

# The University of Reading



## Application of satellite-based rainfall estimates to crop yield forecasting in Africa

Chee-Kiat Teo

A thesis submitted for the degree of Doctor of Philosophy

Department of Meteorology

October 2006

# Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Chee-Kiat Teo

# Acknowledgments

I would like first to express my sincere gratitude to my supervisor Dr. David Grimes for his guidance and encouragement throughout my research. His patience and commitment, in my opinion went well beyond what I expected when I first began my research. It has been a pleasure working with him.

This thesis would not be possible without the advice and guidance from Dr. Andrew Challinor and Dr. Tim Wheeler, particularly in the field of crop modelling and agriculture. Both of them never fail to find time to answer any queries that I had along the way. Prof. Julia Slingo and Dr. Janet Barlow who were on my thesis committee along with David, Tim and Andy have been extremely helpful throughout. I am indebted to them for keeping me on track and giving me suggestions that help shaped the course of my work.

My correspondence with Dr. Eulogio Pardo-Igúzquiza, Rene Gomme and Dr. Sergio Pezzulli has been very useful in many aspect of this research. Many thanks as well to the following people for technical and data support: Mr. Ian Russell from TAMSAT in providing the satellite data; Mr. Bernard Gomez from the Department of Water Resources, The Gambia, for the weather data; Mr. Ebrima Camara and Mr. Komagy Camara from the Department of Planning, The Gambia, for the crop data; Dr. Musa Bojang and his staff in the National Agriculture Research Institute (NARI) for their logistical support and hospitality when I was in the Gambia during October 2003.

My stay in the Reading for the past few years would not have been as enjoyable if not for all the good friends that I have the fortune to meet along the way. In particular, Pascal Mailer and Evangelos Tyrlis whom have given me many intangible rewards from the friendship we share. Pascal, Vagelis, thank you.

Finally, to my mother, my brothers and Aunt Jenny: No word could ever express my gratitude for the support they have given me all these years.

# Abstract

The livelihood of people in many regions of Africa where rainfall is marginal depends on rainfed agriculture. Accurate prediction of crop yield could greatly ameliorate potential famine and allow advanced planning of intervention operations. This thesis explores the feasibility of a combined satellite-based rainfall estimates (RFE) and crop model for crop yield forecasting in Africa, using groundnut yield for the Gambia as a case study. An important focus in the investigation was also to assess the impact of errors in the RFE on the simulated crop yield.

A satellite-based rainfall estimates algorithm, TAMSAT1, based on the original TAMSAT algorithm of Milford and Dudgale (1990) has been developed to better quantify the highly non-Gaussian distribution of the daily Gambian rainfall. The new algorithm outperformed the TAMSAT algorithm in various skill measures. A novel method, SIMU, has been developed to generate spatial rainfall fields stochastically, in order to quantify the random uncertainties in the TAMSAT1 RFE at any arbitrary spatial scale. The results from validating these simulated rainfall fields with gauge estimated pixel rainfall are encouraging; SIMU was able to reproduce the pixel-level statistics and spatial structure of the daily rainfall field, lending support to its ability to simulate realistic rainfall fields, conditional to the satellite data.

TAMSAT1-RFEs are used to drive a process-based crop model, GLAM, and a simpler empirical crop model based on crop water balance (CSWB) for hindcasting divisional groundnut yields in the Gambia. The main results from these crop modelling experiments are as follow:

- 1) The simulated crop yield from GLAM is more sensitive to the intraseasonal rainfall variability than the total seasonal rainfall amount due to non-linear feedbacks between the model-crop development and the water availability. As CSWB does not explicitly simulate crop processes and determines crop yield empirically from total crop water

usage. Hence, simulated crop yield from CSWB is more sensitive to the total seasonal rainfall amount than the intraseasonal variability of rainfall.

2) The effects random error in TAMSAT1-RFE have on both crop model yield predictions are secondary compared to the effects from planting date uncertainty, as well as systematic bias in TAMSAT1-RFE.

3) Since RFE are able to capture the intraseasonal variability of rainfall better than total seasonal amount, the RFE driven processed-based crop model outperforms a RFE driven CSWB model in crop yield predictions.

# Contents

<b>CHAPTER 1 INTRODUCTION .....</b>	<b>1</b>
<b>1.1 Background.....</b>	<b>1</b>
<b>1.2 Motivation and scope of thesis .....</b>	<b>13</b>
<b>1.3 Plan of the thesis.....</b>	<b>15</b>
<b>CHAPTER 2 DESCRIPTION AND PRELIMINARY ANALYSIS OF DATA.....</b>	<b>17</b>
<b>2.1 Introduction .....</b>	<b>17</b>
<b>2.2 Agrometeorological and satellite data.....</b>	<b>17</b>
<b>2.3 Preliminary analysis of Gambian rainfall.....</b>	<b>20</b>
<b>2.4 Gambian groundnut production.....</b>	<b>27</b>
<b>2.5 Conclusion.....</b>	<b>34</b>
<b>CHAPTER 3 SPATIAL ESTIMATION OF RAINFALL FIELD FROM GAUGE DATA .....</b>	<b>36</b>
<b>3.1 Introduction .....</b>	<b>36</b>
<b>3.2 Stochastic spatial interpolation using double kriging.....</b>	<b>38</b>
<b>3.3 Gambian daily rainfall field .....</b>	<b>40</b>
3.3.1 Point validation .....	43
3.3.2 Extension to obtain mean pixel rainfall estimates.....	46
<b>3.4 Conclusion.....</b>	<b>47</b>
<b>CHAPTER 4 SATELLITE RAINFALL ESTIMATION.....</b>	<b>50</b>

<b>4.1 Introduction .....</b>	<b>50</b>
<b>4.2. TAMSAT method.....</b>	<b>56</b>
<b>4.3 TAMSAT1: Extension of TAMSAT .....</b>	<b>59</b>
4.3.1 Modification to TAMSAT for daily rainfall estimation.....	61
<b>4.4 Case study .....</b>	<b>65</b>
4.4.1 Calibration.....	66
4.4.2 Validation .....	75
<b>4.5 Discussion.....</b>	<b>82</b>
<b>4.6 Conclusion.....</b>	<b>89</b>
<b>CHAPTER 5 STOCHASTIC GENERATION OF RAINFALL USING SATELLITE DATA .....</b>	<b>91</b>
<b>5.1 Introduction .....</b>	<b>91</b>
<b>5.2 The problem.....</b>	<b>95</b>
5.2.1 Model and assumptions .....	97
5.2.2 Estimation of conditional parameters.....	100
<b>5.3 Sequential methods of simulating rainfall .....</b>	<b>102</b>
5.3.1 Sequential Indicator Simulation .....	103
5.3.2 Sequential Gaussian Simulation.....	103
5.3.3 Summary of simulation of daily rainfall conditional to CCD .....	104
5.3.4 A simple simulation example.....	108
<b>5.4 Simulation of daily Gambian rainfall by sequential methods.....</b>	<b>113</b>
5.4.1 Covariance Modelling .....	113
5.4.2 Implementation specifics.....	116
5.4.3 Validation .....	119

<b>5.5 Possible improvement .....</b>	<b>127</b>
<b>5.5 Conclusion.....</b>	<b>131</b>
 <b>CHAPTER 6 APPLICATION OF SATELLITE RAINFALL ESTIMATES IN CROP YIELD ESTIMATION.....</b>	 <b>133</b>
<b>6.1 Introduction .....</b>	<b>133</b>
<b>6.2 Crop model descriptions.....</b>	<b>137</b>
<b>6.3 Datasets .....</b>	<b>148</b>
<b>6.4 Model parameters selection and calibration.....</b>	<b>150</b>
<b>6.5 Sensitivity experiments on crop model.....</b>	<b>157</b>
<b>6.6 Crop model prediction using gauge data .....</b>	<b>167</b>
<b>6.7 Application of RFE to crop models .....</b>	<b>167</b>
6.7.1 GLAM.....	170
6.7.2 CSWB .....	179
<b>6.8 Conclusion.....</b>	<b>186</b>
 <b>CHAPTER 7 CONCLUSIONS .....</b>	 <b>188</b>
<b>7.1 Summary of main results.....</b>	<b>188</b>
<b>7.2 Issues and recommendations.....</b>	<b>191</b>
 <b>ANNEX A DETERMINATION OF VARIOUS VARIABLES IN THE FAO PENMAN-MONTIETH EQUATION.....</b>	 <b>195</b>

<b>ANNEX B APPROXIMATION OF THE MEAN SQUARE PREDICTION ERROR OF THE PIXEL RAINFALL ESTIMATION WITH FRACTIONAL COVERAGE USING DOUBLE KRIGING .....</b>	<b>197</b>
<b>ANNEX C PARAMETERS FOR GLAM FOR DIVISIONAL GAMBIAN GROUNDNUT.....</b>	<b>200</b>
<b>REFERENCES.....</b>	<b>202</b>

# Chapter 1 Introduction

## 1.1 Background

Many countries in sub-Saharan Africa are characterised by the persistence of food insecurity (Clover, 2003; UN Millennium Project, 2005). The number of chronically undernourished within the region has been increasing for the past decade and currently accounts for 34% of the population (the Commission for Africa report, 2005). Food insecurity is particularly endemic in the Sahelian region, which stands out as the only region in the world where food production per capital has decreased during the past twenty years (Sultan *et al.*, 2005). The reasons for persistence of food insecurity in this region despite the improvement of global food production capacity are complex. Causes include political instabilities, poor economic infrastructure, poor education, health, extreme weather events and environmental degradation (the Commission for Africa report, 2005). These unfavorable environmental, socio-economic and political factors have often conspired together, resulting in the inability to match food supply with demands. In adverse situations, natural hazards, man-made instabilities or a combination of both, can result in serious food crises or large-scale famine, with associated severe loss of life and disruption of livelihood. This in turn results in further vulnerability to food emergencies (Haile, 2005). Variability of rainfall is one of the most important climatic factors that affect the majority of the food insecure population in Africa through its relation with agriculture and livelihood. Therefore, monitoring of rainfall and its impact on agriculture in Africa is an important area of study.

In the remaining part of this section, a review of the current understanding of the reasons behind the observed rainfall variability in different regions of Africa is provided. This is followed by a discussion on the various impact pathways of such variability on livelihood. The importance of satellite based rainfall estimates in monitoring rainfall related hazards in this region, with emphasis on their uses in crop and hydrological modelling applications, are then discussed. This section concludes with a brief

description of the groundnut crop, used as a case study in this thesis, to provide some context for discussion in later chapters.

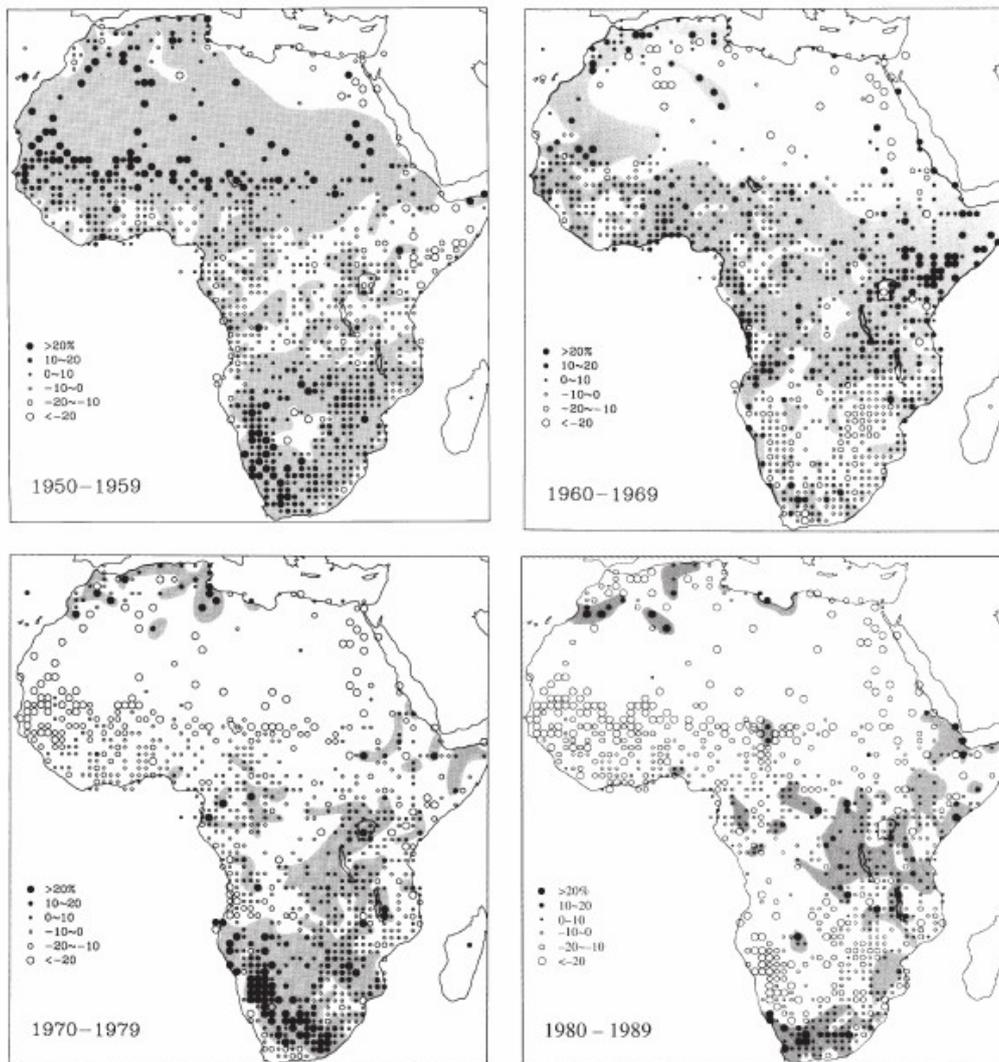


Figure 1.1 Mean rainfall for the decades 1950–59, 1960–69, 1970–79 and 1980–89 (from Nicholson 1993). Rainfall is expressed as a percent departure from the long-term mean, with station data averaged over 1° squares. Positive values are shaded.

### Variability of rainfall in sub-Saharan Africa

The seasonal rainfall pattern for sub-Saharan Africa is determined by the annual meridional migration of the Inter Tropical Convergence Zone (ITCZ), whose character and location depends on the features of larger scale atmospheric circulations (Nicholson,

2001). Over the latter half of the last century, rainfall has seen extreme fluctuations with most parts of the continent experiencing increasing aridity since the 1970s. These changes in rainfall pattern has been drastic, especially in the Sahelian West Africa, where drought condition became prevalent from 1980 onwards, following two decades of anomalously high rainfall (Figure 1.1). Recent observations suggest parts of Sahel could be recovering from the persistent drought condition and increased rainfall along the Guinea coast (Nicholson *et al.*, 2000; Nicholson, 2005). Since the major droughts in the continent in the late 1960s/early 1970s and the 1980s, numerous investigations have focused in identifying the controlling factors in the variability of African rainfall.

Global and regional sea surface temperature (SST) has emerged as an important factor exerting significant control in the atmospheric processes (Sutton *et al.* 2000; Paeth and Hense 2004). Many authors have associated a West African meridional dipole pattern with negative rainfall anomalies in the Sahel, and positive rainfall anomalies along the Guinean coast, with a north-south Atlantic SST dipole and SST anomalies in the Atlantic (Lamb, 1978a,b; Hastenrath, 1984; Lough, 1986; Servain, 1991; Janicot, 1992; Fontaine and Janicot, 1996; Ward, 1998). Various studies also revealed distinct teleconnections between variations in African rainfall and the El Niño-Southern Oscillation (ENSO) phenomena in Equatorial Pacific, particularly Eastern and Southern Africa, with regional difference in its influence; El Niño events are associated with droughts in the main rainy season in Ethiopia (Ininda *et al.*, 1987; Tadesse, 1994; Camberlin, 1995; Seleshi and Demarée, 1995) and southern Africa (Lindesay *et al.*, 1986; Ropelewski and Halpert, 1987; van Heerden *et al.*, 1988; Jury *et al.*, 1994), while ENSO is associated with abnormally heavy rain in the East African ‘short rains’ (Farmer, 1988; Ogallo *et al.*, 1988; Hutchinson, 1992; Hastenrath *et al.*, 1993). The influence of ENSO to the rainfall in the Sahel however is still uncertain (Nicholson, 2001), though some authors suggested the possibility of modulation of rainfall in El Niño years (Rowell *et al.*, 1995; Janicot *et al.*, 1996; Ward, 1998; Camberlin *et al.*, 2001).

Nicholson (2001) had the opinion that it is the SST pattern in the proximity of the continent associated with ENSO, rather than the phenomenon itself, that is the primary

factor in determining the variability of rainfall in these regions. This view had recently been reinforced in several studies. For example, there is emerging evidence of the role of the Indian Ocean Dipole or Zonal Mode (IOZM) (Saji *et al.* 1999; Webster *et al.* 1999), in enhancing East African rainfalls during non El Niño years (Webster *et al.*, 1999) although strong ENSO events may be important in the initialization for some IOZM events (Black *et al.*, 2003).

Apart from large-scale modulation of inter-annual variability of rainfall by SST, other important mechanisms, notably a hydrological feedback, may also play a significant role in the controlling the Sahelian rainfall variability by modifying the thermal and wind structure in the atmosphere through latent heat release from evapotranspiration. Lare and Nicholson (1994) suggested that the persistent drought experienced in the Sahel is a result of such positive feedback between soil moisture and the African Easterly Jet. Evidence of soil moisture-atmospheric feedback in the Sahelian region has been reported; analysis of data obtained during the Hydrologic-Atmospheric Pilot Experiment (HAPEX-Sahel) and the JET2000 experiment suggested positive feedback between soil-moisture and convective rainfall at storm scale (Talyor *et al.*, 1997; Taylor and Lebel, 1998; Taylor *et al.*, 2003). However recent studies suggested that soil-moisture might act to suppress convection at larger scale (Taylor, 2005).

#### Effects of rainfall variability to food availability and livelihood in Sub-Saharan Africa

Agriculture in sub-Saharan Africa is the most important source of both income and food supply for a large portion of the region's population, supporting 70% of the economic activities within the region. As of 2001, 92% of the rural population were agricultural (FAO, 2003), primarily organised into small family holdings practicing subsistence farming, with additional income supplemented through the trading of any surplus of crop during harvest. As agriculture in the region is primarily rainfed, these groups of people are particularly vulnerable to the variability of rainfall in this region (Walker, 1991; Rosenzweig and Binswanger, 1993; Dercon, 2002; Zimmerman and Carter, 2003), in the form of drought and flooding (Dercon, 2004; Haile, 2005) that resulted in crop failure and decimation of livestock. Apart from the immediate problem of hunger, affected

households may be forced to cope with the short-term problem of food provision by consuming seeds, selling livestock, removing children from schools, or abandoning their village and moving to urban areas or refuge camps, thereby degrading their future means of livelihood (Haile, 2005). As crop yields depend, to a large extent, on the distribution of rainfall within the crop growing seasons, inter-annual variability of crop yields can be high (Affholder, 1997; Graef and Haigis, 2001; Ingram *et al.*, 2002; Sultan *et al.*, 2005), causing difficulties in resource allocation decisions for the farmers. This may lead to unwillingness to invest in new technologies, under-utilisation of fertilizers (Sanchez, 2002; Hansen, 2005) and crop price instabilities (Haile, 2005). Apart from direct impact on farmer's produce, variability of rainfall can affect livelihood of the food insecure population through its relations to outbreaks of agricultural pests, like locusts, African armyworms and rodents (Tucker, 1984; Cheke and Holt, 1993; Harvey and Mallya, 1995; Dewhurst *et al.*, 2001, Todd *et al.*, 2002; Davis *et al.*, 2004; Granjon *et al.*, 2005). Rainfall-related disease epidemics afflicting humans like malaria (Loevinsohn, 1994; Kilian *et al.*, 1999; Teklehaimanot *et al.*, 2004; Zhou *et al.*, 2005; Thomson *et al.*, 2005) and Rift Valley Fever (Linthicum *et al.*, 1999) reduce manpower resources for productive activities and in extreme cases cause fatalities, thereby increasing the vulnerability of the afflicted household to other hazards.

#### Satellite rainfall estimates in hazard monitoring and early warning

Many early warning systems (EWS) are currently operating to assist international and national agencies in policy planning and humanitarian relief operations to mitigate the effects of natural and man made hazards. The notable EWS in Africa supported by international organisations like United States Agency for International Development (USAID) and Food and Agricultural Organization (FAO) include the Famine Early Warning System Information Network (FEWSNET), Food Insecurity and Vulnerability Information and Mapping Systems (FIVIMS), Global Information and Early Warning System on Food and Agriculture (GIEWS), and Vulnerability Analysis and Mapping (VAM). In addition, there are number of EWS supported by major sub-regional organisations like the Centre Regional Agrometeorologie-Hydrologie-Meteorologie (AGRHYMET) in West Africa, and the Southern African Development Community

(SADC) and the Inter-Governmental Authority on Development (IGAD) in East Africa that monitor hazards and issue warnings over their sphere of responsibilities. A good review of the organisations and services provided by the EWS for Africa can be found in Vordzorgbe (2003). Because of the inter-relationship between rainfall and the many aspects of livelihood of the food insecure populations noted in the preceding section, real-time rainfall information is inherently important in the overall EWS operation in monitoring of the state of crop production, drought and epidemics. As the rainfall network within Africa is sparse and often reported with significant delays (Washington *et al.*, 2006), satellite based rainfall estimates (RFE) provide a good substitute. Due to their real time availability and good spatio-temporal resolution, satellite RFE provides the possibility of continuous monitoring of rainfall over the continent. At present, real-time daily and dekadal (10-day cummulation) RFE for Africa are routinely processed and made accessible over the internet by various agencies, for example, the Climate Prediction Center (CPC) (<http://www.cpc.ncep.noaa.gov/products/fews/rfe1.html>) and the Tropical Application of Meteorological Satellite (TAMSAT) at the University of Reading (<http://www.met.reading.ac.uk/~tamsat/>).

There is an increasing usage of satellite-based RFE to improve the capability in monitoring weather-environment interactions critical to people whose livelihood are potentially venerable to the vagaries of weather. The benefits of satellite RFE in early detection of malaria outbreaks has been highlighted by case studies conducted in Kenya (Hay *et al.*, 2003), and a real-time internet based RFE malaria monitoring system for Africa is in operation (Grover-Kopec *et al.*, 2005).

The potential for using satellite RFE with crop and hydrological models in monitoring crop production prediction, water resources, and flood prediction within Africa has received much attention by researchers: Verdin and Klaver (2002), and Senay and Verdin (2003) demonstrate the potential of using RFE for the geospatial mapping of the Water Requirement Satisfaction Index (WRSI) (Frère and Popov, 1986) for crop and drought monitoring. Such remotely sensed WRSI products have since been incorporated into FEWSNET EWS operational cycle. More direct applications of RFE to crop yield

prediction are also being investigated. Reynolds *et al.* (2000), and Verdin and Klaver (2002) proposed a GIS-based crop yield estimation system that incorporated dekadal RFE for crop yield forecasting, based on empirical relations between crop yield and the accumulated WRSI at the end of the crop growing seasons, to monitor maize production for Kenya and Zimbabwe. Thornton *et al.* (1997) demonstrated the feasibility of estimating millet production in Burkina Faso by using more sophisticated mechanistic crop models driven by satellite RFE. Various workers had used RFE for river flow predictions over the Senegal River Basin and Nile River regions. The conclusions from these investigations are uncertain. Grimes and Diop (2003), working on the prediction of daily river flow for the Bakoye catchment in West Africa showed that satellite RFE combined with numerical weather model outputs performed better than gauge or satellite RFE only inputs. However, Anderson *et al.* (2002), and Sandholt *et al.* (2003) found that using satellite RFE did not give significantly improved flow prediction as the former authors suggested for the Senegal Basin. Likewise, preliminary investigation of incorporating RFE in the Nile Forecast System (NFS) by Tsintikidis *et al.* (1999) for the Blue Nile region did not show a clear improvement in hydrological response from the NFS.

#### Uncertainties in RFE and its effects on crop and hydrological model response

Satellite rainfall retrieval relies on inference of surface rainfall from the irradiances measured by the satellite sensors operating at the visible, thermal infrared, or microwave region of the electromagnetic spectrum. There are inherent uncertainties in these satellite-based rainfall estimates. Visible and thermal infrared radiometric measurements are broadly speaking, a function of the cloud albedo and cloud top temperatures respectively, and therefore, possess only an indirect relationship with rainfall. Radiometric measurements at microwave channels, on the other hand, depend on the liquid-water content of clouds. Therefore, it is closer in relationship to rainfall than visible and thermal infrared radiometry. However, several issues (Barret and Martin, 1981) imply that the uncertainty is associated with microwave-based rainfall estimates including high and variable microwave emission from land surface masking rainfall signature; low resolution of rain-rate at high brightness temperature (proportional to the detected irradiance); low

spatial resolution, resulting in bias in areal rainfall estimates due to spatial heterogeneity of rainfall.

In the case of tropical Africa where rainfall is spatially intermittent at timescales in the order of several hours to days, the above-mentioned deficiencies translate to the uncertainties in delineation of rainy and non-rainy areas, as well as in the amount of rainfall in the rainy areas measured by satellite. Laurent *et al.* (1998) conducted an inter-comparison of several RFE algorithms over a densely gauged location in Burkina-Faso, and found that the mean error of dekadal rainfall from RFEs over a  $0.5^{\circ} \times 0.5^{\circ}$  spatial resolution is about 35% of its mean areal estimates from rain-gauge observations. A larger error of 60%-84% is reported by Todd *et al.* (1999) for daily rainfall estimates at pixel level (~ 5 km by 5 km) over the Blue Nile Basin. In addition, error in RFE exhibits heterocedasticity – where the error increases with mean rainfall prediction (Grimes *et al.*, 1999).

As sensitivity and non-linearity to rainfall inputs has been reported in crop models (Nonhebel, 1994a,b; Hansen and Jones, 2000; Heinemann *et al.*, 2002; Sultan *et al.*, 2005) and hydrological models (Tsintikidis *et al.*, 1999; Hossian *et al.*, 2004a,b; Kobold and Sugelj, 2005), the question that logically arises in applying RFE to crop and hydrological models is the effect of the uncertainties that the rainfall inputs have on these model responses. Indeed, several studies in different parts of the world have highlighted the differences in hydrological model responses when satellite RFE is used instead of gauge observations. Differences in simulated flow forecasts and soil moisture content are reported for the Des Moines River basin in the United States (Guetter *et al.*, 1996; Tsonis *et al.*, 1996) and the Blue Nile Basin (Tsintikidis *et al.*, 1999). The results from these studies suggested that biases in RFE input, as well as the difference in model calibrations using different type of rainfall data (gauge or RFE), resulted in different antecedent soil moisture in gauge and RFE forced hydrological model. Although much attention has been focused on the use of RFE driven crop model in crop yield estimations for Africa, there is limited literature, to the author's knowledge, pertaining to the effects of uncertainties of satellite RFE on crop model responses. In view of the increasing interest

in using crop models for crop yield estimation in conjunction with satellite RFE in Africa, it is important that attention is paid in assessing the impact of uncertainties of these rainfall estimates have on model responses.

### Groundnut

As groundnut production in the Gambia is used as a case study in this thesis, the remainder of this section briefly describes the groundnut crop – its importance, development, and cultivation practices. This provides useful background for discussions in later chapters.

Groundnut (*Arachis hypogaea* L., Figure 1.2) is an annual legume cultivated in many tropical regions within 40°N to 40°S. It is one of the world's most important oilseed crops (Dwivedi *et al.*, 2003) and an important source of edible oil and protein (Weiss, 1983). The byproduct of groundnut oil extraction is a source for food for human and animal (Savage and Keenan, 1994). Groundnut haulms can also be used as animal feeds (Cook and Crosthwaite, 1994). The average global groundnut production for 2000 to 2005 was 35.5 million metric tons, with developing countries in Asia (66%) and Africa (25%) as the major producers (FAOSTAT data, 2006). The major groundnut producing countries in Africa are Nigeria, Sudan, Senegal, Chad and Ghana, and these contribute 60% of continent's groundnut production by weight.

The share of the global production of groundnut from Africa decreased sharply during the 1970s from about half the global production in the 1960s to the current figure (FAOSTAT data, 2006) of 25%, despite contributing 40% of the global groundnut cultivation area. The reasons attributed to the decline (Chiteka *et al.*, 1992; Schilling, 1996) include environmental factors such as pests, diseases, droughts; exogenous factors, such as instability in the international market price; domestic political factors such as the collapse of marketing organisations, and institutional failures to provide assistance to farmers.

Types (Schilling, 1996)

Cultivated groundnut can be classified into two subspecies: *hypogaea hypogaea*, which corresponds to the Virginia types, and *hypogaea fastigiata*, which includes the Spanish and Valerica types. Notable differences between the two subspecies include:

1) *Plant habit*: Spanish and Valerica types are erect bunch in habit, whereas the Virginias can be either erect or trailing (i.e. runners).

2) *Branching pattern*: Spanish and Valerica types are sequentially branched while Virginias are alternately branched. This resulted in the latter appearing denser in appearance than the former.

3) *Crop duration*: Spanish and Valerica types have a growing period of about 90-110 days whereas Virginias have a longer growing period of 120-140 days. The actual duration of crop development depends on the temperature and rainfall received.



Figure 1.2 *Arachis hypogaea* L.  
©1995-2005 Missouri Botanical Garden  
<http://www.illustratedgarden.org>

Activities	Apr				May				Jun				Jul				Aug				Sep				Oct				Nov				Dec			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Brushing and clearing of farms																																				
Receiving groundnuts from village seed stores																																				
Shelling and picking of quality nuts																																				
Ploughing and sowing of groundnuts																																				
Fertiliser application of groundnuts																																				
Weeding of groundnuts																																				
Guarding of groundnuts																																				
Harvesting and stacking of groundnuts																																				
Trading and marketing of groundnuts																																				
Storage of groundnuts into village seed stores																																				

Figure 1.3 Time schedules of crop activities in The Gambia. (Source: Department of Planning, The Gambia)

Developmental stages (Doorenbos and Kassam, 1979; Schilling, 1996)

Seedlings emerge 3-5 days after germination. Flowering begins 25-30 days after sowing and peaks about 40-60 days after emergence under normal conditions. Vegetative growth is limited before onset of flowering, but follows a period of rapid growth during flowering and pod formation. 8-14 days after pollination, pegs developed under the base of the flower. These aerial pegs then become positively geotropic, and bury themselves 5-7 centimeters into the soil. The tip of these pegs swells and forms pods that subsequently take another 50 – 70 days to mature.

Environmental Requirements (Doorenbos and Kassam, 1979; Schilling, 1996)

Groundnut is best grown in sandy or fine-textured friable soils that are neutral or slightly acidic with good infiltration. The optimum temperature for cultivation is about 25 to 33°C. Germination can be inhibited if temperature is below 15°C or when higher than 45°C. In rainfed cultivation, the crop requires about 500-700mm of reliable rainfall through the growing season to realise a good yield. Water deficit can impede shoot and root growth and consequently, reduced yield (Wright and Nageswara Rao, 1994). Sensitiveness to water deficits depends on the stage of crop development; it is rather drought-resistant before onset of flowering, but enters into a period of high sensitivity during flowering and pod-formation. Relatively dry conditions are favorable from pod formation to maturity. Excess rainfall at this late period can have an adverse impact on yields due to regermination for the non-dormant (Spanish/Valerica) varieties, while too little rainfall may cause cracks in mature pods increasing the risk of aflatoxin contamination.

Cultivation practices and crop activities (Schilling, 1996)

Both mixed cropping and crop rotation are common practices in groundnut cultivation in Africa. The common mixed cropping practices involve groundnut mixed with annual cereals, and for wetter regions, with cassava or other tubers. Crop rotation is more common amongst farmers when groundnut is grown as a cash crop. Figure 1.3 shows a typical schedule of required activity undertaken by groundnut farmers for the Gambia. Before the onset of the rainy season, the selected sites for cultivation are cleared by

burning of crop residues and plant debris that may harbor pest and diseases. Seeds that are collected from the seed stores require shelling followed by sorting to discard any damaged or infected seeds before being chemically “dressed” to protect against pest and diseases. Sowing is done at the onset of rain, often on a day-by-day basis, to reduce the risk of crop failures due to early drought. As young groundnut is highly sensitive to competition from weeds, weeding is performed to ensure the crop is weed-free soon after emergence. At harvest, the crops are lifted, dried and threshed to separate the pods from the haulms. The products are then delivered to the primary collection points where it is stacked and weighed. Most smallholders rely on manual labor for these activities. The use of draught-power for ploughing, seeding and lifting is more common when groundnut is cultivated as cash crop in countries such as Senegal and Nigeria.

## 1.2 Motivation and scope of thesis

### Feasibility of daily satellite RFE for crop yield forecasting in Africa using process-based crop model

Recently there has been a growing interest in using process-based crop models in crop yield prediction. These crop models simulate explicitly the crop development and interaction with environment. Many of these investigations focus on the feasibility of using predictions from numerical models of the atmosphere (GCM) to drive crop model for crop yield prediction (Stone *et al.*, 2000; Challinor *et al.*, 2005a, b; Hansen *et al.*, 2004; Hansen 2005), before or early in the crop growing season. A plausible way to improve these pre-season crop yield prediction is to routinely update the crop yield forecast by using satellite RFE for crop model inputs as the growing season progresses. So far, there is little literature in using satellite RFE for input into a mechanistic crop model for large-scale crop yield predictions in Africa. The only investigation the author is aware of was conducted by Thornton *et al.* (1997), who proposed obtaining daily rainfall time-series as input in the CERES-Millet from dekadal satellite RFE from rainfall generator. The two main questions that this thesis tries to address are:

1) Are daily satellite-based RFEs useful as inputs to such crop models in crop yield prediction?

2) What effects does the uncertainty in daily satellite-based RFE have on crop model responses?

For the present investigation, daily RFEs have been derived from daily cold cloud duration (CCD), obtained from the infrared brightness temperature from METEOSAT data, based on the TAMSAT method originally proposed by Milford and Dugdale (1990). Currently more sophisticated multi-platform satellite RFE exists in literature and operations (Grimes *et al.*, 1999; Xu *et al.*, 1999; Adler *et al.*, 2000; Bellerby *et al.*, 2000; Sorooshian *et al.*, 2000; Todd *et al.*, 2001; Grimes and Diop, 2003; Grimes *et al.*, 2003; Nicholson *et al.*, 2003a, 2003b; Huffman *et al.*, 2003; Kidd *et al.*, 2003; Bellerby, 2004; Marzano *et al.*, 2004; Joyce *et al.*, 2004). Although these algorithms may possibly have better performance, benefiting from additional information including additional satellite data, rain gauge and meteorological fields from numerical weather simulations, the single platform CCD-rainfall estimation model is used in this thesis for the investigating RFE-crop model responses due to its relative simplicity, and the fact that CCD data is readily available. Results from such studies can serve as a basis for future assessment of more sophisticated algorithms in the context of applying RFE to crop models.

With regard to the issue of uncertainty in satellite RFE, investigation within this thesis focuses explicitly on the uncertainty arising from the many possible spatial configurations of daily rainfall field that can be associated with the daily METEOSAT CCD-derived RFE, since sampling error is not a factor in the algorithm employed. From the results of hydrological experiments discussed in the previous section, error from satellite-based RFE is suspected to be important for a crop model operating at daily time-step, where inaccuracies in antecedent rainfall may result in error propagation in the crop model state. Most studies of RFE uncertainties had focused on error in passive microwave rainfall estimates due to limited satellite overpass (Bell, 1987; Bell *et al.*, 1990; Astin, 1997; Steiner *et al.*, 2003; Gebremichael and Krajewski, 2004). Only limited

literature with regard to the error in identifying the actual spatial rainfall field from satellite data is available. The few that the author is aware of includes those reported in Fiorucci *et al.* (2001), Bellerby and Sun (2005), Hossian and Anagnostou (2005), who employed Monte Carlo simulations through suitable stochastic rainfall models to synthesise possible rainfall scenarios consistent with the satellite data. Similar methodology is used to obtain uncertainties in the proposed daily RFE algorithm in this thesis.

The crop model used in the present work is the General Large Area Model for annual crop (GLAM), which has shown some skill in the hindcasting of regional groundnut yields of India, where rainfed agriculture is practiced (Challinor *et al.*, 2004). This model is specifically designed so that a relatively small number of model parameters are required, and is applicable at spatial scales where crop yield have a correlation with rainfall (Challinor *et al.*, 2004). It is therefore appealing for use in a crop yield forecasting system for Africa where high-resolution soil, crop, and management data are unavailable. An empirical crop-yield model adopted from the FAO-CSWB model (Frère and Popov, 1979, 1986; Gommès, 1993), widely used currently in crop yield estimation in Africa, is used as a reference for performance comparison with GLAM.

### 1.3 Plan of the thesis

The thesis starts with the description and preparation of the weather and groundnut production data for the Gambia (Chapter 2), used in the case studies. Some preliminary statistical exploration of the rainfall and regional groundnut yields is performed providing a background understanding for later discussions.

Chapter 3 is devoted to the description of the method used to interpolate the daily gauge rainfall for the Gambia. A kriging method that enables the delineation of intermittent rainfall is used. The purpose for this chapter is three fold: 1) To provide areal rainfall estimates for METEOSAT pixels from gauge rainfall observations. These are used in calibrating the empirical CCD-rainfall relation. 2) For validation purposes in the

following chapters. 3) To obtain daily gauge estimated areal rainfall time-series for input into the crop models as a basis of comparison of crop model responses against satellite RFE.

Chapters 4 and 5 focus respectively on the satellite RFE algorithm and the stochastic model employed to synthesise spatial rainfall fields with METEOSAT data. Chapter 4 describes in detail, the extension of the TAMSAT CCD-rainfall algorithm for the purpose of estimating daily rainfall. The main feature in the chapter is the development of a full statistical model of CCD-rainfall at METEOSAT pixel level that describes the error in identifying rainy and non-rainy pixels, and a probability distribution function of the rainfall amount for rainy pixels. This completes the statistical rainfall model at pixel resolution and serves as the kernel for spatial reconstruction of rainfall field described in the next chapter. Chapter 5 approaches the issue of quantifying the uncertainty in representing the actual rainfall field by the CCD-rainfall estimates through a geostatistical simulation method to reconstruct spatial rainfall field. The main point of the chapter is to propose an alternative to the method of generating spatially intermittent rainfall as those suggested by Fiorucci *et al.* (2001) and Hossian and Anagnostou (2005).

Using the divisional groundnut yield for the Gambia as case study, Chapter 6 attempts to answer the two questions posed earlier that motivate the present investigation namely, 1) the feasibility of using satellite RFE for crop yield prediction, 2) crop model yield response to uncertainties in RFE. The main features of the GLAM and the CSWB model are presented first, followed by a discussion on the calibration of the crop model for the case of Gambian groundnut. Differences in responses of crop model driven by RFE and gauge rainfall are discussed. The sensitivity of simulated crop yield to random and systematic error in RFE is also investigated.

Chapter 7 concludes the thesis with a summary of the main results and the outstanding issues that emerged from the various studies undertaken in previous chapters. Recommendation on future research directions with regards to the application of satellite RFE in crop yield forecasting in Africa using process based crop models are given.

# Chapter 2 Description and preliminary analysis of data

## 2.1 Introduction

This chapter describes the data set used in this thesis; including, the weather and groundnut yield data of the Gambia and the satellite data used in rainfall estimation. Some preliminary statistical analysis of this data is also provided. Section 2.2 describes the preparation of various weather and satellite data to be used in subsequent chapters. Section 2.3 presents some results from the preliminary statistical analysis of rainfall. Section 2.4 describes the groundnut yield of the Gambia and how the technological trends of two most important groundnut-growing divisions are modelled. Section 2.5 presents some conclusions.

## 2.2 Agrometeorological and satellite data

The primary weather data of interest for the work in this thesis include daily maximum and minimum temperatures ( $T_{max}$ ,  $T_{min}$ ), maximum and minimum relative humidity ( $RH_{max}$ ,  $RH_{min}$ ), sunshine hours ( $SS$ ), surface wind speed ( $u$ ) and rainfall ( $Z$ ). Data from the observation network of the Gambia (Figure 2.1) going back to 1971, has been obtained from the Department of Water Resources of the Gambia. There are currently 11 operational meteorological stations and additionally 29 rainfall stations around the country (as of 2002). However, not all station records are available. For the above-mentioned weather parameters, except rainfall, the number of daily reporting meteorological stations for the crop growing seasons (June to November) was about 4 on average during the 1970s, increasing steadily to current number by the late 1980s. The number of stations reporting rainfall before the 1990s was about 19 on average, increasing to about 24 after 1995.

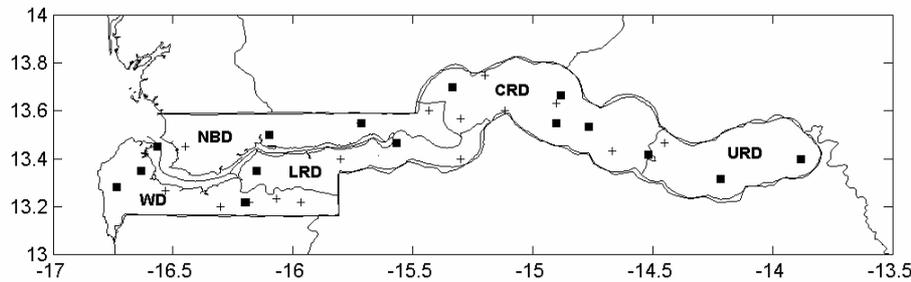


Figure 2.1 Gambian meteorological observations network where data are made available and the division boundaries. Filled-squares represent the locations of the meteorological station and crosses are the locations of rain gauges. Abbreviations of division names are as follow: NBD: North Bank Division; WD: Western Division; LRD: Lower River Division; CRD: Central River Division; URD: Upper River Division.

### Spatial interpolation of weather data

For the purpose of crop modelling and the calibration of the satellite rainfall model, it is necessary to interpolate the various observed weather data to a larger spatial scale. Daily station-observed  $T_{min}$ ,  $T_{max}$ ,  $RH_{max}$ ,  $RH_{min}$ ,  $SS$  and  $u$  are spatially interpolated by inverse square distance weighting to a grid of  $0.05^0 \times 0.05^0$ . Areal averages were obtained by aggregating over the grid within the domain of interest. The inverse square weighting interpolation was deemed sufficient to capture the spatial distribution of these weather parameters as the variability of these parameters across the region are reasonably small compared to the average daily observed values (average daily CVs for these parameters < 50%). As the daily Gambian rainfall exhibited higher spatial variability (average daily CV  $\approx 250\%$ ), a more statistically rigorous interpolation method, following Barancourt *et al.* (1992), was employed to produce daily rainfall fields from rain gauge observations. The details of the spatial interpolating daily Gambian rain gauge data are presented in Chapter 3.

### Potential Evapotranspiration

Gridded daily spatial potential evapotranspiration ( $PET_0$ ) for the reference crop in units of  $\text{mmday}^{-1}$  is derived from meteorological data using the FAO Penman-Monteith equation (Allen *et al.*, 1998):

$$PET_{0j} = \frac{0.408\Delta_j(R_{nj} - G_j) + \gamma \frac{900}{T + 273} u_{2j} VPD_j}{\Delta_j + \gamma(1 + 0.34u_{2j})} \quad 2.1$$

where  $R_n$  is the net radiation at crop surface [ $\text{MJ m}^{-2}\text{day}^{-1}$ ],  $G$  is the soil heat flux [ $\text{MJ m}^{-2}\text{day}^{-1}$ ],  $T$  is the mean temperature at 2m height [ $^{\circ}\text{C}$ ],  $u_2$  is the average wind speed at a 2m height [ $\text{ms}^{-1}$ ],  $VPD$  is the saturated vapour deficit [ $\text{kPa}$ ],  $\Delta$  is the slope of the vapour pressure curve [ $\text{kPa } ^{\circ}\text{C}^{-1}$ ] and  $\gamma$  is the psychrometric constant [ $\text{kPa } ^{\circ}\text{C}^{-1}$ ]. The index  $j$  is the grid index. In the computation, the daily net soil flux is approximately zero,  $T$  is taken as the average of the daily  $T_{max}$  and  $T_{min}$ . Determination of  $R_n$ ,  $VPD$  and  $\Delta$  is given in Annex A.

### Satellite Data

Daily METEOSAT Cold Cloud Duration (CCD) for the Gambian region (13.6W-16.9W, 13N-14N), for June to November from 1988-2002 at  $0.05^{\circ} \times 0.05^{\circ}$  resolution, were obtained from TAMSAT<sup>1</sup>. CCD at threshold temperatures from  $-30^{\circ}\text{C}$  to  $-70^{\circ}\text{C}$  at  $10^{\circ}\text{C}$  intervals are available. Apart from 1988, 1989, 1992, 1993 and 1995, where there were extended periods of days with missing data, the obtained dataset is largely complete, with only intermittent days when CCD data are unavailable.

Table 2.1 summarises the main usage of the agrometeorological data in this thesis.

Data	Purpose	Chapter	Aggregation Scale	
			Spatial	Temporal
$Z$	For determination of the empirical satellite rainfall model.	4	METEOSAT pixel	Daily
	Input for GLAM and CSWB for divisional groundnut yield hindcasting and determination of crop sowing date (together with $PET_0$ )	6	Division	Daily, Dekadal
$T_{max}, T_{min}, RH_{max}, RH_{min}, SS$	Input for GLAM for crop modelling.	6	Division	Daily
$PET_0$	As input into CSWB crop model for groundnut yield hindcasting and determination of sowing date (together with $Z$ )	6	Division	Dekadal

Table 2.1 Reference of agrometeorological data usage. See main text for the meaning of abbreviation of data. Note that a dekad is 10 days.

<sup>1</sup> Tropical Application of Meteorology using Satellite data, University of Reading.

