South African standard time (GMT + 2) each day. The right-hand image shows a tipping bucket gauge. The tips are recorded with a data logger and either downloaded manually or automatically via a cellphone messaging system.

are read by volunteers. The other source of accessible gauge data is the real-time tipping bucket gauge network. This consists of tipping bucket gauges which have been fitted with a cell phone modem and SIM card. The gauges have an embedded micro-controller, programmed to send a message to a central server at 15 min intervals while it's raining and routinely once a day during dry periods (to indicate that the gauge is functional). The real-time tipping bucket gauges are most relevant to Flash Flood Forecasting and will start to become more useful as their numbers are due to be increased by a proposed 200 gauges per year over the next five years (Terblanche, 2005).

Gauges are also subject to several sources of error which occur largely as a result of wind and it's interaction with the obstructions surrounding the gauge site. Manually read gauges suffer from observer errors, the most notable of which are spurious readings on Mondays. This problem typically occurs when gauges are unmanned during the weekend (Deyzel et al., 2004). Wilson and Brandes (1979) cite studies which indicate that gauges may under estimate by up 20% in conditions with strong winds. A detailed discussion of rain gauge errors is given in

Sevruk (1982). The errors can be ameliorated by appropriate siting, maintenance, and data quality control measures. Despite the potential errors, rain gauges are the only measurement device available which directly catches a volume of water. This makes the rain gauge a fundamental part of any rainfall measurement system, and an important tool in rainfall product validation studies.

3.1.2 Weather radars

Weather radar provides (with a single instrument) a highly detailed representation of the spatial structure and temporal evolution of rainfall over large areas. Estimated rainfall rates are indirectly derived from spatially averaged measurements of reflectivity and are therefore subject to a combination of systematic and random errors (e.g Wilson and Brandes, 1979; Austin, 1987; Chumchean et al., 2003 and many others).

At a very superficial level a weather radar works by sending out a pulse of electromagnetic energy and recording how much is returned due to reflection and back-scattering caused by water droplets in the atmosphere. In South Africa the radars scan in azimuth and elevation steps, rapidly emitting pulses of energy to produce a complete volume scan. These data are averaged over bins in polar co-ordinates and subsequently processed into Constant Altitude Plan Position Indicators (CAPPI), stored in a three-dimensional cartesian grid.

Figure 3.2 shows a typical representation of the reflectivity data collected by a weather radar. The polar data collected by the radar has been projected into a Cartesian reference frame (in this case a horizontal surface 2 km above the radar) using the DISPLACE algorithm (Mittermaier and Terblanche, 1997). Each pixel in the image represents the average reflectivity over an area of 1 km^2.

Instantaneous rainfall rates are estimated from the observed reflectivity using the well known Marshall-Palmer relationship (Marshall and Palmer, 1948):

$$Z = AR^b \tag{3.1}$$

where R is the rainfall rate in mm/hr, Z is the reflectivity in units of $10\log(dBZ)$,

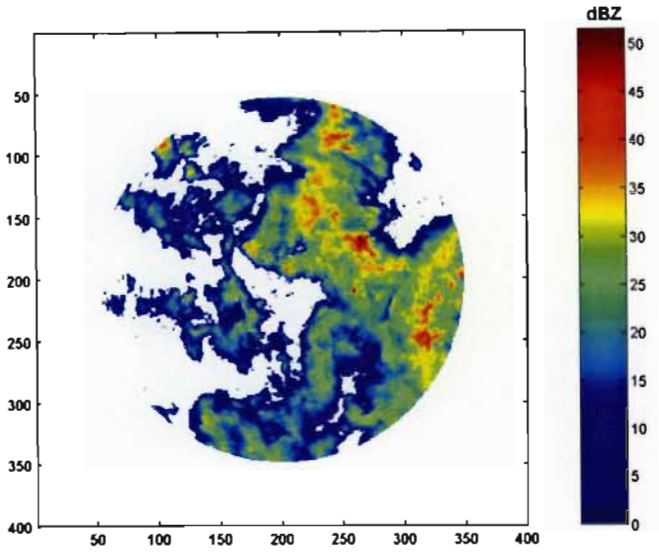


Fig. 3.2: Instantaneous reflectivity data measured by weather radar. The radar is located at the image centre and regions coloured white represent areas with either 'no rain' or 'no data'. Each pixel in this 400×400 image represents an area of 1 km^2. The colour scale to the right represents the magnitude of the reflectivity measurement, which is related to the rainfall rate (see figure 3.3).

A and b are calibration parameters. Typical values of the parameters used for South African radars are, $A = 200$ and $b = 1.6$. Figure 3.3 shows the reflectivity data in figure 3.2 converted into rainfall rate using equation 3.1 (note the change in units and colour scale).

3.1.3 Meteorological satellites

Satellite measurements provide an additional source of data for estimating rainfall. In this thesis the focus is on Meteorological satellites as their rapid update cycle is more suited to providing information for flood forecasting. In particular, the Meteosat-8 (MET-8) platform is considered since it's field of view covers the Southern African region where this work is focused. Low earth orbit satellites can provide additional useful measurements but their overpass intervals (typically

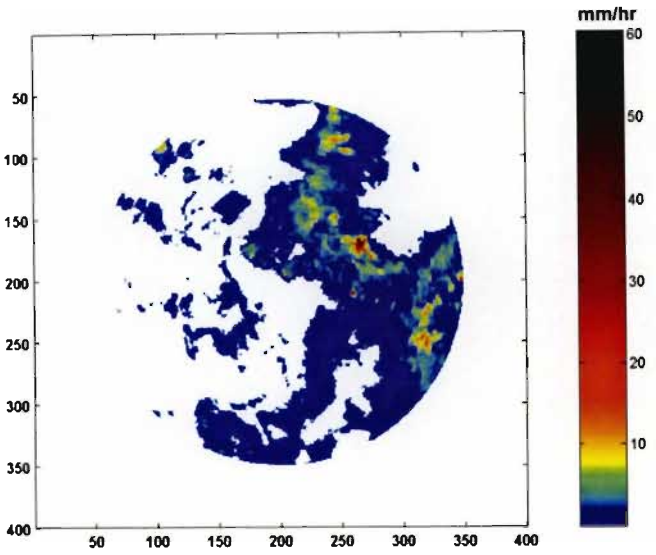


Fig. 3.3: Instantaneous rainfall rate data based on weather radar measurements. The radar is located at the image centre and regions coloured white represent areas with either 'no rain' or 'no data'. Each pixel in this 400×400 image represents an area of 1 km^2. The colour scale to the right represents the intensity of rainfall, associated with the estimated rainfall rate. The rainfall rates have been estimated for the reflectivity data of figure 3.2 using equation 3.1.

upwards of six hours) are too long to be directly useful in a nowcasting context. However, the information from these satellites can be used to inform, correct and update the MET-8 rainfall estimation algorithms at regular intervals and thereby reduce the uncertainty of the estimates (Tapiador et al., 2004a; 2004b).

Radiation measurement

Satellites measure outgoing radiation from the earth by means of radiometers. These instruments record the strength of the outgoing radiation in narrow wavelength (frequency) ranges, typically referred to as channels. Figure 3.4 shows schematically the processes which result in outgoing radiation in the Infrared (IR) portion of the spectrum (1–100 μm wavelengths). The processes for the visible (VIS) part of the spectrum are essentially the same, the only difference being that

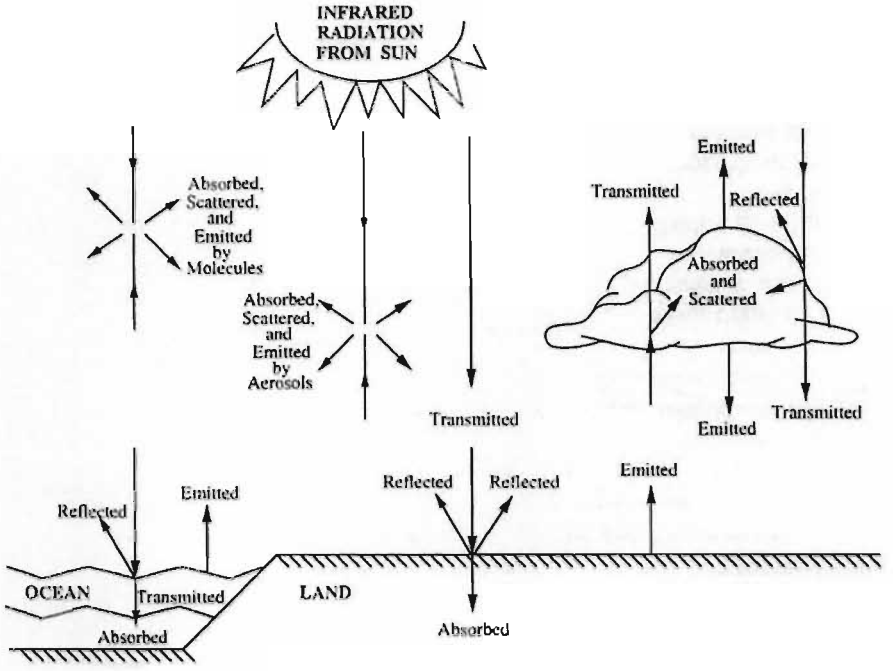


Fig. 3.4: Atmospheric absorption, scattering and transmission of infrared radiation. [After Conway (1997)]

emitted radiation from the earths surface occurs mostly at the longer IR wavelengths.

Since the different materials on the earths surface and in the atmosphere have different emission, scattering, absorbtion and reflection properties for each wavelength in the electro-magnetic spectrum, it is possible to derive considerable information about the state of the earth system from satellite based radiation measurements.

The channels measured by the instruments onboard the MET-8 satellite have been carefully selected to provide insight into cloud, aerosol and surface properties at a rapid update cycle of 15 minutes. This makes the satellite especially useful for meteorological operations and, of course, has potential applications in flood forecasting.

Figure panel	Channel name	Spectral Range (μm)
(a)	Visible 0.6	0.56–0.71
(b)	Visible 0.8	0.74–0.88
(c)	Near-infrared 1.6	1.50–1.78
(d)	Infrared 3.9	3.48–4.36
(e)	Water Vapour 6.2	5.35–7.15
(f)	Water Vapour 7.3	6.85–7.85
(g)	Infrared 8.7	8.3–9.1
(h)	Infrared 9.7	9.38–9.94
(i)	Infrared 10.8	9.8–11.8
(j)	Infrared 12.1	11–13
(k)	Infrared 13.4	12.4–14.4
(l)	High Resolution Visible	0.6–0.9

Tab. 3.1: Meteosat-8 channel descriptions. The letters in the first column of the table refer to the panels shown in figure 3.5.

Meteosat-8 characteristics

Radiance data is collected in 12 spectral channels by the Spinning Enhanced VisiblE and infraRed Imager (SEVERI) instrument onboard the MET-8 satellite. Figure 3.5 shows a set of grey-scale images representing the data collected by the satellite for a single scan time (12:00 GMT on 06 December 2005). This data is collected every 15 minutes and transmitted to user stations with a delay of approximately 15 minutes. The spatial resolution of the data is dependent on the position of a data pixel on the earths surface. MET-8 is a geostationary satellite and maintains a position at 0° Longitude (Greenwich meridian) and 0° Latitude (Equator), directly below this position (the sub-satellite point) the spatial resolution is at its finest. Moving away from this point (in any direction) the viewing angle between the satellite and a point in space increases, this effectively means that the constant scan width at the satellite sees a larger area at the surface, or cloud tops, as the point of interest becomes further away from the sub-satellite point. The sub-satellite resolution is 1 km^2 for the high resolution visible channel and 2.5 km^2 for the remaining eleven channels. These figures increase to about 2.5 km^2 and 4 km^2 respectively over South Africa.

In figure 3.5 panels *(a)* and *(b)* show the two visible band channels which have uses including cloud detection and tracking, scene identification and the monitoring of aerosols and land surfaces. Panel *(c)* shows the near-infrared channel which is used to distinguish between ice and water clouds. Panel *(d)* is the first of the infrared channels with a primary purpose of fog and low cloud detection at night. Panels *(e)* and *(f)* are the two water vapour channels, these are used (as their name suggests) to measure information related to the amount of mid-atmospheric water vapour and are also useful in determining the direction of atmospheric winds. The panels labelled *(g)* to *(k)* are all infrared channels, with the most important for rainfall estimation being *(i)* and *(j)* which provide the basic thermal information used to estimate rainfall rates from geostationary satellites, *(g)* is also useful in this context for its role in the determination of thin Cirrus cloud properties. Finally, panel *(l)* shows the high resolution visible channel. This channel has a sub-satellite spatial sampling resolution of 1 km^2 and a spectral range spanning that of the two visible channels [*(a)* and *(b)*]. Table 3.1 provides the names and more detail on the spectral bands represented in figure 3.5.

The different background colours of the panels in figure 3.5 are a result of the different information derived from each spectral band. The visible channels (figure 3.5, panels a, b, c and l) show the most reflective surfaces (clouds) as white and the least reflective (space) as black, everything else in the images is somewhere between these two extremes. The thermal channels (IR and WV) use an inverted colour scale to keep the clouds looking white, since this fits well with our perception of reality. However, the data are not measuring reflectance but temperature, thus the cold areas, clouds and space both appear white in the inverted colour scale since they are both equated with low temperatures (low emitted radiation in the thermal wavelengths).

Rainfall estimation

Combining the information from three separate channels (each being used to represent either the red, green or blue component of the colour spectrum) can produce simple false colour images like that shown in figure 3.6. The image is produced

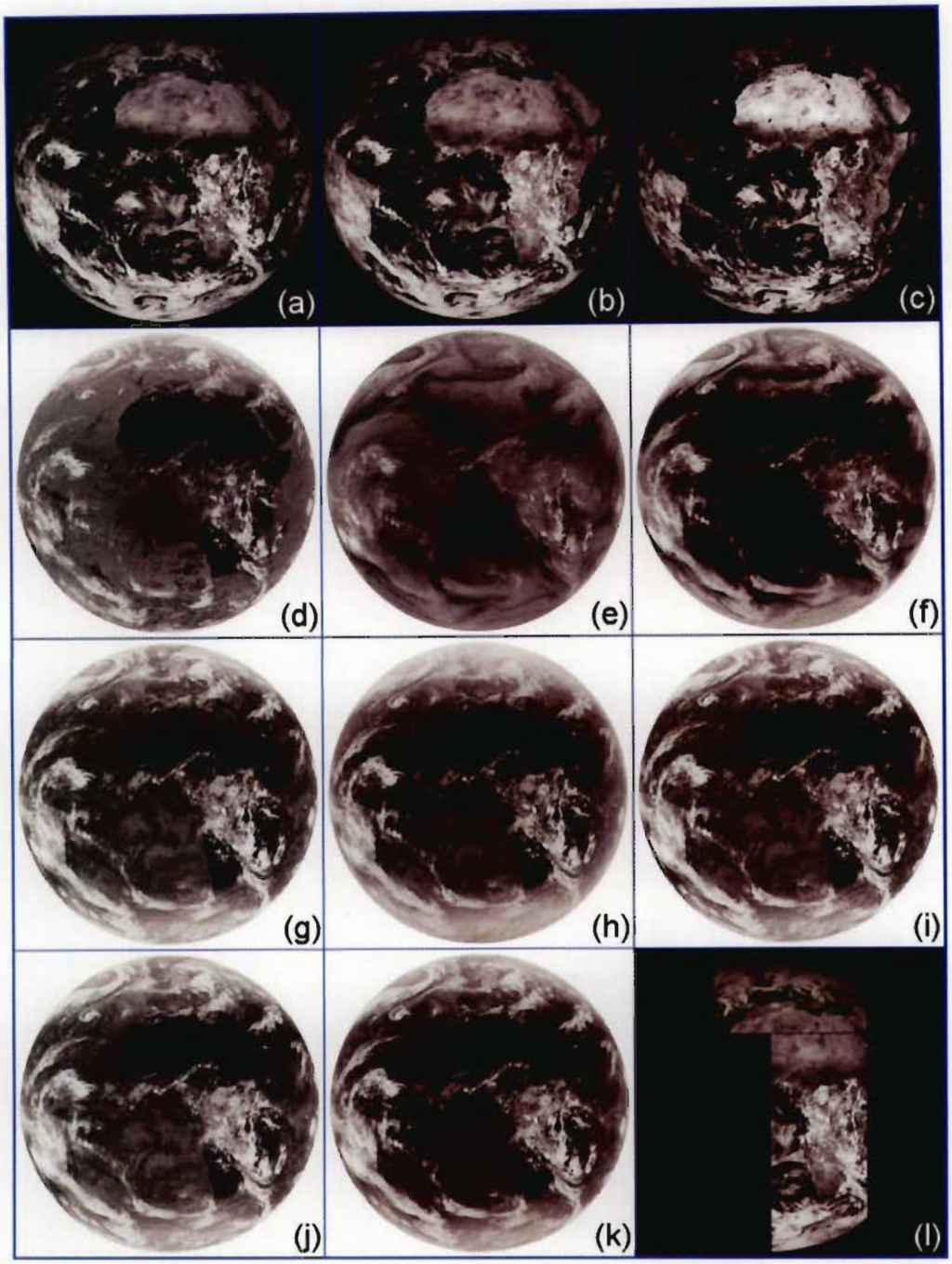


Fig. 3.5: The 12 Meteosat-8 data channels represented as grey-scale images. The data is for the 12:00 GMT scan time on 06 December 2005. The different background colours (panels a, b, c and l) are related to whether the channel is in the visible or infrared wavelength band (see the text for more detail).

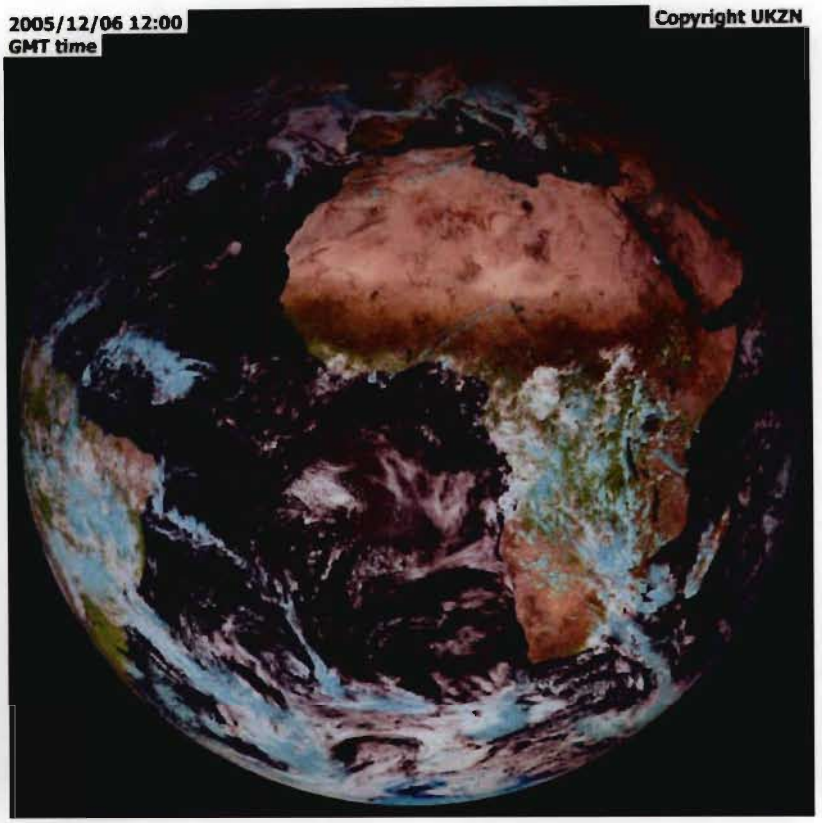


Fig. 3.6: Meteosat-8 false colour image based on channels in the visible wavelength band.

using the data from MET-8 channel 3 (figure 3.5c) to represent the red component, channel 2 (figure 3.5b) to represent the green component and channel 1 (figure 3.5a) the blue component. The turquoise shaded portions of the clouds in the image represent ice clouds which have a weak spectral signature in channel 3 and hence very little red component. False colour combinations can thus be used to highlight important or interesting features in a scene. Obviously, the images can be treated as data arrays and quantitative information relating to the features extracted, using similar reasoning in an algorithmic form.

More complex processing algorithms including remapping of the data to a suitable projection (co-ordinate system) can result in the kinds of information shown in figures 3.7 and 3.8. The first figure shows the result of processing based

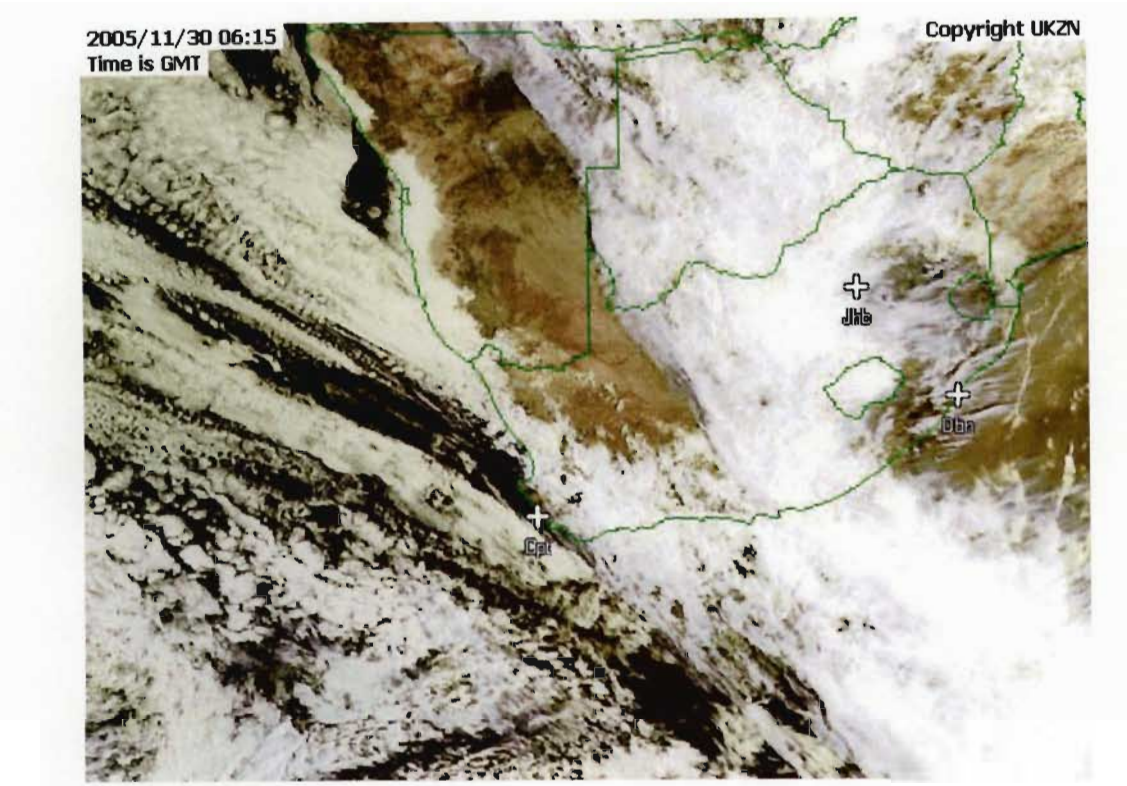


Fig. 3.7: Meteosat-8 false colour image based on channels in the visible wavelength band. The South African cities of Cape Town (Cpt), Johannesburg (Jhb) and Durban (Dbn) are marked.

on data in the visible wavelength (figure 3.7). The figure shows a large frontal system extending over southern Africa, but it is rather difficult to determine the regions in the system where convection is occurring and the potential for rainfall is high.

Using data from the infrared channels (Infrared 10.8 in this case) immediately highlights the convective regions. Figure 3.8 shows the convective portions of the frontal system as the bright orange, red and purple regions. A particularly intense convective region is evident over the western part of Zimbabwe, this detail is barely evident in figure 3.7. Because areas of convection are typically associated with high cloud tops, these equate to very low temperatures, estimated from the thermal (IR) channels. Rainfall rates are inferred from cloud top temperatures on the basis of the correlation between the low temperatures (evidence of

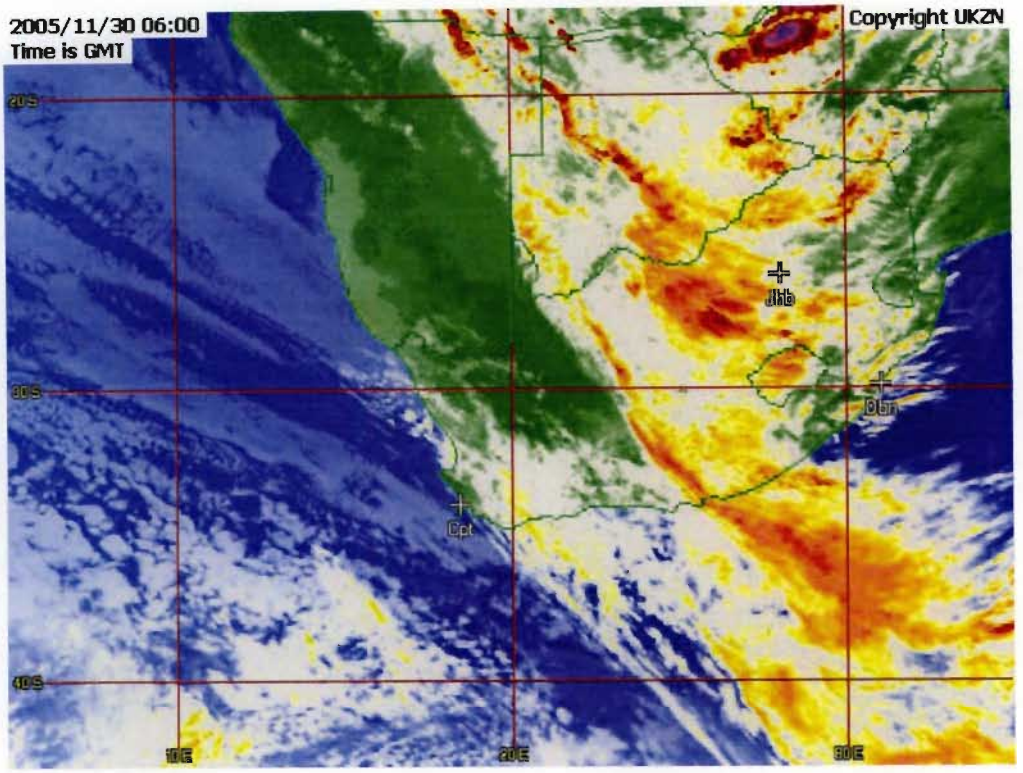


Fig. 3.8: Meteosat-8 false colour image based on channels in the infrared wavelength band. Note the intense convective development over the Western portions of Zimbabwe (upper right hand corner).

convection) and observed rainfall for historical sequences (e.g. Adler and Negri, 1988). Clearly rainfall produced by orographic forcing and other non-convective mechanisms is difficult to estimate based on the geostationary satellite data alone.

A detailed review of satellite based rainfall estimation methods may be found in Barrett and Martin (1981) and Levizzani et al. (2002). The satellite rainfall estimation procedure currently used in South Africa (Deyzel et al., 2004) can be summarized as a two-stage process: (i) mask out non-raining data and (ii) produce IR based rainfall estimates from the data classified as raining. The greatest difficulties are associated with high, cold, cirrus clouds which do not cause rain and coastal rain from warm clouds, mainly due to orographic forcing. Image texture and enhancement techniques were used to develop the final satellite rainfall

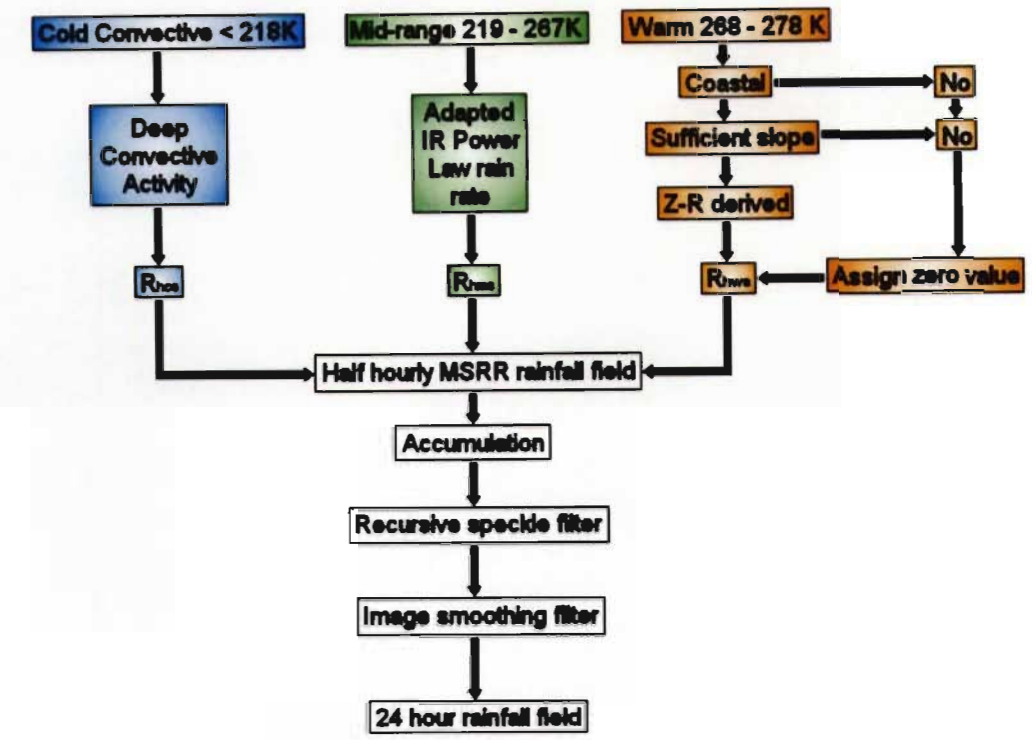


Fig. 3.9: A flow chart showing an overview of the Multi-Spectral Rain Rate algorithm (MSRR). The algorithm was developed by Deyzel et al. (2004).

product. An overview of the algorithm is given in figure 3.9, while a detailed description of the algorithm may be found in Deyzel et al. (2004) and Pegram et al. (2004).

Meteosat-8 reception station

The University of KZN installed a Meteosat-8 reception station in mid-2005. This section describes in some detail how the data get from the satellite to UKZN. With the exception of the installation of the antenna (which was contracted to an external organization), the entire data receiving and processing system was sourced and implemented by the author. This includes hardware and software procurement and set up in consultation with SAWS and the CSIR. Systems to manage the data stream were also set up by the author.

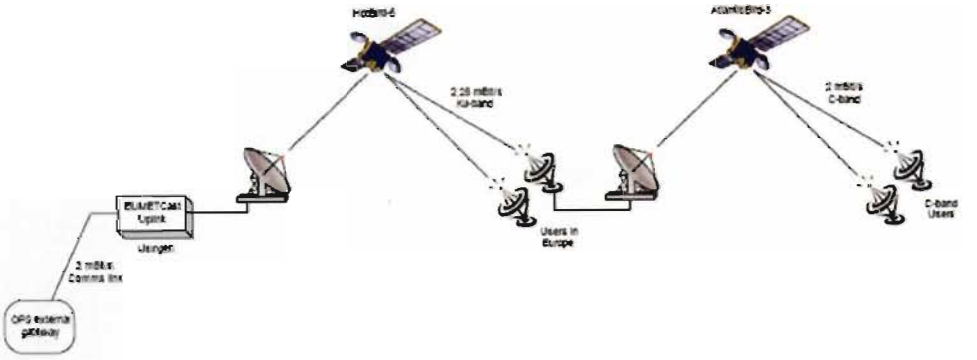


Fig. 3.10: Illustration of the transfer of Meteosat-8 data to C-band users in Africa. [Adapted from EUMETSAT, 2005]

Radiance data collected in 12 spectral channels by the SEVERI instrument on-board MET-8 are transmitted from the satellite to the EUMETSAT ground station in Darmstadt, Germany. The data are processed to produce a spatially rectified and calibrated 10-bit data array for each of the 12 channels. Transmission of the data to users is achieved via standard Digital Video Broadcast (DVB) technology; the same system which is used to view satellite television in private homes. Figure 3.10 gives an overview of the broadcast routing. The data products are transmitted to the HotBird-6 satellite from EUMETSAT's uplink station in Usigen, Germany; the HotBird-6 satellite re-broadcasts the data on a 2 MBit/s data link at a frequency of 10.853 GHz in the Ku-band. European users receive the data using small (30-50 cm) Ku-band antennas. The data stream is received in Fucino, Italy and up-linked to the AtlanticBird-3 satellite from where it is broadcast to African users over a 2 MBit/s C-band data link, at a frequency of 3.7317 GHz. Figure 3.11 shows the C-band coverage over Africa. African users need a larger (2.5m) antenna to receive the data in C-band, figure 3.12 shows the antenna at UKZN, whose software was set up and installed by the author.

In addition to the antenna, the receiving station consists of two computers. One acts as the primary receiver machine and the second is a processing server. In the configuration at UKZN, the second machine also doubles as a data server (shown in figure 3.12). The antenna is connected to the receiver PC via coaxial

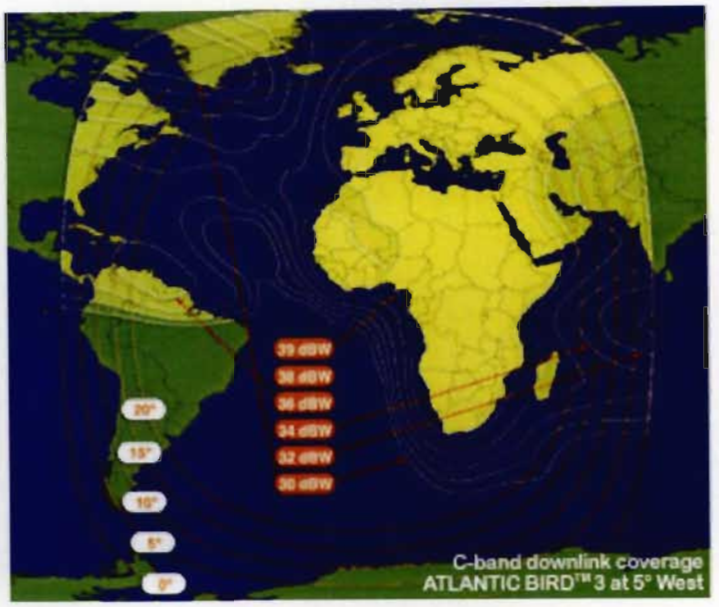


Fig. 3.11: Meteosat-8 C-band down-link coverage over the African region. [Adapted from EUMETSAT, 2005]

Fig. 3.12: Meteosat-8 C-band reception station and processing hardware at UKZN. The receiving PC and data processing server are shown.

cable and a DVB card installed in the PC. While the data is being received by the DVB card it must be decrypted in real time, using the EUMETCast client software and Encryption Key Unit (EKU). Due to the high data volumes this machine

is reserved purely for decoding the data stream. The second machine (processing server) is connected by a direct GBit link to the receiving machine and runs software which manages the raw data files. The raw data are processed into 10-bit PNG images and archived in a three day rolling buffer. All further data processing is carried out using the 10-bit PNG image format as a basis. The processing server has a second network card installed and is accessible, to selected users, via the local area network.

Meteosat-8 rainfall estimation algorithm

The rainfall estimates produced from MET-8 are indirect measurements with precipitation rates being inferred from cloud top temperatures, using the MultiSpectral Rain Rate (MSRR) algorithm (Deyzel et al., 2004). The algorithm classifies the data collected by the satellite into rain/no rain regions. The classification scheme is based on a combination of cloud top temperatures, cloud texture metrics and orography from a digital elevation model. The process has been described in some detail in Deyzel et al. (2004) and Pegram et al. (2006b) and is not repeated here.

Figure 3.13 shows the total accumulated rainfall using the MSRR algorithm and MET-7 data for the 24hr period starting at 06:00 GMT on 21/06/2005 and ending 06:00 GMT on 22/06/2005. The eastern coast of South Africa received substantial rainfall during this time. The accumulated rainfall estimates using the algorithm modified for MET-8 data are shown in figure 3.14. Comparison of the accumulations clearly shows the advantage of the improved temporal resolution provided by MET-8. The precipitation shows a smoother and more realistic texture with fewer artifacts from the masking process employed by the MSRR algorithm.

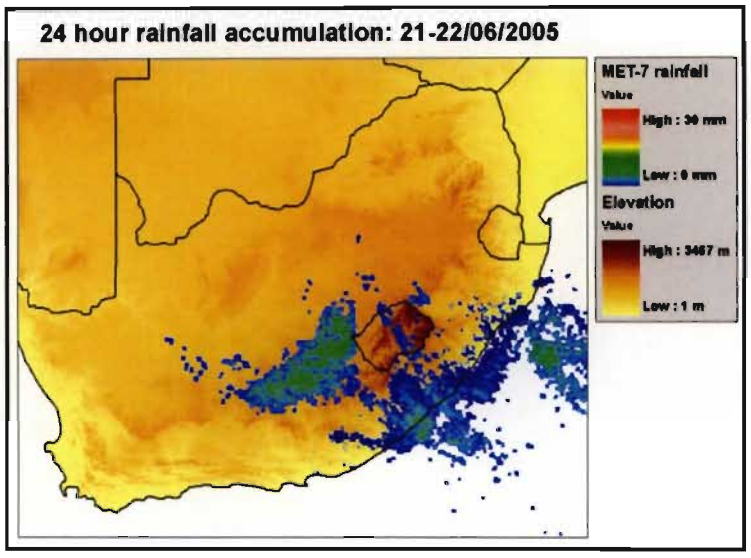


Fig. 3.13: Rainfall accumulation based on Meteosat-7 data. Data captured at 30 minute intervals was used to produce an accumulation over a 24 hour period using the algorithm described in Deyzel et al. (2004).

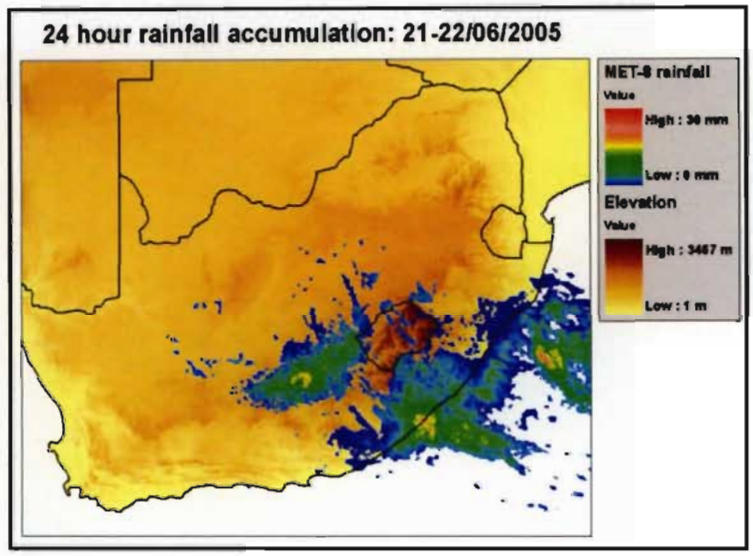


Fig. 3.14: Rainfall accumulation based on Meteosat-8 data. Data captured at 15 minute intervals was used to produce an accumulation over a 24 hour period using the algorithm described in Deyzel et al. (2004). The algorithm has been re-implemented to take advantage of the improved spatial and temporal resolution of MET-8 data, but was originally calibrated for the Meteosat-7 spectral channels.

3.2 Definition of the data merging problem

Several significant sources of complexity arise when attempting to optimally blend rainfall estimates from a variety of different sensors. In this section the major issues pertaining to the data fusion process are discussed and a technique to estimate the true rainfall field is presented. The key issue in obtaining and assessing the accuracy of spatial rainfall estimates is that the true spatial rainfall field cannot be directly measured using current technologies.

Consider modelling the rainfall rate as a spatially correlated random field e.g. as defined by Vanmarcke (1988, pp 30). The field may be described (within a region D) by

$$R(s) : s \in D \subset \mathbb{R}^d$$

at any instant in time, where $R(s)$ is the value of the rainfall rate at location s, and $d = 3$ (2 space and 1 time dimension) if we restrict ourselves to a spatial rainfall field at ground level. For hydrological applications we require an estimate of the areal average rainfall rate $R(A)$ over an area of interest $A \in D$, where

$$R(A) \equiv \begin{cases} \frac{1}{A}\int R(u)du & A > 0 \\ ave\{R(u) : u \in A\} & A = 0 \end{cases}$$

and A may represent a grid cell for a distributed catchment model or a sub-catchment area for a semi-distributed model, or even the entire catchment in the case of a lumped model.

It is usually accepted that the marginal distribution of rainfall rate, when rainfall is non-null, is Log-normal (Kedem et al., 1997; Pegram and Clothier, 1999) therefore it is possible to define $Z(s) = \ln R(s)$, where $Z(s)$ is a normally distributed random variable. If A_i is an element of a regular grid with m rows and n columns, filling D, then the problem becomes one of estimating each $Z(A_i)$ from the available observations. $R(A_i)$ is then easily obtained as $R(A_i) = \exp(Z(A_i))$

It is at this point in the analysis that the difficulties in estimating spatial rainfall begin to become apparent, each of the three sensors we are considering have a different spatial support and temporal sampling frequency. To confound things

further, the estimates of rainfall to be combined are obtained from sensors measuring independent quantities e.g. radar and satellite measurements produce inferred estimates of rainfall rates (based on reflectivity and radiation respectively), while rain gauges measure depths of rainfall as catch volumes over a given period. In order for credible rainfall estimation and validation to be done, equivalent quantities must be obtained. This implies that the estimates must first be transformed to common temporal and spatial scales, or that the merging procedure must account directly for these factors. It is assumed in this section that the best available estimates of rainfall rate have already been derived from the remotely sensed quantities (radar and satellite).

Rainfall is variable in space as well as in time, making it important to capture the spatial structure (in addition to temporal information) to provide a truly representative estimate. Since each instrument samples the rainfall field at a different spatial resolution, blending the data requires consideration of the change of support problem (Cressie, 1991; Gotway and Young, 2002). Cressie (1991, pp 284), in his treatment of the issue, points out that spatial (and temporal) averaging reduces the variance of the resulting averaged data compared to that of the underlying point process. The magnitude of the cross-correlation between the aggregated variables (compared with the cross-correlation between their 'equivalent' point neighbours) is also increased, however the sample mean remains unaffected. The different statistical properties of the processes should therefore be taken into account in the estimation of the unknown process at the spatial and temporal scale of interest. The differing statistical properties arise as a direct result of the scales at which measurements are made.

The spatial estimation of rainfall by combining information from multiple sensors has received considerable attention in the Hydro-meteorological literature. Early work (e.g. Brandes, 1975) focussed on the correction of bias in radar estimates of rainfall using an adjustment factor. Krajewski (1987) suggests a Co-kriging procedure that he demonstrates by numerical experiment. A procedure accounting for the fractional coverage of rainfall (spatial intermittency) is suggested by Seo (1998a) using conditional expectations. Seo (1998b) then goes on

to compute the expectation of rainfall at an ungauged site conditional on the observed rain gauge and radar data in places where these observations are available. A Bayesian merging technique suggested by Todini (2001) relies on a Kalman filtering scheme to remove the error variance (after a prior bias reduction) by using a Block-Kriged spatial estimate based on rain gauges as the observation vector. The Block-Kriged estimate is introduced to account for the spatial sampling differences. The novel techniques developed during the Spatial Interpolation and MApping of Rainfall (SIMAR) project (Kroese, 2004; Deyzel et al., 2004; Pegram, 2004) combine the rainfall estimates from three separate data sources using an explained variance weighting technique. Recent work which is not directly related to Hydrometeorology but which proposes an interesting approach to the problem, is presented by Wikle and Berliner (2005) who suggest a hierarchical Bayesian framework for combining information at different spatial scales and present an example of computing the stream function based on satellite and NWP model based wind observations with different spatial resolutions.

The greatest challenge in implementing the merging techniques discussed here is estimating the structure of the error covariances between the observations and the true (unknown) rainfall process. Optimal blending of the information also requires knowledge of the errors associated with each of the measurements in order to correctly provide a measure of confidence in the combined rainfall estimate. The covariance relationships between the estimates themselves may either be estimated from historical observations on the assumption of stationarity or estimated (and updated) online using a parameter filtering procedure similar to that employed by Anagnostou and Krajewski (1999a, 1999b) for adjusting radar rainfall estimates based on real-time rain gauge observations. It is also possible that remote sensing based estimates be compared to interpolated information derived from rain gauge measurements as is done by Todini (2001). However, the true rainfall field remains unknown and Hydrometeorologists must rely on the relationships between the observational data, to produce space-time rainfall estimates. Work by Habib and Krajewski (2002) has provided a starting point for quantifying the relationship between data sources and similar studies should continue to have

a strong emphasis in the future.

In SIMAR the contributions from each sensor are interpolated onto a common spatial grid. The estimates are then weighted by their expected information content based on the measured covariance structure in South African radar data for both radar and gauge observations and a skill score approach is used to determine the influence of the satellite observations (Deyzel et al., 2004). Given the previous discussion, this approach leaves some room for improvement, as do many of the other approaches cited.

Rain gauges measure "point" rainfall depths at the ground. Remote sensing techniques provide instantaneous rainfall estimates based on information above the ground at the cloud top level (satellite) or in 'bins' at several levels above the surface (radar). Wind effects (and more worryingly Virga - the evaporation of rain before it falls to the surface) can result in spatial mismatches between the estimates derived from information aloft and the rainfall sampled by the gauges at ground level. The significance of these mismatches is difficult to quantify but one which must be acknowledged. A closely related problem is that of rainfall caused by low altitude weather systems. If the radar beam elevations are such that they overshoot these systems then any combined estimates that include the null radar rainfall estimates will contain substantial errors. Cloud top heights estimated from Satellite (or NWP) data could provide a useful switch for determining when to ignore the radar information in these cases.

Before change of support issues can be addressed and blending of different data sources attempted, it is imperative that the data are referenced in a consistent co-ordinate system. The co-ordinate systems in which spatial rainfall data are collected are likely to be different for each data source and may be different from the co-ordinate system in which combined estimates are required for catchment modelling. Transformations between different co-ordinate systems will very likely change the properties of the data (particularly where the transformation results in a significant change of spatial scale) and this aspect of the rainfall estimation problem appears to have received scant attention in the literature.

Clearly the merging algorithm should also account for availability of data.

The question posed is: how does one handle the situation where one or all of the data sources become unavailable? This is a particularly pressing problem when considering automated algorithms that operate in real-time without human intervention. The available options are i) produce no output at all when there is missing data ii) continue to produce a rainfall estimate that is based on the subset of data available iii) attempt to infill the missing data record based on surrounding observations iv) some combination of the preceding, depending on which data source is missing.

Finally, the key to producing a useful rainfall product is performing credible validation exercises that account for the difficulties presented in this section. This validation process is non-trivial and needs to be carefully formulated and interpreted. Since rain gauges are the only direct measurement of rainfall available, validation is usually only possible at points in space and time where gauge measurements are available. This issue remains to be adequately addressed and is the subject of many ongoing research initiatives.

3.3 An algorithm for accumulating spatial rainfall fields measured instantaneously and intermittently

Rain gauges record accumulated rainfall depth. The accumulated totals are recorded either as the total depth in a given time period or as the times over which a given (small) depth is recorded. Rainfall information derived from Radar and Satellite sources are instantaneous samples over large areas, from which the spatial distribution of rainfall rate is estimated.

Temporally accumulated rainfall is the basic input for hydrological models and water resource management applications while many operational spatial rainfall estimation schemes work with accumulated rainfall as a basis (e.g. Smith and Krajewski, 1991; Seo, 1998a; Seo, 1998b; Anagnostou and Krajewski, 1999a; Anagnostou and Krajewski, 1999b).

The temporal accumulation of rainfall estimates needs to be handled carefully, in particular the instantaneous fields from radar and satellite. A first possibility is a

superimposition technique where the rainfall rates from each instantaneous sample are appropriately scaled and summed together. The method used by Fulton et al. (1998) for the NEXRAD radar network is to take an average rainfall rate between consecutive radar scans and apply this rate over the time interval between scans to produce an accumulated value. The accumulations between scans are then summed over the required time-period. A more complex method is described by Hannesen (2002), who suggests a path integral accumulation scheme based on a field of derived motion vectors. Hannesen's (2002) scheme turns out to be in the same spirit as a technique used by Anagnostou and Krajewski (1999a), with the difference being that they view the accumulation as a time integral which is approximated by a summation at fixed time-steps. The decision regarding the complexity of the algorithm as to which to adopt for operational purposes is determined by the trade off between computational speed and the value of information lost due to the sampling interval inadequately capturing changes in the rainfall intensity field. The accumulation method presented here uses the length of computed advection vectors to adaptively select the required computational complexity (providing a trade-off between speed and complexity, which is relevant to the data). The author believes this is a novel approach.

3.3.1 Algorithm description

The accumulation technique developed here accounts for the spatial and temporal evolution of the rainfall field between successive radar (or satellite) scans. The temporal sampling frequency of these instruments is not always short enough to ensure that important evolutionary features of the field are adequately captured. The advection of the field between successive scans is computed using an optical flow algorithm (Bab-Hadiashar et al., 1996), introduced in the context of short term rainfall nowcasting by Seed (2001). The optical flow computation (as implemented here) produces a dense field of advection vectors each associated with a particular grid point (the advection algorithm is introduced in more detail in section 4.1.1). Figure 3.15 shows a portion of the optical flow field computed between two consecutive rainfall intensity images from the SAWS weather radar in

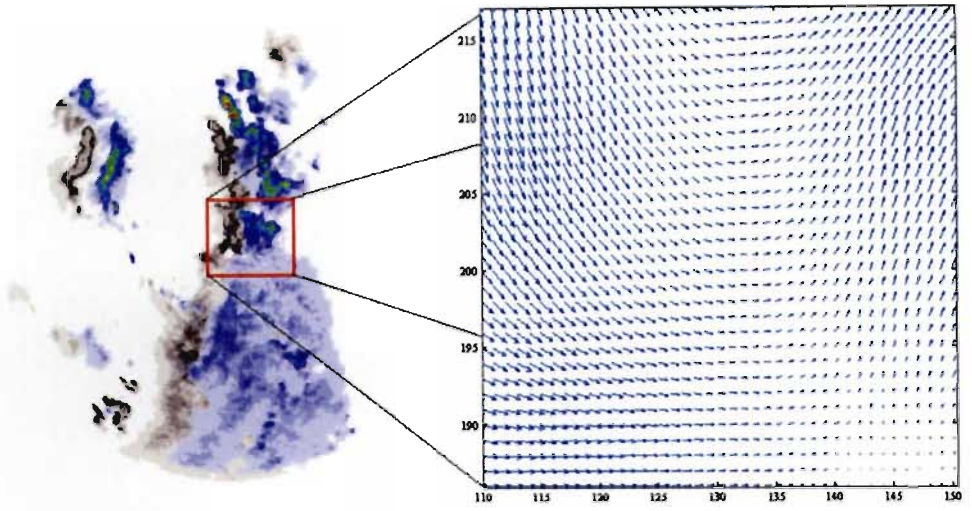


Fig. 3.15: Advection field computed using the optical flow algorithm of Bab-Hadiashar et al. (1996). Two consecutive radar scans are shown, with the first greyed out. The advection vectors in the right hand box are computed from the information contained in these two scans.

Durban, South Africa, during November 2000. The left hand side of Figure 3.15, shows two superimposed radar scans taken five minutes apart. The first scan is in greyscale while the second is in colour. A portion of the field of advection vectors, computed using an optical flow algorithm, is shown on the right of the figure. The complex nature of the advection is evident from the advection vectors shown in the right hand panel.

The accumulation algorithm is developed below (following the exposition of Hannesen, 2002). For any pixel i in the radars field of view, the accumulation on the pixel between times t_0 and t_1 is

$$A_i = \int_{t_0}^{t_1} R(s_i, t)dt \tag{3.2}$$

where $R(s_i, t)$ is the rainfall intensity at any time t on a pixel i located at position $s_i = (x_i, y_i)$ on the radar grid.

$R(s_i, t)$ is only known at the radar observation times, but if we assume that the

radar samples the rainfall field at t_0 and t_1 then it is possible to estimate $R(s_i, t)$ for $t_0 \leq t \leq t_1$ using a linear combination of the radar fields R_a and R_b observed at t_0 and t_1 respectively.

$$R(s_i, t) = \frac{t_1 - t}{\Delta t} R_a \left(s_{a|t}\right) + \frac{t - t_0}{\Delta t} R_b \left(s_{b|t}\right) \tag{3.3}$$

In equation 3.3, $s_{a|t}$ is the location that pixel i would have occupied in R_a at time t, $s_{b|t}$ is the location that pixel i would have occupied in R_b at time t, and $\Delta t = t_1 - t_0$. The locations $s_{a|t}$ and $s_{b|t}$ are determined by the pixels advection vector V_i. All the $s_{a|t}$ must be located on the path between $s_{a|t_0}$ and s_i, and all the $s_{b|t}$ are located along the path between s_i and $s_{b|t_1}$. The weighting factors $\frac{t_1 - t}{\Delta t}$ and $\frac{t - t_0}{\Delta t}$ provide a smooth transition between R_a and R_b while conserving rainfall volume over the target pixel. In order to change the time integral in equation 3.2 into a path integral, the linear weighting factors in equation 3.3 can be replaced by

$$\frac{t_1 - t}{\Delta t} = \frac{|s_{a|t} - s_i|}{\Delta s}$$
$$\frac{t - t_0}{\Delta t} = \frac{|s_{b|t} - s_i|}{\Delta s}$$

where $|s_{a|t} - s_i|$ is the distance between the points $s_{a|t}$ and s_i, similarly $|s_{b|t} - s_i|$ is the distance between $s_{b|t}$ and s_i, and $\Delta s = |s_{a|t_0} - s_i| = |s_{b|t_1} - s_i|$. Substituting into equation 3.3 gives

$$R(s_i, t) = \frac{s_{a|t} - s_i}{\Delta s} R_a \left(s_{a|t}\right) + \frac{s_{b|t} - s_i}{\Delta s} R_b \left(s_{b|t}\right) \tag{3.4}$$

Noting that

$$V_i = \frac{ds}{dt}$$

we can combine equations 3.2 and 3.4 and remove the dependence on time, arriv-

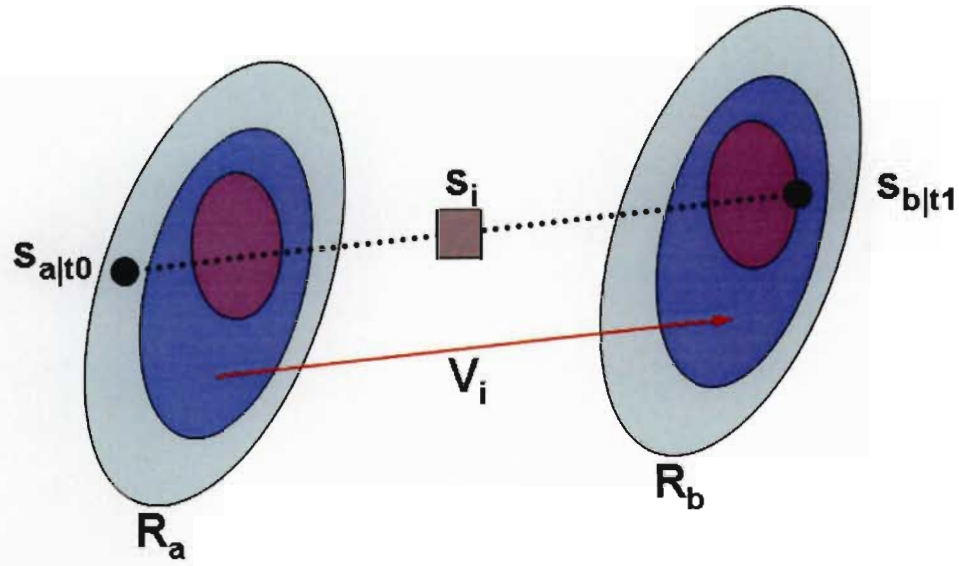


Fig. 3.16: Schematic representation of the accumulation scheme. The figure shows the location of a raincell in two consecutive radar scans (R_a and R_b). The raincell has translated with the distance and direction defined by the motion vector V_i. The total rainfall on the pixel located at s_i is the weighted sum of the integrals along the paths $s_{a|t_0} - s_i$ and $s_i - s_{b|t_1}$. If the advection were not taken into account, the pixel at s_i would not appear to receive any rain during this scan interval.

ing at

$$A_i = \frac{1}{V_i \Delta s} \left[\int_{s_{a|t_0}}^{s_i} |s_a - s_i| R_a(s_a) ds_a + \int_{s_i}^{s_{b|t_1}} |s_b - s_i| R_b(s_b) ds_b \right] \tag{3.5}$$

The total depth of rainfall accumulating on any pixel between scan times is computed from equation 3.5, over the path defined by it's advection vector. The individual depths can then be summed to produce accumulations over longer timescales, as required.

As a simple illustration of the concept figure 3.16 shows a raincell with constant contours of intensity, moving past the target pixel between time t_0 and t_1. The rainfall field measured at time t_0 is R_a and the field measured at time t_1 is R_b.

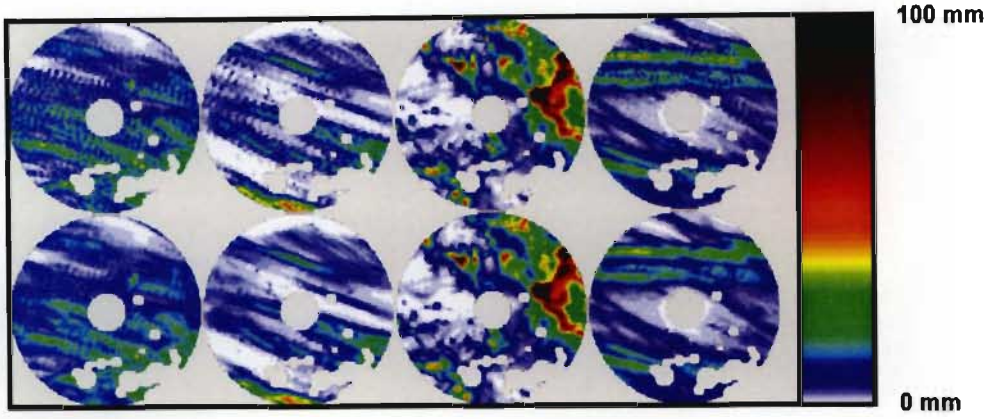


Fig. 3.17: Comparison of accumulated daily rainfall fields, produced by the two different methods. The greyed out regions indicate regions affected, either by ground clutter or missing data. The colour scale on the right indicates the total rainfall depth accumulated in 24 hours.

Since the radar grid is spatially fixed, the rain cell exhibits a translation characterized by the advection vector V_i. The two integrals along the paths $s_{a|t_0}$ to s_i and s_i to $s_{b|t_1}$ are linearly weighted so that the contribution from the data at $s_{a|t_0}$ is 100 % at time t_0 and the contribution from $s_{b|t_1}$ is 100 % at time t_1.

The integrals in equation 3.5 are computed using a discrete Trapezoidal rule approximation. The computational effort of the integration is directly related to the number of interior points that must be evaluated to compute the path integral between consecutive scans. To reduce computation time the number of interior points used is based on the ratio between the length of the advection vector and the spatial resolution of the data. Thus for a zero advection case the accumulation reduces to a simple average between the intensity values for successive scans and for large values of advection, many interior points are chosen at a step size equivalent to the spatial resolution of the data (1 km in this case).

Figure 3.17 shows a qualitative comparison between several accumulated daily rainfall fields using the currently operational simple averaging technique (top row) and the accumulations made using the advection based technique (bottom row). It is clear from qualitative observations that the precipitation swaths produced by the advection accumulation are smoother and more in keeping with what one

would expect the spatial pattern of real precipitation swaths to look like. The technique naturally preserves rainfall volume on the image (through a smoothing of the peak and trough values) and produces a more realistic spatial distribution of the precipitation swathes.

3.3.2 Algorithm performance

The performance of the newly proposed accumulation scheme was compared to the current operational scheme (used for the radar network operated by SAWS) by generating an entire year's worth of daily accumulations using each method and computing the total time taken to complete the computations in each case. This was done so that the performance of the algorithm could be evaluated for a wide variety of weather types. Figures 3.18 and 3.20 summarize the performance of the algorithm relative to the operational SIMAR implementation of the accumulation scheme. Figure 3.18 shows the total time taken to complete the daily accumulations for an entire year's images, where the number of interior points for the integration was fixed between 0 and 8 and in addition, the operational algorithm and scaled routine described in section 3.3.1 above.

Considerable effort was put into optimizing the coding and implementation of the accumulation algorithm. The performance of the first implementation shown in figure 3.18 was improved significantly as a result. The performance of the optimized code, compared to the original, is presented in figure 3.19. Although the scaled advection algorithm is 18.4 times slower than the coarse additive operational algorithm(figure 3.20), its performance is still perfectly reasonable for operational purposes. The average time (during 1999) taken on currently available computers for a daily accumulation (the accumulation of approximately 288 radar scans) is 24.5 seconds. The biggest component of the computing time is due to the optical flow computations, thus refinements in this area will significantly improve the overall performance.

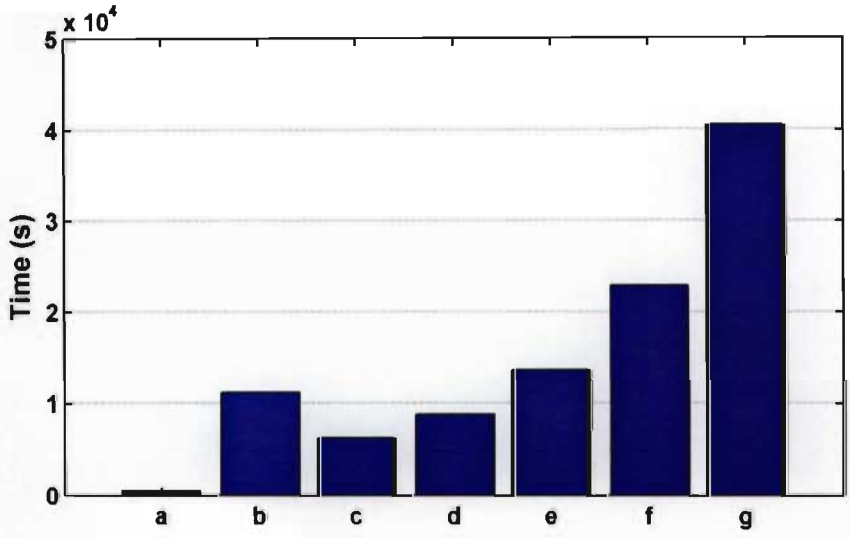


Fig. 3.18: Comparison of timings for a variety of accumulation schemes. The times represent the total time taken to perform one year's worth of daily accumulations for the MRL5 radar, using five minute interval data in 1999. *(a)* is the current coarse operational algorithm, *(b)* is the advection algorithm with adaptive choice of the number of interior points, *(c)* is the advection algorithm with no interior points (i.e. the same as for *(a)* but the advection vectors are computed), *(d)* is the advection algorithm with a single interior point, *(e)* is the advection algorithm with two interior points, *(f)* is the advection algorithm with four interior points and *(g)* is the advection algorithm using eight interior points.

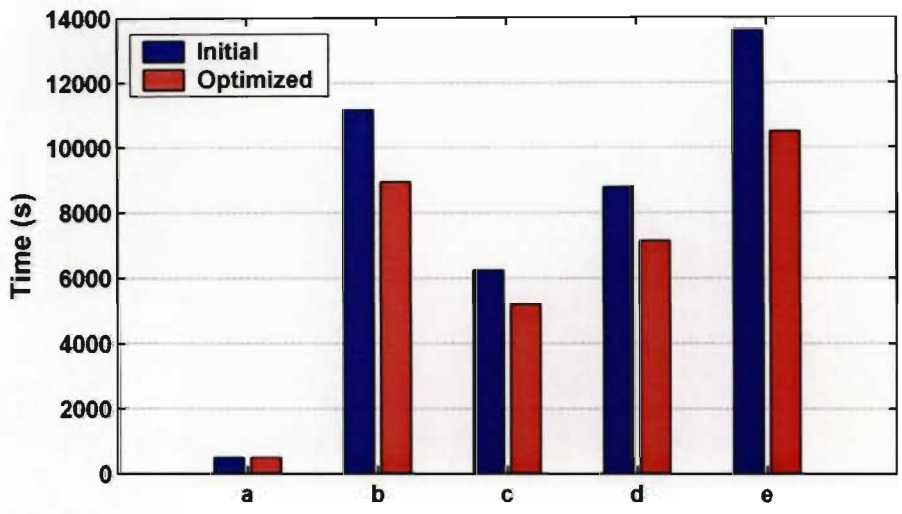


Fig. 3.19: Changes in algorithm performance after careful optimization of the code. *(a)* is the operational algorithm, *(b)* is the advection algorithm with adaptive choice of the number of interior points, *(c)* is the advection algorithm with no interior points (i.e. the same as for *(a)* but the advection vectors are computed), *(d)* is the advection algorithm with a single interior point, *(e)* is the advection algorithm with two interior points.

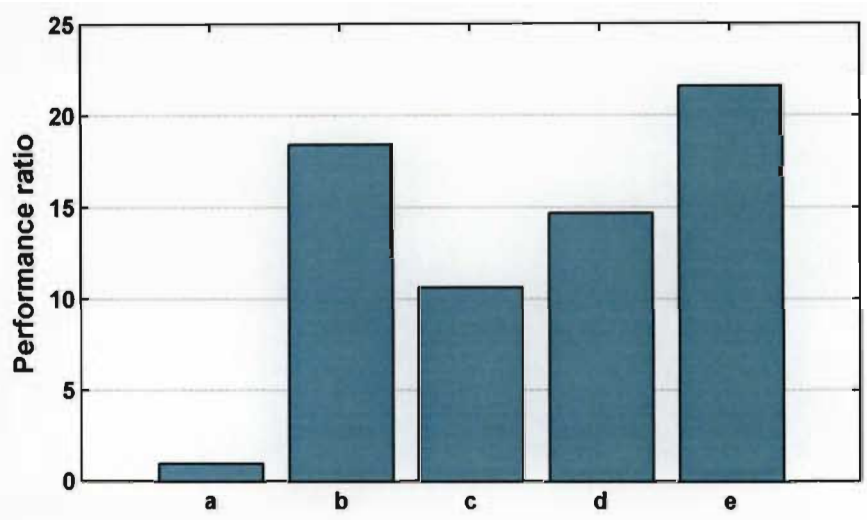


Fig. 3.20: Performance of accumulation schemes relative to the operational algorithm, after code optimization. *(a)* is the operational algorithm, *(b)* is the advection algorithm with adaptive choice of the number of interior points, *(c)* is the advection algorithm with no interior points(i.e. the same as for *(a)* but the advection vectors are computed), *(d)* is the advection algorithm with a single interior point, *(e)* is the advection algorithm with two interior points.

3.4 Bayesian merging

The Bayesian merging technique (Todini, 2001) is a means to provide an optimal combination of radar and rain gauge (or satellite and rain gauge) rainfall estimates. The algorithm requires block Kriging the rain gauge measurements of rainfall onto a grid which has the same spatial resolution as that of the radar rainfall estimates. Following the Kriging step a single recursion of the Kalman filter (Kalman, 1960) produces the optimal (in a Bayesian sense) estimate of the "true" unknown rainfall field. This estimate is only optimal insofar as the "true" rainfall field can be adequately represented by the block Kriged gauge estimate, which is treated as an unbiased observation of the true rainfall process. This assumption is obviously dependent on the spatial scale of the Block Kriged estimates, the spatial sampling of the gauges, the time scale and the nature of the rainfall event.

3.4.1 Algorithm description

The algorithm is most easily described by referring to the sequence of equations 3.6 to 3.13, where the description is limited to the merging of radar and gauges. Merging satellite and gauge estimates is a trivial extension.

- The first step is to Krige the point rain gauge information onto the radar grid to obtain the unbiased and minimum variance estimate of the unknown true rainfall field, given the information content available from the rain gauges (Equations 3.6, 3.7).

- The *a priori* estimates of the true field $\mathbf{y}'_t$ and its error covariance $\mathbf{P}'_t$ are given by equations 3.8 and 3.9, where the radar's expected bias and error structure are derived from historical observations of the errors between the Kriged gauge and radar fields (on the assumption of stationarity).

- The innovations vector $\boldsymbol{\nu}$ is computed from equation 3.10 assuming the Kriged gauge field to be the available observation of the state vector.

- The Kalman gain matrix $\mathbf{K}_t$ is computed from equation 3.11.

- The *a posteriori* estimate of the rainfall field $\mathbf{y}_t''$ and its error covariance $\mathbf{P}_t''$ are computed from equations 3.12 and 3.13 using the gain matrix to modify the *a priori* estimates.

Block Kriging

$$\mathbf{y}_t^G = \Lambda \mathbf{x}_t^G \tag{3.6}$$

$$\begin{bmatrix} \Lambda & \mu \end{bmatrix} = \begin{bmatrix} \Gamma_{lb} & \mathbf{u} \end{bmatrix} \begin{bmatrix} \Gamma & \mathbf{u} \\ \mathbf{u}^T & 0 \end{bmatrix}^{-1} \tag{3.7}$$

Kalman filter

$$\mathbf{y}_t' = \mathbf{y}_t^R - \mu_{\varepsilon_t^R} \tag{3.8}$$

$$\begin{aligned} \mathbf{P}_t' &= \mathbf{V}\varepsilon_t^R \\ &= \mathbf{V}\varepsilon_t + \left\{ \begin{bmatrix} \Lambda & \mu \end{bmatrix} \begin{bmatrix} \Gamma & \mathbf{u} \\ \mathbf{u}^T & 0 \end{bmatrix} \begin{bmatrix} \Lambda^T \\ \mu^T \end{bmatrix} - \Gamma_{ll} \right\} \end{aligned} \tag{3.9}$$

$$\nu_t = \mathbf{y}_t^G - \mathbf{y}_t' \tag{3.10}$$

$$\begin{aligned} \mathbf{K}_t &= \mathbf{V}\varepsilon_t^R \mathbf{V}\varepsilon_t^{-1} \\ &= \mathbf{P}_t' \mathbf{V}\varepsilon_t^{-1} \end{aligned} \tag{3.11}$$

$$\mathbf{y}_t'' = \mathbf{y}_t' + \mathbf{K}_t \nu_t \tag{3.12}$$

$$\mathbf{P}_t'' = \mathbf{P}_t' - \mathbf{K}_t \mathbf{P}_t' \tag{3.13}$$

where; taken in order of appearance:

$\mathbf{y}_t^G$ is the vector of Kriged gauge rainfall estimates on each grid square at time t.

Λ is the matrix of Kriging weights.

$\mathbf{x}_t^G$ is the vector of rain gauge rainfall estimates at time t.

μ is a vector of Lagrange multipliers.

Γ_{lb} is the covariance matrix between the radar grid squares and rain gauge locations.

$\mathbf{u}$ is a vector of ones.

Γ is the covariance matrix among the rain gauge locations.

$\mathbf{y}_t'$ is the *a priori* estimate of the Kalman filter state vector (the "true" rainfall field).
$\mathbf{y}_t^R$ is the vector of radar rainfall estimates on each pixel at time t.
$\boldsymbol{\mu}_{\boldsymbol{\varepsilon}_t^R}$ is the vector of mean radar estimation errors on each grid square (from past observations).
$\mathbf{P}_t'$ is the *a priori* estimate of the Kalman filter state error covariance matrix.
$\mathbf{V}\boldsymbol{\varepsilon}_t^R$ is the covariance matrix of radar estimation errors.
$\mathbf{V}\boldsymbol{\varepsilon}_t$ is the covariance of the measurement difference time series ($\boldsymbol{\varepsilon}_t = \mathbf{y}_t^R - \mathbf{y}_t^G$).
$\boldsymbol{\Gamma}_{ll}$ is the covariance matrix among the radar grid squares.
$\boldsymbol{\nu}_t$ is the vector of Kalman filter innovations.
$\mathbf{K}_t$ is the gain matrix of the Kalman filter.
$\mathbf{y}_t''$ is the *a posteriori* estimate of the filter state (the "true" rainfall field).
$\mathbf{P}_t''$ is the *a posteriori* estimate of the state error covariance matrix.

Because the algorithm is computationally intensive it is useful to list the number of operations required for each step. Denoting the number of radar pixels (the 1km x 1km radar grid squares) by m and the number of gauge observations by n, the operation count is as follows, with typical numbers of floating point operations (FLOPS) given in parentheses for $m = 40000$ (a 200×200 mesh) and $n = 50$ (a dense network of rain gauges):

Equation 3.6, $m \times n$ matrix by n-vector multiplication (2×10^6).
Equation 3.7, $(n+1) \times (n+1)$ matrix inverse and $m \times (n+1)$ by $(n+1) \times (n+1)$ matrix multiplication ($625000, 2 \times 10^6$).
Equation 3.8, m-vector subtraction or scalar subtraction from an m-vector (40000).
Equation 3.9, $m \times (n+1)$ by $(n+1) \times (n+1)$ and $m \times (n+1)$ by $(n+1) \times m$ matrix multiplications and two $m \times m$ matrix subtractions ($10^8, 4 \times 10^{12}$).
Equation 3.10, m-vector subtraction (40000).
Equation 3.11, $m \times m$ matrix inverse and $m \times m$ by $m \times m$ matrix multiplication ($64 \times 10^{12}, 256 \times 10^{16}$).
Equation 3.12, $m \times m$ matrix by m-vector multiplication and m-vector addition (64×10^{12}).

Equation 3.13, $m \times m$ by $m \times m$ matrix multiplication and $m \times m$ matrix subtraction (256×10^{16}).

The magnitude of matrix arithmetic increases with the size of the radar field that is to be conditioned, making the method computationally infeasible for very large fields. The obvious workaround is to use moving window methods and split the field into a number of smaller sub-regions. An extreme case of this thinking is to treat each pixel in the field separately, as was done in the large scale (128×128) experiments reported in section 3.6.1. This modification reduces the size of m to one, making significant savings in the computational load.

The disadvantage of the moving window approach is the loss of information from neighbouring pixels. Examination of equation 3.12 shows that the estimate of the "true" field at each pixel is a combination of the a priori estimate at that pixel and a weighted sum of the innovations at each pixel in the field. The combination of the innovations at each pixel is determined by the gain matrix $\mathbf{K}_t$ computed in equation 3.11. If it is safe to assume that the correlation structure is such that the spatial dependence dies off quickly then it makes sense to use the more computationally efficient moving window construct. A 7×7 lattice was used to determine the error to be expected from using the method suggested here (pixel at a time). The pixel-wise implementation is compared to the full method described by Todini (2001), who used a similar lattice. The results appearing in section 3.4.3 show that there appears to be no substantive difference between the two methods for the 7×7 lattice. It is postulated that as the fields become larger the computational advantages will outweigh the loss of information, particularly once the dimensions of the field become much larger than the expected spatial correlation lengths associated with common rainfall field structures. Pegram and Clothier (2001) cite typical correlation lengths for instantaneous radar rainfall measured in South Africa as ranging between 12 and 25 km. Creutin and Obled (1982) cite correlation lengths of approximately 50 km for event accumulations upwards of 6 hours in length. This suggests that the resulting information loss will therefore be greater for accumulations than for instantaneous rainfall data. Therefore it makes sense to do the merging for instantaneous data (or short accumulation periods)

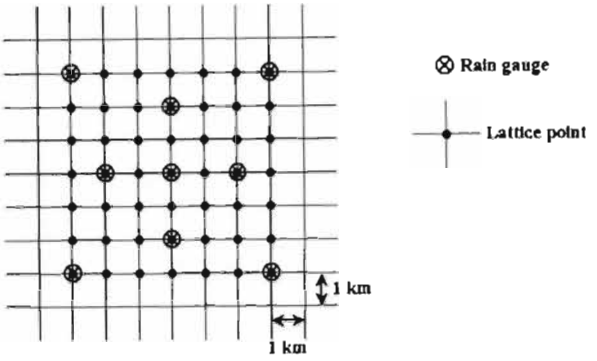


Fig. 3.21: A Schematic of the lattice showing the positions of the gauge measurement points.

and accumulate the merged fields when using this algorithm.

In the following two sections, Todini's (2001) experiment is replicated (section 3.4.2), then repeated (section 3.4.3) using moving windows of single pixel size, for comparison.

3.4.2 Verification of the Bayesian merging implementation

Todini (2001) reports a numerical experiment to demonstrate the potential of the proposed algorithm. This section describes the calculations done to replicate Todini's experiment and therefore verify that the coding of the relatively complex algorithm was accomplished correctly. This exploratory computation demonstrates the results achieved in terms of bias and variance reduction. The first step was to generate 2000 independent standard normally distributed, $N(0,1)$, realizations of a random field on a 7×7 lattice and sample the lattice at 9 gauge measurement points as shown in figure 3.21. The lattice points are labelled row-wise from 1 to 49, starting at the top left hand corner. The crossed circles represent the "rain gauge" sampling points.

Next, a spatial correlation structure was imposed on the random field in the following manner.

Compute the matrix $\mathbf{B}$ where $\mathbf{BB}^T$ is defined by

$$\mathbf{BB}^T = \mathbf{\Gamma}_{ll}$$

$\mathbf{\Gamma}_{ll}$ is the covariance matrix amongst the lattice points and is computed from equation 3.14 using an isotropic Gaussian covariance function (e.g. Cressie, 1991).

$$v(h) = \sigma^2 - \left[p + \omega\left(1 - e^{-\frac{h^2}{a}}\right)\right] \tag{3.14}$$

where σ^2 is the simulated field variance, p is the nugget, ω the sill, a the range and h the distance between points.

$\mathbf{B}$ is computed by singular value decomposition (SVD), which states (e.g. Press et al., 1992) that any matrix $\mathbf{A}$ may be decomposed in the following way:

$$\mathbf{A} = \mathbf{UWV}^T \tag{3.15}$$

if we define $\mathbf{A} = \mathbf{BB}^T$, then it follows from equation 3.15 that

$$\mathbf{B} = \mathbf{UW}^{\frac{1}{2}}\mathbf{V}^T$$

We can easily verify by multiplication that $\mathbf{BB}^T = \mathbf{A}$ as required

$$\mathbf{BB}^T = \mathbf{UW}^{\frac{1}{2}}\mathbf{V}^T\mathbf{VW}^{\frac{1}{2}}\mathbf{U}^T = \mathbf{UWV}^T = \mathbf{A}$$

since $\mathbf{U} = \mathbf{V}$ and $\mathbf{V}^T = \mathbf{V}^{-1}$ if $\mathbf{A}$ is a square symmetric matrix. The simulated rainfall field $\mathbf{y}_t$ is computed from the 49-element vector of $N(0,1)$ random noise δ_t by premultipling by $\mathbf{B}$ to impose the correlation structure.

$$\mathbf{y}_t = \mathbf{B}\delta_t$$

The gauge measurements are considered unaffected by measurement errors, but a noise field (with a different covariance structure) is generated on the lattice and added to the observations to produce a simulation of radar measurements. The

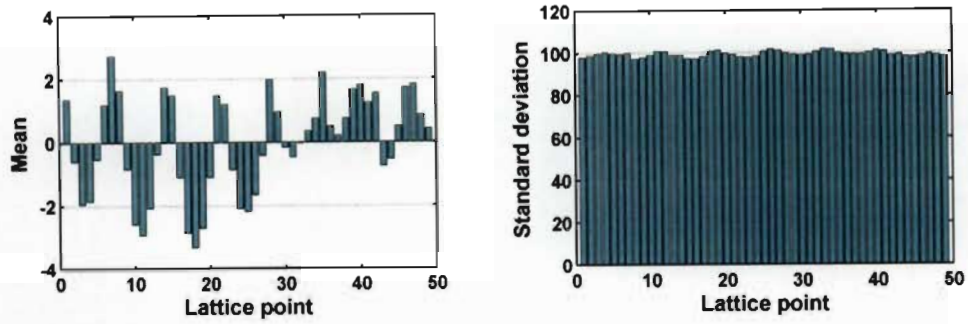


Fig. 3.22: Statistics of the simulated rainfall field $\mathbf{y}_t$ generated on the lattice points. Since the lattice points are numbered row-wise, the apparent cyclic structure is due to the spatial correlation structure imposed for each realization.

noise field is generated in the same way as the "observed" field, but with different variogram parameters (the parameters for both cases are shown in table 3.2). The

	Mean (μ)	**Variance** (σ^2)	**Nugget** (p)	**Sill** (ω)	**Range** (a)
Rainfall field parameters	0	10000	0	10000	10^7
Radar noise field parameters	40	3000	0	3000	10^6

Tab. 3.2: The Gaussian Variogram parameters used for the numerical experiment.

statistics of the simulated rainfall field and the noisy radar measurement of the true field are summarized in figures 3.22 and 3.23. Figure 3.22 shows the average mean and standard deviation for each pixel on the lattice computed over 2000 realizations of the simulated rainfall field $\mathbf{y}_t$. The apparent cyclic behaviour in the means is due to the consistent spatial correlation structure imposed for each realization. Figure 3.23 shows the mean and standard deviations for the noise affected radar observations $\mathbf{y}_t^R$ of the simulated rainfall field.

Figure 3.24, shows a single realization of a pure Gaussian noise field, a spatially correlated simulated rainfall field and a contaminated "radar" field respectively, for visualization purposes.

The coding and implementation of the Bayesian merging algorithm was carefully checked for errors, prior to its use. The first check was of the Kalman filter-

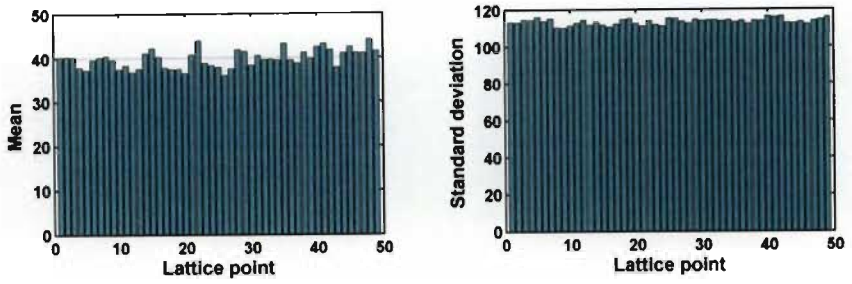


Fig. 3.23: Statistics of the noise affected radar observations $\mathbf{y}_t^R$ of the simulated rainfall field. As noted for figure 3.22, the apparent cyclic structure is due to the consistently imposed spatial structure.

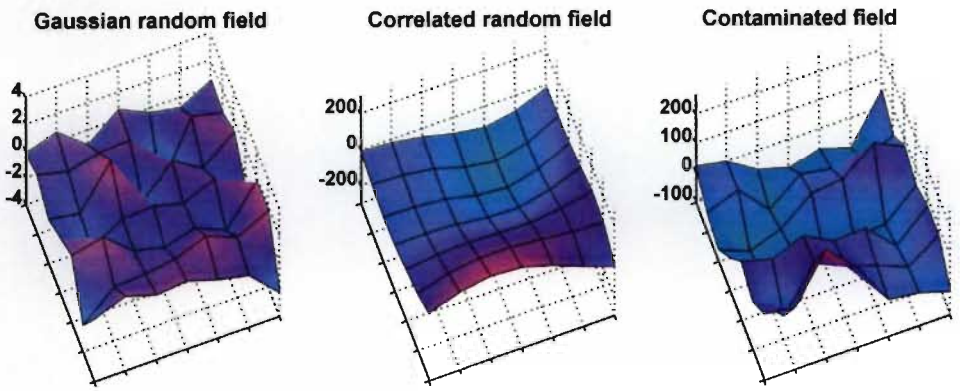


Fig. 3.24: A single realization of a spatially uncorrelated gaussian random field, the simulated rainfall field (spatially correlated) and the simulated radar rainfall field (contaminated field).

ing routine on its own. The simulated $\mathbf{y}_t$ was substituted in place of the Kriged gauge field $\mathbf{y}_t^G$. In this way the efficiency of the Kalman filter in reducing the *a posteriori* estimation error $(\mathbf{y}_t^G - \mathbf{y}_t'')$ was tested.

Figure 3.25 shows a plan view of the simulated rainfall field $\mathbf{y}_t$ (top left) and a noisy radar estimate $\mathbf{y}_t^R$ (top right) of the true field for a single realization on the lattice. The figure also shows the prior $\mathbf{y}_t'$ (bottom left) and posterior $\mathbf{y}_t''$ (bottom right) estimates of $\mathbf{y}_t$. The prior field is obtained from equation 3.8 by subtracting $\mu_{\varepsilon_t^R}$, the mean of the radar noise, from the radar field $\mathbf{y}_t^R$. The plots in figure 3.25

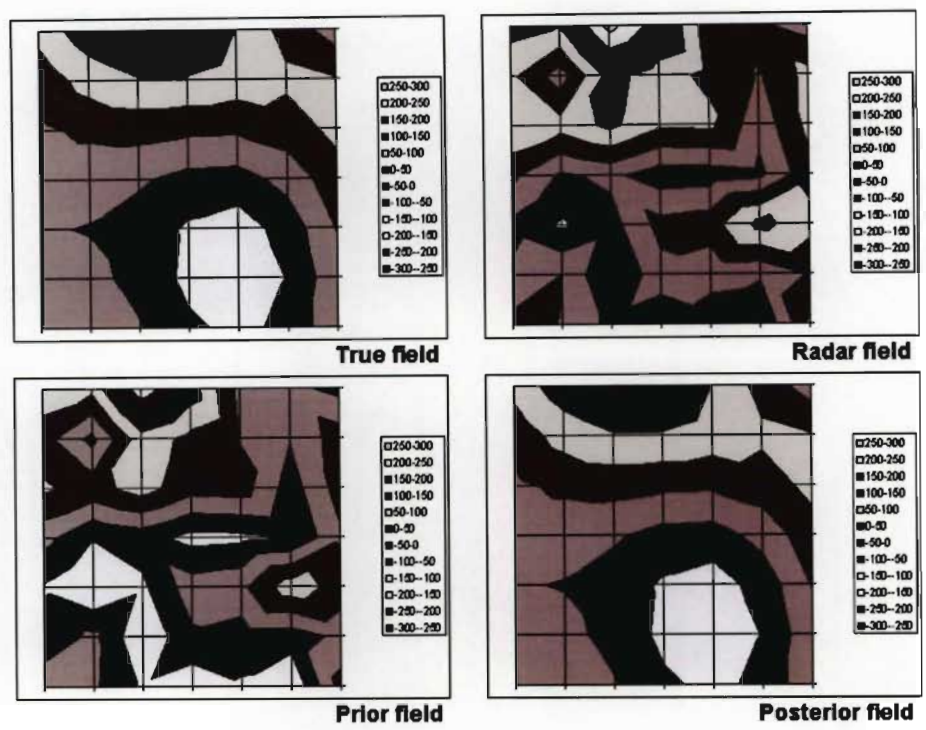


Fig. 3.25: Verification of algorithm implementation: Example fields representing a single time realization, with the Kriged gauge field replaced by the true field.

indicate that a large part of the bias is removed by subtraction of the radar error mean, to produce the prior estimate of the field. The application of the Kalman filter removes all traces of bias and also reduces the standard deviation between the posterior estimate and the true field to zero, as it should do in this case, where the statistics of the ε_t time series are know *a priori* and an "error free" observation of the rainfall field is available.

A second check was carried out to verify that the Kriging portion of the code did not contain any errors. In this case the set of gauge $\mathbf{x}_t^G$ measurements was replaced with the 49 simulated rainfall field values on the 7×7 lattice. The Kriging procedure was employed to Krige these values back onto the lattice points. Evidently the original field should be recovered exactly, as the matrix of Kriging weights $\mathbf{\Lambda}$ must become the identity matrix in this case. This was achieved and for the single realization of the two fields compared in figure 3.26, there is no visible

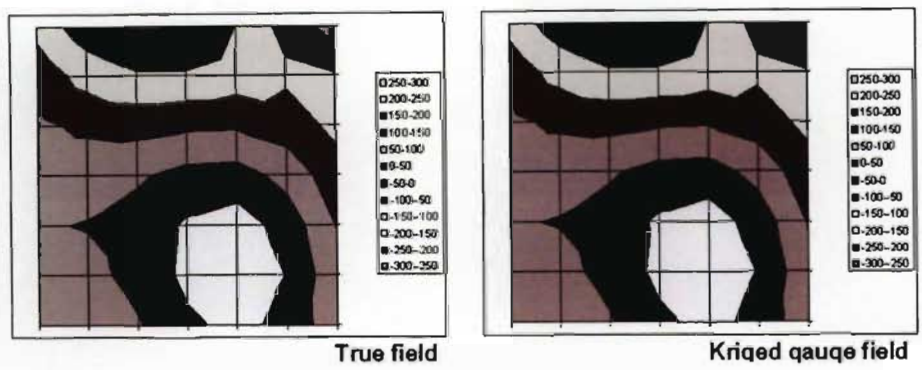


Fig. 3.26: Verification of Kriging algorithm implementation: Example fields represent a single time realization and $\mathbf{x}_t^G$ was set to be the "true" simulated rainfall field.

difference between them.

The mean and standard deviation for the Kriged residuals (the difference between $\mathbf{y}_t$ and the Kriged gauge field $\mathbf{y}_t^G$) were also computed for 2000 applications of the Kriging routine. The resulting mean over 2000 realizations on the lattice has a maximum standard deviation, about a zero mean, of 2×10^{-6} over the 49 lattice points. The negligible errors of estimation can be attributed to floating point precision errors in the computation, therefore there is a high degree of confidence that the code performs as designed by Todini (2001).

Application of Bayesian merging on the lattice

Having carefully checked the coding of the algorithm, the Bayesian merging technique was then applied on the 7×7 lattice described earlier, the mean and covariance structure of the residuals were measured *a priori*. The time series of residuals $\boldsymbol{\varepsilon}_t$ between the radar estimate of the true field $\mathbf{y}_t^R$ and the Kriged field estimate $\mathbf{y}_t^G$ was examined and the mean $\boldsymbol{\mu}_{\boldsymbol{\varepsilon}_t}$ and covariance $\mathbf{V}_{\boldsymbol{\varepsilon}_t}$ computed from these residuals. The Kalman filter equations were then applied with

$$\boldsymbol{\mu}_{\boldsymbol{\varepsilon}_t^R} = \boldsymbol{\mu}_{\boldsymbol{\varepsilon}_t}$$

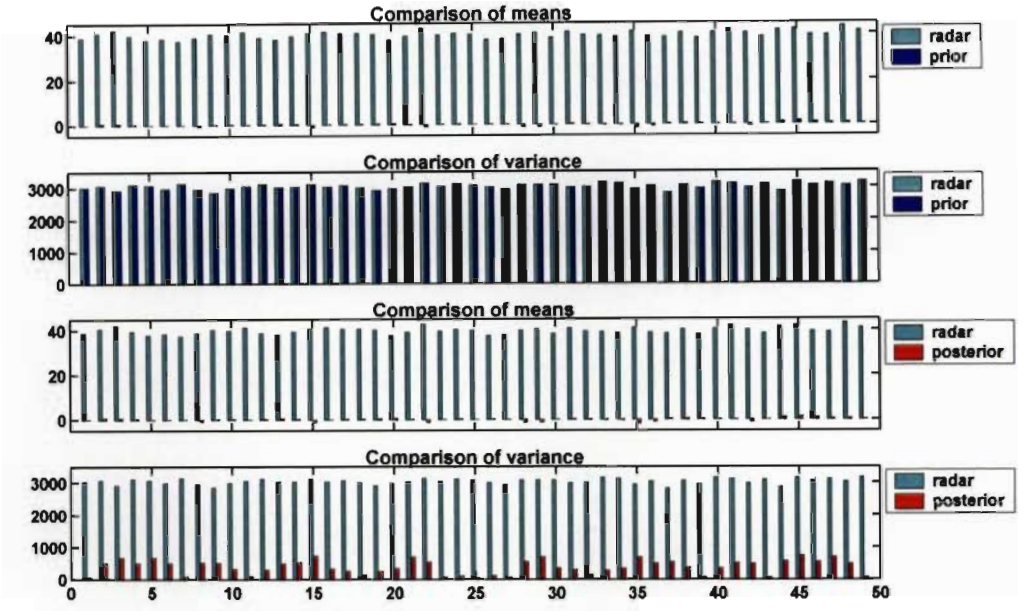


Fig. 3.27: Comparison of statistics of residuals. The entire lattice was used in the calculation of $\mathbf{y}_t''$. Note the significant reduction in the *posterior* error variance.

since the Kriged estimate $\mathbf{y}_t^G$ is assumed unbiased. The known covariance structure (see table 3.2) was used for $\mathbf{V}_{\boldsymbol{\epsilon}_t^G}$. The results for 2000 realizations on the lattice are summarized in figures 3.27 and 3.28. From these figures it is shown that the mean errors become close to zero and that the variance of their errors shows a significant reduction in magnitude. Figures 3.27 and 3.28 qualitatively match (the random number sequences are different) those presented by Todini (2001) very well, verifying and corroborating the implementation of the Bayesian merging technique.

3.4.3 Testing the moving window implementation

The computationally efficient moving window method suggested in section 3.4.1 was tested on the same set of 2000 random fields to compare the results to those achieved by the full field Bayesian merging technique. The results are presented and compared in terms of bias and variance reduction from the noisy "radar" estimates $\mathbf{y}_t^R$ of the simulated rainfall field $\mathbf{y}_t$.

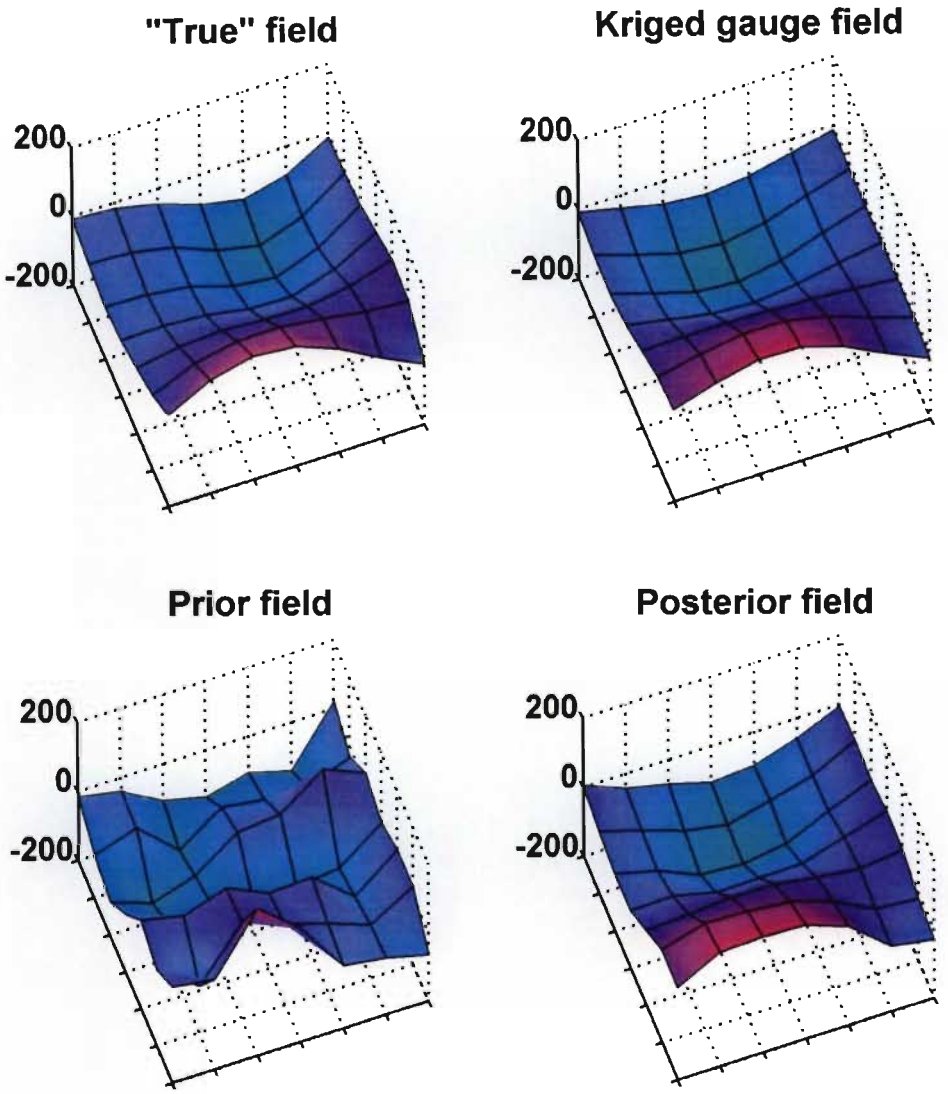


Fig. 3.28: 3D view of fields representing a single time realization. The entire lattice was used in the calculation of y''_t.

Figures 3.29 and 3.30 show the equivalent comparisons to figures 3.27 and 3.28 for a single realization on the lattice. The results are similar to those presented in the previous figures.

Figure 3.31 compares the statistics of the posterior residuals for the two meth-

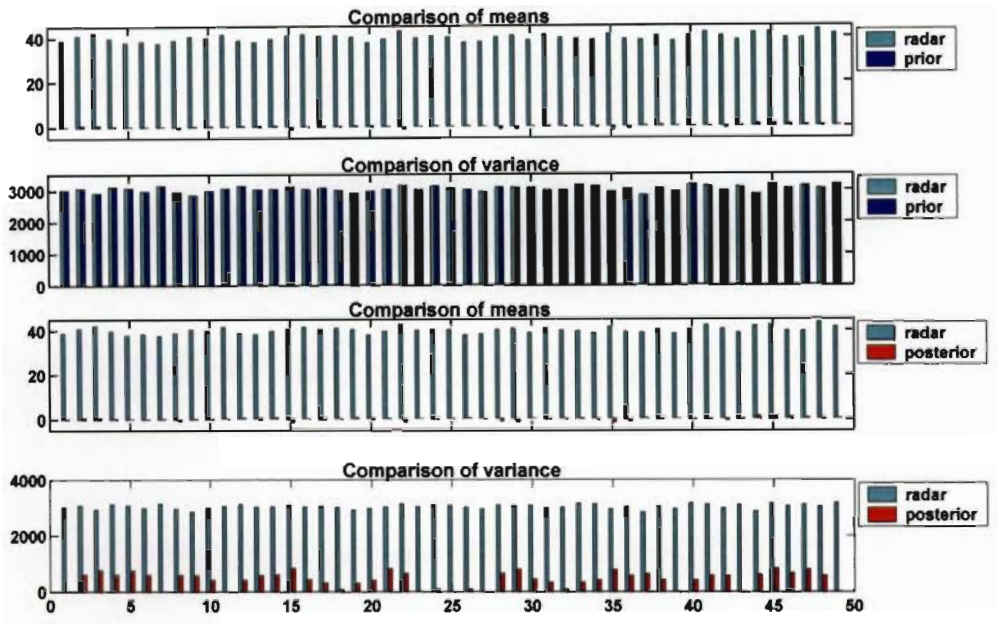


Fig. 3.29: Comparison of statistics of residuals. The *a posteriori* estimate of the rainfall field $\mathbf{y}_t''$ was computed using a pixel-wise approximation of the Bayesian merging algorithm.

ods. The figure shows that the means on each lattice square are identical (this is expected since the bias reduction step is identical) while the variance on each lattice point is very similar for each method, with neither method appearing to have a clear advantage.