### 2.3 Preliminary analysis of Gambian rainfall

The basic statistics used in the following discussion are defined and estimated as follows:  $z(x,t)$  represents the daily gauge rainfall observed at location  $x$  on date  $t$ . The mean daily rainfall for a given day,  $d$ , denoted as  $m_z(d)$  and is estimated by

$$m_z(d)^* = \frac{\sum_{i=1}^{N_d} z(x_i, d)}{N_d} = \langle z(x, t) | t = d \rangle \quad 2.2$$

where  $N_d$  is the total number of available gauges within the region of interest. The shorthand notation  $\langle \cdot | Q \rangle$  in Equation 2.2 denotes the arithmetic averages for all available gauge rainfall observations satisfying the condition  $Q$ . The asterisk on the LHS of Equation 2.2 denotes an estimation of the value. For large sample size, the standard deviation of the daily rainfall,  $\sigma_z(d)$  is then estimated as:

$$\sigma_z(d)^* = \sqrt{\langle (z(x, t) - m_z(d)^*)^2 | t = d \rangle} \quad 2.3$$

Similarly, for day  $d$ , the mean daily rainfall over rainy areas  $m_{z>0}(d)$ , and the fraction of area covered by rainfall  $m_f(d)$  is estimated by

$$m_{z>0}(d)^* = \langle z(x, t) | t = d, z(x, t) > 0 \rangle \quad 2.4$$

and

$$m_f(d)^* = \langle I(z(x, t)) | t = d \rangle \quad 2.5$$

respectively. In Equation 2.5,  $I(z) = 1$  if  $z > 0$ , and 0 if  $z = 0$ . Note that  $m_f(d)$  can alternatively be interpreted as the probability of rainfall for the day  $d$ .

Likewise, the estimated mean daily rainfall over rainy area for a given year  $Y$  is:

$$m_{z>0}(Y)^* = \langle z(x, t) | year(t) = Y, z(x, t) > 0 \rangle \quad 2.6$$

The averaging operation in Equation 2.6 is performed over the collection of gauge rainfall amount for year  $Y$  that is positive. The mean fractional coverage by rainfall for year  $Y$ ,  $m_f(Y)$ , is estimated in similar fashion as in Equation 2.6 using the transformation  $I$  of Equation 2.5.

Mean rainfall at a given location  $a$  is denoted by  $m_Z(a)$  and is estimated by

$$m_Z(a)^* = \langle z(x,t) \mid x = a \rangle \quad 2.7$$

Similar estimates can be given to the mean rainfall when raining,  $m_{Z>0}(a)$ , and probability of rain,  $m_I(a)$ .

### Seasonal rainfall characteristics

The rainy season for the Gambia usually starts around the beginning of June and ends in late October (Figure 2.2a). The rainfall follows a unimodal distribution, with the main rainy seasons around late July/early August to late September. Rainfall peaks around mid-August to early-September. August and September account for 60% of the total seasonal rainfall on average. Total seasonal rainfall for a given year  $Y$ ,  $Z_T(Y)$ , was obtained by adding all  $m_Z(d)$  from June to October of year  $Y$ . The total seasonal rainfall amount is 758 mm on average with high inter-annual variability ( $CV = 21\%$ ). Figure 2.2b shows the time series of the total seasonal rainfall and the five-year moving average for the years 1971-2002. Two distinct periods of persistent lower than the 32-year average can be identified; a consecutive 6 years from 1979-1985 (“A” in figure 2.2b) and another less pronounced 4-year period from 1990-1993. The five-year moving average showed that rainfall has been steadily on the increase from 1985 to present (2002).

Rainfall in the Gambia during the rainy seasons is punctuated by days of no rain and the rainfall fields are often “patchy” on rainy days. For August and September when rainfall is climatologically the most abundant, there are about 5 days of no rain (all gauge observations reported no rain) on average each year, with  $CV = 70\%$ . The mean fractional coverage of rainfall on rainy days (at least one gauge is reporting rain) for the same period is 0.50, with  $CV = 16\%$ . Therefore, total seasonal rainfall may depend on two factors: 1) Rainfall received during rainy event. 2) Numbers of dry days and rain area coverage. For example, a year of negative (positive) rainfall anomaly may be due to a larger (smaller) number of dry days and a smaller (larger) rain area, or due to a lower (higher) than normal rainfall during rainy event, with number of dry days and the rain area coverage being normal.

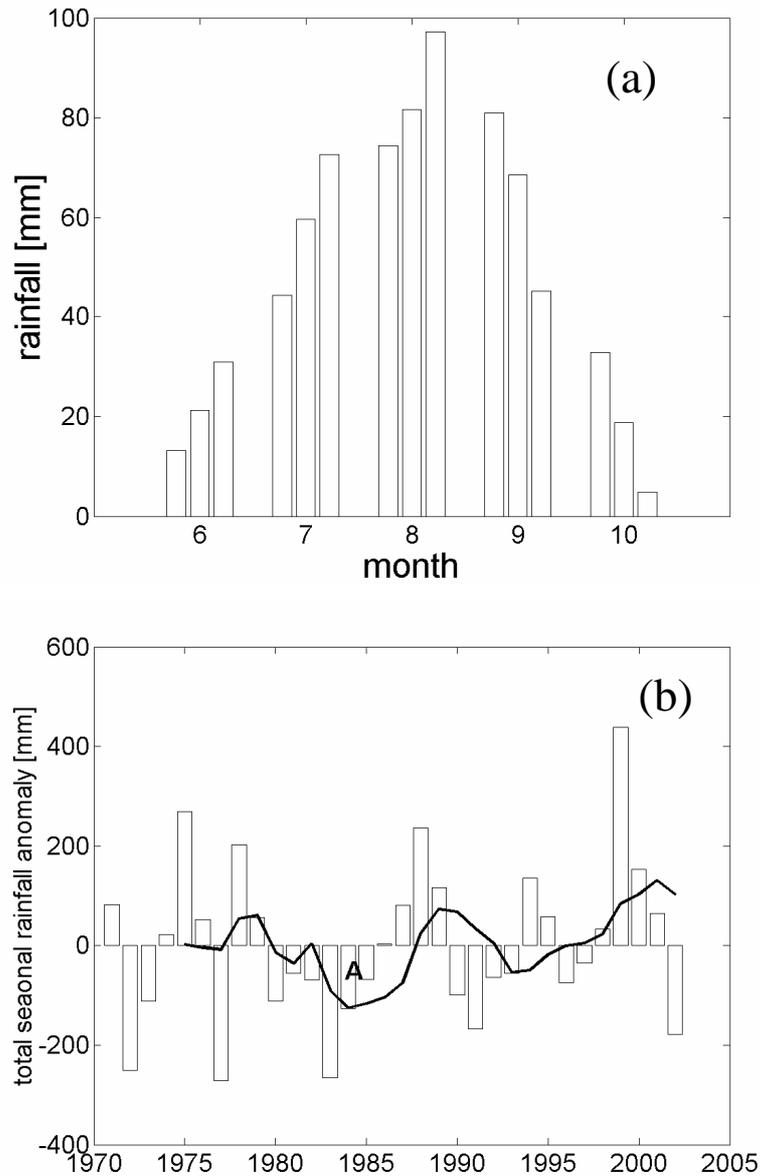


Figure 2.2 Seasonal rainfall characteristics for the Gambia. **(a)** Seasonal rainfall distribution represented by the mean dekadal (10-day) rainfall. Note that the third dekad of each month contains the total rainfall for either 10 or 11 days, depending on the number of days in that month. **(b)** Time series of total seasonal rainfall anomaly,  $Z_T - \langle Z_T \rangle$ , with respect to the 1971-2002 mean. Solid line is the 5-year moving average of the rainfall anomaly. “A” highlights a continuous period of negative rainfall anomalies.

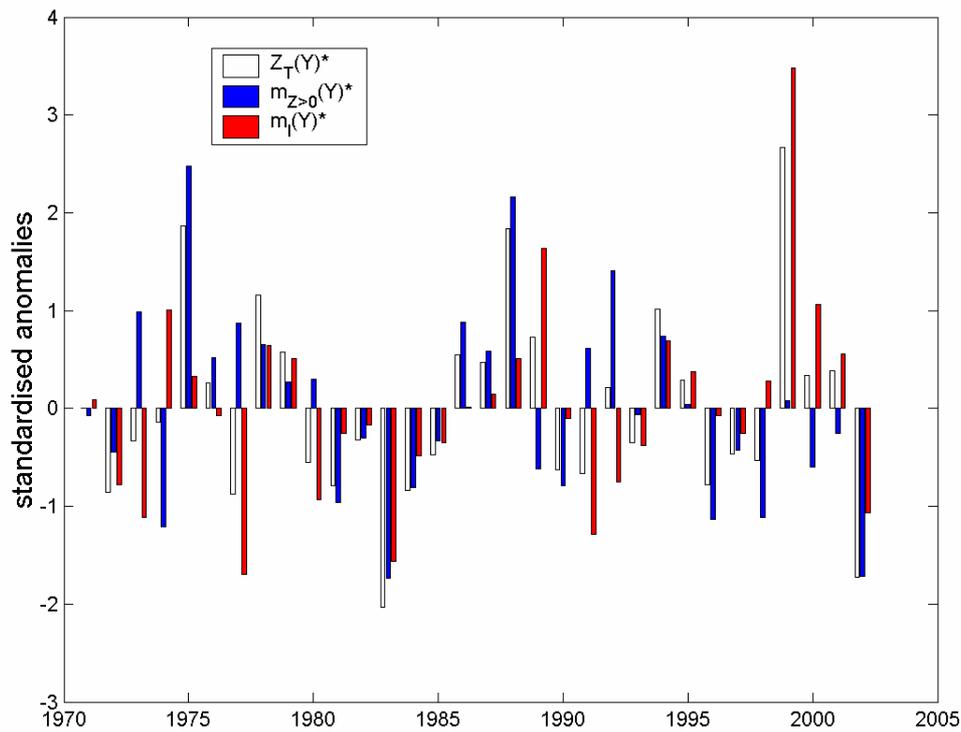


Figure 2.3 Standardised anomaly of total seasonal rainfall,  $Z_T(Y)^*$ , mean daily rainfall in rainy area,  $m_{Z>0}(Y)^*$ , and probability of rain at a given day  $m_l(Y)^*$ , using three gauges.

To investigate these possibilities, 3 gauges with the most complete observations, and distributed evenly over the country, were selected for analysis to avoid preferential sampling. For each year  $Y$ , the mean daily positive rainfall  $m_{Z>0}(Y)$ , and probability of a rainy day  $m_l(Y)$ , were obtained using the 3 reference gauges. Figure 2.3 shows the time series of the two statistics compared with the total seasonal rainfall after scaling. It can be observed that the persistent negative seasonal rainfall anomalies “A” of Figure 2.2 are due to a combination of a lower than normal rainfall amount within the rainy area, and lower than normal probability of rainfall, except in 1979. Both the mean daily rainfall amount and the probability of rainfall have considerable interannual variability (CV is 12% and 15% respectively). Correlation ( $r^2$ ) between  $m_l(Y)$  and  $Z_T(Y)$  is 0.61, while correlation between  $m_{Z>0}(Y)$  and  $Z_T(Y)$  is 0.42, suggesting a stronger link between seasonal rainfall and the probability of rain than the rainfall amount when raining.

However, the rainfall time-series available is not long enough to make this observation statistically significant ( $p = 0.15$ ).

The above discussions highlight the importance of intermittency in Gambian rainfall. Therefore, the intermittent character of the daily Gambian rainfall field is treated explicitly when spatially interpolating the raingauge observations (Chapter 3), in the satellite RFE algorithm (Chapter 4), and the stochastic rainfall simulation (Chapter 5). A further point from the above discussions is the importance in capturing both the mean daily rainfall amount and the probability of rain in satellite RFE, when accurate determination of the interannual variability of seasonal rainfall is critical. In Chapter 6, the impact on crop yield of bias in estimating seasonal rainfall using satellite RFE is demonstrated.

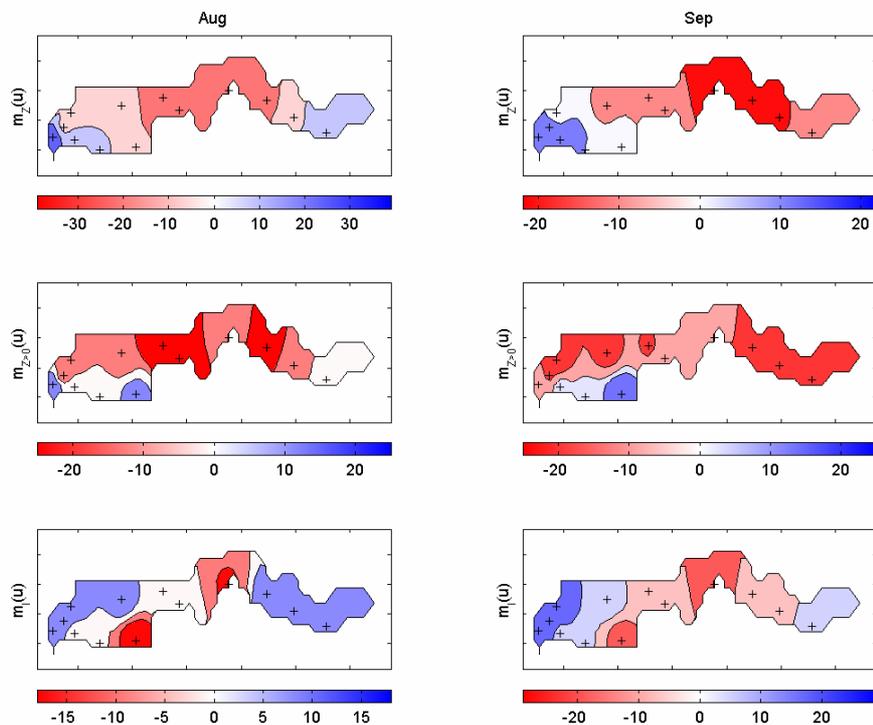


Figure 2.4 Spatial variability of relative difference from global mean expressed in percentages for mean daily rainfall  $m_z(x)$ , mean rain amount within rainy areas,  $m_{z>0}(x)$ , and frequency of rain,  $m_f(x)$ . Crosses mark the gauge locations where the statistics are obtained. The data used is August and September of 1995-2002. The relative difference of a variable  $X$  is defined as  $[X - \langle X \rangle] / \langle X \rangle$ .

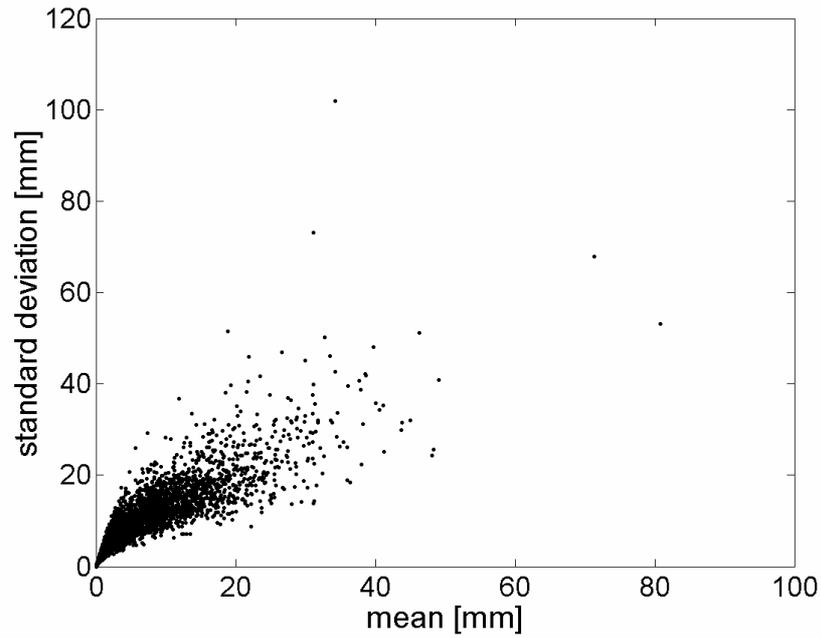


Figure 2.5 Scatter plot of  $\sigma_Z(d)$  against  $m_Z(d)$  using all available days of June-October from 1971-2002.

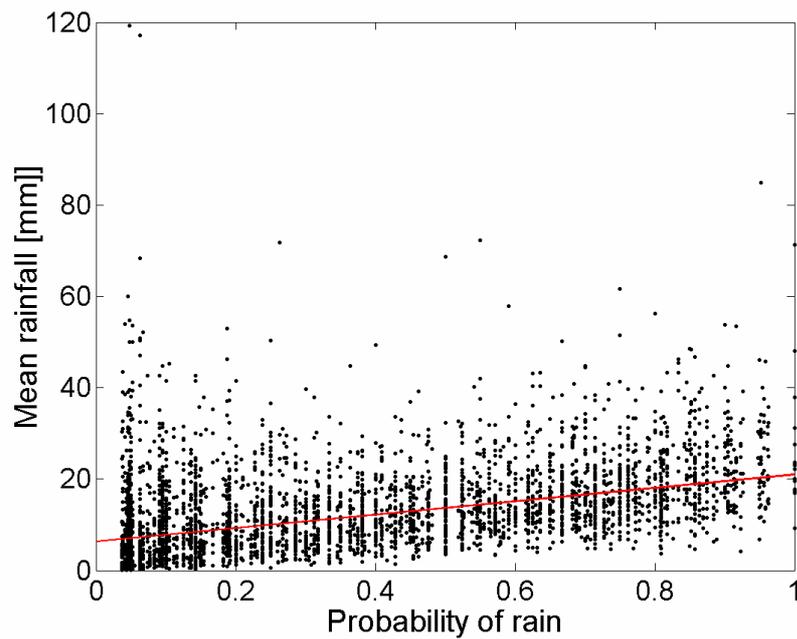


Figure 2.6 Scatter plot of  $m_I(d)$  against  $m_{Z>0}(d)$  using all available days of June-October from 1971-2002. The regression line is shown in red.

### Daily rainfall characteristics

To examine the spatial variability of the daily rainfall field,  $m_Z(x)$ , is obtained from the subset of observations from the gauge network from 1995 to 2002, corresponding to the period with highest amount of gauge data available. The contour plots of the mean daily rainfall for August and September, together with the corresponding daily mean rainfall amount within rainy area,  $m_{Z>0}(x)$ , and the probability of rainfall,  $m_I(x)$  are shown in Figure 2.4. As the inter-annual variability of the rainfall is high, only gauges with a 95% availability were included in the computation. For these two months, it appears that a longitudinal trend in  $m_Z(x)$  and  $m_I(x)$  exists, with Western parts of the country having more rainfall, while  $m_{Z>0}(x)$  has a more latitudinal dependence.

It is unclear whether there is a physical reason behind these spatial patterns, or they are artefacts due to the small number of rain gauges. One reason for the increasing frequency of rainfall towards the Western region could be due to coastal effects, where differential heating along the coast may encourage convective activities. Due to the elongated geometry of the Gambia, inclusion of more observations to the north and south of the boundaries of the country would be required to ascertain any larger scale spatial coherence of the mean positive rainfall field, and to potentially provide a possible physical explanation in the apparent latitudinal variation, particularly over Western Gambia. Unfortunately this data is not available. Apart from spatial heterogeneity, two notable characteristics observed from the data relevant to the discussion in later chapters are as follows:

- 1) Scatter plot of  $\sigma_Z(d)$  against  $m_Z(d)$  (Figure 2.5) shows that variability of daily rainfall increases with the mean rainfall amount, implying that the daily rainfall is heteroscedastic. Such heteroscedasticity of rainfall is taken into account in the statistical CCD-rainfall model in Chapter 4. Heteroscedasticity of this form means that the daily rainfall must be non-ergodic as well (Barancourt *et al.*, 1992). Spatial interpolation of non-ergodic rainfall fields using kriging is discussed in the next chapter.

2) Scatter plot of  $m_I(d)$  against  $m_{Z>0}(d)$  (Figure 2.6) shows that fractional coverage of rainfall increases with rainfall amount. The correlation of  $m_I(d)$  and  $m_{Z>0}(d)$  is 0.38 ( $p < 0.01$ ), suggesting that the probability of rain at any given location increases with the mean rainfall. This feature is used in Chapter 5 to validate the statistical properties of the stochastic rainfall field generated, conditional to daily CCD.

## 2.4 Gambian groundnut production

Groundnut is an important cash crop for the Gambia. The export of groundnut accounts for over 40% of the foreign exchange earnings for the country (Bojang, 2001). Cultivation of the crop is primarily rainfed (Fyhri, 1998). The two major varieties of groundnut grown in the Gambia are the 28/206 and 73/33. 28/206 is a Virginia variety with a crop duration of about 120 days, while the 73/33 is a Spanish-Virginia cross with intermediate crop duration of 105 days.

The groundnut yield data and production area at divisional-level (see Figure 2.1 for the various divisional boundaries) from 1974-2002 were obtained from the Department of Planning, Ministry of Agriculture, The Gambia. The cultivation area and production of the various divisions are shown in Figure 2.7. Average All-Gambian annual groundnut production by weight for the period is 102 metric tons, with an average yield of 1150 kg/ha. The average yield for all West African countries for the same period is 873 kg/ha (FAOSTAT data, 2006). North Bank Division (NBD) and Central River Division (CRD) are the two most important divisions in groundnut production: together, the two divisions account for 58% and 60% of the total area cultivated and production weight respectively for the Gambia. In particular, the area of groundnut cultivation for NBD, as a fraction of country's total groundnut cultivation, increased by 10% in the past decade from the pre-1990 average of 27%.

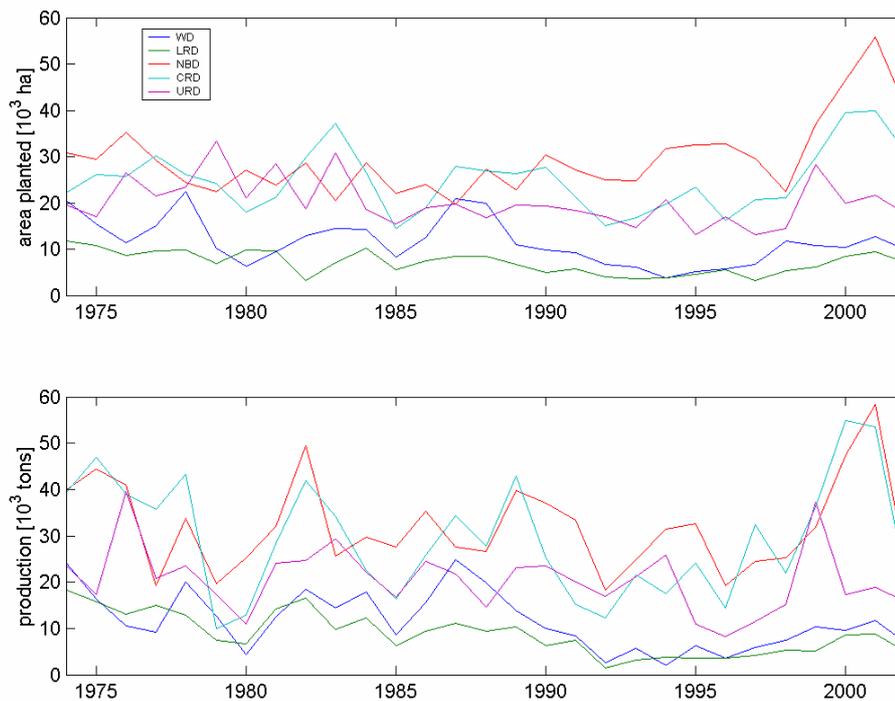


Figure 2.7 Area of cultivation (top) and production by weight (bottom) of divisional groundnut from 1974-2002.

Of the two major groundnut cultivars grown, 28/206 is the major groundnut crop cultivated. For 2001 and 2002, the average area used for cultivating the 73/33 variety ranges from about 14% (NBD) to 35% (Upper River Division, URD) of the total divisional area for groundnut cultivation. Central River Division (CRD), the second most important groundnut production division by weight, has 32% of the groundnut cultivating area growing the new variety (predominantly in the southern region of the division recently). Since 73/33 was introduced specifically in response to the increasingly erratic rainfall in the Gambia, it would be ideal if the yields of the two different varieties were treated separately. However, the records of divisional groundnut yield for the individual varieties are not available. Hence, for the remainder of this work, 28/206 is assumed to be the representative groundnut crop for all divisions.

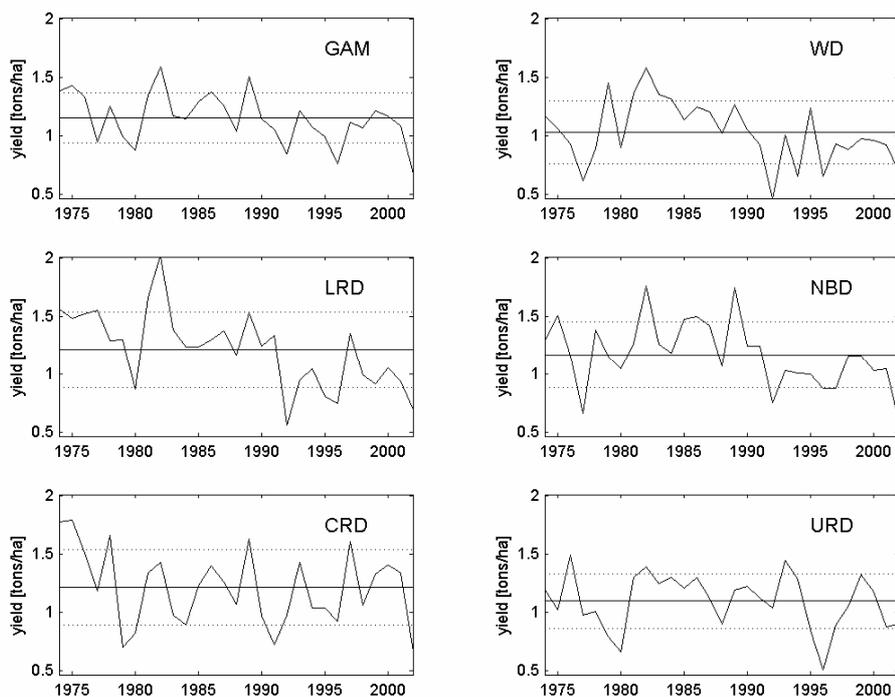


Figure 2.8 All-Gambia (GAM) and divisional ground yield [tons/ha] from 1974 to 2002. Horizontal solid line is the 1974-2002 mean for the individual yield time series. The dotted line is one standard deviation away from the mean.

The All-Gambian and divisional yields are shown in Figure 2.8. Two features can be identified from these yield time series.

- 1) A rapid increase in yield for all divisional yields, apart from CRD, at the beginning of the 80s followed by a decline. Except for CRD, above average yields were recorded for most of the years during the 1980's.
- 2) For NBD, Western Division (WD) and LRD, there was a sharp decline in yields between 1990-1991 and yield have been mostly below average since.

As groundnut cultivation in the Gambia is primarily rain-fed, it is reasonable to expect that the inter-annual variability of groundnut yields be to some degree correlated with the inter-annual variability of rainfall. Figure 2.9 shows the All-Gambia groundnut yield

anomaly and the total seasonal rainfall anomaly from 1974-2002. Both anomaly time series were formed by first linearly detrending the primary time series, and then standardising with the mean and the standard deviation of the detrended time series. The correlation coefficient ( $r$ ) between the two time series is 0.42 ( $p = 0.02$ ). Visual inspection of Figure 2.9 indicates stronger relations between yield and total seasonal rainfall during mid 70s, and from the late 80s, in particular the later part of the last decade. Near or above normal yields are attained in the early 80s, despite persistent lower-than normal seasonal rainfall.

There are reasons to believe that yields were not predominantly weather-limited during the early 80's: A new round of 5-year agriculture policies introduced from 1981-1985 resulted in a general improvement in farming technology, and the Economic Recovery Program in 1986 implemented policies aiming to increase groundnut production (Fyhri, 1998), may have been important in sustaining above average yields, despite the low rainfalls. The rapid decline of yields to mainly below long-term divisional average at the beginning of the 90s for NBD, WD and LRD may be a combined response to a long term decrease in productivity related to lower soil fertility caused by the removal of fertilizers subsidies in 1987 (Fyhri, 1998), and below average rainfall during the earlier 90s (Figure 2.9).

The relationship between rainfall and yield is more complex than is conveyed by correlation of annual rainfall amounts and yields. Duration and intensity of moisture stress at different stages of groundnut development can have a different impact on the final yield. This has been observed in field experiments (Pallas *et al.*, 1979; Nageswara Rao *et al.*, 1985; Williams *et al.*, 1985) and crop modelling studies (Rao *et al.*, 2000; Challinor *et al.*, 2005a). Sensitivity of groundnut yield for the Gambia to water stress at different growth stages is discussed in more detail in Chapter 6.

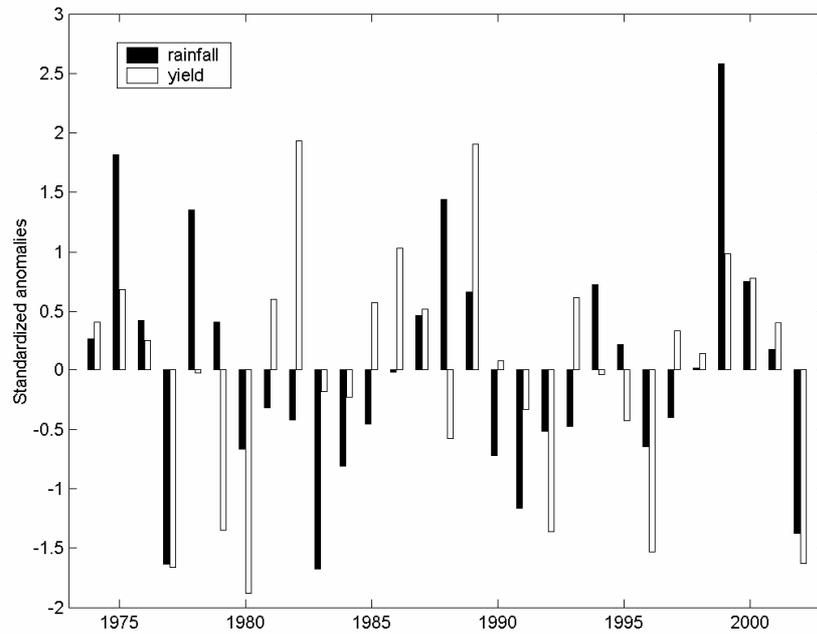


Figure 2.9 Standardised anomalies of ALL-Gambia total seasonal rainfall and groundnut yields from 1974-2002.

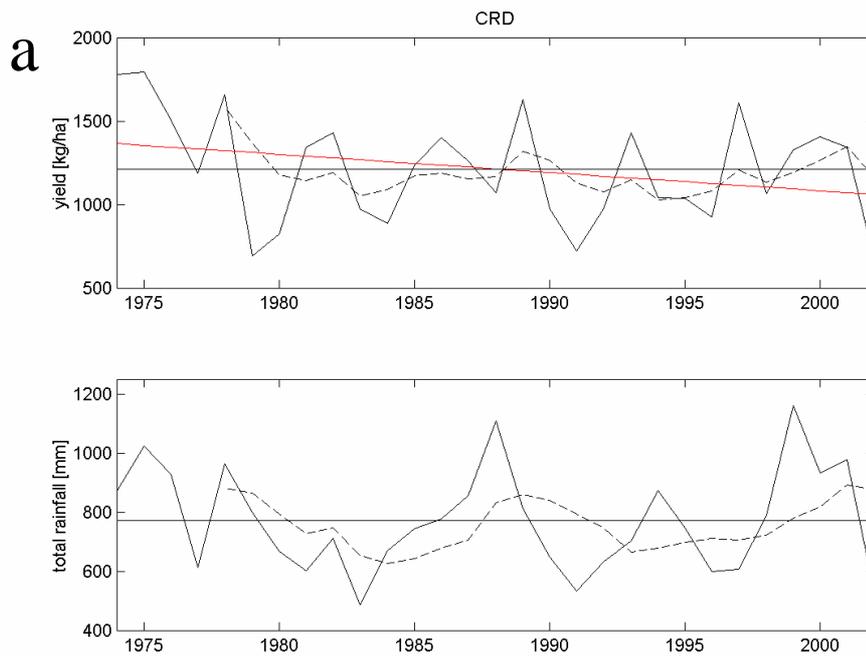


Figure 2.10 Groundnut yield and total seasonal rainfall from 1974-2002 for CRD (a) and NBD (b). Dashed lines are the 5-year moving average. Modelled trend is in red.

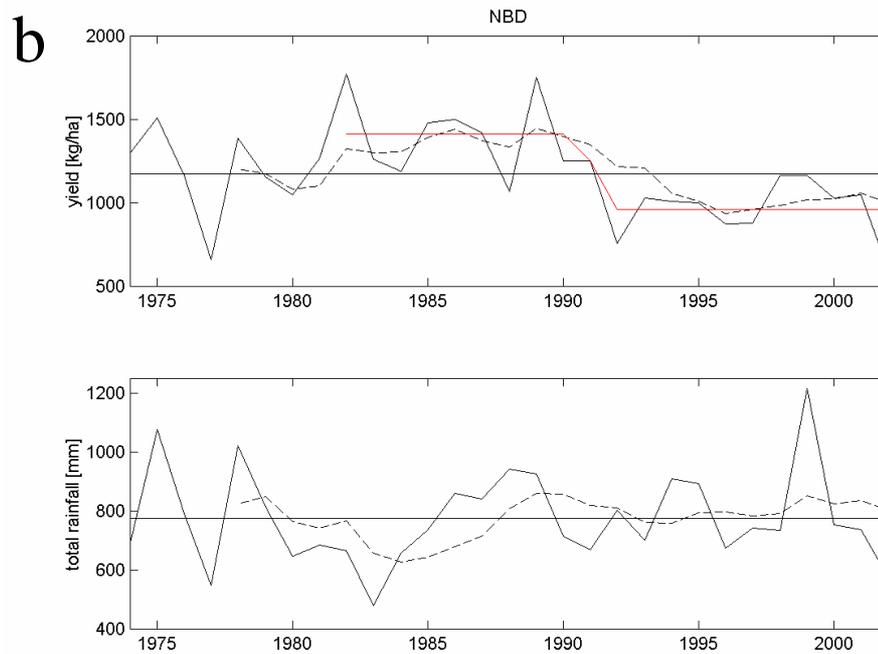


Figure 2.10 (Continue)

### Removal of technological trend of groundnut yield

In Chapter 6, groundnut yields are simulated using crop models for the two most important divisions for groundnut production for the country, CRD and NBD. As these agrometeorological crop models assume stationarity in all other factors that affect crop yields apart from weather, it is necessary to filter out any non-weather related variability in yields that the crop models are unable to reproduce (Challinor *et al.*, 2005a). Since it appears from the preceding discussion that various policies and management practices implemented may have different impact on the groundnut yields at divisional level, removal of the non-weather related trends in the divisional yields is considered separately. For CRD, visual comparison of yield and corresponding total seasonal rainfall in Figure 2.10a indicates relatively good correspondence between the trends of rainfall and yield in terms of their temporal variation; correlation between the 5-year moving averages of total season rainfall and yield is 0.5 ( $p < 0.01$ ). The 5-year moving averages of yield time series indicates decreasing yields prior to the early 1980s before stabilising,

with some slight decreases in the early to mid 1990s. For this division, a linear trend applied to the yield time series should suffice.

For NBD (Figure 2.10b), the 5-year yield and rainfall trend appeared to be negatively correlated ( $r = -0.31$ ,  $p = 0.13$ ). Rainfall was below average throughout the 80s while 5-year yield trend attained a broad maximum after the mid-80s. It has been suggested in the preceding section that improved policies implemented in this decade may have strong positive effect on the divisional yield. Here, the year 1982 is chosen as the starting year where the implemented policies were obviously effective; for that year, the yield attained is the highest for the division from the available data, despite the lower than normal rainfall. Although the 5-year moving average of rainfall since the late 1980s tends to be above or near normal, yield decreases sharply during the early part of the 90s, and from visual inspection of the 5-year moving averages, it stabilises to a below average value after 1994. Therefore, it appears that the recent NBD groundnut yield for the period (especially after 1993-1994) on average is different than its pre-1990 average. A most obvious contrast is seen when one compares the rainfall and groundnut yield for the periods 1982-1990 and 1994-1999: For the earlier period, the yield is above normal for all years apart from 1988, despite unfavourable rainfall conditions, especially for 1982-1985 (Section 2.3). For the period 1994-1999, seasonal rainfall for the division is above normal, while yields are below normal (either with respect to the 5-year moving average or the mean of the time series). In fact for 1999, with the highest seasonal rainfall on record available in this study, yields improved only to just below the long-term mean. There is also no evidence from the monthly rainfall distributions (not shown) to suggest that the lower than average yield for the latter period is due to an uneven distribution of rainfall.

As no yield data prior to 1974 is available, no detrending is attempted for NBD yields before 1982 due to the high inter-annual variability of yields. To account for the higher yield regime in the 1980s, the lower yield regime after 1994, as well as the transitional period in between the two yield regimes, a sigmoid model is adopted to describe the yield trend after 1982:

$$\bar{y}(t) = A - \frac{B}{(1 + e^{-k(t-t_0)})} \quad t \geq 1982 \quad 2.8$$

where  $A$ ,  $B$ ,  $k$  and  $t_0$  are to be determined from non-linear least square regression using the observed yield data. The fitted values are:  $A = 1409.2$  kg/ha,  $B = 452.1$  kg/ha,  $k = 22.8$  yr<sup>-1</sup> and  $t_0 = 1991$ .

The modelled trends for both divisions are superimposed on their respective yield time series for reference in Figure 2.10.

## 2.5 Conclusion

The data used in this thesis has been described. Preliminary analysis of rainfall and the Gambian groundnut yields has also been performed. Several features of rainfall are highlighted from the analysis:

- 1) Inter-annual variability of the seasonal total rainfall for the Gambia is high. This variability has been shown to be related to the inter-annual variability of frequency of rainy days, mean coverage of rainy areas, as well as the average intensity of rainfall within the rainy regions.
- 2) Rainfall within the country during the rainy seasons is characterised by “patchy” rain during rainy days and is punctuated by days of no rain. On rainy days, rain area coverage increases with the mean rainfall amount within rainy area.
- 3) Daily rainfall fields are non-stationary (varying spatial mean) and heteroscedastic.

The following summarised the main results from analysing the Gambian groundnut yields data for the two important groundnut-growing divisions:

- 1) For NBD, annual yields during most of the 80s were above average, followed by a rapid decrease during the early 90s, and it remains below average thereafter. This

variability of yield is unaccountable by weather. Such a pattern in yield trends is not evident in CRD. The reason for the differences in the yield patterns between the two divisions may be attributed to the different impact of agricultural policies introduced during the 80s.

2) A linear and a sigmoid non-weather related trend is modelled respectively for the yield time series for CRD and NBD. The detrended yields are used for yield predictions using weather-crop models in Chapter 6.

# Chapter 3 Spatial estimation of rainfall field from gauge data

## 3.1 Introduction

Daily average rainfalls for the Gambia at METEOSAT pixel resolution and at divisional scale are required for calibration of the satellite-rainfall model and crop modelling respectively in the subsequent chapters. One popular technique used for point or mean areal rainfall estimation given a set of raingauge observations is the application of an optimal linear interpolation method, commonly known as kriging. In the kriging paradigm, the rainfall field is considered as a regionalised random function (Goovaerts, 1997, Section 3.2.2). Many examples of the application of kriging for spatial interpolation of rainfall fields exist in literature; for flood monitoring (Sun *et al.*, 2000), gauge rainfall network design (Rodríguez-Iturbe and Mejía, 1974; Pardo-Igúzquiza, 1998; Campling *et al.*, 2001), remote sensor calibration (Brady, 1978; Creutin *et al.*, 1987, 1988; Grimes *et al.*, 1999), drought monitoring (Cheng *et al.*, 2000) and climatological applications (Duchon *et al.*, 1995; Pardo-Igúzquiza, 1998; Yuan and Duchon, 2001; Jeffrey *et al.*, 2001). Although being demonstrated in many studies to be better than, or at least as good as several other rainfall interpolation methods like Thiessen polygon, multiquadratic interpolation, inverse-square (Dirks *et al.*, 1998; Goovaerts, 2000; Syed *et al.*, 2003), one of the assumption for statistical inference of the kriging estimates requires the underlying random field (the rainfall field in this case) to be multivariate-Gaussian (Cressie, 1991, Chapter 3). This is obviously violated for many cases of rainfall accumulations at a short timescale, which are usually highly non-symmetric, as well as possessing a finite probability mass at zero – the daily Gambian rainfall being a good example.

One consequence of departure from a Gaussian distribution is that the kriging estimates no longer represent the optimal predictor for the underlying random field (Cressie, 1991, Chapter 3); for rainfall with fractional coverage (underlying rainfall distribution has a

finite probability mass at 0), kriging without consideration of the finite probability of zero rainfall often produces estimates that are positively biased outside the rainy region, and negatively biased within the rainy area. This results in a misrepresentation of the heterogeneity in the rainfall (Barancourt *et al.*, 1992; Seo, 1998).

For the Sahelian region, high spatial-temporal variability of rainfall at a scale of orders of kilometers, is known to affect crop yield (Graef and Haigis, 2001), and has been suggested to be important in the dynamics of convective storm development (Taylor and Lebel, 1998; Taylor *et al.*, 2003). Sivapalan *et al.* (1997) also highlighted that fractional coverage is important in hydrological processes in medium to large catchments.

One method of addressing non-normality is disjunctive kriging (Matheron, 1976) which involves the construction of optimal non-linear estimator. Although it is shown to provide better results than linear estimator (Azimi-Zonooz *et al.*, 1989), disjunctive kriging is only appropriate in situations when no intermittencies are expected. Barancourt *et al.*, (1992) and Seo (1998) proposed simpler estimation methods for rainfall with fractional coverage by using a two stage process (double kriging) in which 1) the rainfall map is estimated using the indicator kriging and 2) estimating separately the rainfall amount within the rainy regions through Simple or Ordinary Kriging. In this chapter, the method of Barancourt *et al.* (1992) (BCR92 hereafter), which is the simpler of the two methods, is adopted for the construction of the spatial rainfall field for the Gambia.

The remaining chapter is structured as follows: Section 3.2 provides a brief overview of kriging and a description of the rainfall interpolation method proposed in BCR92. Section 3.3 describes the application of the method to estimate daily Gambian rainfall, and presents the result of point validation before concluding the chapter in Section 3.4.

### 3.2 Stochastic spatial interpolation using double kriging

#### Overview of kriging

For point rainfall estimation, kriging attempts to obtain the linear estimator rainfall,  $Z(\mathbf{u})^*$  at an unsampled location  $\mathbf{u}$ , given  $n(\mathbf{u})$  observation  $\{Z(\mathbf{u}_j) \mid j = 1, \dots, n(\mathbf{u})\}$ , by solving the linear regression equation:

$$Z(\mathbf{u})^* - m(\mathbf{u}) = \sum_{j=1}^{n(\mathbf{u})} \lambda_j(\mathbf{u}) [Z(\mathbf{u}_j) - m(\mathbf{u}_j)] \quad 3.1$$

for the kriging weights  $\lambda_j(\mathbf{u})$  such that the error variance  $\text{Var}[Z(\mathbf{u})^* - Z(\mathbf{u})]$  is minimum, subjected to the constraint that the estimator is unbiased ( $E[Z(\mathbf{u})^*] = E[Z(\mathbf{u})]$ ).  $E[\ ]$  and  $\text{Var}[\ ]$  are the expectation and variance operators.  $n(\mathbf{u})$  in Equation 3.1, is the number of sampled observation used in predicting  $Z(\mathbf{u})^*$ .

Different variants of kriging have been formulated depending on the structure of  $m(\mathbf{u})$ . A widely used type of kriging in spatial rainfall estimation is Ordinary Kriging (OK), where  $m(\mathbf{u})$  is considered a constant but unknown within a neighbourhood around  $\mathbf{u}$  (Goovaerts, 1997, Section 5.3). In this case, the optimisation of the error variance under the unbiased constraint leads to a set of linear algebraic equations for determining the kriging weights:

$$\begin{aligned} \sum_{k=1}^{n(\mathbf{u})} \lambda_k(\mathbf{u}) \gamma(\mathbf{u}_j - \mathbf{u}_k) - \mu(\mathbf{u}) &= \gamma(\mathbf{u}_j - \mathbf{u}) \\ \sum_{k=1}^{n(\mathbf{u})} \lambda_k(\mathbf{u}) &= 1 \end{aligned} \quad 3.2$$

where  $\mu(\mathbf{u})$  is the Lagrange multiplier, and  $\gamma(\mathbf{h})$  is the semivariogram of  $Z(\mathbf{u})$  defined as  $2\gamma(\mathbf{h}) = E[(Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h}))^2]$ . The semivariogram quantifies the spatial coherence of the rainfall field, and is normally inferred from the observations. The optimised error variance corresponding to Equation 3.2 is

$$\sigma_{OK}^2(\mathbf{u}) = \sum_{j=1}^{n(\mathbf{u})} \lambda_j(\mathbf{u}) \gamma(\mathbf{u}_j - \mathbf{u}) - \mu(\mathbf{u}) \quad 3.3$$

If  $Z$  is isotropic, the semivariogram depends only on the distance  $|\mathbf{h}|$  (Goovaerts, 1997, Section 3.2.2).

From hereafter, OK estimates of a variable will be denoted by a superscript OK. For example, OK estimates of  $Z(\mathbf{u})$  is  $Z(\mathbf{u})^{OK}$ .

The Indicator Kriging estimates for the indicator transform of  $Z(\mathbf{u})$  can be expressed in a similar form as Equation 3.1:

$$I(\mathbf{u})^* - E[I(\mathbf{u})] = \sum_{j=1}^{n(\mathbf{u})} \lambda_j(\mathbf{u})(I(\mathbf{u}_j) - E[I(\mathbf{u}_j)]) \quad 3.4$$

where  $I(\mathbf{u})$  is the indicator random function defined as

$$I(\mathbf{u}) = \begin{cases} 1 & Z(\mathbf{u}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad 3.5$$

For the case where  $E[I(\mathbf{u})]$  is unknown, solutions for kriging weights of  $I(\mathbf{u})^{OK}$  using Ordinary Kriging can be obtained through Equation 3.2, using the semivariogram for  $I(\mathbf{u})$ . It can be shown that  $I(\mathbf{u})^{OK}$  is an estimate of the conditional probability  $\text{Prob}(I(\mathbf{u})=1 | \{n\})$  with respect to the observed  $n(\mathbf{u})$  indicator values. More details about Ordinary Kriging and Indicator Kriging can be found in many excellent references including Goovaerts (1997) and Cressie (1991).