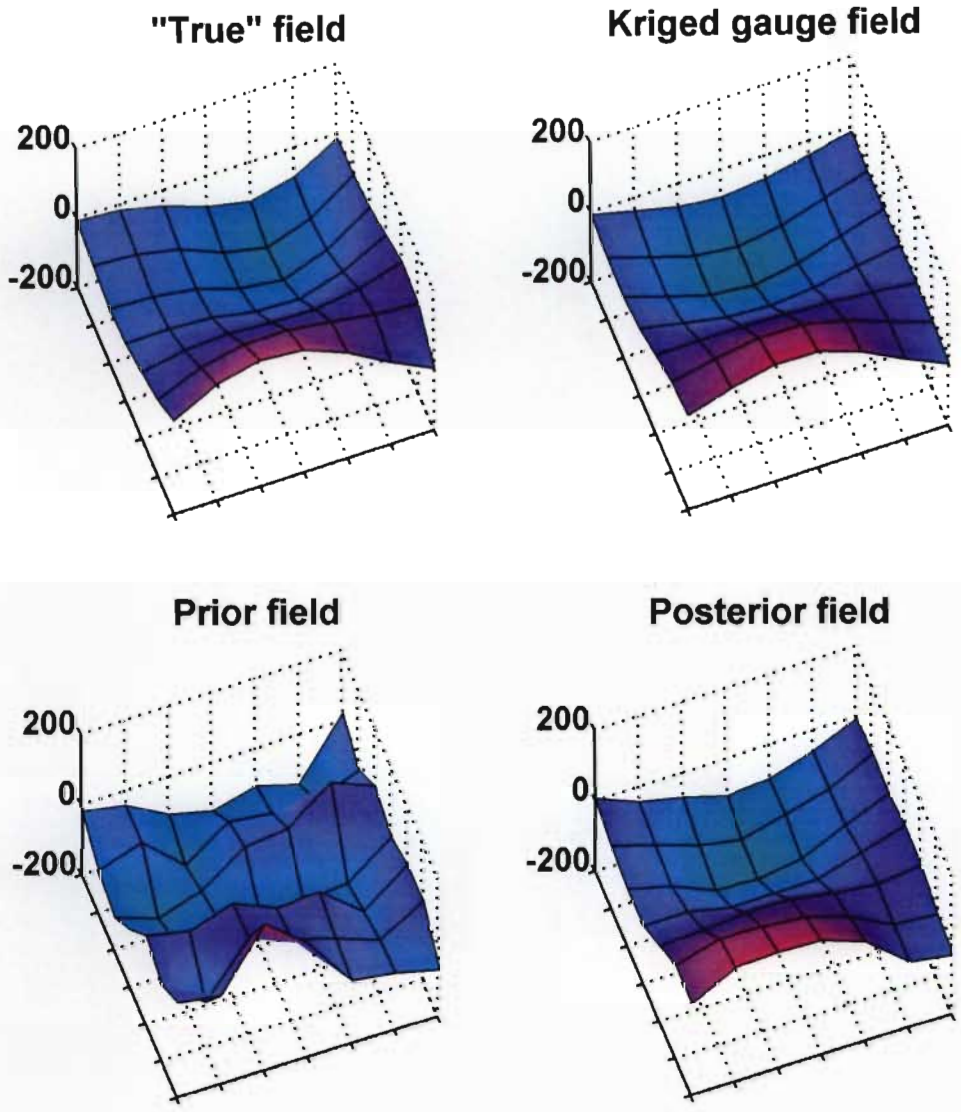


Fig. 3.30: 3D view of fields representing a single time realization. The *a posteriori* estimate of the rainfall field y_t'' was computed using a pixel-wise approximation of the Bayesian merging algorithm.

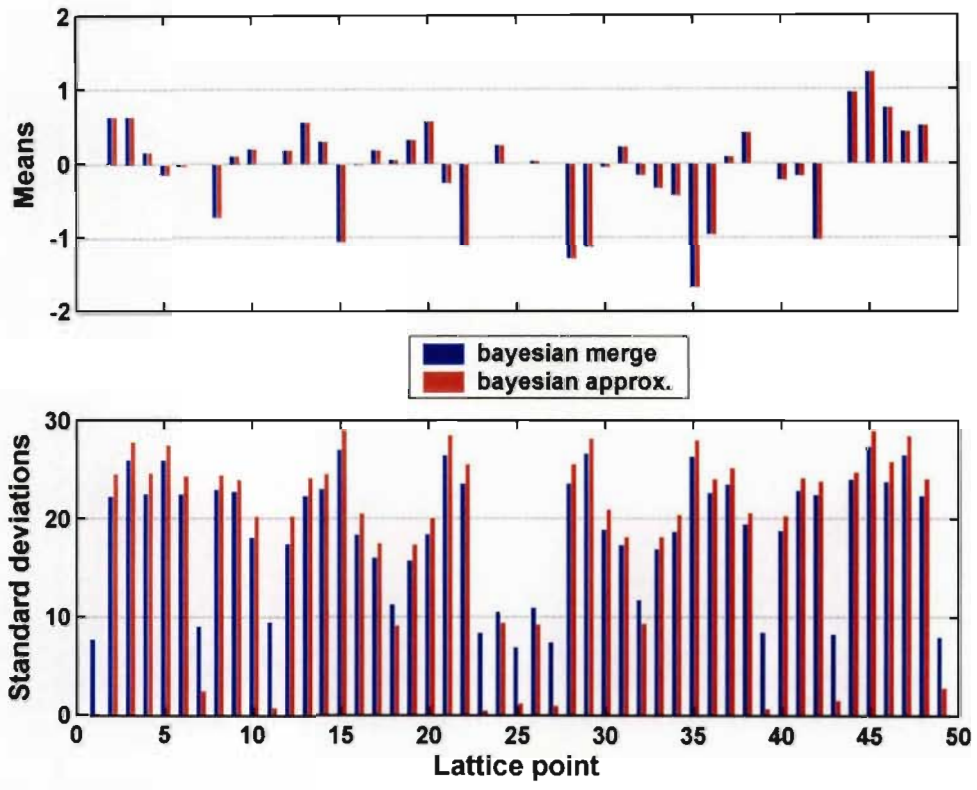


Fig. 3.31: Comparison of Posterior residuals computed using Todini's (2001) full method and the pixel-by-pixel approximation.

3.5 Conditional merging

In Sinclair and Pegram (2004a) a comparison was presented between the Bayesian merging technique of Todini (2001) and the conditional merging technique described in this section. Later Mazzetti (2004) used a small-scale numerical experiment to show that the Bayesian merging technique outperformed several others (Brandes, 1975; Krajewski, 1987; Koistinen and Puhakka, 1981). The experiment has been discussed in sections 3.4.1, 3.4.2 and 3.4.3. Sinclair and Pegram (2004a) showed that conditional merging performed competitively with Bayesian merging in a comparison experiment (the results of that study are repeated here in section 3.6.1). Further refinement of the conditional merging technique and case studies using observed data from South Africa and Spain are presented in sections 3.6.2 and 3.6.3.

Weather radar produces an observation of the unknown rainfall field that is subject to several well-known sources of error (e.g Wilson and Brandes, 1979; Chumchean et al., 2003) but which retains the general covariance structure of the true precipitation field. The spatial information from the radar can be used to condition the spatially limited information obtained by interpolating between rain gauges and produce an estimate of the rainfall field that contains the appropriate spatial structure while being constrained to match the rain gauge data (where it is available). The conditional merging technique of Ehret (2002) makes use of ordinary Kriging (e.g Journel and Huijbregts, 1978; Cressie, 1991) to extract the unbiased minimum variance rainfall estimate, given the information content of the observed data. Figure 3.32 gives an overview of the technique (for the one dimensional case) which is adapted from Ehret's (2002) work.

The spatial structure of the merged rainfall field is obtained from the radar while the rainfall values are "stitched down" to the gauge observations of the true rainfall field. The approach taken here is similar to the technique of conditional simulation by Kriging discussed in Chilès and Delfiner (1999, pp 452). However, the key difference in this case is that the radar rainfall estimate is not simply a simulation, unrelated (except by it's statistical properties) to the rainfall field to be estimated, but is in fact an independent observation of the true unknown

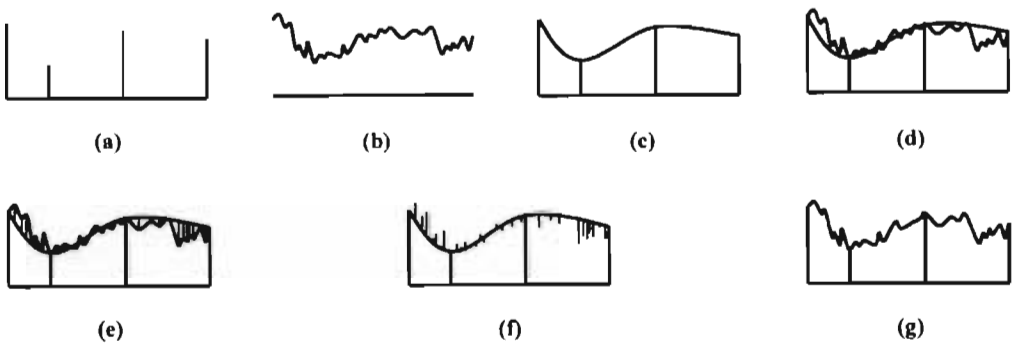


Fig. 3.32: One dimensional overview of the Conditional Merging algorithm. *(a)* The rainfall field is observed at discrete points by rain gauges. *(b)* The rainfall field is also observed by radar on a regular, volume-integrated grid. *(c)* Kriging of the rain gauge observations is used to obtain the best linear unbiased estimate of rainfall on the radar grid. *(d)* The radar pixel values at the rain gauge locations are interpolated onto the radar grid using Kriging. *(e)* At each grid point, the deviation $\varepsilon_R(s)$ between the observed and interpolated radar value is computed. *(f)* The field of deviations obtained from (e) is applied to the interpolated rainfall field obtained from Kriging the rain gauge observations. *(g)* A rainfall field that follows the mean field of the rain gauge interpolation, while preserving the mean field deviations and the spatial structure of the radar field is obtained.

rainfall field. This important link means that the radar data provides an estimate of the Kriging error (a concept which is distinct from the Kriging variance) and in particular the spatial structure of this error.

The rainfall field is observed at discrete points by rain gauges and is also observed by radar on a regular, volume-integrated grid (Figure 3.32a, b). Kriging of the rain gauge observations is used to obtain the best linear unbiased estimate of rainfall on the radar grid (Figure 3.32c). This can be described by

$$Z(s) = G_K(s) + \varepsilon_G(s) \tag{3.16}$$

where $Z(s)$ is the true (unknown) rainfall field at location s and $G_K(s)$ is the Kriged (unbiased field) estimate of $Z(s)$ from the rain gauge data. The term $\varepsilon_G(s)$ in equation 3.16 cannot be estimated (except at the gauges where it is exactly zero – if we assume for simplicity that the gauges have no error in estimating $Z(s)$)

since $Z(s)$ is unknown. The radar pixel values at the rain gauge locations are interpolated onto the radar grid using Kriging (Figure 3.32d), this operation may be described as

$$R(s) = R_K(s) + \varepsilon_R(s) \tag{3.17}$$

where $R(s)$ is the measured radar rainfall estimate and $R_K(s)$ is the Kriged (unbiased field) estimate of $R(s)$ using the radar values at rain gauge locations. At each grid point, $\varepsilon_R(s)$ the deviation between the observed and interpolated radar value (Figure 3.32e) is computed using equation 3.17. On the basis that $R(s)$ is a measurement of $Z(s)$, the field of deviations $\varepsilon_R(s)$ is applied to $G_K(s)$ the interpolated rainfall field obtained from Kriging the rain gauge observations (Figure 3.32f) to produce $M(s)$, the merged estimate of $Z(s)$

$$M(s) = G_K(s) + \varepsilon_R(s) \tag{3.18}$$

Equation 3.18 results in a rainfall field that follows the mean field of the rain gauge interpolation, while preserving the mean field deviations and the spatial structure of the radar field (Figure 3.32g).

The expected value of the error between the merged estimate and the true field is zero if the fields are Gaussian, since the Kriged estimates are unbiased in this case, i.e.

$$E\left[Z(s) - M(s)\right] = E\left[\varepsilon_G(s) - \varepsilon_R(s)\right] \tag{3.19}$$

and $E\left[\varepsilon_G(s)\right] = 0$, $E\left[\varepsilon_R(s)\right] = 0$. The variance of the merged estimation errors can be examined by considering

$$\begin{aligned} var\left[Z(s) - M(s)\right] &= var\left[\varepsilon_G(s) - \varepsilon_R(s)\right] && (3.20)\\ &= \sigma^2_{\varepsilon_G(s)} + \sigma^2_{\varepsilon_R(s)} - 2cov\left[\varepsilon_G(s), \varepsilon_R(s)\right] && (3.21)\\ &= \beta - 2\sigma_{\varepsilon_G(s)}\sigma_{\varepsilon_R(s)}\rho && (3.22) \end{aligned}$$

where $\beta = \sigma^2_{\varepsilon_G(s)} + \sigma^2_{\varepsilon_R(s)}$

The variance of the error estimate given by equation 3.20 can be decomposed as shown in equations 3.21 and 3.22. The variance of the error is (trivially) zero

at the gauged points, while at any other position in the field it is bounded by a maximum value of β for positive correlations ρ between the errors $\varepsilon_G(s)$ and $\varepsilon_R(s)$. If $\varepsilon_G(s)$ and $\varepsilon_R(s)$ are strongly (positively) correlated as one would expect since both gauges and radar are measurements of $Z(s)$, then the variance of the error will be significantly less than β as suggested in equation 3.22. If $\sigma_{\varepsilon_G(s)} = \sigma_{\varepsilon_R(s)} = \sigma$ and $\rho > 0.5$, then $var\,[Z(s) - M(s)] = 2\sigma^2[1 - \rho]$ showing that the variance of the error between the merged and real rainfall field will be less than that of either field (Kriged gauge or radar estimate) alone. If $\rho = 0$ (the errors are uncorrelated), then the error is the sum of the errors of the individual fields as expected.

3.6 Comparison and application of merging methods

In this section, three case studies are presented. The first compares the relative performance of the Bayesian merging algorithm (Section 3.4) and the conditional merging algorithm (Section 3.5) using simulated rainfall fields. The experiment and results of the comparison are presented in section 3.6.1. In section 3.6.2 the conditional merging method is used to estimate daily rainfall on the Liebenbergsvlei catchment in South Africa. The quality of the estimated rainfall fields is assessed by comparing the merged fields to rain gauges, using a cross-validation procedure. Lastly, a similar study is carried out using storm rainfall from a radar site in Catalunya, Spain (Section 3.6.3).

3.6.1 Case study - Comparing the Bayesian and conditional merging techniques using simulated rainfall fields

The experiments on small (7×7) lattices reported in sections 3.4.1, 3.4.2 and 3.4.3 were not thought to be particularly representative of realistic rainfall fields. To test the performance of the merging techniques the experiment presented here was devised, because it is impossible to measure a real rainfall field to test the

algorithms performance in an objective way.

A sequence of 1000 independent 128×128 pixel rainfall fields was produced using the "String of Beads" rainfall simulation model (Pegram and Clothier, 2001). These rainfall fields were treated as the "true" rainfall field and "observed" radar estimates, produced by adding bias and noise. Each pixel in the field represents an area of size 1×1 km. The "true" field was sampled at 83 "rain gauge" locations chosen randomly on the pixel grid. This gives an average coverage of one gauge for every 198 km^2, representing a gauge spacing of 14 km, which is a fairly dense but realistic network (although in South Africa this gauge density is very seldom realized). The "rain gauge" measurements were assumed to be without error. A single realization of the "true" field as well as the corresponding "observed" radar estimate, the location of the gauges and the Kriged gauge estimate of the field are shown in figure 3.33, while a more detailed description of the experiment follows.

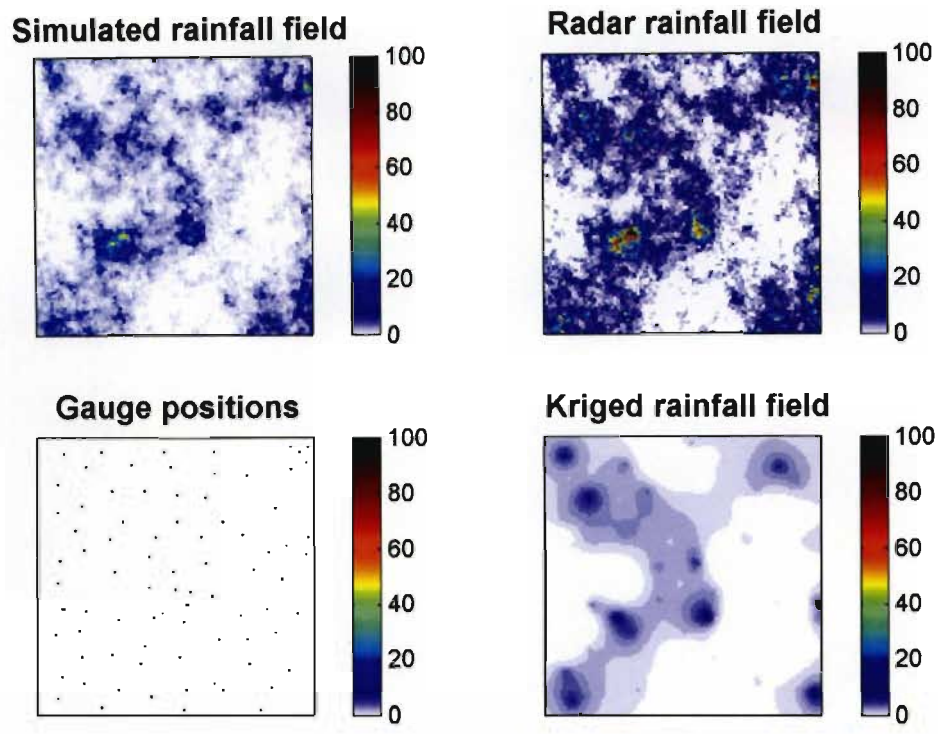
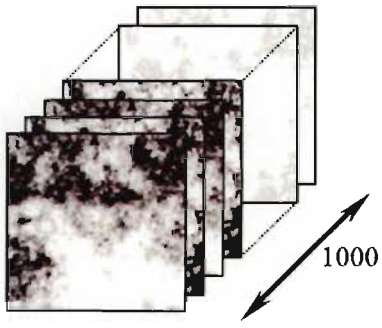
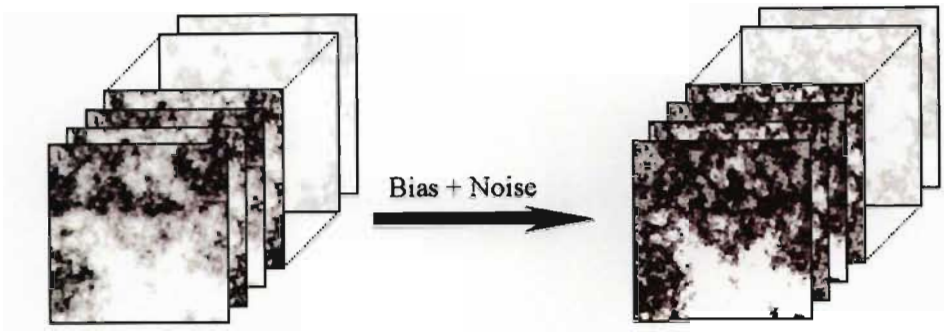


Fig. 3.33: A single realization of the modelled instantaneous rainfall fields (mm/hr) and rain gauge locations. Note that the Kriged estimate captures the general structure quite well due to the high sampling frequency, but fails to represent the finer details.

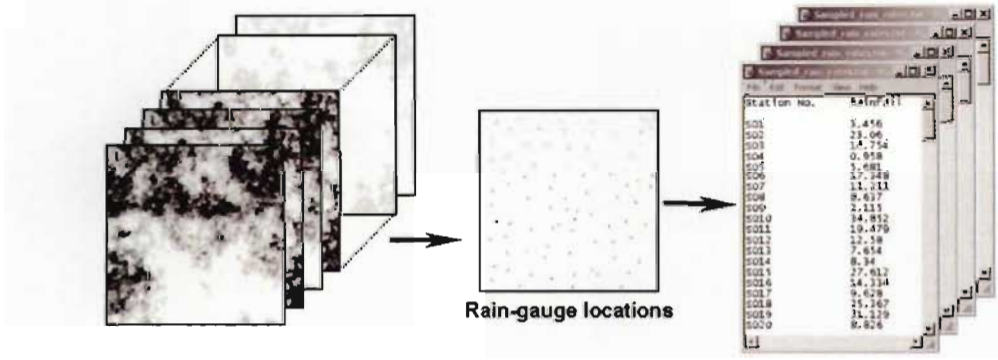
i) Generate 1000 independent rainfall fields using the "String of Beads" model.



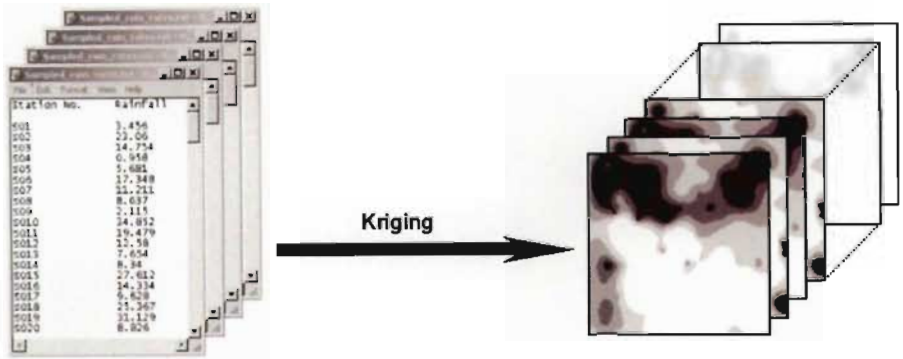
ii) Add bias and noise to simulate radar measurements of the "true" rainfall field.



iii) Sample the "true" rainfall field (at 83 random locations) to get a set of unbiased and error free rain gauge observations.



iv) Krige the rain gauge observations onto the unobserved radar pixels.



v) Apply the merging procedure to estimate the original field on the basis of the "observed" gauge and radar estimates.

vi) Compute the mean error $\bar{\epsilon}_i$ $(i = 1, 2, ..., m)$ at each pixel over the 1000 realizations.

$$\bar{\epsilon}_i = \frac{1}{N} \sum_{k=1}^{N} y_i(k) - m_i(k)$$

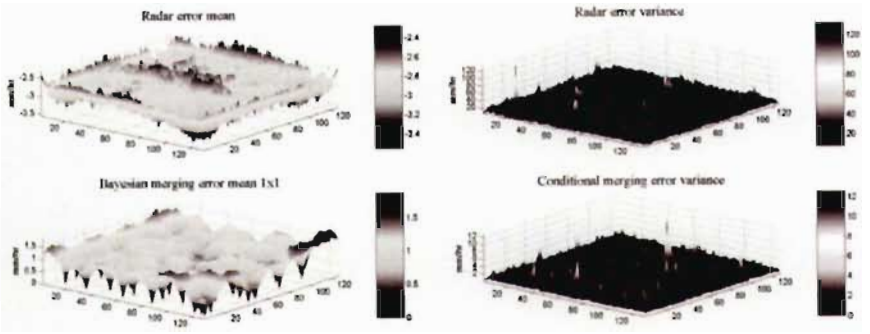
where $y_i(k)$ is the "true" value of the rainfall field at pixel i for realization k and $m_i(k)$ is the merged estimate of the "true" value of the rainfall field at pixel i for realization k.

vii) Compute the variance of the errors at each pixel over the 1000 realizations.

$$\sigma_i^2 = \frac{1}{N} \sum_{k=1}^{N} \{y_i(k) - m_i(k) - \epsilon_i(k)\}^2$$

where σ_i^2 is the variance of error between the "true" value of the rainfall field $y_i(k)$ and $m_i(k)$, the merged estimate of the "true" value at pixel i. $\epsilon_i(k)$ is the error at pixel i for realization k.

viii) Compare the mean and variance of the estimate's deviations from the "true" field, computed from steps vi) and vii) to those computed from the radar errors in order to quantify the improvement (relative to the radar estimate) gained from each merging technique.



A single realization of the true field, radar observation together with the estimates via Bayesian and conditional merging is shown in figure 3.34.

Figure 3.35 shows histograms of the mean errors on each field. The radar error mean was computed by taking the mean value (over 1000 realizations) of the residuals between the "true" field and the radar observation at each pixel. This results in the 16384 values plotted as a histogram. The mean errors for each merging technique were computed in the same way. It is clear from the figure that both techniques provide an improvement over the radar observation and that the Conditional merging technique performs somewhat better in terms of bias reduction. The spike of values at zero in each case is due to the 83 rain gauge observations that are (as noted earlier) without any measurement bias.

The average variance of the errors at each pixel is reported in histogram form in figure 3.36. Once again both merging techniques give a considerable improvement relative to the "radar", with the conditional merging technique once again performing better in terms of variance reduction.

All of the Kriging and merging computations were done on the logarithms of the variables in order to transform the log-normally distributed rainfall rate simulations to a Gaussian space. Comparisons, however, were done on the back transformed variables in rainfall rates (mm/hr).

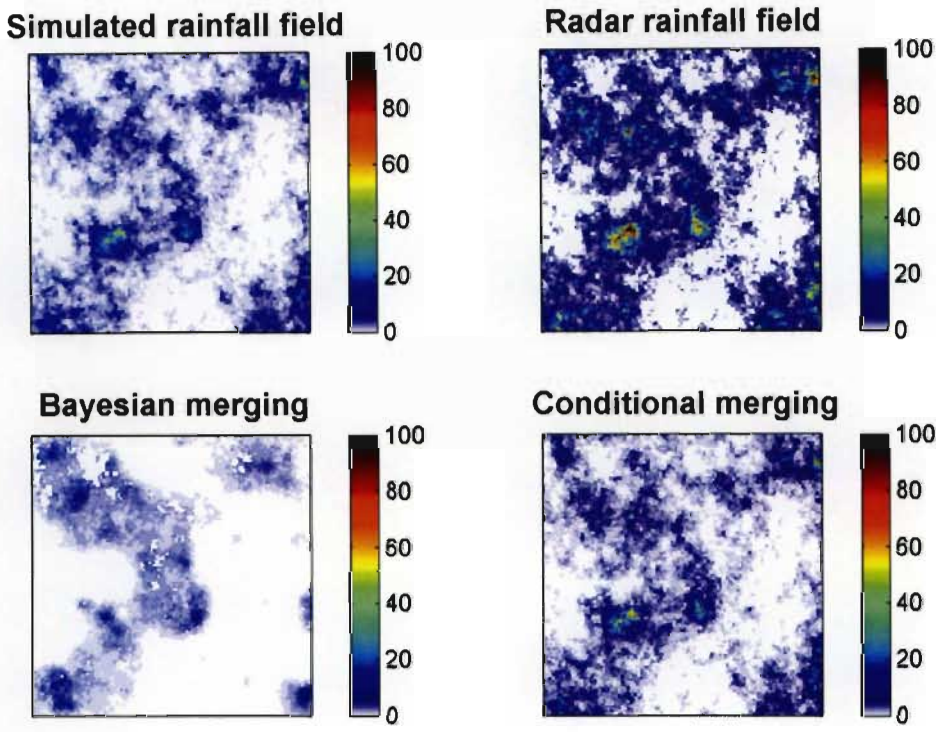


Fig. 3.34: Comparison of Bayesian and Conditional merged fields for a single realization. The estimated rainfall field produced by Bayesian merging appears very similar to the Kriged estimate shown in the bottom right field in figure 3.33, whereas the Conditionally merged field more closely resembles the "truth" in the upper left simulated rainfall field. The reason for this is that the Bayesian merging algorithm uses the Kriged gauge estimate as the *a posteriori* observation in it's Kalman filter.

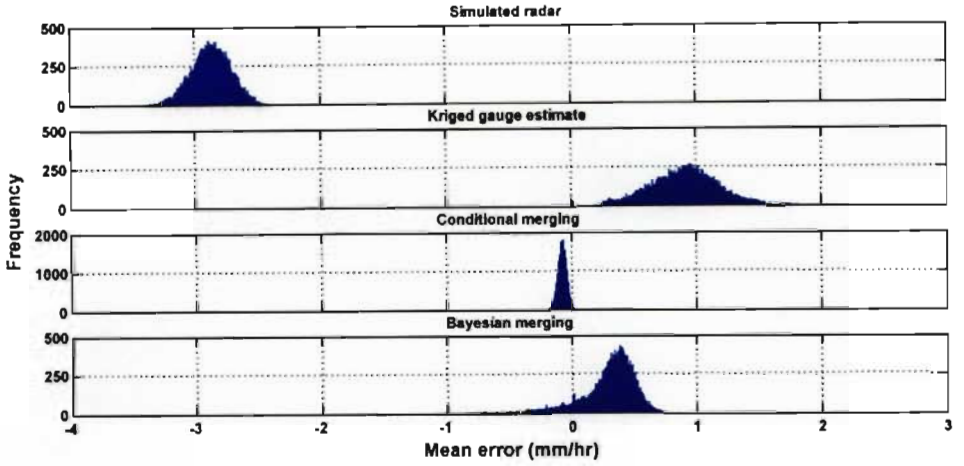


Fig. 3.35: Histograms of mean pixel errors relative to the simulated rainfall field. The error is defined as $\epsilon_i(k)$, the difference between the simulated and estimated fields at pixel i for realization k. The average over all the realizations at each pixel was computed and the resulting 16384 average values for the entire field are plotted here as histograms.

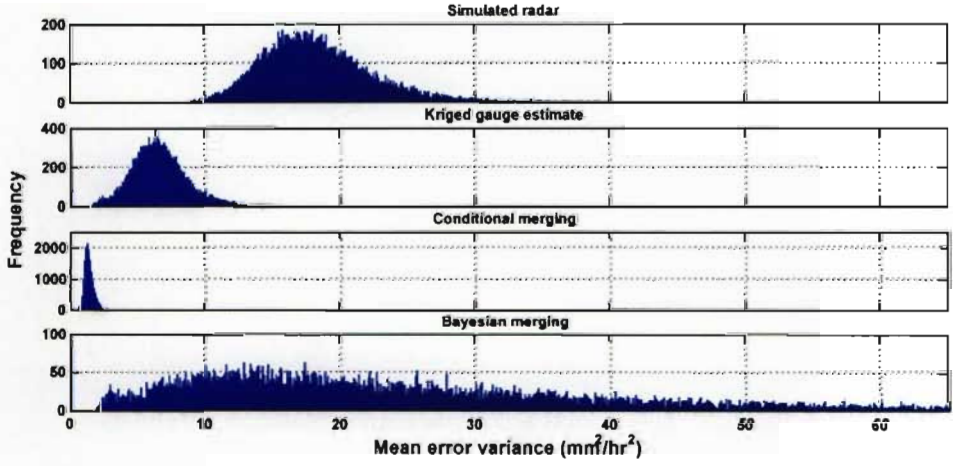


Fig. 3.36: Histograms of the variance of the mean pixel errors relative to the simulated rainfall field. The error is defined as for figure 3.35. The average over all realizations at each pixel was computed and the resulting 16384 average values for the entire field are plotted here as histograms.

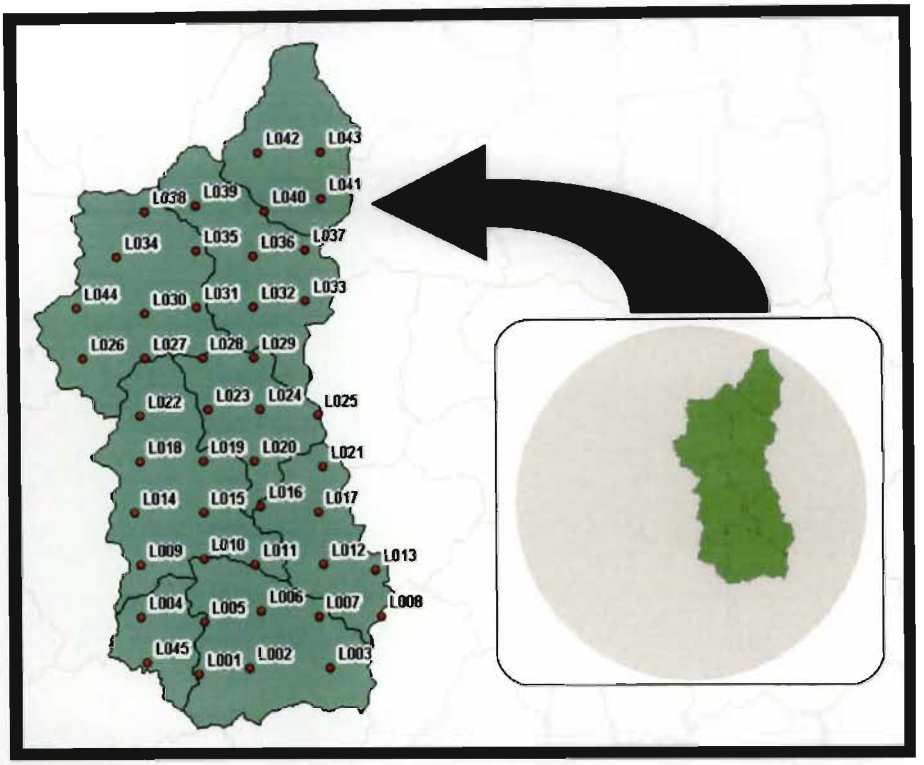


Fig. 3.37: The Liebenbergsvlei catchment superimposed on the MRL5 radar's 100 km radius footprint. The rain gauge network and numbering system are shown by the red dots in the expanded view of the catchment.

3.6.2 Case study - Conditional Merging, Liebenbergsvlei, South Africa

The conditional merging algorithm was applied to radar and tipping bucket rain gauge data from the Liebenbergsvlei catchment in South Africa. The data analysed consisted of several daily accumulations from the South African Weather Services MRL5 radar (situated approximately 5 km from the catchment at it's closest point) and daily totals obtained from a network of 45 tipping bucket rain gauges. The gauge network and radar coverage are shown in figure 3.37.

The radar collects its data in volume scans at 5 minute intervals. A three dimensional Kriging algorithm, described in Wesson (2005), as well as Wesson and Pegram (2006), was used to produce a ground level estimate of rainfall rate

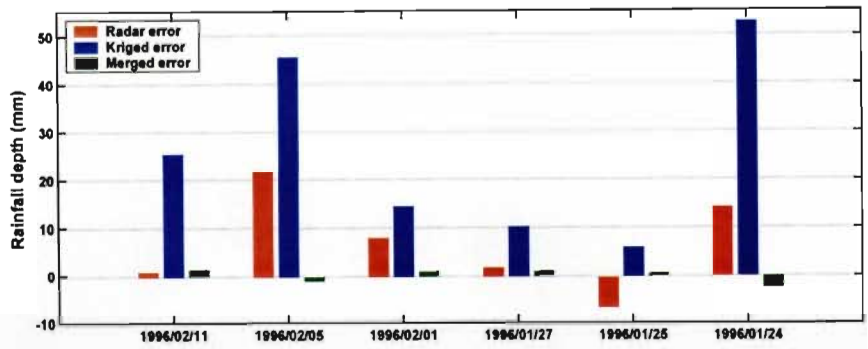


Fig. 3.38: Comparison of the cross validation errors for daily rainfall totals. The histogram shows the error in rainfall depth for various rainfall estimates, relative to the tipping bucket gauge totals. The mean error over all of the gauges is shown in each case.

from each volume scan. These rainfall estimates were accumulated over a number of 24 hour periods using the algorithm described in section 3.3. Daily totals for the same 24 hour periods were computed from the tipping bucket rain gauge data.

Each of the 45 rain gauge totals was compared with the radar rainfall pixel covering it. To try and determine whether the merging technique gave an improved estimate, relative to Kriging using the gauges or direct use of the radar rainfall estimates, a cross-validation technique was used. The cross-validation was carried out by repeatedly "hiding" one of the gauges and computing the estimated rainfall field (either by Kriging or using conditional merging) with the remaining gauges. The hidden gauge value was then directly compared with the estimated pixel value covering the gauge. This was repeated until each of the 45 gauges had been treated in turn. The results shown in figure 3.38 ignore the point to area conversion error because it was not considered to play a significant role with a small pixel size of 1×1 km. The conclusion is that conditional merging is a practical method to apply in this field.

A second investigation was performed, which takes account of the change of support issues. A 3×3 km block was selected, surrounding each of the gauges. A block Kriged rainfall estimate was computed over this region using the informa-

tion from the gauges, where the block Kriged estimate is defined as

$$Z_A = \frac{1}{A}\int_A \hat{z}(s)ds$$

Z_A is the block Kriged estimate over the area A and $\hat{z}(s)$ is the Kriging estimate at point s in A. The average value of the merged field is computed by hiding the relevant gauge as before and taking the average of the resulting merged estimate over the 3×3 pixel region. Clearly the Kriged gauge estimate and the merged estimate both contain information from the remaining gauges, but this was ignored on the basis that the influence of the hidden gauge would dominate the Block Kriged estimate over this relatively small region. Figure 3.39 shows scatter plots representing the comparison between merged average rainfall values and Block Kriged rainfall estimates for daily totals. It is clear that there is relatively good correlation between the estimates, although bias is exhibited on 10/02/1996 in the bottom left panel.

3.6.3 Case study - Conditional Merging, Catalunyan radar data set

The conditional merging technique described in Sinclair and Pegram (2005a) [see second paper in the Appendix] was applied to a data set provided by Professor Daniel Sempere-Torres and Carlos Velasco from GRAHI in Catalunya, Spain, an experiment which arose from a conversation between Professors Sempere-Torres and Pegram at the Symposium on Hydrological Applications of Weather Radar in Melbourne, 2004. The data set consists of twenty co-located rainfall rate estimates for rain gauge and radar, measured at 10-minute intervals for the duration of a 13.5-hour rainfall event. The event began at 17:00 GMT on 28 September 2000 and ended at 06:30 GMT on 29 September 2000. The radar data consist of the 1×1 km radar pixel above each rain gauge and its 8 surrounding neighbours as indicated schematically in figure 3.40. The pixel labelled 0 is the radar pixel which has it's centre point located nearest to the rain gauge, the numbering scheme is relevant as it is used in the figures which follow.

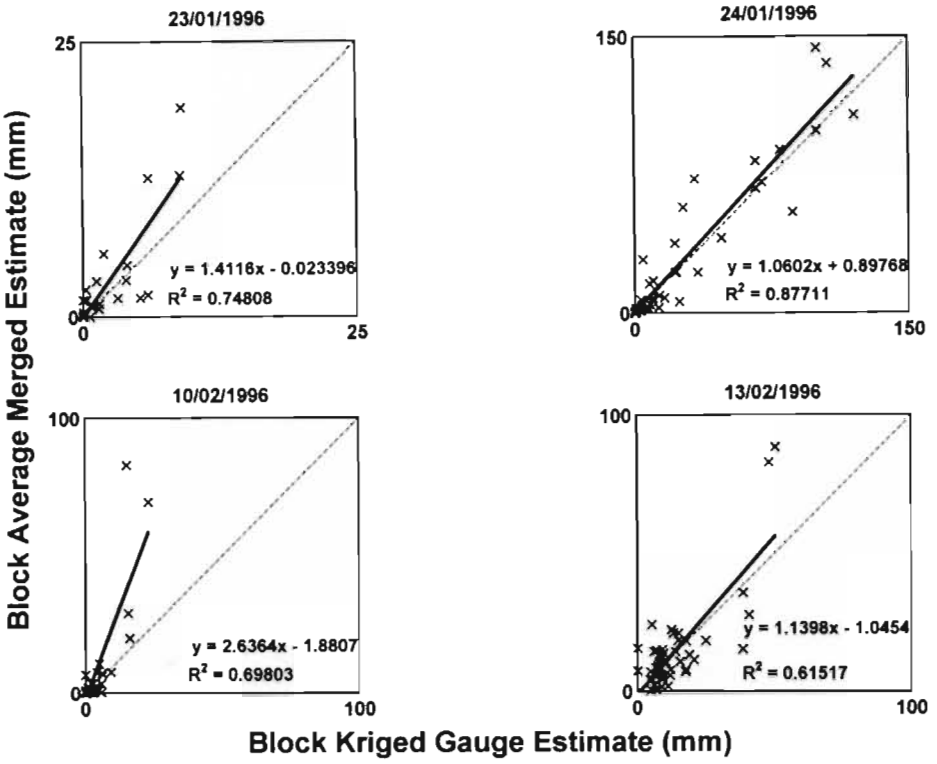


Fig. 3.39: Scatter plots comparing block Kriged gauge estimates and Conditionally merged average values over the 3×3 km regions surrounding each gauge.

The first step in the data analysis was to compute accumulated event totals for each of the gauges. The event total rainfall depth was also computed for the 9 radar pixels surrounding each gauge (for the pixel layout see figure 3.40). A simple accumulation was made which does not account for the effects of advection and sampling frequency in the radar estimations. The Optical Flow accumulation scheme described in section 3.3 could not be implemented due to the limited portions of radar data available. Figure 3.41 shows a sample of the gauge and radar accumulations, for 4 of the 20 gauges examined. In this figure and those following, the single rain gauge value is plotted as a constant line against each of the pixels in it's vicinity. In most cases the radar under-estimated the rainfall accumulation measured by the gauges. There were only 5 cases (out of 20) in which the radar rainfall estimates were higher than the gauge estimate.

6	5	4
7	0	3
8	1	2

Fig. 3.40: Schematic local cut-out of the Catalunyan radar data, surrounding each of the 20 rain gauges. The pixel labelled zero has its centre point closest to the rain gauge being considered.

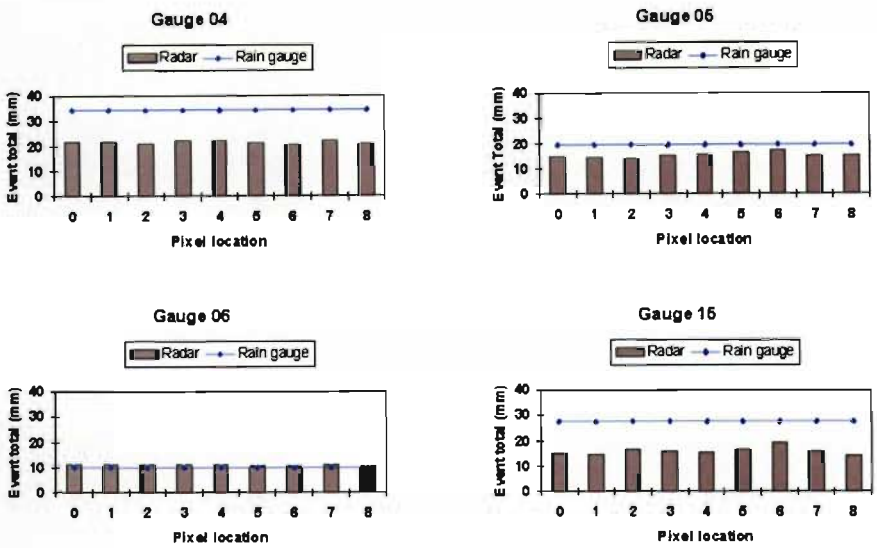


Fig. 3.41: Comparison of the rainfall accumulations over 13.5 hours for a number of the gauges and the radar rainfall estimates from the 9 surrounding pixels, centered on the gauges.

A cross-validation experiment was carried out to compare the rainfall estimates obtained by *i)* Kriging the rain gauge data, *ii)* Conditional Merging and *iii)* the raw radar estimates. Figures 3.42 and 3.43 show a sample of comparisons between gauge rainfall and the estimated rainfall at surrounding pixels computed by Kriging and conditional merging when the gauge in question is excluded from the computation. The main purpose of showing these figures is to highlight the

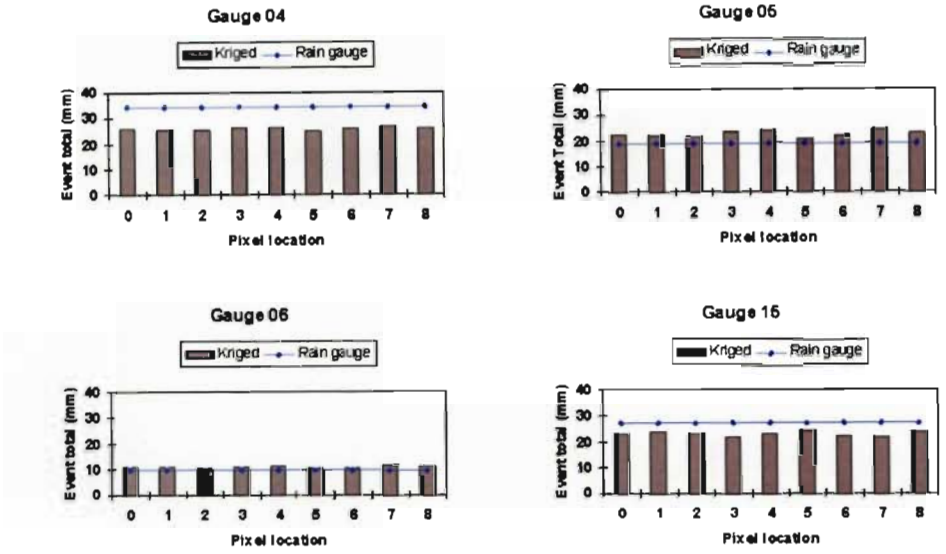


Fig. 3.42: Catalunyan gauge and Kriged rainfall estimate cross-validation comparison. The gauge accumulations are compared with the Kriged estimate at the gauges computed by leaving out the gauge in question and using the remaining 19 gauges.

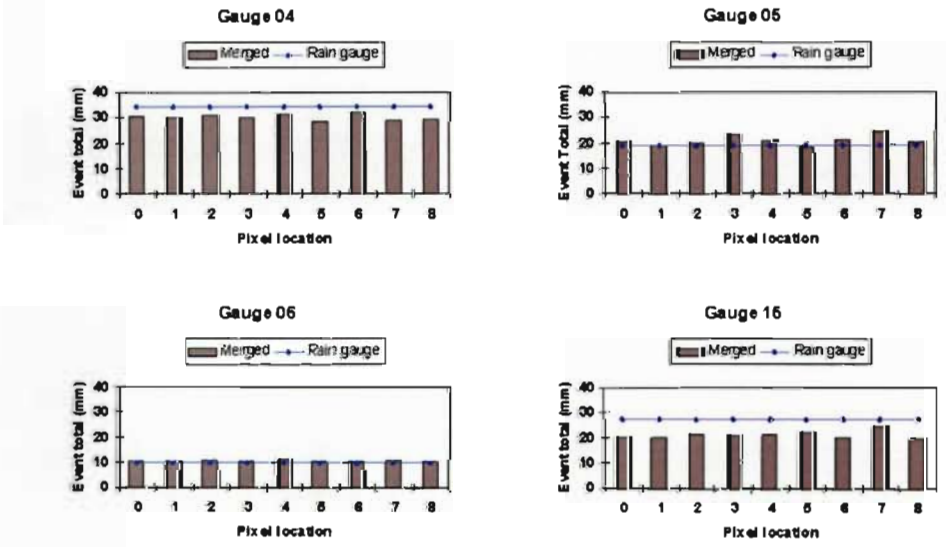


Fig. 3.43: Catalunyan gauge and Conditionally merged rainfall estimate cross-validation comparison. The gauge accumulations are compared with the merged estimate at the gauges computed by leaving out the gauge in question and using the remaining 19 gauges.

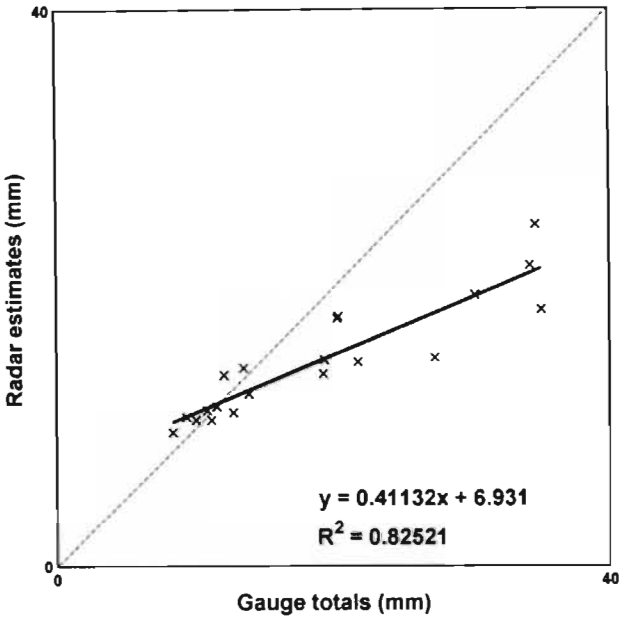


Fig. 3.44: Scatter plots comparing the gauge and radar rainfall estimates for the Catalunyan data set. The gauge totals are compared to the estimate from the radar pixel which is located directly above the gauge in question.

lack of variation in the radar rainfall estimates over the 3 × 3 km block of radar data. Figures 3.44, 3.45 and 3.46 summarize the results of this comparison. The merged estimates clearly sit closer to a 1:1 relationship despite a slightly smaller R^2 value (indicating a wider scatter about their best fit line) than the raw radar estimates. The Kriged estimates do not perform as well as the merged estimates for this event and exhibit a wide scatter about their "best fit" line as indicated by the low R^2 value of 0.49, compared to 0.68 for the latter. The radar is expected to be better correlated to the gauges even though biased. Figure 3.44 should not be directly compared to the cross-validation figures following it as it provides a measure of the strength of ρ – see the discussion in section 3.5.