#### Double Kriging method of BCR92

Following BCR92, the intermittent rainfall field,  $Z(\mathbf{u})$ , is defined as the product of two random fields (RF):

$$Z(\mathbf{u}) = I(\mathbf{u}) \times F(\mathbf{u}) \quad 3.6$$

where  $I(\mathbf{u})$  is the indicator field defined in Equation 3.5, and  $F(\mathbf{u})$  is a non-zero continuous RF, representing the rainfall amount if the location is raining. A ‘‘Double Kriging’’ estimate for the rainfall is given by

$$Z(\mathbf{u})^{DK} = I(\mathbf{u})^* \times F(\mathbf{u})^{OK} \quad 3.7$$

$F(\mathbf{u})^{OK}$  being the ordinary kriging estimates of  $F(\mathbf{u})$  and

$$I(\mathbf{u})^* = \begin{cases} 1 & I(\mathbf{u})^{OK} > I_c \\ 0 & \text{otherwise} \end{cases} \quad 3.8$$

where  $I(\mathbf{u})^{OK}$  is the ordinary kriging estimate of  $I(\mathbf{u})$ , and  $I_c$  is a threshold such that

$$E[I(\mathbf{u})^*] = E[I(\mathbf{u})^{OK}] \quad 3.9$$

If  $I(\mathbf{u})$  and  $F(\mathbf{u})$  are independent,  $I(\mathbf{u})^{OK}$  and  $F(\mathbf{u})^{OK}$  can be determined using indicator kriging and performing kriging using the set of rainy observations respectively. Mean areal rainfall over the  $j^{th}$  pixel support,  $Z_v^j$ , can then be approximated as an average of  $N$  point rainfall estimates over a  $N$ -points representation of the pixel area  $\{\mathbf{u}_\alpha^j \mid \alpha = 1, \dots, N\}$ ,

$$Z_j^{BDK} \approx \frac{1}{N} \sum_{\alpha=1}^N I(\mathbf{u}_\alpha^j) * \times F(\mathbf{u}_\alpha^j)^{OK} \quad 3.10$$

The superscript BDK denotes a ‘Block’ average of the Double Kriging estimates over the area.

### 3.3 Gambian daily rainfall field

The double kriging (DK) method was used for spatial interpolation from daily raingauge observations, for the Gambia during the 1974 – 2002 rainy seasons (June to November). Preliminary statistical analysis using gauge observations in Chapter 2 shows spatial variation in the  $E[I(\mathbf{u})]$  and  $E[F(\mathbf{u})]$ , as well as dependency of rainfall variability on its mean amount (Figures 2.4 and 2.5). These results suggest that both  $I(\mathbf{u})$  and  $F(\mathbf{u})$  for the daily Gambian rainfall are strictly non-stationary and non-ergodic. Several kriging methods exist when underlying RF have non-stationary means, namely, kriging with a trend model (Goovaerts, 1997, Section 5.4), and median polish kriging (Cressie, 1991, Section 3.5). However, with the small number of daily gauge observations and high variability of the rainfall, it is doubtful whether the trend can be sufficiently captured by such methods. Certainly, it is not advisable to use the climatological mean at various gauge location as the daily mean due to the non-ergodic nature of daily rainfall fields. In spite of these difficulties, it is not unreasonable to treat  $I(\mathbf{u})$  and  $F(\mathbf{u})$  for Gambian daily rainfall as a stationary isotropic RF over the entire domain as a compromise and employ Ordinary Kriging with the following arguments:

1) With regards to stationarity of the random fields, it is noted that for several case studies in orographic rainfall, a locally constant mean (underlying RF has a constant mean for some finite neighbourhood) is found to be a good approximation when applying

stochastic interpolation methods, despite the overall orographic variation of the mean rainfall (Chua and Bras, 1982; Creutin and Obled, 1982; BCR92).

2) If the non-ergodic daily rainfall field  $Z$  is of a form  $Z(\mathbf{u}) = Y(\mathbf{u})\sigma_Z(d) + m_Z(d)$ , where  $Y(\mathbf{u})$  is a stationary RF with zero mean and unit variance,  $\sigma_Z(d)$  and  $m_Z(d)$  are respectively the variance and mean of the  $Z$  over the domain at day  $d$ , the variogram of  $Z$ ,  $\gamma_Z(h) = E[\sigma_Z^2(d)]\gamma_Y(h)$ , as long as  $Y(\mathbf{u})$  and  $\sigma_Z(d)$  are independent (BCR92). In this case, the spatial variability of  $Z$  depends only on the underlying climatological variogram  $\gamma_Y(h)$  (Lebel *et al.*, 1987). This argument is extended to both  $I(\mathbf{u})$  and  $F(\mathbf{u})$  for the daily rainfall fields of the Gambia. The spatial correlation structures of  $I(\mathbf{u})$  and  $F(\mathbf{u})$  are assumed to be fully defined by their respective climatological variograms,  $\hat{\gamma}_I(h)$  and  $\hat{\gamma}_F(h)$ .

To obtain the monthly climatological variogram for the daily Gambian rainfall, daily observed  $I(\mathbf{u})$  and  $F(\mathbf{u})$  were first standardised as follows:

At each location  $\mathbf{u}$  for a given day  $d$ , the standardised  $I(\mathbf{u})$  and  $F(\mathbf{u})$  are computed as

$$\hat{I}(\mathbf{u}) = \frac{I(\mathbf{u}) - m_I(d)}{\sigma_I(d)} \quad 3.11$$

and

$$\hat{F}(\mathbf{u}) = \frac{F(\mathbf{u}) - m_F(d)}{\sigma_F(d)} \quad 3.12$$

where

1)  $m_I(d)$  and  $\sigma_I(d)$  are respectively the average indicator values and standard deviation of the indicator for day  $d$ .

2)  $m_F(d)$  and  $\sigma_F(d)$  are respectively the average positive rainfall amount and standard deviation of the positive rainfall amount for day  $d$ .

Next, experimental climatological variograms (Goovaerts, 1997, Section 4.1.3) for  $\hat{I}(\mathbf{u})$  and  $\hat{F}(\mathbf{u})$  were computed for each month from June to November. The monthly

climatological variograms were then fitted to an exponential model (Goovaerts, 1997, Section 4.2.1) using weighted non-linear least square (Pardo-Igúzquiza, 1999).

Figure 3.1 shows the standardised experimental variograms and the fitted climatological variogram for  $I(\mathbf{u})$  and  $F(\mathbf{u})$  for each month. As the number of rainfall events for November (end of rainy season) was small, it is difficult to infer the variogram model parameters from the sample; the variogram model for October was adopted instead. The nugget effects in the indicator variograms are large (70%-80% of the sill), suggesting the existence of smaller scale intermittencies unresolved by the raingauge network. Pardo-Igúzquiza *et al.* (2006), working with a denser gauge network (104 gauges) in the Kenya over a region similar in size as the present case study, obtained a nugget of similar magnitude (about 66%).

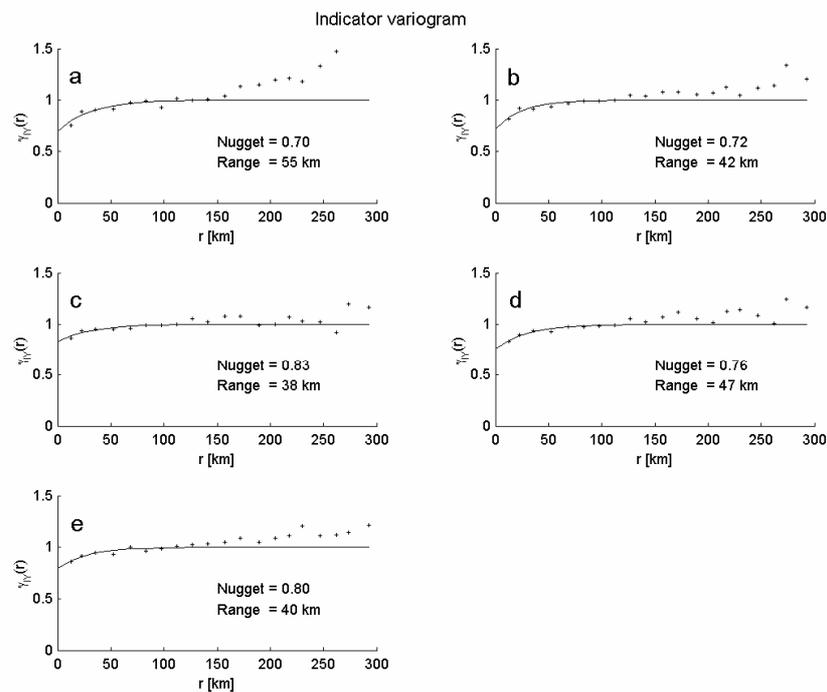


Figure 3.1 Climatological variograms of  $I$  and  $F$  for daily Gambian rainfall. a) June; b) July; c) August d) September; e) October. Range is defined as the distance where the semivariance is 95% of the sill.

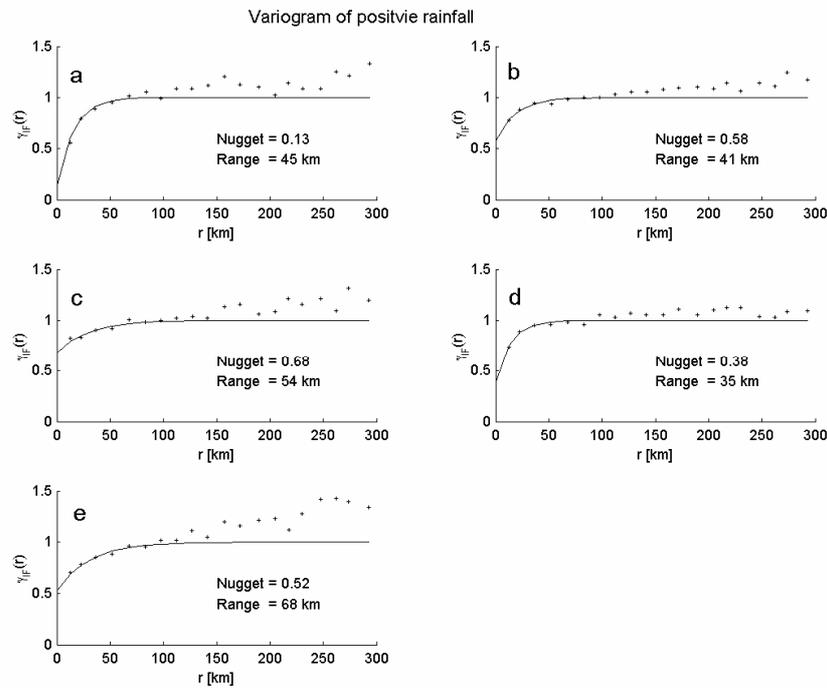


Figure 3.1 (Continue)

In order to perform Double Kriging, the Ordinary Kriging estimates  $I^{OK}(\mathbf{u})$  and  $F^{OK}(\mathbf{u})$  were first obtained. For days with 3 or less rainy gauges, it is doubtful that  $F^{OK}(\mathbf{u})$  are representative over the entire region. For such days, all sampled locations inclusive of zero rainfall were used in the kriging of  $F(\mathbf{u})$ . Estimates of  $E[I(\mathbf{u})^{OK}]$  of the daily rainfall field were determined implicitly from the Indicator Kriging algorithm. The appropriate  $I_c$  satisfying Equation 3.9 was obtained through iteration, given that  $E[I(\mathbf{u})^*]$  is the arithmetic average  $I(\mathbf{u})^*$  over the domain for a given  $I_c$ .

### 3.3.1 Point validation

As the number of valid daily gauge observations is small, separating a subset of the gauge network for validation purpose is not feasible. A cross-validation method has been used instead to assess the performance of the Double Kriging method for daily Gambian rainfall: For each day, one gauge observation is sequentially omitted for validation. The rainfall for the validating location was also estimated using OK, chosen as an alternate

estimation method for comparison. Days when all gauges reported zero rainfall and all gauges reported rainfall were omitted; the earlier case produced trivial results, while both methods are equivalent for the latter (since DK degenerates to OK when all gauges are rainy).

The following skill measures were used to compare the two method of rainfall estimation:

Overall skill measures are 1) the mean error of rainfall estimation:

$$ME_Z = \frac{1}{N} \sum_{i=1}^N z_i - z_i^+ \quad 3.13$$

where  $z^+$  is the estimated rainfall (either by DK or OK), and  $N$  is the number of validation data, and 2) the correlation coefficient between observed and estimated rainfall:

$$r_Z = \frac{\sum_{i=1}^N (z_i^+ - \bar{z}^+)(z_i - \hat{z})}{\sqrt{\sum_{i=1}^N (z_i^+ - \bar{z}^+)^2 \sum_{i=1}^N (z_i - \hat{z})^2}} \quad 3.14$$

where  $\hat{z}$  is the sample mean of the validation data, and  $\bar{z}^+$  is the mean of the rainfall estimates.

Following BCR92, two measures are used to assess the skill in reproducing intermittencies in the rainfall. These are: 1) the mean error in collocating the rainy area defined as

$$ME_I = \frac{1}{N} \sum_{i=1}^N I(z_i) - I(z_i^+) \quad 3.15$$

and 2) the Rousseau Index,  $I_R$ , defined as (Rousseau, 1980)

$$I_R = \frac{4p_{11}p_{00} - (p_{10} + p_{01})^2}{(2p_{11} + p_{01} + p_{10})(2p_{00} + p_{01} + p_{10})} \quad 3.16$$

where  $p_{ij} = \text{Prob}(I(z)=i, I(\hat{z})=j)$ . Note that from the definition,  $I_R$  equals 1 for a perfect prediction ( $p_{01} = p_{10} = 0$ ), and  $-1$  for a worst possible prediction ( $p_{00} = p_{11} = 0$ ). For a random prediction constrained to the observed rainfall frequencies,  $I_R = 0$ . Skill in

estimating the rainfall amount within the rainy area is assessed using the mean error in the rainfall amount,

$$ME_F = \frac{1}{N} \sum_{j=1}^N z_j - z_j^+ \text{ where } z_j, z_j^+ > 0 \quad 3.17$$

Table 3.1 summarises the various performance measures of  $Z^{DK}$  and  $Z^{OK}$  for daily Gambian rainfall. OK performs better than DK in terms of mean estimation error and correlation. The bias in  $ME_Z$  is due to the interdependence of  $I(\mathbf{u})$  and  $F(\mathbf{u})$ , which is indicated in the preliminary analysis of the rainfall in Chapter 2 (Figure 2.6). However, as this bias is about 3% of the mean rainfall, this deficiency should not be critical in practice. Performance of DK in reproducing the mixed distribution characteristics of rainfall is better than OK: Distribution of  $Z^{DK}$  is comparatively closer to that of the observed than the distribution  $Z^{OK}$  for the lower rainfall amount (Figure 3.2). The mean positive rainfall amount from DK is near observed compare to OK ( $E(Z | Z>0) = 16.4$  mm,  $E(Z^{DK} | Z^{DK} > 0) = 15.5$  mm,  $E(Z^{OK} | Z^{OK} > 0) = 6.53$  mm), though both methods still underestimate the variance of the positive rainfall amount ( $\text{Var}(Z | Z>0) = 307$  mm<sup>2</sup>,  $\text{Var}(Z^{DK} | Z^{DK} > 0) = 99$  mm<sup>2</sup>,  $\text{Var}(Z^{OK} | Z^{OK} > 0) = 53$  mm<sup>2</sup>). The probability of rainfall from DK (OK) is 0.40 (0.99), agreeing well with observed value of 0.39.

Apart from better reproduction of global rainfall probability, collocation of rainy area in terms of  $ME_I$  and  $I_R$  (Table 3.1) by DK is superior. Estimation of positive rainfall amounts within the rainy region is better for DK than OK (Table 3.1). The positive  $ME_F$  and the negative  $ME_I$  for  $Z^{OK}$  are due to the tendency of OK to maintain zero bias on the overall spatial rainfall field (constant mean rainfall over the domain), by underestimating rainfall over the rainy areas and overestimating rainfall over non-rainy areas. For intermittent rainfall, OK therefore reduces the heterogeneity in the rainfall field.

	$ME_Z$ [mm]	$r_z$	$ME_I$	$I_R$	$ME_F$ [mm]
$Z^{DK}$	0.24	0.39	-0.01	0.30	0.6
$Z^{OK}$	0.00	0.49	-0.60	-0.44	6.3

Table 3.1 Various skill measurements of the Double Kriging and Ordinary Kriging for daily Gambian rainfall field.

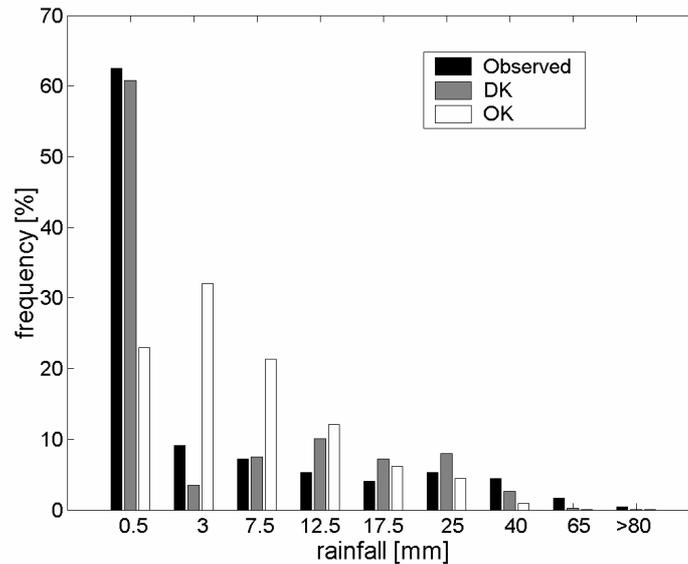


Figure 3.2. Frequency distribution of observed and estimated rainfall using DK and OK.

### 3.3.2 Extension to obtain mean pixel rainfall estimates

The previous subsection demonstrated the superiority of DK over OK. DK was therefore adopted to obtain the mean areal rainfall at METEOSAT pixel support ( $0.05^0 \times 0.05^0$ ) for the Gambia. This was achieved by interpolating rainfall from the available daily gauge network observation over a 16 points (4 points  $\times$  4 points) discrete representation of each pixel using Equation 3.10. One issue encountered was that BCR92 did not provide the estimation variance of the Double Kriging method due to the uncertainty in quantifying the error in estimating the hard indicator  $I(\mathbf{u})$ . Recently, Pardo-Igúzquiza *et al.* (2006) proposed a method for estimating variance for DK for point estimates using maximum likelihood method. Although promising, the extension of the estimation variance for average areal rainfall is not available. However, as uncertainties in the rainfall estimates are required in the subsequent chapter, a simple approximation to the error was obtained by ignoring the uncertainties in the identification of the rainy areas by the DK algorithm. This is a reasonable simplification since the estimated collocation error (Table 3.1) of

rainy region for point rainfall using the DK algorithm is about 17% on average. In this case, the error in the estimated rainfall field is assumed to be confined to the rainy areas whose mean square error (MSE) for pixel  $j$  can be approximated by

$$E[(Z_{Vj} - Z_j^{BDK})^2] \approx I_{Vj} \sigma_{jBOK}^2 \quad 3.18$$

where  $Z_{Vj}$  is the mean pixel rainfall for pixel  $j$ ,  $Z_j^{BDK}$  is the DK estimates of  $Z_{Vj}$  (Equation 3.10),  $I_{Vj}$  is the fractional coverage of rain for pixel  $j$ , using the arithmetic mean of the indicator value of the discretised points in  $j$ , and  $\sigma_{jBOK}^2$  is the Ordinary Block Kriging variance (Goovaerts, 1997, Section 5.5) of the mean pixel positive rainfall amount for pixel  $j$ . The derivation of Equation 3.18 is given in the Annex B. Estimates for point rainfall using DK, in the case of Gambian rainfall, are shown to be biased from the preceding discussion, implying that  $Z_j^{BDK}$  is biased as well. The magnitude of the bias for each pixel estimate is the average of the bias of the point estimates representing the pixel. As there is no data to validate the pixel rainfall estimates, the magnitude of this bias is not known. It is however, not unreasonable to assume that the ratio of the bias to the average pixel rainfall remains about the same order of that for point rainfall estimates. Therefore the computed MSE using Equation 3.18 is considered a valid the estimation variance of  $Z_j^{BDK}$ . An example of daily areal rainfall at METEOSAT pixel resolution and its associated error estimates for the Gambia is shown in Figure 3.3.

### 3.4 Conclusion

The mean Gambian daily rainfall at METEOSAT pixel resolution has been obtained by applying Double Kriging to the observed raingauge data, so as to estimate the intermittencies of rainfall and the rainfall amount. This method gives a reasonable representation of the fractional coverage, which is a dominant feature of the daily rainfall in the region. For point rainfall estimation, the method has better skill than Ordinary Kriging in delineating rainy and non-rainy locations, as well as in estimating rainfall amounts within rainy area. However DK estimates showed a slight bias and lower correlation in the overall rainfall estimation. An exact method for computing the

estimation variance for DK areal estimates is not available. As the estimation error is required for the development of satellite rainfall estimates in the next chapter, a simple approximation has been calculated by ignoring the error in collocating rainy pixels to provide some measure of the uncertainty in the DK estimates.

Although the method presented in this chapter is adequate for subsequent investigations presented in the later chapters, several recommendations can be made to improve the accuracy of rainfall estimates:

1) All gauges within the Gambia, as far as this author is aware, were utilised in the interpolation. However, gauge data from regions in Senegal near the Senegalese-Gambian border were not available at the time of the investigation. Due to the elongated geometry of the Gambia, the inclusion of this extra information if available, could be useful in better representation of the rainfall especially for the North Bank Division and the Eastern-most region of the country.

2) As there is evidence of interdependence of  $I(\mathbf{u})$  and  $F(\mathbf{u})$ , other variants of kriging that take account of this relation can be employed in DK. For estimating  $I(\mathbf{u})$ , one possible alternative is to replace Indicator Kriging with Indicator Cokriging (Goovaerts, 1997, Section 7.3) to take account of the rainfall amount information. Although such method is superior to Indicator Kriging in principle, it is more computationally expensive and does not always guarantee better performance than Ordinary Indicator Kriging (Goovaerts, 1994). For estimating  $F(\mathbf{u})$ , Seo (1998) suggested an optimised linear estimation of the positive rainfall taking into account all the gauge information (including zero rainfall). This could be a candidate to replace the present method of estimating  $F(\mathbf{u})$ , which uses only the positive rainfall.

3) Current error variance for the areal rainfall is only an approximation. Further investigation needs to be undertaken to generalise the DK point estimation variance (Pardo-Igúzquiza *et al.*, 2006) for areal rainfall estimates. However, interdependence of intermittencies and rainfall amount within the DK would also render the estimation

variance as proposed by Pardo-Igúzquiza *et al.* (2006) invalid. A point estimation variance in the case where  $I(\mathbf{u})$  and  $F(\mathbf{u})$  are dependent has been suggested by Seo (1998). Further investigation could be undertaken to generalise the estimation variance for areal estimates.

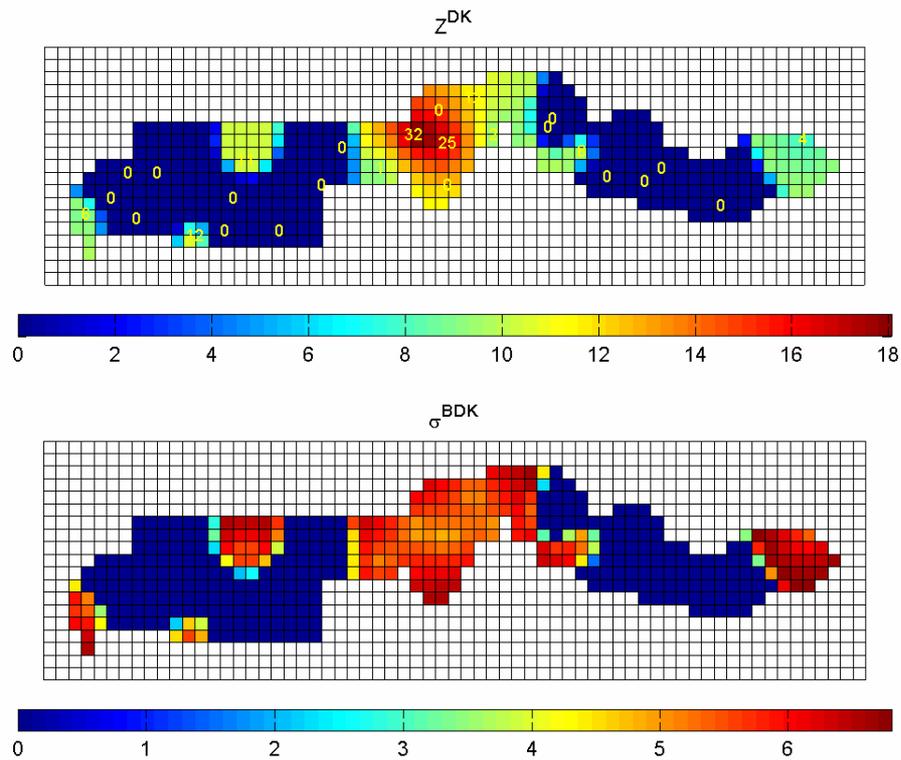


Figure 3.3 Example of the estimated daily mean pixel rainfall using DK for the Gambia (a) and the corresponding adopted standard error (b). The corresponding gauge rainfall observations are included in (a) for comparison. All units are in mm.

# Chapter 4 Satellite rainfall estimation

## 4.1 Introduction

Satellite-based estimates of rainfall have found wide applications in agriculture and hydrology both operationally and in research. The attraction of these rainfall estimates comes from its better spatial and temporal coverage when compared to raingauge-based observations. Although ground-based weather radar, when available, gives superior rainfall observations, a network dense enough to capture regional rainfall is still limited to developed nations owing to the high cost in deployment and maintenance of such radars. In regions where ground based observations are sparse or non-existent, satellite based rainfall estimates become vital. In the following paragraphs, a brief overview of the science, technologies and techniques involved in the remote sensing of rainfall from space are provided. As there are many reviews on various aspect of this subject, the intention in this introduction is to provide a concise background discussion. Excellent detailed reviews can be found in Barrett and Martin (1981), Kidd (2001) and Levizzani *et al.* (2002).

### Overview

Observations of terrestrial weather systems on space-borne platforms began with the launch of the first weather observation satellite TIROS-1 in 1960. However, it is the subsequent deployment of the constellation of geostationary and sun-synchronous polar-orbiting satellites beginning in the mid to late 1970's that herald the beginning of satellite-based rainfall measurements, at acceptable temporal and spatial resolutions globally. Most useful for remote sensing of rainfall are the satellites with on-board sensors operating in the microwave (at several wavelengths between 6 GHz to 85.5GHz), thermal infrared (8 to 12.5  $\mu\text{m}$ ) and the visible (0.5-0.7  $\mu\text{m}$ ) band of the electromagnetic spectrum. Both the thermal infrared (TIR) and visible band (VIS) are useful for estimating convective rainfall. The physical basis of TIR and VIS rainfall inference is that convective cloud precipitations are often associated with high (and hence cold) cloud

tops as well as high cloud albedo. Hence the cloud top temperature and cloud brightness of precipitating clouds are indirect measures of precipitation intensity (Woodley and Sancho, 1971; Martin and Suomi 1972; Scofield and Oliver, 1977; Griffith *et al.*, 1978; Stout *et al.*, 1979). The underlying basis for rainfall retrieval from microwave radiometry rests on the fact that the absorption and scattering characteristics of cloud particles in the microwave spectrum depends on their shape, size and phase (liquid or ice). In principle, modification to the upwelling terrestrial microwave radiation by precipitation-size particles allows rainfall retrieval from the microwave channels (Barrett and Martin, 1981).

Geostationary	Country of origin	Sensors	Sub-satellite point
Meteostat-8	Europe	VIS/IR	0°
Meteosat-7	Europe	VIS/IR	0°
Meteosat-6	Europe	VIS/IR	10°E
Meteosat-5	Europe	VIS/IR	63°E
GOMS-1	Russia	VIS/IR	76°E
INSAT-1D	India	VIS/IR	74°E
INSAT-2E	India	VIS/IR	83°E
FY-2	China	VIS/IR	105°E
MTSAT-1	Japan	VIS/IR	140°E
GOES-10	USA	VIS/IR	135°W
GOES-12	USA	VIS/IR	75°W

Low-Earth orbits	Country of origin	Sensors	Coverage
NOAA	USA	Vis/IR	90°N-90°S
Meteor	Russia	Vis/IR	90°N-90°S
FY-1	China	VIS/IR	90°N-90°S
TRMM	USA/Japan	Vis/IR/MW/radar	40°N-40°S
DMSP	USA	Vis/IR/MW	89°N-89°S

Table 4.1 Current meteorological satellites. Adopted and updated from Kidd (2001).

Table 4.1 shows the meteorological satellites currently in operation that have been used for rainfall retrieval in operations and research. Crucial to the task of satellite rainfall estimation is the network of geostationary meteorological satellites providing continuous

monitoring of global weather systems at a sampling rate of 30 minutes, using on-board thermal infrared and visible sensors. Spatial resolution typically is about 5 and 2.5 km at sub-satellite point (SSP) for the TIR and VIS channel respectively. The newest generation of geostationary satellites like Meteosat-8 (previously known as MSG-1) have improved on the resolution of the TIR channel to 3 km at SSP, increased sampling frequency of 15 minutes, and an increased number of channels in the TIR spectrum, offering better cloud classification and spatio-temporal description of the cloud field. Earth observing polar orbiters, though lacking the temporal resolution of geostationary satellites, have better spatial resolution at their Field of View (FOV) for both TIR and VIS channels. For example, the Advanced Very High Resolution Radiometers (AVHRR) onboard the NOAA series of satellites, operating at TIR and VIS band have resolution of about 1 km. More importantly several polar orbiters, carry passive microwave sensors, arguably the most important being the Special Sensor Microwave/Imager (SSM/I) sensors on the Defense Meteorological Satellites (DMSP) series of satellite which provide three dual polarized frequency channels and a vertical polarized channel with resolution at the highest frequency at about  $13 \times 15$  km.

Rainfall retrievals from the different frequency channels have different strengths and weaknesses. As TIR and VIS rainfall retrieval solely relies on the indirect relations between cloud-top temperatures or cloud brightness with surface rainfall, they are subjected to error in misidentification of precipitating cloud and rainfall amounts. High-level non-precipitating cloud, when present, may sometimes be indistinguishable from convective cold cloud tops in TIR images, leading to overestimation of the rainy area. Although visible images are useful in such circumstances in delineation of rainy/non rainy regions since sufficiently thin ice clouds are almost transparent in the visible spectrum, they are only available during daytime. In addition, the complex dependence of cloud brightness in the visible spectrum with cloud geometry and orientation with respect to solar angle (Reynolds and Vonder Haar, 1973; Griffith and Woodley, 1973; McKee and Cox, 1974) requires a more sophisticated algorithm, or human intervention. Unlike TIR and VIS, microwave radiation is cloud penetrating, modified mainly by precipitation-size particles. Hence, microwave (MW) radiometry provides a more direct

measurement of rainfall, in principle. MW rainfall retrieval however, suffers from various problems including the so-called ‘beam-filling’ effect due to unresolved rainfall heterogeneity within sensor FOV, as well as having a non-linear relationship between brightness temperatures and high rainfall rate which can result in bias in estimated rainfall intensity (Barrett and Martin, 1981; Chiu *et al.*, 1990); infrequent sampling due to the limited overpasses of the polar orbiters, which contributes to the error in accumulated rainfall estimates (Bell *et al.*, 1990). Furthermore, the emissivity of the land surface at the microwave spectrum is higher and more variable in space and time than over the ocean. The latter problem resulted in additional complexity of rainfall retrieval over semiarid regions of Africa. In these regions, emissivity of surface with partial vegetative cover is strongly dependent on soil moisture, which in turn have strong spatio-temporal variability depending on antecedent rainfall (Morland *et al.*, 2001).

There are a multitude of techniques developed for the retrieval of rainfall from satellites available in the literature. Most of these techniques can be broadly classified under two categories.

*TIR technique* utilise the TIR channel of geostationary satellites, taking advantage of the good coverage and spatio-temporal resolution from the constellation of geostationary weather satellites. Many such TIR rainfall estimation algorithms are different in terms of the sophistication of the algorithm employed for rainfall retrieval. The simplest rainfall retrieval method is the GPI (Arkin and Meisner, 1987), which uses a single TIR brightness temperature for delineating rainy/non-rainy pixels, and a fixed rainfall rate is assigned to the rainy regions. The TAMSAT method (Milford and Dugdale, 1990) determines the threshold temperature and rainfall rate based on historical raingauge observations, giving it more accountability in the regional climatic variation of rainfall than GPI. Many methods have been proposed to improve the shortcomings of such simple TIR-techniques; the inclusion of real-time or near real-time gauge data (Herman *et al.*, 1997; Laurent *et al.*, 1998; Grimes *et al.*, 1999; Todd *et al.*, 1999); the inclusion of meteorological field data from numerical weather models (Grimes *et al.*, 2003; Funk *et al.*, 2003; Funk and Michaelsen, 2004); the inclusion of larger-scale structure in the TIR

imageries in the neighbourhood of the pixel (Adler and Negri, 1988; Bellerby, 2004), as well as ‘bi-spectral techniques’ that include corresponding VIS channel data (Scofield and Oliver, 1977; Lovejoy and Austin, 1979; Martin and Howland, 1986; Tsintikidis *et al.*, 1999).

*MW-TIR techniques* utilise both MW and TIR channels of satellites in rainfall retrieval. Many such algorithms have been investigated (Alder, 1993; Bellerby and Barrett, 1993; Xu *et al.*, 1999; Hsu *et al.*, 1999; Sorooshian *et al.*, 2000; Bellerby *et al.*, 2000; Todd *et al.*, 2001; Kidd *et al.*, 2003; Joyce *et al.*, 2004). These algorithms are based on the premise that such combinations capitalise on the more direct rainfall retrieval from MW sensors on polar orbiters, as well as the better spatio-temporal coverage from TIR sensors on geostationary satellites. The MW rainfall estimates, whenever available, enable dynamic recalibration of TIR rainfall retrieval. Therefore, MW-TIR RFE is potentially useful for regions where ground observations are sparse or non-existent. In addition, methods to merge raingauge network data, as those proposed by Xie and Arkin (1996), could be useful in further improving the rainfall estimation of combined MW-TIR techniques over land, where rainfall retrieval from MW sensors is complicated by the high and variable emissivity at the microwave spectrum. Recently real-time daily rainfall estimates over Africa from the African Rainfall Estimation Algorithm (RFE version 2.0) based on such combined gauge-network observations and multi-platform multi-sensor satellite rainfall estimates has been made available routinely by the NOAA Climate Prediction Center.

In this chapter, a simple TIR technique using METEOSAT TIR data modified from the TAMSAT algorithm, originally proposed by Milford and Dugdale (1990), is presented. The TAMSAT algorithm (Section 4.2) uses a linear TIR-rainfall relationship at pixel level, empirically determined using historical raingauge observations. Currently, TAMSAT dekadal (10-day) rainfall estimates for Africa are routinely generated for drought monitoring, primarily for the context of food security. Recently, interests are emerging in the possibility of using satellite-based rainfall estimates at a daily time scale as inputs for crop models or hydrological models. Grimes and Diop (2003) reported the

potential of TAMSAT at daily time scale for riverflow prediction for the Bakoye catchment in West Africa using a hydrological model. Hence, it is of interest to investigate the feasibility of using simple daily rainfall estimation algorithms, such as those based on TAMSAT, in driving weather-crop models for crop yield prediction in Africa. Although TIR-techniques with real-time gauge inputs and MW-TIR techniques potentially give better accuracy in their rainfall estimates than TIR-rainfall algorithm, the latter is still relevant in Africa for the following reasons: 1) Rain-gauge observations are sparse and often suffer from significant delays in reporting. These would imply that algorithms that require real-time or near real-time calibration, or adjustment using rain gauges, would not be effective or applicable for some regions. 2) For operational purposes, there are attractions in a simple algorithm. 3) With the implementation of Global Precipitation Measurement (GPM) in the near future, realising a 3 hour return-rate of microwave soundings globally (Smith, 2003), it is reasonable to believe that multiplatform TIR-MW rainfall retrieval techniques will gain prominence in rainfall monitoring in Africa. Since most TIR-MW algorithms rely on some form of TIR-rainfall retrieval algorithm in between MW sensors overpass, deficiencies in MW rainfall retrieval could be transferred to the TIR rainfall estimates (Kidd *et al.*, 2003). Therefore TIR-MW rainfall estimation over many regions of the continent still requires careful validation. Investigations in hydrological and crop yield model performance using simpler TIR-rainfall estimates can provide a good starting point to focus potential areas of improvement in such multi-sensor rainfall retrieval techniques, in the context of practical hydrological and agricultural applications in Africa.

As TAMSAT algorithm relies on a statistical TIR-rainfall relationship, it is inevitable that a certain amount of uncertainty is associated with the rainfall estimates. In regions where daily rainfall is intermittent and highly skewed, TAMSAT may be inadequate for two reasons: 1) There are significant uncertainties in delineating the non-rainy areas at the optimum temperature threshold at daily timescale, which are considered as negligible in TAMSAT for dekadal rainfall. 2) The variability of the rainfall associated with a rainfall estimate is high, and the error distribution is non-Gaussian: These characteristics are not sufficiently captured by TAMSAT's limited error description (standard deviation). The

main intention of this chapter is to address this shortcoming by introducing a modified version of TAMSAT, abbreviated as TAMSAT1 hereafter, that provides a detailed description of the estimation uncertainties at the native spatial scale. As discussed in the following chapter, this in turn serves as a basis for quantifying uncertainties for area-averaged rainfall through Monte Carlo simulations, useful in sensitivity studies in hydrological or crop models.

The remainder of this chapter is arranged as follows: Section 4.2 gives an overview of the TAMSAT method. Section 4.3 describes the development of TAMSAT1, as an extension of TAMSAT method for estimation of daily rainfall at native pixel level. In Section 4.4, the daily rainfall during the rainy season in the Gambia is used as a case study to evaluate the performance of TAMSAT1. Section 4.5 further discusses several issues regarding the performance of TAMSAT1, and Section 4.6 concludes the chapter.

## 4.2. TAMSAT method

For regions where rainfall is predominately convective, the TAMSAT method of satellite rainfall estimation provides a simple way of estimating rainfall from satellite derived Cold Cloud Duration (CCD). The CCD of a pixel is defined as the duration within a given time interval, that a pixel temperature (determined from pixel TIR irradiance) is colder than a predefined temperature. Apart from the coastal monsoon areas of West Africa, the Southern region of South Africa, as well as the Mediterranean regions, such CCD estimated rainfall performs well in most parts of Africa where convective rainfall due to the passage of the ITCZ predominates (Thorne *et al.*, 2001). Dekadal rainfall determined from METEOSAT data using the TAMSAT method has been used by regional meteorological services for drought monitoring purpose in Africa (Grimes *et al.*, 1999). The method essentially consists of a linear CCD-rainfall relationship at the scale of the satellite pixel:

$$Z = \begin{cases} a_0 + a_1 \times D_T + e & D_T > 0 \\ 0 & D_T = 0 \end{cases} \quad 4.1$$

where  $D_T$  and  $Z$  are respectively the pixel CCD at a threshold temperature  $T$  and the corresponding pixel-averaged rainfall respectively.  $a_0$  and  $a_1$  are empirical constants.  $e$  represents the stochastic component of the CCD-rainfall model, with  $E[e] = 0$  and  $\text{Var}[e] = \sigma^2$  (a constant).

It is implicit from the model assumption in Equation 4.1 that  $T$  must be selected so that pixels with positive  $D_T$  are coincident with most of the rainy pixels. Selection of this optimal threshold temperature is achieved by constructing  $2 \times 2$  contingency tables for a range of threshold temperature. An example of the contingency table is shown in Table 4.1.

	$D_T = 0$	$D_T > 0$
$Z = 0$	$n_{11}$	$n_{12}$
$Z > 0$	$n_{21}$	$n_{22}$

Table 4.1 Contingency table for selecting optimal CCD threshold temperature  $T$ .

In Table 4.1,  $n_{ij}$  are the number of pixels satisfying the various CCD and rainfall conditions for the  $ij^{\text{th}}$  cell of the contingency table for a given threshold temperature. The off-diagonal terms  $n_{12}$  and  $n_{21}$  are the number of ‘false-alarms’ and ‘mis-detections’. The selection criteria for optimality (Milford and Dudgeon, 1990; Grimes *et al.*, 1999) are

$$\frac{n_{11} + n_{22}}{n_{12} + n_{21}} \gg 1 \quad \text{and} \quad \frac{n_{21}}{n_{12}} \approx 1 \quad 4.2$$

In principle,  $T$ ,  $a_0$ ,  $a_1$  (and  $e$ ) may depend on seasonal, as well as regional rainfall characteristics. In practice, seasonal dependence of these parameters is usually taken into account by separate calibrations for individual months. The method in identifying homogenous regions of similar calibrations or calibration zones, depends on the complexity of the region’s rainfall characteristics; calibration zones for Sahelian Africa are largely invariant for all months within the rainy season, whereas for other regions a

different set of calibration zones is needed for each month (Thorne *et al.*, 2001). Figure 4.1 shows the TAMSAT calibration zones for Northern Africa.

Once  $T$  is obtained from the contingency table analysis (Equation 4.2), linear regression is employed to determine model parameters  $a_0$  and  $a_1$  in Equation 4.1. Due to non-constant error variance (heterocedasticity) and a positive skewness in rainfall distribution, various robust regression methods have been used to estimate the model parameters. Milford and Dugdale (1990) suggested countering the influence of skewness in rainfall distribution by grouping CCD into various classes and using weighted regression to fit the median of the observed rainfall for each CCD class. Grimes *et al.* (1999) account for heterocedasticity by proposing a power model for the error variance:

$$\sigma^2 = \kappa[E(Z)]^\theta \quad 4.3$$

where  $\kappa$  and  $\theta$  are constants to be determined. The model parameters in Equations 4.1 under a variance model of Equation 4.3 can then be determined using iterative weighted least square methods of Carroll and Ruppert (1988).

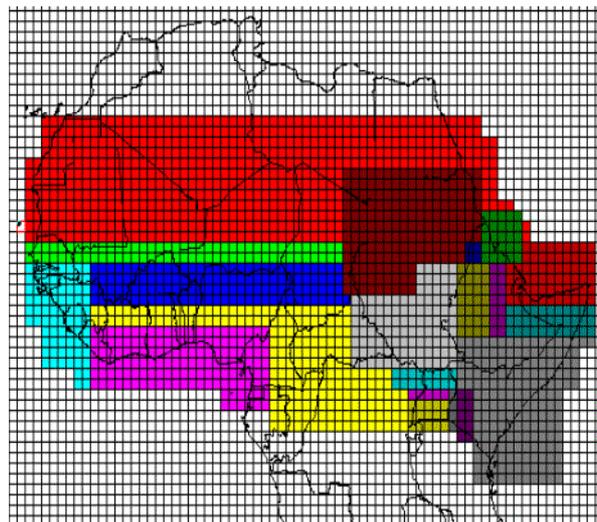


Figure 4.1 TAMSAT calibration zones for Northern Africa for August. (Courtesy of TAMSAT)

### 4.3 TAMSAT1: Extension of TAMSAT

By design, TAMSAT does not require contemporaneous rainfall observations for operation. Hence it is suitable in for operational applications requiring real-time data in regions where ground observations are not available in real-time (Grimes *et al.*, 1999). Thorne *et al.*, (2001) found that TAMSAT performed relatively better than CPC rainfall estimates (Herman *et al.*, 1997) over flat, arid regions where rainfall gauge data is sparse or non-existent, while the latter were better over regions with better gauge network. In addition, once calibrated, the simple algorithm has the advantage of requiring only the minimum of computational resources.

The original TAMSAT algorithm has been extensively tested for dekadal rainfall. For the application of rainfall estimates to hydrological and crop modelling, daily rainfall amounts are often required. Two issues arise from TAMSAT when used for daily rainfall estimation:

- 1) Delineation of rainy areas using any available CCD temperature is not optimal, due to the significant number of mis-detections and false alarms.
- 2) Providing only the estimated mean rainfall as well as a standard deviation of the estimation is not sufficient in representing the rainfall due to the highly non-Gaussian distribution of daily rainfall.

In the following subsections, an alternative method of estimating daily rainfall using satellite RFE is proposed. The proposed method aims to 1) retain the simplicity of TAMSAT– using CCD at a single temperature threshold as the only predictant and using linear relationships whenever possible, 2) to provide a description for mis-detections and false-alarms, and 3) to provide a framework so as to be able to compute the uncertainties of rainfall estimates upon spatial aggregation.

**Box 1: Notation of key parameters of TAMSAT1**

$a_0$	Positive rainfall amount at zero CCD
$a_1$	Increment of rainfall amount per unit increase of CCD
$b_0$	Right limit of log-odds of pixel being rainy at zero CCD
$b_1$	Increment of log-odds of pixel being rainy per unit increase of CCD for non-zero CCD
$D_T$	CCD of pixel at threshold temperature $T$
$g(a,b)$	Gamma distribution with mean $a$ and standard deviation $b$
$p$	Probability of pixel being rainy given pixel CCD
$p_0$	Probability of pixel of zero CCD being rainy
$\kappa$	Scale factor of the variance ( $\sigma_+$ ) model for positive rainfall retrieval
$\kappa_g$	Scale factor for the gauge-estimated positive rainfall variance ( $\sigma_g$ ) model
$\mu_+$	Mean rainfall amount if pixel is rainy given pixel CCD
$\theta$	Shape factor of the variance ( $\sigma_+$ ) model for positive rainfall retrieval
$\theta_g$	Shape factor of the gauge-estimated variance ( $\sigma_g$ ) model
$\sigma_+$	Standard deviation of the satellite estimated positive rainfall for given CCD
$\sigma_R$	Standard deviation of the regression of gauge-estimated positive rainfall against CCD
$\sigma_g$	Standard deviation of the gauge-estimated positive rainfall for a given CCD

**Box 2: Key Equations of the TAMSAT1****Pixel rainfall model given CCD**

Probability distribution function of positive rainfall given pixel CCD      Equation 4.5 b-d

Probability of pixel being rainy given pixel CCD      Equation 4.5a

**Accounting estimation error in gauge-rainfall estimates in modelled variance**

Relation between the regression variance, estimation variance and  
variance in rainfall      Equation 4.8

Variance model for gauge-estimated pixel rainfall      Equation 4.9

### 4.3.1 Modification to TAMSAT for daily rainfall estimation

Modification of TAMSAT to daily time scale proceeds with the aim of capturing features of daily rainfall-CCD relationship not explicitly modelled by Equation 4.1, while striving to retain the simplicity of the original TAMSAT methodology (linear regression model with CCD as only predictant). Features of mis-detections and false-alarms for daily rainfall is best illustrated using Gambian daily pixel rainfall-CCD pairs for June and September as an example. These two months are chosen for their different rainfall characteristics; June being the beginning of the rainy season with lower than average daily rainfall, while average rainfall peaks in September. The pixel rainfall used was estimated from raingauge observations as discussed in Chapter 3. Only pixels collocated with a rain gauge are selected. The optimal temperature following TAMSAT method is found to be  $-50^{\circ}\text{C}$  for both June and September (see section 4.4 for details).

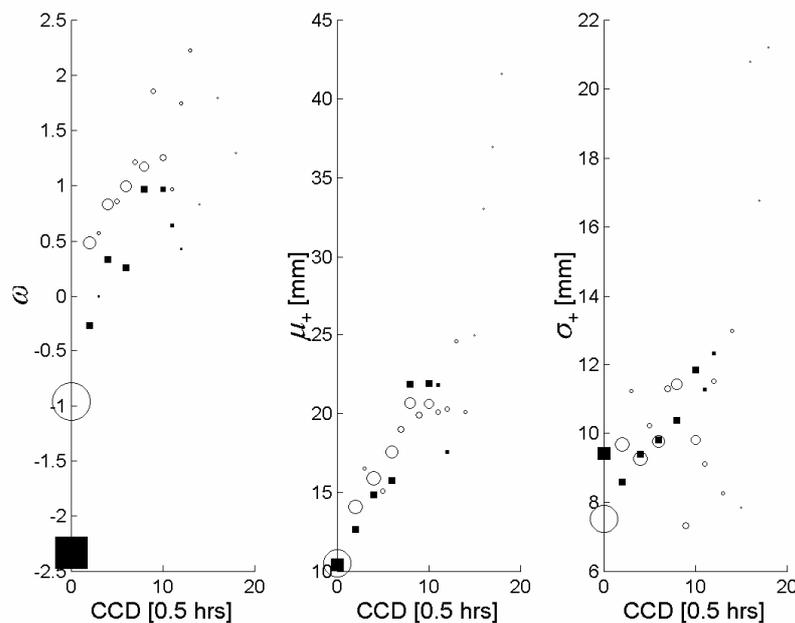


Figure 4.2 Scatter plots of sample estimates of  $\hat{\omega}$  (left),  $\hat{\mu}_+$  (center),  $\hat{\sigma}_+$  (right) versus pixel CCD for the Gambia for June (square) and September (circle). The size of markers is proportional to the number of data used to compute the samples estimates. Data used are from 1988 to 2002.

Figure 4.2 shows the sample estimated log-odds for pixel being rainy  $\hat{\omega} = \log\left(\frac{\hat{p}}{1-\hat{p}}\right)$ , where  $\hat{p}$  is the sample estimate of the probability of a pixel being rainy, the mean positive rainfall  $\hat{\mu}_+$  and the standard deviation of positive rain  $\hat{\sigma}_+$ , for different CCD values for both months.  $\hat{\omega}$  is used instead of  $\hat{p}$  in the above plot, since the  $p$ -CCD relationship is suspected to be logistic in form, in which case,  $\omega$ -CCD has a linear relationship.

From these scatter plots, several features are noted: 1) The presence of a significant positive rainfall for zero CCD is evident for both months. 2) Linearity in the  $\hat{\mu}_+$ -CCD relationship, particularly for  $\text{CCD} \leq 10$  hours. 3)  $\hat{\omega}$  at  $\text{CCD} = 0$  tends to be significantly lower than those for  $\text{CCD} > 0$ . This is due to the selection of optimal threshold temperature using the TAMSAT method. 4) Arguably, a linear  $\hat{\omega}$ -CCD relationship for  $\text{CCD} > 0$ . 5)  $\hat{\sigma}_+$  increases with CCD values for both months.

A statistical model that fits such observations is a mixed distribution model where 1) the occurrence of rain at a pixel is a Bernoulli process, with probability of a pixel being rainy determined by the pixel CCD, and 2) a non-negative, continuous probability density function (pdf) fully parameterised by the pixel CCD, describing the observed distribution of rainfall amount for a given CCD value under the condition of rain occurring. A possible modification of TAMSAT to include the mixed distribution of rainfall is to 1) quantify the probability that a pixel being rainy through an explicit  $\omega$ -CCD relation, and 2) generalizing the linear CCD-rainfall model of Equation 4.1 to a pdf of rainfall amount, with mean and variance as functions of CCD:

$$\begin{aligned} P(Z > 0 | \text{CCD} = D_T) &= \psi(D_T) \\ f(Z | Z > 0, \text{CCD} = D_T) &= g(\alpha(D_T), \beta(D_T)) \end{aligned} \tag{4.4}$$

In Equation 4.4,  $\psi$  is assumed to be a monotonic increasing function of  $D_T$ ,  $f$  is the pdf of positive rainfall amount conditioned to the pixel CCD, and  $g$  is the Gamma Distribution

with two parameters  $\alpha$  and  $\beta$ . The model for the conditional probability of rain,  $P(Z>0|CCD = D_T)$  follows immediately from the observations above. The assumption that  $f$  can be modelled by a two parameter Gamma Distribution can be demonstrated with sample rainfall distribution for various CCDs using the Gambian rainfall/CCD data pairs for the various rainy months (June to Oct). Figure 4.3 shows examples of a fitted gamma distribution compared with the sample rainfall distribution for various CCD values for August. Since it is well known that  $g$  is fully defined if  $\mu_+$  and  $\sigma_+$  are known, the second equation of Equation 4.4 means that dependence of  $f$  on CCD can be defined through  $\mu_+$  and  $\sigma_+$ .

To complete the model description, three more assumptions are made: (1)  $\psi$  follows a logistic model for  $CCD>0$ , (2)  $\mu_+$  is a linear function of CCD, and (3) heterocedasticity of positive rain is described by a power model as in Equation 4.3. The complete model of the daily pixel CCD rainfall model is described by the following equations:

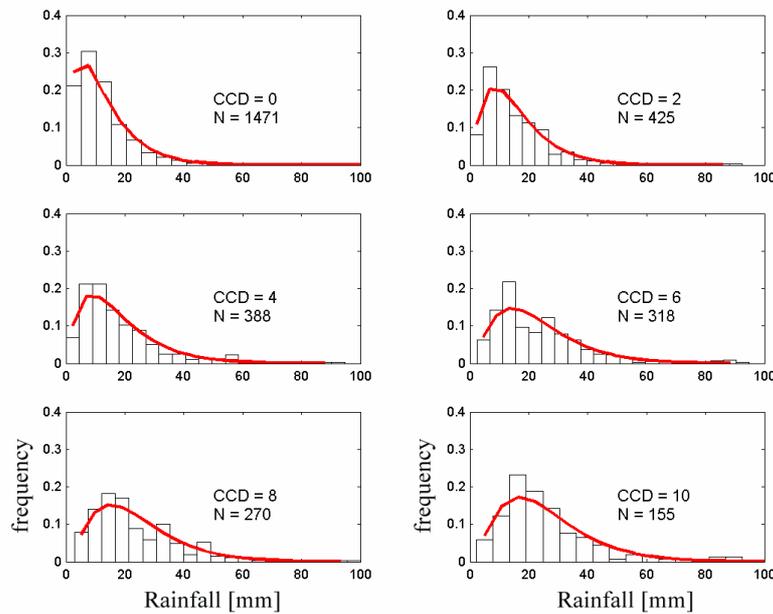


Figure 4.3 Frequency histogram of daily positive rainfall conditional to pixel CCD values for the month of August from Gambian rainfall and CCD data. Solid line is the corresponding fitted gamma distribution for each CCD value.  $N$  is the number of data.

$$P(Z > 0 | CCD = D_T) = p = \begin{cases} \frac{1}{1 + e^{-(b_0 + b_1 \times D_T)}} & D_T > 0 \\ p_0 & D_T = 0 \end{cases} \quad 4.5a$$

$$f(Z | Z > 0, CCD = D_T) = g(\mu_+(D_T), \sigma_+(\mu_+)) \quad 4.5b$$

$$\mu_+ = a_0 + a_1 \times D_T \quad 4.5c$$

$$\sigma_+ = \kappa \mu_+^\theta \quad 4.5d$$

where  $a_0$ ,  $a_1$ ,  $b_0$ ,  $b_1$ ,  $p_0$ ,  $\kappa$  and  $\theta$  are empirical constants to be determined. This rainfall model is denoted as TAMSAT1 hereafter. In essence, TAMSAT1 relegates TAMSAT to the description of the first and second order statistics of the positive rainfall amount and at the same time, removes the constraint of a deterministic delineation of rainfall for  $CCD = 0$ .

This section concludes with the following remarks:

1) Although the logistic model for  $CCD > 0$  in Equation 4.4a is non-linear for nonzero  $CCD$ , TAMSAT1 can be considered a “linear” model, since the logistic regression model obtained through logarithmic transformation of Equation 4.4a for  $CCD > 0$  is linear:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 \times D_T \quad 4.5e$$

Logistic regression has also been applied by Chiu and Kendem (1990) in estimating the probability of rain rate using MW radiometric measurements from the Nimbus 7 during the GATE experiment.

2) Equations 4.1 and 4.3 are a special case of Equation 4.5 with  $b_1 \gg 1$ ,  $b_0 = 0$  and  $p_0 = 0$ . Hence TAMSAT1 is a consistent generalisation of TAMSAT.

#### 4.4 Case study

TAMSAT1 was calibrated for Gambian daily rainfall for the rainy season, June to October, and its performance was compared with daily rainfall estimates from TAMSAT. Daily gauge estimated pixel rainfall (Chapter 3) for the Gambia and its corresponding pixel CCD from 1988 to 2002 were used in both calibration and subsequent validations. The available CCD images for the entire period have threshold temperatures from  $-70^{\circ}\text{C}$  to  $-30^{\circ}\text{C}$  at an increment of  $10^{\circ}\text{C}$ . Although pixel rainfall estimates from rain gauge data were available over the entire country (Chapter 3), rainfall at pixels containing a gauge had the highest accuracy. Therefore only these pixels are used in the calibration and validations.

Calibration methodology adopted in this case study follows those of TAMSAT as described in Grimes *et al.* (1999); CCD threshold temperatures and the set of unknown model parameters in Equations 4.5 were determined separately for individual months from June to October to account for variability within the rainy seasons. These are discussed in section 4.4.1. Skill comparison between TAMSAT1 and TAMSAT for the case of Gambian rainfall is discussed in Section 4.4.2.

For clarity, the following notational rules are adopted: Estimates of model parameters  $a_0$ ,  $a_1$ ,  $b_0$ ,  $b_1$ ,  $p_0$ ,  $\kappa$  and  $\theta$  for TAMSAT1 are denoted with an asterisk. Similarly, estimates of  $\mu_+$ ,  $\sigma_+$  and  $p$  of TAMSAT1 through the fitted model parameters are denoted with an asterisk, while the ‘hatted’ symbols denote sample estimates. For example,

$\mu_+^* = a_0^* + a_1^* \times D_T$ , while  $\hat{\mu}_+ = \sum_{j=1}^N z_j / N$ , where  $\{z_j | j = 1, \dots, N\}$  is the set of  $N$  observed

pixel rainfall values from gauge data with corresponding pixel CCD =  $D_T$ . The entire set of CCD-rainfall observations used for calibration is denoted by  $C$ .  $C^m$  (a subset of  $C$ ) represents the collection of the observations for the month  $m$ . The set of CCD-rainfall data with positive rainfall is  $C_+$ .  $C_+^m$  represents the subset of  $C_+$ , with elements from month  $m$ .

### 4.4.1 Calibration

#### Selection of optimal CCD threshold temperature

Monthly contingency tables (Section 4.1) for the available threshold temperature for the Gambia were constructed. The optimal  $T$  for each month was selected using the criterions listed in Equation 4.2. The optimal temperature determined were  $-60^{\circ}\text{C}$  for all months except August, where  $T = -50^{\circ}\text{C}$  was found to be more appropriate. As the spatial extent of Gambia is relatively small ( $\sim 300$  km East-West), no additional calibration zoning was performed. This range of optimal threshold temperatures for delineating daily rainfall is lower than those reported for the rainy seasons in Zambia ( $-50^{\circ}\text{C}$  to  $-30^{\circ}\text{C}$ ) and Mali ( $-40^{\circ}\text{C}$ ) (Grimes and Diop, 2003; Grimes *et al.*, 2003), suggesting that the Gambia is within a different climatic zone, possibly due to its proximity to the coast.

#### Positive rain model parameters estimation

In order to account for heteroscedasticity and non-normality in the data as discussed in Section 4.3.1, the iterative weighted least square (IWLS) of Carroll and Ruppert (1988) is used to estimate the parameters  $a_0$ ,  $a_1$ ,  $\kappa$  and  $\theta$  in Equations 4.5c and 4.5d for each month. The method is outlined as follows: In matrix formulation, the weighted least solution of  $A^* = [a_0^* \ a_1^*]^T$  is given by

$$A^* = (X^T W X)^{-1} X^T W Y \quad 4.6$$

where  $X$  is a  $2 \times n$  matrix with  $X_{i1} = 1$  and  $X_{i2} = d_i$ ,  $Y$  is a  $n \times 1$  vector with  $Y_i = z_i$  and  $(d_i, z_i)$  is the pixel CCD and gauge estimated rainfall pair ( $\in C_+^m$ ). Weight matrix  $W$ , is a  $n \times n$  diagonal matrix with  $W_{ii} = 1/\sigma_{+i}^2$ , where  $\sigma_{+i}$  is given by Equation 4.5d. Since  $\sigma_{+i}$  is unknown, a variance modelling approach approximates the weight by providing estimates  $\sigma_{+i}^*$  for  $\sigma_{+i}$ . Using the logarithm transform of Equation 4.5d,

$$\log \sigma_+ = \log \kappa + \theta \log \mu_+ \quad 4.7$$

Equation 4.7 is used to estimate model parameters for the variance model by Ordinary Least Square. In the regression, the absolute residuals,  $r^* = |z - \mu_+^*|$  are used as a proxy for the predicted response,  $\sigma_+$ . To avoid the problem of asymptotic infilling, where the logarithm of small  $r^*$  may have undue influence on the estimates using Equation 4.7,  $r^*$  less than a certain amount are not included in the regression. Analysis of Cook distances (Cook and Weisberg, 1982) of  $\log r^*$  for  $C_+$  suggested that about 75% of possible infill asymptotic can be attributed for  $r^* < 1$  mm alone. Thus, this value is adopted as the cutoff for all the months in this study.  $a_0^*$ ,  $a_1^*$ ,  $\kappa^*$  and  $\theta^*$  are obtained from Equations 4.6 and 4.7 iteratively starting with an OLS estimates for  $a_0^*$  and  $a_1^*$ . The iteration terminates when the changes in  $a_0^*$  and  $a_1^*$  are less than a prescribed tolerance.

	$a_0^*$ [mm]	$a_1^*$ [mm/hr]	$\kappa^*$ [mm]	$\theta^*$	$\kappa_g^*$ [mm]	$\theta_g^*$
June	10.72	0.95	3.85	0.36	1.65	0.45
July	12.70	1.40	2.24	0.61	2.33	0.38
Aug	10.72	1.30	1.94	0.66	1.67	0.51
Sep	10.66	1.17	1.98	0.57	1.81	0.43
Oct	9.45	1.29	1.84	0.64	1.22	0.57

Table 4.2 Estimates of the positive rainfall model parameters for Gambia

Estimated parameters for the various months are listed in Table 4.2 and model fit for both the mean response and the variance are shown in Figures 4.4 and 4.5 respectively. The fitted models for all months compare well with the sample means for CCD values with large number of observations. For the months from July to September where rainfall is climatologically most significant, the fitted models are satisfactory for CCD value of less than 8 hours. For the larger CCD values, the model tends not to do as well. For September, consistent under-prediction of mean rainfall by the fitted model suggested that a non-linear model may be more appropriate, while no discernible trend can be seen to discredit the current model for July and August. Such non-linearity may arguably be suggested for June from observing the data plot. However, no significant advantage may be gained from fitting a more sophisticated model for the following reasons:

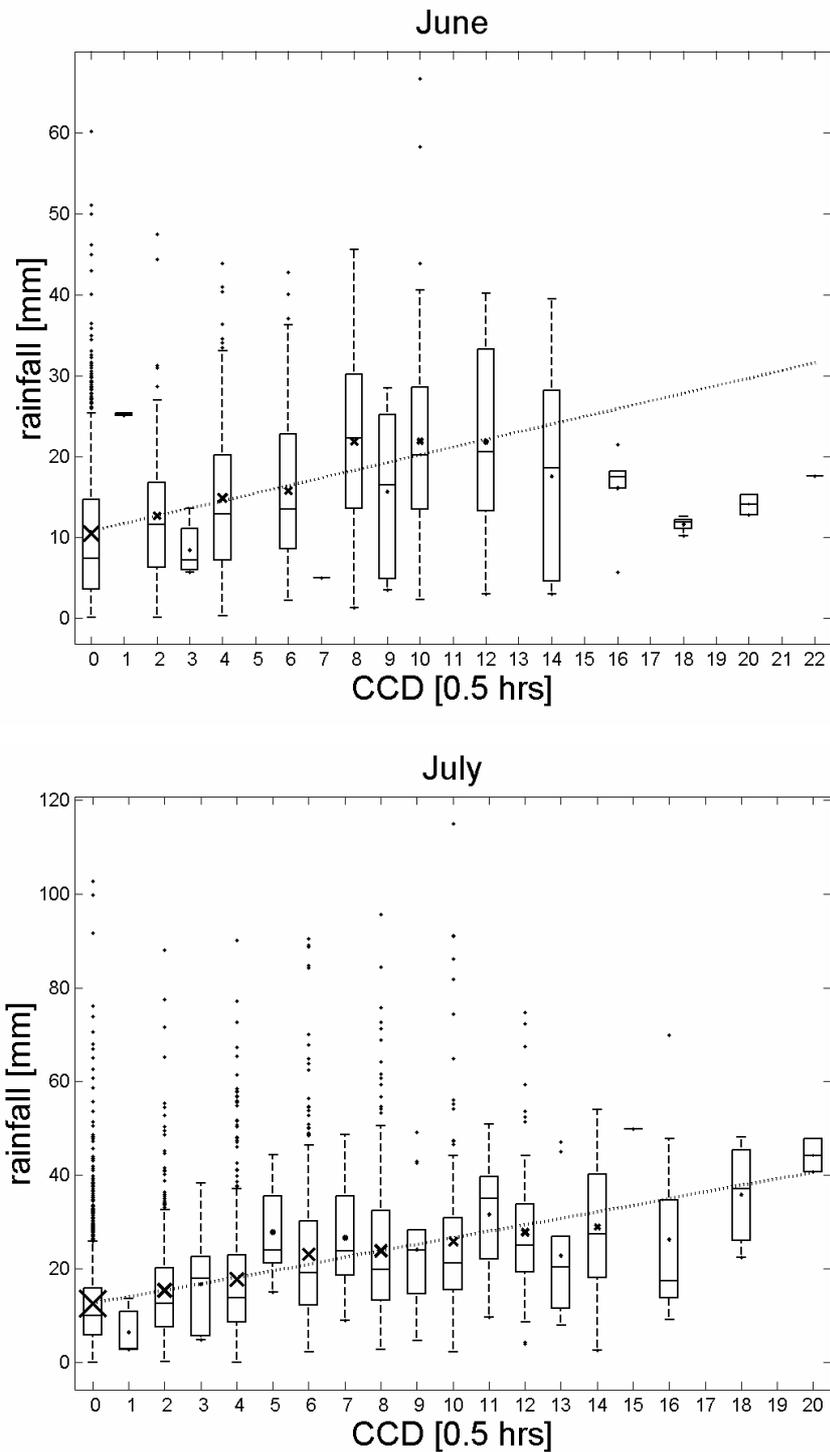


Figure 4.4 Boxplots of daily pixel rainfall versus CCD and fitted model for various months: 'x' – sample mean for each CCD bin. Size of marker is proportional to the number of data available for each CCD bin; dotted line: regression fit.

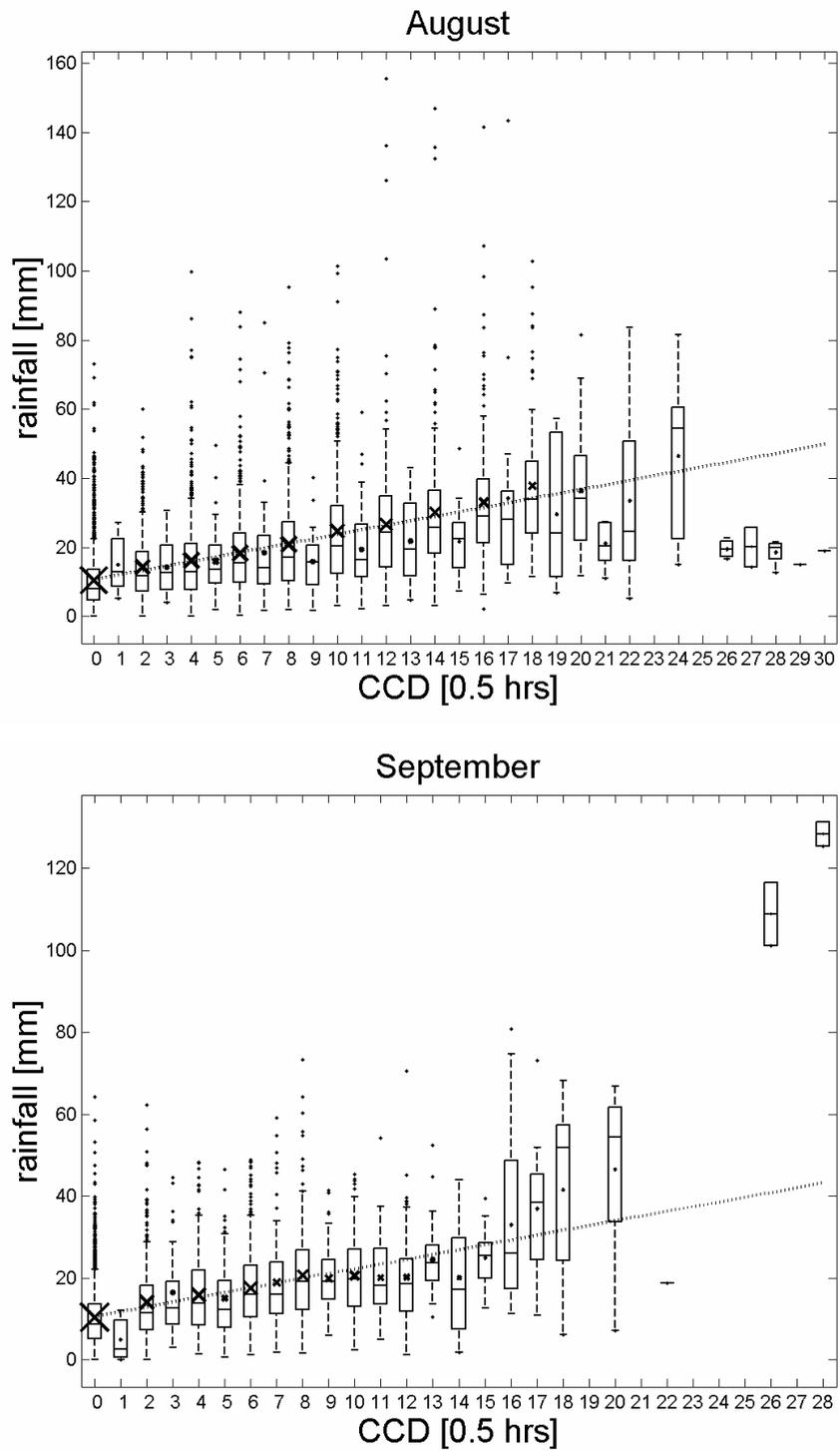


Figure 4.4 (Continue)

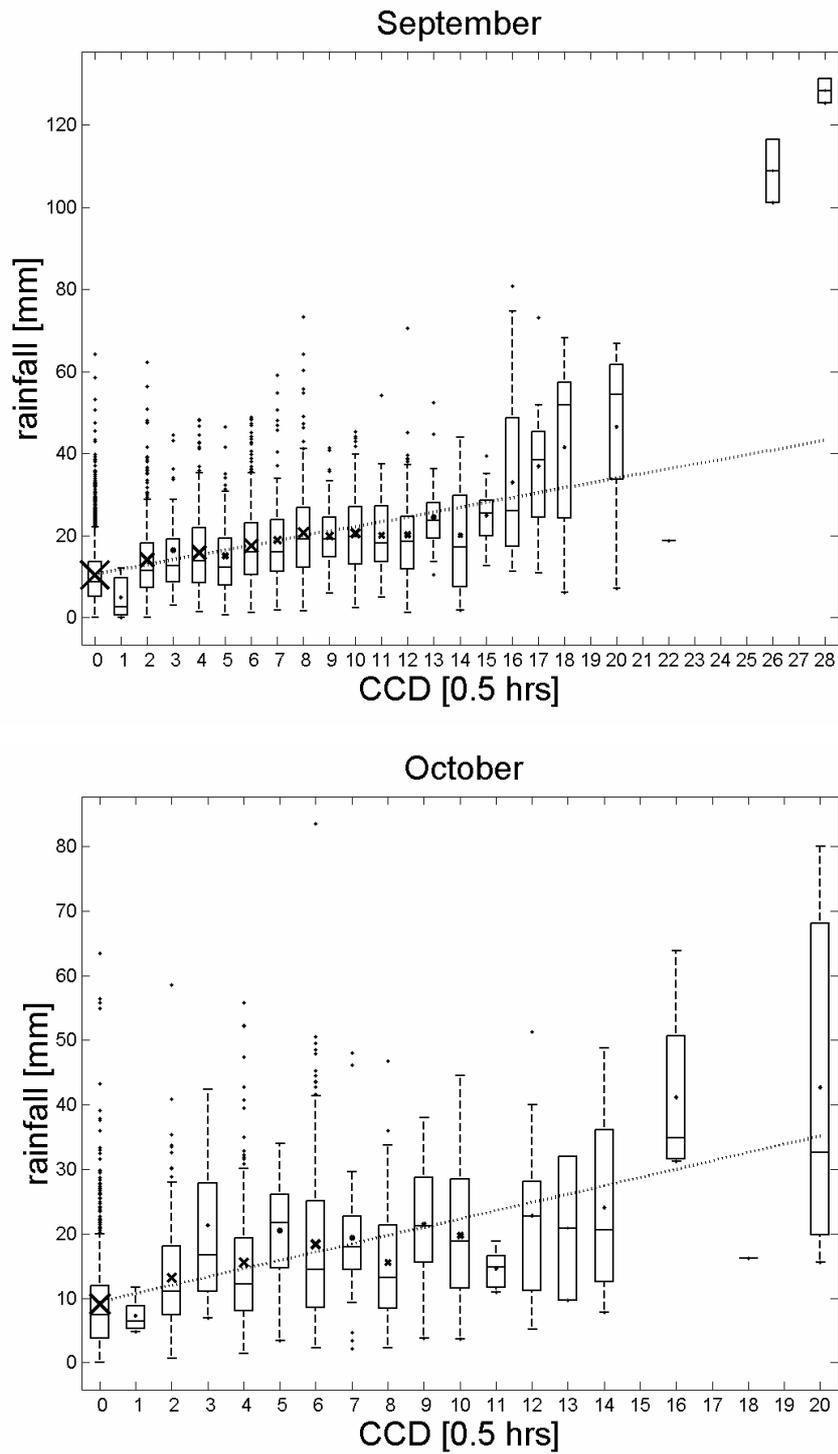


Figure 4.4 (Continue)

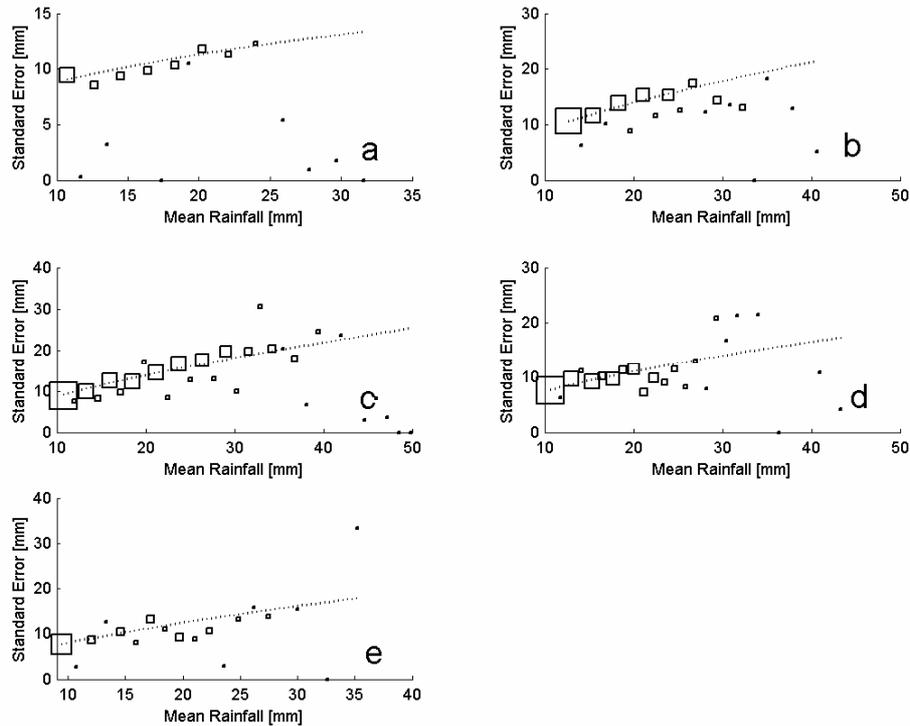


Figure 4.5 Scatter plots of sample standard error (open square) and modelled standard error of rainfall (dotted line) versus mean rainfall for various months. a) June; b) July; c) August; d) September; e) October. The size of the square marker is proportional to the sample size.

1) Most of the CCD values encountered falls within the region where the linear model works best: Around 90% of CCD values encountered are less than 5 hours for all months except August (< 7 hrs). 2) The low number of observations at the high CCD values needs to be interpreted with care. Over-fitting of the rainfall model with a more sophisticated model only complicates the current model without any practical improvement. It is noted however that the misfit of the linear model at September warrants further investigation.

As the observations for some CCD bins are large, the sample standard error provides good estimates for the standard error of the rainfall distribution at these CCD bins. From Figure 4.5, it can be seen that for all months, the fitted standard error agrees well with the sample standard errors for CCD when the number of observations are larger.

Accounting estimation error in gauge rainfall estimates in modelled variance

As there are no actual observations of pixel-averaged rainfall in reality, the mean pixel rainfall within the calibration sets,  $C_+^m$ , were estimated from the gauge rainfall observations, and therefore, possess uncertainties. For each of these gauge-based estimates,  $z_i$ , the estimated standard error,  $s_i$ , is available (see Chapter 3). The effect of such uncertainties is to cause an overestimation of the stochastic component in the CCD-rainfall relationship in the following way: For a given CCD =  $D_T$ , the rainfall amount given by Equations 4.5b-d can be rewritten as  $Z(D_T) = \mu_+(D_T) + e(\mu_+)$ , where  $e$  reflects the stochastic nature of the CCD-rainfall relation, with  $E(e) = 0$  and  $Var(e) = \sigma_+^2$ .

In the present case, with no actual  $Z$  samples available, a ‘noise’ due to the usage of estimates of  $Z$  needs to be included in the above relation:  $Z(D_T) = \mu_+(D_T) + e(\mu_+) + e_g$ , where  $e_g$  is the error from the gauge-based estimates. The preceding relationship is reasonable since for an ideal model, where CCD-rainfall relation is deterministic,  $e = 0$ . In this case,  $Z(D_T) = \mu_+(D_T) + e_g$ . Since  $e$  and  $e_g$  are independent,  $Var(Z) = Var(e(\mu_+) + e_g) = \sigma_+^2 + \sigma_g^2$ , where  $Var(e_g) = \sigma_g^2$ . Hence the regression variance  $\sigma_R^2$ , which estimates the total variance of  $Z$ , is

$$\sigma_R^2 = \sigma_+^2 + \sigma_g^2 \quad 4.8$$

Accounting for uncertainties in the gauge rainfall estimates in obtaining  $\sigma_+^2$  through Equation 4.8 has been attempted by Grimes *et al.* (1999). As in their work,  $\sigma_g$  in this present study is assumed to have power dependence on mean rainfall:

$$\sigma_g = \kappa_g \mu_+^{\theta_g} \quad 4.9$$

Justification of functional dependence on the mean rainfall can be seen from Figure 4.6, where the moving average of the sample standard error estimates,  $s_i$ , increases with the predicted mean rainfall. A nonlinear least square regression is used for estimating  $\kappa_g$  and  $\theta_g$ , and their model fitted agrees well when compared with the sample mean of the  $s_i$  for large sample size (Figure 4.6).

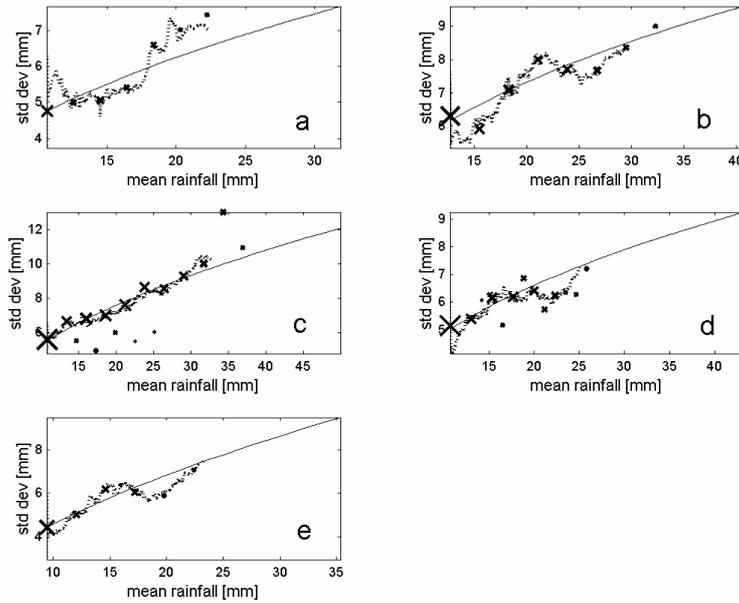


Figure 4.6 Plots of standard deviation ( $s$ ) versus mean ( $\mu_+$ ) of positive rainfall for various months. Legend: 'x': sample mean of  $s$  for  $\mu_+$  with samples number  $> 30$ . The size of marker is proportional to the number of data. Dotted line: Running mean of  $s$  versus running mean of  $\mu_+$  with a window of 250 samples. Solid line: Fitted model for  $\sigma_g$ . a) June; b) July; c) August; d) September; e) October.

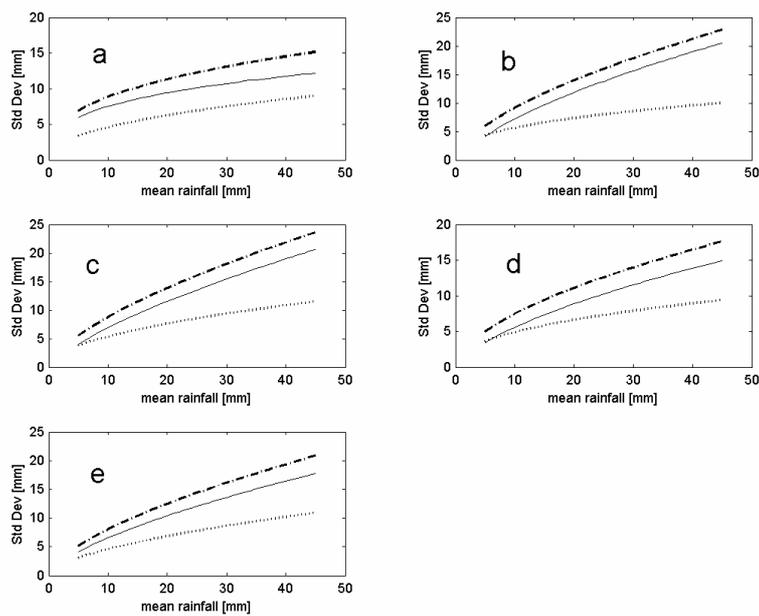


Figure 4.7 Modelled standard error for regression (dot-dash line), gauge estimates (dotted line) and the adjusted  $\sigma_+$  (solid line). a) June; b) July; c) August; d) September; e) October.

The fitted values for  $\kappa_g$  and  $\theta_g$  are given in Table 4.2. Once  $\sigma_g$  is determined,  $\sigma_+$  can then be computed from Equation 4.8, with  $\sigma_R$  taking the role of unadjusted  $\sigma_+$  determined from Equation 4.7. Figure 4.7 shows the adjusted  $\sigma_+$  compared with the model regression standard error and standard error of the gauge rainfall estimates.

#### Model parameter estimation for probability of rainy pixels

Estimates of  $b_0$  and  $b_1$  of Equation 4.5a are determined from Equation 4.5e using logistic regression. Indicator transform  $\{z, \text{CCD}\} \rightarrow \{i, \text{CCD}\}$  is performed for each  $\{z, D\} \in C^m$ , where  $i$  is 1 if  $z > 0$ , and 0 if  $z = 0$ . For a given  $\text{CCD} = D$ , sample probability of rain,  $\hat{p}_D$  is computed as  $\sum_{j=1}^{N_D} i_j / N_D$ , where  $N_D$  is the number of observations with  $\text{CCD} = D$  and the set  $\{i_j | j = 1, \dots, N_D\}$  is the collection of corresponding indicator values.  $p_0$  of Equation 4.5a, was estimated by  $\hat{p}_0$ . Table 4.3 shows the fitted results of  $p_0$ ,  $b_0$  and  $b_1$ , as well as some measure of model performance from the logistic regression for  $\text{CCD} > 0$ . Model Chi Squares indicates improvement (for  $p$  at  $\text{CCD} > 0$ ) from a constant probability model for all months. However, the deviance for all months suggests that the  $\text{CCD}-p$  model for non-zero  $\text{CCD}$  did not fit as well as initially envisaged. An alternative model using weighted least square cubic spline fit (smoothing parameter = 0.8) in place of the logistic model for  $\text{CCD} > 0$  was computed for comparison. The coefficient of determination  $R^2$ , for the proposed and alternative model (all months combined) is 0.51 and 0.57 respectively. Though the alternative model performed better in terms  $R^2$ , visual comparison of the two models with observations (Figure 4.8) suggested that simplicity of the original model outweighs the slight improvement of fit by the more sophisticated model.

	$p_0^*$	$b_0^*$	$b_1^*$	Model Chi Square	Deviance at solution
June	0.09	-0.29	0.11	37.95	35.79 (14)
July	0.22	0.40	0.09	44.45	32.66 (16)
August	0.29	0.32	0.09	153.34	55.83 (26)
September	0.28	0.38	0.10	62.66	30.17 (20)
October	0.10	-0.10	0.10	28.51	8.12 (15)

Table 4.3: Fitted  $p_0$ ,  $b_0$ ,  $b_1$  for the various months as well as Model Chi Square and Deviance for fitted logistic model for  $p$  when  $\text{CCD} > 0$ . Bracketed values beside the deviance are the degree of freedom.

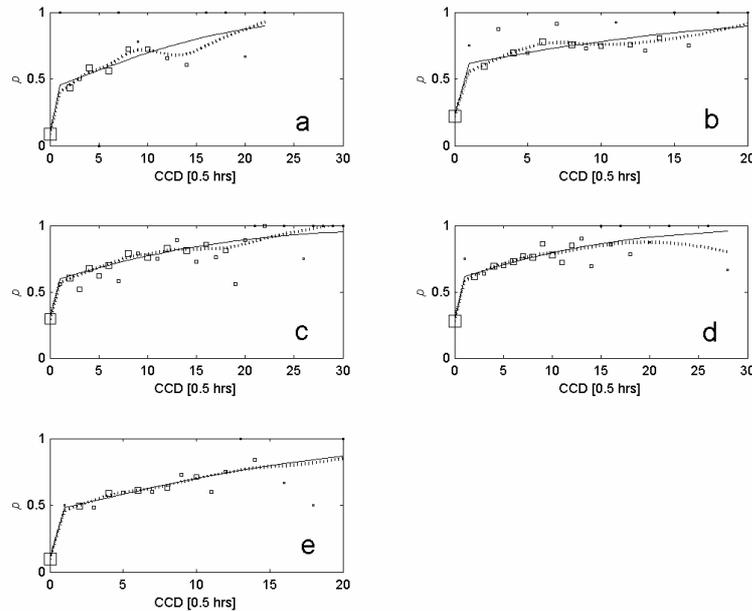


Figure 4.8 Scatter plots of the observed rainfall occurrence frequencies as well as the model of rainfall probability as a function of CCD. Solid line: Proposed model; Dotted line: Alternative model using a cubic smoothing spline when  $CCD > 0$ .

#### 4.4.2 Validation

Validation of TAMSAT1 has been carried out using the Gambian CCD and rainfall dataset described in the preceding section. The validation is carried out with two broad motivations. First, skill comparison with TAMSAT algorithm (Section 4.2) is made in order to gain an appreciation of its overall performance with the expectation that as an extension of TAMSAT algorithm, TAMSAT1 should perform at least as well with TAMSAT statistically. Second, one reason for developing TAMSAT1, as discussed in Section 4.3, is to enable CCD-rainfall estimates to be included as input data for downstream applications, like crop or hydrological modelling experiments, or operational systems setup in a stochastic framework. The ability of TAMSAT1 to capture features crucial for such purposes, namely, the probability of rain and distribution of rainfall amount at pixel level need to be demonstrated.

Comparing TAMSAT1 and TAMSAT

The relative performance of TAMSAT1 and TAMSAT was assessed using daily pixel and gauge rainfall estimates, for the case of Gambia rainfall from 1988 to 2002. The skill scores for quantitative assessments include the bias (BIAS), the root-mean-square error (RMSE), and the coefficient of determination ( $R^2$ ). These quantities are defined below:

$$BIAS = \frac{\sum_{i=1}^N (z_i - z_i^*)}{N} \quad 4.10a$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (z_i - z_i^*)^2}{N}} \quad 4.10b$$

$$R^2 = \frac{\sum_{i=1}^N (z_i^* - \bar{z})^2}{\sum_{i=1}^N (z_i - \bar{z})^2} \quad 4.10c$$

where  $N$  is the number of validation data available,  $z_i$  is the gauge-based rainfall estimate,  $z_i^*$  is the CCD rainfall estimate, and  $\bar{z}$  is the sample mean of the gauge-based rainfall estimates.

To assess the skill for both RFE methods in capturing the proportion of rainy pixels, the probability of a pixel being rainy given by TAMSAT and TAMSAT1 were computed as follows:

For a given set of  $N$  observations of CCD,  $\{D_i \mid i = 1, \dots, N\}$ , the predicted probability of pixels being rainy,  $p^*$ , in TAMSAT algorithm is simply the proportion of non-zero CCD to the number of CCD observations. For TAMSAT1,  $p^*$  is computed as

$$p^* = \frac{\sum_{i=1}^N p^{T1}(D_i)}{N} \quad 4.11$$

where  $p^{T1}(D_i)$  denotes the estimated probability of pixel with  $CCD = D_i$  being rainy from TAMSAT1. Once  $p^*$  is computed for both TAMSAT and TAMSAT1,  $p - p^*$ , where  $p$  is the observed proportion of rainy pixels, gives the bias of the estimator  $p^*$ .

A cross-validation was adopted: rainfall/CCD observations for each year in turn are left out in calibrating the rainfall models. The calibrated model parameters for both TAMSAT and TAMSAT1 for the cross-validation are summarised in Table 4.4. The estimates for  $a_0$  for TAMSAT1 were found to be approximately 2 to 3 times that of TAMSAT. This is not surprising since  $a_0$  in TAMSAT1 is by definition the mean rainfall amount when a pixel with zero CCD is rainy; a more relevant comparison with  $a_0$  of TAMSAT is  $p_0 \times a_0$  of TAMSAT1.

		TAMSAT1									TAMSAT	
		$a_0^*$	$a_1^*$	$p_0^*$	$b_0^*$	$b_1^*$	$\kappa^*$	$\theta^*$	$\kappa_K^*$	$\theta_K^*$	$a_0^*$	$a_1^*$
		[mm]	[mm/0.5hr]			[0.5hr <sup>-1</sup> ]	[mm]		[mm]		[mm]	[mm/0.5hr]
Jun	min	10.23	0.84	0.08	-0.39	0.10	10.16	0.34	2.17	0.30	3.67	0.91
	max	11.07	1.19	0.09	-0.21	0.12	16.82	0.43	5.75	0.49	4.68	1.18
Jul	min	12.21	1.32	0.21	0.35	0.08	4.11	0.58	3.78	0.36	6.84	1.20
	max	12.94	1.45	0.23	0.45	0.10	5.93	0.65	6.34	0.44	7.92	1.41
Aug	min	10.45	1.23	0.28	0.26	0.08	2.99	0.64	2.29	0.49	5.35	1.17
	max	10.88	1.34	0.30	0.39	0.10	4.02	0.70	3.11	0.54	6.19	1.35
Sep	min	10.40	1.11	0.26	0.29	0.09	2.70	0.55	2.62	0.39	5.82	1.06
	max	10.92	1.25	0.28	0.47	0.11	4.46	0.65	4.15	0.47	6.60	1.25
Oct	min	8.89	1.23	0.09	-0.35	0.08	2.70	0.56	1.13	0.47	3.87	0.74
	max	9.69	1.33	0.10	-0.05	0.12	4.92	0.69	2.52	0.63	6.16	1.05

Table 4.4 Range of calibrated model parameters for TAMSAT1 and TAMSAT for validation years from 1988 to 2002.