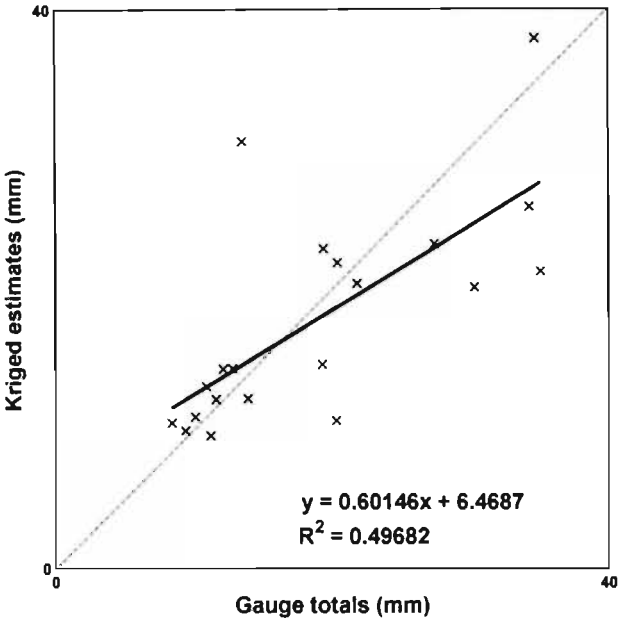


Fig. 3.45: Catalunyan gauge and Kriged rainfall estimate cross-validation scatter plot. The gauge accumulations are compared with the Kriged estimate at the gauges computed by leaving out the gauge in question and using the remaining 19 gauges.

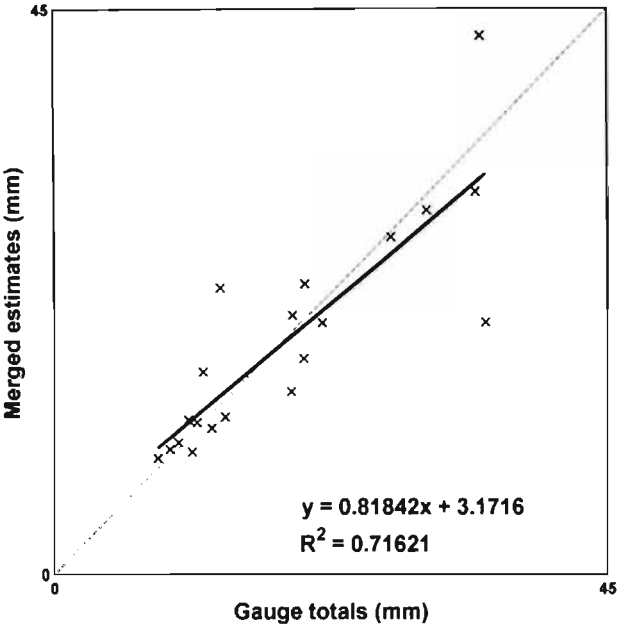


Fig. 3.46: Catalunyan gauge and merged rainfall estimate cross-validation scatter plot. The gauge accumulations are compared with the merged estimate at the gauges computed by leaving out the gauge in question and using the remaining 19 gauges.

Chapter Summary: In this chapter, the three sources of real-time rainfall measurements available (at present) for use in South African flood forecasting systems have been discussed. The advantages and disadvantages of each have been presented and a distinction drawn between direct and indirect methods of rainfall measurement using these instruments.

The concept of blending different rainfall estimates to produce an optimum estimate of the unknown rainfall field has been introduced and a selection of the extensive literature on this subject has been reviewed.

A technique for accumulating instantaneous spatial rainfall estimates has been presented that allows a direct comparison to gauge based accumulations. The algorithm takes account of the field advection between sampling intervals. The resulting accumulated fields are then used as input to the conditional merging algorithm, which has been tested using cross–validation techniques.

The Bayesian merging algorithm (Todini, 2001) has been described and an implementation coded and tested. The algorithm was found to produce results comparable to those published by Todini (2001), providing a measure of confidence in the implementation.

In addition, the Conditional merging algorithm (Ehret, 2002) has been revisited and it's error structure investigated.

Using a large number of simulated rainfall fields, the Conditional Merging algorithm was compared with Bayesian merging and found to be the most competitive in terms of bias and variance reduction on this simulated data set.

The Conditional merging algorithm has also been tested on observed data sets from South Africa and Spain, and was found to perform well in cross validation experiments. On this basis the Conditional Merging algorithm is proposed as the technique to employ for producing a best estimate of distributed rainfall for flood forecasting systems in South Africa.

CHAPTER 4

RAINFALL FIELD NOWCASTING

Forecasts of rainfall hold promise for extending the lead time of streamflow forecasts. Although Numerical Weather Prediction (NWP) models offer the possibility of significant forecast lead time, they suffer from two serious limitations. First, the spatial resolution of these models must be relatively coarse, for computational reasons; this limits their use for flash flooding which occurs by definition on small catchments. Second, the quantitative accuracy of precipitation estimates by NWP is limited. For short lead times (up to 2 hours ahead) nowcasts based on advection and stochastic models prove more useful. In this chapter a review of available nowcasting techniques is presented and critically examined. Two stochastic nowcasting models are compared, these are the S-PROG model and a nowcasting version of the SBM. The final sections of the chapter focus on a new method for examining the spatial structure of rainfall fields, on the basis that correctly separating the rainfall field into it's spatial components might improve forecast accuracy. This is shown to be a potentially fruitful strategy when comparing the SBM and S-PROG forecasting models. The method decomposes the observed data into a sequence of high through low frequency spatial components using a set of Wavelet like, data driven, decomposition bases. The technique is called Empirical Mode Decomposition (EMD) and a novel use of the two-dimensional extension of the method is presented here.

4.1 Overview of nowcasting techniques and models

Rainfall forecasts applicable to flood forecasting (ideally) span a wide range of timescales from seasonal outlooks through several days ahead, right down to nowcasts for the next few minutes or hours. In order that the longer range forecasts

are of significant use, the uncertainties associated with them need to be reduced by decreasing the spatial resolution (increasing the basic area from pixel-scale) at which the forecasts can be made. This unfortunately has the effect of reducing their usefulness for flood forecasting, even for relatively large catchments.

Rainfall forecasts can be produced using two broad categories of model, these are; physically based models and stochastic models. Physically based models attempt to directly describe the complex physics of the atmosphere, which would be an ideal solution if computing resources were sufficient to rapidly perform the computations at space and time resolutions which properly characterize the variable nature of rainfall. Global circulation models (GCM) are only able to perform operational simulations at synoptic scale spatial resolutions. Nesting limited area models (LAM) in a GCM can provide greater spatial detail up to resolutions of a few (60×60 km) kilometres (Engelbrecht et al., 2002). it is certain that future advances in computing power will allow such models to provide simulations and forecasts at much higher resolutions than is currently possible. Until such a time, the output of these models is mostly useful as a qualitative early warning for possible heavy rains which may occur over sensitive catchments. In South Africa SAWS has recently (October 2006) begun running the Unified Model (UM) with a spatial resolution of 12×12 km. The quality of the Quantitative Precipitation Forecasts (QPF) outputs from this model have not yet been ascertained.

Stochastic models try to match key statistics of rainfall conditioned on past observations; for example an application of the Neyman-Scott shot noise model presented by Cowpertwait (1991) models the arrival times and duration of rainfall events as measured by rain gauges, while the "String of Beads" model (Pegram and Clothier, 2001) models the spatial and temporal characteristics of rainfall fields measured by radar. For the reasons described in section 1.2 (advocating spatially distributed rainfall fields as input to the catchment model) a decision was taken to limit this study to models capable of modelling spatial rainfall fields rather than considering point process models.

Grecu and Krajewski (2000) compared the results obtained from persistence, advection and a nowcasting scheme based on neural networks. A large radar data

set was used and the results compared at a range of space and time resolutions. The key findings were that advection appeared to be the most important factor in determining the quality of forecast results and that integration in both space and time reduced forecast errors by reducing the variance of the data.

Harris et al. (2001) compare 3×3 km forecasts from a nested numerical model to observed radar rainfall data. The model forecasts were found to compare favourably with the observed data for large spatial scales, but the small scale variability evident in the radar data was not reproduced. Harris et al. (2001) suggest that this is a deficiency of the model for applications where the small scale variability is of importance, but acknowledge that the model forecasts would be useful in applications where the small scale variability and intermittency do not play a significant role.

Berenguer et al. (2005) show that S-PROG (Seed, 2003) rainfall forecasts are able to extend the useful lead time of streamflow forecasts compared to using persistence and simple advection based forecasts of the rainfall fields. Since S-PROG filters out small scale variability with lead time this suggests one of two things. Either small scale variability isn't important relative to areal rainfall totals for streamflow forecasting or, incorrect forecasting of the small scale variability (as is likely to occur with persistence or pure advection of the rainfall fields) is much worse than ignoring it.

Ganguly and Bras (2003) present a modelling strategy for producing precipitation forecasts using a combination of the outputs from an NWP and advection of radar observations. They show that the resulting model forecasts perform better than NWP or radar forecasts alone and suggest that NWP outputs might benefit more from the assimilation of external data (e.g. radar) than from increased spatial resolution.

Germann and Zawadzki (2002) present a methodology for quantifying the predictability of composite radar rainfall data at a continental scale. They find that the range at which rainfall features are predictable is extended by ignoring the smaller scale features in the fields and note that

"If the predictability is predominantly limited by the smallest scale

present in an image, then filtering efficiently increases the lifetime."

It is with this in mind that Empirical Mode Decomposition (EMD) based scale decomposition is studied in sections 4.2 and 4.3. Later, in the first of two follow-up papers, Germann and Zawadzki (2004) examine how the methodology can be extended to produce probabilistic forecasts of the distribution of rainfall rates at any given point in the forecast field. In a further extension of the methodology Turner et al. (2004) examine the use of optimal forecast filters which are designed to remove the most unpredictable scales such that a near-optimal forecast (in terms of the root mean squared errors) is produced.

Figure 4.1 indicates the way in which a generic model's forecast skill varies with time, relative to Eulerian (persistence) and Lagrangian (advection) forecasts as well as the long term average. The figure shows that a rainfall field forecasting model is only useful if there is a period of time for which the model forecast skill is greater than the skill achieved by Lagrangian forecasting of the most recent observed fields or simply using the estimates obtained from the long term mean. Clearly it is prudent to consider more than one model as different types of model will exhibit different levels of skill at any given lead time (e.g. NWP may be most appropriate for larger lead-times). The most successful approach to forecasting rainfall is therefore to choose a different method or model as a function of the forecast lead-time, in order to remain at the upper bound of realizable forecast skill (dotted line). A good forecasting system must therefore attempt to maximize the area under this bounding curve through the proper selection of appropriate models (from an ensemble) for each forecast lead-time.

Two stochastic models, focussed on single radars, were investigated in this study. These were the "String of Beads" Model (SBM) (Pegram and Clothier, 2001; Clothier and Pegram, 2002) and the "Spectral PROGnosis" (S-PROG) model (Seed, 2001; Seed, 2003).

4.1.1 Estimating rainfall field advection

Several studies (e.g. Grecu and Krajewski, 2000; Germann et al., 2006) show that advection is an important aspect in rainfall nowcasting. The approaches typi-

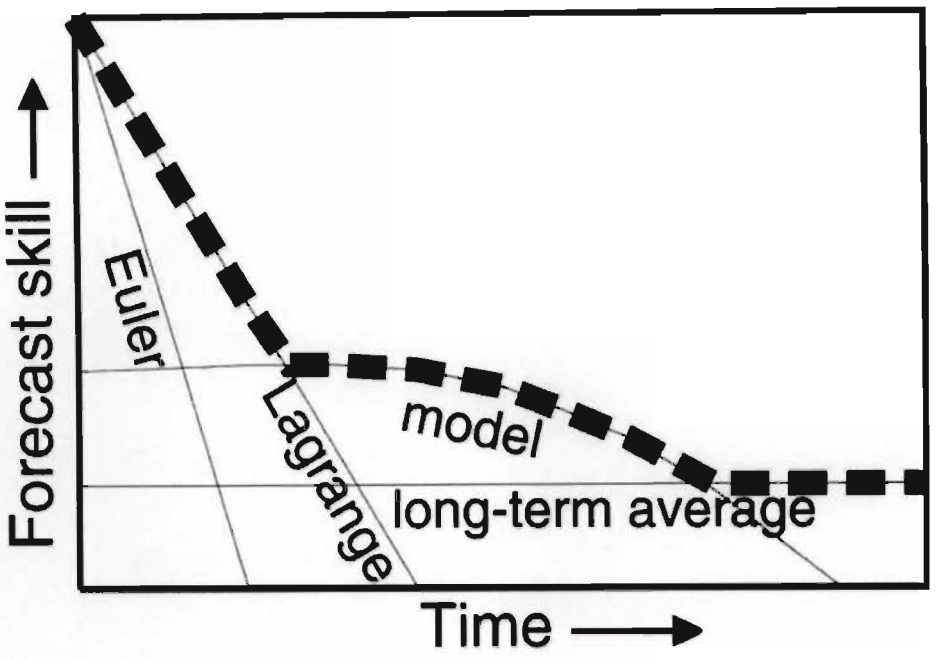


Fig. 4.1: Representation of relative forecast skill. [After Germann et al. (2006)].

cally adopted for determining the advection based on radar data can be classified according to one of two broad categories. These are either object based or field based, with the best recognized object tracking technique being the TITAN system described by Dixon and Wiener (1993). The field based techniques rely on correlation methods (Seed, 2003) or solutions to the Optical Flow Constraint (OFC) well-known from image processing literature (Bab-Hadiashar and Suter, 1998) and implemented, in the context of rainfall field nowcasting, by Seed (2001) and later, in an improved form, by Bowler et al. (2004). The Optical flow method was used in section 3.3 for the accumulation of remotely sensed rainfall fields.

The OF constraint equation is given by equation 4.1 and relates changes in the image brightness I (or reflectivity value in this case) to the magnitude of the velocity components u_x and u_y in the x and y directions, at a particular position in the image.

$$\frac{\delta I}{\delta x}u_x + \frac{\delta I}{\delta y}u_y + \frac{\delta I}{\delta t} = 0 \tag{4.1}$$

Clearly the OFC cannot be uniquely determined at a single grid point in the field as there is one equation in two unknowns. There are several techniques that have been employed to increase the number of constraints, for example Bowler et al. (2004) impose a smoothness constraint on the spatial second derivatives of u_x and u_y (this constrains the advection field to be smoothly varying in space).

An alternative approach is to find the least squares estimate of u_x and u_y over blocks, on the assumption that the motion field is reasonably constant within these blocks. The OFC for each pixel i within a block is given by

$$a_{i1}u_x + a_{i2}u_y = b_i, \qquad i = 1, 2, \cdots, n$$

where a_{i1} and a_{i2} are the partial derivatives ($\frac{\delta I}{\delta x}$ and $\frac{\delta I}{\delta y}$ respectively) for pixel i and b_i is $\frac{\delta I}{\delta t}$ for pixel i. The spatial derivatives are calculated by a central difference approximation while the temporal derivative is calculated using a backward difference approximation. The number of pixels in the block is given by n. The set of constraints within the block can be visualized as a set of straight lines, which must all intersect at a single point if the advection is constant throughout the block and the assumptions implicit in the OFC are valid. Figure 4.2 shows a schematic representation of this situation.

The least squares solution, which minimizes

$$U = \sum_{i=1}^{n} (a_{i1}u_x + a_{i2}u_y - b_i)^2$$

is well known and given by

$$\mathbf{x} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b} \tag{4.2}$$

where $\mathbf{x}$ is the vector $[u_x, u_y]^T$ containing the least squares estimate, $\mathbf{A}$ is the matrix containing the n pairs of a_{ik} and $\mathbf{b}$ the n-vector of b_i. Thus the estimate of

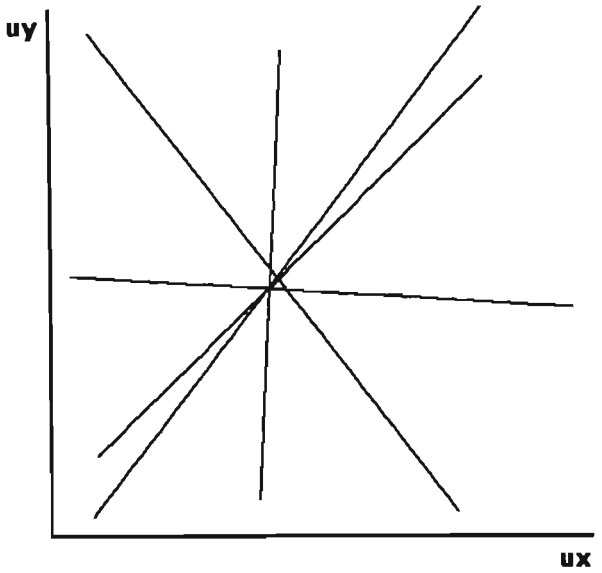


Fig. 4.2: Representation of Optical Flow Constraints. Since all of the lines do not intersect at precisely the same point a best estimate must be made using a least squares approach.

advection for each block is given by the solution of equation 4.2. The estimates for each block are then linearly interpolated onto the data grid.

4.1.2 Adapting SBM for nowcasting

The "String of Beads" Model (SBM) was developed for the simulation of rainfall fields measured by radar. Developing a forecasting implementation of the SBM rainfall field simulation model resulted in some improvements to the models advection routines. A comprehensive description of the SBM in simulation mode is given by Clothier and Pegram (2002) and subsequently Clothier (2003). A basic outline is presented here to provide a stand-alone description of the modelling process and to illustrate the differences between SBM in forecast and simulation modes.

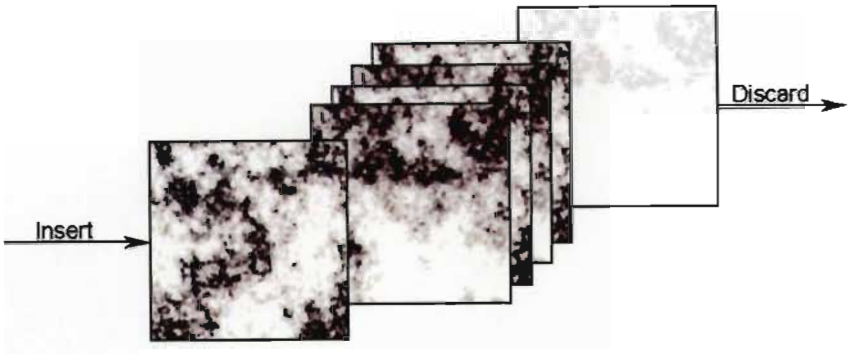


Fig. 4.3: An illustration of the SBM noise stack.

Rainfall field simulation

A typical simulation produces a series of wet and dry periods using an alternating renewal process. The wet periods are the "beads" and the dry periods the "string" connecting the wet periods, hence the name "String of Beads" model.

The simulation of each "bead" begins with the generation of a stack of five Gaussian random fields on a rectangular grid with dimensions $m \times n$. The values at each point in a field are independent and normally distributed with a zero mean and unit variance and each field is independent of the others. A new field is constructed (pixel by pixel) using an AR(5) model and placed at the first position on the stack, with the previous fields moving backwards one position and the final field falling away (Figure 4.3). The random number generation requires thought. At five minute intervals on a 128×128 grid, one hour of simulation requires generating 196608 independent random numbers. It is thus important to ensure that a generator with a long cycle period is used for lengthy simulations on large fields. The SBM uses a fast algorithm by Wichmann and Hill (1982), which has a cycle length of more than 6.95×10^{12}. This equates to approximately 4000 years of simulation at 5 minute resolution for the 128×128 km grid mentioned.

A mean field advection vector is used to maintain the appropriate temporal alignment of the pixel values in each field. A warm up period is required to ensure that the sequence of fields used in the generation is properly conditioned – i.e. the correct serial correlation structure, as defined by the AR(5) model param-

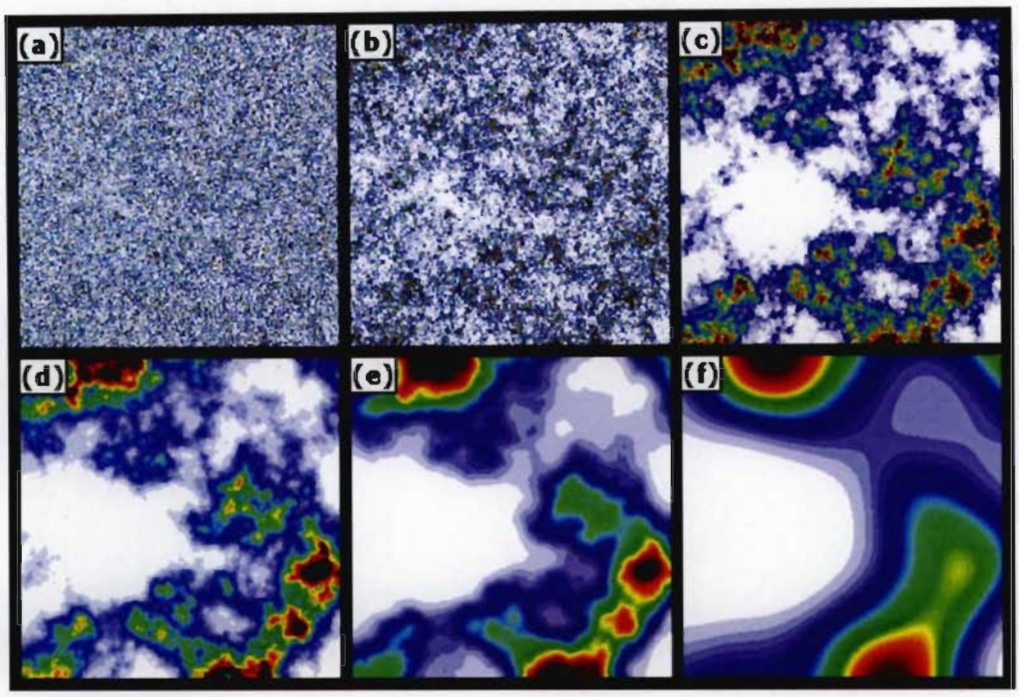


Fig. 4.4: Simulated rainfall fields from SBM, using an increasing degree of spatial correlation. Each field is 256×256 km and β the exponent of the power law filter in the Fourier domain is for panel *(a)* 0.5, *(b)* 1.5, *(c)* 3, *(d)* 3.75, *(e)* 5 and *(f)* 10.

eters, exists. Once the stack has been given a sufficient number of recursions to be correctly conditioned, a copy of the most recently generated field is made. At this point the copied field has no spatial correlation. A suitable spatial correlation structure is imposed, by applying a spatial filter to increase the spatial correlation to the desired level. The filter is defined by a power-law function in the Fourier domain, the structure and parameters of the filter were defined by Pegram and Clothier (1999) on the basis of observed radar rainfall data in South Africa. Once the field has appropriate spatial correlation, the simulated image scale statistics "Wetted Area Ratio" (WAR) and "Image Mean Flux" (IMF) are imposed on the field by a thresholding and scaling process. In this thesis, IMF will be referred to as Spatial Mean Flux (SMF) to avoid confusion with the concept of "Intrinsic Mode Functions" (introduced later in section 4.2.3). The resulting field is exponentiated to produce a field of simulated rainfall rates. Figure 4.4 shows six

different simulated rainfall fields produced using SBM. Each of the fields has the same WAR, SMF and underlying noise field, but the degree of spatial correlation increases row-wise from top left. The degree of spatial organization exhibited by observed radar rainfall fields (Pegram and Clothier, 1999) lies somewhere between the third and the fourth panels and is defined by the exponent of the power law filter function in the Fourier domain.

Rainfall field forecasting

In forecast mode the process has been streamlined. The spatial correlation structure from the observed fields is retained and the pixel scale development, now using an AR(2) model, computed directly from the observed fields. The image scale statistics are forecast using the bivariate AR(5) WAR-SMF process of the simulation mode but this time conditioned on the values from the previous five observations. Additionally, a sophisticated motion-tracking algorithm (section 4.1.1) is used to estimate the field advection for each pixel, rather than assuming a mean advection vector as is done in simulation mode. A dense grid of advection vectors is computed at each time step and used to advect the forecasts as well as for maintaining the appropriate temporal alignment between pixels. The smoothed advection grid is updated, at each time step, as new information becomes available. The computation of the advection vectors is efficient enough to allow it to be used for real-time applications.

4.1.3 The S-PROG model

The S-PROG model exploits the idea that rainfall fields exist as structures encompassing a range of spatial scales, where the persistence of structures within the field is proportional to their spatial scale i.e. larger scale structures have a longer persistence time than those at smaller scales (Seed et al., 1999). This idea is explored further in sections 4.2 and 4.3, using a new technique for the spatial decomposition of rainfall fields.

The first step in the generation of forecast rainfall fields using the S-PROG

model is to compute an advection field using the Optical Flow technique described in section 4.1.1. The advection field is computed using spatially smoothed copies of the observed data and the result is a dense grid of advection vectors for each pixel.

The observed field is then disaggregated into a multi-layered hierarchy of fields each of which describes the features of the data for a range of different spatial scales. This disaggregation is achieved by applying a bandpass filter in the Fourier domain to separate the field into a number of components with well defined spectral ranges as shown in figure 4.5. The detail of this Fourier filtering technique is described further in section 4.2.1.

Each of the fields in the hierarchy is modelled as a dynamically fitted temporal AR(2) process evolving at the relevant space scale. The AR(2) model (equation 4.3 is applied to each pixel in the field, with the pixels temporally matched using the advection field computed from the smoothed data.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + a_t \tag{4.3}$$

The parameters of the AR(2) model are estimated by solving the Yule-Walker equations, given for an AR(2) model as (e.g Box and Jenkins (1970)):

$$\phi_1 = \frac{\rho_1(1-\rho_2)}{1-{\rho_1}^2}$$

$$\phi_2 = \frac{\rho_2-{\rho_1}^2}{1-{\rho_1}^2}$$

Where ρ_1 and ρ_2 are replaced with the temporal correlation coefficients r_1 and r_2 computed from the two most recent fields at that level in the cascade (matched in a Lagrangian framework, according to the computed advection vectors).

Once a new field has been estimated for each level in the cascade, these are combined to construct the forecast estimate. Figure 4.5 shows an example of a reconstructed field, in panel (b). The forecast fields are produced in the same way as the reconstructed estimate, but the structure of the AR(2) models ensures that each level in the cascade decays to a mean value as the lead time increases. Since

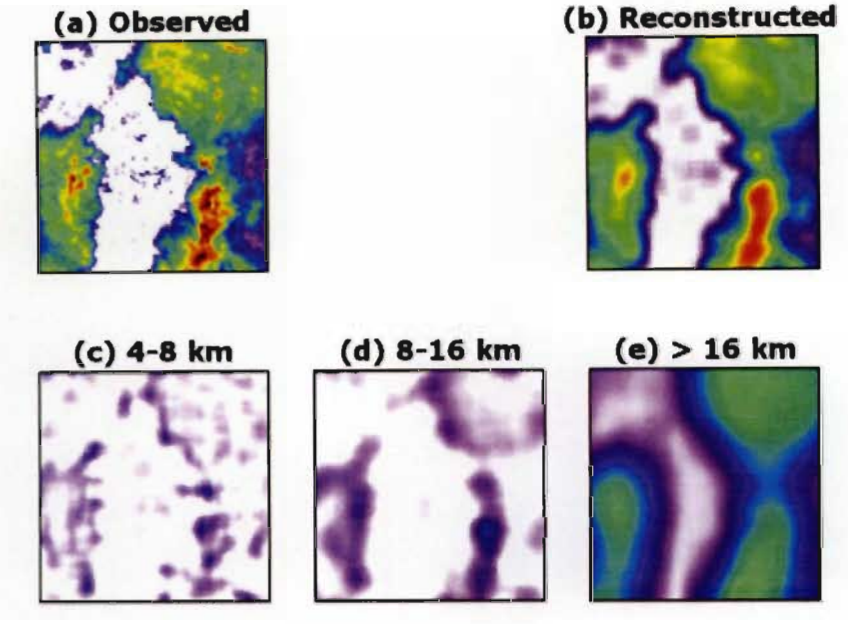


Fig. 4.5: An example of the decomposition of a rainfall field into contributions at various scales, using a Fourier notch filter. This technique is employed to produce the S-PROG cascade. Panel (a) shows the original reflectivity field. Panel (c) shows the components with scales between 4 and 8 km, panel (d) the components with wavelengths between 8 and 16 km and panel (e) shows the components with scales greater than 16 km. Panel (b) is a reconstruction of the original field using the components shown in (c), (d) and (e) and therefore is the portion of the data represented by spatial scales greater than 4 km.

the levels representing the smaller spatial scales have a much smaller life-time (correlation length) relative to the larger scales, the forecast quickly decays to a smooth field dominated by the information representing the larger scale structures.

The forecast field at each time-step is conditioned to have the same mean rainfall rate and wetted area as the most recent observed data, in order to prevent complete degradation of the forecasts. This is in contrast to the SBM approach where the evolution of these quantities is predicted using a bivariate AR(5) model and the forecast fields conditioned to the forecast WAR and SMF.

The implementation of the S-PROG model used in this study was kindly provided by Alan Seed of the Australian Bureau of Meteorology and included an implementation of an optical flow algorithm that formed the basis of the code used by the author for computing advection fields.

4.1.4 Case study - Comparing S-PROG and SBM in nowcasting mode

A series of qualitative and quantitative comparisons were made between the nowcasts produced by SBM and S-PROG. Both models were configured to produce nowcasts conditioned on two different observed rainfall events measured by weather radar in South Africa. The first event is a sequence of 42 instantaneous measurements by the Durban radar on 08 November 2000. Since the radar scans are at 5 minute time steps this equates to a 3.5 hour observation period. This event is referred to as Event01 in the figures. The second event is a sequence of 40 instantaneous measurements by the Bethlehem radar on 24 January 2002. The total observation period was therefore 3.2 hours. This event has been dubbed Event02 in the figures. For each observed radar scan during the events, nowcasts were produced by each model out to a lead time of one hour. The nowcasts were produced in a sequential way, treating each observation in turn as a new time origin. The resulting nowcasts were compared to the radar observations at a number of lead-times as described below.

Figure 4.6 shows a short sequence of qualitative comparisons between the observed data and corresponding nowcasts made using the SBM and S-PROG models, scaled to the same colour palette. The nowcasts have been produced at time steps of approximately 5 minutes (the nowcasts are for the same time as the next available observed field as the radar scan intervals vary slightly [+- 15 s] but scans are approximately 5 minutes apart and this interval will be used for simplicity in the text), thus the forecasts in figure 4.6 are between 5 and 20 minutes ahead. The images shown in each panel represent the observed and nowcast reflectivity fields measured in dBZ. All fields have been set to zero below a threshold value of 10

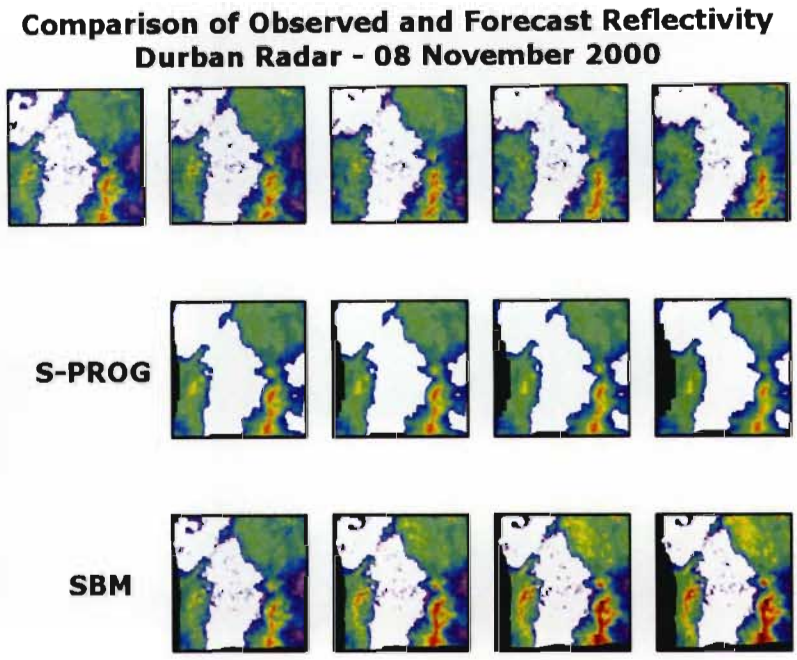


Fig. 4.6: Comparative forecasts and observed reflectivity fields (Event01): top row – Observed data, second row – S-PROG nowcasts, third row – SBM nowcasts. Each frame represents the instantaneous reflectivity at 5 minute intervals starting simultaneously. The blacked out portions of the nowcast images indicate places where new (unknown) information from outside the data region is advected into the view.

dBZ as the associated rainfall rates below this level are considered to be negligible ($\leq$ 0.15 mm/hr) for the purposes of comparing nowcast skill.

The forecast fields have been compared with the observed fields in terms of two "goodness of fit" measures. The first is the square root of the mean sum of squared errors over all pixels in the field (RMSSE)

$$RMSSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[y_{obs}(i) - y_{for}(i)\right]^2}$$

where $y_{obs}(i)$ is the observed reflectivity for pixel i and $y_{for}(i)$ is the forecast value

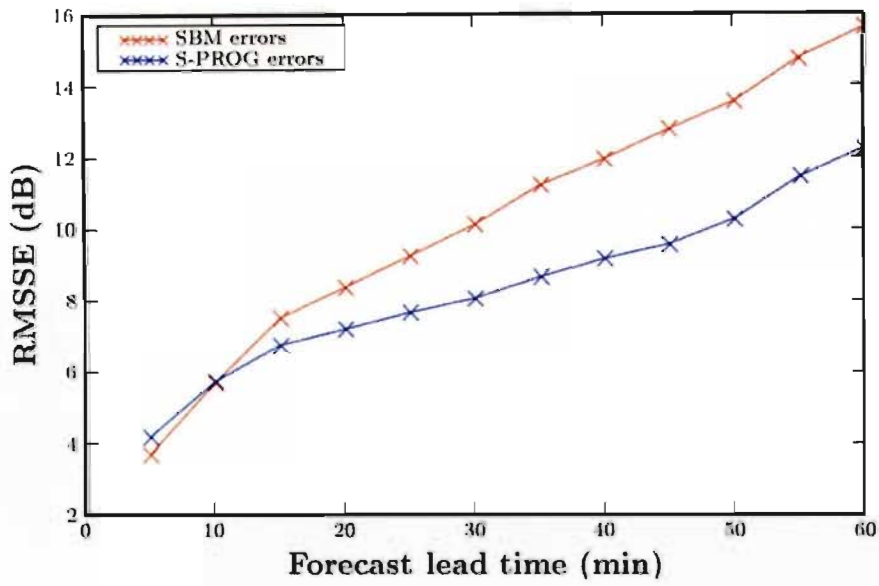


Fig. 4.7: Root mean sum of squared forecast errors for lead-times up to 1 hour. The forecasts are made for 5-min intervals based on the same set of initial data. This data is produced from a sequence during Event01.

of reflectivity for pixel i at the same time. The second criterion used was the mean absolute deviation (MAD)

$$MAD = \frac{1}{N}\sum_{i=1}^{N}|y_{obs}(i) - y_{for}(i)|$$

this is closely related to the SMF in the rainfall rate domain and gives an indication of whether the mean flux over the data domain is comparable to the observed flux. Figures 4.7 and 4.8 compare the magnitude of these two measures out to an hour ahead for a single forecast origin during Event01.

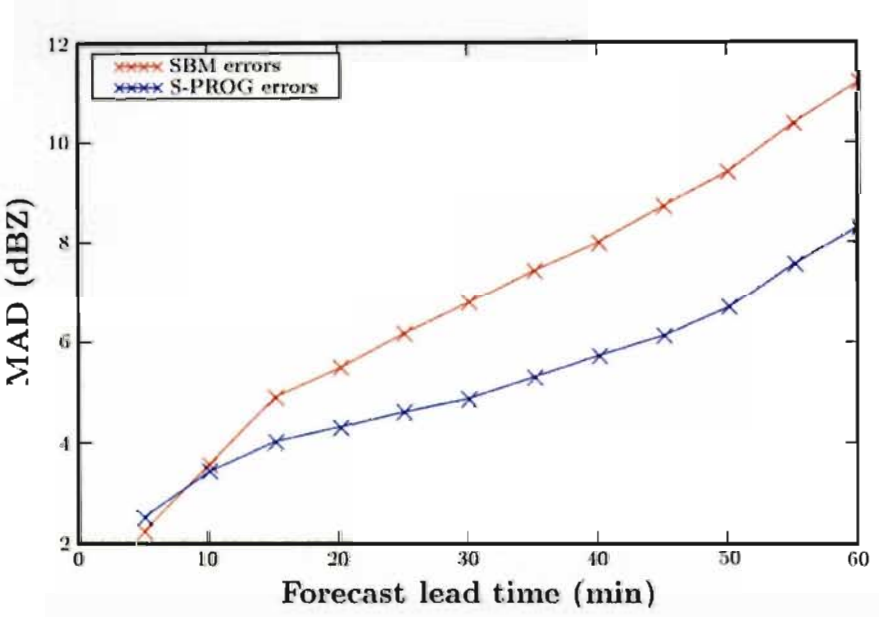


Fig. 4.8: Mean Absolute Deviation of squared forecast errors for lead-times up to 1 hour. The forecasts are made for 5-min intervals based on the same set of initial data. This data is produced from a sequence during Event01.

Figures 4.9 and 4.10 show in detail the nowcast errors computed for each pixel. The regions which are blacked out were ignored in the computation of the mean field "goodness of fit" measures. Histograms of the error fields have been included in the figures and show two interesting features. First, the bulk of the pixels exhibit close to zero error. This can be attributed to the non-raining pixels which match for both the observed and forecast fields. Second, there is evidence of bi-modality of the errors in several cases, which can be attributed to the spatial mismatch in raining areas between observations and nowcasts. This is most pronounced for the SBM nowcasts and a possible reason is discussed below.

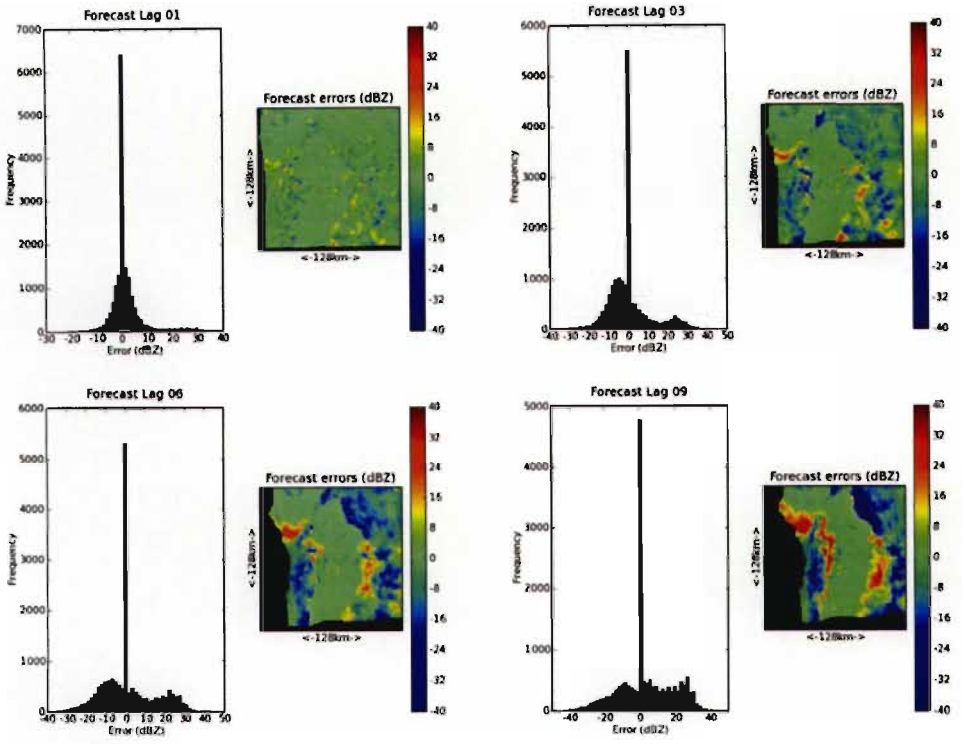


Fig. 4.9: A comparison of the SBM forecast errors with increasing lead-time for Event01. The lag shown in each of the four panels refers to the number of 5-min intervals ahead for which the forecast is made. Therefore, lag 1 refers to 5-mins ahead, lag 3 15-mins ahead and so on. The Histogram of the forecast error field in shown in the left hand side of each panel. The blacked out portions of the field indicate places where new (unknown) information from outside the data region has been advected into the view. The error statistics (and Histograms) ignore these portions of the field.

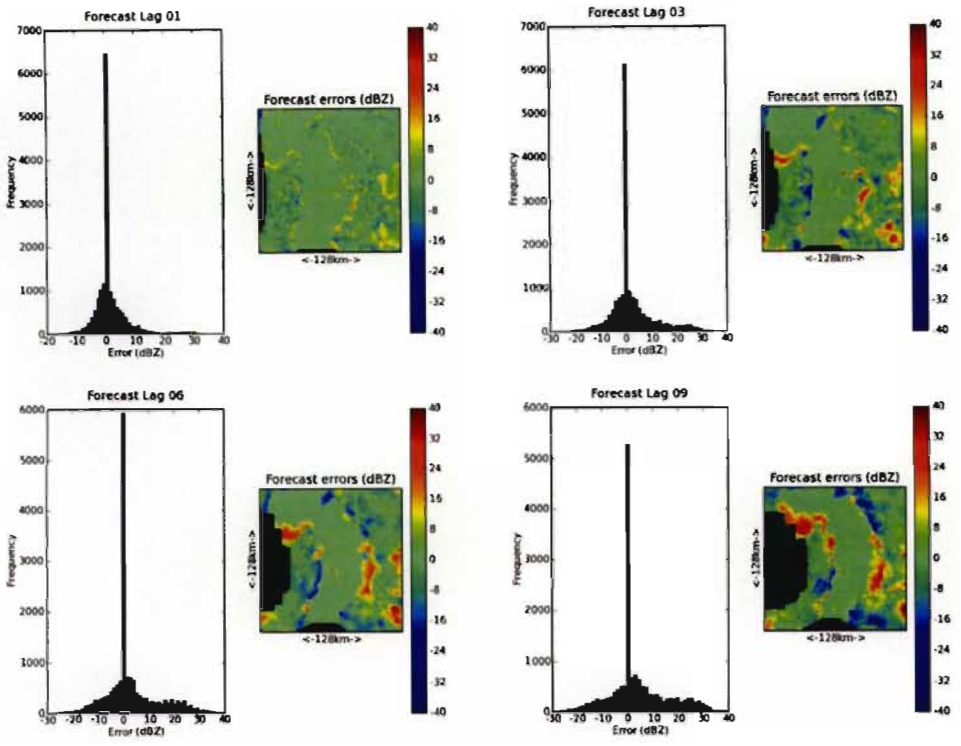


Fig. 4.10: A comparison of the SPROG forecast errors with increasing lead-time for Event01. The lag shown in each of the four panels refers to the number of 5-min intervals ahead for which the forecast is made. Therefore, lag 1 refers to 5-mins ahead, lag 3 15-mins ahead and so on. The Histogram of the forecast error field in shown in the left hand side of each panel. The blacked out portions of the field indicate places where new (unknown) information from outside the data region has been advected into the view. The error statistics (and Histograms) ignore these portions of the field.

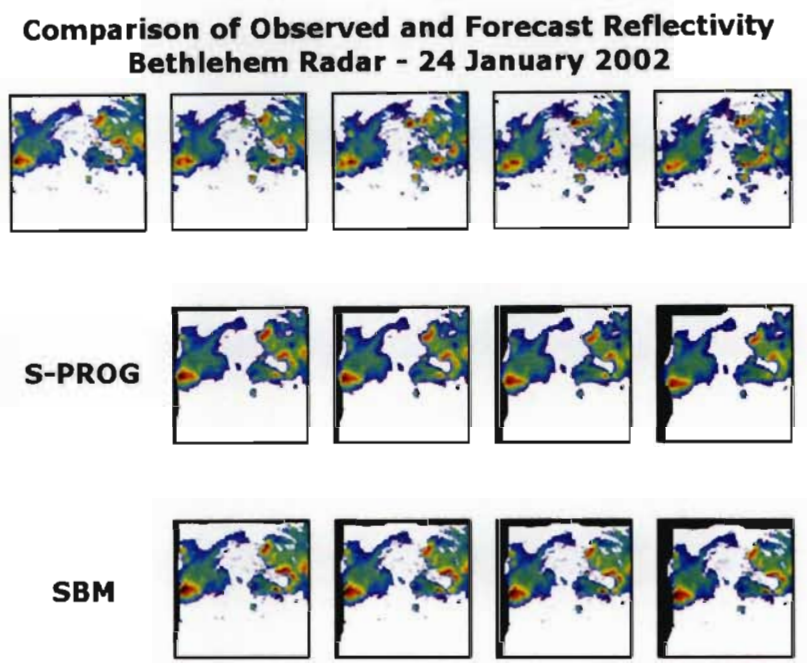


Fig. 4.11: Comparative forecasts and observed reflectivity fields (Event02): top row – Observed data, second row – S-PROG nowcasts, third row – SBM nowcasts. Each frame represents the instantaneous reflectivity at 5 minute intervals starting simultaneously. The blacked out portions of the nowcast images indicate places where new (unknown) information from outside the data region is advected into the view.

Figure 4.11 shows a similar qualitative comparison to that shown in figure 4.6, but this time for the Bethlehem radar on 24 January 2002. A comparison of the mean field error measures for each model is shown in figures 4.12 and 4.13, while the detailed errors per pixel are presented in figures 4.14 and 4.15.

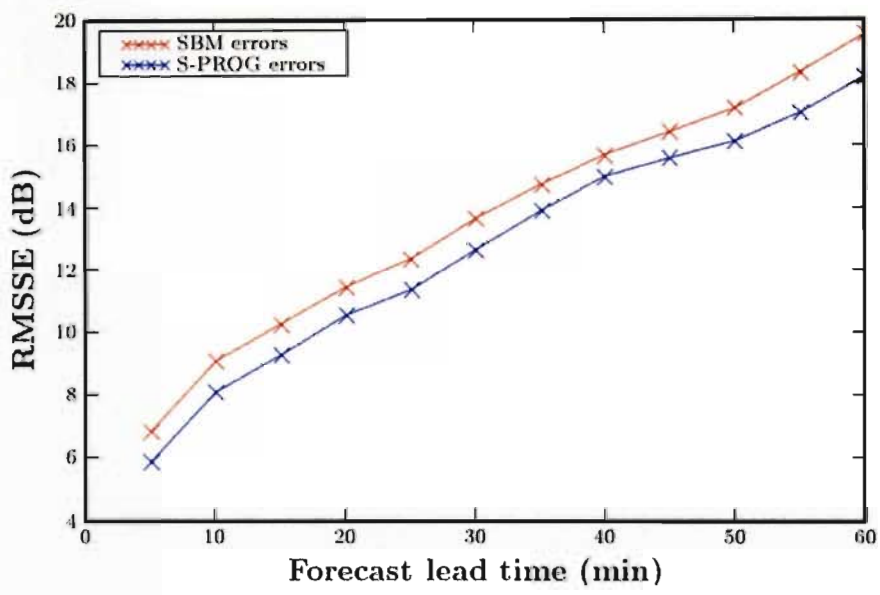


Fig. 4.12: Root mean sum of squared forecast errors for lead-times up to 1 hour. The forecasts are made for 5-min intervals based on the same set of initial data. This data is produced from a sequence during Event02.

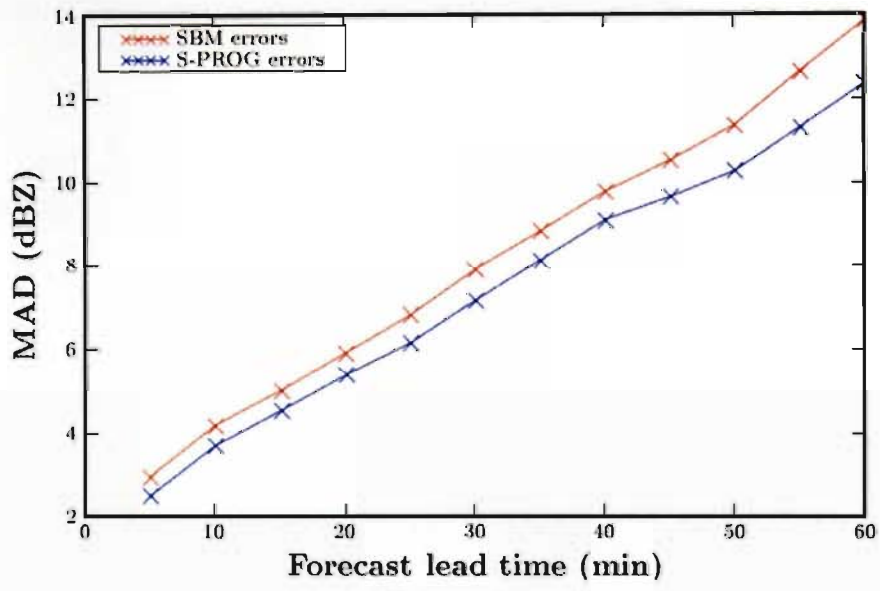


Fig. 4.13: Mean Absolute Deviation of squared forecast errors for lead-times up to 1 hour. The forecasts are made for 5-min intervals based on the same set of initial data. This data is produced from a sequence during Event02.