Test statistics defined in Equations 4.10 were computed for both methods for each validation year (1988 - 2002), and their results are tabulated in Table 4.5. To evaluate the statistical significance of the difference in model performance, a Student's  $t$ -test (Pansofsky and Brier, 1965) was used on the individual test statistics, using the paired values for the 15 years cross-validation experiments. The null hypothesis is that there is

no difference in the mean of the test statistic for both TAMSAT and TAMSAT1. The  $t$ -test value was computed as follows:

$$t = \frac{\bar{\Delta}}{s} (N - 1)^{1/2} \quad 4.12$$

where  $N$  is the number of validating years,  $s$  is the standard deviation of the difference in the samples, and  $\bar{\Delta}$  is the sample mean of the difference in the samples.

The overall performance of TAMSAT1 was better than that TAMSAT; the difference in skill measure between the two algorithms was significant ( $p < 0.05$ ) for each of the four skill scores used. TAMSAT1 consistently outperforms TAMSAT in the RMSE and  $R^2$  for all validating years. Reduction of bias of TAMSAT1 is observed for most years, except 1990, 1993, 1997 and 2002. The difference between sample estimates of the probability and that estimated from each model is shown in Table 4.5. Overall skill in estimating  $p$  in TAMSAT1 is better than TAMSAT, with the exceptions of 1990, 1993, 1997 and 2002.

#### Pixel rainfall distribution

Assessment of TAMSAT1's ability in replicating the observed rainfall distribution was done using Monte Carlo simulation with modelled distributions given in Equations 4.5. The algorithm for a single simulation is as follows:

- 1) For the set of pixel-CCD observations of a validating year, obtain the set of unique pixel-CCD,  $D_U(m)$ , for each month  $m$  ( $m = 6, \dots, 10$ ).
- 2) For each month in the validating year, compute the number of observations,  $n(d; m)$ , for each CCD value  $d$  in  $D_U(m)$ .
- 3) Determined the probability of rain  $p^*(d; m)$ , the mean positive rainfall  $\mu_+^*(d; m)$ , and the standard error of the positive rainfall  $\sigma_+^*(d; m)$ , from Equations 4.5a, 4.5c and 4.8, for each month  $m$  in the validating year respectively. Data from the validating year is omitted when computing these parameters.

Validation Year	Method	BIAS	RMSE	$R^2$	$\hat{p} - p^*$
1988	TAMSAT1	<b>1.78</b>	<b>12.26</b>	<b>0.35</b>	<b>0.028</b>
	TAMSAT	3.34	12.59	0.31	0.083
1989	TAMSAT1	<b>-0.35</b>	<b>9.28</b>	<b>0.35</b>	<b>0.024</b>
	TAMSAT	1.26	9.56	0.31	0.058
1990	TAMSAT1	-1.16	<b>7.64</b>	<b>0.29</b>	-0.038
	TAMSAT	<b>0.25</b>	7.76	0.27	<b>0.005</b>
1991	TAMSAT1	<b>-0.52</b>	<b>7.16</b>	<b>0.26</b>	<b>-0.026</b>
	TAMSAT	1.20	7.40	0.21	0.074
1992	TAMSAT1	<b>0.09</b>	<b>9.73</b>	<b>0.24</b>	<b>-0.029</b>
	TAMSAT	1.62	10.00	0.20	0.034
1993	TAMSAT1	-1.37	<b>8.50</b>	<b>0.27</b>	-0.052
	TAMSAT	<b>0.15</b>	8.64	0.24	<b>-0.012</b>
1994	TAMSAT1	<b>0.82</b>	<b>10.02</b>	<b>0.28</b>	<b>0.032</b>
	TAMSAT	2.29	10.33	0.24	0.084
1995	TAMSAT1	<b>0.96</b>	<b>9.72</b>	<b>0.20</b>	<b>0.057</b>
	TAMSAT	2.18	10.00	0.16	0.096
1996	TAMSAT1	<b>-0.72</b>	<b>7.47</b>	<b>0.37</b>	<b>-0.023</b>
	TAMSAT	0.82	7.62	0.34	0.050
1997	TAMSAT1	-0.73	<b>8.83</b>	<b>0.30</b>	-0.034
	TAMSAT	<b>0.69</b>	8.92	0.29	<b>0.007</b>
1998	TAMSAT1	<b>0.48</b>	<b>9.23</b>	<b>0.27</b>	<b>-0.002</b>
	TAMSAT	1.92	9.52	0.22	0.055
1999	TAMSAT1	<b>1.90</b>	<b>11.16</b>	<b>0.32</b>	0.096
	TAMSAT	3.16	11.35	0.30	0.096
2000	TAMSAT1	<b>0.42</b>	<b>8.93</b>	<b>0.33</b>	<b>0.023</b>
	TAMSAT	1.74	9.07	0.31	0.045
2001	TAMSAT1	<b>-0.07</b>	<b>7.65</b>	<b>0.38</b>	<b>0.006</b>
	TAMSAT	1.31	7.80	0.35	0.039
2002	TAMSAT1	-1.68	<b>6.96</b>	0.31	-0.065
	TAMSAT	<b>-0.15</b>	6.98	0.31	<b>-0.013</b>
Overall	TAMSAT1	<b>0.00<sup>+</sup></b>	<b>9.00<sup>+</sup></b>	<b>0.31<sup>+</sup></b>	<b>0.000<sup>+</sup></b>
	TAMSAT	1.45	9.20	0.28	0.047

Table 4.5. Skill scores comparison between TAMSAT and TAMSAT1. Numbers are bolded when skill scores are better. Value with superscript '+' denotes that the difference in skill score between TAMSAT and TAMSAT1 is significant at  $p < 0.05$ .

4) For each month  $m$  in the validating year, rainfall amounts (including zero rainfall) for  $n(d; m)$  pixels are sampled for each  $d$  in  $D_U(m)$  as follows:

i) A value  $n^{rain}(d; m)$ , representing the number of pixels with positive rainfall, is first sampled from a Binomial Distribution of  $n(d; m)$  trials, with  $p^*(d; m)$  chances of success (raining).

ii) Next, the rainfall amounts for each of the  $n^{rain}(d; m)$  rainy pixels is sampled from a Gamma Distribution, with mean  $\mu_+^*(d, m)$  and variance  $[\sigma_+^*(d, m)]^2$ . The remaining  $1 - n^{rain}(d; m)$  pixels are assigned as non-rainy.

This method of simulation was applied for each validating year from 1988-2002. The 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 95<sup>th</sup> percentiles of the simulated and corresponding observed rainfall amount are summarised in Figure 4.9. Comparison with various observed percentiles demonstrates TAMSAT1 can reproduce the pixel rainfall distribution given pixel-CCD.

For all years, the 25<sup>th</sup> and 50<sup>th</sup> percentiles are both zero for the observed rainfall distribution, and were well captured by the simulations. The higher percentiles (75<sup>th</sup> and 95<sup>th</sup> percentiles) of the rainfall distributions for each year was also reasonably reproduced by the simulations, though the tail of the observed distributions tends to be underestimated by the simulated distributions. The main reason for discrepancies between the simulated and observed rainfall distribution appears to be the effect of errors in the estimated TAMSAT1 model parameters. This can be demonstrated by comparing the 75<sup>th</sup> and 95<sup>th</sup> percentiles of annual rainfall distributions generated from TAMSAT1, calibrated using only the validating year. The simulated percentiles of these regenerated distributions improved, in particular for the 95<sup>th</sup> percentiles (Figure 4.9). Hence, accurate determination of TAMSAT1 parameters is important in capturing the rainfall distributions.

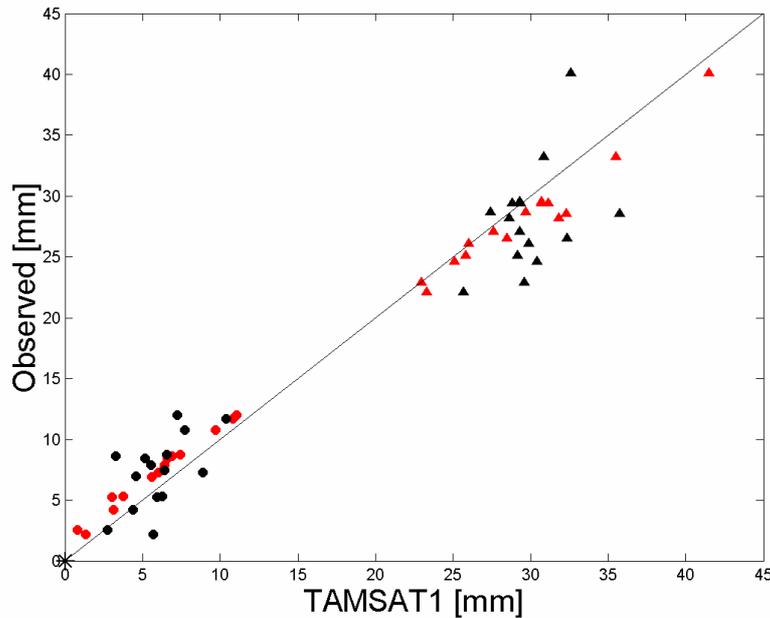


Figure 4.9 Scatterplot of the 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 95<sup>th</sup> percentiles of the observed and simulated rainfall distribution for 1988 to 2002. Legend: '+' : 25<sup>th</sup> percentile; 'x' : 50<sup>th</sup> percentile; '•' : 75<sup>th</sup> percentile; '▲' : 95<sup>th</sup> percentiles. Black marker: the cross-validated percentiles. Red markers: the percentiles when TAMSAT1 is calibrated using only the validating year's data. Solid line is the 1-1 line where the markers should lie if both distributions are identical.

The modelled positive pixel rainfall distributions were simulated in similar fashion as described above, but omitting Step 4ii. Only pixels where the coincidental gauge-estimated rainfall amount was positive were used to generate the rainfall distribution. As before, a leave-one-year-out validation was adopted in modelling the distribution parameters within the simulation. The simulated and observed rainfall distribution was then compared using quantile-quantile plots. The points in these plots represent the quantiles of the two distributions, from 0% to 100% with an increment of 1%. Figure 4.10 shows such plots for a few representative years. For most cases, simulated rainfall distribution was found to be in good agreement with the observed rainfall distribution. For 2002 however, the bias in the model parameters resulted in a significant over-prediction of rainfall, over the entire range of values.

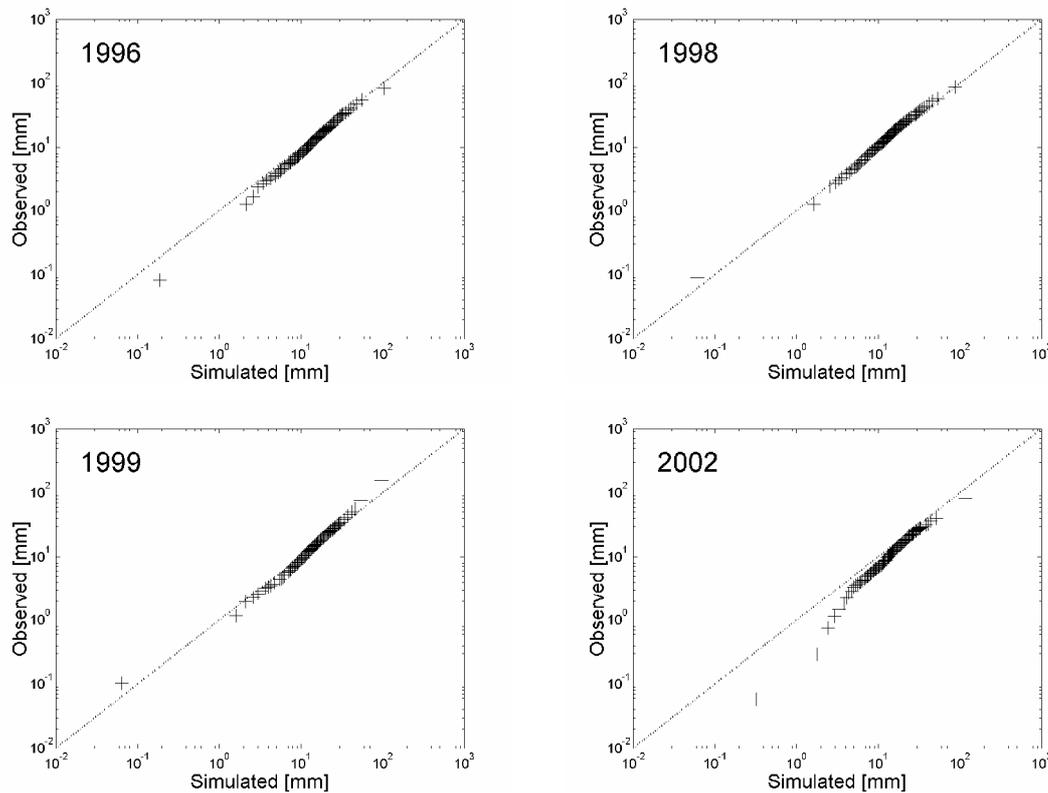


Figure 4.10 Quantile-Quantile plot of simulated versus observed positive rainfall for 4 years. Note that the plot is in logarithmic scale in both axes. Solid line is the ideal relationship between the quantiles if both distributions are identical.

## 4.5 Discussion

### Aggregation in space and time

Since TAMSAT1 rainfall is estimated from the pixel CCD, the relationship between satellite estimates and observed rainfall, at various spatial and temporal scales, depends on the underlying correlation between the spatio-temporal aggregated rainfall and CCD. Using the set of pixels containing rainfall gauges observations, correlations between CCD and observed rainfall, as well as correlations between TAMSAT1 estimates and observed rainfall for different temporal accumulations,  $\tau$ , at pixel scale were determined. Similar correlations were recomputed at a larger spatial scale, “ALL GAMBIA”, by averaging rainfall over all pixels (containing gauge observations) for each day.

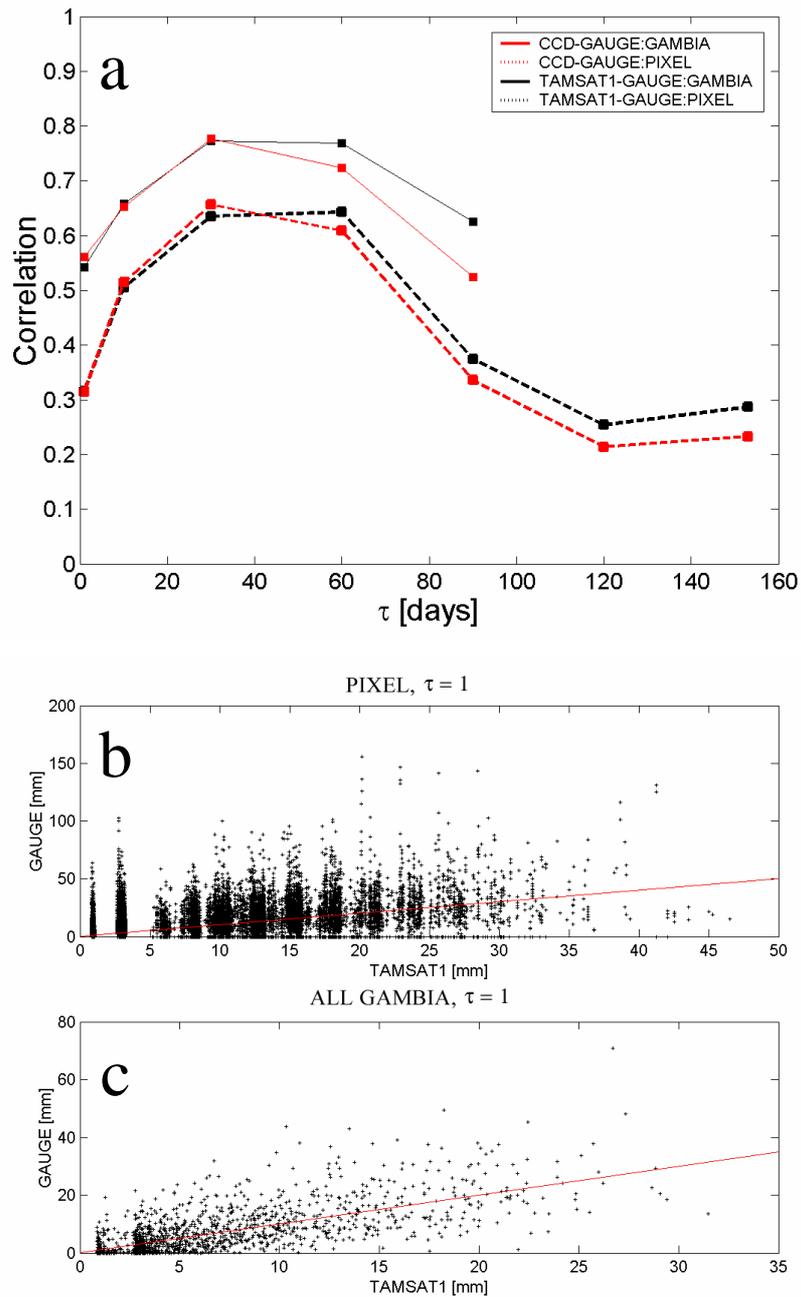


Figure 4.11 Effects of spatial-temporal aggregation of pixel rainfall. (a) Correlations ( $r^2$ ) between CCD and observed rainfall (CCD-GAUGE) and correlations between TAMSAT1 estimates and observed rainfall (TAMSAT1-GAUGE) at pixel scale and ALL GAMBIA scale for different temporal aggregations,  $\tau$ . Only correlations with  $p$ -value  $< 0.05$  are shown. (b) and (c): Scatter plots of daily observed rainfall against TAMSAT1 estimates at pixel and ALL GAMBIA scale respectively. Red line in (b) and (c) is the 1:1 line.

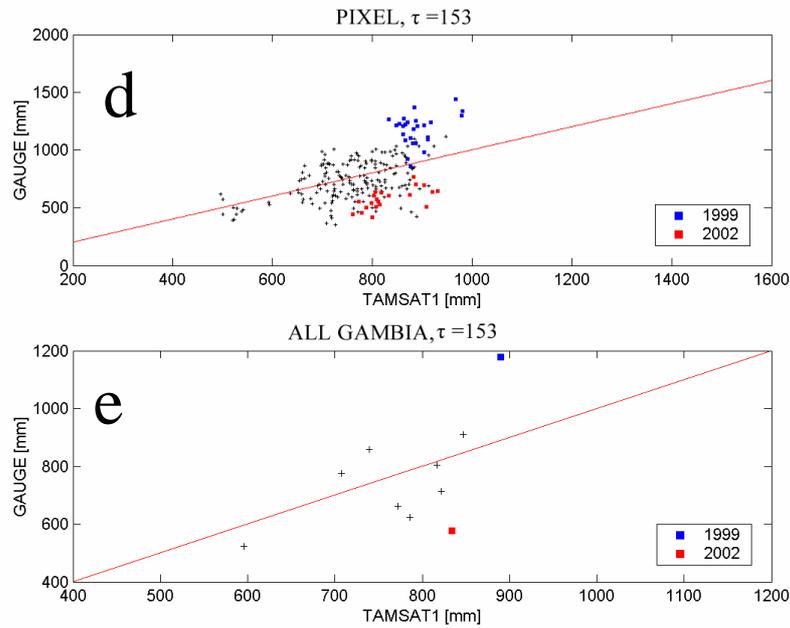


Figure 4.11 Effects of spatial-temporal aggregation of pixel rainfall (Continue). (d) and (e): Scatter plots of observed seasonal (June - Oct) rainfall against TAMSAT1 estimates at pixel and ALL GAMBIA scale. Red line is the 1:1 line.

$\tau$ [day]	1	10	30	60	90	120	153
A: CV for CCD [%]	114	55	50	32	13	12	12
B: CV for rainfall [%]	199	85	65	46	31	28	28
A/B	0.57	0.65	0.77	0.69	0.42	0.43	0.42

Table 4.6 Coefficient of variance (CV) of pixel-scale rainfall and CCD for Gambia at various temporal aggregates.

The results are summarised in Figures 4.11a-e. Correlation between TAMSAT1 and observed rainfall at the ALL GAMBIA scale is consistently higher at all temporal aggregation compared to pixel rainfall due to the reduction of variance in the CCD-rainfall relationship after spatial averaging (compare Figures 4.11b and 4.11c). The CCD-rainfall correlation (and therefore TAMSAT1-observed rainfall correlation) at both pixel and ALL GAMBIA scales increases for shorter time aggregates, reaching a maximum at  $\tau \approx 30$ -60 days and decreases for larger time aggregates.

The main reason for the observed CCD-rainfall correlation under temporal aggregation is due to the difference in the variance reduction of both variables when aggregated in time. Table 4.6 shows the CV of the rainfall and the CCD at pixel scale for various time aggregates. Daily rainfall exhibited higher variability when compare to that of daily CCD (see also Figure 4.11a). As  $\tau$  increases up to 60 days, the variability of rainfall decreases more than that of the CCD, improving the CCD-rainfall correlation. For the larger time aggregates of 90 days and beyond, the variability of rainfall remains relatively high CV (~30%) compared to CCD (~12%) – consequently, the prediction error variance, in particular for the higher rainfall estimates, is large (Figure 4.11d).

#### Effects of systematic bias

Since the parameters in Equations 4.5 are determined using historical gauge data, one implicit assumption of TAMSAT1 is that there is a certain degree of stationarity in the rainfall processes. Unfortunately, for many regions in Africa, interannual variability of rainfall can be significant. One consequence is that for abnormally ‘dry’ or ‘wet’ years, TAMSAT1 may introduce systematic bias in the rainfall estimates. This is best illustrated for the Gambian rainfall for 1999 and 2002, which are two years where rainfall were above and below normal respectively (see Chapter 2). From Table 4.5, the bias in the model parameters resulted in an underestimation and overestimation of the daily rainfall by 1.90 mm/day and 1.68 mm/day on average. The effects of such bias are obvious in Figure 4.11d; the biases in TAMSAT1 can be clearly seen from the clustering of the gauge-TAMSAT1 scatterplots. However, it is important to note that such systematic bias in TAMSAT1 for a given year may not be even from month to month: For 2002, the biases of TAMSAT1 estimates for Gambian rainfall were primarily confined to July and August (Table 4.7).

	June	July	August	September	October
TAMSAT1	0.52	-3.25	-3.20	-1.97	-0.43
TAMSAT1*	0.52	-0.82	0.34	-1.97	-0.43

Table 4.7 Bias in mm of rainfall estimates for each month of 2002 for TAMSAT1 and TAMSAT1\*(with bias correction for July and August). See text for details.

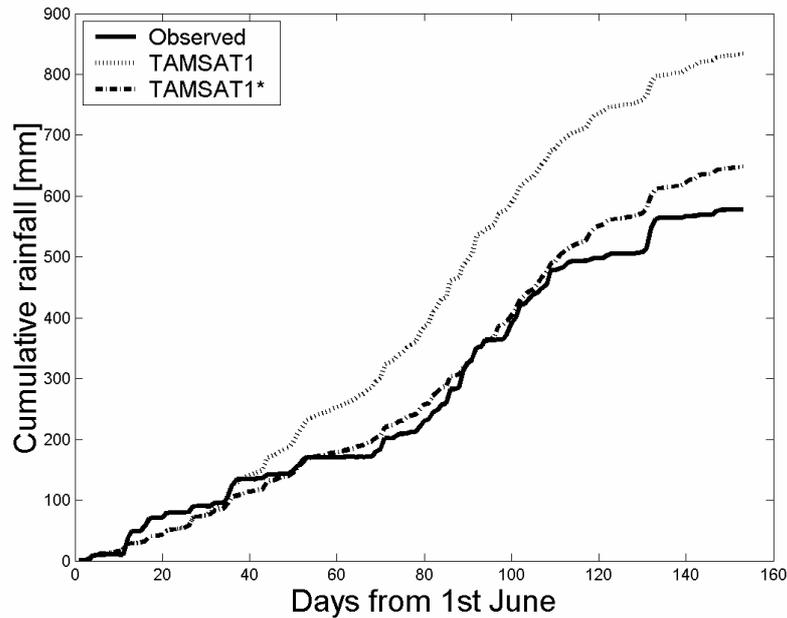


Figure 4.12 Effect of bias in TAMSAT1 for 2002. Cumulative ALL-GAMBIA rainfall from June to October. TAMSAT1\* in the figure denotes TAMSAT1 with bias correction for July and August.

The effect of bias in the TAMSAT1 is best seen in the cumulative mean areal rainfall for the Gambia for 2002. After applying a simple bias correction by reducing the magnitude of the TAMSAT1 model parameters that affect mean pixel rainfall estimates, namely  $a_0$ ,  $a_1$ ,  $b_0$ ,  $b_1$ ,  $p_0$  of Equations 4.5 individually by 30% for July and August, the pixel rainfall bias reduced substantially for the two months (Table 4.7). The bias-corrected cumulative rainfall estimates subsequently also showed a marked improvement over TAMSAT1 cumulative rainfall estimates (Figure 4.12). Apart from introducing systematic error in the mean pixel rainfall, such bias in mean pixel rainfall caused systematic error in the higher order statistics (variance, skewness) in the pixel rainfall distribution (Equation 4.5b). Even when there is no systematic error in mean pixel rainfall, errors in the shape and scale parameters in the variance model (Equations 4.5d and 4.9) can lead to misrepresentation of the rainfall distribution, and consequently caused error in quantitative assessment of uncertainties in the rainfall estimates.

It is important in the context of practical application of satellite-based rainfall estimates in hydrology and crop yield forecasts to realise that contribution of overall systematic error by the various months may not be even. In crop modeling, sensitivity to crop water satisfaction at different crop developmental stages may have unequal impact (see for example Sultan *et al.*, 2005). The sensitivities to such intra-seasonal biases in TAMSAT1 on Gambian groundnut yield prediction are explored further in Chapter 6.

#### Possible improvement

For TAMSAT, as well as many other TIR-rainfall estimation algorithms available in literature, threshold temperature is used for delineation of rainy and non-rainy regions. One disadvantage of TAMSAT1 compared to these RFE-algorithms is that it does not offer a direct delineation of rainy and non-rainy regions, since estimated rainfall from TAMSAT1 is always greater than zero ( $= p\mu_+$ ). This leads to over prediction of rainfall for days when there is obviously no rain over large extent of the domain. For example, Figure 4.13 shows the daily rainfall overall Gambia for October 1996, obtained from TAMSAT1 and gauge estimates: For days after the 19<sup>th</sup> of July, no rainfall was observed over all rain gauges, except on three days, from the 24<sup>th</sup> to the 26<sup>th</sup>, while TAMSAT1 estimated a mean rainfall of about 1mm per day for the entire period. Although such overestimations are in the order of several mm per day and its effect can often be neglected in practice, it highlights some deficiencies of TAMSAT1 in describing situation of large-scale intermittency. Although additional information for example, satellite images from the visible channel could be use to discern such situation, it would add an additional layer of complexity to the rainfall estimation model, especially when automation is desired.

One possible improvement to TAMSAT1 in this aspect is to utilise larger scale TIR-rainfall characteristics to identify such instances of domain-wide zero-rainfall. For the Gambia, it was observed from the available records that 90% of the days when no rainfall was recorded by gauges country-wide, corresponded to days when 95% of the CCD at the pre-determined optimum threshold temperatures were zero within a domain encompassing the country (bounded from 13N to 14N and 13.6W to 16.9W). The

situation when 95% of pixel CCDs are zero is termed as Large Domain Zeros (LDZ) in this thesis. A simple improvement to TAMSAT1 for use in the Gambia is as follows: The TAMSAT1 algorithm is used for all non-LDZ days. Hence it is necessary to re-calibrate TAMSAT1 parameters in Equations 4.5a-d, excluding the LDZ days. For LDZ days, a single probability of rain, determined from the gauge-estimated pixel rainfall data, is assigned to all pixels regardless of the pixel-CCD values; positive rainfall are still determined using Equations 4.5 b-d. Although this improvement still cannot identify explicitly no-rainfall days, it achieved lower bias in days when no rainfall was observed (Figure 4.13). In the next chapter, consideration of LDZ is shown to improve the spatial statistics of rainfall field generated using TAMSAT1.

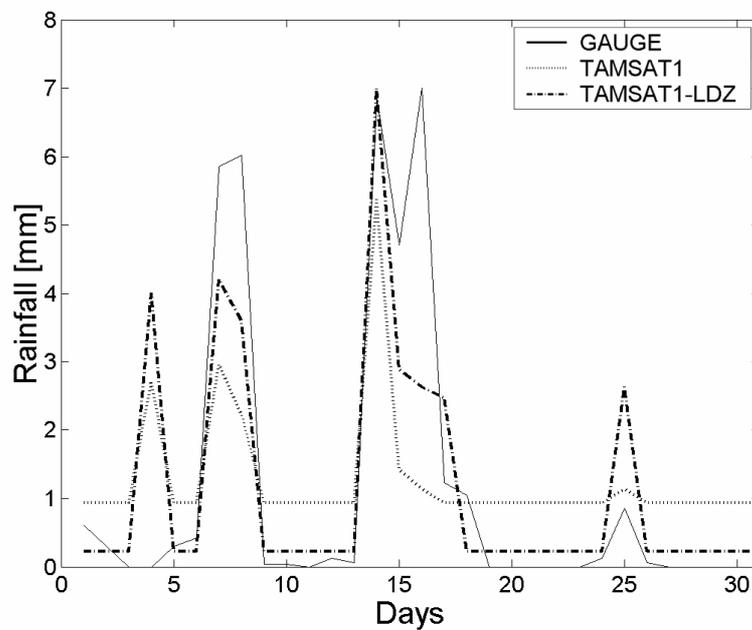


Figure 4.13 All Gambia mean daily rainfall for October 1996.

## 4.6 Conclusion

A new rainfall estimation algorithm, TAMSAT1, has been introduced. It is a generalisation from the TAMSAT method, commonly used in estimating dekadal rainfall using satellite CCD data. It retains the operational simplicity (uni-variate and linear CCD-rainfall relationship) of the TAMSAT algorithm, but includes a complete statistical model for the pixel CCD-rainfall relationship. The daily rainfall for the Gambia was used as a case study to validate the proposed TAMSAT1 algorithm. For the Gambia, TAMSAT1 performed as well as, or slightly better than TAMSAT in terms of the basic skill scores in daily pixel rainfall estimates. The algorithm has been demonstrated to reproduce the pixel rainfall distribution for the Gambia through Monte Carlo simulations. This is particularly useful for accessing the uncertainties of such rainfall estimations in view of the highly non-Gaussian rainfall distribution. The ability of TAMSAT1 to be used in stochastic simulation of a spatial rainfall field is further discussed in the following chapter. TAMSAT1 is especially useful in spatial aggregated rainfall estimation for timescale between 10 to 60 days.

For anomalous rainfall years, the variability in TAMSAT1 parameters can result in systematic bias in rainfall. The case for the Gambian rainfall through the rainy seasons for 2002 demonstrated how systematic error in two months can lead to overestimating the total seasonal rainfall by as much as 34%. The ability to capture such anomalous low rainfall is vital in the context of drought monitoring and food security in regions practicing rain-fed agriculture. It is inevitable through its statistical formulation that TAMSAT1 should give a non-zero rainfall, even for days that obviously have no rainfall. Although such bias is slight and should not pose any serious problems in most practical applications of these rainfall estimates, it highlights the deficiency of TAMSAT1 to estimate rainfall with large-scale intermittencies. Further improvement to TAMSAT1 was achieved by separately considering the CCD values at a larger spatial scale, to better identify days where no rainfall is expected.

To reduce the effect of systematic biases, real-time or near real time gauge rainfall observations can be included in the model calibration, as those proposed by Todd *et al.* (1999), or a direct combination of the real-time rainfall network observations (Grimes *et al.*, 1999). Such methods may not be feasible for regions where there are significant delays in the availability in gauge reports. With the introduction of Global Precipitation Monitoring program in the near future, increasing the frequency of microwave soundings over the region, real-time calibration of the rainfall models using microwave rainfall estimates, as proposed in the literature (example Todd *et al.*, 2001), is another alternative. However, more validations on microwave-based rainfall estimates over Africa are required to ascertain the quality of these rainfall estimates. As interannual variability of rainfall in many regions of Africa has been known to relate to larger scale phenomena such as regional SST and ENSO, it may be also worthwhile in pursuing the feasibility of including such information into the calibration methodology of TAMSAT1 for regions where SST or ENSO signal is evident in rainfall.

# Chapter 5 Stochastic generation of rainfall using satellite data

## 5.1 Introduction

Attempting to obtain the spatio-temporal distribution of rainfall using statistical models has been a preoccupation of many researchers for decades, driven by the need to quantifying the uncertainties associated with rainfall estimates due to finite sampling in space and time. These rainfall models first saw application in hydrological research in the 1960s and 1970s (LeCam, 1961; Mejía and Rodríguez-Iturbe, 1974; Bras and Rodríguez-Iturbe 1976a), motivated primarily by the need to understand the impact of rainfall uncertainties on hydrology due to interpolation from gauge observations to ungauged location. As satellite derived rainfall data becomes available, statistical rainfall models have also been used in investigating the uncertainties of satellite-derived rainfall due to finite sampling in space (Huffman, 1997) and time (Bell, 1987; Bell *et al.*, 1990). Downscaling of NWP predicted rainfall using stochastic rainfall models for hydrological research has also been attempted (Wilks, 1989; Perica and Foufoula-Georgiou, 1996; Kyriakidis *et al.*, 2004). Without loss of generality, one might describe stochastic rainfall models as numerical models that provide statistical spatial and/or temporal representations of rainfall with a varying degree of sophistication in their description of the characteristics of the associated storms. Various classifications of such rainfall models in literature have been attempted. A classification frequently used is given by Rhenals-Figueroa *et al.* (1974), who suggested the following classification of rainfall models:

- 1) *Point rainfall models* are those that generate a time sequence of rainfall for a single location.
  
- 2) *Multivariate rainfall models* are those that generate rainfall for several locations simultaneously preserving the spatial structure of the observed rainfall at a given time.

3) *Areal or multi-dimensional rainfall models* are those that are able to generate rainfall for everywhere within a domain of interest.

Bras and Rodríguez-Iturbe (1976) suggested an alternate classification of all the above rainfall models:

1) *Rainfall exterior models*: these generate the exterior characteristics of the storms like the total rainfall amount, and duration, without involving the explicit statistical modelling of individual storm events.

2) *Rainfall interior models*: those that resolve the time distribution of rainfall amount within each storm event. Early examples of these models include that of Bras and Rodríguez-Iturbe (1976) and subsequently, more sophisticated rainfall models which emerged during the 1980s that emphasised the physical characteristics of storms (e.g. Gupta and Waymire, 1979; Waymire *et al.*, 1984; Lovejoy and Mandelbrot, 1985; Kavvas *et al.*, 1987).

Much attention has been focused on multidimensional (2-space, 1-time) rainfall exterior models due to their relative simplicity in application. Rainfall exterior models are in general, less sophisticated in their model description than rainfall interior models, and therefore, are not expected to generate as many features of observed rainfall compared to the more sophisticated rainfall interior models. However, the low parameter requirements and good reproduction of rainfall fields makes the rainfall exterior models appealing in practice (Lanza, 2000), while model parameter estimations remain a major obstacle for rainfall interior models for practical application (Sivapalan and Wood, 1987), although some progress has been achieved in this respect (Smith and Karr, 1985; Mellor, 1996; Mellor and Metcalfe, 1996; Mellor and O'Connell, 1996; Shah *et al.*, 1996).

### Monte-Carlo simulation of rainfall field

Synthetic rainfall fields generated from multidimensional rainfall models using Monte Carlo methods, have been used in sensitivity studies of hydrological model response to rainfall uncertainties due to finite spatial sampling (Shah *et al.*, 1996) and temporal sampling (Krajewski *et al.*, 1991; Hossain and Anagnostou, 2004; Hossain *et al.*, 2004a). The main constraints imposed on these simulations are the spatio-temporal covariance structures of the rainfall fields and any available rainfall observations (raingauge measurements, ground based radar or satellite based rainfall estimates). Another classification of rainfall models that is useful for our present discussion is that suggested by Smith and Krajewski (1987):

- 1) *Gaussian random field models and their variants*: These are models whose rainfall fields are linked to an underlying multidimensional Gaussian random field.
- 2) *Cluster models*: These are models based on statistical point processes, which follow from the work of LeCam (1961), and includes those of Gupta and Waymire (1979), Waymire *et al.* (1984), Smith and Karr (1985) and Rodríguez-Iturbe *et al.* (1986).

For this present review, the discussion is limited to stochastic generation of Gaussian random field models. In Gaussian random field models, the actual rainfall random field is associated with a Gaussian random field via a suitable transformation. The Gaussian random field is assumed to possess covariance structures, such that the covariance of the actual rainfall field is preserved after transformation. The advantage of using such models is that the rainfall simulation reduces to generating realisations of the underlying Gaussian random field; transforming these Gaussian realisations to the actual rainfall fields is straightforward once the transformation is determined. Several techniques of generating realizations of Gaussian random fields with prescribed covariance structures have been used. These include the Turning Band Method (Shah *et al.*, 1996; Bellerby and Sun, 2005); Spectral Methods (Mejía and Rodríguez-Iturbe *et al.*, 1974; Bell *et al.*, 1990; Lanza, 2000); Wavelet Transform (Pericia and Foufoula-Georgiou, 1996); Sequential Gaussian Simulation (Kyriakidis *et al.*, 2004).

### Proposed model

The main purpose of this chapter is to develop a stochastic rainfall model based on satellite data. The motivation is to use stochastically generated rainfall from this rainfall model under the condition that only satellite information is available, to drive crop models for the following purpose:

- 1) To understand the sensitivity of the crop yield prediction due to uncertainties of satellite rainfall estimates.
- 2) To obtain probabilistic crop yield prediction.

The rainfall model developed represents daily rainfall at satellite pixel level (~5 km × 5 km for METEOSAT TIR) with two independent random fields; a multi-dimensional Gaussian random field model for the positive rainfall and a binary random field to distinguish rainy and non-rainy pixels. The actual rainfall is the scalar product of the two simulated fields analogous to the method of delineating rainfall under fractional coverage by Barancourt *et al.* (1992). This is in contrast with the rainfall models of Bell (1987), Lanza (2000) and Bellerby and Sun (2005), where the rainfall field (zero rainfall included) is obtained from transforming a single Gaussian random field.

Similar to the model of Kyriakidis *et al.* (2004), the daily rainfall field was not assumed to be stationary: Statistics of the rainfall over a pixel depend on the pixel-CCD as prescribed by TAMSAT1 rainfall estimates. Both the rain/no-rain map and positive rainfall fields were generated using sequential simulation with the prior distribution of daily pixel-averaged rainfall obtained from TAMSAT1. The spatial structure of the proposed stochastic rainfall field is defined by the two-point covariance of the indicator transform of the rainfall, and the two-point covariance of the positive rainfall. The adoption of a sequential simulation algorithm within the simulation has two advantages: 1) The algorithm is relatively simple to implement. 2) If gauge observations are available, pixel rainfall estimated from the gauges can be incorporated into the simulation without much modification to the algorithm.

The remainder of this chapter is structured as follows: Section 5.2 formalises the rainfall models. Section 5.3 describes the algorithm of the conditional simulation and discusses some aspect of the simulation with a simple example. Section 5.4 provides a case study of using the algorithm to simulate daily rainfall for the Gambia. Section 5.5 discusses some possible improvement to the algorithm and Section 5.6 concludes the chapter with several recommendations.

## 5.2 The problem

For clarity, Box 3 and Box 4 summaries the key parameters and equations of the stochastic model being formulated later in this section. Consider a spatial domain  $\xi$  that is partitioned into  $N$  pixels, each with known CCD values. The rainfall estimates for these  $N$  pixels are obtained through the TAMSAT1 algorithm (Chapter 4, Equations 4.5a-d), which is assumed to provide unbiased estimates for the pixel-level rainfall in the absence of any other observations. Let the daily pixel rainfall be a random field  $Z(\mathbf{u}_j)$ , where  $\mathbf{u}_j$  represents a 2-dimensional positional vector for pixel  $j$ . The object of interest is the conditional distribution,  $g(\bar{Z} | D_\xi)$ , of  $\bar{Z}$ , the averaged daily rainfall over  $\xi$ , given CCD observations where

$$\bar{Z} = \frac{\sum_{j=1}^N Z(\mathbf{u}_j)}{N} \quad 5.1$$

and  $D_\xi$  denotes the set of pixel CCD over  $\xi$ . Although the distribution for an individual pixel given its CCD is known (Equations 4.5a-d), an analytic expression of  $g(\bar{Z} | D_\xi)$  is not trivial due to the spatial correlation of  $Z$ . The method used here to obtain an estimate of  $g(\bar{Z} | D_\xi)$  is the direct estimation through multiple simulations of possible rainfall within  $\xi$ . For the  $k^{th}$  simulation, a set of pixel-rainfall realisation  $\{Z^{(k)}(\mathbf{u}) | \mathbf{u} \in \xi\}$  constrained to the observed CCD is obtained, and the corresponding averaged rainfall,  $\bar{Z}^{(k)}$ , is then computed. The proposed simulation method is a two-stage process:

First, the rainy areas within  $\xi$  are determined. Next, the rainfall amounts within these rainy areas are calculated. For a large enough number of simulations, say,  $M$ ,  $\{\bar{Z}^{(k)} | k=1, \dots, M\}$  can be used to estimate  $g(\bar{Z} | D_\xi)$ .

### Box 3: Notation of key parameters of stochastic simulation model

$\hat{\mathbf{C}}$	Normalised covariance matrix of $R(\mathbf{u})$
$\hat{\mathbf{C}}_0$	Normalised covariance vector of $R(\mathbf{u})$
$D_T$	Pixel CCD for temperature threshold $T$
$f(\mathbf{u})$	Sample of the $F(\mathbf{u})$
$F(\mathbf{u})$	Rainfall amount for pixel at location $\mathbf{u}$ if pixel is rainy
$G(\cdot)$	Cumulative distribution function of a Standard Normal
$I(\mathbf{u})$	Indicator transformed of pixel rainfall located at $\mathbf{u}$
$p(\mathbf{u})$	Probability of pixel located at $\mathbf{u}$ being rainy given the pixel CCD
$Q(\cdot; a, b)$	Cumulative distribution function of a Gamma Distribution with mean $a$ and variance $b^2$
$r(\mathbf{u})$	Sample of $R(\mathbf{u})$
$S$	Set of pixel indicator realisation
$S_y$	Set of realisation of normal transformed positive rainfall
$R(\mathbf{u})$	Indicator residual $I(\mathbf{u}) - p(\mathbf{u})$
$w(\mathbf{u})$	Sample of $W(\mathbf{u})$
$W(\mathbf{u})$	Residual of the positive rainfall $F(\mathbf{u}) - \mu_+(\mathbf{u})$
$y(\mathbf{u})$	Realisation of $Y(\mathbf{u})$
$Y(\mathbf{u})$	Normal transform of positive rainfall $F(\mathbf{u})$
$\phi$	Normal transformation of $W(\mathbf{u})$ to $Y(\mathbf{u})$
$\Lambda$	Covariance matrix of $Y(\mathbf{u})$
$\Lambda_0$	Covariance vector of $Y(\mathbf{u})$
$\mu_+(\mathbf{u})$	Mean positive rainfall amount at pixel located at $\mathbf{u}$ given the pixel CCD
$\sigma_+(\mathbf{u})$	Standard deviation of positive rainfall at pixel located at $\mathbf{u}$ given pixel CCD

**Box 4: Key Equations of the stochastic simulation model****Pixel level rainfall model**

Probability distribution function of $F(\mathbf{u})$ given pixel CCD	Equation 4.5 b-d, 4.8, 4.9
Probability of pixel being rainy given pixel CCD	Equation 4.5a

**Simulation of indicator field  $I(\mathbf{u})$** 

Estimation of probability of pixel being rainy given pixel CCD and a set of realisation	Equation 5.2
Estimation of $R_j$ given a set of realisation	Equation 5.8, 5.9

**Simulation of positive rainfall field  $F(\mathbf{u})$** 

Estimation of mean and variance of $Y(\mathbf{u})$ given a set of realisation	Equation 5.10, 5.11
Conditional distribution of $Y(\mathbf{u})$ given a set of realisation	Equation 5.12
Transformation of $Y(\mathbf{u})$ to $W(\mathbf{u})$	Equation 5.14
Obtaining $F(\mathbf{u})$ from $W(\mathbf{u})$	Equation 5.3

**5.2.1 Model and assumptions**

Following Barancourt *et al.* (1992), the finite probability mass of zero rain may be accounted for by representing  $Z$  as a scalar product of two independent variable  $I$  and  $F$ :

$$Z(\mathbf{u}_j) = I(\mathbf{u}_j)F(\mathbf{u}_j)$$

where  $I$  is a indicator random field defined as

$$I(\mathbf{u}_j) = \begin{cases} 0 & Z(\mathbf{u}_j) = 0 \\ 1 & Z(\mathbf{u}_j) > 0 \end{cases}$$

$I(\mathbf{u}_j)$  is an indicator of whether the pixel at  $\mathbf{u}_j$  is rainy. Hence, it can be represented by a Bernoulli trial with a certain probability of ‘success’ (rain).  $F(\mathbf{u}_j)$  is a random field representing positive rainfall amount, and has a non-negative distribution, which is

usually non-Gaussian. For ease of discussion, the following mathematical notation is adopted hereafter:

1) Functional dependence on the pixel location will be indexed as a subscript whenever possible. For example  $Z_j = Z(\mathbf{u}_j)$ .

2) A set within a parenthesis of a function means the set of functions at each elements of the set. For example, for a function  $A$  and  $U = \{x_j | j = 1, \dots, N\}$ ,  $A(U) = \{A(x_j) | j = 1, \dots, N\}$ .

Unbiasedness in TAMSAT1 requires that for pixel  $j$  with  $CCD = D_j$ ,  $E[I_j] = P(Z_j > 0 | D_j) = p_j$  (Equation 4.5a), and  $E[F_j] = \mu_{+j}$  (Equation 4.5c). However, when the realisations of rainfall for several pixels within  $\xi$  are available, it is reasonable to assume that the conditional expectation of  $I_j$  and  $F_j$  for an unsampled pixel  $j$  will be perturbed away from  $p_j$  and  $\mu_{+j}$  respectively because of spatial continuity. Several assumptions are made in order to estimate these quantities:

1. Assume  $n$  pixels located at  $U' = \{\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_n\}$  within  $\xi$  where the rainfall including zero rain, is known. If  $m$  of the  $n$  pixels within  $U'$  are rainy, we can define  $U_F'$ , a subset of  $U'$ , containing all the rainy pixels:  $U_F' = \{\mathbf{u}' | \mathbf{u}' \in U', Z(\mathbf{u}') = F(\mathbf{u}') > 0\}$ .

2. Defining  $S$  as the set of indicator values  $I$  for all  $n$  pixels in  $U'$ :  $S = \{I(\mathbf{u}') | \mathbf{u}' \in U'\}$ , an estimator of the conditional probability of rain at an unsampled pixel  $\mathbf{u}_j$  can be expressed as a linear combination of the residuals,  $R(\mathbf{u}')$ , where  $R(\mathbf{u}') = I(\mathbf{u}') - p(\mathbf{u}')$ ,  $\mathbf{u}' \in U'$ :

$$[P(Z_j > 0 | S)]^* = p_j + \sum_{\alpha=1}^n \lambda_{j\alpha} R_\alpha \quad 5.2$$

where  $R_\alpha = R(\mathbf{u}'_\alpha)$ ,  $\mathbf{u}'_\alpha \in U'$ .

3. If  $U_F'$  is non-empty, define  $S_w$  as the set of residuals of the positive rainfall in  $U_F'$ :  $S_w = \{W(\mathbf{u}) = F(\mathbf{u}) - \mu_{+}(\mathbf{u}) | \mathbf{u} \in U_F'\}$ , the positive rainfall amount at an unsampled pixel  $\mathbf{u}_j$ ,

can be represented as a combination of a zero-mean residual random field  $W(\mathbf{u})$  about  $\mu_+(\mathbf{u})$ :

$$F_j = \mu_{+j} + W_j \quad 5.3$$

where  $W_j$  depends on the sampled residual  $S_w$  as described below.

4. There exists an invertible transformation,  $\phi$ , such that the transformed residual,  $Y = \phi(W)$  is multivariate Gaussian. It follows then that the conditional distribution of  $Y_j$  is Gaussian. A useful conditional distribution of  $Y$ , following the assumption of  $Y$  being multivariate Gaussian is (Goovaerts, 1997, Section 7.2.1)

$$G(y_j | S_y) = G\left(\frac{y_j - E\{Y_j | S_y\}}{\sqrt{\text{Var}(Y_j | S_y)}}\right) \quad 5.4$$

where  $G$  is the cumulative distribution of the standard normal, and  $S_y$  represents the normal transform of the conditioning set  $S_w$ :  $S_y = \{Y(\mathbf{u}) | \mathbf{u} \in U_F\} = \phi(S_w)$ . Furthermore, it is assumed that  $Y_j$  can be expressed as a linear combination of the transformed residual  $S_y$ ,

$$Y_j = \sum_{\alpha=1}^m \lambda_{j\alpha} Y_\alpha \quad 5.5$$

where  $Y_\alpha$  are elements of  $S_y$ .

5. Both  $Y(\mathbf{u})$  and  $R(\mathbf{u})$  are stationary and isotropic. Hence the two-point covariances  $\text{Cov}(Y_p, Y_q)$  and  $\text{Cov}(R_p, R_q)$  are only dependent on the distance between the pixel  $p$  and  $q$ ,  $h_{pq}$  (Goovaerts, 1997, Section 3.2.2):

$$\text{Cov}(Y_p, Y_q) = \Lambda(h_{pq}) \quad 5.6$$

$$\text{Cov}(R_p, R_q) = C(h_{pq}) \quad 5.7$$

### 5.2.2 Estimation of conditional parameters

The above assumptions allows Simple Kriging (SK) to be used in estimating the conditional probability of rainfall and the conditional distribution of positive rainfall amount at an unsampled pixel.

*Conditional probability of rain:* For the given set of  $n$  indicator transformed values  $S$ , the corresponding indicator residuals can be obtained as  $r_\alpha = i(\mathbf{u}'_\alpha) - p(\mathbf{u}'_\alpha)$ , where  $\mathbf{u}'_\alpha \in U'$ . The conditional probability of rain for an unsampled pixel  $j$  can be estimated by first rewriting Equation (5.2) as

$$R_j^{SK} = \sum_{\alpha=1}^n \lambda_{j\alpha} r_\alpha \quad 5.8$$

where  $R_j^{SK} = [P(Z_j > 0 | S)]^* - p_j$ . The above can be solved using Simple Kriging where the kriging weights  $\lambda_j = [\lambda_{j1}, \lambda_{j2}, \dots, \lambda_{jn}]$  are given by

$$\lambda_j = \mathbf{C}_0 \mathbf{C}^{-1}$$

where  $\mathbf{C}_0$  and  $\mathbf{C}$  are the  $1 \times n$  vector and  $n \times n$  matrix, whose  $p^{th}$  and  $pq^{th}$  elements are given by  $Cov(R_j, R_p)$  and  $Cov(R_p, R_q)$  respectively. This formulation of obtaining estimates of the conditional probability is the well-known ‘‘Simple Kriging with local means’’ (Goovaerts, 1997, Section 6.1.2). For the purpose of later development, it is noted that since  $r(\mathbf{u})$  is assumed to be stationary and isotropic,  $Cov(R_p, R_q) = C(h_{pq}) = \sigma_R^2 \hat{C}(h_{pq})$ , where  $h_{pq} = |\mathbf{u}_p - \mathbf{u}_q|$ ,  $\sigma_R^2 = C(0)$  (e.g. Goovaerts, 1997, Section 3.2.3.).  $\hat{C}(h)$  is the normalised covariance of  $R$ . Then the kriging weights can alternatively be expressed as

$$\lambda_j = \hat{\mathbf{C}}_0 \hat{\mathbf{C}}^{-1} \quad 5.9$$

where  $\hat{\mathbf{C}}_0$  and  $\hat{\mathbf{C}}$  are the  $1 \times n$  and  $n \times n$  matrices whose  $p^{th}$  and  $pq^{th}$  elements given by  $\hat{C}(h_{jp})$  and  $\hat{C}(h_{pq})$  respectively.

*Conditional distribution of positive rainfall:* Similarly, with the set of normal transformed residuals of positive of rainfall,  $S_y$ , SK can be used in Equation 5.5 to obtain estimates of the conditional mean and variance for  $Y$  at an unsampled pixel  $j$ :

$$[E(y_j | S_y)]^* = (y_j)_{SK}^* = \sum_{y_\alpha \in S_y} \lambda_{j\alpha} y_\alpha \quad 5.10a$$

$$[Var(y_j | S_y)]^* = (\sigma_j)_{SK}^* = \Lambda(0) - \sum_{u_\alpha \in U_F'} \lambda_{j\alpha} \Lambda(|\mathbf{u}_j - \mathbf{u}_\alpha|) \quad 5.10b$$

where the kriging weights  $\lambda_j = [\lambda_{j1}, \lambda_{j2}, \dots, \lambda_{jm}]$  are given by

$$\lambda_j = \Lambda_0 \Lambda^{-1} \quad 5.11$$

$\Lambda_0$  and  $\Lambda$  are the  $1 \times m$  and  $m \times m$  covariance matrices whose  $r^{th}$  and  $rs^{th}$  elements are  $\Lambda(h_{jr})$  and  $\Lambda(h_{rs})$  of Equation 5.6 respectively. Then, an estimated conditional distribution for  $y_j$  follows immediately from Equation 5.4

$$[G(y_j | S_y)]^* = G\left(\frac{y_j - (y_j)_{SK}^*}{(\sigma_j)_{SK}}\right) \quad 5.12$$

This can then be transformed back to give  $w_j$  from which a realisation of  $F$  can be obtained.

*Normal-score transform and its inverse:* The existence of a multivariate normal transformation,  $\phi$ , for the residual rainfall amount,  $W$ , is a stringent condition (Section 5.2.1). Although given the pixel CCD, the pixel rainfall distribution over the domain is known (Equations 4.5 b-d), there is no analytical construct of such transformation at present that can guarantee multivariate normality of the transformed rainfall residual  $y$ . The common approach (Deutsch and Journel, 1992, Section 5.2.1; Goovaerts, 1997, Section 7.2.2) is to provide instead, a univariate normal transformation such that

$$y = \phi(w) = G^{-1}(K_W(w))$$

where  $K_W$  is the univariate cdf of  $W$ ,  $G^{-1}(\cdot)$  is the inverse of the standard normal cdf  $G(\cdot)$ . Given the pixel CCD =  $D_T$ , the distribution of  $w$  is fully defined through Equations 4.5c, Equation 4.5d and Equation 5.4. Thus

$$y = \phi(w; D_T) = G^{-1}(Q(f; \mu_+(D_T), \sigma_+(D_T))) \quad 5.13$$

$Q$  is the cdf of a Gamma Distribution with a mean  $\mu_+(D_T)$  and variance  $\sigma_+(D_T)$  derived from Equations 4.5c and 4.5d, and  $f = w + \mu_+(D_T)$ . Multi-normality of  $y$  cannot be comprehensively tested. In practice, it is suggested that the bivariate normality of the transformed be verified (Goovaerts, 1997, Section 7.2.3) as a compromise, though it is by itself a necessary but not sufficient condition of multivariate normality. However, in the present rainfall model proposed, it is assumed that the transformation (Equation 5.13) is multi-normal as well. It follows immediately from Equation 5.13 that the inverse transformation of  $y$  back to  $w$  is

$$w = \phi^{-1}(y; D_T) = Q^{-1}(G(y); \mu_+(D_T), \sigma_+(D_T)) \quad 5.14$$

where  $Q^{-1}$  is the inverse of  $Q$ .

### 5.3 Sequential methods of simulating rainfall

The proposed rainfall model, known as SIMU hereafter, is formulated with the following principles in mind: For a given set of daily pixel-CCD at selected threshold temperature, the aim is to generate corresponding rainfall amount such that 1) the distribution of pixel-level rainfalls approaches Equations 4.5 in the limit of large number of realisation, and 2) spatial rainfall distribution for each realisation is consistent with the spatial correlation structure of the rainfall defined by the two-point covariances (Equations 5.6 and 5.7).

For the purpose of application in crop/hydrological model where upscaling of rainfall data is often necessary, rainfall distribution at these larger spatial extents can be obtained by aggregating the relevant pixel rainfall for each realisations. The principle behind the simulation procedure is as follows: 1) Simulate rain occurrence (rainy or not rainy) for individual pixels via Sequential Indicator Simulation (sIs), and 2) Simulate the rainfall amount for pixels that are raining using Sequential Gaussian Simulation (sGs). Detailed discussion of Sequential Indicator Simulation and Sequential Gaussian Simulation can be found in various references, including Goovaerts (1997) and Deutsch and Journel (1992). For clarity, the following subsections outline the sIs and sGs algorithm individually before the SIMU algorithm is summarised.

### 5.3.1 Sequential Indicator Simulation

To initialise the simulation, a set of initial sampled indicator values is obtained by selecting a random set of pixels with indicator values assigned (see Section 5.3.3). After selecting this set of ‘seed’ pixels,  $S$ , the algorithm of sIs to generate a realisation of indicator values conditional to  $S$ , over the remaining unsampled pixels in the domain, is as follows:

- 1) Define a random path  $U$  covering the entire set of unsampled pixels.
- 2) Compute the set of indicator residuals  $r$  for the sampled locations.
- 3) At the first pixel in  $U$  (say located at  $\mathbf{u}_j$ ), obtain  $[P(Z_j > 0 | S)]^*$  by adding the prior probability  $p_j$  from Equation 4.5a and a kriging estimate of the probability residual at  $j$  (Equation 5.8). Then, use this estimated probability for a Bernoulli trial to sample an indicator value (0 or 1) for the pixel.
- 4) Add this sampled value to the set of sampled indicator  $S$ .
- 5) Repeat step 2 through 4 for each pixel in  $U$  until all elements are sampled.

### 5.3.2 Sequential Gaussian Simulation

To initialise the simulation, sample a set of  $m$  ‘seed’ pixels from the set of rainy pixels (indicator value = 1),  $U_F$ . Assign rainfall amount to these pixels (See Section 5.3.3). Starting with the initial set of  $m$  assumed known positive pixel rainfall amounts, obtain the set of corresponding residuals,  $S_w = \{w_\alpha = f_\alpha - \mu_{+\alpha} | \alpha = 1, \dots, m\}$ . Next, perform a normal transformation (Equation 5.13) on all elements of  $S_w$  to obtain the set of normal-transformed residual  $S_y = \{y_\alpha = \phi(w_\alpha) | \alpha = 1, \dots, m\}$ .

The recipe for sGs to generate positive rainfall values for the rainy pixels is as follows:

- 1) Define a random path  $U$  covering the unsampled pixels in  $U_F$ .
- 2) Starting with the first pixel in  $U$ , determine the conditional (to  $S_y$ ) mean and variance of the normal transformed residual  $Y$  at the location via Simple Kriging (Equations 5.10a,b).
- 3) Sample a value,  $y$ , from the normal distribution given in Equation 5.12 with the estimated mean and variance from step 2 and add  $y$  to  $S_y$ .
- 4) Proceed to the next pixel in  $U$  and repeat step 2 and 3 for all pixels in  $U$ .
- 5) Finally, back transform  $S_y$  using Equation 5.14 into rainfall amount residual and recover the rainfall using Equation 5.3.

### 5.3.3 Summary of simulation of daily rainfall conditional to CCD

Given the set of CCD,  $D_\xi$ , the procedure for one simulation cycle of rainfall in the SIMU simulation algorithm is described below. A schematic is shown in Figure 5.1 to illustrate diagrammatically the various steps of the algorithm.

- 1) Randomly select  $n$  pixels located at  $U' = \{\mathbf{u}'_k \mid k = 1, \dots, n\}$  (see Remark 1). A set of indicator realisation,  $S = \{i(\mathbf{u}) \mid \mathbf{u} \in U'\}$ , is sampled using a Bernoulli trial and the probability of rain  $\{p(\mathbf{u}) \mid \mathbf{u} \in U'\}$  is determined through Equation 4.5a.
- 2) A realisation of the rain/no rain pixels within  $\xi$ ,  $I(\xi)$ , is then generated using sIs described in Section 5.3.1, starting with the initial 'seed' set of sampled indicator,  $S$ .
- 3) From the set of all pixels identified as rainy from step 2 called  $U_F$ , randomly select a subset of  $m$  pixels from  $U_F$ , called  $U_1$  (See Remark 1). Sample the rainfall amount for

each pixel in  $U_1$  from a Gamma Distribution with parameters given by Equation 4.5c to obtain an initial set of rainfall residual  $S_w = \{f(\mathbf{u}) - \mu_+(\mathbf{u}) \mid \mathbf{u} \in U_1\}$ .

4) Perform a normal score transform on the set of sampled rainfall amount residuals to obtain  $S_y = \phi(f(U_1) - \mu_+(U_1))$ .

5) With  $S_y$ , perform sGs described in Section 5.3.2 to obtain the residual of positive rainfall for all rainy pixels,  $y(U_F)$  within  $\xi$ .

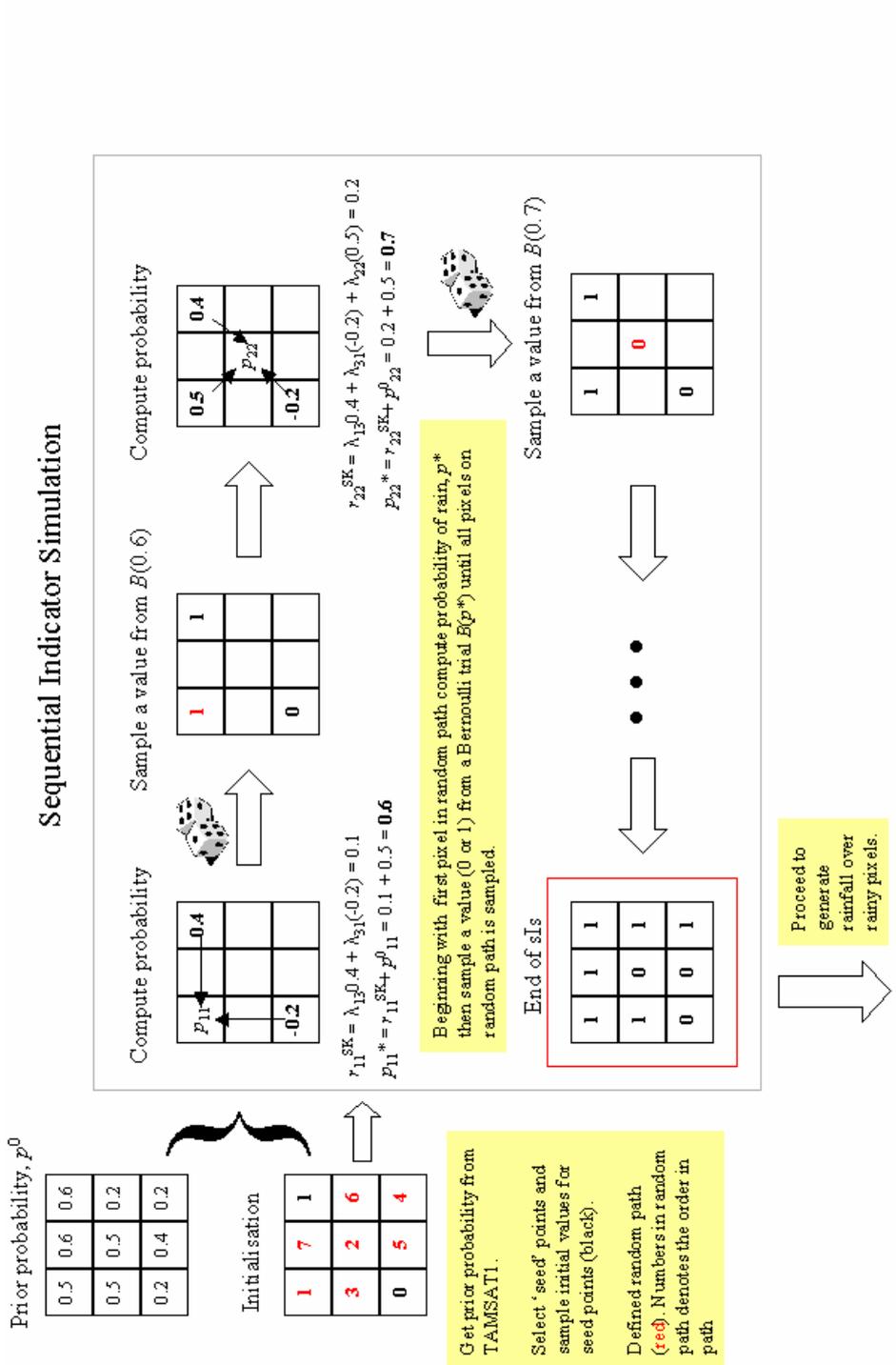
6) Obtain the rainfall from the inverse normal transform (Equation 5.14):

$$f(U_F) = \phi^{-1}(y(U_F)) + \mu_+(U_F).$$

7) Finally, the area-averaged rainfall for  $\xi$ ,  $\bar{Z}$ , can be obtained by averaging each realisation of  $Z$  over  $\xi$ , and the conditional distribution of the area-averaged rainfall amount  $g(\bar{Z} \mid D_\xi)$  can be obtained from large enough number of realisations.

*Remarks:*

1) Selection of the pixels in step 1 and 3 for the initial assignment of the indicator values and rainfall amounts through the use of Equations 4.5 should be such that the pixels are relatively far apart from one another. This is to ensure that they are ‘sufficiently’ independent from influence of each other. There are no definite rules on how far apart is sufficient. A rule of thumb is that the separation between these pixels should be at least the range of their semivariograms, but with enough pixels selected to proceed with the simulation.



Proceed to generate rainfall over r rainy pixels.

Figure 5.1 Schematic of one cycle of rainfall simulation using the SIMU algorithm illustrated with a fictitious 3x3 pixel domain. a) Simulation of the rain/no rain map using Sequential Indicator Simulation. b) Simulation of the non-zero rainfall using Sequential Gaussian Simulation.

Sequential Gaussian Simulation

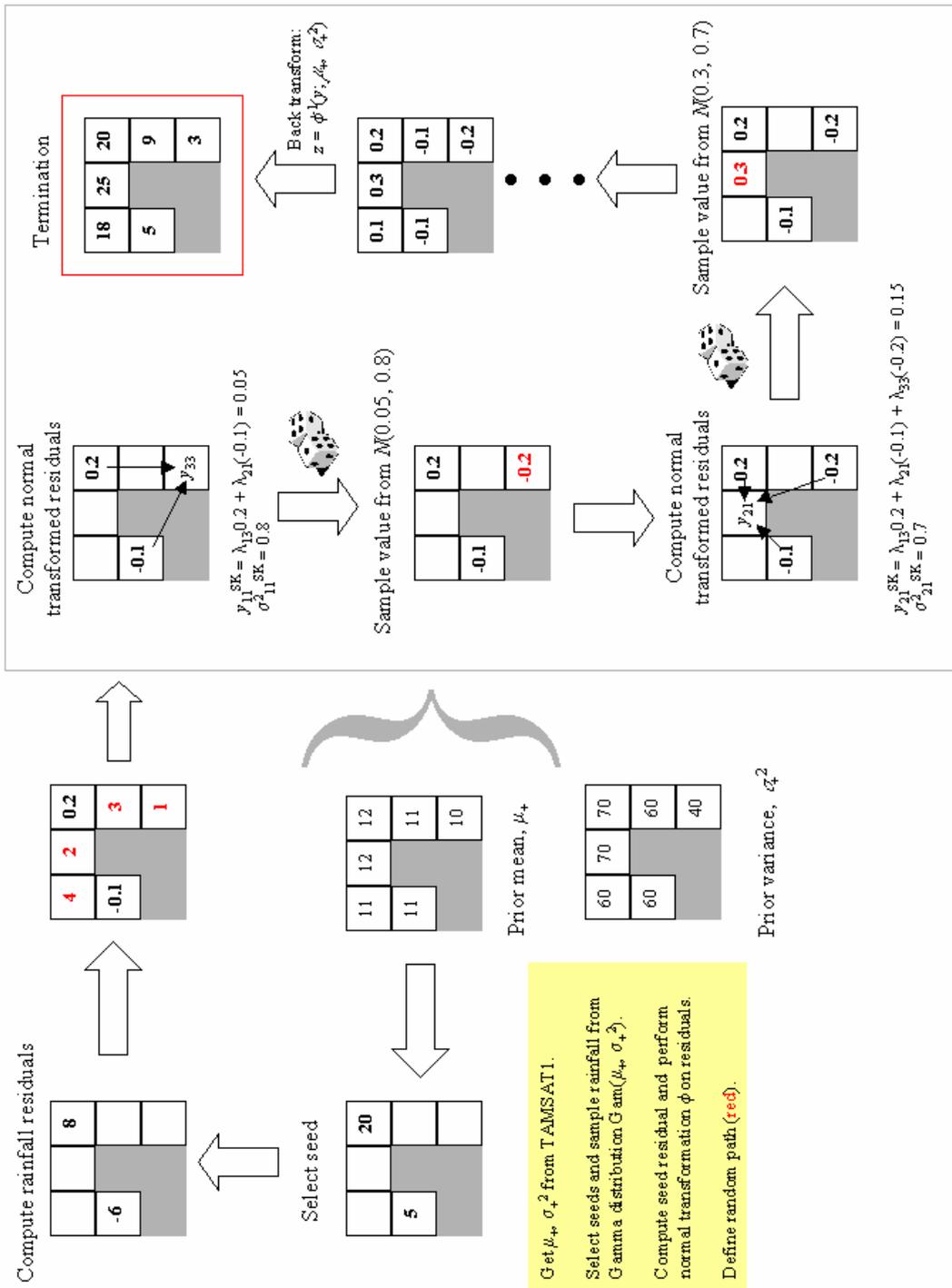


Figure 5.1b (Continue)

2) Each realisation depends not only on the CCD information but also on the initial allocation of rainy location (step 1) and rainfall amount (step 3). Therefore, a significant number of realisation need to be generated to remove any biases due to a preferential allocation of the rainy location and the rainfall amount. When the number of realisations is large, a reasonable representation of the distribution for the rainfall, conditioned solely on the CCD information, can be obtained.

3) In practice, to achieve a large number of realisations within a reasonable time, sIs is performed over a fixed random path. This means that the kriging weights within the sIs for each pixel in the random path is computed only once at the start of the simulation. In this way, a large number of indicator realisations can be achieved in a very short time. For each of these indicator realisations, the final rainfall realisation is generated through sGs, where again, the random path is fixed and kriging weights are calculated only once for each pixel along the random path.

### 5.3.4 A simple simulation example

A simple example of simulation using SIMU was performed to demonstrate the various features of the algorithm. For this example, the simulation of rainfall was carried out for a 25-pixel by 25-pixel rectangular region with a uniform CCD. Typical values for the rainfall estimates corresponding to the CCD value found in Chapter 4 were assigned namely,  $p(\mathbf{u}) = 0.6$ ,  $\mu_+(\mathbf{u}) = 12$  mm and  $\sigma_+^2(\mathbf{u}) = 60$  mm<sup>2</sup>. The indicator residual field,  $R(\mathbf{u})$ , was assumed to be stationary and isotropic, and hence the two-point covariance  $Cov(R_r, R_s) = C(h_{rs})$ , where  $h_{rs}$  is the distance between the  $r^{th}$  and  $s^{th}$  pixel. The exponential model of the form  $\sigma^2 e^{-h/L}$ , was adopted for both  $C(h)$  and the covariance function,  $\Lambda(h)$ , for the normal transformed residual,  $Y(\mathbf{u})$  (Equation 5.6). Here,  $\sigma^2$  is the variance of the random field, and the length scale of both covariances,  $L$ , were set to 4 pixels-unit for both  $R$  and  $Y$ . The initial seed pixels for assignment of indicator values were fixed as a rectangular grid of spacing  $2L$  (8 pixels-unit). A maximum number of sampled elements used in the kriging equations,  $n_K$ , used within both sIs and sGs were about 100. The left-panel of Figure 5.2 shows the results of several simulations of

rainy/non-rainy pixels from the sIs algorithm. In the next step, the pixels that are marked as rainy (with indicator values of 1) were used to sample the initial set of positive rainfall values through a Gamma Distribution with a mean of 12 mm and variance of  $60 \text{ mm}^2$ . sGs was then performed to obtain the rainfall field as shown on the right panel of Figure 5.2.

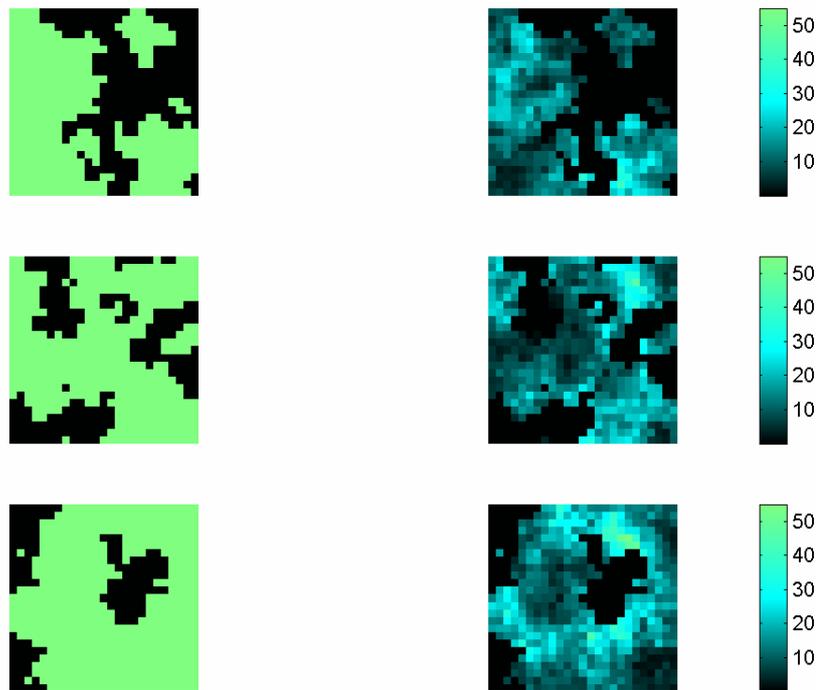


Figure 5.2 Daily pixel-rainfall of three simulation of a 25-pixel by 25-pixel region with  $n_K \sim 100$  with uniform CCD value over the domain. Left panel: Simulated rainy/no-rainy pixels at the end of the sIs algorithm. Right panel: Synthesized rainfall field at the end of the simulation cycle corresponding to the rain map to the left.

For a large enough number of simulations, the sample estimates of the various pixel-level rainfall statistics approached the satellite estimates as required; Figure 5.3 shows the plot of the estimated probability of rain, mean and variance of the positive rainfall against the number of simulations,  $N$ , for a pixel location not used in the initial assignment of indicator or rainfall values (Step 1 and 3 in Section 5.3.3). There is a general tendency for the sampled statistics to vary before converging. Similar results are obtained for several other unsampled pixels. For  $N = 300$  (sIs)  $\times$  2 (sGs), the sampled estimates for  $p$ ,  $\mu_+$  and

$\sigma_+^2$  are 0.6, 12.06 mm and 58.22 mm<sup>2</sup> respectively. It is essential that enough simulations are performed; it takes about 200 realisations before  $\mu_+$  and  $\sigma_+^2$  start to converge in this experiment.

By design, the spatial structure of the simulated rainfall fields is determined by the imposed covariances of the rainfall (Equations 5.9 and 5.11) within the simulation. Since  $R(\mathbf{u})$  and  $Y(\mathbf{u})$  are stationary, the semivariograms of both fields have simple relationships with their covariances. The semivariogram of  $R(\mathbf{u})$  for two locations  $\mathbf{u}_i$  and  $\mathbf{u}_j$ , separated by a distance  $h$ , is  $\gamma_R(\mathbf{u}_i, \mathbf{u}_j) = \gamma_R(h) = \sigma_R^2 \gamma(h)$ , where  $\sigma_R^2$  is the variance of  $R$ , and  $\gamma(h) = (1 - e^{-h/L})$ . Similarly, the semivariogram of  $Y(\mathbf{u})$ ,  $\gamma_Y(h) = \sigma_Y^2 \gamma(h)$ ,  $\sigma_Y^2$  being the variance of  $Y$  which in theory is unity. Rainfall spatial structure of the simulated rainfall is assessed using semivariograms instead of their covariances as sample semivariograms are more easily computed. For this example, the normalised sample semivariogram,  $\hat{\gamma}_R(h)$  and  $\hat{\gamma}_Y(h)$ , are computed from 10 randomly selected simulated rainfall fields. Both sample semivariograms were found to be in good agreement with the imposed structure,  $\gamma(h)$  (Figure 5.4).

To test the effect of the size of the kriging neighborhood used on the simulations, the experiment was conducted with various different  $n_K$ . Simulations with  $N = 600$  ( $300 \times 2$ ) at  $n_K \sim 25$  and  $n_K \sim 50$  showed similar pixel-level statistics ( $p$ ,  $\mu_+$  and  $\sigma_+^2$ ) compared with those of  $n_K \sim 100$ , and the sampled variograms compare well with  $\gamma(h)$  for all  $n_K$  (Figure 5.4). The distribution of the mean rainfall over the domain for the various  $n_K$  was found to be similar (Figure 5.5). The insensitivity of the simulations to  $n_K$  is mainly due to the high degree of continuity of the semivariogram model (zero nugget), which provided a significant screening effect within the kriging procedures; a small amount of sampled information nearest to the target location is sufficient to obtain the kriged estimates. This gives some impetus to select the smallest  $n_K$  possible since the simulation time increases as a function of  $n_K^2$ .

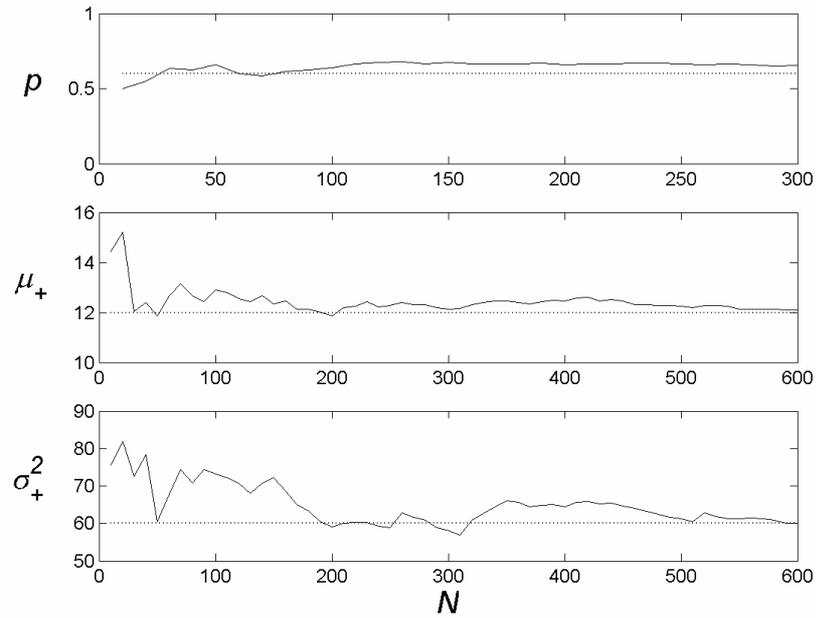


Figure 5.3 Plots of sampled rainfall statistics for a selected pixel against number of simulations,  $N$ . Top: Probability of rain ( $p$ ); Middle: Mean of positive rainfall ( $\mu_+$ ); Bottom: Variance of positive rainfall ( $\sigma_+^2$ ). Dotted lines are the expected values.

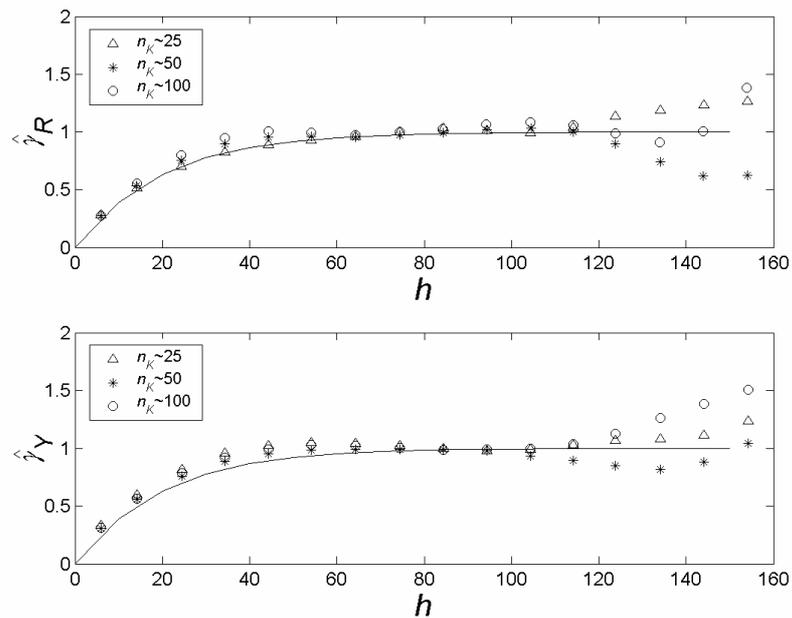


Figure 5.4 Sampled for various neighbourhood sizes ( $n_K$ ) and imposed normalised semivariogram (solid line) for  $R$  (top) and  $Y$  (bottom).

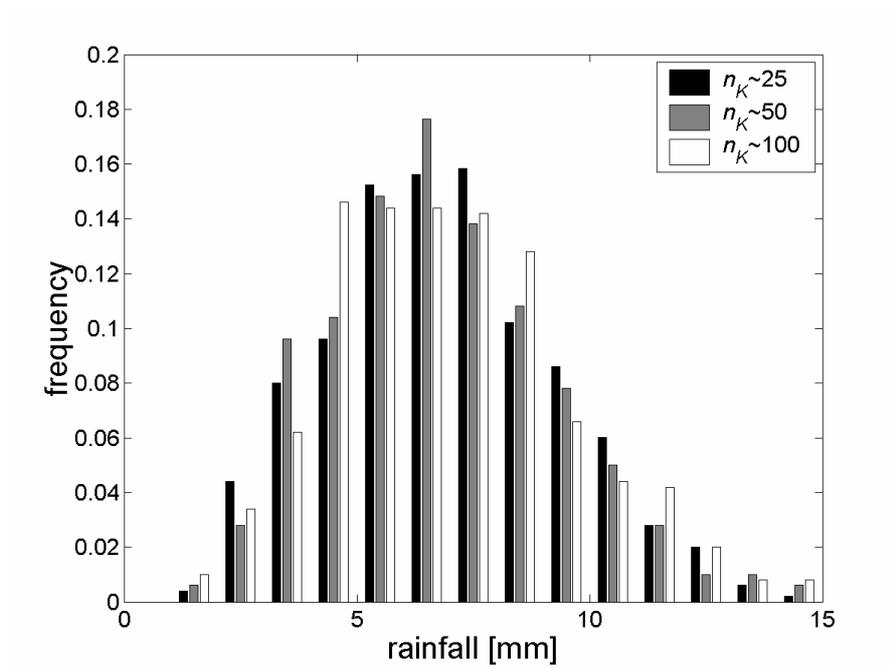


Figure 5.5 Frequencies histogram of mean rainfall over the 25 pixels by 25 pixels domain for the three set of simulations with different neighbourhood size ( $n_k$ ).

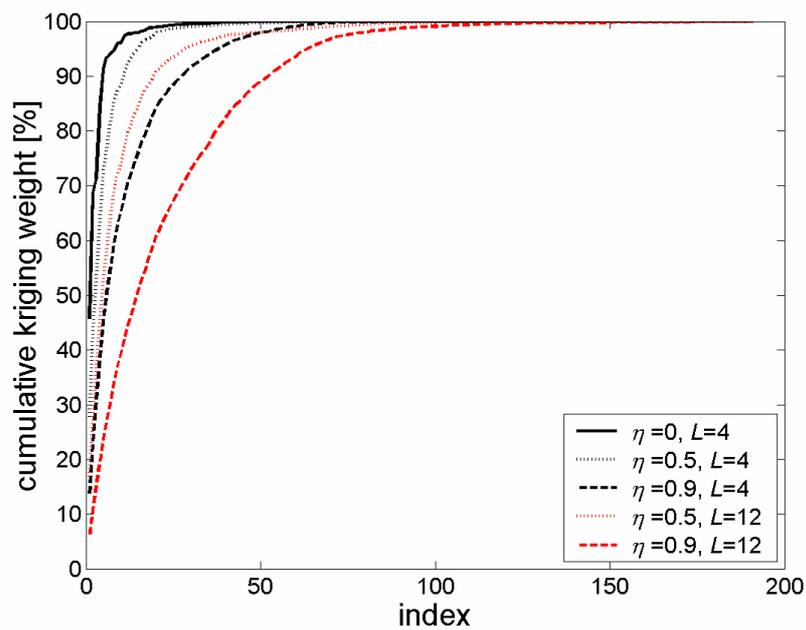


Figure 5.6 Plots of cumulative absolute kriging weights ratio for exponential semivariogram with various nugget ( $\eta$ ) and range ( $\sim 3L$ ).

To further consider the effect of a limited  $n_K$  on kriging estimates, it is useful to note that the proportion of kriging weights attributed to each of the sampled locations depends on the range, and the nugget-to-sill proportion of the semivariogram, as well as the spatial distribution of the sampled location. Figure 5.6 shows the contribution of kriging weights for the variograms of various nuggets and ranges in a Simple Kriging system. The sampled locations are indexed in ascending order from the nearest to the furthest from the target location. When the semivariogram with zero nugget and short range (4 units), screening effect is most pronounced with the nearest 10 sampled locations contributing to about 90% of the total kriging weights. Increasing the nugget-to-sill ratio and the range generally distributes the kriging weights more to the further sampled locations: For the case with semivariogram of nugget-to-sill ratio of 0.9 and a range of 60 units, about 60 of the sampled locations are required to contribute 90% of the total kriging weights. Overly restrictive search neighbourhoods can result in conditional biases in the kriging estimates (Krige, 1994, 1996a,b). This implies that in practice, a suitable search neighbourhood is one that is large enough to ensure sufficient accuracy, but small enough to be computationally efficient.

## 5.4 Simulation of daily Gambian rainfall by sequential methods

Simulations of daily Gambian rainfall were carried out using the available CCD data from 1996 to 2002 for the rainy months from June to October, using the SIMU algorithm described in Section 5.3. The CCD image for this work is bounded from  $13^{\circ}$  N to  $14^{\circ}$  N in latitude and  $13^{\circ}$  W to  $16.9^{\circ}$  W in longitude, covering the entire country with pixel resolution at  $0.05^{\circ} \times 0.05^{\circ}$ . See Figure 3.3 of Chapter 3 for an example of the domain for which daily rainfall fields were simulated.

### 5.4.1 Covariance Modelling

The covariance model for both indicator residuals  $R(\mathbf{u})$ , the normal transformed rainfall residuals  $Y(\mathbf{u})$ , needed in the simulation were derived from their respective semivariogram models. Monthly climatological semivariograms of  $R(\mathbf{u})$  and  $Y(\mathbf{u})$ ,

denoted by  $\gamma_R(h;m)$  and  $\gamma_Y(h;m)$ , respectively are required.  $m$  is included in the semivariogram notation to emphasise the dependence on the month  $m$ . Since there are no actual pixel-averaged rainfall observations,  $Cov(R_p, R_q)$  cannot be inferred through the construction of sample covariance in the usual sense. As a first approximation, the pixel level indicator values were obtained from the gauge-estimated pixel rainfall described in Chapter 3. Next, the climatological semivariogram,  $\gamma_R(h;m)$  was modelled from the sample semivariogram of  $r_j^* = I_j^* - p_j$ , where  $I_j^*$  is the estimated pixel indicator and  $p_j$ , the prior probability, was estimated as the sample proportion of rain from pixels within the same month and having the same CCD. Only pixels containing gauge observations are used in computing the sample semivariance of  $r^*$ . An exponential model was fitted for each month using weighted non-linear least square (Pardo-Igúzquiza, 1999) as shown in Figure 5.7. The corresponding normalised climatological semivariogram,  $\hat{\gamma}_R^*(h;m)$  was obtained by dividing the modelled semivariograms by their respective sill. The normalised covariance was obtained as  $\hat{C}(h;m) = 1 - \hat{\gamma}_R^*(h;m)$  assuming stationarity in  $R(\mathbf{u})$ .

Semivariograms of the normal transform of non-zero rainfall residuals were likewise modelled using the gauge estimated pixel rainfall: For a given month, the rainfall residual for a pixel  $j$  with coincident gauge observation were obtained as  $w_j = f_j - \mu_{+j}$ . Here,  $\mu_{+j}$  denotes the average of the non-zero rainfall amount associated with pixels having the same CCD as pixel  $j$  within the month. The set of transformed rainfall residuals,  $y$ , was then obtained from  $w$  by normal score transform (Goovaerts, 1997, Section 7.2.2), and the sample semivariance was then computed. The monthly semivariograms of  $Y$  is fitted using the spherical model (Goovaerts, 1997, Section 4.2.1) as shown in Figure 5.8.

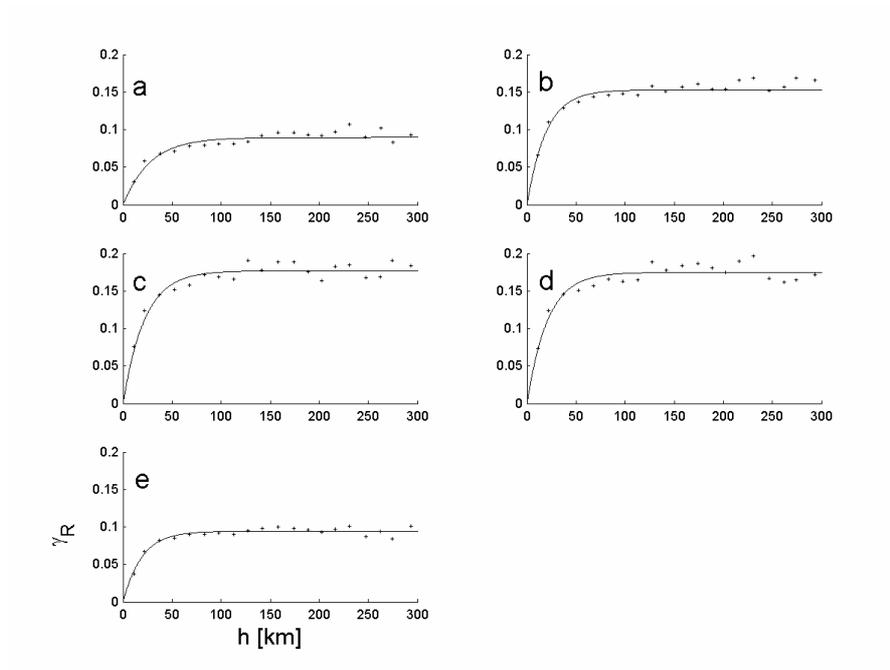


Figure 5.7 Climatological indicator residual semivariogram and fitted model for the different months. a) June; b) July; c) August; d) September; e) October

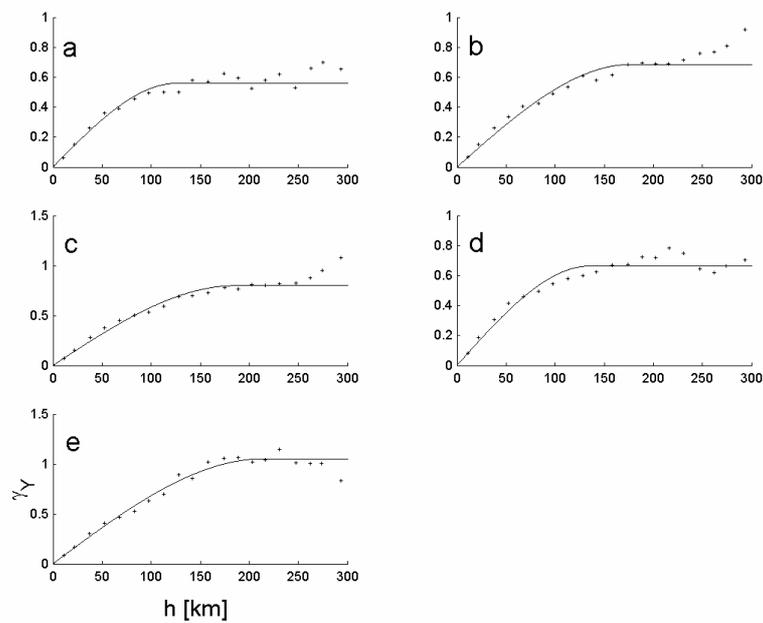


Figure 5.8 Climatological normal-transformed rainfall semivariogram and fitted model for the different months. a) June; b) July; c) August; d) September; e) October

### 5.4.2 Implementation specifics

Daily rainfall at METEOSAT CCD-pixel level ( $0.05^0 \times 0.05^0$ ) bounded within 13 N to 14N and 13.6W to 16.9W, was simulated from available daily CCD images. The maximum number of kriging weights,  $n_K$ , was set to 50. From the discussion from Section 5.3.4,  $n_K = 50$  is a reasonable value taking into account the low nugget-to-sill ratio for both  $\gamma_R$  and  $\gamma_Y$ , where the screening effect should be significant.