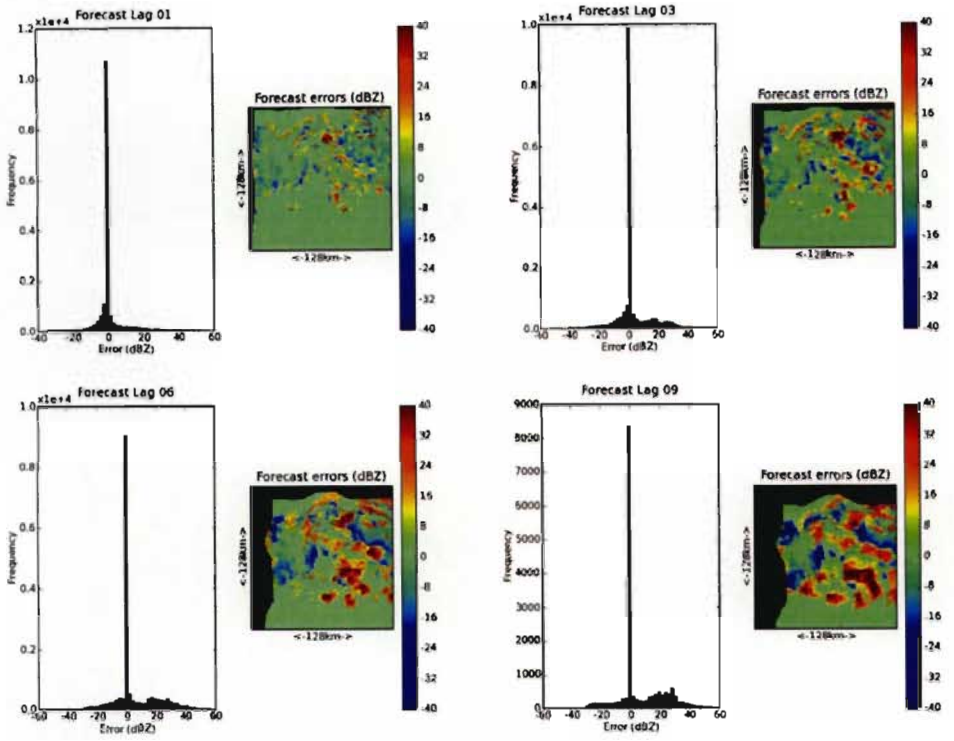


Fig. 4.14: A comparison of the SBM forecast errors with increasing lead-time for Event02. The lag shown in each of the four panels refers to the number of 5-min intervals ahead for which the forecast is made. Therefore, lag 1 refers to 5-mins ahead, lag 3 15-mins ahead and so on. The Histogram of the forecast error field in shown in the left hand side of each panel. The blacked out portions of the field indicate places where new (unknown) information from outside the data region has been advected into the view. The error statistics (and Histograms) ignore these portions of the field.

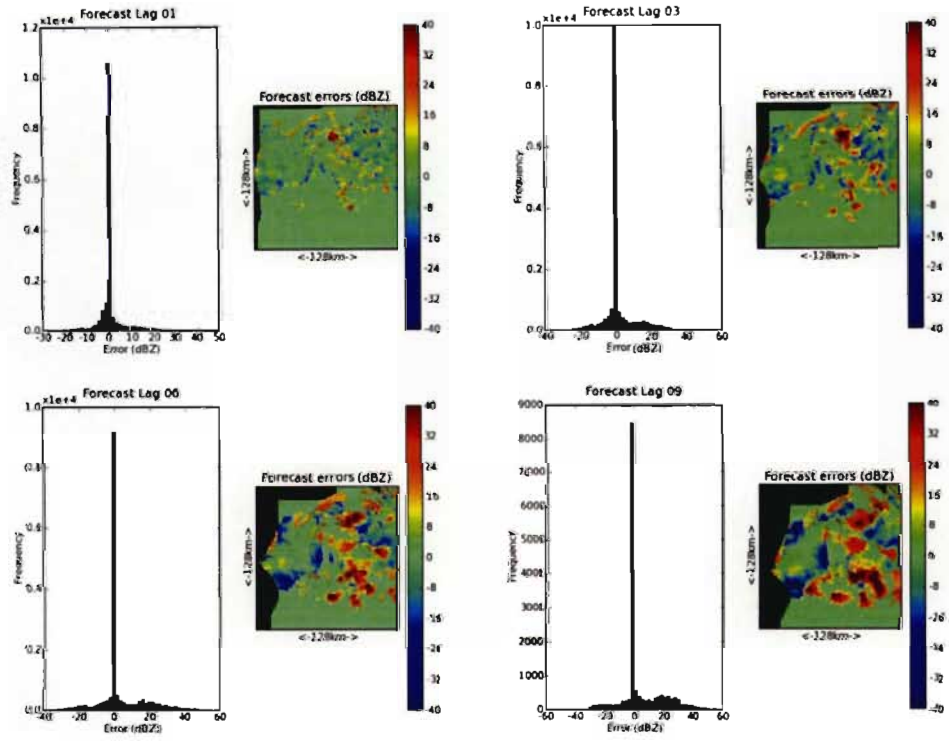


Fig. 4.15: A comparison of the SPROG forecast errors with increasing lead-time for Event02. The lag shown in each of the four panels refers to the number of 5-min intervals ahead for which the forecast is made. Therefore, lag 1 refers to 5-mins ahead, lag 3 15-mins ahead and so on. The Histogram of the forecast error field in shown in the left hand side of each panel. The blacked out portions of the field indicate places where new (unknown) information from outside the data region has been advected into the view. The error statistics (and Histograms) ignore these portions of the field.

Figures 4.7 and 4.8 (for Event01) show the increasing error with lead time for nowcasts made from a single time origin. If the origin is incremented to simulate receiving a new radar scan under operational conditions (for example), then a new pair of lines may be plotted for RMSSE and MAD. Figures 4.16 and 4.17 show a collection of many pairs of lines for each event.

These two figures clearly show the general tendency for S-PROG to perform better than SBM. What is also evident from these figures is the increase in the uncertainty of the forecast errors with lead time. This is expected and is shown in the widening of the error range (with lead time) for nowcasts made from different time origins.

It makes sense that S-PROG performs better (in terms of this criterion) since the model is formulated to give an optimum nowcast in the least squares sense, with the nowcast values being degraded towards a mean field value as the lead-time increases. The resulting errors have a smaller magnitude. On the other hand, SBM does not behave in this way because the image scale parameters (WAR and SMF) evolve according to a bivariate AR(5) model. This means that dynamic changes in intensity and wetted area are evident in the nowcasts. It appears that this is only a good strategy for error reduction if the forecasts of these parameters are highly accurate and the placement of raining areas in nowcasts close to the observed.

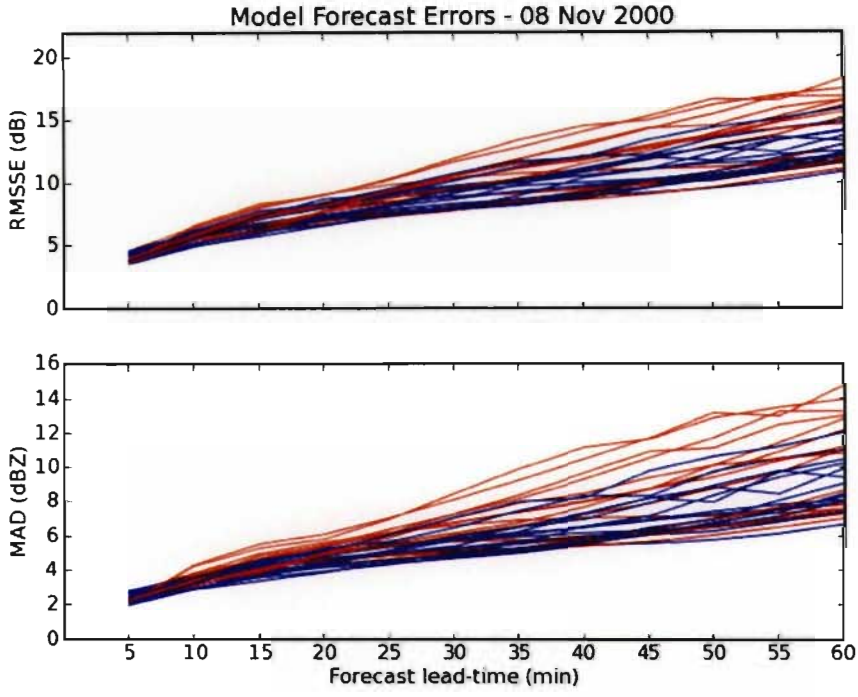


Fig. 4.16: Multiple error comparison (Event01) for lead-times up to 1 hour. The forecasts are made for 5-min intervals based on the same set of initial data. Red lines are the SBM results while the blue lines are the S-PROG results.

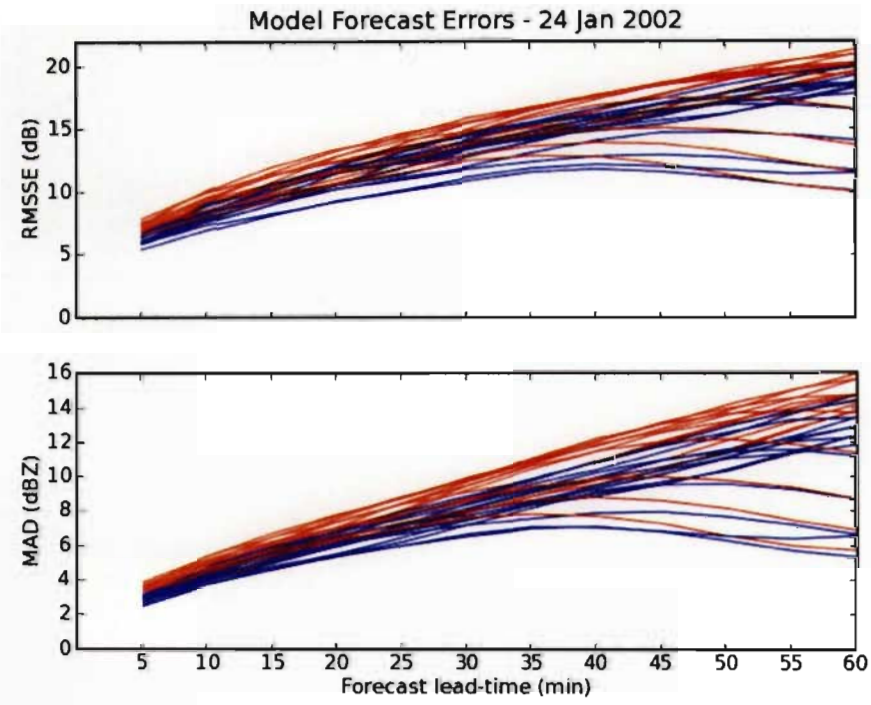


Fig. 4.17: Multiple error comparison (Event02) for lead-times up to 1 hour. The forecasts are made for 5-min intervals based on the same set of initial data. Red lines are the SBM results while the blue lines are the S-PROG results.

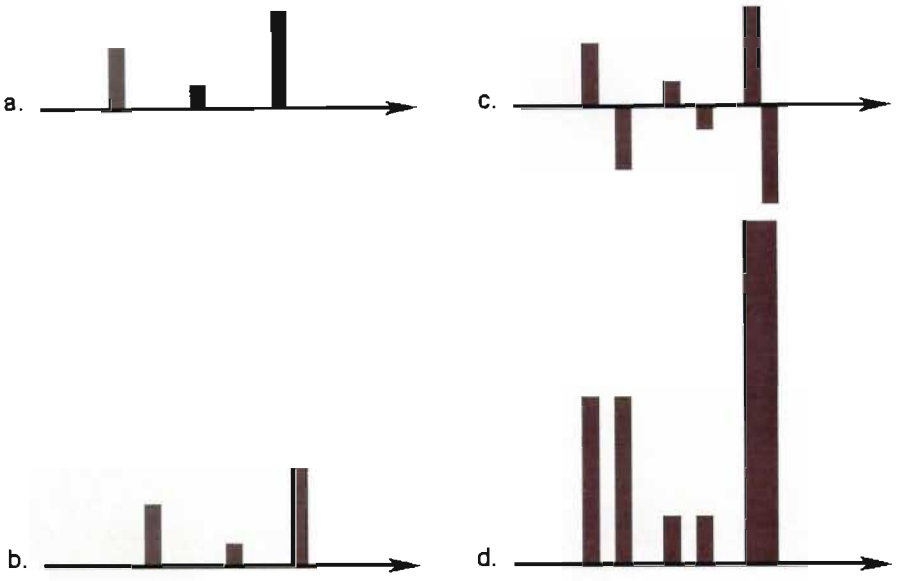


Fig. 4.18: Sensitivity of RMSSE to the position of peaks.

The RMSSE criterion is very sensitive to discrepancies in the relative positioning of peaks in the fields. Squaring the error values removes the effects of sign from the comparison, which further increases the apparent error. Figure 4.18 shows this effect graphically for a single dimension (e.g. time).

4.18a shows the arrival times of three observed rainfall pulses while 4.18b shows the nowcast arrival times, the magnitude of the rainfall is exactly correct (an unlikely situation in practise) but the arrival times of the pulses are not correctly forecast. Figure 4.18c shows the result of taking the difference between 4.18a and 4.18b. It's obvious from this that the mean error (over the period shown) will be zero since the rainfall volume is the same in both cases. However, 4.18d shows the squared errors and it's clear from inspection of the figure that the sum of squared errors will be large and non-zero. In this case the RMSSE will be larger (a poorer "fit" to the observations) than the RMSSE that would result from making a nowcast that expects no rainfall at all. It is for this reason that the MAD was considered in addition to RMSSE for comparing the model performance. The MAD criterion gives a better indication of whether a reasonable rainfall volume has been forecast. Providing the arrival time of the rainfall is not too poorly fore-

cast, getting the correct volume is more significant for producing a good flood forecast.

On the basis of the nowcast comparisons performed in this section, it seems that the scale separation approach followed by S-PROG is an appropriate choice for nowcasting rainfall fields in the presence of many uncertainties. The implications of this approach are investigated further in sections 4.2 and 4.3.

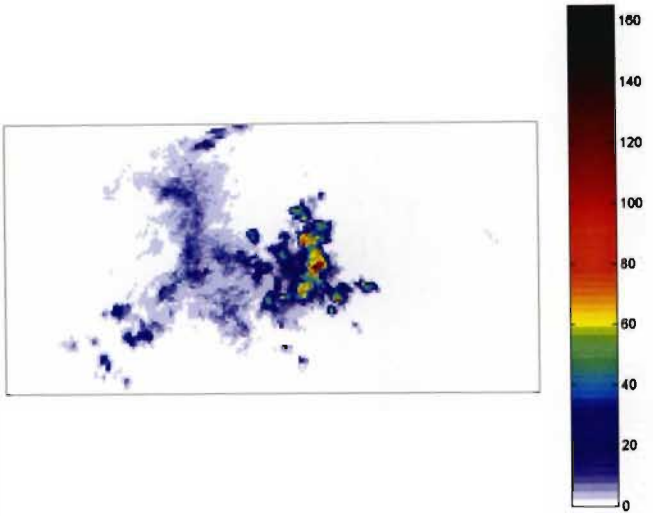


Fig. 4.19: An observed convective rainfall field measured by S-Band weather radar at Bethlehem, South Africa (colour scale indicates instantaneous rain rate in mm/hr). The image is 100×200 with 1 km^2 pixels.

4.2 Extracting spatial structure from rainfall data

Spatial rainfall data contain information at a broad range of spatial scales (Schertzer and Lovejoy, 1987; Harris et al., 2001; Pegram and Clothier, 2001). It has been suggested in the literature (Seed, 2003; Turner et al., 2004) that the larger spatial components exhibit more temporal persistence than the smaller ones; this premise is used here to prepare the data for nowcasts based on the evolution of the different spatial scale components of space-time rainfall sequences. A typical instantaneous estimate of rainfall rate obtained by weather radar is shown in figure 4.19. Examination of the (radially averaged) power spectrum (Figure 4.20) derived from these data indicates that most of the power, hence potential for deterministic prediction in the context of nowcasting, is contained in the high wavelength components, as exploited by S-PROG.

In this section, the focus is on developing a data-driven technique to remove the short wavelength (less persistent in time) modes as the first step towards a rain-

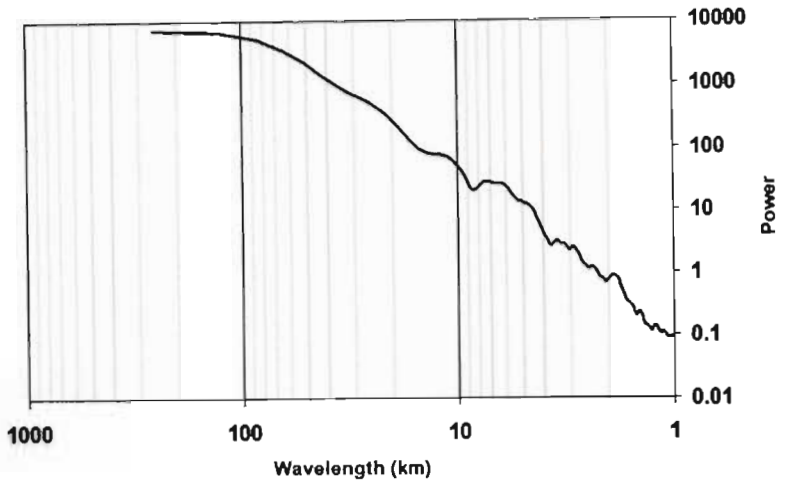


Fig. 4.20: Radially averaged power spectrum of instantaneous rainfall rate from typical radar rainfall data shown in figure 4.19.

fall nowcasting scheme and as a useful tool for studying the structure of rainfall fields. The technique employed is a two-dimensional (2D space) generalization of the one-dimensional Empirical Mode Decomposition (EMD) technique introduced by Huang et al. (1998). Alternative techniques for extracting the spatial structure are also noted, those considered are Fourier and Wavelet analysis.

4.2.1 Fourier analysis

Time varying signals may be described in both the time and frequency domains. In the time domain, the signal $f(t)$ is described by it's amplitude as a function of time. If this signal is transformed into the frequency domain, the transformed signal is described by it's amplitude as a function of frequency. This can be easily generalized to other units. For example, the signal of interest may be described as a function of position x in metres, the analogue of this function in the "frequency" domain would be amplitudes as a function of inverse wavelength (wavenumber). The Fourier transform provides a means of switching between these representa-

tions, using the following relationships.

$$F(k) = \int_{-\infty}^{\infty} f(x)e^{2\pi ikx}dx \tag{4.4}$$
$$f(x) = \int_{-\infty}^{\infty} F(k)e^{-2\pi ikx}dk$$

where $f(x)$ is the signal and $F(k)$ is it's Fourier transform.

Data encountered in practice is discretely sampled, not continuous. A discrete analogue of the continuous Fourier transform is therefore required. Following the exposition in Press et al. (1992), consider sampling a function $f(x)$ at n equally spaced points

$$f_j = f(x_j), \qquad x_j = j\Delta, \qquad j = 1, 2, \cdots, n$$

where Δ is the length of the sampling interval.
Define

$$k_s \equiv \frac{s}{n\Delta}, \qquad s = -\frac{n}{2}, \cdots, \frac{n}{2}$$

Using the preceding, the integral in equation 4.4 can be approximated by a discrete sum

$$\begin{aligned} F(k_s) &= \int_{-\infty}^{\infty} f(x)e^{2\pi ik_s x}dx \\ &\approx \sum_{j=0}^{n-1} f_j e^{2\pi ik_s x_j}\Delta = \Delta\sum_{j=0}^{n-1} f_j e^{2\pi ijs/n} \end{aligned}$$

The discrete Fourier transform is thus given by

$$F_s = \sum_{j=0}^{n-1} f_j e^{2\pi ijs/n}$$

and can be viewed as a discretely sampled version of the continuous transform

$$F(k_s) \approx \Delta F_s$$

The transform can be efficiently implemented using the Fast Fourier Transform (FFT) algorithm developed by Cooley and Tukey (1965). There are many computationally expensive operations (such as convolution) that have more efficient analogues in the frequency domain. This makes the FFT a popular and powerful tool for numerical analysis.

Two dimensional Fourier transform

The focus in this chapter is on spatial rainfall data, represented as two dimensional fields. Thus a two dimensional discrete Fourier transform is required. Consider a function $f(x_1, x_2)$ defined on a discrete grid $0 \leq x_1 \leq n-1,\ 0 \leq x_2 \leq m-1$. The two dimensional Fourier transform is defined as

$$F(k_1, k_2) = \sum_{x_2=0}^{m-1} \sum_{x_1=0}^{n-1} e^{2\pi i x_2 k_2/m}\, e^{2\pi i x_1 k_1/n}\, f(x_1, x_2)$$

This amounts to performing two passes of the transform, first along one dimension and subsequently along the second, because the exponentials can be separated and computed independently. Thus spatial information between data in adjacent columns (or rows) is ignored due to the directional orthogonality of the transform.

Fourier filtering to extract spatial structure

The Discrete Fourier approximation of a signal can be fundamentally defined in terms of the Euler-Fourier coefficients (a_0, a_k, b_k) with k=1, 2, ..., m (Equation 4.5). The signal is approximated as the sum of sine and cosine terms. The coefficients are all that are required to reconstruct the series and any signal can be well approximated (as long as it satisfies the Dirichlet conditions), provided m is sufficiently large. In equation 4.5, $F(x_j)$ is the Fourier approximation of the signal y_j at each of the n discrete (evenly spaced) data points x_j. L is the range of values

x_j over which the data set is assumed periodic.

$$\begin{aligned} F(x_j) &= \tfrac{a_0}{2} + \sum_{k=1}^{m} \{a_k \cos(2\pi k x_j/L) + b_k \sin(2\pi k x_j/L)\} \\ a_k &= 2\sum_{j=1}^{n} \cos(2\pi k x_j/L) y_j/n \\ b_k &= 2\sum_{j=1}^{n} \sin(2\pi k x_j/L) y_j/n \end{aligned} \tag{4.5}$$

This approach does not often yield much practical information about the signal (unless it truly is periodic) as there is an infinite number of components in a continuous signal, and n in a finite sample. A potentially more useful approach is to use band pass filtering to separate the signal into a small number of components, each of which represents the information contained over a distinct frequency range (wavenumber in this case). The filtering process is achieved by the convolution of a filter function with the data (an efficient operation, consisting of an element wise multiplication, in the frequency domain). This approach is employed in the S-PROG model described earlier (section 4.1.3) and a typical decomposition is shown in figure 4.5. FFT filtering is used in a slightly different sense to impose spatial correlation structure on the random noise fields for SBM rainfall simulations.

4.2.2 Wavelets

The Wavelet transform provides a means to perform a space-scale decomposition of a signal using a basis function with finite spatial support. The basis function is called a Wavelet and needs to satisfy certain criteria in order to qualify. These criteria include (e.g Daubechies, 1992)

- **Admissibility** In order to be a wavelet the basis function $\psi(\mathbf{x})$ must satisfy

$$C_\psi < \infty$$

where

$$C_\psi = (2\pi)^n \int_{\mathbb{R}^n} |\hat{\psi}(\mathbf{k})|^2 \frac{d^n\mathbf{k}}{|\mathbf{k}|^n}$$

and $\hat{\psi}(\mathbf{k})$ is given by

$$\hat{\psi}(\mathbf{k}) = (2\pi)^{-n} \int_{\mathbb{R}^n} \psi(\mathbf{x}) e^{-\sqrt{-1}\mathbf{k}\cdot\mathbf{x}} d^n\mathbf{x}$$

and n is the number of spatial dimensions. If $\psi(\mathbf{x})$ is an integrable function this implies that the wavelet must have a zero mean.

- **Similarity** The set of basis functions used to perform the decomposition should all be scaled and translated derivatives of the same waveform ("mother" wavelet). This means that each of the analysing wavelets will have the same shape, proportional to the support (scale) of the wavelet. This results in the well known properties of the Wavelet transform which dictate that good spatial resolution is available for the small scale (high wavenumber) features and good scale resolution is available for large scales. Figure 4.21 summarizes this idea schematically.

- **Invertibility** The invertibility condition requires that there be at least one formula for exact reconstruction of the signal from it's Wavelet transform coefficients. This means that there must be a way to perform an inverse transform.

- **Regularity** The wavelet function should be continuous and have a finite spatial support.

- **Cancelations** The wavelet should not only have a zero mean as required by the admissibility condition, but also have at least some high-order vanishing moments. This is to prevent the wavelet from reacting to low order variations in the signal when the higher order fluctuations are of interest. If fluctuations of order M are of importance then

$$\int_{\mathbb{R}^n} \psi(\mathbf{x}) \mathbf{x}^m d^n\mathbf{x} = 0, \qquad \text{for all } m \leq M$$

The Wavelet coefficients $\chi(a, b)$ are computed by performing a convolution between the function (data) being analysed and the wavelet corresponding to a

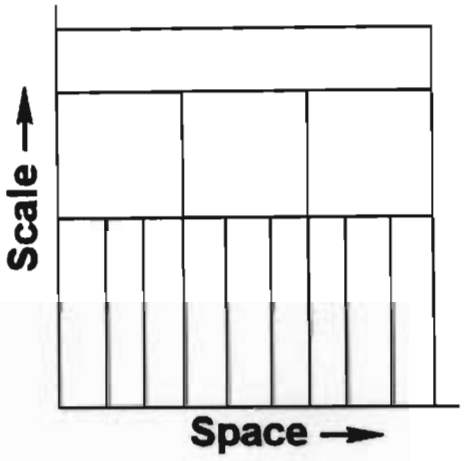


Fig. 4.21: Schematic indicating the information contained in a typical space-scale plot of wavelet coefficients. Note that there is good spatial resolution at small scales and increasing scale resolution (with poor spatial resolution) at larger scales.

particular scale and position

$$\chi(a,b) = \int_{\mathbb{R}^n} f(\mathbf{x})\psi(a,b,\mathbf{x})d^n\mathbf{x}$$

where $\psi(a,b,\mathbf{x})$ is a scaled and translated version of the mother wavelet $\psi(\mathbf{x})$

$$\psi(a,b,\mathbf{x}) = \frac{1}{\sqrt{a}}\psi(\frac{\mathbf{x}-b}{a})$$

A signal analysed using Wavelets can be decomposed into components that are a function of both position and scale. The Wavelet coefficients at a particular scale can be used to construct band pass filters and decompose a rainfall field according to spatial scale (Turner et al., 2004).

4.2.3 Empirical mode decomposition (EMD)

In a single dimension, EMD analysis produces a set of Intrinsic Mode Functions (IMF) that are very nearly orthogonal; in two dimensions a set of Intrinsic

Mode Surfaces (IMS) is produced with similar quasi-orthogonal properties. Two-dimensional EMD appears to have first been introduced by Linderhed (2002) in the context of image compression; the key contribution in this section is to introduce 2D EMD in a Hydrometeorological context as a tool for the analysis of space-time rainfall data. More specifically, the focus is on the implementation and development of the two-dimensional extension of the EMD algorithm in this context, decomposing spatial rainfall data into its intrinsic spatial scale components.

In the application presented here, the least persistent IMS (exhibiting the highest local wavenumber and least amount of spatial correlation – hence nearly white noise) is computed and removed from the raw rainfall data leaving a residual composed of the more persistent lower frequency structural components in the data. This process is equivalent to applying a low-pass spatial filter, based on the observed properties of the data rather than the predefined basis functions used in traditional Fourier or Wavelet decompositions. The idea can be easily extended to produce the EMD equivalent of the Fourier notch filters employed by Seed (2003) to decompose rainfall fields into their characteristic spatial frequency components. In sections 4.2.4 and 4.2.5, simple theoretical examples, showing the power of EMD in one and two dimensions, are presented as a "proof of concept" before applying the procedure to observed radar rainfall data from Bethlehem, South Africa (Section 4.3). These complement and extend the original presentation by Huang et al. (1998) and the work of Flandrin et al. (2004). Computational aspects relating to image processing and surface fitting are covered in detail and conclusions drawn.

4.2.4 Empirical mode decomposition in a single dimension

The basic idea embodied in EMD analysis, is to allow for an adaptive and unsupervised representation of the intrinsic components of linear and non-linear signals based purely on the properties observed in the data without requiring that the signals exhibit stationarity. As Huang et al. (1998) point out in their abstract

"This decomposition method is adaptive and therefore highly effi-

cient. Since the decomposition is based on the local characteristic time scale of the data, it is applicable to non-linear and non-stationary processes."

Few sequences of observations of natural phenomena are long enough to justify the hypothesis of stationarity and frequently the phenomena are patently non-stationary. This tacitly applies in the measurement of rainfall at a point or in space-time because sequences of rain are interspersed with dry periods (intermittency) and during the raining periods, the variability of the intensity due to mixtures of rainfall type (stratiform, convective, frontal) confound the homogeneity definition. The EMD algorithm copes with stationarity (or the lack of it) by ignoring the concept, embracing non-stationarity as a practical reality. For a fuller discussion of the genesis of these ideas, see the introduction of Huang et al. (1998), who also heuristically demonstrate the implicit orthogonality of the sequences of IMFs defined by the EMD algorithm.

In the application of the EMD algorithm, the possibly non-linear signal, which may exhibit varying amplitude and local frequency modulation, is linearly decomposed into a finite number of (zero mean) frequency and amplitude modulated signals, as well as a residual function which exhibits a single extremum, is a monotonic trend or is simply constant. Although EMD is a relatively new data analysis technique, its power and simplicity have encouraged its application in a myriad of fields. It is beyond the scope of this thesis to give an exhaustive review of the applications, however a few interesting examples are cited here to give the reader a feeling for the broad scope of applications. Chiew et al. (2005) examine the one-dimensional EMD of several annual streamflow time series to search for significant trends in the data, using bootstrapping to test the statistical significance of identified trends. The technique has been used extensively in the analysis of ocean wave data (Huang et al., 1999; Hwang et al., 1999) as well as in the analysis of polar ice cover (Gloersen and Huang, 2003). EMD has also been applied in the analysis of seismological data by Zhang et al. (2003) and has even been used as an aid in diagnosing heart rate fluctuations (Balocchi et al., 2004).

Computing the one-dimensional EMD

The EMD algorithm recursively extracts the oscillatory mode that exhibits the highest local frequency from the data ("detail" in the Wavelet context or the result of a high-pass filter in Fourier analysis), leaving the remainder as a "residual" ("approximation" in Wavelet analysis). Successive applications of the algorithm on the sequence of residuals produce a complete frequency decomposition of the data. The final residual is a constant, a monotone trend or a curve which only has a single extremum.

The EMD of a one-dimensional data set $z(k)$ is obtained using the following procedure:

1. Set $r_0(k)=z(k)$ and set $i=1$.
2. Identify all of the extrema (maxima and minima) in $r_{i-1}(k)$.
3. Compute a maximal envelope, $\max_{i-1}(k)$, by interpolating between the maxima found in step 2. Similarly compute the minimal envelope, $\min_{i-1}(k)$. Cubic splines (as suggested by Huang et al., 1998) appear to be the most appropriate interpolation method for deriving these envelopes in one dimension (Flandrin et al., 2004).
4. Compute the mean value function of the maximal and minimal envelopes $m_{i-1}(k)=\frac{[\max_{i-1}(k)+\min_{i-1}(k)]}{2}$.
5. The estimate of the IMF is computed from $\mathrm{IMF}_i(k)=r_{i-1}(k)-m_{i-1}(k)$. Each IMF is (by definition) supposed to oscillate about a zero mean and in practice it is necessary to perform a "sifting" process by iterating steps 2–5 (setting $r_{i-1}=\mathrm{IMF}_i$ before each iteration) until this is achieved.
6. Once the IMF_i has a mean value that is sufficiently close to zero over the length of the data (defined by a stopping criterion within some predefined tolerance ε) the residual $r_i(k)=r_{i-1}(k)-\mathrm{IMF}_i(k)$ is computed. Alternatively the sifting procedure can be stopped when the difference in the standard deviation of successive estimates of IMF_i falls below a critical threshold (Huang et al., 1998).

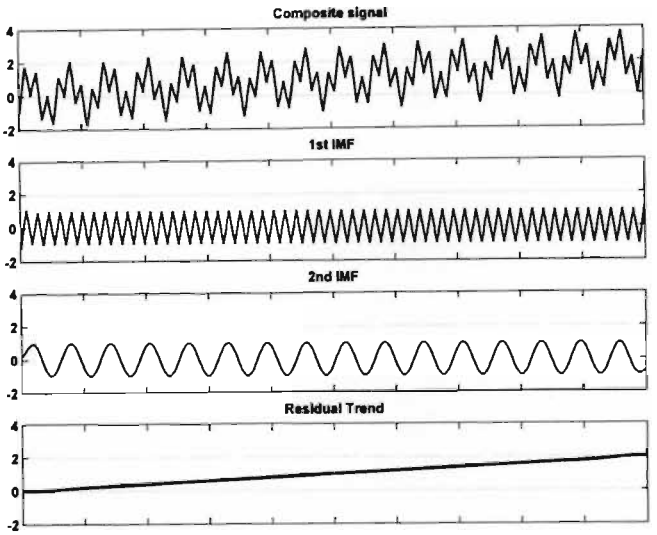


Fig. 4.22: EMD based signal separation; all IMFs are plotted to the same vertical scale. Top panel is the combined signal; lower 3 panels are the decomposition which recaptures, almost exactly, the original components.

7. If the residual $r_i(k)$ is a constant or trend then stop; else increment i and return to step 2.

Figure 4.22 shows the EMD of a composite data series (shown in the first panel) that is the summation of a sine wave, a triangular waveform and a slowly varying trend. The compact representation obtained by EMD extracts (almost perfectly – except near the ends) the three separate data series (shown in panels 2 to 4) that make up the composite signal, without resorting to Fourier or Wavelet techniques with restrictive assumptions about the form of the underlying oscillatory modes (in the form of predefined basis functions). Figure 4.23 shows the analysis of the same data, using Wavelet decomposition. Here a fifth order Daubechies wavelet basis was (arbitrarily) chosen for illustration purposes; this choice of basis function may not be optimal for detrending but serves to demonstrate a typical decomposition. Seven levels of decomposition were required before the trend became apparent; this decomposition is clearly far less compact and physically meaningful than the EMD results in this case.

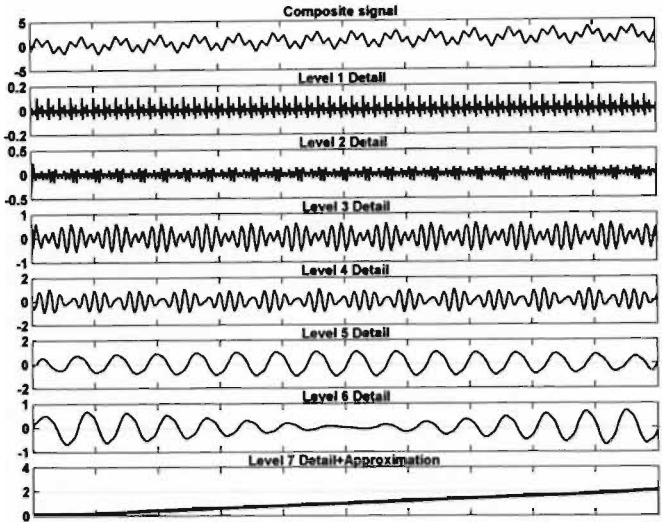


Fig. 4.23: Wavelet based signal separation – The "data" are the same as in figure 4.22, the vertical scale has been compressed for a compact presentation. An arbitrarily chosen db5 wavelet decomposition basis has been used for illustrative purposes.

A similar decomposition analysis can be carried out using Fourier techniques. Figures 4.24 and 4.25 show the result of decomposing the data using a finite Fourier series approximation (see equation 4.5). Figure 4.24 shows the first 5 harmonics; while figure 4.25 shows the series reconstruction by accumulating the lower harmonics up to m. Computing the Euler-Fourier coefficients provides a compact approximation of the original signal (useful for data compression) but fails to extract physically meaningful information. The ability to determine meaningful structural information is clearly important in a nowcasting context, which cannot be bound by the periodicity assumption implicit in Fourier methods.

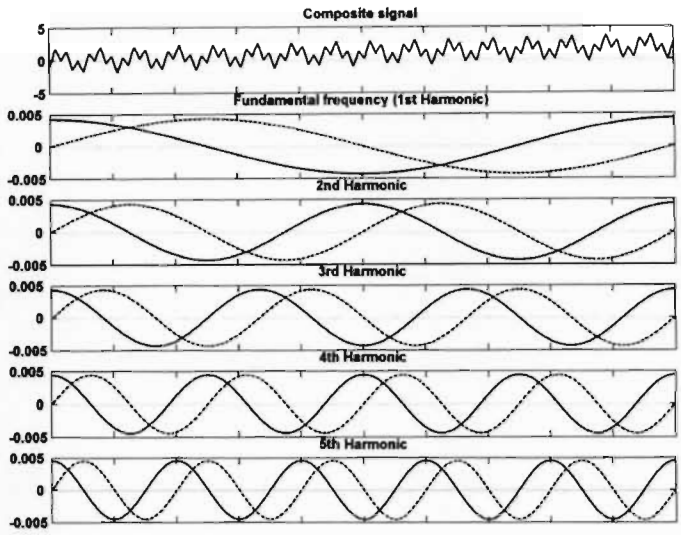


Fig. 4.24: Fourier based signal separation, the first 5 of 75 components – The dashed lines show the sine component and the solid lines the cosine component.

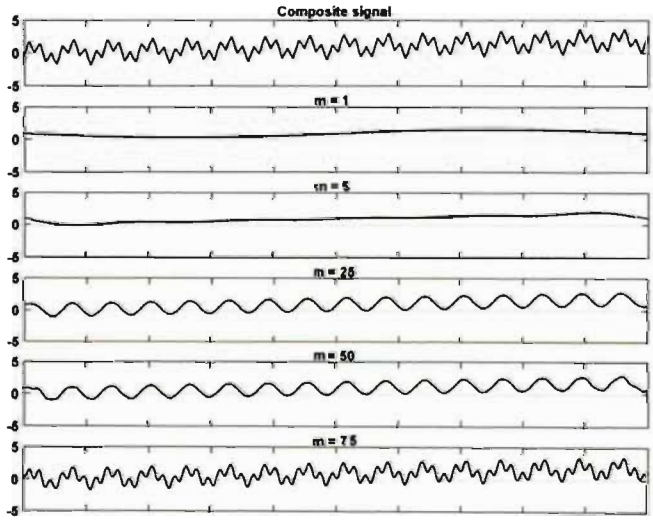


Fig. 4.25: Reconstruction of the signal from the sine and cosine components, m represents the number of Euler-Fourier coefficients used in each reconstruction.

4.2.5 Extending empirical mode decomposition to two dimensions

In two dimensions the EMD process is conceptually the same as for a single dimension, except that the curve fitting exercise becomes one of surface fitting and the identification of extrema becomes (slightly) more complicated. Very little work appears to have been done which applies the EMD technique to two-dimensional data. Han et al. (2002) use EMD in one dimension along four different directions to smooth Synthetic Aperture Radar (SAR) images and remove speckle. Nunes et al. (2003) develop a technique, which they term "Bidimensional Empirical Mode Decomposition" (BEMD) in the context of texture analysis in image data where they demonstrate several examples of intrinsic mode extraction from image data. Linderhed (2002, 2004) examined the use of EMD in two dimensions for image compression. Both of these implementations are very similar to what is proposed here, with the exception that in the rainfall field context there are raining and non-raining areas on the same image. The 2D EMD provides a truly two-dimensional analysis of the intrinsic oscillatory modes inherent in the data. Two-dimensional Fourier and Wavelet analyses are really applications of their one-dimensional counterparts in a number of principal directions. Fourier analysis concentrates on orthogonal "East-West" and "North-South" directions (e.g. Press et al., 1992). Wavelet analysis can, in general, consider any direction of the wavelet relative to the data, however a typical 2D Wavelet analysis examines only horizontal, vertical and diagonal orthonormal wavelet basis functions (Daubechies, 1992, pp. 313; Kumar and Foufoula-Georgiou, 1993). In contrast, EMD produces a fully two-dimensional decomposition of the data, based purely on spatial relationships between the extrema, independent of the orientation of the coordinate system in which the data are viewed.

Description of the algorithm

The algorithm follows intuitively from the one-dimensional case and may be briefly summarized as follows:

1. Locate the extrema in the 2D space including maximal and minimal plateaus.
2. Generate the bounding envelopes using appropriate surface fitting techniques. Conical Multiquadrics are suggested (for reasons explained later).
3. Compute the mean surface function as the average value of the upper and lower envelopes.
4. Determine the first estimate of an IMS by subtracting the mean surface from the data.
5. Iterate until the IMS mean surface function is close to zero everywhere.
6. Estimate the IMS and Residual.
7. If the Residual is a constant or a monotone trend, then stop; else return to step 2.

Surface fitting for extremal envelope generation

The generation of maximal and minimal envelopes is of key importance to a successful 2D EMD implementation and is the most computationally intensive task. The problem is a familiar one of collocating a smooth surface to randomly scattered data points in two-dimensions. There are several options available to achieve this. Ultimately the fitting procedure reduces to computing the unknown value of the surface at a point $s_i=(x_i, y_i)$, by some linear (or nonlinear) weighting of the known data. In general, a basis function determines the influence of each known data point based on its spatial position relative to the unknown point s_i. Nunes et al. (2003) use radial basis functions while Linderhed uses bi-cubic splines (Linderhed, 2002) and later (Linderhed, 2004) chooses the more suitable option of Thin Plate Splines. The choice here is radial basis functions (technically, conical Multiquadrics), which are identical to Kriging with a purely linear semi-variogram model (Borga and Vizzaccaro, 1997). It could perhaps be argued that it would be more appropriate to fit a semi-variogram model to the maxima and minima, but this would be over-elaborate and presumptuous, as the extrema are only related by

distance and cannot be considered drawn from a homogeneous correlated random field. Invoking Occam's razor in the spirit of Huang's original derivation of EMD, the idea is to let the data do the talking and conical Multiquadrics assume the least structure of any linear surface fitting algorithm.

The Ordinary Kriging estimate $\hat{z}_i$ at any point i based on n observed data points is

$$\hat{z}_i = \sum_{k=1}^{n} \lambda_k z_k$$

where z_k are the observations and λ_k are weights associated with each observation and the target point. The mean is assumed unknown and the weights λ_k are constrained to sum to unity. The vector of weights $\boldsymbol{\lambda}$ is obtained by solving the linear system in equation 4.6

$$\begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{u} \\ \mathbf{u}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \mu \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma} \\ 1 \end{bmatrix} \tag{4.6}$$

where $\boldsymbol{\gamma}$ is a vector of semivariogram values, in this application simply defined by the linear distance basis function $\gamma(s_{ij}) = |s_{ij}|$ with s_{ij} the distance between point i and the j=1, 2, ..., n observation locations. $\boldsymbol{\Gamma}$ is the matrix of distances between the observations, $\mathbf{u}$ is a vector of n ones and μ is a Lagrange multiplier ensuring that the Kriging weights $\boldsymbol{\lambda}$ sum to unity, as required. The solution of equation 4.6 is obtained using Singular Value Decomposition (SVD), to ensure that a stable solution is assured (when the matrix is ill conditioned). This is achieved by truncating singular and near-singular components. Although SVD is computationally less efficient than (for example) LU decomposition as a means of solving a dense linear system, it's use is preferred here because of robustness in the face of the near-singular Kriging systems which are frequently encountered in gridded data applications (Wesson and Pegram, 2004).

A more efficient choice of interpolation technique would be useful and more work could be done in this regard, however care is required. Moving-neighbourhood Kriging (a possible alternative to reduce the number of control points) can produce unwanted discontinuities in regions that are data sparse (Chilès and Delfiner,

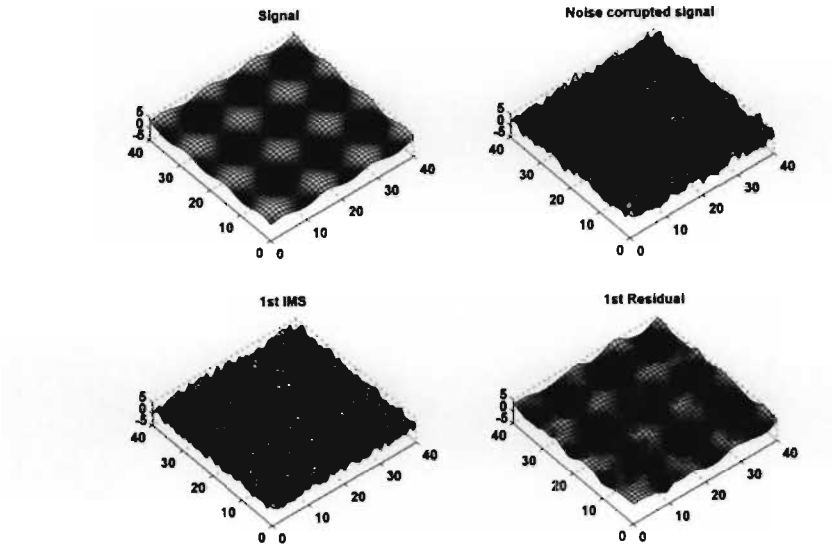


Fig. 4.26: Example of EMD used for noise removal on a 2D sine wave. The bulk of the additive white noise in the corrupted signal is well captured by the first IMS.

1999, pp 201), such discontinuities would be amplified through the EMD sifting process. In addition, the particular choice of Ordinary Kriging as a method of generating the bounding envelopes was (partially) directed by the property that the estimates decay asymptotically to the mean of the observed extrema.

Simple two-dimensional EMD

In this section, applications of the 2D EMD technique are presented. As an artificially constructed example figure 4.26 shows the successful removal of noise added to a synthetically generated two-dimensional sine signal. The noise (with it's high local spatial frequency) is almost completely described by the first IMS leaving a residual, which is closely representative of the underlying signal.

Turning to a more realistic example, figure 4.19 showed an instantaneous radar rainfall field with dimensions 100×200 km. A complete EMD of this field is shown in figure 4.27 using a direct application of the 2D EMD process described in this section; note the change in scale of the individual IMS. The final residual

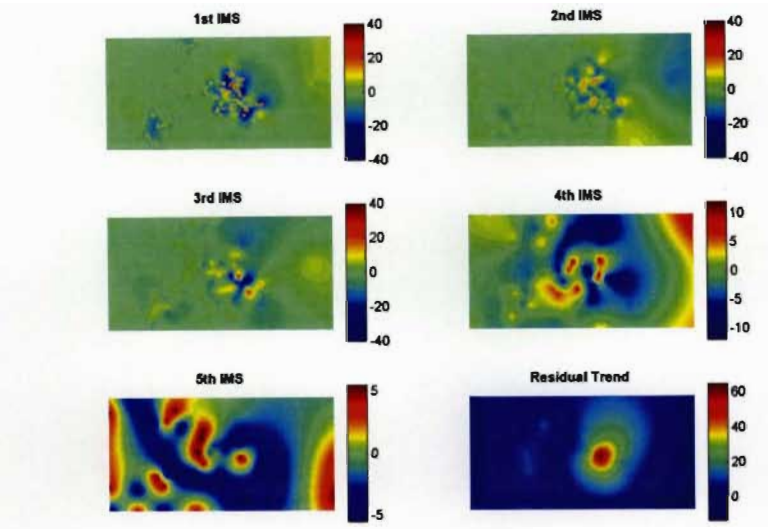


Fig. 4.27: Naïve EMD of the observed rainfall field shown in figure 4.19 – note the change in scale of the rain rates in the IMS.

(with a single extremum) gives a clear indication of the position of the largest convective raincell evident in the field. Co-incidently it turns out that this particular raincell is in fact the feature which persists for the longest time in the observed rainfall sequence.

4.3 Case Study - Application of two dimensional EMD to rainfall data

The simple 2D EMD application presented in the previous section is computationally burdensome when applied to rainfall data. In this section, to overcome this drawback, a number of specific refinements are presented which combine to make EMD tractable in practical real-time situations.

4.3.1 Image processing techniques and optimizations

Since an application of 2D EMD requires the use of surface fitting techniques, large linear systems must be solved. The size of a system is determined by the number of known data points which are to be used in combination to find the unknown values of the surface at each remaining position in the field. The highly variable nature of rainfall data means that the fields may contain a large number of extrema from which the bounding envelopes must be constructed. Additionally there are a large number of zero (no rain) data, which constitute minima. By only considering raining areas, the size of the linear systems requiring a solution are greatly reduced since each raining area (if more than one exists) will contain a considerably smaller number of extrema than the entire data region and each can be treated separately. Furthermore, it makes no sense to consider an EMD in areas where the variable of interest does not exist, in this case the areas with no rain.

A number of well-known image processing techniques are implemented to isolate and process each raining area. Figure 4.28 summarizes the steps taken in processing the data with the boxes numbered 1–7 indicating different steps in the process. First a mask is generated to separate the raining and non-raining pixels (Figure 4.28, Box 1) in the instantaneous radar image; pixels below a threshold of 1 mm/hr are considered as non-raining and the remaining pixels are marked as raining.