Due to the amount of simulation that is required (1071 days in total), the recommended algorithm described in Section 5.3.3 is still too computationally expensive with the available resources. In order to accelerate the simulations, the non-zero rainfall field  $F(\xi)$  within the domain was produced using sGs assuming all pixel were rainy. The final rainfall,  $Z(\xi)$ , was taken as product of the simulated rain/no-rain mask (from sIs) and positive rainfall field,  $Z(\xi) = I(\xi) \times F(\xi)$ . In this way, the same set of kriging weights can be reused in both the sIs and sGs algorithms after the first cycle of simulation. This was found to dramatically reduce computational overheads in practice, even though it involves synthesising non-zero rainfall over all pixels that would eventually be masked out. For each day, 250 realisations of rain/no rain maps were obtained using sIs. For each of these realisations, 2 corresponding rainfall realisations were obtained using sGs, resulting a total of  $N=500$  simulations per day. The numbers of sIs and sGs simulations employed (250 and 500 respectively) uses the results of Section 5.3.4 as a guide (See Figure 5.3).

Figure 5.9 shows several realisations of rainfall field from SIMU for a selected day. A necessary condition for the validity in the various parameters, as well as the modification to the simulation algorithm adopted, is that the ensemble of synthesised rainfall fields for the selected day must be able to replicate the primary statistical constrains imposed within the simulation algorithm. These include the statistical relations of CCD and pixel rainfall defined by TAMSAT1, as well as the two-point statistics of the rainfall field defined by the semivariograms of the probability residuals and normal transformed non-zero rainfall residuals. To test the agreement in pixel level statistics between the

simulations and those predicted by TAMSAT1, the mean and variance of pixel rainfall for non-rainy pixels as well as the probability of pixel being rainy were computed for all CCD levels available for the selected day. The sample semivariograms for the probability residuals and normal transformed non-zero rainfall residuals were computed using the methods described in Section 5.4.1 from all available pixel rainfall within the ensemble. From Figure 5.10 it can be seen that the statistics from the simulated rainfall were in good agreement with TAMSAT1 as well as the semivariograms model imposed, lending support to algorithm and the chosen parameters.

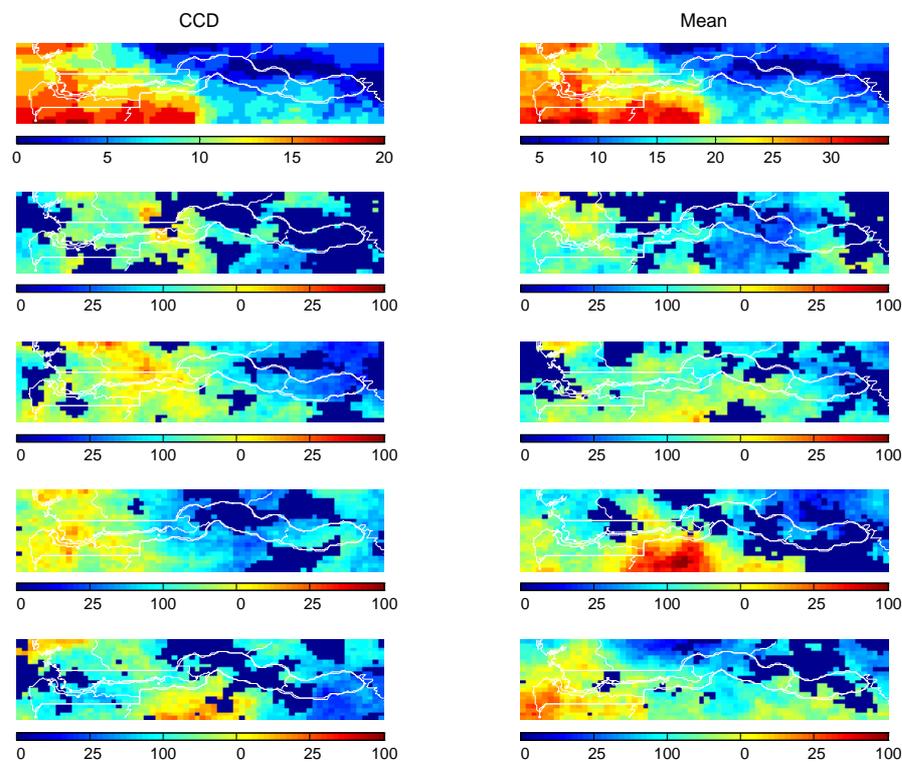


Figure 5.9 CCD image and realisations from simulating Gambian rainfall on 7<sup>th</sup> Aug 2001 using SIMU. Top two images are the daily CCD and ensemble mean rainfall, followed by 8 realisations from the simulation. Unit: CCD [0.5hr]; rainfall [mm].

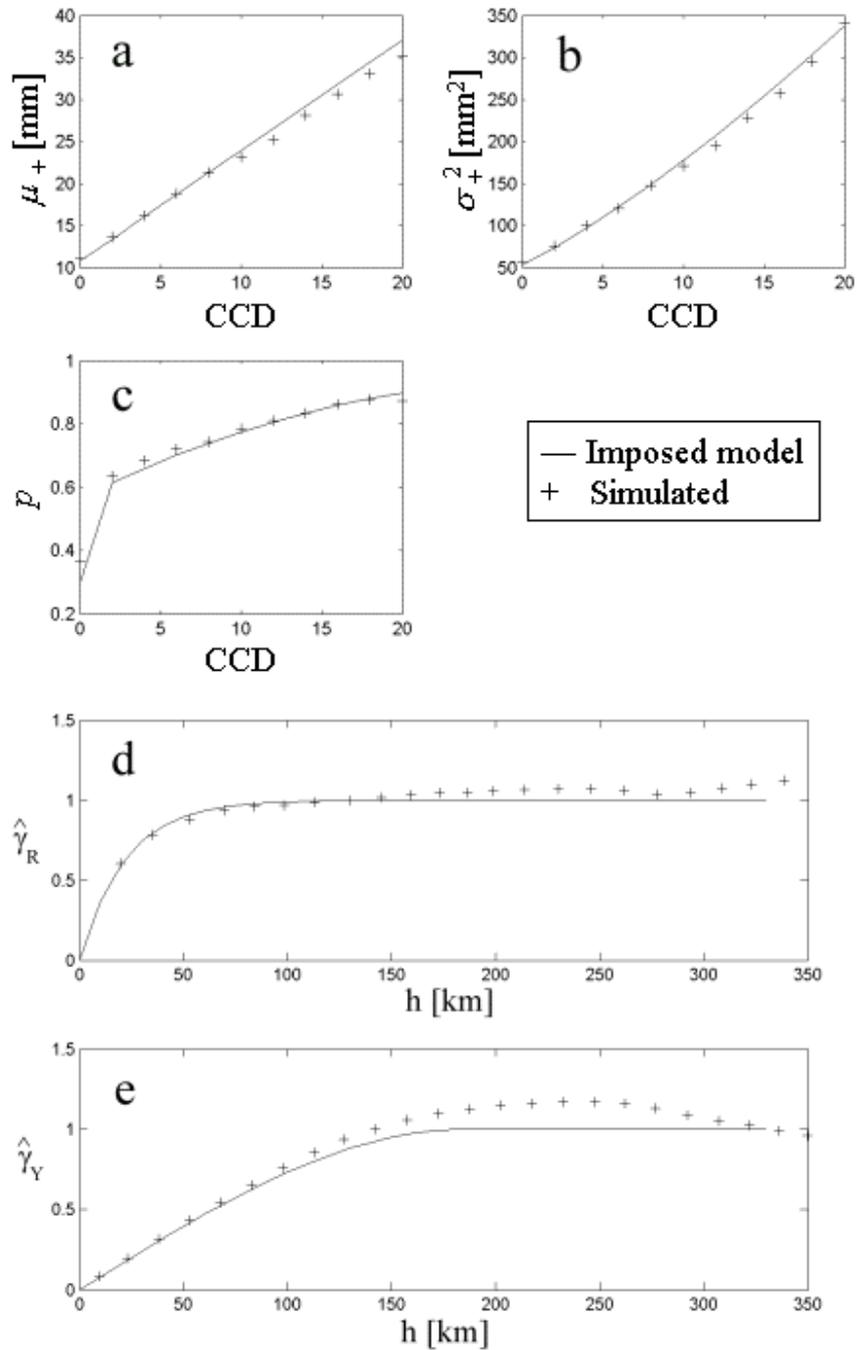


Figure 5.10. Internal consistency of simulations for the same day as Figure 5.9 by comparing pixel rainfall characteristics with TAMSAT1, and sample variogram from simulation with semivariogram models used. a) Mean positive rainfall; b) Variance of positive rainfall; c) Probability of rain; d) Normalised indicator semivariogram; e) Normalised normal-transformed rainfall semivariogram.

### 5.4.3 Validation

The previous section demonstrated the internal consistency in reproducing the constraints by the SIMU algorithm, namely the pixel level rainfall distribution (TAMSAT1), the indicator, and normal transformed rainfall variograms. It remains a question of how valid the algorithm is when compared to the gauge-based estimated rainfall. A validation experiment has been conducted to test the ability of the SIMU algorithm to replicate the daily pixel rainfall distribution for the Gambia, and the spatial structure of the rainfall field as obtained from gauge-estimated pixel rainfall.

Daily rainfall for the Gambia region has been generated using SIMU from 1996 to 2002 for the month of June through October. Implementations details of the simulations of Gambian daily rainfall are described in Section 5.4.2. Although CCD data from 1988 was available, the years from 1996 to 2002 were chosen for completeness of data. A leave-one-year-out cross-validation has been adopted for calibrating the TAMSAT1 model parameters used within the SIMU algorithm by leaving out the validating year, but included all available years going back to 1988 in the calibration. Although it is also necessary in principle to obtain separate climatological variograms for the residuals  $R(\mathbf{u})$  and  $Y(\mathbf{u})$  for each validating year in the same manner, it was found that they remain stable across all validating years. Hence a common set of climatological variograms (5 monthly variograms for each residuals fields) was used throughout the validation for ease of implementing the cross-validation. For the purpose of validation, only pixels containing valid raingauge observation were used. This applies to both the observed data as well as the simulated data. The results of the validation are discussed below.

#### Rainfall distribution at pixel scale

The distribution of daily pixel rainfall was evaluated by comparing rainfall probabilities at various rainfall thresholds using the procedure adopted from Bellerby and Sun (2005). In this method of assessment, the exceedence probabilities - probability of rainfall larger

or smaller than a given threshold<sup>2</sup> ( $P(Z > 5\text{mm})$ , for instance) - are computed for the set of simulated daily rainfall at the validating pixels, everyday. These probabilities are then binned, and the pixels associated with each probability bins are identified (with the date and the pixel locations). For each of these probability bins, the corresponding exceedence probability is computed from the sampled rainfall distribution, constructed from the gauge-estimated rainfall associated with pixels identified with the probability bin. In the ideal situation where the rainfall simulation is perfect, the exceedence probabilities computed from the gauge-estimated rainfall equal the corresponding bin values.

The results from implementing the method outlined above for different rainfall amount thresholds are shown in Figure 5.11. From the scatter plots of the exceedence probabilities for the simulated and the observed rainfall, it can be seen that the distribution of the rainfall is reasonably reproduced for daily rainfall for a broad range of rainfall threshold. Simulated rainfall however appears to reproduce a higher exceedence probability for the larger rainfall threshold. This implies that simulated rainfall distribution has a longer tail than the observed.

#### Spatial structure of daily rainfall field: Pixel level

As mentioned previously, one problem faced in raingauge-sparse Gambia is that it is not possible to infer daily variograms of the rainfall from the observed rainfall. Here, the spatial structure was assessed based on the assumption that the normalised climatological variograms defined the spatial structure of the daily rainfall (Lebel *et al.*, 1987). The simulations had shown to reproduce the variogram of the  $Y(\mathbf{u})$  and  $R(\mathbf{u})$  at the beginning of this section. A more stringent assessment is to compare the variograms of the positive rainfall at pixel scale,  $F(\mathbf{u})$ , and the indicator transform,  $I(\mathbf{u})$ , determined from the observed data with those obtained from simulations.

---

<sup>2</sup> The term “exceedence probability” takes a wider meaning in the present context than that of Bellerby and Sun (2005), by including the probability of rainfall *smaller* than a given threshold, within its definition.

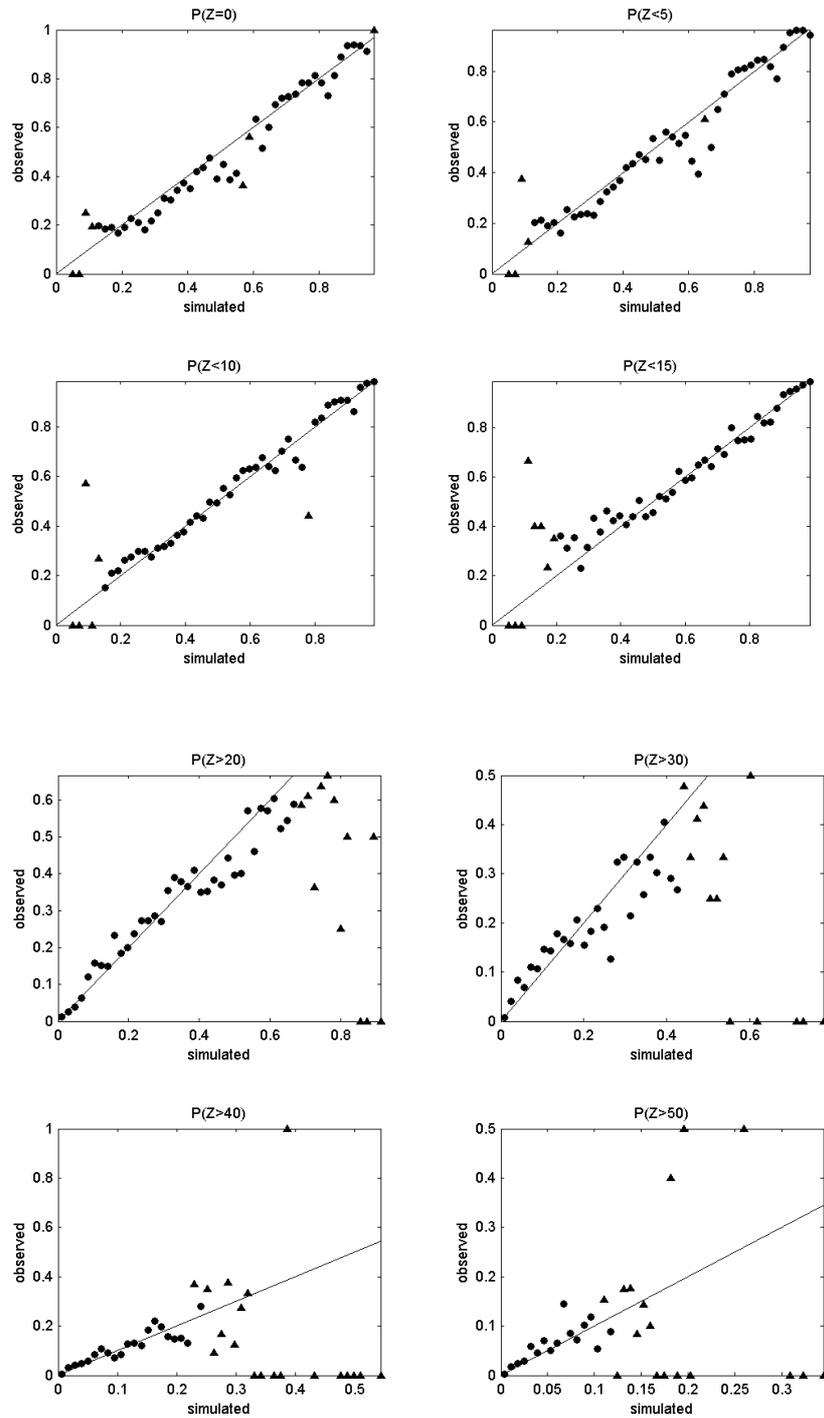


Figure 5.11 Comparison of pixel rainfall probabilities of exceedence at various rainfall amount thresholds between simulations and the corresponding gauge estimated from 1996 to 2002. Solid line is the 1-1 line. Legend: ●:  $\geq 30$  observed gauge estimates; ▲:  $< 30$  gauge estimates.

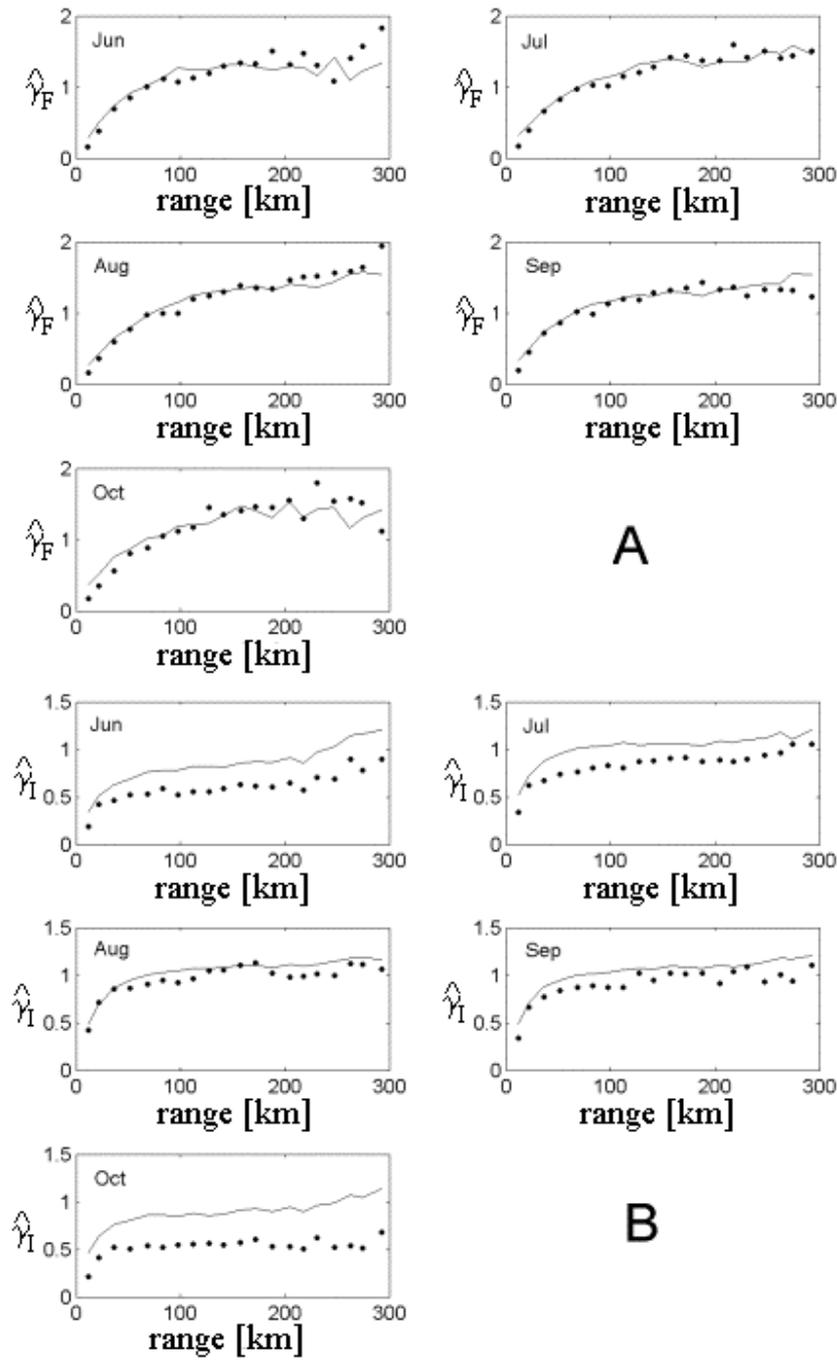


Figure 5.12 Standardised monthly semivariogram computed from simulation (solid lines) and gauge estimates (filled circles). A) Positive rainfall amount variogram; B) Indicator variogram using all data. C) Indicator variogram omitting LDZ cases.

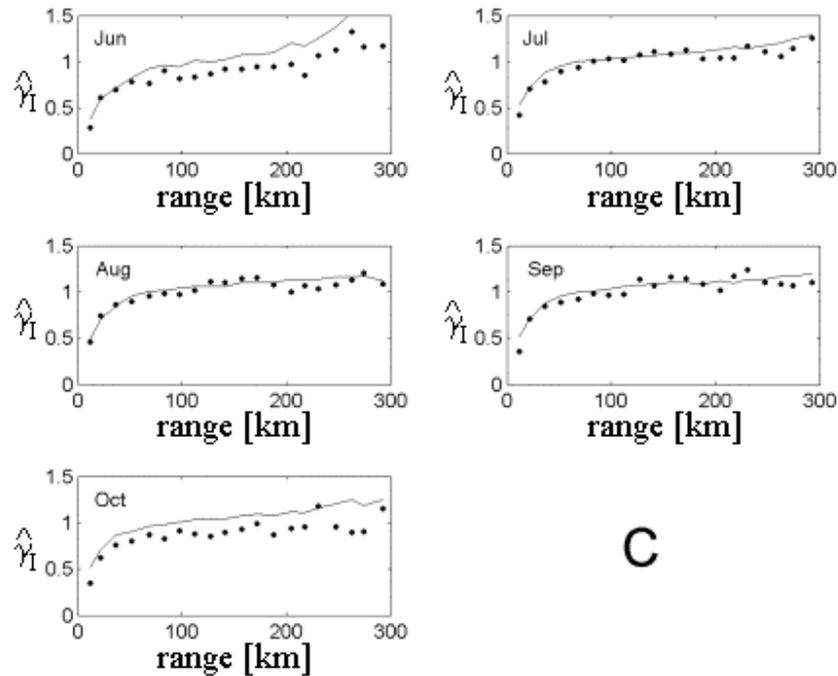


Figure 5.12 (Continue)

Monthly standardised sample variograms of  $F(\mathbf{u})$  and  $I(\mathbf{u})$  from June and October for Gambian daily rainfall, were computed using the gauge estimated pixel rainfall. Only pixels containing gauge observations were used in the computation. From the generated rainfall fields, 10 realisations of each monthly variograms were obtained. Each sample variogram realisation was computed using equal amount of data as used in computing the observed variogram.

Figure 5.12 compares the observed sample variograms with the average of the sample variogram realisations obtained from the simulation. The rainfall amount variograms were well reproduced for all months. The observed indicator variograms for June, July and October have a lower sill than those computed from the simulated rainfall fields. The primary reason for the difference of these indicator variograms can be traced to the fact that the TAMSAT1 rainfall estimates tend to overpredict the probability of rain when  $CCD = 0$  over a large fraction of the area – a situation which occurs frequently at the beginning and end of the rainy seasons. For such cases, the simulations consequently

overestimated the rain coverage. Indeed, by using the definition of Large Domain Zero (LDZ) as 95% of the pixels CCD containing within the domain of interest (see Chapter 4, Section 4.5), and filter out case of such extended zero CCD days, the monthly indicator variograms determined from simulated rainfall field agrees well with the observed as shown in Figure 5.12c. It is noted that although the effective range of the residual indicator used in the simulation is about 75 km, the simulated indicator variograms were able to reproduce both short and long range effects of the observed variograms – most obvious in the July indicator variogram in Figure 5.12c.

#### Spatial structure of daily rainfall field: Larger scale structures

A consequence of the spatial structure of rainfall is a proportional relationship between the mean areal rainfall amount and rainfall coverage at larger spatial scale than the scale at which the rainfall field is simulated. This relationship is evident for the Gambian daily rainfall (Chapter 2, see Figure 2.6). Such relationship often seen in observed rainfall (Doneaud *et al.*, 1981; Kedem and Pavlopoulos, 1991) is not explicitly defined within the stochastic rainfall model but should manifest itself from the simulations. Therefore it can be used to further validate the ability of the simulations to reproduce the observed rainfall spatial structure (Onibon *et al.*, 2004).

Daily ALL GAMBIA mean rainfall and the corresponding fractional coverage with rainfall greater than a given rainfall amount threshold were computed from both the observed and simulated rainfall fields following Onibon *et al.* (2004). For both data, the areal mean rainfall and coverage were computed using only pixels with coincidental gauge observations. The scatterplots of the ALL GAMBIA rainfall against area coverage for different rainfall thresholds are shown in Figure 5.13. Visual inspection of the plots and rain area coverage-rainfall correlations (Table 5.1) show good agreement between simulated and observed for all prescribed rainfall threshold.

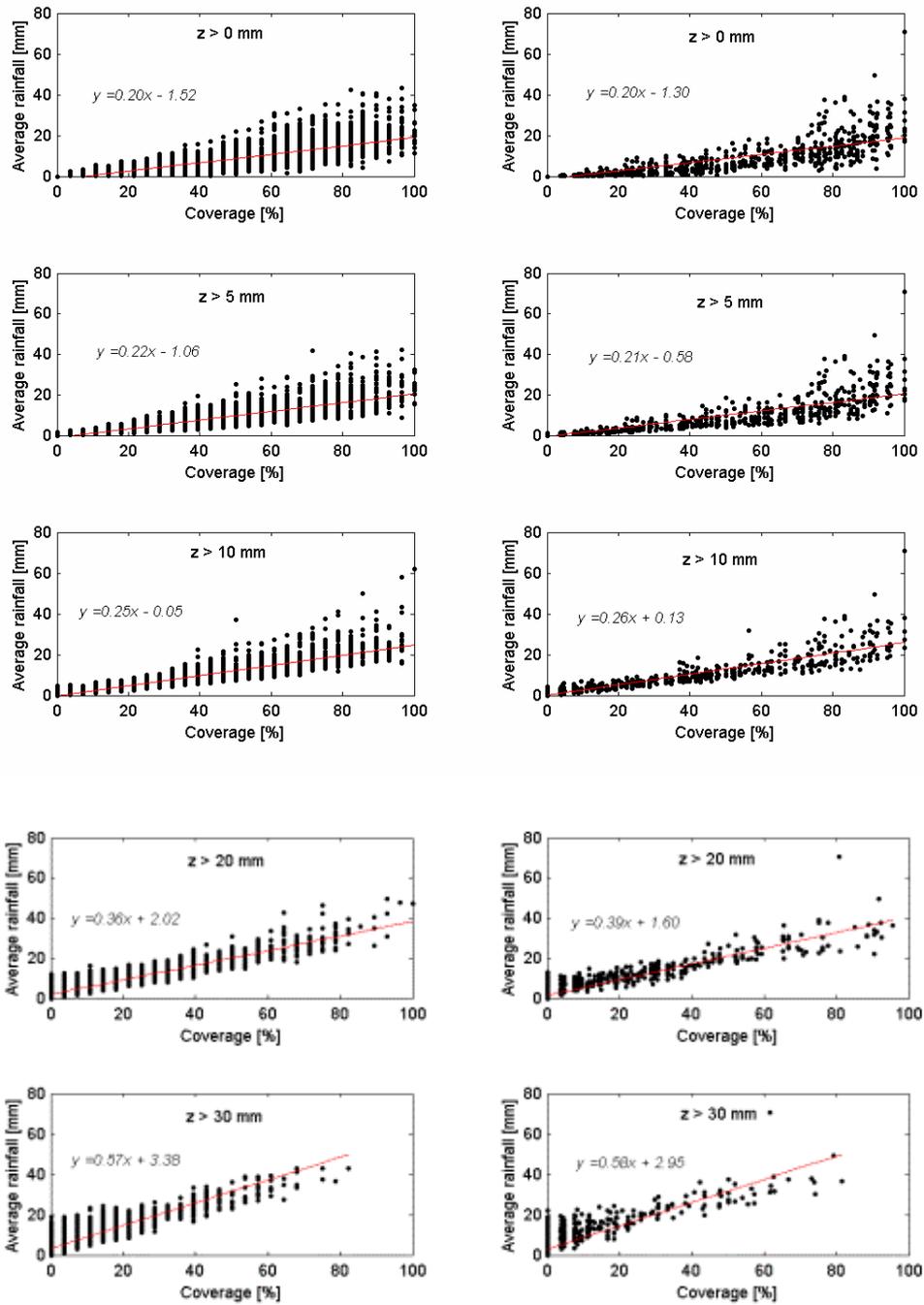


Figure 5.13 Scatterplots of ALL GAMBIA rainfall and fractional coverage of rainfall above various thresholds. Left column: Simulated; Right column: Observed. Red line: Linear regression line.

Threshold [mm]	$r^2$	
	Simulated	Observed
0	0.70	0.68
5	0.78	0.75
10	0.87	0.86
20	0.87	0.87
30	0.72	0.71

Table 5.1 Correlations of ALL GAMBIA rainfall and fractional coverage of rainfall beyond various thresholds.

### Inter-year comparison

The performance of the stochastic model has also been validated using monthly rainfall totals, determined by totaling ALL GAMBIA daily rainfall (see above paragraph) for each month in a given year. Correlation ( $r^2$ ) between ensemble mean of the simulated and observed monthly ALL GAMBIA rainfall is 0.76. A lower than expected value of 48% of the observed, lies within the 10<sup>th</sup> and 90<sup>th</sup> percentile of the prediction (Figure 5.14). The discrepancies can be attributed primarily to the errors in the TAMSAT1 model parameters: Biases in TAMSAT1 parameters affects the mean pixel rainfall, effectively shifting the centroid of ensemble away from the observed mean value. The three months where simulations deviate most from the observed are from 1999 and 2002 where TAMSAT1 have the largest biases in magnitude amongst the validating years (Table 4.5, Chapter 4). For 2002, such bias in the TAMSAT1 parameters consequently results in the large overestimation of the monthly mean areal rainfall for July and August (Chapter 4, Section 4.5). For August 1999, in addition to the systematic biases in TAMSAT1 model parameters, one day of anomalous high rainfall of 70 mm in the observed data also contributes to the large disparity between the observed and the simulated rainfall amount.

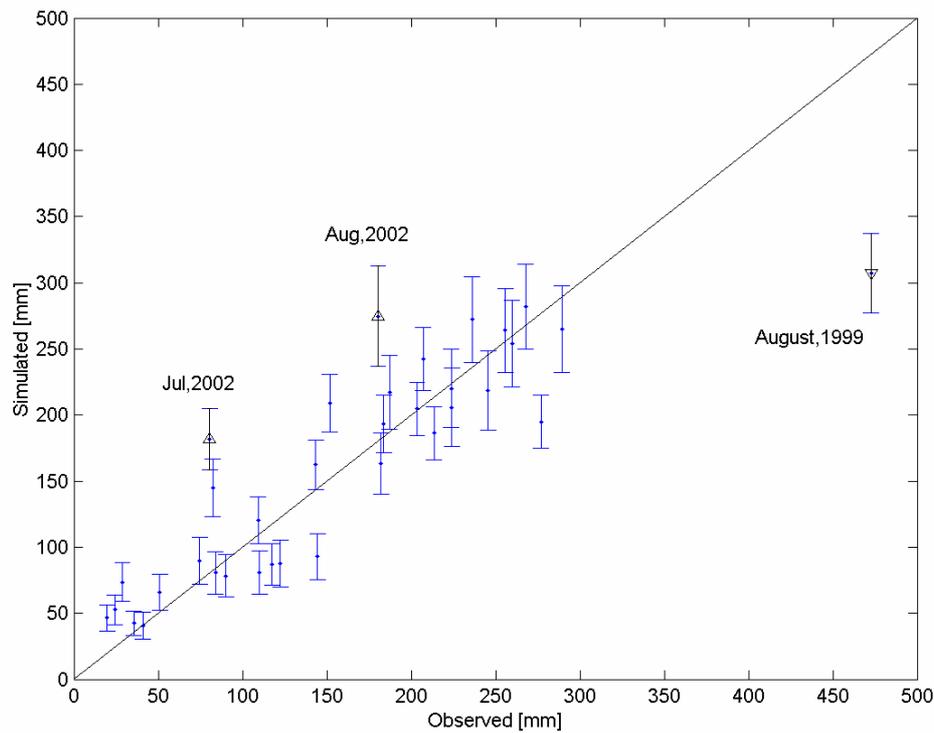


Figure 5.14 Scatter plot of ALL GAMBIA monthly total rainfall for 1996 to 2002. Error bar indicate the 10<sup>th</sup> and 90<sup>th</sup> percentile of rainfall obtained from the simulation. Solid Line is the 1-1 line.

## 5.5 Possible improvement

For days when there are extended regions of zero CCD, TAMSAT1 tends to overestimate the probability of rain at pixel level (Chapter 4, Section 4.5). This caused the SIMU algorithm to generate too large a rainfall coverage in such a situation, which resulted in the discrepancies in the indicator variograms obtained from the simulated and the gauge-estimated rainfall (Figure 5.12). It is therefore of interest to investigate the improvement to SIMU if the underlying pixel rainfall distribution (ie. TAMSAT1) is modified to account for LDZ situations (Chapter 4, Section 4.5). The modified algorithm abbreviated as TAMSAT1-LDZ hereafter, is identical in form with TAMSAT1 (Equations 4.5a-d) for days with sufficient nonzero CCD pixels (not LDZ situations). For days satisfying the LDZ condition, an additional estimation of the probability of a pixel being rainy is used. To avoid introducing too many variables into the satellite-rainfall relationship, the rainfall

amount within the rainy regions during LDZ days was determined by Equations 4.5b-d. This should not affect the realism of the simulation since the rain coverage on such days is generally low. Normalised variograms for  $Y(\mathbf{u})$  and  $R(\mathbf{u})$  were obtained as before, but with an additional set of monthly variograms for LDZ days.

Simulations of the Gambian daily rainfall were repeated with the TAMSAT1-LDZ as the underlying satellite rainfall model at pixel level. Figures 5.15 and 5.16 show the results of the pixel rainfall distribution and variograms obtained from the generated rainfall using the SIMU algorithm with TAMSAT1-LDZ under a similar cross-validation procedure as before. The simulated rainfall field showed overall improvement when compared to the experimental variogram of  $I(\mathbf{u})$ , although the range of variogram for July was still underestimated. Similar performance was maintained in representing the spatial structure of  $F(\mathbf{u})$  when compared to simulations using TAMSAT1 (see Figure 5.12). Overall improvement in reproducing the extended region of zero rainfall by using TAMSAT1-LDZ in the stochastic rainfall model was also observed when compared to the simulated distribution of ALL GAMBIA rainfall with those determined from gauge estimates. Plots of the exceedence probabilities for various rainfall thresholds obtained from simulation and observed ALL GAMBIA daily rainfall from 1996 to 2002 are shown in Figure 5.17. The ability in simulating the observed distribution of ALL GAMBIA rainfall using TAMSAT1-LDZ is similar to that using TAMSAT1 for most of the range of rainfall amount. However, simulations using TAMSAT1-LDZ reproduced a larger range of probabilities of no rain compared to those using TAMSAT1.

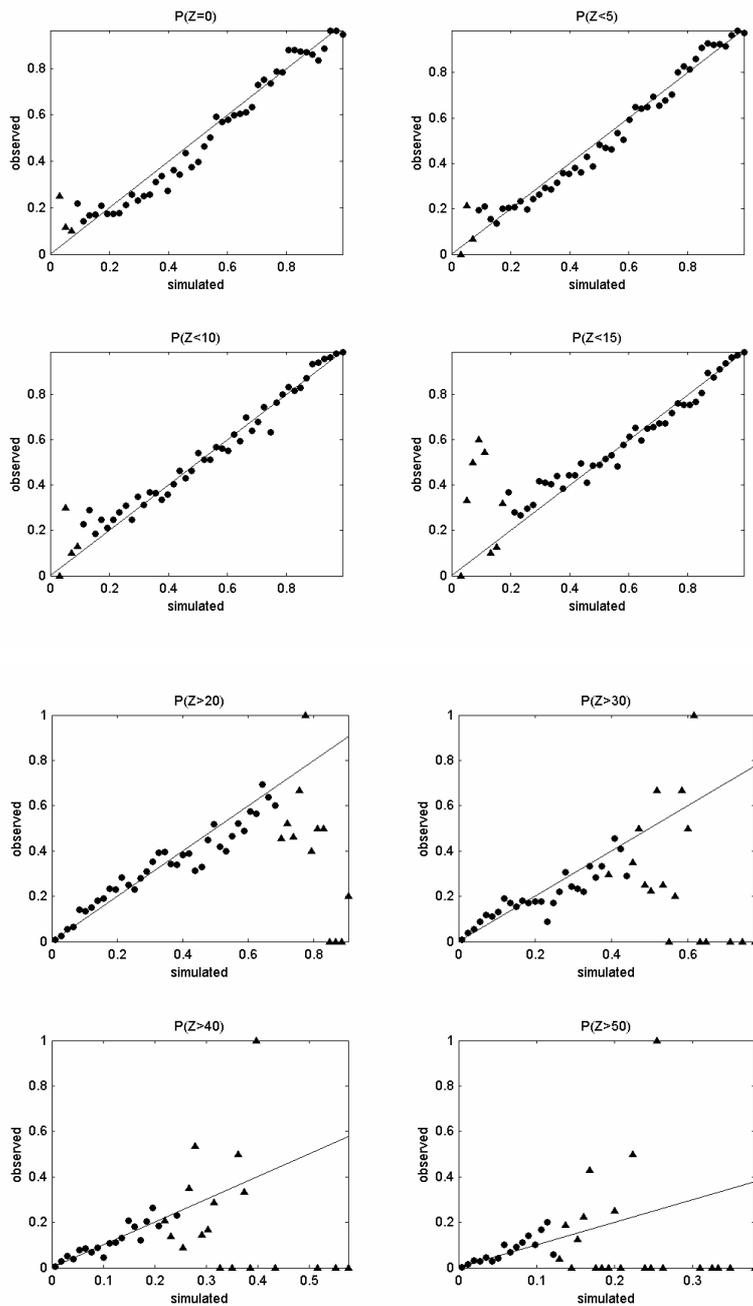


Figure 5.15 Comparison of pixel rainfall probabilities at various thresholds between simulations using TAMSAT1-LDZ from 1996 to 2002 and the corresponding gauge estimated. Solid line is the 1-1 line. Legend: ●:  $\geq 30$  observed gauge estimates; ▲:  $< 30$  gauge estimates.

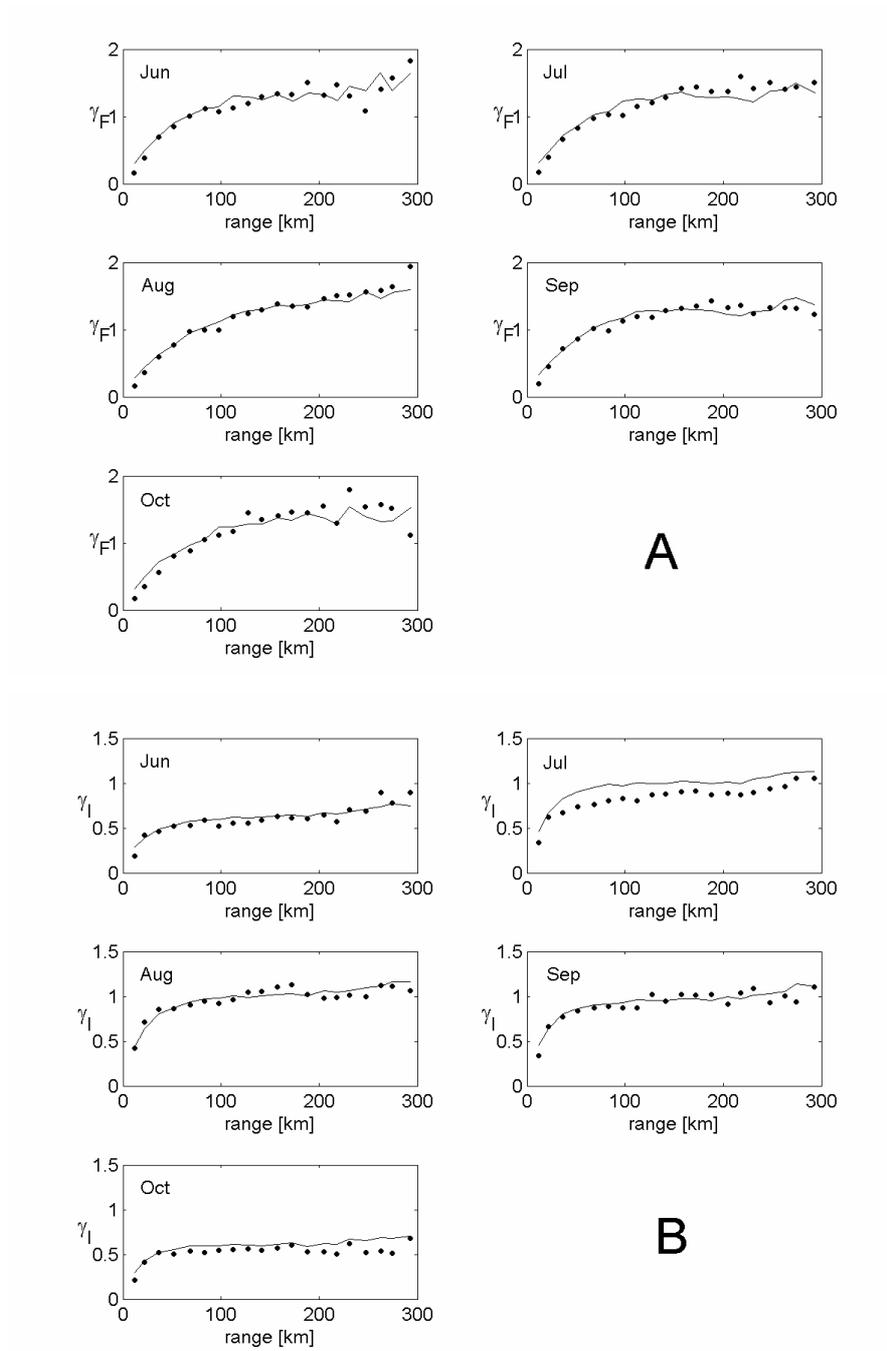


Figure 5.16 Standardised monthly semivariogram computed from simulation using TAMSAT1-LDZ (solid lines) and gauge estimates (filled circles). A) Positive rainfall amount variogram; B) Indicator variogram using all data.

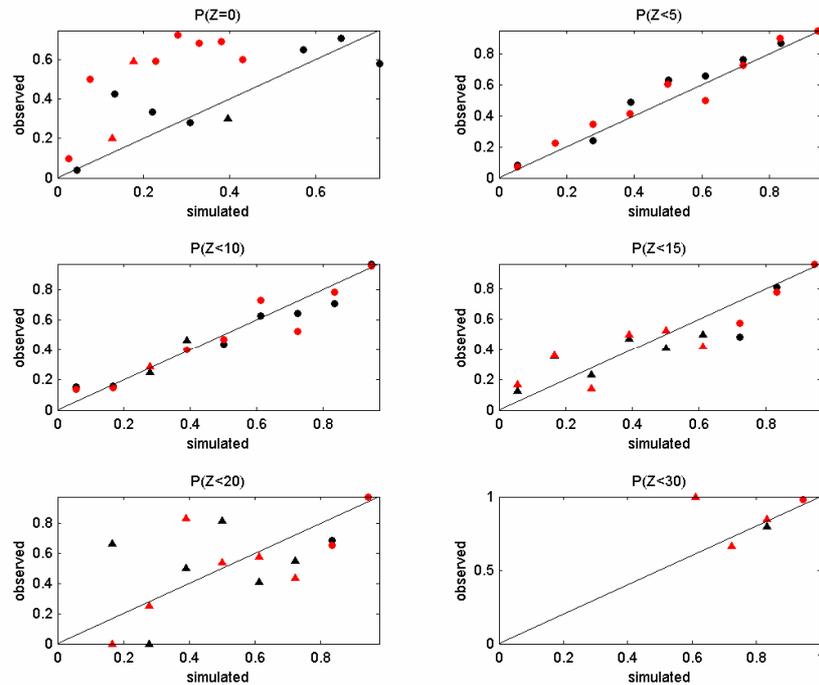


Figure 5.17 Comparison of ALL GAMBIA rainfall probabilities at various thresholds between simulations from 1996 to 2002 and the corresponding gauge estimated. Solid line is the 1-1 line. Legend: ●:  $\geq 30$  observed gauge estimates; ▲:  $< 30$  gauge estimates. Red: TAMSAT1; Black: TAMSAT1-LDZ.

## 5.5 Conclusion

A stochastic method of generating intermittent rainfall fields conditional to METEOSAT-CCD has been developed based on a dual sequential simulation methodology that generates the fractional rain coverage and the positive rainfall amount separately. The proposed model was used to reproduce daily rainfall fields for the Gambia.

The distribution of pixel rainfall from the simulations agreed well with that from gauge-estimated pixel rainfall. Although the spatial structure of the rainfall was defined only by the residual variograms of the normal transformed rainfall amount and rainfall indicator transforms, the model nevertheless, was able to reproduce short-range and long-range features observed in the spatial structures of actual rainfall and intermittencies, as well as

rainfall-coverage characteristics at aggregated spatial scales. However, at the beginning and end of the rainy season when large-scale intermittencies are frequent, SIMU tends to over-simulate the rainy area. The reason was due to TAMSAT1 overestimating the probability of rainfall for days when CCD is zero over large spatial extent. By modifying the TAMSAT1 algorithm to account for days with extended area of zero CCD, the simulated spatial structure of the rainfall intermittencies improved.

One unresolved issue in SIMU is the lower than expected skill in capturing the monthly mean areal rainfall. Since the calibration of the underlying pixel rainfall distribution model (TAMSAT1) uses climatology, SIMU implicitly assumes a certain degree of stationarity in the rainfall process. As interannual variability is high, stationarity may be violated for anomalous rainfall years. In such cases, bias in TAMSAT1 would inevitably be transferred to the simulated rainfall fields. These systematic errors cannot be significantly reduced without introducing real-time or near real-time rainfall observations.