An outer boundary border-tracing algorithm (Sonka et al., 1999) is used to establish a boundary "fence" around each raining area (Figure 4.28, Box 2) and a flood-fill procedure is then used to fill each raining area with a unique identifier, resulting in separately labelled raining regions (Figure 4.28, Box 3). To reduce the computational burden of the algorithm even further, the boundary "fence" is decimated by a factor of 5 to reduce the continuous string of border points to "fence posts" while retaining the gross shape of the raining areas (Figure 4.28, Box 4). The next step in the processing of the data is to isolate the extrema in the rainfall field (Figure 4.28, Boxes 5 and 6). There are numerous possible techniques for identifying extrema in the rainfall field. Nunes et al. (2003) use a morphological reconstruction technique. One alternative, which was explored, is based on image

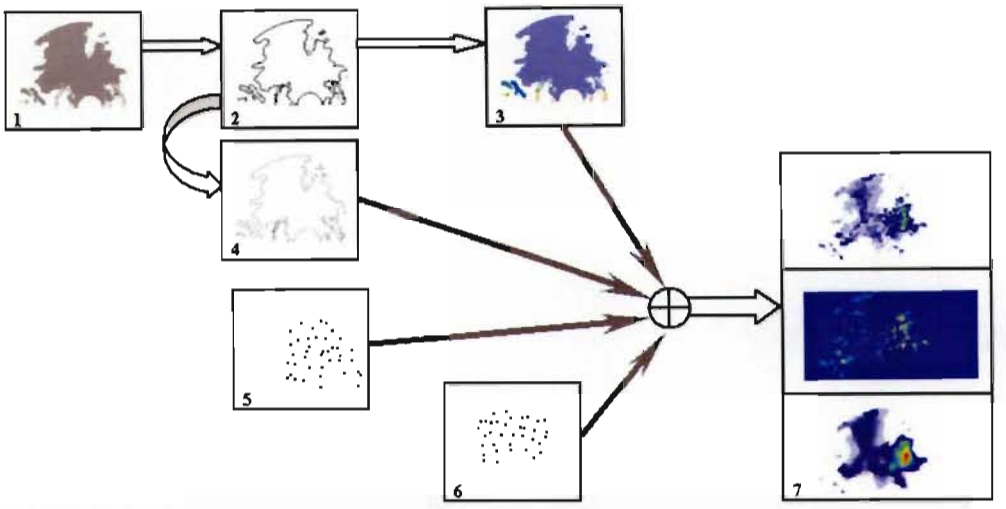


Fig. 4.28: Summary of data processing 1. Mask the wet and dry areas; 2. Trace the boundary of each wet region; 3. Separately label each wet region; 4. Decimate the fence by a factor of 5, isolating the "fence posts"; 5. Isolate the maxima in each sub-region; 6. Isolate the minima in each sub-region; 7. EMD analysis decomposes the data into the first IMS and the first residual using the maximal and minimal envelopes defined using the points in 4, 5 and 6.

segmentation and detection of extremal plateaus. However, the chosen method was to use a simple 8 neighbour search routine for identification of pixels with extreme values as done by Linderhed (2004). The choice was partly for convenience and computational speed, but also because the majority of the (non-zero) extreme values in the rainfall fields studied turned out to consist of single pixels. There is a rich literature on image processing techniques and an introductory text such as Sonka et al. (1999) will provide the detail omitted here. Finally, the EMD analysis is carried out using the extrema within each raining area and the zeros at the "fence posts" of non-raining border pixels to specify the extremal envelopes (Figure 4.28, Box 7). Only one step of decomposition is shown – the data is decomposed into the noisy first IMS and the first residual.

4.3.2 Results

An analysis of over 800 individual radar scans, embodying mixtures of various ratios of Stratiform and Convective rainfall types, was carried out to determine the effectiveness of the 2D EMD algorithm in separating the high wavenumber spatial components from the low wavenumber components of the original rainfall data. Working on the basis that the average characteristics of the data over a range of spatial scales summarized by the power spectrum is intuitively useful, the (radially averaged) power spectra of *(i)* the original data, *(ii)* the first IMS and *(iii)* the first residual of each image were examined and compared. Figure 4.29 shows a typical result; the power spectrum of the residual shows a very close correspondence with that of the original data at large wavelengths while it contains far less power at the shorter wavelengths (note the logarithmic scale on both axes of the figure). In contrast, the spectrum of the first IMS has very little power relative to the data's spectrum at high wavelengths but shows a strong correspondence at the lowest wavelengths. Figure 4.29 clearly indicates how the 2D EMD technique moves the bulk of the high frequency components in the original data into the first IMS and leaves the high power, lower frequencies in the first residual. The decomposition behaves as a low pass spatial filter, without presupposing the shape of the filter function. Figure 4.30 shows a time average of this behaviour by plotting the mean values at each wavelength of the three spectra over five consecutive radar scans (beginning with the data used to produce figure 4.29). The radar scans are captured at approximately five-minute intervals. It is interesting to observe that the average of the spectra of the first IMS is flat (constant mean) for wavelengths longer than 10 km, suggesting nearly white noise over this range.

The temporal persistence exhibited at the spatial scales represented in each of the three sequences of: *(i)* the data, *(ii)* the first IMS and *(iii)* the first residual was examined by considering the persistence of their temporally consecutive power spectra. The notion of "spectral persistence" was used to determine how variable the spatial structure (at a particular spatial scale) is in time and hence to give an indication of the temporal predictive capability at each spatial scale. A summarized example of the analysis of a sequence of 5 radar rainfall images is presented

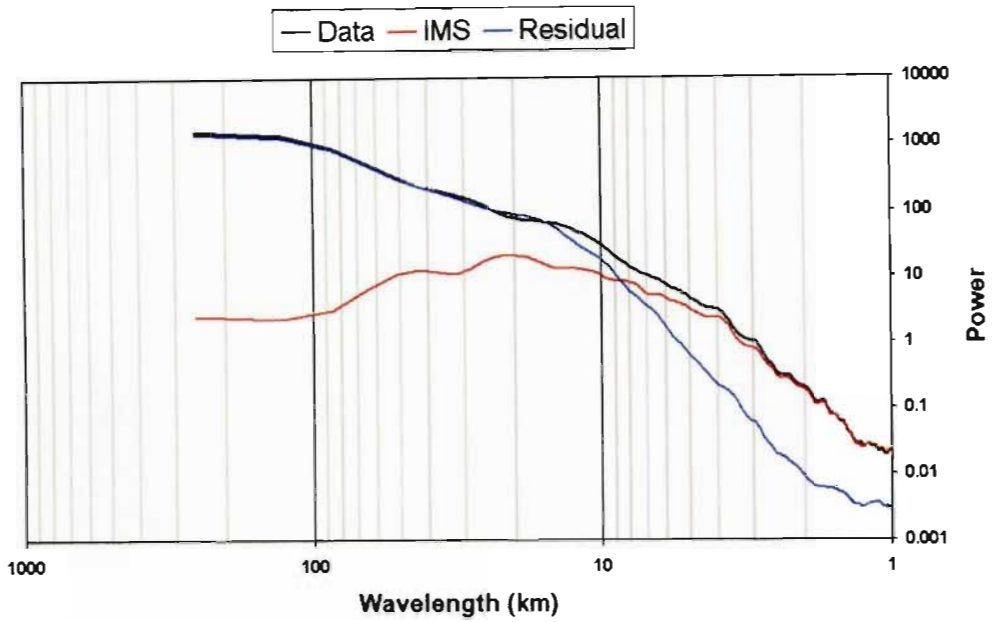


Fig. 4.29: Comparison of individual radially-averaged power spectra of the radar rainfall data (of figure 4.19) with its EMD components: the first IMS and the first residual. Note the logarithmic scale on both axes.

in figures 4.31, 4.32 and 4.33 where a "matrix" of scatter plots is shown in each case. Scatter-plots of the pairs of power values at each discrete wavelength for five consecutive spectra (with the 1:1 line indicated) are shown for the original data (Figure 4.31), the first IMS (Figure 4.32), and the first residual (Figure 4.33). The rows and columns of the scatter-plot matrices are labelled from T_0 to T_4 and indicate separate radar scans between time T=0 and time T=4. Each block in the scatter-plot matrix represents a scatter-plot of the power at each wavelength for the spectrum computed at T_i versus that of the spectrum computed at T_j. Clearly the plots on the "matrix" diagonal each compare a spectrum to itself and a perfect 1:1 relationship is observed in this case. For the off-diagonal plots, the degree of scatter amongst the data points indicates the degree of similarity between the spectra at individual wavelengths at increasing time lags with a large scatter indicating a weak similarity. The trends shown here are typical of the data analysed and show how the first (high average wavenumber) IMS has a temporally incoherent spatial structure, while the first (low average wavenumber) residual shows

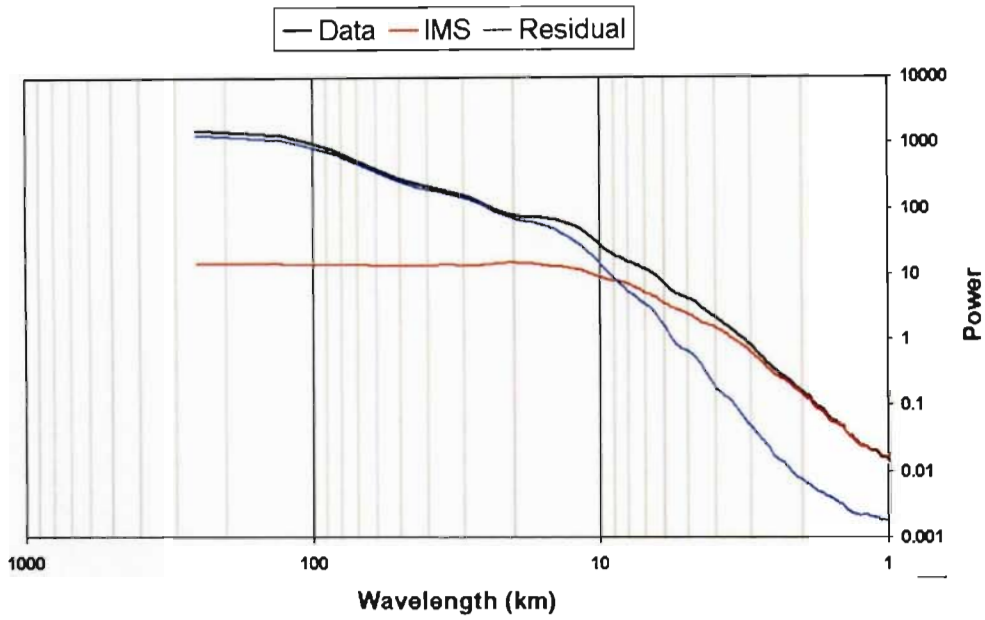


Fig. 4.30: The same as figure 4.29 but for the mean of individual power spectra for five consecutive, radar scans – Beginning with the spectra shown in figure 4.29. Note the logarithmic scale on both axes.

a temporally consistent structure. The behaviour shown in Figures 4.29 - 4.33 suggests that the high frequency IMS components in spatial rainfall data do not contain much predictive capability. This observation supports the suggestions of Seed (2003) and Turner et al. (2004) who propose to increase the degree of spatial smoothing and give more credibility to the information contained in the lower frequency components as forecast lead times increase.

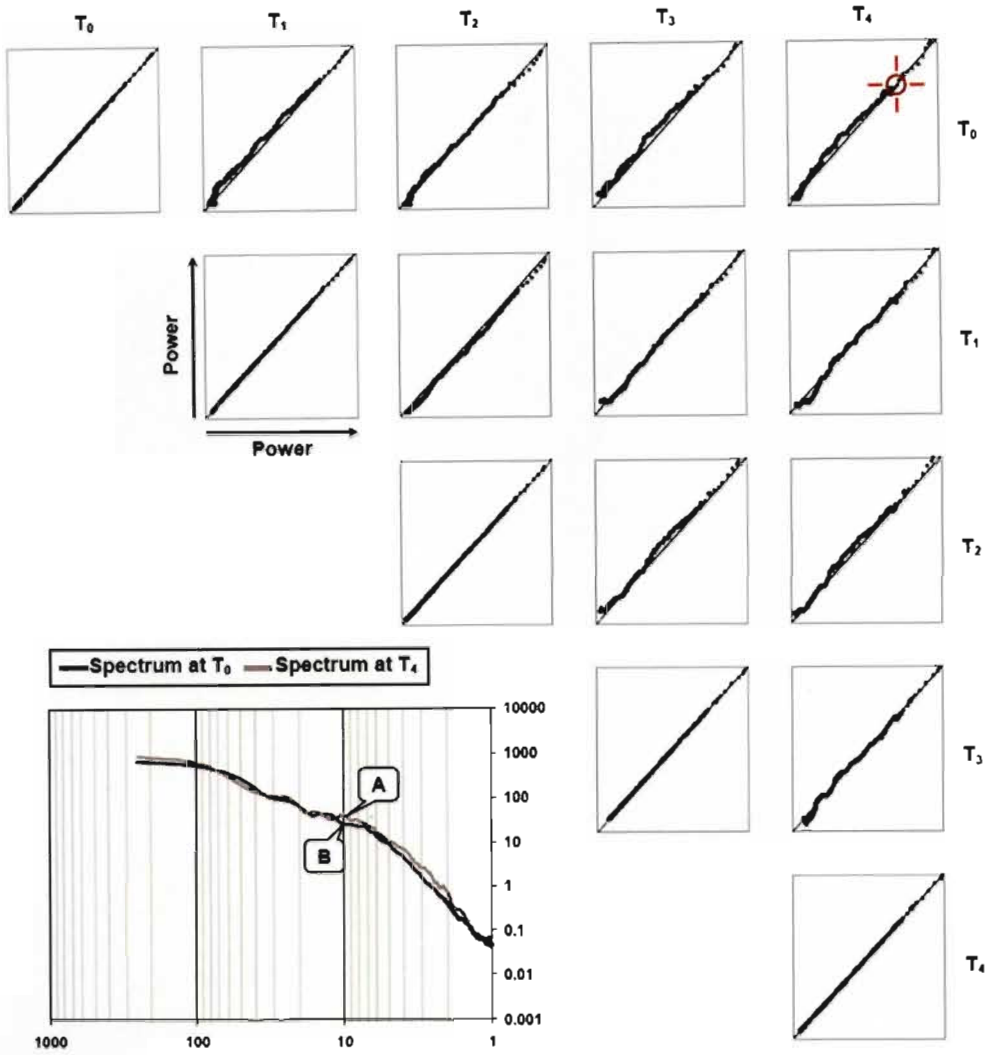


Fig. 4.31: Spectral persistence scatter plots of the original data for a sequence of rainfall fields and those at successive intervals. This is constructed by plotting the values of power for each field at corresponding wavelengths coaxially. For example points A and B at the 10 km wavelength are plotted against each other and appear ringed in the upper right diagram.

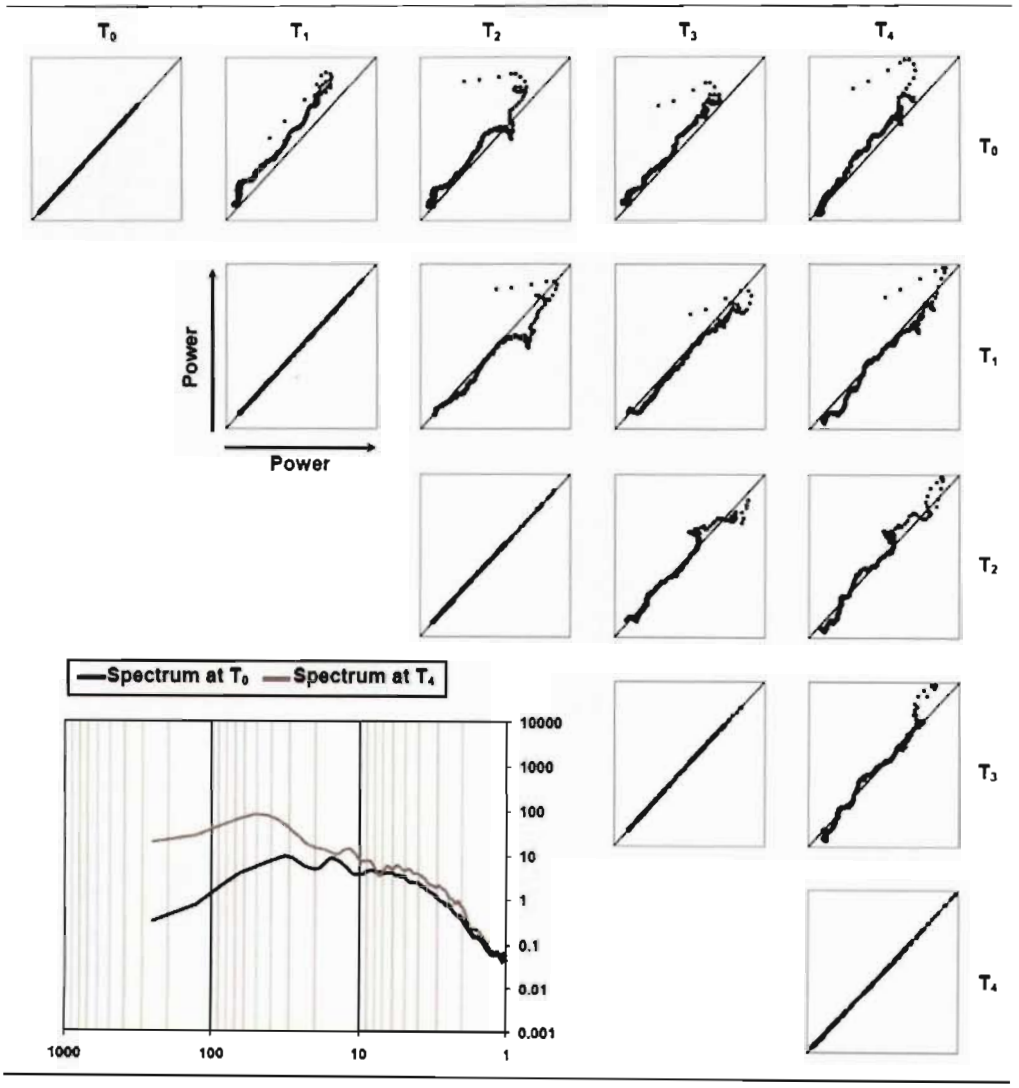


Fig. 4.32: Spectral persistence scatter plots of the sequence of 1st IMS of each pair of rainfall fields $T_0, \cdots, T_4$.

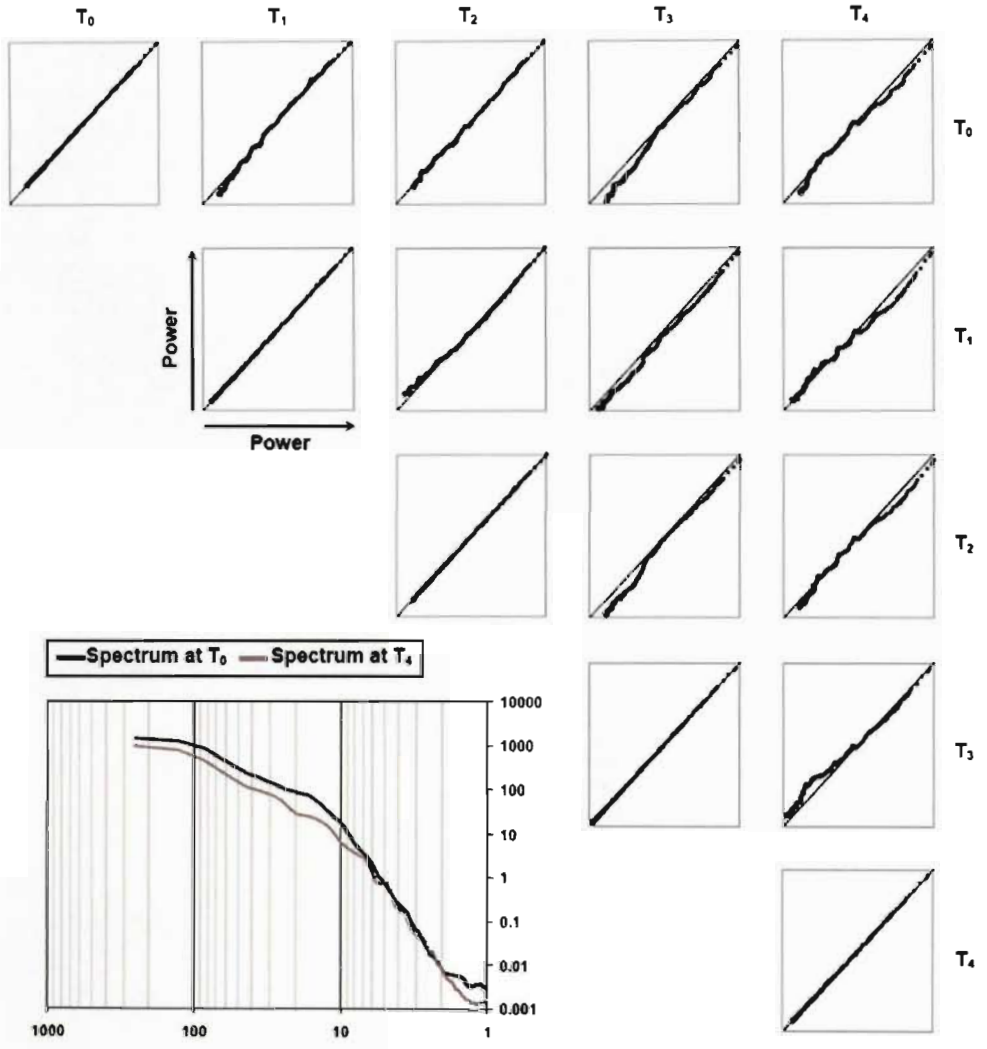


Fig. 4.33: Spectral persistence scatter plots of the sequence of 1st Residual of each pair of rainfall fields $T_0, \cdots, T_4$.

Chapter Summary: This chapter began with a brief overview of some of the nowcasting techniques and models in the literature. This was followed by a discussion of the Optical Flow method as a technique for determining the advection of rainfall fields measured by radar or satellite. This technique is a core component of the S-PROG and SBM stochastic nowcasting models, as well as the accumulation scheme presented in section 3.3.

The SBM rainfall simulation model (Pegram and Clothier, 1999) was described and extended to provide a nowcasting implementation. The differences between the simulation and nowcasting modes were highlighted. In addition, a well known nowcasting model called S-PROG (Seed, 2001) was described and it's main features discussed.

With the nowcasting models introduced some comparitive investigations were carried out using observed radar reflectivity data from two South African radars. It turns out that S-PROG's strategy of decomposing the observed rainfall fields according to pre-specified spatial scales gave it a performance advantage when compared with SBM in nowcasting mode for two mean field error statistics.

With the results of the nowcast comparisons in mind, a new technique for analysing the spatial scaling structure of rainfall fields was presented. The technique is a two dimensional extension of Empirical Mode Decomposition for the analysis of non-linear and non-homogeneous time series. An EMD analysis in two dimensions linearly decomposes the spatially distributed rainfall data into a set of Intrinsic Mode Surfaces, which are approximately mutually orthogonal (Huang et al., 1998) and sum back to the original data. Each IMS contains an oscillatory mode inherent in the data at a different (narrow) range of spatial frequencies. The EMD analysis successively extracts the IMS with the highest local wavenumbers in a recursive way, which is effectively a set of successive low-pass spatial filters based entirely on the properties exhibited by the data. The utility of the EMD technique for signal separation has been demonstrated in both one and two dimensions and applied to the analysis of a large set of 800 radar rainfall images in South Africa. The 2D EMD technique is proposed here in the context of rainfall nowcasting to separate the less persistent high wavenumber components from the

more persistent low wavenumber ones in the data. The aim is to isolate the high wavenumber components, which do not exhibit a strong temporal correlation and add little structural information to nowcasting algorithms. The scale separation achieved by 2D EMD has been analysed using radially averaged power spectra to summarize the spatial structure of the data and filter outputs. In addition these power spectra have also been used to examine the temporal persistence of the spatial structure exhibited by the first IMS and residual. The results presented here support other work in the Hydrometeorological literature, which suggests that the low frequency spatial components in rainfall data are most useful in a nowcasting context.

CHAPTER 5

COMBINING THE METHODOLOGIES FOR FLASH FLOOD FORECASTING

Earlier chapters in this thesis have presented techniques for overcoming a number the challenges presented in relation to flash flood forecasting. In this chapter the various methodologies are drawn together into a coherent system to provide a road-map for the implementation of a flash flood forecasting system in South Africa. Particular emphasis is given to the organizational structures which will be required to implement a flood forecasting system. Acquisition, processing and transfer of data are discussed in terms of a prototype system in the eThekwini municipality, Durban, South Africa and shortcomings noted for future improvements.

In section 1.2 a brief discussion was presented relating to a proposed system for flood forecasting in South Africa. With the exception of a model for portions of the Orange river and Vaal dam catchment maintained by the Department of Water Affairs and Forestry (DWAF) hydrology group, there are no operational flood warning systems anywhere in South Africa. This is a dire situation in a country where so many are impoverished and without the means to recover from disasters. The Vaal system model uses the Sacremento rainfall runoff model calibrated for the region and rainfall data is manually fed into the model during periods of heavy rain. This model is mainly intended to assist with the management of Water Resources and was not designed with disaster management in mind. The model is suitable for the large Vaal catchment but it is not automated and would be cumbersome to run on smaller catchments with shorter response times. In addition

skilled personal to man the system are in short supply in South Africa. This critical personnel situation will need to receive urgent attention.

Sinclair and Pegram (2004a) provide the details of a prototype system set-up for the Mgeni and Mlazi catchments near Durban, South Africa. The implementation provides several of the components of a flood forecasting system shown in figure 5.1 and these are described, along with suggested ways to "complete the puzzle" by providing the remaining components of the system.

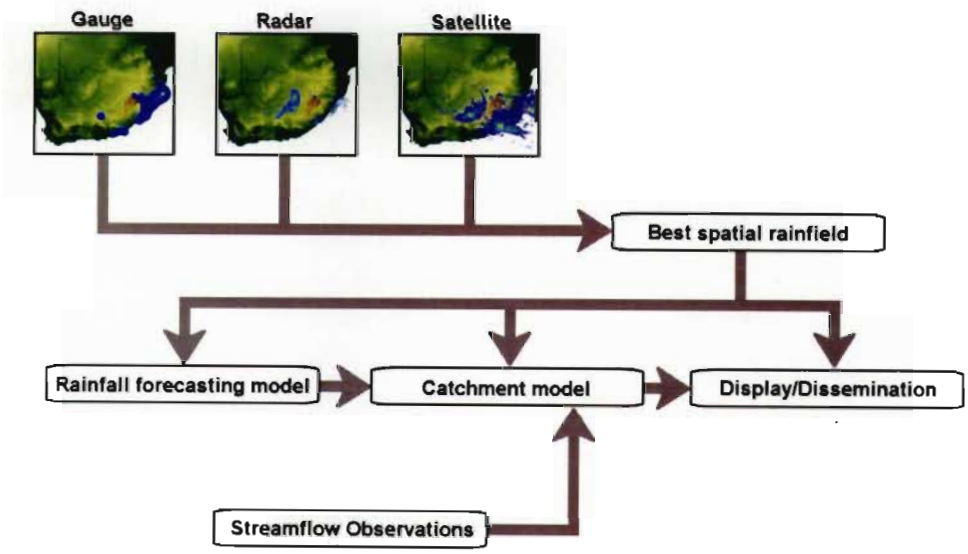


Fig. 5.1: A schematic overview of the main components required for a successful flood forecasting system. The "best" option for each of the components is dependant on factors such as the size (hence response time) of the catchment, available data and skills within the team tasked with implementing the system.

5.1 Rainfall estimation

The most important input to any flash flood forecasting system is precipitation. In Southern Africa the influence of snow can safely be ignored (except in a few select

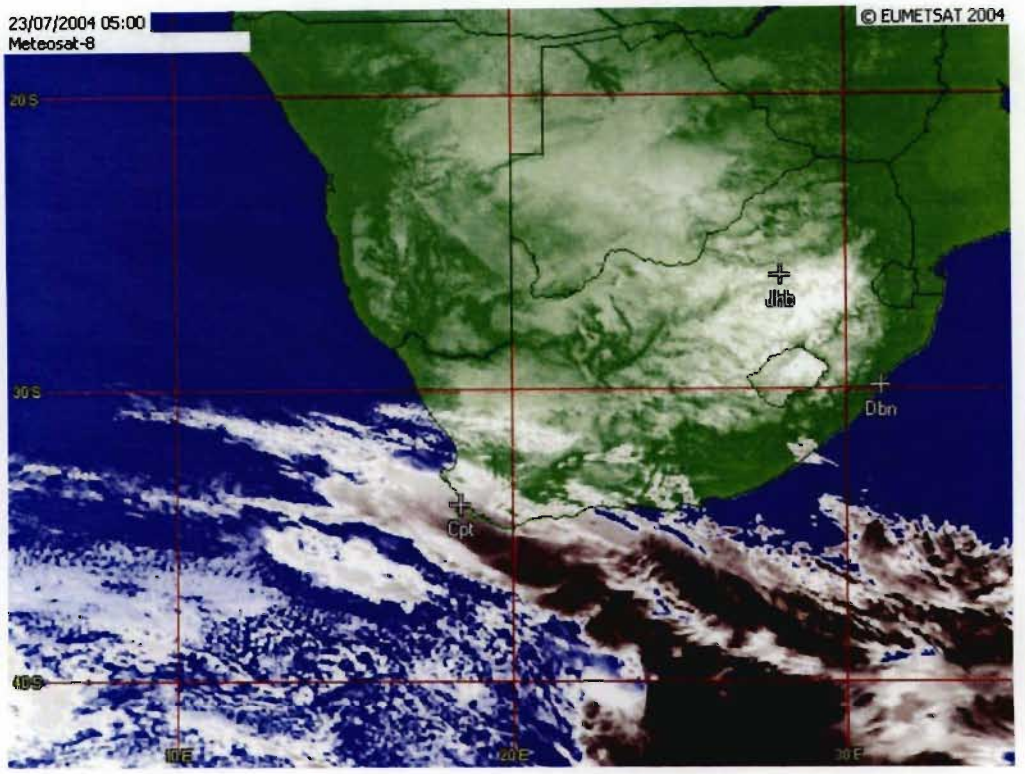


Fig. 5.2: 23 July 2004 Met-8 false colour image. A typical frontal system over the Western Cape Peninsula, South Africa.

high areas) and the measurement of rainfall becomes the most important factor in determining the input to the catchment. As described in detail in chapter 3, there are three measurement devices available in South Africa which provide estimates of rainfall at suitable spatial and temporal resolution for flash flood forecasting. In the short term any flood forecasting system will be forced to rely on these.

The rainfall estimates produced from rain gauges, weather radar and meteorological satellite can be combined in an optimal way using conditional merging to produce the best spatial estimate of rainfall for the catchment of interest. The resulting combined rainfall estimate can be fed into display systems for direct visualization of instantaneous and accumulated rainfall (section 5.5.1). Figure 5.2 shows an example of the kind of useful visual information that may be provided in real time via remote sensing data. The estimates are also used as input to catch-

ment models and converted to streamflow, the core output of the flood forecasting system. It is essential that processes are put in place to produce rainfall estimates in real-time using the most current available data streams.

The responsibility for this will ultimately lie with SAWS, but a countrywide implementation at the short time-scales relevant to flash flood forecasting is still some way off. The most pragmatic approach at this juncture will be site specific implementations of the appropriate merging algorithms. These specific implementations will direct the rollout to a wider area and provide operational experience with the algorithms and data systems.

5.2 Catchment model

The nature of the catchment model is not dictated here as its implementation and effective use are (almost entirely) dependent on the expertise and data which are available when the forecasting system is implemented. In South Africa the necessary hydrological expertise does not exist in many of the municipal structures which are responsible (by law, Disaster Management Act 2002) for ensuring appropriate flood mitigation strategies, including flood forecasting and warning systems. For useful real-time operation, the catchment models need to be informed by real-time streamflow observations (Section 5.4) which provide a means of updating the models performance and improving forecasts. The concepts related to model updating are discussed in chapter 2.

Models should also be suitable for automation, or be available in a form which exposes a suitable Application Programming Interface (API) for systems development. Once again, the level of sophistication in the system implementation will be highly dependent on a combination of funding and expertise.

Most of the larger Metros have by now completed the process of producing flood vulnerability assessments for the areas where they have responsibility. A key component in the vulnerability assesment is the production of floodlines for multiple return periods. In some cases this may have entailed Hydrological modelling. In these cases it may be possible to re-use the existing model if the relevant

data related to its set-up and calibration can be retrieved from the consultants hired to produce the floodlines.

5.3 Rainfall forecasts

Nowcasting of rainfall fields holds considerable promise for improving the lead-time of streamflow forecasts. While accurate quantitative estimates remain troublesome using radar and satellite information, properly forecasting the arrival times of runoff producing rainfall will likely provide some assistance to the Hydrological forecasting efforts. Short term rainfall nowcasting is dealt with in chapter 4, where two stochastic nowcasting models are investigated as a means for extending the information from current best spatial rainfall fields into the future. Operational implementation of such schemes in South Africa falls within the purview of SAWS. The nowcasting schemes will be well complemented by the outputs of NWP models, which provide advance warning of heavy falls.

5.4 Real-time streamflow observations

Real-time streamflow observations, upstream of vulnerable areas provide many advantages for a flood forecasting system. If a Hydraulic model has been set up for the river reaches downstream of the gauging station it is possible to get an estimate of the flood levels in real-time. This is possible either through a lookup system, or by running the Hydraulic model online. This is discussed in more detail in section 5.5.2. The Hydraulic modelling effort becomes far more useful if forecasts of streamflow can be used as input. Feedback to improve Hydrological model outputs using filtering techniques (e.g. Kalman filters) is an important way to produce better forecasts and therefore real-time streamflow observations hold great importance.

The current situation is South Africa is that DWAF has instrumented a large number of flow gauging structures with telemetering flow gauges. The data are recorded at short time intervals (12 minutes in most cases) and transmitted to a

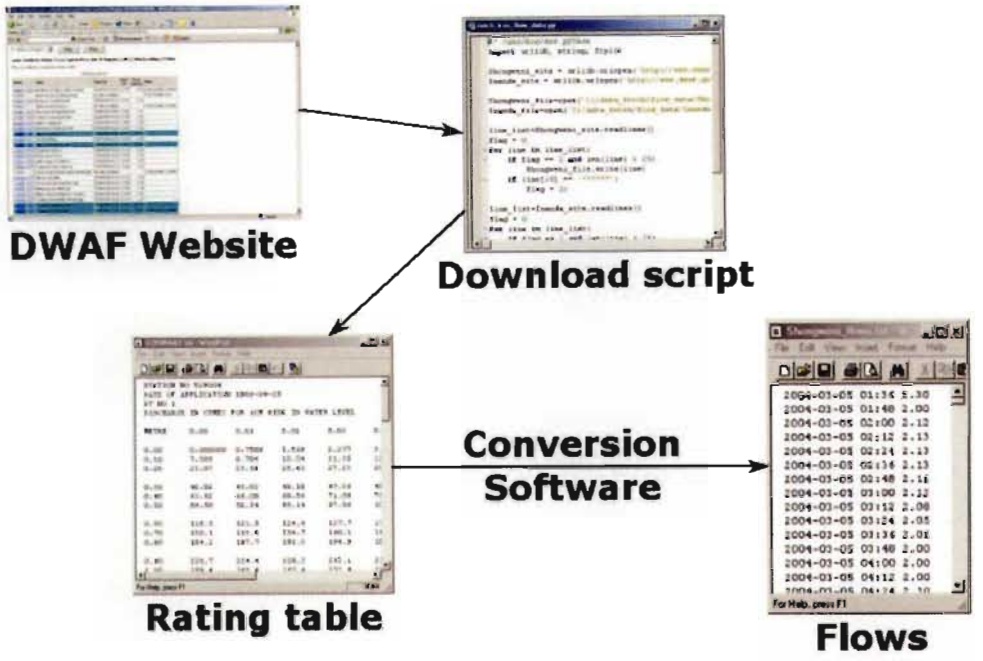


Fig. 5.3: An outline of the procedure for downloading and extracting real-time flow records from the DWAF website. The system can be set up to run at regular intervals, depending on the application.

central server with a frequency of five times daily (on average). The frequency of data transmissions is limited by cost and increased cost can only be justified by increased demand. In times of high flows the data could conceivably be requested on demand at a higher temporal frequency. The data are freely accessible via DWAF's website and returned in an HTML format as a rolling buffer of the most recent data. Quality controlled historical records for each gauge are also available through a similar interface. The author (Sinclair and Pegram, 2004a) developed software to request the web page containing data for a relevant gauge, read and parse the page and extract the relevant flow (or stage) data. Where dams are used as the gauging structure (and flow rates are not provided by DWAF), the author developed additional software to convert the dam levels to flows via the appropriate rating table. Figure 5.3 gives a diagrammatic representation of the data retrieval process.

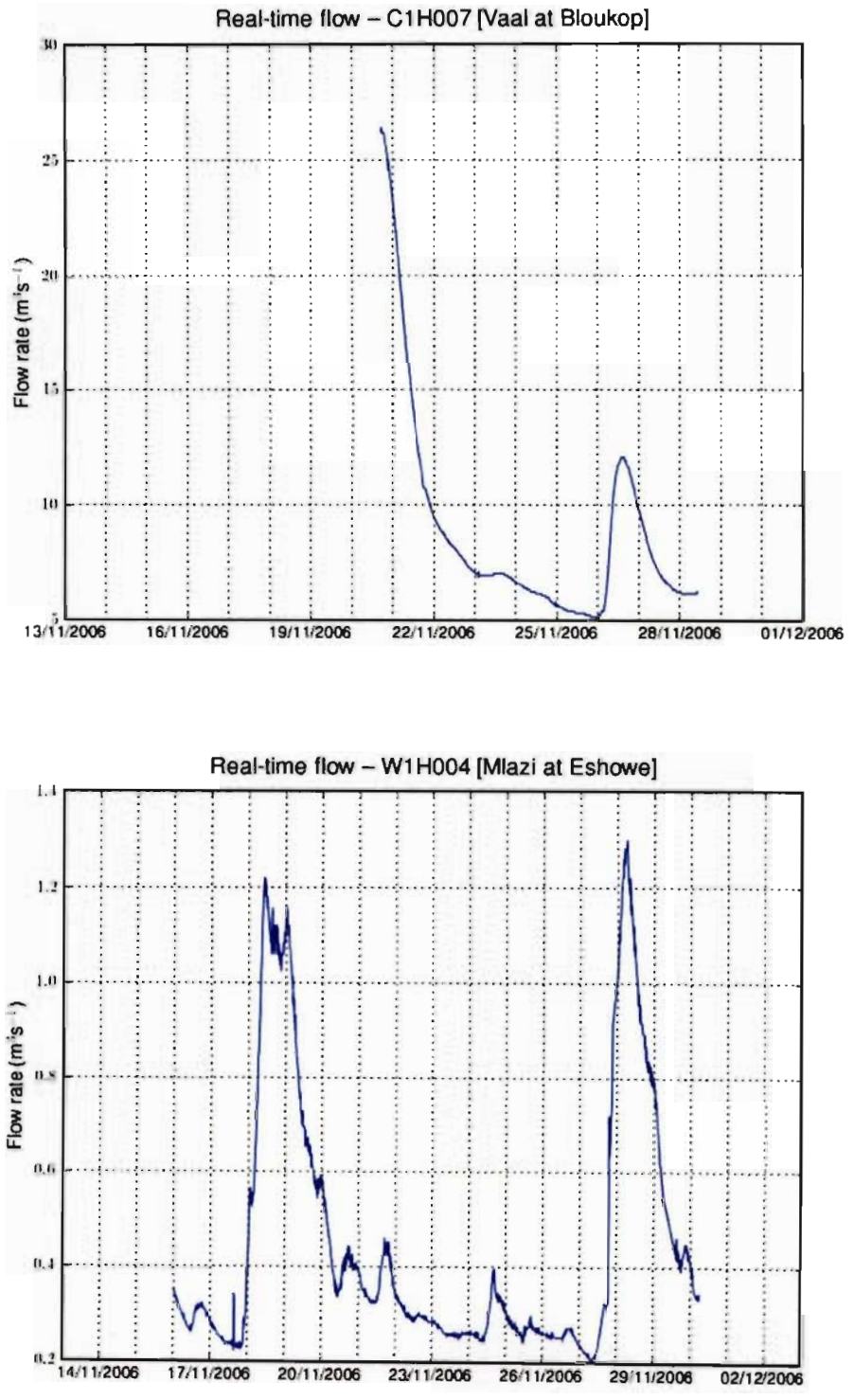


Fig. 5.4: Examples of automatically downloaded real-time flow records. The data have been downloaded from the DWAF website and converted into graphical form. The software is set-up to run at regular intervals, which can be varied depending on the specific application.

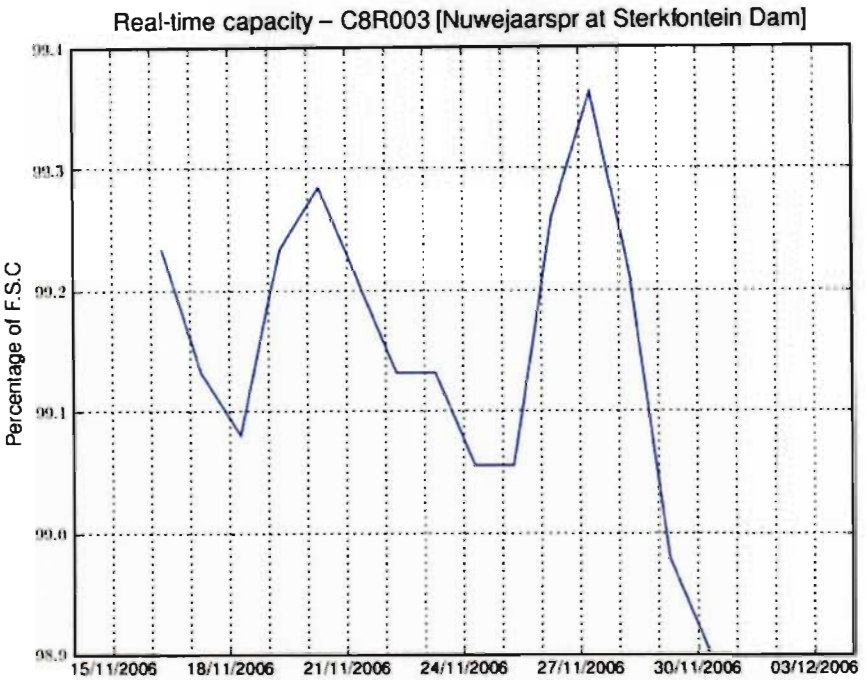


Fig. 5.5: Example of an automatically downloaded real-time reservoir level record. The data have been downloaded from the DWAF website and converted into graphical form. The software is set-up to run at regular intervals, which can be varied depending on the specific application.

The resulting flow record can be used to provide information about current flow conditions as well as feedback (through a suitable filter) for the catchment model. This system has been successfully implemented and run at UKZN and in the eThekwini municipality's disaster management centre. Since the software is easily scheduled to run at any time interval, improved frequency of data collection from gauging stations on selected catchments will make this a reasonably robust and effective method for obtaining real-time streamflow estimates. Figures 5.4 and 5.5 show recent examples of the data collected automatically using the software.

5.5 Visualizing the output

In this section the kind of information presented to the end user of the flood forecasting system, in this case the disaster managers (DMs) is discussed. The products which are relevant have been informed by consultation with the members of the disaster management team at eThekwini Metro and other role players such as SAWS and DWAF. The role of the flood forecasting system in the process of disaster management is to provide a warning of the possibility and seriousness of an impending flood event, the actions and decisions taken on the basis of that information are then in accordance with the action plans of the relevant authorities. The consultation process yielded two major requirements from the DMs point of view; some warning should be provided ahead of a likely flood and the affected areas should be clearly defined (preferably in a graphical format). To meet these requirements, the Arcview GIS was chosen as a display tool since there is already considerable skill and experience within the major Metros in South Africa with regard to this system. Images of current and historical rainfall were made available for real-time display in the eThekwini Metro disaster management control room and a system devised for the dynamic selection of flood lines in response to currently observed or forecast streamflows. These systems are discussed in more detail in sections 5.5.1 and 5.5.2.

5.5.1 Integration of the radar rainfall images into a GIS

When radar rainfall images are presented in an animated spatial context they can provide a simple means for providing advance warning of heavy rainfall approaching sensitive (flood-prone) catchments. Figure 5.6 shows an example of an instantaneous radar image of a major storm event that moved over the Mlazi catchment and the lower reaches of the Mgeni catchment.

The image data were made available in an Arcview compatible format at the eThekwini municipality's disaster management centre in near real-time. The data transfer process and relevant software were developed by the author in consultation with the DMs and SAWS. A brief description of the transfer methodology is

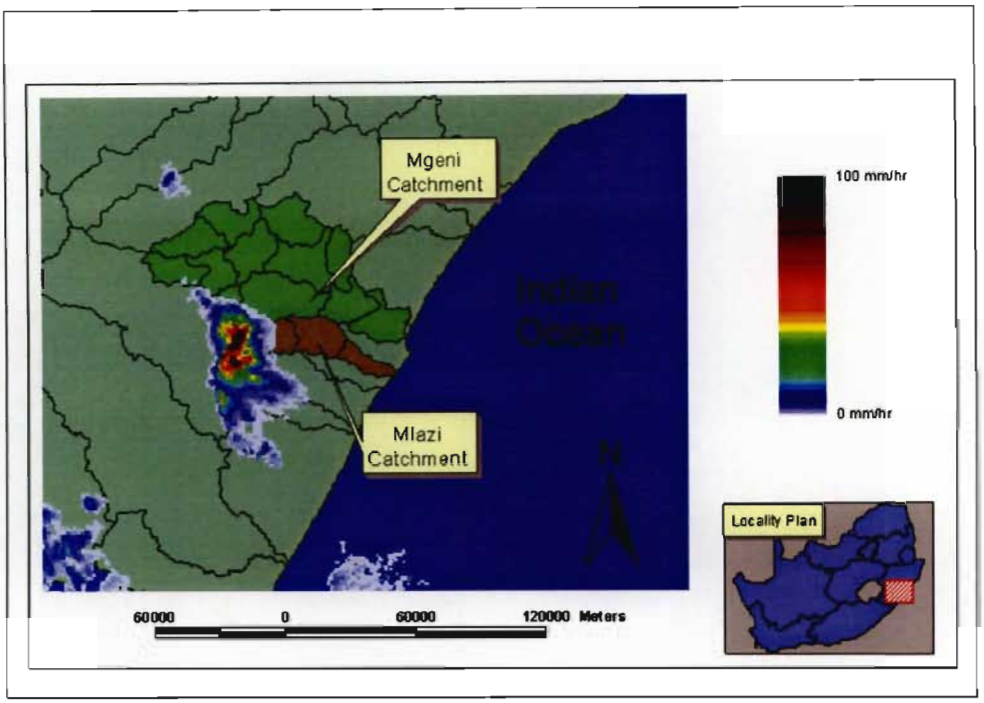


Fig. 5.6: Visualization of rainfall in the Disaster Management centre. This figure shows a large convective storm over the upper Mlazi catchment (17 December 2002 at 14:58 SA standard time).

presented in the following paragraphs.

SAWS make use of the Meteorological Data Volume (MDV) file format to transfer and archive spatial radar data. In order to make use of this primary dataset it is necessary to have a means of extracting the data from MDV files.

An overview of the MDV file format

MDV is a flexible binary file format designed for the storage and transfer of gridded meteorological data. The SAWS radar data processing system is built around the MDV format and uses the TITAN system extensively. The main advantages of the MDV format are the fast data access speeds that result from it's binary nature, portability of the format between operating systems (the libraries handle issues such as Endianness) and the possibilities for data compression included as part

of the format. Each MDV file consists of a file header which defines the layout and contents of the data portion which follows. The data sets are organized as a hierarchy of "Fields" each field may (for example) represent a different variable. Within each field there is the capacity to store a number of "Levels". The structure of the format is designed to be appropriate for storing data sets collected at various levels relative to the sensor.

The South African radar data is stored in a single MDV file for each instantaneous scan. The raw radar data is resampled onto a stack of 18 Cartesian grids with a 1×1 km grid cell size and 1 km vertical discretisation.

Using the MDV data

The C language source code for MDV file handling was obtained from METSYS and adapted to compile under Microsoft Windows using the MSVC compiler. A set of statically linked libraries was produced. Linking to these libraries and calling the routines contained therein allows the user to make use of the MDV file handling routines in custom software used to extract and manipulate the MDV data. A full understanding of the MDV routines has been achieved and they are now an integral part of many of the software products developed within the research group at UKZN.

The raw radar data-stream is processed into MDV format at each of the 11 radar sites around the country (Figure 5.7). The MDV data are then transferred to the METSYS server in Bethlehem where SAWS publishes data products on their web page and archives the data in real-time. The relatively static products currently available on the web page do not provide the powerful data analysis capabilities of a GIS. Software developed during this project was run routinely on SAWS' METSYS server to produce ArcGIS compatible geolocated images from the Durban radar. The software accepts the MDV format data as input and produces radar rainfall images in a spatially referenced format. The data are compressed to reduce the size and stored on the METSYS server in a rolling buffer with 2 days of data stored before deletion. There is no need to preserve these data as the products can readily be extracted from the archived MDV data if required.

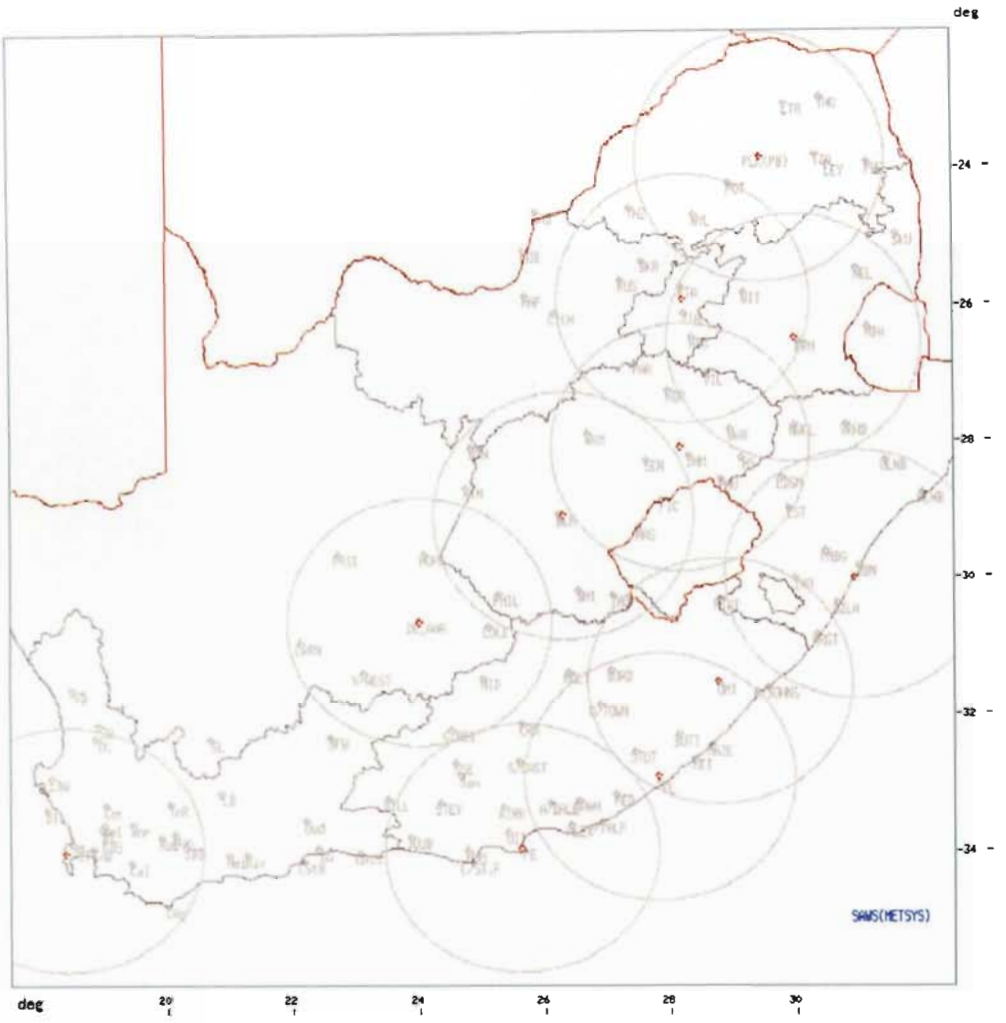


Fig. 5.7: The South African weather radar network. The location of each of the 11 radars is indicated by the red points and the range is indicated by the rings which surround the points. The range shown in this figure is the 200 km range, some of the radars (notably Cape Town) have an increased range.

Software developed in this project using the Python language was tested for several months at the Durban Metro DM offices to retrieve the latest images from METSYS. The software makes use of the File Transfer Protocol (FTP) to access the METSYS server remotely via the Internet at five to ten minute intervals. The METSYS FTP server is queried for new data at frequent intervals and when new data are found they are downloaded to the client machine; data that reside on the

client but are no longer on the server are automatically deleted (but could be stored if this was required).

5.5.2 Generation and display of floodlines and inundation depth

It's all very well having a sequence of observed streamflows and possible forecasts of the future flows but without these numbers being interpreted in terms of flood levels, and depths of inundation, they are of no real significance to the DMs. The process of translating flows in the Mlazi river into inundation depths for the eThekwini Municipality is dealt with in Mkwananzi and Pegram (2004). The details of the methods used for the determination of flood lines and inundation depths will not be repeated here as they are well known. Instead a brief description of the process and the application of these products in the prototype flood forecasting system are presented.

In general the computation of floodlines is a fairly laborious process which, despite the proliferation of software packages, requires a large degree of human interaction. This does not encourage online computation of current and forecast inundation depths as a first choice of modus operandi.

A possible process for producing floodlines is as follows, based on the Mlazi river study. A Hydraulic model of the river channel and adjacent flood plains must be produced, using a Digital Elevation Model (DEM) in combination with field surveys. This is a once-off procedure since the same model may be used to model the effects of many different flow rates (this assumes that there are no disturbances to the geometry of the channel, which would require revising the model). The hydraulic model used in this study was the well-known HEC-RAS model (Brunner, 2001). HEC-RAS routes the flood-wave along the modelled channel and computes flood levels at each of the modelled cross-sections. A typical output from HEC-RAS (steady state) is shown in figure 5.8. The model can also output the flood levels (to a file) as a set of points in a three-dimensional co-ordinate system. In order to interpret these points as flood lines and inundation levels an interpolation between channel cross-sections is required, preferably in conjunction with additional information from a DEM. A polygon can then be produced in a suit-

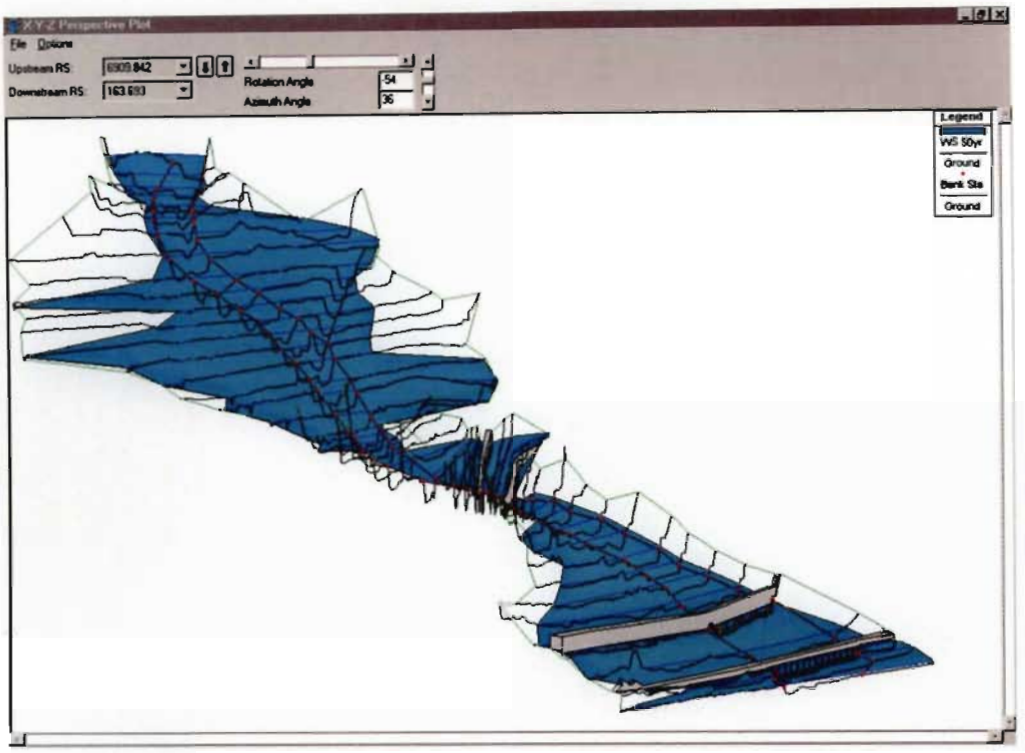


Fig. 5.8: HEC-RAS output for the lower reaches of the Mgeni river, Durban, South Africa.

able GIS format for viewing in the disaster management centre. In this case an Arcview shapefile like that show in figure 5.9 needs to be produced. A separate floodline must be produced for each flow rate considered. In typical flood hazard studies the flow rates are chosen by recurrence interval.

The main issue with automating this process is that the DEM terrain data have limited precision in following the topography and the resulting flood maps may not be entirely realistic. The problem can be partially resolved using improved resolution for the DEM data and suitable mapping software. This approach is still not foolproof in automation and would require considerable technical skill (and therefore expense) to set up.

For the purposes of this study a simple alternative solution was proposed as a starting point. Flood lines at various recurrence intervals are available for the Mgeni and Mlazi rivers through studies commissioned by the eThekwini munici-

Fig. 5.9: A 2% Annual exceedance probability (50 year Recurrence Interval) flood line: (Mgeni catchment Downstream of Inanda dam), Durban, South Africa.

pality. These serve as the starting point and an Arcview script was written which can select the appropriate flood lines to display based on the most recent observed streamflow, or alternatively the forecast flow.

Although the flood-wave is dynamic, it is assumed (for steep channels with negligible off-channel storage) that the inundation levels traced by the flood peak will be closely approximated by the corresponding steady-state peak, whose inundation level is the dynamic wave's upper bound.

Since steep coastal catchments are the most flood prone in South Africa (Pegram et al., 2006a). Using the peak of the inflow Hydrograph gives a conservative estimate of inundation depths and if there is little off-channel storage they should effectively behave as a channel and experience a Kinematic flood wave.

CHAPTER 6

CONCLUSION

The first section of this chapter summarizes the technical contributions presented in this thesis. Discussion and recommendations are given in the second section.

6.1 Summary of main results and contributions

The purpose of this section is to summarize the main results and highlight the new contributions which have been made by this work. The work presented in this thesis has been carried out in the context of providing flood forecasting capabilities in South Africa. In this context, contributions have (of necessity) been split between science and implementation. The main contributions are outlined below.

- A two-dimensional extension of Empirical Mode Decomposition (EMD) has been presented and applied to the analysis of spatial rainfall data. EMD analysis explicitly accepts non-homogeneity in the data and does not rely on the predefined basis functions integral to other methods (e.g. Fourier, Wavelet). The work is a new contribution, published in Sinclair and Pegram (2005b) and discussed in detail in section 4.2 of the thesis. The significance of this contribution is that the spatial wavenumber components inherent in the data can be separated in an unsupervised manner. The resulting cascade of fields can be used for forecasting using the ideas of Seed (2003) and Turner et al. (2004).

- A novel technique for the accumulation of spatial rainfall fields has been presented. Advection vectors computed between successive radar scans are used to generate path integrals. The approach presented adds a refinement to those reported by Anagnostou and Krajewski (1999a) and Hannesen (2002). The accumulation scheme is presented in section 3.3 and it's significance is that it provides a way to compare rain gauge accumulations and accumulations from instantaneous fields in a sensible way. The advection routine used also has benefits in rainfall field nowcasting.

- The String of Beads Model (SBM) presented in Pegram and Clothier (1999) and Clothier and Pegram (2002), has been adapted for short term nowcasting and compared with the Spectral PROGnosis (S-PROG) model (Seed, 2001). SBM in nowcasting mode is a new contribution, described in detail in section 4.1.2. It turns out that it is useful for ensemble forecasting (Berenguer et al., 2006) but does not perform as well as S-PROG for mean field forecasts.

- The conditional merging technique of Ehret (2002) has been adopted and validated using cross-validation techniques for synthetic as well as observed rainfall data. The technique has been compared with an implementation of the Bayesian merging technique (Todini, 2001) and found to perform competitively. The conditional merging technique has also been extended to provide a more formal discussion of its error structure. The conditional merging algorithm was discussed in section 3.5.

- Although adaptive time series forecasting is certainly not new (Haykin, 2001; Ljung, 1987), the application of these techniques is useful in the Hydrometrological context where traditional time series methods (Box and Jenkins, 1970) and filters (Kalman filters, Extended Kalman filter) are quite common (e.g. Seed, 2003; Szollosi-Nagy and Mekis, 1987; Todini 1978). Adaptive filtering techniques are used to forecast non-stationary time series data. The adaptive filtering techniques and model fitting procedures adopted in this study are presented in chapter 2 and are suggested as an alternative

to stationary models where these fail to provide structural information.

- Implementation of data transfer mechanisms. Components of a prototype flood forecasting system were set up in the eThekwini disaster management centre (Durban, South Africa). This system is a first in South Africa and has laid the ground work for the eventual implementation of a more robust countrywide flood warning system. This work was expanded upon in chapter 5 and included rainfall and flood line visualization in a GIS environment, as well as the automated collection of real-time stream flow data.

6.2 Discussion and recommendations

The focus of this thesis has largely been on the estimation and nowcasting of spatial rainfall fields for the purposes of flash flood forecasting. A second theme has been to suggest strategies (partially based on working implementations) for installing and operating practical flash flood forecasting systems in a country where there are virtually no existing mechanisms in place. This is a dire situation in a country where so many are impoverished and without the means to recover from disasters. The government has recognized this and therefore the South African Disaster Management Act (Act 57 of 2002) advocates a paradigm shift from the current "bucket and blanket brigade" response-based mind-set to one where disaster prevention or mitigation are the preferred option.

The Act mandates that local municipalities in South Africa have in place effective and properly financed mitigation strategies for disaster management, this specifically includes flood warning systems. It is in the context of mitigating the effects of flood events that the development and implementation of a reliable flood forecasting system has major significance. The research reported on in this thesis is part of a broader contribution aimed at solving these problems.

Perhaps the most important aspect of the development and finally implementation of flood forecasting systems in South Africa is the provision of reliable and robust data gathering and distribution links. The capacity does not exist, in general, at local government level to perform this task and it is here that SAWS and

DWAF play a key role in realizing the letter of the laws set out in the Act. Most of the work presented in this thesis has been carried out in support of the roles that SAWS and DWAF play and the focus is therefore on pragmatism rather than rigorous attention to fine details. Clearly, where these details are important they have received the attention that they deserve.

Finally, the greatest challenge facing provision of appropriate flood forecasting and warning systems in South Africa is not the technological limits but a critical lack of people with relevant skill sets. Progress will become far more rapid if this situation can be adequately addressed.

REFERENCES

Adler, R. and Negri, A. (1988). A satellite infrared technique to estimate tropical convective and stratiform rainfall. *Journal of Applied Meteorology*, 27:30–51.

Akaike, H. (1973). Information Theory and an extension of the Maximum Likelihood Principle. In Petrov, B. and Csaki, F., editors, *2nd International Symposium on Information Theory*, pages 267–281, Budapest, Hungary.

Anagnostou, E. and Krajewski, W. (1999a). Real-Time Radar Rainfall Estimation. Part I: Algorithm Formulation. *Journal of Atmospheric and Oceanic Technology*, 16:189–197.

Anagnostou, E. and Krajewski, W. (1999b). Real-Time Radar Rainfall Estimation. Part II: Case Study. *Journal of Atmospheric and Oceanic Technology*, 16:198–205.

Austin, P. M. (1987). Relation between Measured Radar Reflectivity and Surface Rainfall. *Monthly Weather Review*, 115:1053–1070.

Bab-Hadiashar, A. and Suter, D. (1998). Robust optic flow computation. *International Journal of Computer Vision*, 29(1):59–77.

Bab-Hadiashar, A., Suter, D., and Jarvis, R. (1996). 2D motion extraction using an image interpolation technique. *Proceedings of SPIE, Applications of Digital Image Processing XVIII*, 2564:271–281.

Balocchi, R., Menicucci, D., Santarcangelo, E., Sebastiani, L., Gemignani, A., Gellarducci, B., and Varanini, M. (2004). Deriving the respiratory sinus arrhythmia from the heartbeat time series using Empirical Mode Decomposition. *Chaos, Solitons and Fractals*, 20:171–177.

Barrett, E. and Martin, D. (1981). *The use of satellite data in rainfall monitoring*. Academic Press, New York.

Berenguer, M., Corral, C., Sánchez-Diezma, R., and Sempere-Torres, D. (2005). Hydrological validation of a radar-based nowcasting technique. *Journal of Hydrometeorology*, 6:532–549.

Berenguer, M., Sempere-Torres, D., Sanchez-Diezma, R., Pegram, G., Zawadzki, I., and Seed, A. (2006). Modelization of the uncertainty associated to radar-based nowcasting techniques. Impact in flow simulation. In *ERAD 2006, Proceedings of the 4th European conference on Radar in Meteorology and Hydrology*, Barcelona, Spain.

Borga, M. and Vizzaccaro, A. (1997). On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging. *Journal of Hydrology*, 195:160–171.

Bowler, N. E., Pierce, C. E., and Seed, A. (2004). Development of a precipitation nowcasting algorithm based upon optical flow techniques. *Journal of Hydrology*, 288:74–91.

Box, G. and Jenkins, G. (1970). *Time Series Analysis forecasting and control.* Holden Day, San Francisco, California.

Brandes, E. (1975). Optimizing rainfall estimates with the aid of radar. *Journal of Applied Meteorology*, 14:1339–1345.

Buys, L. (2005). The SA National Disaster Management Framework. Pers. Comm.

Chiew, F., Peel, M., Amirthanathan, G., and Pegram, G. (2005). Identification of oscillations in historical global streamflow data using empirical mode decomposition. In *Regional Hydrological Impacts of Climatic Change – Hydroclimatic Variability, Proceedings of symposium S6 held during the Seventh IAHS Scientific Assembly*, Foz do Iguau, Brazil.

Chilès, J. and Delfiner, P. (1999). *Geostatistics: Modelling spatial uncertainty.* Wiley, New York.

Chumchean, S., Sharma, A., and Seed, A. (2003). Radar rainfall error variance and its impact on radar rainfall calibration. *Physics and Chemistry of the Earth*, 28:27–39.

Clothier, A. and Pegram, G. (2002). Space-time modeling of Rainfall using the String of Beads model: Integration of Radar and Raingauge data. Technical Report 1010/1/02, Water Research Commission, Pretoria, South Africa.

Clothier, A. N. (2003). *High resolution space-time modelling of rainfall: The String of Beads Model.* PhD thesis, University of Natal, Durban, South Africa.

Conway, E. D. (1997). *An Introduction to Satellite Image Interpretation.* John Hopkins University Press, Baltimore, Maryland.

Cooley, J. and Tukey, J. (1965). An algorithm for the machine computation of the complex Fourier series. *Mathematics of Computation*, 19:297–301.

Cowpertwait, P. (1991). Further developments of the Neyman-Scott clustered point process for modeling rainfall. *Water Resources Research*, 27(7):1431–1438.

Cressie, N. (1991). *Statistics for Spatial Data*. Wiley, New York.

Creutin, J. and Obled, C. (1982). Objective analyses and mapping techniques for rainfall fields: An objective comparison. *Water Resources Research*, 18(2):413–431.

Daubechies, I. (1992). *Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics*. Philadelphia, Pennsylvania.

Deyzel, I., Pegram, G., Visser, P., and Dicks, D. (2004). Spatial Interpolation and Mapping of Rainfall (SIMAR) Volume 2: Radar and satellite products. Technical Report 1152/1/04, Water Research Commission, Pretoria, South Africa.

Dixon, M. and Wiener, G. (1993). TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting – A Radar-based Methodology. *Journal of Oceanic and Atmospheric Technology*, 10(6):785–797.

Dooge, J. (2003). *Linear Theory of Hydrologic Systems*. European Geosciences Union, Katlenburg-Lindau, Germany.

Ehret, U. (2002). *Rainfall and flood nowcasting in small catchments using weather radar*. PhD thesis, University of Stuttgart, Stuttgart, Germany.

Engelbrecht, F., deW Rautenbach, C., McGregor, J., and Katzfey, J. (2002). January and July climate simulations over the SADC region using the limited-area DARLAM. *Water SA*, 28(2):361–374.

Flandrin, P., Rilling, G., and Goncalves, P. (2004). Empirical Mode Decomposition as a filter bank. *IEEE Signal Processing Letters*, 11:112–114.

Fulton, R. A., Breidenbach, J. P., Seo, D.-J., Miller, D. A., and O'Bannon, T. (1998). The WSR-88D rainfall algorithm. *Weather and Forecasting*, 13:377–395.

Ganguly, A. R. and Bras, R. L. (2003). Distributed quantitative precipitation forecasting using information from radar and numerical weather prediction models. *Journal of Hydrometeorology*, 4:1168–1180.

Gelb, A. (1974). *Applied Optimal Estimation.* MIT Press, Cambridge, Massachusetts.

Germann, U. and Zawadzki, I. (2002). Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology. *Monthly Weather Review*, 130:2859–2873.

Germann, U. and Zawadzki, I. (2004). Scale-dependence of the predictability of precipitation from continental radar images. Part II: Probability forecasts. *Journal of Applied Meteorology*, 43:74–89.

Germann, U., Zawadzki, I., and Turner, B. (2006). Predictability of Precipitation from Continental Radar Images. Part IV: Limits to Prediction. *Journal of Atmospheric Sciences*, 63:2092–2108.

Gloersen, P. and Huang, N. (2003). Comparison of Interannual Intrinsic Modes in Hemispheric Sea Ice Covers and Other Geophysical Parameters. *IEEE Transactions on Geoscience and Remote Sensing*, 41:1062–1074.

Gotway, C. and Young, L. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97:632–648.

Grecu, M. and Krajewski, W. (2000). Multiquadric equations and optimal areal rainfall estimation. *Journal of Hydrology*, 239:69–84.

Habib, E. and Krajewski, W. (2002). Uncertainty Analysis of the TRMM Ground-Validation Radar-Rainfall Products: Application to the TEFLUN-B Field Campaign. *Journal of Applied Meteorology*, 41:558–572.

Han, C., Guo, H., Wang, C., and Fan, D. (2002). A novel method to reduce speckle in SAR images. *International Journal of Remote Sensing*, 23:5095–5101.

Hannesen, R. (2002). An enhanced surface rainfall algorithm for Radar Data. Technical Report Progress report for MUSIC, contract No. EVK1-CT-2000-00058, European Commission.

Harris, D., Foufoula-Georgiou, E., Droegemeier, K. K., and Levit, J. J. (2001). Multiscale statistical properties of a high-resolution precipitation forecast. *Journal of Hydrometeorology*, 2:406–418.

Haykin, S. (2001). *Adaptive Filter Theory.* Prentice-Hall, 4th edition.

Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., Tung, C., and Liu, H. (1998). The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. *Proceedings: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995.

Huang, N. E., Shen, Z., and Long, S. R. (1999). A new view of Nonlinear water waves: the Hilbert Spectrum. *Annual Review of Fluid Mechanics*, 31:417–457.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.

Hwang, P. A., Huang, N. E., and Wang, D. W. (1999). A note on analyzing nonlinear and nonstationary ocean wave data. *Applied Ocean Research*, 25:187–193.

Journel, A. and Huijbregts, C. (1978). *Mining Geostatistics*. Academic Press, New York.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82:35–45.

Kedem, B., Pfeiffer, R., and Short, D. (1997). Variability of space-time mean rain rate. *Journal of Applied Meteorology*, 36:443–451.

Koistinen, J. and Puhakka, T. (1981). An improved spatial gauge-radar adjustment technique. In *20th Conference on RADAR Meteorology*, AMS Boston USA.

Krajewski, W. F. (1987). Cokriging radar-rainfall and rain gage data. *Journal of Geophysical Research*, 92:9571–9580.

Kroese, N. (2004). Spatial Interpolation and Mapping of Rainfall (SIMAR) Volume 1: Maintenance and upgrading of radar and rain gauge infrastructure. Technical Report 1151/1/04, Water Research Commission, Pretoria, South Africa.

Kumar, P. and Foufoula-Georgiou, E. (1993). A Multicomponent Decomposition of Spatial Rainfall Fields 1. Segregation of Large- and Small-Scale Features Using Wavelet Transforms. *Water Resources Research*, 29(8):2515–2532.

Levizzani, V., Amorati, R., and Meneguzzo, F. (2002). A review of satellite-based rainfall estimation methods. Technical Report Progress report for MUSIC, contract No. EVK1-CT-2000-00058, European Commission.

Linderhed, A. (2002). 2-D empirical mode decompositions – in the spirit of image compression. *Proceedings of SPIE, Wavelet and Independent Component Analysis Applications IXI*, 4738.

Linderhed, A. (2004). Variable sampling of the empirical mode decomposition of two-dimensional signals. Available http://www.icg.isy.liu.se/~anna/emdsamp.pdf.

Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, New Jersey.

Marshall, J. and Palmer, W. (1948). The distribution of raindrops with size. *Journal of Meteorology*, 5:165–166.

Mazzetti, C. (2004). Report on performance of the new methodology. the application of the block kriging and bayesian combination technique. Technical Report Final report for MUSIC, contract No. EVK1-CT-2000-00058, European Commission.

Mittermaier, M. and Terblanche, D. (1997). Converting weather radar data to cartesian space: a new approach using DISPLACE averaging. *Water S.A.*, 23(1):45–50.

Mkwananzi, N. and Pegram, G. (2004). A Flood Nowcasting System for the eThekwini Metro. Volume 2: Modelling Flood Inundation in the Mlazi river under uncertainty. Technical Report 1217/2/04, Water Research Commission, Pretoria, South Africa.

Mkwananzi, N., Pegram, G., and Sinclair, S. (2003). Modelling flood inundation in the Mlazi river under uncertainty. In *Proceedings of the 11th South African National Hydrology Symposium*, Port Elizabeth, South Africa.

Nunes, J. C., Bouaoune, Y., Delechelle, E., Niang, O., and Bunel, P. (2003). Image analysis by bidimensional empirical mode decomposition. *Image and Vision Computing*, 21:1019–1026.

Pegram, G. (2004). Spatial Interpolation and Mapping of Rainfall (SIMAR) Volume 3: Data Merging for rainfall map production. Technical Report 1153/1/04, Water Research Commission, Pretoria, South Africa.

Pegram, G. and Clothier, A. (1999). High resolution space-time modeling of rainfall: the "String of Beads" model. Technical Report 752/1/99, Water Research Commission, Pretoria, South Africa.

Pegram, G. and Clothier, A. (2001). High Resolution Space-Time Modelling of Rainfall: The "String of Beads" Model. *Journal of Hydrology*, 241:26–41.

Pegram, G., Deyzel, I., Sinclair, S., Visser, P., Terblanche, D., and Green, G. (2004). Daily mapping of 24 hr rainfall at pixel scale over South Africa using satellite, radar and raingauge data. In *Proceedings of the 2nd IPWG Workshop*, Monterey, CA, USA.

Pegram, G., Seed, A., and Sinclair, D. (2002). Comparison of Methods of Short-term Rainfield Nowcasting. In *Proceedings of the 27th EGS general assembly*, Nice, France.

Pegram, G. and Sinclair, D. (2002). A linear catchment model for real-time flood forecasting. Technical Report 1005/1/02, Water Research Commission, Pretoria, South Africa.

Pegram, G. and Sinclair, S. (2004). National Flood Nowcasting System towards an integrated mitigation strategy in South Africa. In *Proceedings of the 6th International Symposium on Hydrological Applications of Weather Radar*, Melbourne, Australia.

Pegram, G., Sinclair, S., Parak, M., Sakulski, D., and Nxumalo, N. (2006a). National Flood Nowcasting System: Towards an integrated mitigation strategy. Technical Report 1429/1/06, Water Research Commission, Pretoria, South Africa.

Pegram, G., Sinclair, S., and Wesson, S. (2006b). Daily Rainfall Mapping over South Africa: Mapping. Technical Report 1425/1/06, Water Research Commission, Pretoria, South Africa.

Pegram, G. and Terblanche, D. (1998). Radar Hydrology developments in South Africa. In *Proceedings of Radar Hydrology for real time Flood Forecasting: An advanced study course*, University of Bristol, United Kingdom.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C – The art of scientific computing, Second Edition*. Cambridge University Press, 2nd edition.

Schertzer, D. and Lovejoy, S. (1987). Physical modelling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *Journal of Geophysical Research*, 92:9693–9714.

Seed, A. (2001). A dynamic and spatial scaling approach to advection forecasting. In *Fifth International Symposium on Hydrological Applications of Weather Radar - Radar Hydrology*, Kyoto - Japan.

Seed, A., Srikanthan, R., and Menabde, M. (1999). A space and time model for design storm rainfall. *Journal of Geophysical Research*, 100:31623–31630.

Seed, A. W. (2003). A Dynamic and Spatial Scaling Approach to Advection Forecasting. *Journal of Applied Meteorology*, 42:381–388.

Seo, D. (1998a). 1. Real-time estimation of rainfall fields using rain gage data under fractional coverage conditions. *Journal of Hydrology*, 208:25–36.

Seo, D. (1998b). 2. Real-time estimation of rainfall fields using radar rainfall and rain gage data. *Journal of Hydrology*, 208:37–52.

Sevruk, B. (1982). Methods of Correction for Systematic Error in Point Precipitation Measurement for Operational use. Technical Report 21 Operational Hydrology, World Meteorological Organization, Geneva, Switzerland.

Sinclair, S., Ehret, U., Bardossy, A., and Pegram, G. (2003). Comparison of Conditional and Bayesian Methods of Merging Radar and Rain gauge Estimates of Rainfields. In *EGS - AGU - EUG joint assembly*, Nice, France.

Sinclair, S. and Pegram, G. (2003a). Combining Traditional and remote sensing techniques as a tool for Hydrology, Agriculture and Water Resources Management. In *Proceedings of the 11th South African National Hydrology Symposium*, Port Elizabeth, South Africa.

Sinclair, S. and Pegram, G. (2003b). The Design and Implementation of a Real-Time Flood Forecasting System in Durban, South Africa. In *EGS - AGU - EUG joint assembly*, Nice, France.

Sinclair, S. and Pegram, G. (2004a). A Flood Nowcasting System for the eThekwini Metro. Volume 1: Umgeni Nowcasting using Radar - An integrated pilot study. Technical Report 1217/1/04, Water Research Commission, Pretoria, South Africa.

Sinclair, S. and Pegram, G. (2004b). Combining radar and rain gauge rainfall estimates for flood forecasting in South Africa. In *Proceedings of the 6th International Symposium on Hydrological Applications of Weather Radar*, Melbourne, Australia.

Sinclair, S. and Pegram, G. (2005a). Combining radar and rain gauge estimates using conditional merging. *Atmospheric Science Letters*, 6:19–22.

Sinclair, S. and Pegram, G. (2005b). Empirical Mode Decomposition in 2-D space and time: a tool for space-time rainfall analysis and nowcasting. *Hydrology and Earth System Sciences*, 9:127–137.

Sinclair, S. and Pegram, G. (2005c). Empirical Mode Decomposition in 2-D space and time: A tool for space-time rainfall analysis and nowcasting. In *Proceedings of the 12th South African National Hydrology Symposium*, Pretoria, South Africa.

Sinclair, S. and Pegram, G. (2005d). Space-time rainfall analysis and nowcasting using Empirical Mode Decomposition in 2D. In *EGU general assembly*, Vienna, Austria.

Smith, J. A. and Krajewski, W. F. (1991). Estimation of the Mean Field Bias of Radar Rainfall Estimates. *Journal of Applied Meteorology*, 30:397–412.

Sonka, M., Hlavac, V., and Boyle, R. (1999). *Image Processing, Analysis and Machine Vision*. PWS Publishing, Pacific Grove, 2nd edition.

Sorenson, H. W., editor (1985). *Kalman Filtering: Theory and Application*. IEEE Press, New York.

Szollosi-Nagy, A. and Mekis, E. (1987). Comparative analysis of three recursive real-time river flow forecasting models: deterministic, stochastic, and coupled deterministic-stochastic. *Stochastic Environmental Research and Risk Assessment*, 2(1):17–33.

Tapiador, F., Kidd, C., Levizzani, V., and Marzano, F. (2004a). A maximum entropy approach to satellite quantitative precipitation estimation (QPE). *International Journal of Remote Sensing*, 25(21):4629–4639.

Tapiador, F. J., Kidd, C., Levizzani, V., and Marzano, F. S. (2004b). A Neural NetworksBased Fusion Technique to Estimate Half-Hourly Rainfall Estimates at 0.1° Resolution from Satellite Passive Microwave and Infrared Data. *Journal of Applied Meteorology*, 43:576–594.

Todini, E. (2001). A bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements. *Hydrology and Earth System Sciences*, 5(2):187–199.

Turner, B., Zawadzki, I., and Germann, U. (2004). Predictability of precipitation from continental radar images. Part III: Operational nowcasting implementation (MAPLE). *Journal of Applied Meteorology*, 43:231–248.

Vanmarcke, E. (1988). *Random fields, analysis and synthesis*. MIT Press, Cambridge, Massachusetts.

Wesson, S. M. and Pegram, G. G. S. (2004). Radar rainfall image repair techniques. *Hydrology and Earth System Sciences*, 8:220–234.

Wichmann, B. and Hill, I. (1982). *Applied Statistics Algorithms*, chapter Algorithm AS183. Royal Statistical Society.

Wikle, C. and Berliner, L. (2005). Combing information across spatial scales. *Technometrics*, 47:80–91.

Wilson, J. and Brandes, E. (1979). Radar measurement of rainfall - A summary. *Bulletin of the American Meteorological Society*, 60:1048–1058.

WPDM (1998). *White paper on disaster management*. Department of constitutional development, South Africa.

Young, P. (1984). *Recursive Estimation and Time-Series Analysis. An Introduction*. Springer-Verlag, Berlin.

Zhang, R. R., Ma, S., and Hartzell, S. (2003). Signatures of the seismic source in EMD-based characterization of the 1994 Northridge, California, earthquake recordings. *Bulletin of the Seismological Society of America*, 93:501–518.

APPENDICES

ATMOSPHERIC SCIENCE LETTERS
Atmos. Sci. Let. **6**: 19–22 (2005)
Published online 22 February 2005 in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/asl.85

Combining radar and rain gauge rainfall estimates using conditional merging

Scott Sinclair* and Geoff Pegram
Civil Engineering, University of KwaZulu-Natal, Durban, South Africa

*Correspondence to:
Scott Sinclair, Civil Engineering,
University of KwaZulu-Natal,
Durban, South Africa.
E-mail: sinclaird@ukzn.ac.za

Abstract

The Hydrologist's traditional tool for measuring rainfall is the rain gauge. Rain gauges are relatively cheap, easy to maintain and provide a direct and suitably accurate estimate of rainfall at a point. What rain gauges fail to capture well is the spatial variability of rainfall with time, an important aspect for the credible modelling of a catchment's response to rainfall. This spatial variability is particularly evident at short timescales of up to several days. As the period of accumulation increases, the expected spatial variability is reduced and rain gauges provide improved spatial rainfall estimates. Because of the fractal variability of rainfall in space, simple interpolation between rain gauges does not provide an accurate estimate of the true spatial rainfall field, at short time scales.

Weather radar provides (with a single instrument) a highly detailed representation of the spatial structure and temporal evolution of rainfall over a large area. Estimated rainfall rates are derived indirectly from measurements of reflectivity and are therefore subject to a combination of systematic and random errors.

This article describes a recently proposed merging technique and presents an application to simulated rainfall fields. The technique employed is Conditional Merging (Ehret, 2002), which makes use of Kriging to extract the optimal information content from the observed data. A mean field based on the Kriged rain gauge data is adopted, while the spatial detail from the radar is retained, reducing bias, but keeping the spatial variability observed by the radar. The variance of the estimate is reduced in the vicinity of the gauges where they are able to provide good information on the true rainfall field. **Copyright © 2005 Royal Meteorological Society**

Received: 1 July 2004
Accepted: 12 November 2004

Keywords: **Kriging; conditional merging; multi-sensor rainfall estimation**

1. Introduction

The spatial estimation of rainfall by combining information from multiple sensors has received considerable attention in the Hydro-meteorological literature. Early work (e.g. Brandes, 1975) focussed on the correction of bias in radar estimates of rainfall using an adjustment factor. Krajewski (1987) suggests a Cokriging procedure that he demonstrates by a numerical experiment. A procedure that accounts for the fractional coverage of rainfall (spatial intermittency) is suggested by Seo (1998a,b) using conditional expectations. Seo (1998b) computes the expectation of rainfall at an ungauged site, conditional on the observed rain gauge and radar data. A Bayesian merging technique suggested by Todini (2001) relies on a Kalman filtering scheme to remove the error variance (after a separate bias reduction) by taking a Block-Kriged spatial estimate based on rain gauges as the observation vector representing the true rainfall field.

The greatest challenge in implementing the merging techniques mentioned above is the difficulty of estimating the error structure of rain gauge and radar observations. This structure may either be estimated from historical observations on the assumption of stationarity or estimated (and updated) online using some parameter updating procedure similar to that employed by Anagnostou and Krajewski (1999a,b) for adjusting radar rainfall estimates based on real-time rain gauge observations. However, the true rainfall field remains unknown and Hydro-meteorologists must rely on the relationships between the observational data to draw conclusions regarding the unknown error structure.

2. Conditional merging

Radar produces an observation of the unknown true rainfall field that is subject to several well-known sources of error (e.g. Wilson and Brandes, 1979; Habib and Krawjewski, 2002), but retains the general covariance structure of the true precipitation field. The information from the radar can be used to condition the spatially limited information obtained by interpolating between rain gauges and produce an estimate of the rainfall field that contains the correct spatial structure while being constrained to the rain gauge data (where it is available). The conditional merging technique of Ehret (2002), suggested by Pegram (2002), does just

Copyright © 2005 Royal Meteorological Society

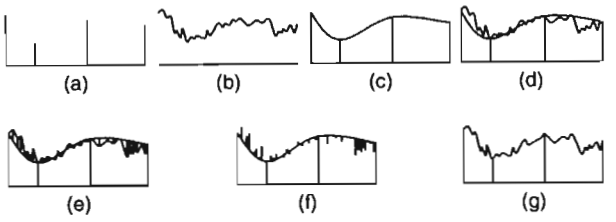


Figure 1. The conditional merging process. (a) The rainfall field is observed at discrete points by rain gauges. (b) The rainfall field is also observed by radar on a regular, volume-integrated grid. (c) Kriging of the rain gauge observations is used to obtain the best linear unbiased estimate of rainfall on the radar grid. (d) The radar pixel values at the rain gauge locations are interpolated onto the radar grid using Kriging. (e) At each grid point, the deviation C between the observed and interpolated radar value is computed. (f) The field of deviations obtained from (e) is applied to the interpolated rainfall field obtained from Kriging the rain gauge observations. (g) A rainfall field that follows the mean field of the rain gauge interpolation, while preserving the mean field deviations and the spatial structure of the radar field is obtained

this, making use of Ordinary Kriging (e.g Cressie, 1991) to extract the information content from the observed data. Figure 1 and the text following give an overview of the technique (for the one-dimensional case), which is adapted from Ehret's (2002) work.

The spatial structure of the merged field is therefore obtained from the radar, while the rainfall values are 'stitched' to the gauge observations of the true rainfall field. The approach taken here has great similarities to the technique of Conditional Simulation by Kriging discussed by Chiles and Delfiner (1999). However, the key difference in this case is that the radar rainfall estimate is *not* a simulation, unrelated (except by its statistical properties) to the rainfall field we want to observe, but is in fact an observation of the true unknown field. This important link means that the radar data provides an estimate of the *actual* Kriging error and, in particular, its spatial structure.

The error structure of the merged estimate can be examined by considering Equations 1 to 7 below:

$$Z(s) = G_K(s) + \varepsilon_G(s) \tag{1}$$

$$R(s) = R_K(s) + \varepsilon_R(s) \tag{2}$$

$$M(s) = G_K(s) + \varepsilon_R(s) \tag{3}$$

$$E[Z(s) - M(s)] = E[\varepsilon_G(s) - \varepsilon_R(s)] \tag{4}$$

$$\begin{aligned} \mathrm{var}[Z(s) - M(s)] &= \mathrm{var}[\varepsilon_G(s) - \varepsilon_R(s)] \\ &= \sigma^2_{\varepsilon_G(s)} + \sigma^2_{\varepsilon_R(s)} \\ &\quad - 2\mathrm{cov}[\varepsilon_G(s), \varepsilon_R(s)] \\ &= \beta - 2\sigma_{\varepsilon_G(s)}\sigma_{\varepsilon_R(s)}\rho \end{aligned} \tag{5}$$

$$\varepsilon_G(s) = Z(s) - G_K(s) \tag{6}$$

$$\varepsilon_R(s) = R(s) - R_K(s) \tag{7}$$

$$\beta = \sigma^2_{\varepsilon_G(s)} + \sigma^2_{\varepsilon_R(s)}$$

where $Z(s)$ is the true rainfall field at location s, $G_K(s)$ is the Kriged (mean field) estimate of $Z(s)$ from the rain gauge data, $R(s)$ is the radar rainfall estimate, $R_K(s)$ is the Kriged (mean field) estimate of $R(s)$ using the radar values at rain gauge locations and $M(s)$ is the merged estimate of $Z(s)$.

The term $\varepsilon_G(s)$ in Equation 1 is unknown since $Z(s)$ is unknown. In Equation 2, $\varepsilon_R(s)$ is known and, on the basis that $R(s)$ is a measurement of $Z(s)$, Equation 3 can be used to estimate $Z(s)$. Equation 4 shows that the expected value of the error between the merged estimate and the true field is zero if the fields are Gaussian, since the Kriged estimates are unbiased in this case. The variance of the error estimate given by Equation 5 can be decomposed, as shown in Equation 6. The variance of the error is (trivially) zero at the gauged points, while at any other position in the field, it is bounded by a maximum value of β, for positive correlations ρ between $\varepsilon_G(s)$ and $\varepsilon_R(s)$. If $\varepsilon_G(s)$ and $\varepsilon_R(s)$ are strongly (positively) correlated as one would expect since both gauges and radar are measurements of $Z(s)$, then the variance of the error will be *significantly* less than β, as suggested by Equation 7.

3. An illustration based on the 'string of beads' model

An artificial experiment was carried out to test the efficiency of the technique at estimating the true rainfall field. Since the true rainfall is unknowable in reality, a sequence of 1000 independent 128 × 128 pixel rainfall fields was produced using the 'string of beads' simulation model (Pegram and Clothier, 2001). Each pixel is representative of a 1 × 1 km area and all 1000 simulated fields were generated with identical

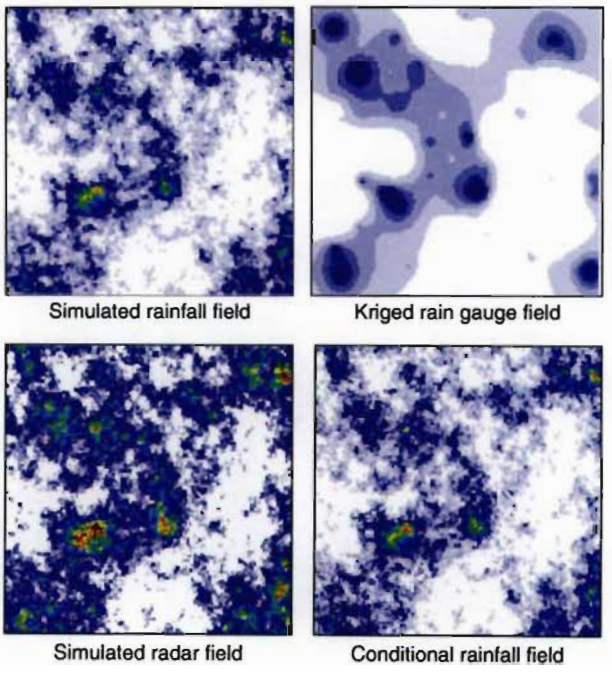


Figure 2. A single realization of the simulated and merged rainfall fields

Copyright © 2005 Royal Meteorological Society

statistical properties. The simulated rainfall fields were treated as the 'true' rainfall and 'observed' radar estimates produced by adding systematic (bias) and random (noise) error components to the simulated fields. The 'true' field was then sampled at 83 'rain gauge' locations distributed randomly on the pixel grid. This gives an approximate coverage of one gauge for every 200 km^2. A single realization of the 'true' field as well as the corresponding 'observed' radar estimate and the computed Kriged gauge field are shown in Figure 2, while a more detailed description of the experiment and the results obtained follows. Figure 3 shows the correlation structure in the simulated fields.

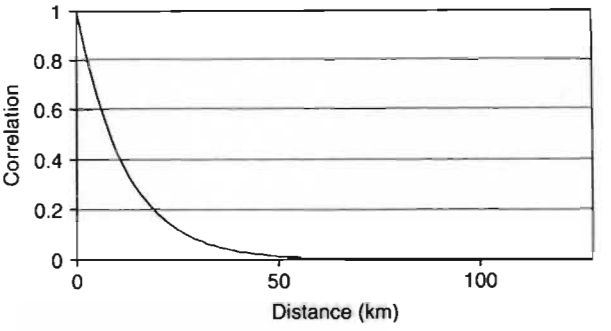


Figure 3. A plot of the exponential correlation function used in generating rainfields

For each of the 1000 simulated radar rainfall fields and rain gauge observations, the conditional merging technique was applied and the subsequent merged field compared with the true simulated rainfall field in terms of the mean and variance of its deviations from the true field. All of the Kriging and merging computations were done on the logarithms of the variables in order to transform the log-normally distributed rainfall rate simulations to a Gaussian space. Comparisons, however, were done on the back transformed variables. The mean error of the merged field (over 1000 realizations) was computed for each of the 16 384 pixels and plotted as a histogram (Figure 4). Also plotted in Figure 4 are histograms for the simulated

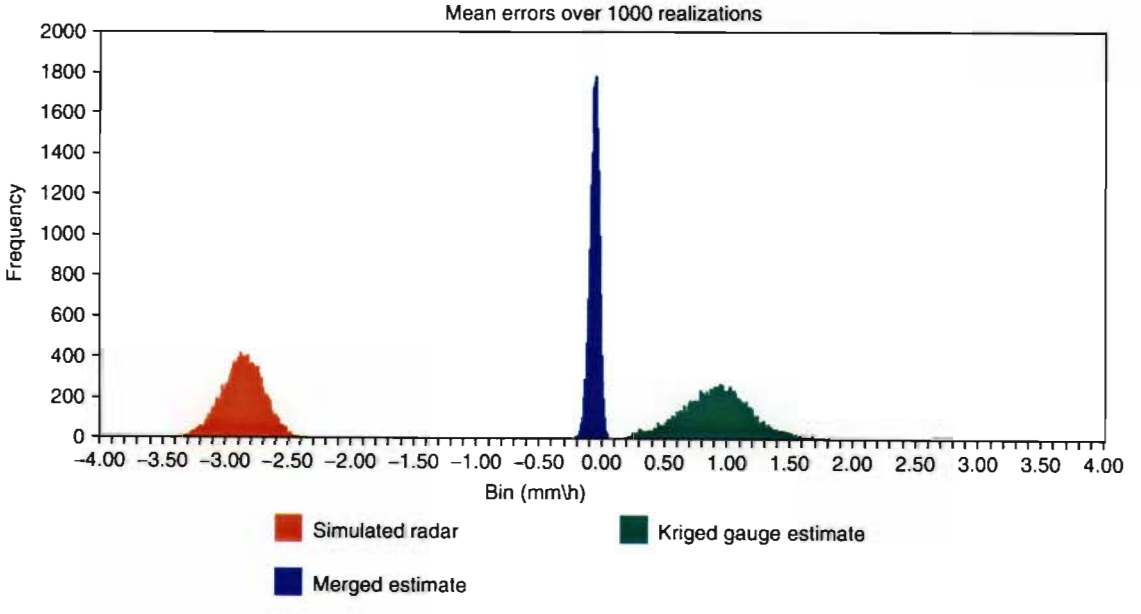


Figure 4. Comparison of the mean errors

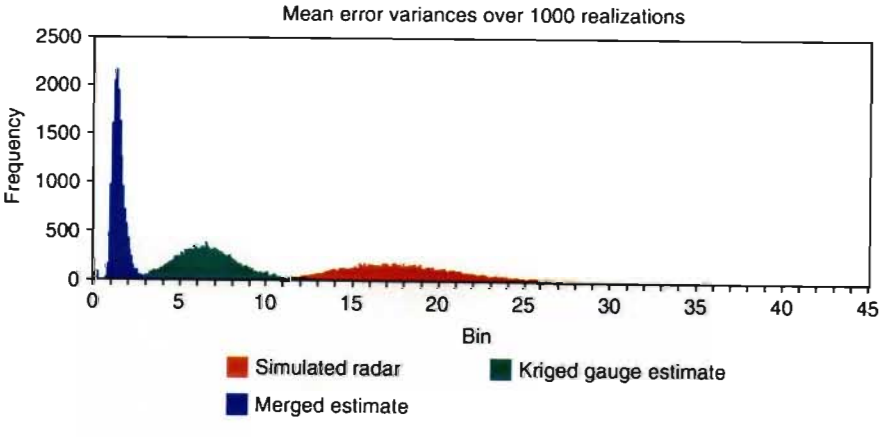


Figure 5. Comparison of the variance of the errors

Copyright © 2005 Royal Meteorological Society

radar deviations from the true field and the Kriged rain gauge deviations for comparison. The figure clearly shows the improvement of the merged fields in estimating the true field when compared with the simulated radar and Kriged gauge fields. Figure 5 shows similar plots for the variance of the errors, again the merged estimate performs well.

4. Conclusion

A conditional merging technique for combining rainfall information from radar and rain gauges has been presented and its efficiency in terms of reducing the bias and variance of error estimates has been shown using an artificial simulation experiment. The technique is to be used for spatial rainfall estimation at short timescales relevant to a National Flood Forecasting and Warning System in South Africa. Rain gauge observations provide good point rainfall accuracy, but poor spatial detail. The technique presented in this article uses radar observations of rainfall fields to estimate the errors associated with using Kriging to interpolate between the rain gauge observations and condition the Kriged gauge field accordingly. Thus, the spatial detail of the final merged field is improved while maintaining the mean field characteristics measured by the gauges.

Acknowledgements

The authors would like to acknowledge the support of the Water Research Commission of South Africa for supporting the work presented herein through contracts K5/1249 and K5/1425/2. The South African Weather Service (SAWS): Meteorological systems and technology section (METSYS), the National Disaster Management Centre (NDMC) and Department of Water Affairs and Forestry (DWAF): Hydrology are also acknowledged for their input as research partners.

References

Anagnostou EN, Krajewski WF. 1999a. Real-time radar rainfall estimation. Part I: algorithm formulation. *Journal of Atmospheric and Oceanic Technology* **16**: 189–197.

Anagnostou EN, Krajewski WF. 1999b. Real-time radar rainfall estimation. Part II: case study. *Journal of Atmospheric and Oceanic Technology* **16**: 198–205.

Brandes EA. 1975. Optimizing rainfall estimates with the aid of radar. *Journal of Applied Meteorology* **14**: 1339–1345.

Cressie NAC. 1991. *Statistics for Spatial Data*. Wiley: New York.

Chiles JP, Delfiner P. 1999. *Geostatistics: Modelling Spatial Uncertainty*. Wiley: New York.

Ehret U. 2002. Rainfall and flood nowcasting in small catchments using weather radar. PhD Thesis, University of Stuttgart.

Habib E, Krajewski WF. 2002. Uncertainty analysis of the TRMM ground-validation radar-rainfall products: application to the TEFLUN-B field campaign. *Journal of Applied Meteorology* **41**: 558–572.

Krajewski WF. 1987. Cokriging radar-rainfall and rain gauge data. *Journal of Geophysical Research* **92**: 9571–9580.

Pegram GGS. 2002. Spatial interpolation and mapping of rainfall: 3. Optimal integration of rain gauge, radar & satellite-derived data in the production of daily rainfall maps. Progress report to the water research commission, for the period April 2001 to March 2002 on contract K5/1153, April 2002.

Pegram GGS, Clothier AN. 2001. High resolution space-time modelling of rainfall: the string of beads model. *Journal of Hydrology* **241**: 26–41.

Seo DJ. 1998a. Real-time estimation of rainfall fields using rain gage data under fractional coverage conditions. *Journal of Hydrology* **208**: 25–36.

Seo DJ. 1998b. Real-time estimation of rainfall fields using radar rainfall and rain gage data. *Journal of Hydrology* **208**: 37–52.

Todini E. 2001. A Bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements. *Hydrology and Earth System Sciences* **5**(2): 187–199.

Wilson JW, Brandes EA. 1979. Radar measurement of rainfall — a summary. *Bulletin of the American Meteorological Society* **60**(9): 1048–1058.

Copyright © 2005 Royal Meteorological Society

Hydrology and Earth System Sciences, 9, 127–137, 2005
www.copernicus.org/EGU/hess/hess/9/127/
SRef-ID: 1607-7938/hess/2005-9-127
European Geosciences Union

Hydrology and Earth System Sciences

Empirical Mode Decomposition in 2-D space and time: a tool for space-time rainfall analysis and nowcasting

S. Sinclair and G. G. S. Pegram

Civil Engineering, University of KwaZulu-Natal, Durban, South Africa

Received: 23 December 2004 – Published in Hydrology and Earth System Sciences Discussions: 10 February 2005
Revised: 20 May 2005 – Accepted: 12 July 2005 – Published: 22 July 2005

Abstract. A data-driven method for extracting temporally persistent information, at different spatial scales, from rainfall data (as measured by radar/satellite) is described, which extends the Empirical Mode Decomposition (EMD) algorithm into two dimensions. The EMD technique is used here to decompose spatial rainfall data into a sequence of high through to low frequency components. This process is equivalent to the application of successive low-pass spatial filters, but based on the observed properties of the data rather than the predetermined basis functions used in traditional Fourier or Wavelet decompositions. It has been suggested in the literature that the lower frequency components (those with large spatial extent) of spatial rainfall data exhibit greater temporal persistence than the higher frequency ones. This idea is explored here in the context of Empirical Mode Decomposition. The paper focuses on the implementation and development of the two-dimensional extension to the EMD algorithm and it's application to radar rainfall data, as well as examining temporal persistence in the data at different spatial scales.

1 Introduction

Spatial rainfall data contain information at a broad range of spatial scales (Schertzer and Lovejoy, 1987; Harris et al., 2001; Pegram and Clothier, 2001). It has been suggested in the literature (Seed, 2003; Turner et al., 2004) that the lower frequency components exhibit more temporal persistence than the higher ones; this premise is used here to prepare the data for nowcasts based on the evolution of the lower frequency components of space-time rainfall sequences. Examination of the (radially averaged) power spectrum (Fig. 1b) derived from a typical instantaneous estimate of rainfall rate obtained by weather radar (Fig. 1a) indicates that most of the power, hence potential for deterministic prediction in the context of nowcasting, is contained in the low frequency components.

In this paper we focus on a data-driven technique to extract the high frequency (less persistent in time) modes as the first step in a rainfall nowcasting scheme. The technique employed is a two-dimensional (2-D space) generalization of the one-dimensional Empirical Mode Decomposition (EMD) technique introduced by Huang et al. (1998). In a single dimension, EMD analysis produces a set of Intrinsic Mode Functions (IMFs) that are very nearly orthogonal; in two dimensions a set of Intrinsic Mode Surfaces (IMSs) is produced with similar quasi-orthogonal properties. Two-dimensional EMD appears to have been first introduced by Linderhed (2002) in the context of image compression; the key contribution in this paper is to introduce the concept of 2-D EMD to the Hydrometeorological literature as a tool for the analysis of space-time rainfall data. This paper focuses on the implementation and development of the two-dimensional extension of the EMD algorithm in this context, decomposing spatial rainfall data into its intrinsic spatial scale components.

In the application presented here, the least persistent IMS (exhibiting the highest local spatial frequency and least amount of spatial correlation – hence nearly white noise) is computed and removed from the raw rainfall data leaving a residual composed of the more persistent low frequency structural components in the data. This process is equivalent to using a low-pass spatial filter, based on the observed properties of the data rather than the predefined basis functions used in traditional Fourier or Wavelet scale decompositions. In Sects. 2 and 3, simple theoretical examples, showing the power of EMD in one and two dimensions, are presented as a "proof of concept" before applying the procedure to observed radar rainfall data from Bethlehem, South Africa (Sect. 4). These complement and extend the original presentation by

Correspondence to: S. Sinclair
(sinclaird@ukzn.ac.za)

© 2005 Author(s). This work is licensed under a Creative Commons License.

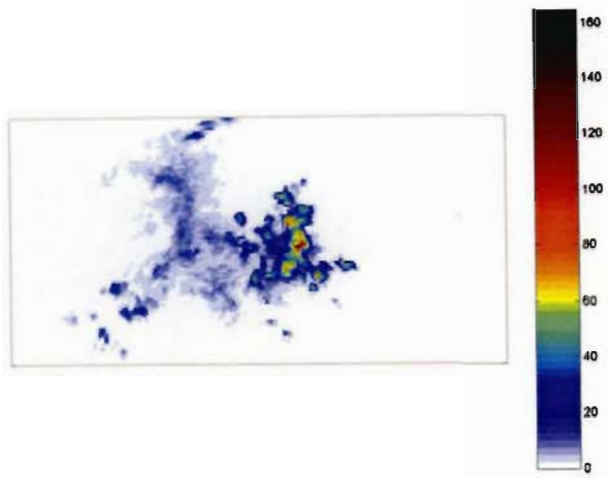


Fig. 1. **(a)** An observed convective rainfall field measured by S-Band weather radar at Bethlehem, South Africa (colour scale indicates instantaneous rain rate in mm/h). The image is 100×200 with 1 km^2 pixels.

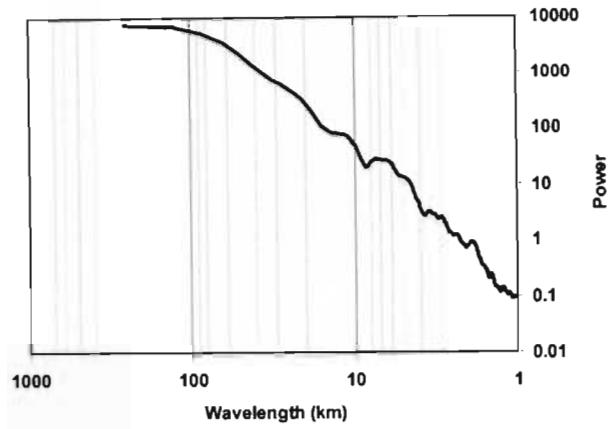


Fig. 1. **(b)** Radially averaged power spectrum of instantaneous rainfall rate from typical radar rainfall data shown in **(a)**.

Huang et al. (1998) and Flandrin et al. (2004). Computational aspects relating to image processing and surface fitting are covered in detail (Sect. 4) and conclusions are drawn in Sect. 5.

2 Empirical Mode Decomposition in a single dimension

The basic idea embodied in the EMD analysis, as introduced by Huang et al. (1998), is to allow for an adaptive and unsupervised representation of the intrinsic components of linear and non-linear signals based purely on the properties observed in the data without appealing to the concept of stationarity. As Huang et al. (1998) point out in their abstract: "This decomposition method is adaptive and therefore highly efficient. Since the decomposition is based on the local characteristic time scale of the data, it is applicable to nonlinear and non-stationary processes."

Few sequences of observations of natural phenomena are long enough to test the hypothesis of stationarity and frequently, the phenomena are patently non-stationary. This tacitly applies in the measurement of rainfall at a point or in space-time because sequences of rain are interspersed with dry periods and during the raining periods, the variability of the intensity due to mixtures of rainfall type (stratiform, convective, frontal) confound the stationarity definition. The EMD algorithm copes with stationarity (or the lack of it) by ignoring the concept, embracing non-stationarity as a practical reality. For a fuller discussion of the genesis of these ideas, see the Introduction of Huang et al. (1998), who also heuristically demonstrate the implicit orthogonality of the sequences of Intrinsic Mode Functions (IMFs) defined by the EMD algorithm.

In the application of the EMD algorithm, the possibly nonlinear signal, which may exhibit varying amplitude and local frequency modulation, is linearly decomposed into a finite number of (zero mean) frequency and amplitude modulated signals, as well as a residual function which exhibits a single extremum, is a monotonic trend or is simply a constant. Although EMD is a relatively new data analysis technique, its power and simplicity have encouraged its application in a myriad of fields. It is beyond the scope of this paper to give a complete review of the applications, however a few interesting examples are cited here to give the reader a feeling for the broad scope of applications. Chiew et al. (2005) examine the one-dimensional EMD of several annual streamflow time series to search for significant trends in the data, using bootstrapping to test the statistical significance of identified trends. The technique has been used extensively in the analysis of ocean wave data (Huang et al., 1999; Hwang et al., 2003) as well as in the analysis of polar ice cover (Gloersen and Huang, 2003). EMD has also been applied in the analysis of seismological data by Zhang et al. (2003) and has even been used to diagnose heart rate fluctuations (Balocchi et al., 2004).

2.1 Computing the one-dimensional EMD

The EMD algorithm extracts the oscillatory mode that exhibits the highest local frequency from the data ("detail" in the Wavelet context or the result of a high-pass filter in Fourier analysis), leaving the remainder as a "residual" ("approximation" in Wavelet analysis). Successive applications of the algorithm on the sequence of residuals produce a complete decomposition of the data. The final residual is a constant, a monotone trend or a curve with a single extremum.

The EMD of a one-dimensional data set $z(k)$ is obtained using the following procedure:

1. Set $r_0(k)=z(k)$ and set $i=1$.

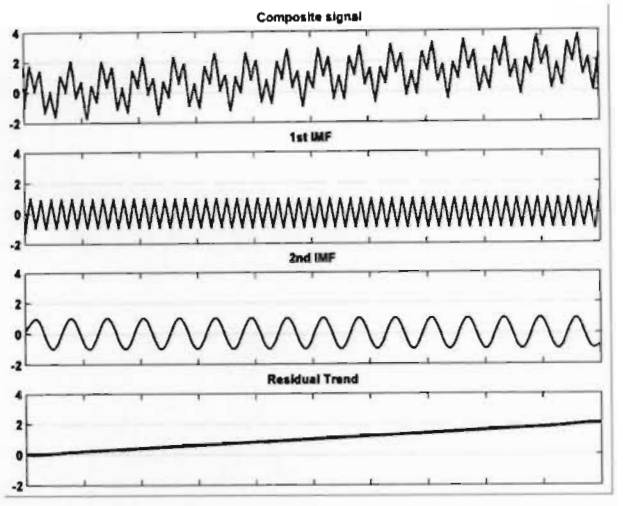


Fig. 2. EMD based signal separation; all IMFs are plotted to the same vertical scale. Top panel is the combined signal; lower 3 panels are the decomposition which recaptures, almost exactly, the original components.

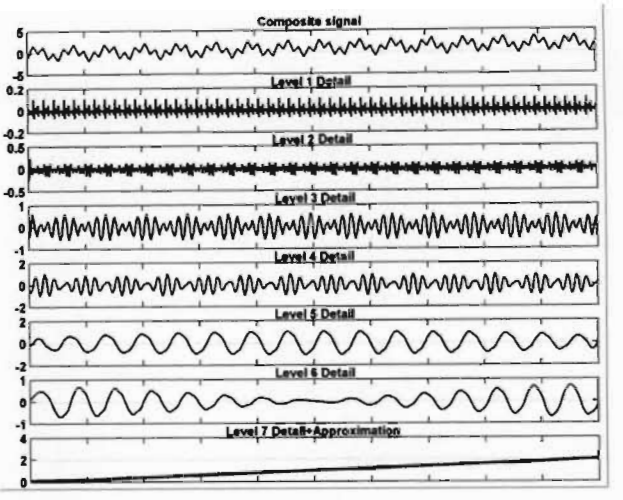


Fig. 3. Wavelet based signal separation – The "data" are the same as in Fig. 2, the vertical scale has been compressed for a compact presentation. An arbitrarily chosen db5 wavelet basis has been used.

2. Identify all of the extrema (maxima and minima) in $r_{i-1}(k)$.

3. Compute a maximal envelope, $\max_{i-1}(k)$, by interpolating between the maxima found in step 2. Similarly compute the minimal envelope, $\min_{i-1}(k)$. Cubic splines (as suggested by Huang et al., 1998) appear to be the most appropriate interpolation method for deriving these envelopes in one dimension (Flandrin et al., 2004).

4. Compute the mean value function of the maximal and minimal envelopes $m_{i-1}(k)=\frac{[\max_{i-1}(k)+\min_{i-1}(k)]}{2}$.

5. The estimate of the IMF is computed from $\mathrm{IMF}_i(k)=r_{i-1}(k)-m_{i-1}(k)$.

   Each IMF is supposed to oscillate about a zero mean and in practice it is necessary to perform a "sifting" process by iterating steps 2–5 (setting $r_{i-1}=\mathrm{IMF}_i$ before each iteration) until this is achieved.

6. Once the IMF_i has a mean value that is sufficiently close to zero over the length of the data (defined by a stopping criterion within some predefined tolerance ε) the residual $r_i(k)=r_{i-1}(k)-\mathrm{IMF}_i(k)$ is computed. Alternatively the sifting procedure can be stopped when the difference in the standard deviation of successive estimates of IMF_i falls below a critical threshold (Huang et al., 1998).

7. If the residual $r_i(k)$ is a constant or trend then stop; else increment i and return to step 2.

Figure 2 shows the EMD of a composite data series (shown in the first panel) that is the summation of a sine wave, a triangular waveform and a slowly varying trend. The compact representation obtained by EMD extracts (almost perfectly – except near the ends) the three separate time series (shown in panels 2 to 4) that make up the composite signal, without resorting to Fourier or Wavelet techniques with restrictive assumptions about the form of the underlying oscillatory modes or basis functions. Figure 3 shows the analysis of the same data, using Wavelet decomposition. Here a fifth order Daubechies wavelet basis was (arbitrarily) chosen for illustration purposes; this choice of basis function may not be optimal for detrending but serves to demonstrate a typical decomposition. Seven levels of decomposition were required before the trend became apparent; this decomposition is clearly far less compact and physically meaningful than the EMD results in this case.

A similar decomposition analysis can be carried out using Fourier techniques. The Discrete Fourier approximation of a signal can be defined in terms of the Euler-Fourier coefficients (a_0, a_k, b_k) with $k=1, 2, \ldots, m$ (Eq. 1). The coefficients are all that are required to reconstruct the series and any signal can be well approximated (as long as it satisfies the Dirichlet conditions), provided m is sufficiently large. In Eq. (1), $F(x_j)$ is the Fourier approximation of the signal y_j at each of the n discrete (evenly spaced) data points x_j. L is the range of values x_j over which the data set is assumed periodic.

$$F(x_j) = \frac{a_0}{2} + \sum_{k=1}^{m} \left\{ a_k \cos(2\pi k x_j / L) + b_k \sin(2\pi k x_j / L) \right\} \quad (1)$$

$$a_k = 2 \sum_{j=1}^{n} \cos(2\pi k x_j / L) y_j / n \quad (2)$$

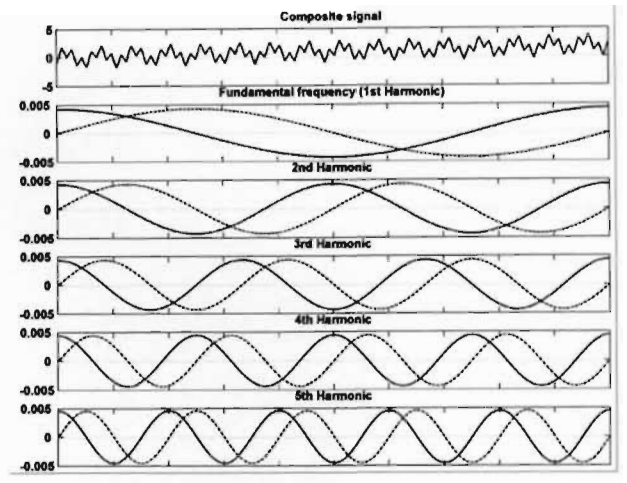


Fig. 4. Fourier based signal separation, the first 5 of 75 components – The dashed lines show the sine component and the solid lines the cosine component.

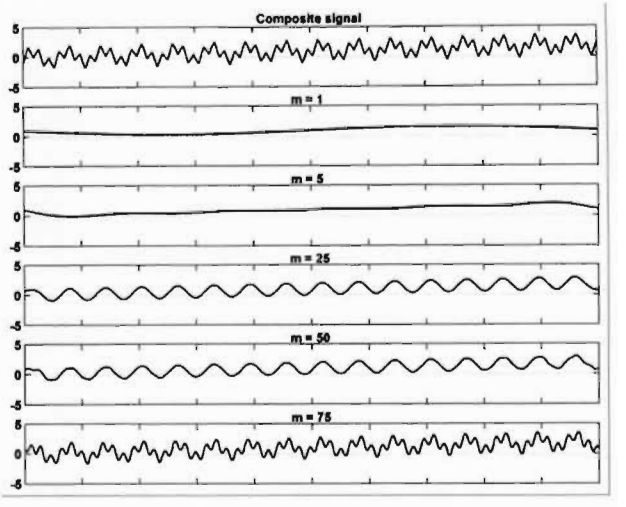


Fig. 5. Reconstruction of the signal from the sine and cosine components, m represents the number of Euler-Fourier coefficients used in each reconstruction.

$$b_k = 2\sum_{j=1}^{n} \sin(2\pi k x_j/L) y_j/n. \quad (3)$$

Figures 4 and 5 show the result of decomposing the data using a finite Fourier series. Figure 4 shows the first 5 harmonics; while Fig. 5 shows the series reconstruction by accumulating the lower harmonics up to m. Computing the Euler-Fourier coefficients provides a compact approximation of the original signal but fails to extract physically meaningful information. The ability to determine meaningful structural information is clearly important in a nowcasting context, which cannot be bound by the periodicity assumption implicit in Fourier methods.

3 Empirical Mode Decomposition in two dimensions

In two dimensions the EMD process is conceptually the same as for a single dimension, except that the curve fitting exercise becomes one of surface fitting and the identification of extrema becomes (a little) more complicated. Very little work appears to have been done which applies the EMD technique to two-dimensional data. Han et al. (2002) use EMDs in one dimension along four different directions to smooth Synthetic Aperture Radar (SAR) images and remove speckle. Nunes et al. (2003) develop a technique, which they term "Bidimensional Empirical Mode Decomposition" (BEMD) in the context of texture analysis in image data where they demonstrate several examples of intrinsic mode extraction from image data. Linderhed (2002, 2004[1]) examined the use of EMD in two dimensions for image compression. Both of these implementations are very similar to what we propose in this paper. The 2-D EMD provides a truly two-dimensional analysis of the intrinsic oscillatory modes inherent in the data. Two-dimensional Fourier and Wavelet analyses are really applications of their one-dimensional counterparts in a number of principal directions. Fourier analysis concentrates on orthogonal "East-West" and "North-South" directions (e.g. Press et al., 1992). Wavelet analysis can, in general, consider any direction of the wavelet relative to the data, however a typical 2-D Wavelet analysis examines only horizontal, vertical and diagonal orthonormal wavelet basis functions (Daubechies, 1992, pp. 313; Kumar and Foufoula-Georgiou, 1993). In contrast, EMD produces a fully two-dimensional decomposition of the data, based purely on spatial relationships between the extrema, independent of the orientation of the coordinate system in which the data are viewed.

3.1 Description of the algorithm

The algorithm follows intuitively from the one-dimensional case and may be briefly summarised as follows:

1. Locate the extrema in the 2D space including maximal and minimal plateaus.
2. Generate the bounding envelopes using appropriate surface fitting techniques. We suggest conical Multiquadrics (for reasons explained in Sect. 3.2).
3. Compute the mean surface function as the average value of the upper and lower envelopes.
4. Determine the first estimate of an IMS by subtracting the mean surface from the data.
5. Iterate until the IMS mean surface function is close to zero everywhere.

[1] http://www.icg.isy.liu.se/~anna/emd-samp.pdf

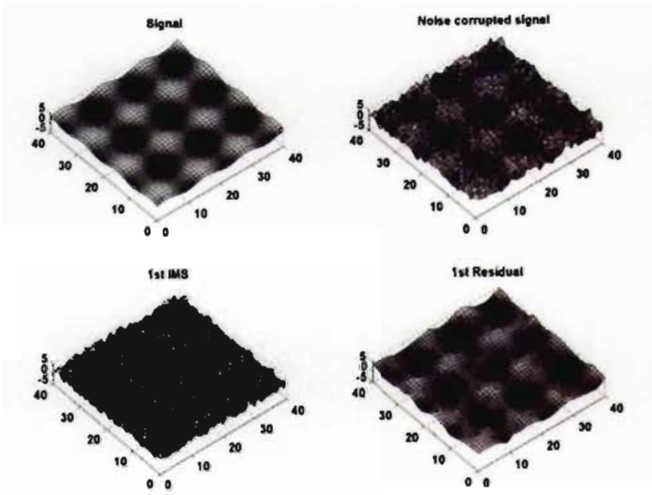


Fig. 6. Example of EMD used for noise removal on a 2-D sine wave. The bulk of the additive white noise in the corrupted signal is well captured by the first IMS.

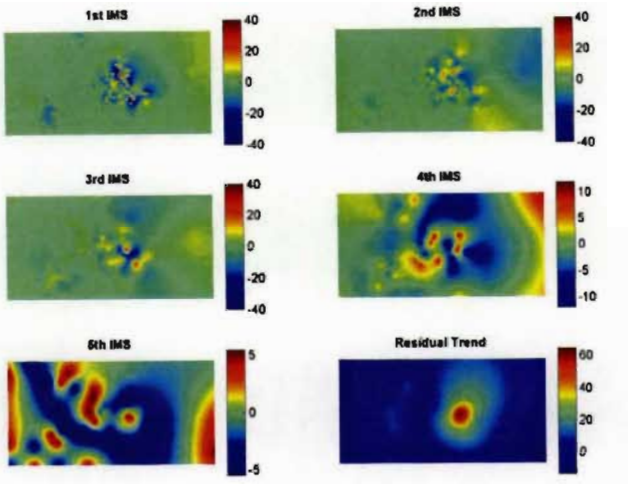


Fig. 7. Naïve EMD of the observed rainfall field shown in Fig. 1a – note the change in scale of the rain rates in the IMSs.

6. Estimate the IMS and Residual.

7. If the Residual is a constant or a monotone trend, then stop; else return to step 2.

3.2 Surface fitting for extremal envelope generation

The generation of maximal and minimal envelopes is of key importance to a successful 2-D EMD implementation and is the most computationally intensive task. The problem is a familiar one of collocating a smooth surface to randomly scattered data points in two-dimensions. There are several options available to achieve this. Ultimately the fitting procedure reduces to computing the unknown value of the surface at a point $s_i=(x_i, y_i)$, by some linear (or nonlinear) weighting of the known data. In general, a basis function determines the influence of each known data point based on its spatial position relative to the unknown point s_i. Nunes et al. (2003) use radial basis functions while Linderhed uses bi-cubic splines (Linderhed, 2002) and later chooses the more suitable option of Thin Plate Splines (Linderhed, 2004). We use radial basis functions (technically, conical Multiquadrics), which are identical to Kriging (Cressie, 1991) with a purely linear semi-variogram model. It could perhaps be argued that it would be more appropriate to fit a semi-variogram model to the maxima and minima, but we feel this would be over-elaborate and presumptuous, as the extrema are only related by distance and cannot be considered drawn from a stationary correlated random field. Invoking Occam's razor in the spirit of Huang's original derivation of EMD, we wish to let the data do the talking and conical Multiquadrics assume the least structure of any linear surface fitting algorithm.

The Ordinary Kriging estimate $\hat{z}_i$ at any point i based on n observed data points is

$$\hat{z}_i = \sum_{k=1}^{n} \lambda_k z_k, \tag{4}$$

where z_k are the observations and λ_k are weights associated with each observation and the target point. The mean is assumed unknown and the weights λ_k are constrained to sum to unity. The vector of weights λ is obtained by solving the linear system in Eq. (5)

$$\begin{bmatrix} \Gamma & u \\ u^T & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma \\ 1 \end{bmatrix}, \tag{5}$$

where γ is a vector of semivariogram values, in this application simply defined by the linear distance basis function $\gamma(s_{ij})=|s_{ij}|$ with s_{ij} the distance between point i and the $j=1, 2, ..., n$ observation locations. Γ is the matrix of distances between the observations, $\boldsymbol{u}$ is a vector of n ones and μ is a Lagrange multiplier ensuring that the Kriging weights λ_k sum to unity, as required. The solution of Eq. (5) is obtained using Singular Value Decomposition (SVD) in this application to ensure that a stable solution is assured (when the matrix is ill conditioned). This is achieved by truncating singular and near-singular components. Although SVD is computationally less efficient than (for example) LU decomposition as a means of solving a dense linear system, it's use is preferred here because of it's robustness in the face of the near-singular Kriging systems which are frequently encountered in gridded data applications (Wesson and Pegram, 2004).

A more efficient choice of interpolation technique would be useful and more work could be done in this regard, however care is required. Moving-neighbourhood Kriging (a possible alternative to reduce the number of control points)

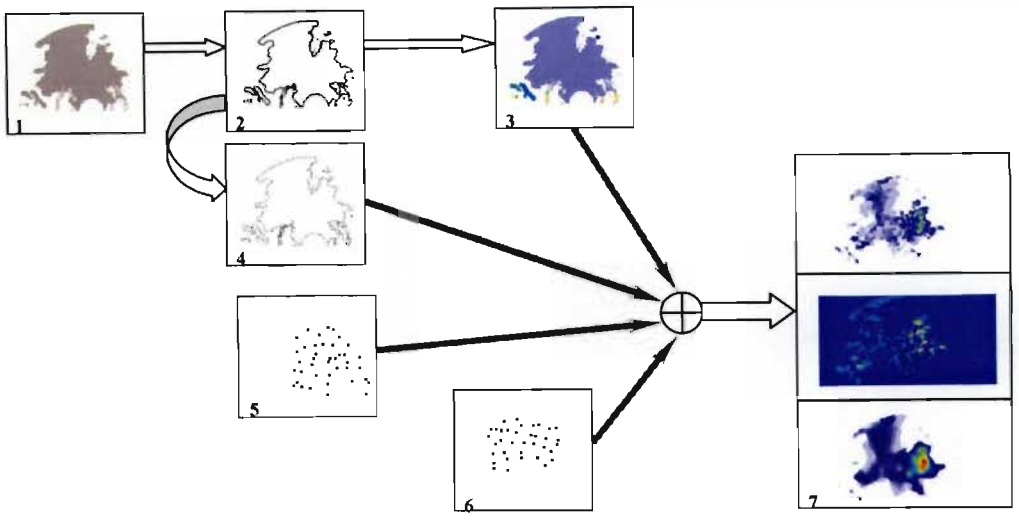


Fig. 8. Summary of data processing 1. Mask the wet and dry areas; 2. Trace the boundary of each wet region; 3. Separately label each wet region; 4. Decimate the fence by a factor of 5, isolating the "fence posts"; 5. Isolate the maxima in each sub-region; 6. Isolate the minima in each sub-region; 7. EMD analysis decomposes the data into the first IMS and the first residual using the maximal and minimal envelopes defined using the points in 4, 5 and 6.

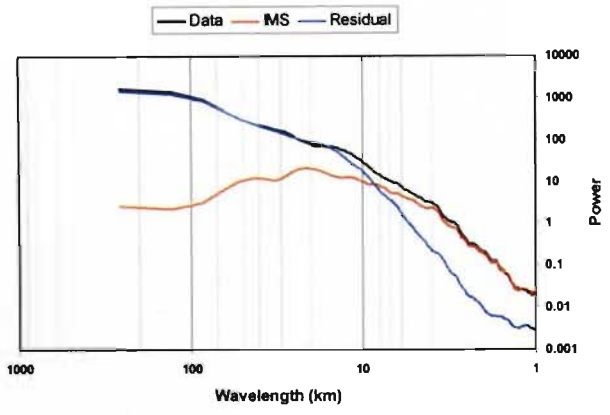


Fig. 9. Comparison of individual radially-averaged power spectra of the radar rainfall data (of Fig. 1a) with its EMD components: the first IMS and the first residual.

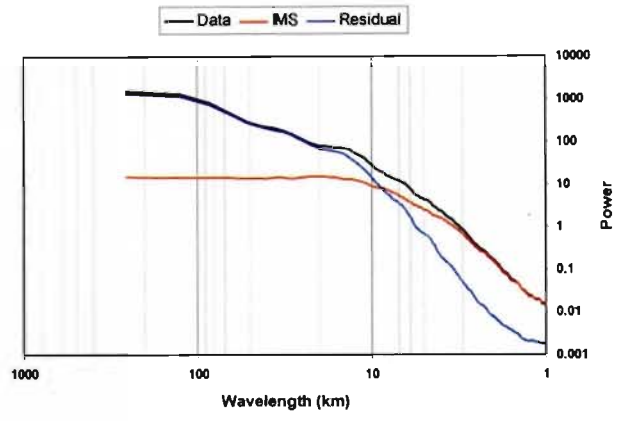


Fig. 10. The same as Fig. 9 but for the mean of individual power spectra for five consecutive, radar scans – Beginning with the spectra shown in Fig. 9.

can produce unwanted discontinuities in regions that are data sparse (Chiles and Delfiner, 1999, pp. 201), such discontinuities would be amplified through the EMD sifting process. In addition, the particular choice of Ordinary Kriging as a method of generating the bounding envelopes was (partially) directed by the property that the estimates decay asymptotically to the mean of the observed extrema.

3.3 Simple two-dimensional EMDs

In this section, applications of the 2-D EMD technique are presented. As an artificially constructed example Fig. 6 shows the successful removal of noise added to a synthetically generated two-dimensional sine signal. The noise (with it's high local spatial frequency) is almost completely described by the first IMS leaving a residual, which is closely representative of the underlying signal.

Turning to a realistic example of the type we have been aiming for, Fig. 1a showed an instantaneous radar rainfall field with an area of 100×200 km. A complete EMD of this field is shown in Fig. 7 using a direct application of the 2-D EMD process described in Sect. 3.1; note the change in scale of the individual IMSs. The final residual (with a single extremum) gives a clear indication of the position of the largest convective raincell evident in the field.

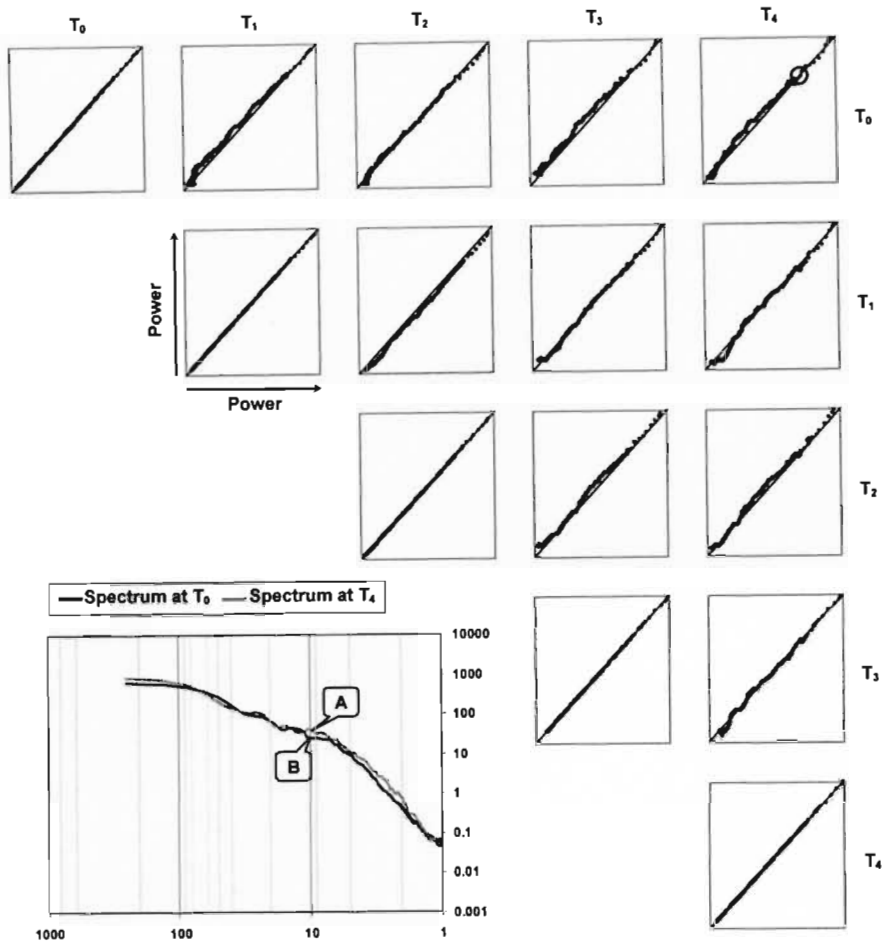


Fig. 11. Spectral persistence scatter plots of the original data for a sequence of rainfall fields and those at successive intervals. This is constructed by plotting the values of power for each field at corresponding wavelengths coaxially. For example points *A* and *B* at the 10 km wavelength are plotted against each other and appear ringed in the upper right diagram.

4 Application of 2-D Empirical Mode Decomposition to rainfall data

The simple 2-D EMD application presented in the previous section is computationally burdensome when applied to rainfall data. In this section, to overcome this drawback, a number of specific refinements are presented which combine to make EMD tractable in practical real-time situations.

4.1 Image processing techniques and optimisations

Since an application of 2-D EMD requires the use of surface fitting techniques, large linear systems must be solved. The size of a system is determined by the number of known data points which are to be used in combination to find the unknown values of the surface at each remaining position in the field. The highly variable nature of rainfall data means that the field contains a large number of extrema from which the bounding envelopes must be constructed. Additionally there are a large number of zero (no rain) data, which constitute minima. By only considering raining areas, the size of the linear systems requiring a solution are greatly reduced since each raining area (if more than one exists) will contain a considerably smaller number of extrema than the entire data region and each can be treated separately. Furthermore, it makes no sense to consider an EMD in areas where the variable of interest does not exist, in this case the areas that are not raining.

A number of well-known image processing techniques are implemented to isolate and process each raining area. Fig. 8 summarises the steps taken in processing the data with the boxes numbered 1–7 indicating different steps in the process. First a mask is generated to separate the raining and non-raining pixels (Fig. 8, Box 1) in the instantaneous radar image; pixels below a threshold of 1 mm/hr are considered as non-raining and the remaining pixels are marked as raining.

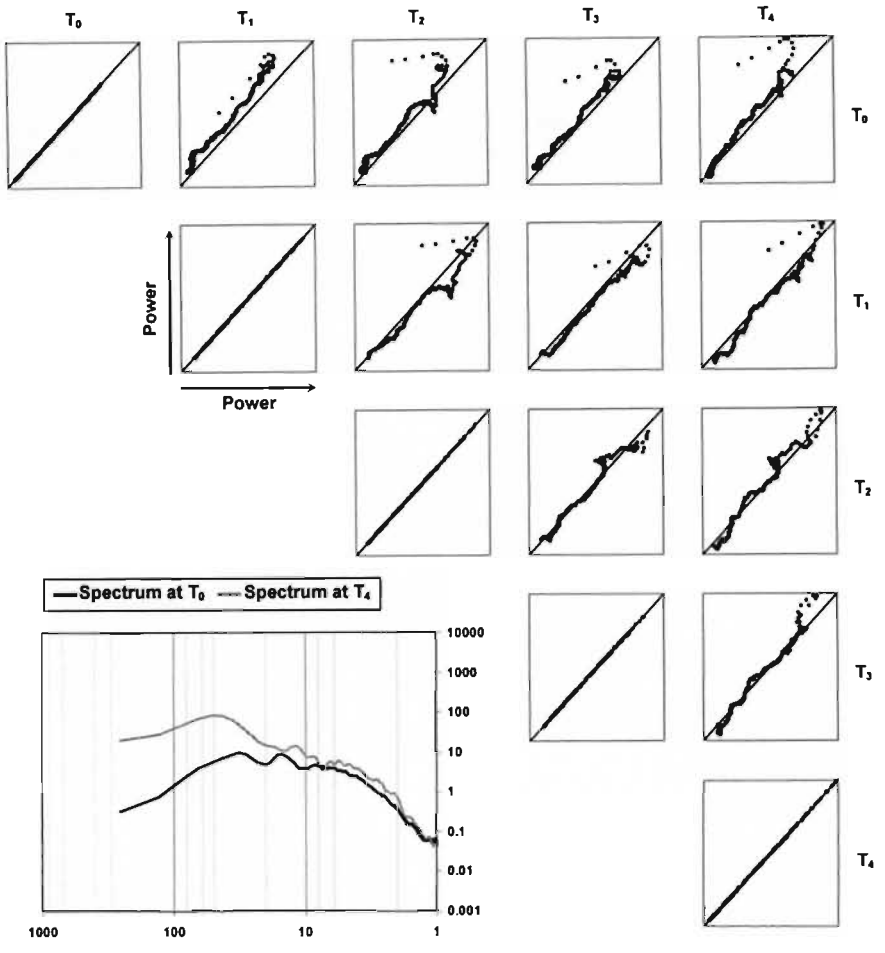


Fig. 12. Spectral persistence scatter plots of the sequence of 1st IMSs of each pair of rainfall fields $T_0, \cdots, T_4$.

An outer boundary border-tracing algorithm (Sonka et al., 1999) is used to establish a boundary "fence" around each raining area (Fig. 8, Box 2) and a flood-fill procedure is then used to fill each raining area with a unique identifier, resulting in separately labelled raining regions (Fig. 8, Box 3). To reduce the computational burden of the algorithm even further, the boundary "fence" is decimated by a factor of 5 to reduce the continuous string of border points to "fence posts" while retaining the gross shape of the raining areas (Fig. 8, Box 4). The next step in the processing of the data is to isolate the extrema in the rainfall field (Fig. 8, Boxes 5 and 6). There are numerous possible techniques for identifying extrema in the rainfall field. Nunes et al. (2003) use a morphological reconstruction technique. One alternative, which was explored, is based on image segmentation and detection of extremal plateaus. However, our method of choice was to use a simple 8 neighbour search routine for identification of pixels with extreme values as done by Linderhed (2004). The choice was partly for convenience, but also because the majority of the (non-zero) extreme values in the rainfall fields studied turned out to consist of single pixels. There is a rich literature on image processing techniques and the reader is referred to an introductory text such as Sonka et al. (1999) to explore the field further. Finally, the EMD analysis is carried out using the extrema within each raining area and the zeros at the "fence posts" of non-raining border pixels to specify the extremal envelopes (Fig. 8, Box 7). Only one step of decomposition is shown here – the data is decomposed into the noisy first IMS and the first residual.

4.2 Results

An analysis of over 800 individual radar scans, embodying mixtures of various ratios of Stratiform and Convective rainfall types, was carried out to determine the effectiveness of the 2-D EMD algorithm in separating the high frequency spatial components from the original rainfall data. Working on the basis that the average characteristics of the data over a range of spatial scales summarised by the power spectrum is intuitively useful, the (radially averaged) power spectra

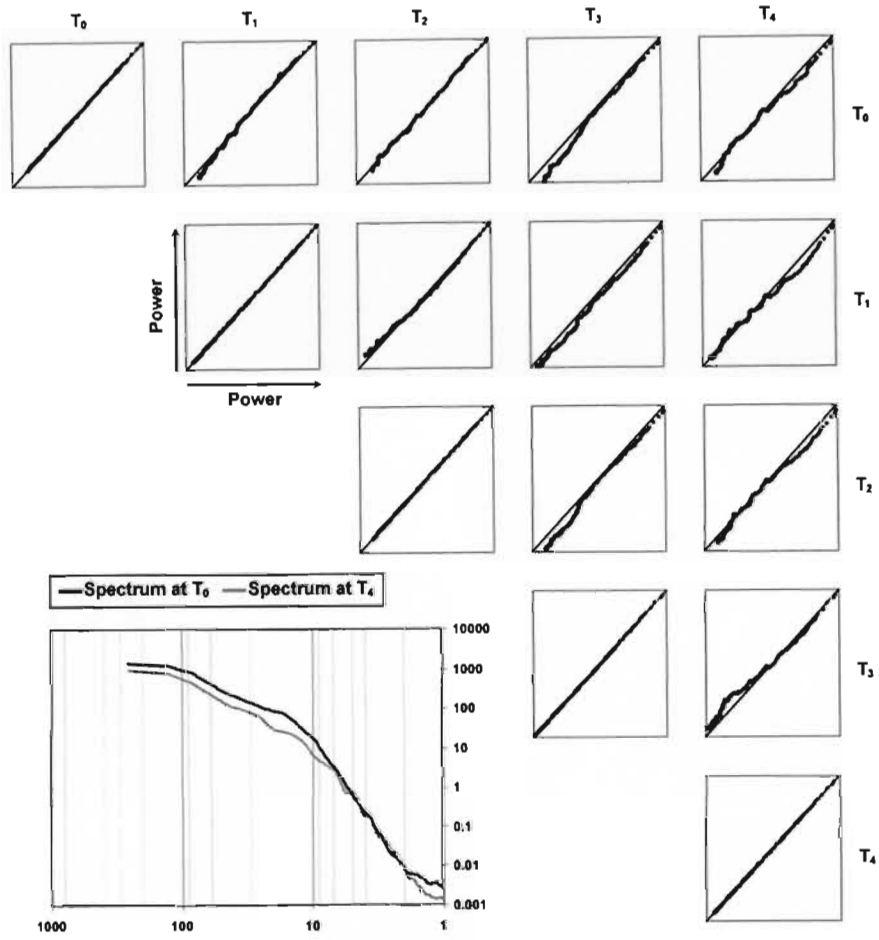


Fig. 13. Spectral persistence scatter plots of the sequence of 1st Residual of each pair of rainfall fields $T_0, \cdots, T_4$.

of (i) the original data, (ii) the first IMS and (iii) the first residual of each image were examined and compared. Fig. 9 shows a typical result; the power spectrum of the residual shows a very close correspondence with that of the original data at high wavelengths while it contains far less power at the lower wavelengths. In contrast, the spectrum of the first (noisy) IMS has very little power relative to the data's spectrum at high wavelengths but shows a strong correspondence at the lowest wavelengths. Fig. 9 clearly indicates how the 2-D EMD technique moves the bulk of the high frequency components in the original data into the first IMS and leaves the high power, lower frequencies in the residual. Fig. 10 shows a time average of this behaviour by plotting the mean values at each wavelength of the three spectra over five consecutive radar scans (beginning with the data used to produce Fig. 9). The radar scans are captured at approximately five-minute intervals. It is interesting to observe that the average of the spectra of the first IMSs is flat for wavelengths longer than 10 km, suggesting nearly white noise over this range.

The temporal persistence exhibited at the spatial scales represented in each of the three sequences of: (i) the data, (ii) the first IMS and (iii) the first residual was examined by considering their temporally consecutive power spectra. The notion of "spectral persistence" was used to determine how variable the spatial structure (at a particular spatial scale) is in time and hence to give an indication of the temporal predictive capability at each spatial scale. A summarised example of the analysis of a sequence of 5 radar rainfall images is presented in Figs. 11, 12 and 13 where a "matrix" of scatter plots is shown in each case. Scatter-plots of the pairs of power values at each discrete wavelength for five consecutive spectra (with the 1:1 line indicated) are shown for, the original data (Fig. 11) the first IMS (Fig. 12) and the first residual (Fig. 13). The rows and columns of the scatter-plot matrices are labelled from T_0 to T_4 and indicate separate radar scans between time $T=0$ and time $T=4$. Each block in the scatter-plot matrix represents a scatter-plot of the power at each wavelength for the spectrum computed at T_i versus that of the spectrum computed at T_j. Clearly the plots on the

"matrix" diagonal each compare a spectrum to itself and a perfect 1:1 relationship is observed in this case. For the off-diagonal plots, the degree of scatter amongst the data points indicates the degree of similarity between the spectra at individual wavelengths at increasing time lags with a large scatter indicating a weak similarity. The trends shown here are typical of the data analysed and show how the first (high average frequency) IMS has a temporally incoherent spatial structure, while the first (low average frequency) residual shows a temporally consistent structure. The behaviour shown in Figs. 9–13 suggests that the high frequency IMS components in spatial rainfall data do not contain much predictive capability, supporting the suggestions of Seed (2003) and Turner et al. (2004) to increase the degree of spatial smoothing and rely more on the information contained in the lower frequency components as forecast lead times increase.

5 Conclusions

A new technique for analysing the spatial scaling structure of rainfall fields has been presented. The technique is a two dimensional extension of Empirical Mode Decomposition for the analysis of non-linear and non-stationary time series. An EMD analysis in two dimensions linearly decomposes the spatially distributed rainfall data into a set of Intrinsic Mode Surfaces, which are approximately mutually orthogonal and sum back to the original data. Each IMS contains an oscillatory mode inherent in the data at a different (narrow) range of spatial frequencies. The EMD analysis successively extracts the IMS with the highest local spatial frequencies in a recursive way, which is effectively a set of successive low-pass spatial filters based entirely on the properties exhibited by the data. The utility of the EMD technique for signal separation has been demonstrated in both one and two dimensions and applied to the analysis of a large set of 800 radar rainfall images in South Africa. The 2-D EMD technique is proposed here in the context of rainfall nowcasting to separate the less persistent high frequency components from the more persistent low frequency ones in the data. The aim is to remove the noisy high frequency components, which do not exhibit a strong temporal correlation and add little structural information to nowcasting algorithms. The scale separation achieved by 2-D EMD has been analysed using radially averaged power spectra to summarise the spatial structure of the data and filter outputs. In addition these power spectra have also been used to examine the temporal persistence of the spatial structure exhibited by the first IMS and it's residual. The results presented in this paper agree with other work in the hydrometeorological literature, which suggests that the low frequency spatial components in rainfall data are most useful in a nowcasting context. This methodology is being exploited in ongoing research into rainfall nowcasting.

Acknowledgements. The authors would like to thank the Water Research Commission of South Africa for supporting the work presented herein through contracts K5/1249 and K5/1425. The contributions of two anonymous reviewers were also helpful in improving the submission.

Edited by: M. Sivapalan

References

Balocchi, R., Menicucci, D., Santarcangelo, E., Sebastiani, L., Gemignani, A., Gellarducci, B., and Varanini, M.: Deriving the respiratory sinus arrhythmia from the heartbeat time series using empirical mode decomposition, Chaos, Solitons and Fractals, 20, 1, 171–177, 2004.

Daubechies I.: Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.

Chiew, F. H. S., Peel M. C., Amirthanathan, G. E., and Pegram, G. G. S.: Identification of oscillations in historical global streamflow data using empirical mode decomposition. Regional Hydrological Impacts of Climatic Change – Hydroclimatic Variability, Proceedings of symposium S6 held during the Seventh IAHS Scientific Assembly at Foz do Iguaçu, Brazil, IAHS Publ. 296, 53–62, 2005.

Chiles, J. P. and Delfiner P.: Geostatistics: Modelling spatial uncertainty, Wiley, 1999.

Flandrin, P., Rilling, G., and Goncalves, P.: Empirical Mode Decomposition as a filter bank, IEEE Signal Processing Letters, 11, 2, 112–114, 2004.

Gloersen, P. and Huang, N.: Comparison of Interannual Intrinsic Modes in Hemispheric Sea Ice Covers and Other Geophysical Parameters, IEEE Transactions on Geoscience and Remote Sensing, 41, 5, 1062–1074, 2003.

Han, C., Guo, H., Wang, C., and Fan, D.: A novel method to reduce speckle in SAR images, Int. J. Remote Sensing, 23, 23, 5095–5101, 2002.

Harris, D., Foufoula-Georgiou, E., Droegemeier, K. K., and Levit, J. J.: Multiscale Statistical Properties of a High-Resolution Precipitation Forecast, J. Hydrometeor., 2, 406–418, 2001.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N., Tung, C. C., and Liu, H. H.: The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis, Proceedings: Mathematical, Physical and Engineering Sciences, Royal Society London, 454, 1971, 903–995, 1998.

Huang, N. E., Shen, Z., and Long, S. R.: A new view of Nonlinear water waves: the Hilbert Spectrum, Ann. Rev. Fluid Mech., 31, 417–457, 1999.

Hwang, P. A., Huang, N. E., and Wang, D. W.: A note on analyzing nonlinear and nonstationary ocean wave data, App. Ocean Res., 25, 4, 187–193, 1999.

Linderhed, A.: 2-D empirical mode decompositions – in the spirit of image compression, Proceedings of SPIE, Wavelet and Independent Component Analysis Applications IXI, Volume 4738, 2002.

Nunes, J. C., Bouaoune, Y., Delechelle, E., Niang, O., and Bunel, P.: Image analysis by bidimensional empirical mode decomposition, Image and Vision Computing, 21, 1019–1026, 2003.

Pegram, G. G. S. and Clothier, A. N.: High Resolution Space-Time Modelling of Rainfall: The "String of Beads" Model, J. Hydrol., 241, 26–41, 2001.

Press, W. H., Teukolsky S. A., Vetterling W. T., and Flannery, B. P.: Numerical Recipes in C – The art of scientific computing, Second Edition, Cambridge University Press, 1992.

Schertzer, D. and Lovejoy, S.: Physical Modelling and Analysis of Rain and Clouds by Anisotropic Scaling Multiplicative Processes, J. Geophys. Res., 92, D8, 9693–9714, 1987.

Seed, A. W.: A Dynamic and Spatial Scaling Approach to Advection Forecasting, J. Appl. Meteor., 42, 381–388, 2003.

Sonka, M., Hlavac, V., and Boyle, R.: Image Processing, Analysis and Machine Vision, 2nd Edition, PWS Publishing, Pacific Grove, 1999.

Turner, B. J., Zawadzki, I., and Germann, U.: Predictability of precipitation from continental radar images. Part III: Operational nowcasting implementation (MAPLE), J. Appl. Meteor., 43, 231–248, 2004.

Wesson, S. M. and Pegram, G. G. S.: Radar rainfall image repair techniques, Hydrol. Earth Syst. Sci., 8, 220–234, 2004, **SRef-ID: 1607-7938/hess/2004-8-220**.

Zhang, R. R., Ma, S., and Hartzell, S.: Signatures of the seismic source in EMD-based characterization of the 1994 Northridge, California, earthquake recordings, Bulletin of the Seismological Society of America, 93, 1, 501–518, 2003.