As SIMU relies mainly on a pixel-level rainfall distribution conditional to the satellite CCD, it can be adapted with other more suitable satellite-based pixel rainfall distribution model other than TAMSAT1 if required. Finally it should be highlighted that although the discussion in this chapter assumes no real time data is available, additional information can be incorporated into SIMU. For example, daily raingauge data whenever available, can be directly assimilated (after spatially averaging to satellite pixel scale) without any alternation to the simulation algorithm. In this case, pixels with the real-time rainfall information can be used to initialise the sIs and the sGs algorithms in SIMU instead of the 'seed' pixels. It would be interesting to compare the performance of SIMU in such situations to other stochastic rainfall models that are able to assimilate multisensor data, such as those proposed by Lanza (2000) and Fiorucci *et al.* (2001).

# Chapter 6 Application of satellite rainfall estimates in crop yield estimation

## 6.1 Introduction

Utilisation of satellite remote sensing in agriculture in Africa has been a subject of intensive research. This is primarily due to the large spatial coverage and good temporal resolution of satellite information, which complements ground-based agrometeorological observations for the monitoring of crop growth. This data is vital for many regions of Africa because of the sparse and often poorly maintained hydrometeorological networks (Washington *et al.*, 2006). National, regional and international agencies concerned with food securities in the continent have adopted satellite-based information with various levels of sophistication, for the in-season monitoring of crop growth as well as crop production forecasting for the planning of food relief operations, when necessary (Rijiks *et al.*, 2003; Verdin *et al.*, 2005).

Satellite data is used in crop monitoring in several ways. Satellite-based vegetation indices, in particular, the Normalised Difference Vegetation Index (NDVI) data derived from the Advance Vertical High Resolution Radiometer (AVHRR) of the NOAA polar orbiting satellites have been used for crop yield prediction in Africa based on NDVI-yield regression models (Rasmussen, 1992; Groten, 1993; Maselli *et al.*, 1993, 2000; Mkhabela *et al.*, 2005). However, there are some difficulties in using NDVI. These include data loss due to cloud cover, data response to atmospheric moisture and spatial heterogeneity, calibration mismatches between satellites and within sensor calibration drift (Holben, 1986; Gutman, 1991, 1999; Justice *et al.*, 1991, Maselli *et al.*, 1993; Hayes and Decker, 1996; Baez-Gonzalez, 2002). Hence, careful post-processing of satellite data is required before applying it to crop forecasting (Groten, 1993; Maselli *et al.*, 2000; Mkhabela *et al.*, 2005).

As reliable satellite-based rainfall estimates are available over much of the continent, using satellite-based rainfall estimates in weather-driven crop models provides an attractive alternative for crop yield forecasting for Africa, especially for regions where crop yield is water-limited and ground measurement of rainfall is not available in a timely fashion. One such method of crop yield forecasting uses the crop water requirement satisfaction index (WRSI) (Frère and Popov, 1986) as an indicator of crop development. In contrast with vegetation indices, which measure crop growth through inference from the electromagnetic signatures of the vegetation itself, the WRSI is based on the crop specific water balance model (CSWB) developed by Frère and Popov (1979) that accounts for the crop water satisfaction during the crop growing season. Incorporating satellite rainfall estimates from METEOSAT in the CSWB for large-scale crop yield prediction in Africa have been attempted by Reynolds *et al.* (2000), Verdin and Klaver (2002) and Senay and Verdin (2003). As the CSWB does not involve simulating detailed crop development processes, it is a relatively simple model. The advantage of this is that the model requires few parameters and hence easy to implement. The disadvantage is that the model outputs are crop water usages rather than crop yield. An empirical relation that links model outputs to crop yield is necessary.

An alternative method of using satellite-based rainfall estimates for crop production prediction is to assimilate this information into process-based crop models that are able to simulate crop processes. These are more complex models, which in theory, are capable of capturing more subtleties in the intraseasonal variability of weather that may elude the more empirical CSWB models. Descriptions of many different process-based crop models are available in the literature. A good review can be found in Matthews *et al.* (2000). Thornton *et al.* (1997) demonstrated the potential of using satellite RFE and other ground-based data in a mechanistic crop model (CERES-millet) for millet yield prediction in Burkina Faso. Although the potential of such crop models has been demonstrated, the study also highlighted the overhead of assembling the necessary large amount of input information.

An additional issue in using these mechanistic crop models in operational crop yield forecasting is that most of these models are designed for small homogeneous spatial units (field or plot). Therefore, they are sensitive to spatial heterogeneity in agrometeorological conditions, soil types and management practices. Consequently, exhaustive spatial information is required to implement these crop models for yield prediction over a large spatial extent. To apply such mechanistic crop models directly on a spatial scale larger than its original working scale by spatially upscaling model inputs, can result in bias in model prediction. This happens as a result of the non-linearity of model processes and the emergence of new processes due to the spatial interaction not accounted for within the models (Hansen and Jones, 2000; Faivre *et al.*, 2004). For Africa, regions in which such crop forecasting systems are vital often lack the spatially high-resolution data required for such crop models to be practical in operation. Hence, opportunities for applying these crop models for regional crop production prediction are at present limited.

The limitations of using process-based crop models in data-sparse region could potentially be overcome. Recent advances in assimilating satellite-derived Vegetation Index fully into mechanistic crop models (Moulin *et al.*, 1998) offers the potential of offsetting errors in simulated crop development due to uncertainties in input and model calibrations, by nudging some of the model crop state variables (example, biomass or Leaf Area Index) towards those derived from satellite imageries, in the course of the model simulation (Pellenq and Boulet, 2003). However, most of these investigations are largely confined to developed countries in Europe and the US. Another plausible solution for the problem of utilising process-based crop models in the absence of exhaustive crop model inputs is by model reduction; Roebeling *et al.* (1999) uses satellite derived solar radiation and evapotranspiration in the EARS-CGS crop model which does not rely on a soil hydrological sub-model to simulate crop transpiration. They reported satisfactory yield predictions for maize at provincial level in Zambia and Zimbabwe. Another crop model with potential for application for data scarce regions in Africa is The General Large Area Model for annual crops (GLAM) (Challinor *et al.*, 2004), a process-based crop model that is designed for large area application, combining the advantages of empirical models and mechanistic models to achieve low input requirement, while

maintaining the ability to simulate crop processes. GLAM has shown good performance in hindcasting of district-level rainfed groundnut yields for India (Challinor *et al.*, 2004, 2005a,b). However, its performance in Africa using satellite-based rainfall estimates as inputs has not been tested.

The main purpose of this chapter is to investigate the feasibility using GLAM for crop yield prediction for regions where crop production is water-limited, focusing primarily on the model's ability to predict crop yields with TAMSAT1 RFE as input. Yield predictions from the simpler crop model, based on the CSWB are used for skill comparison in yield prediction. The CSWB model is also useful as a reference when exploring the various aspects of rainfall-yield interactions in GLAM. The two most important groundnut production regions of the Gambia, North Bank Division (NBD) and Central River Division (CRD) were used as case studies.

As crop model yield sensitivity to inter-seasonal and intra-seasonal rainfall has been reported in various studies (Semenov and Porter, 1995; Mearns *et al.*, 1996; Riha *et al.*, 1996; Hansen and Jones, 2000; Challinor *et al.*, 2005a), another area to be explored concerns the crop model yield responses to errors in satellite RFE. Of particular relevance to the present discussion is the investigation by Challinor *et al.* (2005a), who reported improvement in GLAM yield prediction by correcting systematic bias in rainfall for the case of Indian groundnut. In the present case study of Gambian groundnut, several cases of GLAM crop processes with different rainfall inputs were analysed in detail to understand how crop model yield error can arise from differences in rainfall inputs. This analyses served to compliment those observations in Challinor *et al.* (2005a). To this end, a set of preliminary experiments has been conducted to gain a basic understanding on both GLAM and CSWB model yield responses to simple rainfall distributions. The observations from these preliminary experiments are used as a basis for analysing a few cases of satellite-based rainfall estimates and gauge-rainfall driven crop models. The effect of systematic bias in TAMSAT1 RFE on model yield are discussed through a detailed analysis of one year when bias in TAMSAT1 RFE leads to overestimation of NBD groundnut yield for both GLAM and CSWB. In addition, sensitivity of the

simulated yields from TAMSAT1 driven crop models due to stochastic uncertainties in TAMSAT1 RFE are quantified using ensemble crop yield prediction methods.

The remainder of this chapter is arranged as follows: Section 6.2 gives a brief description of GLAM and the CSWB model. Section 6.3 describes the dataset used in this study. Section 6.4 describes the calibration of the crop models for groundnut cultivation in the two Gambian divisions. Section 6.5 provides the results of preliminary crop model yield responses to simple rainfall distribution. Discussions of the performance in crop yield prediction using gauge rainfall and satellite-based rainfall driven crop models is given respectively in Section 6.6 and 6.7 and the conclusions are drawn in Section 6.8.

## 6.2 Crop model descriptions

### GLAM

The key parameters and equations for GLAM are listed in Box 5 and Box 6. GLAM is a one-dimensional process-based crop model that simulates at a daily timestep, the development of the crop and its interaction with its environment, from planting to maturity. The model primarily consists of a crop development sub-model and a soil hydrological submodel. An outline of the model is given below. Detail description of GLAM can be found in Challinor *et al.* (2004) abbreviated as CH04 hereafter.

*Soil hydrology:* The model soil is multi-level defined by its drained upper limit (soil water at field capacity), drained lower limit (soil water at wilting point) and saturation limit (the saturated water content) which is depth independent. Surface runoff is calculated using the Soil Conservation Method (Equation 6.1; USDA-SCS, 1964) and the change in soil moisture at each layer due to gravitational drainage is obtained using the method due to Suleiman (1999) (Equation 6.2).

## Box 5: Notation for key parameters in GLAM

$C_{YG}$	yield gap parameter	$T_T$	transpiration
$e_{sat}$	saturated vapor pressure	$T_{Tpot}$	potential transpiration
$E$	evaporation	$T_T^e$	‘energy-limited’ transpiration
$E_T$	normalised transpiration efficiency	$T_T^p$	‘physiologically-limited’ transpiration
$E_{TN,max}$	maximum transpiration efficiency	$v_{EF}$	extraction front velocity
$E^e$	‘Energy-limited’ evaporation	$V$	vapor pressure deficit
$E_{POT}^s$	‘soil-structure’ limited potential evaporation	$W$	above-ground biomass
$H_I$	harvest index	$Y$	yield
$l_v$	volumetric root length density	$z$	depth of soil from surface
$L$	leaf area index	$z_{max}$	maximum soil depth
$Q$	incoming water flux	$Z$	rainfall
$R$	runoff	$\alpha$	Priestly-Taylor coefficient
$R_N$	net all-wave radiation	$\gamma$	psychometric constant
$S$	soil stress factor	$\lambda$	latent heat of vaporisation of water
$t$	time	$\theta$	soil water content
$t_e$	time of arrival of extraction root front	$\theta_{dll}$	drained-lower limit
$t_R$	time since first rain	$\theta_{dul}$	drained-upper limit
$t_{TT}$	thermal time	$\theta_{sat}$	saturation limit
$T$	air temperature	$\theta_{pe}$	potentially extractable water
$\bar{T}$	average daily air temperature	$\left(\frac{\partial L}{\partial t}\right)_{max}$	maximum rate of LAI increase
$T_b$	base cardinal temperature		
$T_m$	maximum cardinal temperature		
$T_o$	optimum cardinal temperature		
$T_{eff}$	effective temperature		

**Box 6: Key equations for GLAM**

Meaning of the key symbols is given in Box 5. Undefined symbols are understood as prescribed constants in the model. Refer to Challinor *et al.* (2004) or Annex C for the definition of these parameters

**Soil hydrology**

$$\text{Surface runoff} \quad R = \frac{Z^2}{Z + k_{sat}}, \quad k_{sat} = K_{ks} \left( \frac{\theta_{sat} - \theta_{dul}}{\theta_{dul}} \right)^2 \quad 6.1$$

$$\text{Drainage} \quad \frac{\partial \theta}{\partial t} = -FD(\theta_s - \theta_{dul}) \quad 6.2$$

$$D = C_{d1}\theta_{dul}^2 + C_{d2}\theta_{dul} + C_{d3}, \quad F = 1 - \frac{\ln(Q_i + 1)}{\ln(k_{sat} + 1)}$$

**Crop development**

$$\text{Thermal time} \quad t_{TT} = \int_i^r (T_{eff} - T_b) dt \quad 6.3$$

$$T_{eff} = \begin{cases} \bar{T} & T_b \leq \bar{T} \leq T_0 \\ T_0 - (T_0 - T_b) \left( \frac{\bar{T} - T_0}{T_m - T_0} \right) & T_0 < \bar{T} < T_m \\ T_b & \bar{T} > T_m, \bar{T} < T_b \end{cases}$$

**Crop growth**

$$\text{Biomass and yield} \quad \frac{\partial W}{\partial t} = T_T \min \left( \frac{E_T}{V}, E_{TN, \max} \right) \quad 6.4$$

$$\frac{\partial H_I}{\partial t} = \text{prescribed constant}, \quad Y = H_I W$$

$$\text{LAI} \quad \frac{\partial L}{\partial t} = \begin{cases} \left( \frac{\partial L}{\partial t} \right)_{\max} C_{YG} \min \left( \frac{S}{S_{cr}}, 1 \right) & \text{crop stage} > 3 \\ 0 & \text{crop stage} = 3 \end{cases} \quad 6.5$$

$$S = \frac{T_T}{T_{Tpot}}$$

**Box 6: Key equations for GLAM**

$$\begin{aligned} \frac{\partial l_v(z=0)}{\partial L} &= \text{prescribed constant} \\ \text{Root } v_{EF} &= \text{prescribed constant} \\ l_v(z=z_{ef}) &= \text{prescribed constant} \end{aligned} \quad 6.6$$

**Potential Evaporation**

$$\begin{aligned} E^e &= (1 - C_G) E_{\max}^T e^{-kL} \\ \text{Energy-limited } E_{\max}^T &= \frac{\alpha \Delta R_N}{\lambda(\Delta + \gamma)}, \alpha = 1 + 0.26 \frac{V}{V_{ref}} \text{ and } \Delta = \frac{\partial e_{sat}}{\partial T} \end{aligned} \quad 6.7$$

$$\text{Soil structure-limited } E_{POT}^s = \frac{E^e}{t_R}$$

**Potential Transpiration**

$$\begin{aligned} \text{Energy-limited } T_T^e &= E_{\max}^T (1 - e^{-kL}) \\ \text{Physiologically limited } T_T^p &= \begin{cases} T_{T\max} \left( 1 - \frac{L_{cr} - L}{L_{cr}} \right) & L < L_{cr} \\ T_{T\max} & L \geq L_{cr} \end{cases} \\ \text{Potential transpiration } T_{Tpot} &= \min(T_T^p, T_T^e) \end{aligned} \quad 6.8$$

**Evaporation and transpiration**

$$\begin{cases} T_T = T_{Tpot} \text{ and } E = E_{POT}^s & \text{for } \theta_{pe} \geq E_{POT}^T \\ T_T = \theta_{pe} \frac{T_T^e}{E_{POT}^T} \text{ and } E = \theta_{pe} \frac{E^e}{E_{POT}^T} & \text{for } \theta_{pe} < E_{POT}^T \end{cases} \quad 6.9$$

$$E_{POT}^T = E^e + T_T^e$$

$$\begin{aligned} \text{Potential extractable moisture } \theta_{pe} &= \int_0^{\infty} \theta_{cr} (1 - e^{-k_{DIF} l_v(z) t_e(z)}) dz \\ \theta_{cr} &= \theta_{ll} + C_\theta (\theta_{dul} - \theta_{ll}) \end{aligned} \quad 6.10$$

*Crop development:* The crop development (groundnut in this case) is represented by four sequential stages: Planting to flowering (Stage 0), Flowering to Pod Initiation (Stage 1), Pod Filling to LAI maximum (Stage 2) and LAI maximum to Maturity (Stage 3). The duration of the crop at a given developmental stage depends on the thermal time elapsed ( $t_{TT}$ ; Equation 6.3) in that stage. In Equation 6.3, the effective temperature ( $T_{eff}$ ) is a function of the surface air temperature and the crop specific cardinal temperatures; air temperature away from the optimal temperature ( $T_0$ ) will result in slower crop development. Once this thermal time is greater than a prescribed value (which are stage dependent in general), the crop progresses to the subsequent developmental stage or is harvested (simulation terminated) if mature (Stage 3).

*Crop growth:* The crop is characterised by its biomass, Leaf Area Index (LAI), and root profile. Rate of biomass mass accumulation, LAI development and root growth are given in Equation 6.4 to 6.6. The rate of biomass accumulation is obtained from transpiration via the transpiration efficiency, which below a maximum value of crop-dependent maximum transpiration efficiency ( $E_{TN,max}$ ) is determined by the crop-dependent normalised transpiration efficiency ( $E_T$ ) and the vapor pressure deficit,  $V$ . The useful crop yield is determined as a fraction of the biomass via the harvest index  $H_I$ . The maximum rate leaf growth is given by the crop-dependent maximum rate of LAI increase  $\left(\frac{\partial L}{\partial t}\right)_{max}$ , before Stage 3. LAI growth rate is attenuated by a yield gap parameter  $C_{YG}$  ( $\leq 1$ ), which represent crop development due to non-optimal crop management practices. Water deficits will also begin to affect the leaf development when the water stress  $S$  is below a crop-dependent threshold  $S_{cr}$ . The volumetric root length density at surface ( $z = 0$ ) and at the root extraction front ( $z = z_{ef}$ ) defines the root systems in the model; the vertical profile for the volumetric root length density is given by a linear interpolation of  $l_v$  between the surface and the extraction front. In the model, the rate of root density increase depends on the rate of leaf growth; non-optimal environmental condition thus affects the root development indirectly through environmental stresses on leaf development.

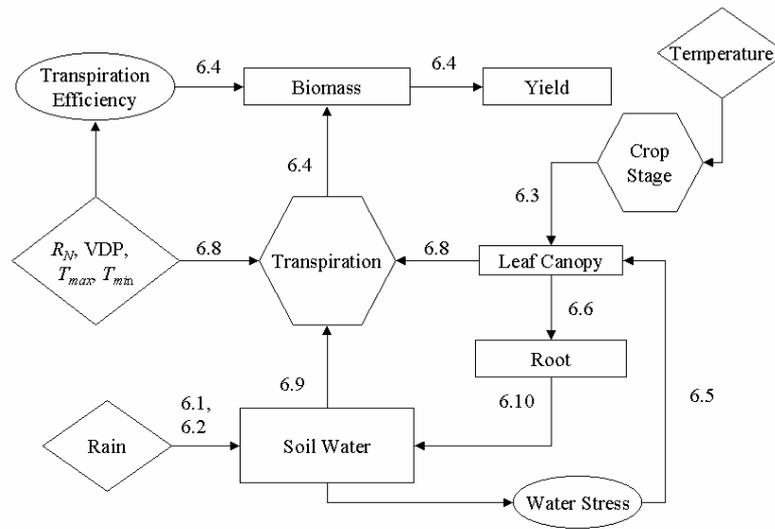


Figure 6.1 Relationships between environmental and crop components in GLAM. The numbers are the reference for the equations. The arrow represents the dependences between the components.

*Evaporation and Transpiration:* Evaporation ( $E$ ) and transpiration ( $T_T$ ) are determined separately in the model. Potential evaporation ( $E_{POT}^s$ ) is constrained by both energy availability and soil structure (Equation 6.7) while the potential transpiration ( $T_{Tpot}$ ) is constrained by energy availability and physiology, whichever is lower (Equation 6.8). In Equation 6.7, the reciprocal dependence of potential evaporation on number of days since a wetting event ( $t_R$ ) reflects the decrease in potential evaporation as when soil moisture decrease. Transpiration is limited physiologically by the size of leaf canopy; beyond a critical size ( $L_{cr}$ ) transpiration does not increase. The ‘energy-limited’ evaporation and evaporation is defined such that their sum is consistent with evapotranspiration obtained using the Priestly-Taylor equation (Priestly-Taylor, 1972). The effect of wind speed on evapotranspiration is taken into account in the Priestly-Taylor coefficient ( $\alpha$ ) through the vapor pressure deficit. Evaporation and transpiration are determined from their respective potential values (Equation 6.9). For non-water limiting situation (extractable soil water  $\geq$  potential evapotranspiration), transpiration is at potential, limited only by crop physiology and energy availability. When the extractable soil water is limited (extractable

soil water < potential evapotranspiration), partitioning of the limited extractable soil water to transpiration depends on relative competition with evaporative demands.

One feature of GLAM to note is the soil-crop feedback within the model. This is best illustrated with the schematic of the relations of the major model components of GLAM shown in Figure 6.1. In the model, soil water availability affects crop development through its ability to satisfy the transpiration demands as well as its effect on leaf canopy and root development. The soil-leaf-root feedback loop is important in determining the impact of rainfall variability in the final crop yield at harvest. Soil water stress at the early stages of crop development (before LAI attained its maximum value) results in lower LAI and volumetric root density as the crop matures. This limits transpiration due to lower soil water extraction by the roots (Equation 6.10) and smaller leaf canopy. Even when soil moisture is sufficient later in the crop season, the crop is unable to utilise the available soil moisture. On the other hand, when soil moisture is sufficient in the earlier crop stages, near optimum leaf canopy and root growth can be attained and consequently, is able to better utilise soil moisture in the later crop stages compared to the preceding case. Hence, for GLAM, the impact of soil moisture deficits on crop development can be unequal depending on their occurrence with respect to the crop stages. Although other input weather parameters affect the biomass accumulation, their effects on crop development in general is comparatively lower than that of rainfall in practice. This is primarily due to the absence of coupling between these inputs, except temperature, and the leaf and root development, as well as the lower variability of these parameters compared to that of rainfall (see for example Chapter 2, Section 2.2). Although the duration of each crop stage depends on temperature, the relatively low variability of daily average temperature usually has a lower impact than rainfall for GLAM in practice. The inclusion of effect of crop development has been recently addressed in GLAM (Challinor *et al.*, 2005c) in response to the importance of temperature stress on various crops (Ferris *et al.*, 1998; Ismail and Hall, 1999; Prasad *et al.*, 2000; Matsui *et al.*, 2001) especially during flowering (Wheeler *et al.*, 2000). However, this is not included in the current discussion.

**Box 7: Notation for key parameters in CSWB model**

$D$	soil water surplus
$E_0$	reference evapotranspiration from Penman-Montieth method
$E_T$	evapotranspiration
$E_{POT}$	potential evapotranspiration
$k_y$	yield respond coefficient
$K_c$	crop coefficient
$S$	available soil water
$W_{HC}$	soil water holding capacity
$Y$	crop yield
$Y_m$	maximum non-water limiting yield under optimal conditions
$\beta$	yield reduction factor
$\varepsilon$	rainfall attenuation factor

**Box 8: Key Equations of CSWB model****Soil Water balance**

Soil water surplus  $D_i = \varepsilon Z_i + S_i - E_{POTi}$  6.11

Updating soil moisture  $S_{i+1} = \begin{cases} \min(D_i, W_{HC}) & D_i \geq 0 \\ 0 & D_i < 0 \end{cases}$  6.12

**Evapotranspiration**

Potential evapotranspiration  $E_{POT} = K_c E_0$  6.13

Crop coefficient  $K_c(i) = K_{stage-1} + \left[ \frac{i - \sum L_{prev}}{L_{stage}} \right] \Delta K$  6.14

$$\Delta K = (K_{stage} - K_{stage-1}),$$

$\sum L_{prev}$ : total duration of all previous stages

Evapotranspiration  $E_{Ti} = \begin{cases} E_{POTi} & D_i \geq 0 \\ \varepsilon Z_i + S_i & D_i < 0 \end{cases}$  6.15

**Box 8: Key Equations of CSWB model****Yield**

$$Y = \beta Y_m \left( 1 - k_y \left( 1 - \frac{\sum_i E_{Ti}}{\sum_i E_{POTi}} \right) \right) \quad 6.16$$

CSWB model

The key parameters and equations within the CSWB model are listed in Box 7 and Box 8 respectively. The relatively simple CSWB model operates at dekadal (10 day) timestep. It primarily book-keeps the amount of soil moisture recharging by precipitation,  $Z$ , after soil moisture lost through crop evapotranspiration, surface runoff and gravitational drainage to determine crop water satisfaction after each timestep. Crop yield is determined from the overall crop water satisfaction.

*Soil water balance:* Soil water is computed using a ‘bucket model’ where the soil hydrology is parameterised only by its water holding capacity,  $W_{HC}$ . At each timestep  $i$ , the soil water surplus ( $D_i$ ) is first determined after accounting for recharging by rain ( $Z_i$ ), the potential loss through evapotranspiration demands ( $E_{POTi}$ ), as well as antecedent soil moisture ( $S_i$ ) (Equation 6.11). A schematic of the water budget for the bucket model used in the CSWB is shown in Figure 6.2. Surface runoff and the any other non-optimal extraction of soil water by the roots not explicitly captured by the bucket model are reflected by attenuating the rainfall before infiltration through an attenuation factor ( $\varepsilon$ ). Soil moisture at the end of the water balance cycle is updated according to Equation 6.12: In the case when the evapotranspiration demands is higher than the total combined rainfall and antecedent soil moisture, all water available (rainfall and antecedent soil moisture) are assumed to be used for evapotranspiration. If available water is in excess, any surplus in excess of the soil water holding capacity is assumed to be loss through gravitational drainage away from the root zone.

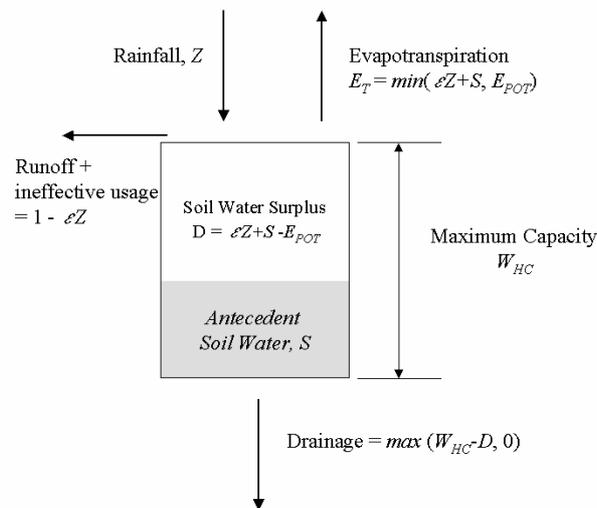


Figure 6.2 Schematic of the water budget in the bucket soil model in CSWB.

*Evapotranspiration:* No crop processes are represented in the CSWB model; Potential evapotranspiration is determined from the reference potential evapotranspiration ( $E_0$ ) through a crop coefficient,  $K_c$  (Dorenbos and Pruitt, 1977; Equation 6.13).  $E_0$  is computed using the FAO Penman-Monteith equation and determination of  $K_c$  follows from the FAO recommendation (Allen *et al.*, 1998). For a given crop, crop development is divided into 4 stages namely, Initial (planting to 10% of groundcover), Vegetative Development (10% to maximum ground cover), Mid Season (maximum ground cover to start of maturity) and Late Season (maturity to harvest).  $K_c$  is crop and time dependent. In general,  $K_c$  is dependent on crop phenology reflecting the different crop water demand for different crop stages as well as the environmental condition as shown in Figure 6.3. In this study,  $K_c$  for a time step is numerically determined by the following (Allen *et al.*, 1998, Equation 6.13). Adjustments of  $K_c$  due to environmental conditions (e.g. water stress, increases of evaporative demands after rainfall during the initial stages) are not taken into account in the current model. Actual water-limiting evapotranspiration is determined through Equation 6.15.

*Crop yield:* Several methods of linking the CSWB model outputs to crop yield exist. In the current work, the actual water-limited crop yield from the CSWB model was computed using a yield response function of Doorenbos and Kassam (1979) (Equation 6.16). Reduction from the maximum yield ( $Y_m$ ) is due to non-optimal management practices is through the yield reduction factor ( $\beta$ ) analogous to the yield gap parameter in GLAM. This way of determining of yield from the CSWB model is similar in form to that used by Reynolds *et al.* (2000). In their work,  $\beta = 1$  and  $Y_m$  is assumed to take into account of non-optimal management practices.

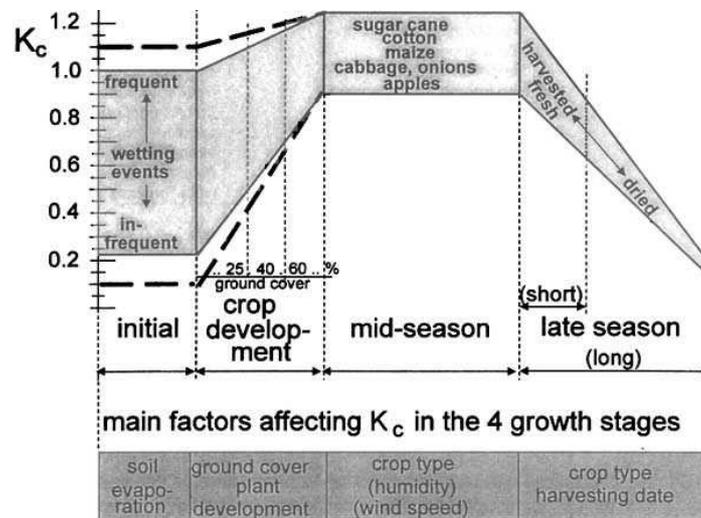


Figure 6.3 Dependence of  $K_c$  with crop developmental stages for a typical crop. (From Allen *et al.*, 1998)

To aid discussion in the following sections, several yield-water relations for the CSWB are further noted. It can be shown from Equation 6.16 that the difference in predicted yield between two CSWB simulations using two rain inputs (call them  $a$  and  $b$ ), with other inputs remaining the same, is directly proportional to the difference in the total crop evapotranspiration for the growing season:

$$Y_a - Y_b = \frac{\beta Y_m}{E_m} k_y \left[ \left( \sum E_T \right)_a - \left( \sum E_T \right)_b \right] \quad 6.17$$

where  $k_y$  is the constant crop yield coefficient and  $E_m$  is the total potential evapotranspiration. Equation 6.6 can be further expressed in the terms of the crop-environment water budget from the bucket model of CSWB:

$$Y_a - Y_b = \alpha [\varepsilon(Z_a - Z_b) - (D_{Ra} - D_{Rb}) - (\Delta S_a - \Delta S_b)] \quad 6.18$$

where  $\varepsilon$  is the rainfall attenuation coefficient,  $\alpha = \frac{\beta Y_m}{E_m} k_y$  is the water use efficiency,  $Z$  is the total rainfall amount received,  $D_R$  is the total drainage of excess water through deep percolation,  $\Delta S$  is the change in soil moisture at harvest from planting. If it is further assumed that the differences in soil moisture between  $a$  and  $b$  at planting can be ignored, the above relationship becomes

$$Y_a - Y_b = \alpha [\varepsilon(Z_a - Z_b) - (D_{Ra} - D_{Rb}) - (S_{Ha} - S_{Hb})] \quad 6.19$$

where  $S_H$  is the remaining soil water at harvest.

Equations 6.18 and 6.19 are useful when comparing crop yields prediction when the CSWB model is driven by gauge-based rainfall and satellite rainfall estimates, or when investigating the effects of errors in satellite rainfall estimates on CSWB yield predictions (Section 6.7.2).

### 6.3 Datasets

#### Yield and weather data

Divisional groundnut yield data from 1974 to 2002 for NBD and CRD were available. Although other cultivars have been introduced in recent years, the major cultivar of groundnut grown in the Gambia is 28/206, which has a crop-duration of approximately 120 days. As the yield records available did not differentiate the contribution for the individual cultivars, it was assumed that the groundnut yield data are predominately those of the 28/206 cultivar. The divisional yield time series were detrended using a linear trend model and a sigmoid model for CRD and NBD respectively. For NBD, due to the uncertainty of trend before 1982, the current crop modelling focuses only on the groundnut yields from 1982. Further details of Gambian groundnut production data and discussion of the detrending of the divisional groundnut yield time series is given in Chapter 2.

Weather data for the same period as the yield records were available as inputs for crop modelling. Gridded ( $0.05^0 \times 0.05^0$ ) daily maximum and minimum temperatures, sunshine hours, relative humidity and surface winds (Chapter 2) are spatially aggregated for the two divisions. Likewise, daily divisional gauge-based and satellite-based rainfall timeseries were obtained by spatially aggregating the gauge-pixel rainfall estimates (Chapter 3), TAMSAT1 rainfall estimates (Chapter 4), as well as stochastic daily rainfall realisations from the SIMU algorithm (Chapter 5). For ease of discussion, the daily divisional rainfall data obtained from TAMSAT1 algorithm and raingauge observations is abbreviated respectively as TAMSAT1 and GAUGE whenever permissible.

### Soil Data

Soil in the Gambia crop cultivation region is generally loamy sand and sandy loam for the top 50 cm of the soil with an increasing amount of clay with depth (Williams, 1975). Drained lower limits and drained upper limits were adopted as 0.04 – 0.14 and 0.14 – 0.24 from the range of values for soil water content at wilting point and field capacity recommended by Allen *et al.* (1998) for the two soil series. These are reasonable ranges considering the total available water in the soil is 150-200 mm for NBD and Southern CRD, while Northern CRD is about 100-150 mm (FAO, 1995). Several field experiments conducted in the Gambia (Williams, 1975) reported saturation water content for agricultural soils at about 0.3 – 0.35 for deeper soils; as bulk density is generally lower at the top soil of interest (~1m), a higher saturation water content was used. Hence, a range of saturation limit of 0.3 – 0.5 was adopted.

### Planting dates

The annual sowing dates that best represent the aggregated sowing activity on a divisional scale were estimated solely from the weather data. Following FAO (1978), the criterion used to determine the day favourable for planting was when the daily rainfall was greater than half of the potential evapotranspiration. For each division, three such possible sowing dates satisfying the above criteria were determined each year, within a broad window from mid-June to end-July. This was done to account for the variability in planting dates over the entire division. The average of these dates was then taken to be

the sowing date for the division used in the crop models. Such method of determining planting dates may be too simplistic. However, without other information, it is prudent to keep the algorithm as simple as possible.

For the purpose of investigating crop model sensitivities to sowing date uncertainties, a 20-days window ( $\pm 10$  days around the estimated sowing date) was adopted as the uncertainty of the computed sowing date. This window was chosen by analysing the limited sowing date data available for Gambian groundnut: For 1996 and 1997 where the sowing data on a national level was available, both CRD and NBD sowing dates estimated from the algorithm coincided with the week when the countrywide cumulative sowing percentages from data were reported to be just above 60 percent. A window of 10 days of either side of the estimated sowing dates include the 50<sup>th</sup> and 70<sup>th</sup> percentiles of the countrywide sowing percentages for both years. Unfortunately no other information was available to further validate such estimation. More sophisticated methods of determining planting dates using NDVI (Reynolds et. al., 2000) can probably be devised to give a more complete validation of the present method.

#### 6.4 Model parameters selection and calibration

The general philosophy in calibrating the model parameters for both GLAM and CSWB model is as follows: all crop growth related parameters were assumed to be space-time invariant (global parameters). Spatio-temporal variation of crop yield after removing any non-weather related trends should be determined by the spatio-temporal variability of weather, and the spatial variation of soil hydrology and management practices. The set of crop-related parameters was obtained as far as possible from available literature. The yield reduction factor and soil parameters for the two divisions were allowed to vary over a valid range of values to select the set of optimised value that gave the minimum root mean square error (RMSE) in the predicted yield. The yield reduction factor representing management practices ( $C_{YG}$  for GLAM and  $\beta$  for CSWB) was first calibrated for each division before optimising the soil parameters.

For each division, separate calibrations were carried out with two non-overlapping subsets of the observed groundnut yield time series to assess the time-stability of the model parameters adopted. Invariance with location of the global parameters was confirmed by ensuring that the individual (global) parameter prescribed produced a minimum RMSE in the simulated yield for the two divisions. The details of the calibrations for the two models are given below.

### GLAM

All global parameters, namely those relating to crop development and water usage for groundnut were adopted from CH04, except the set of parameters that were related to the thermal times of the groundnut which determined the crop duration. The reason being the groundnut being modelled in CH04 had a shorter crop-duration. The thermal time for the individual phenological stages required was estimated by proportionally increasing all thermal times from the mean value of the range given in CH04 by a common factor, such that the period from planting to maturity was about 120 days. The yield gap parameter ( $C_{YG}$ ) was determined from a wide range of values from 0.1 to 1.0 at an interval of 0.05. Soil parameters were determined by only calibrating the drained upper ( $\theta_{dul}$ ) and saturation limit ( $\theta_{sat}$ ) of the soil. The drained lower limit of soil ( $\theta_{dll}$ ) was fixed at the average of the range of values (0.09) suitable for the Gambia (see Section 6.3). For a more complete model evaluation, various model calibrations for different prescribed values of the drained lower limit to reflect the uncertainties of the soil parameters should be performed. However, these were not carried out in the present study for the sake of simplicity. Values of the parameters used for GLAM are given in Annex C.

Calibrations were carried out for two equal intervals of the time series for both divisions: 1974-1988 and 1989-2002 for CRD, and 1982-1992 and 1993-2002 for NBD. The divisional calibration curves for the each calibration time interval are shown in Figure 6.4. For clarity, the earlier and later calibrating time interval for both divisions is denoted as T1 and T2. For example, T1 for CRD is the 1974-1988 calibrating time interval, while T1 for NBD is the 1982-1992 calibrating time interval.

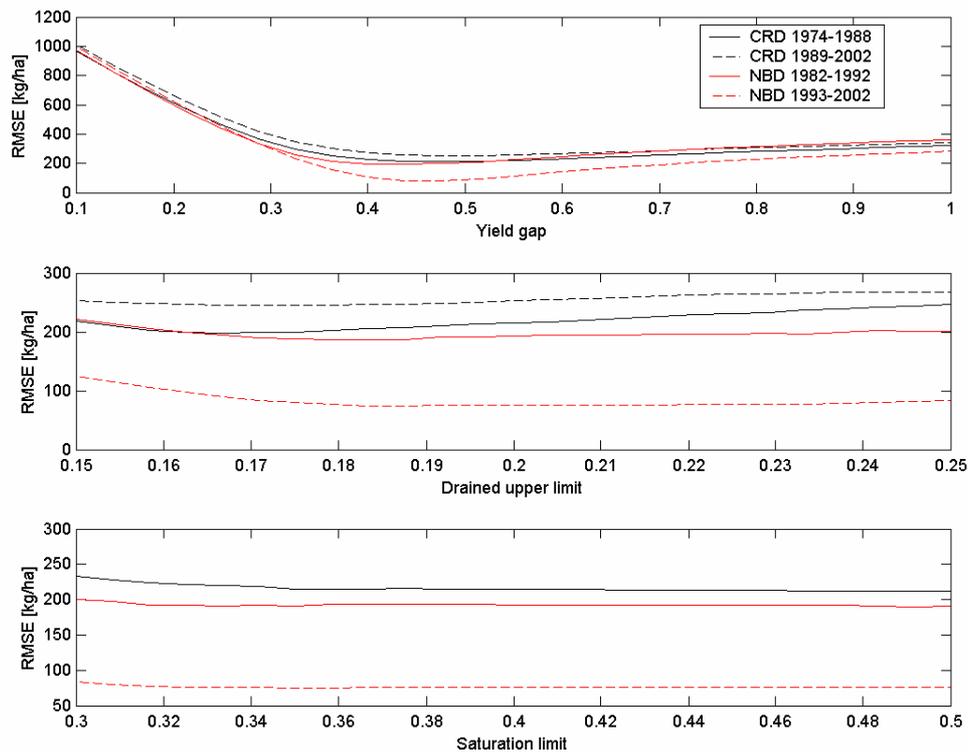


Figure 6.4 RMSE of crop yield functions of the different parameter values for CRD (black) and NBD (red) for different calibrating time periods. The yield RMSE functions with yield gap parameter were obtained when the average value of range of possible values was assigned to the drained upper limit and saturation of the soils (0.2 and 0.4 respectively). The yield RMSE functions with drained upper limit and saturation of the soil were obtained under the optimal yield gap parameters.

*Yield gap parameter:* The optimal yield gap parameter for CRD was found to be 0.5 for the two calibrating time periods. For NBD, there was a slight difference in the optimal yield gap parameter obtained — 0.45 and 0.4 for T1 and T2 respectively. This difference did not impact the optimality of the model calibration significantly; for T1, there was only an increase of 3% increase in the RMSE of the predicted yield for the range of yield gap parameters between 0.4-0.5.

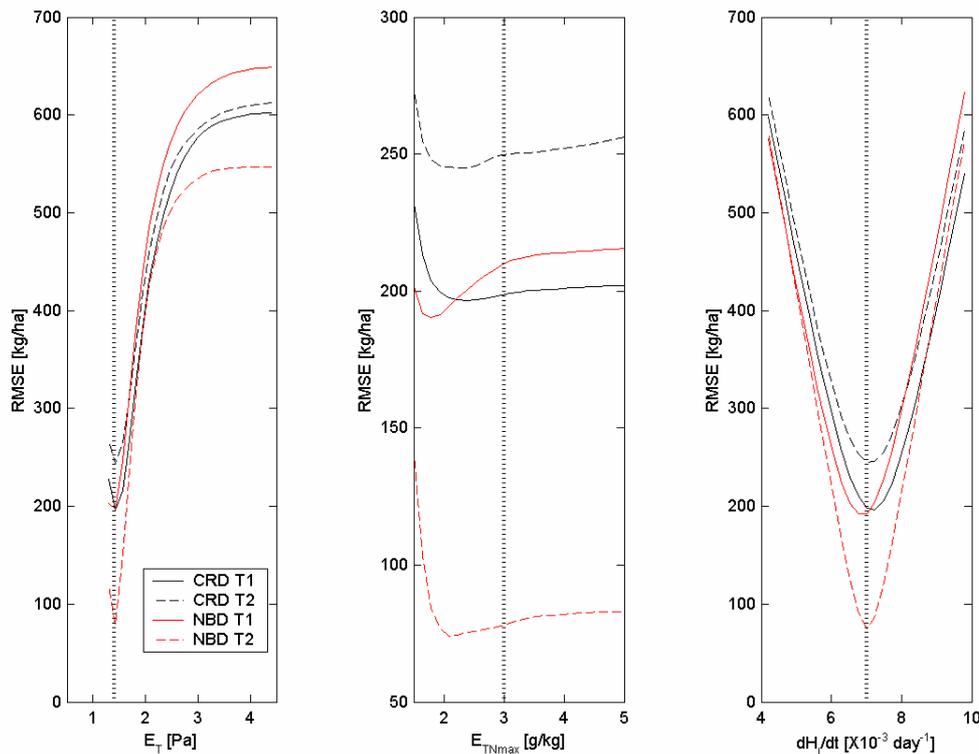


Figure 6.5 RMSE of crop yield as a function of normalised transpiration efficiency ( $E_T$ ), maximum transpiration efficiency ( $E_{TN,max}$ ) and rate of change of harvest index ( $dH_i/dt$ ) for CRD and NBD over a range of parameter values. Dotted line is the adopted value used in the model. In the legend, T1 in the legend is 1974-1988 and 1982-2002 for CRD and NBD respectively and T2

*Soil parameters:* Simulated crop yields for both divisions showed little sensitivity to the range of the drained upper limit and the saturation limit assigned. For CRD, the optimum drained upper limit was about 0.17 and 0.16-0.19 for T1 and T2 respectively. For NBD, the optimal drained upper limit obtained was between 0.17-0.19 for T1. For T2, there was little variation (<2%) in RMSE in the simulated yield for values between 0.18-0.22. Optimal saturation limits for CRD were identical for both calibration time periods, while for NBD no significant difference in optimality was achieved beyond a saturation limit value of 0.35 (<2% change in RMSE of predicted yield).

*Stability of parameters in time and space:* Once the model was calibrated for the two divisions, stability in space and time of the global parameters adopted in the model were confirmed by verifying that the RMSE of the predicted yield attained a minimum about

the adopted values for individual parameters. This was done by computing the RMSE of predicted yield for both divisions over a permitted range of values adopted from CH04 for each of the global parameters. The optimality and stability across divisions of three such parameters namely, the normalised transpiration efficiency, maximum transpiration efficiency and rate of change of harvest index are shown Figure 6.5. All three optimal parameters for both divisions were found to be close to or coincidental to the values adopted except for the maximum transpiration efficiency  $E_{TN,MAX}$ , whose optimal values appeared to be about 2 g/kg. However, the changes in RMSE in the observed yields were not large when  $E_{TN,MAX}$  varies from 2 to 3 g/kg; the largest fractional change of RMSE of yield was less than 15% (T2 for NBD). Hence, no further optimisation was conducted in order to avoid over calibration.

#### CWSB model

For the CSWB model, the parameters related to the crop coefficients ( $K_c$ ) and yield response coefficient ( $k_y$ ) for groundnut were adopted from Allen *et al.* (1998) and Doorenbos and Kassam (1979) respectively. The depth of the bucket model of the soil was set to 1 meter as the fully grown crop is able to extract most of the soil water at a depth of about 0.5 to 1 meter (Doorenbos and Kassam, 1979). The model parameter values and the calibration ranges adopted are shown in Table 6.1. As the recommended values for the stage durations for groundnut in Allen *et al.* (1998) are for groundnuts of 130-day duration, the duration of each crop stage ( $L_{stage}$ ) were reduced proportionally by a common factor from the recommended values to attain a crop duration of 120 days. Maximum attainable yield ( $Y_m$ ) was set at 2500 kg/ha, taking the mean of the range of values given in Doorenbos and Kassam (1979). The precise value of  $Y_m$  is not too important for yield forecasting since the magnitude of the reduction factor ( $\beta$ ) is relative to  $Y_m$ .

For each division, calibration was performed on the yield reduction factor, the water holding capacity and rainfall attenuation factor. The yield reduction factor and the rainfall attenuation factor were given a broad range (0.1-1 and 50%-100% respectively). The

optimisation curves for the yield reduction factor, the water holding capacity and the rain attenuation factor are shown in Figure 6.6.

*Yield reduction factor:* For both divisions, the optimal value for  $\beta$  was found to be about 0.6 for all calibration time periods.

*Water holding capacity:* For NBD, the calibration curves of the water holding capacity ( $W_{HC}$ ) for the two different time periods attained a broad minimum beyond 100 mm. For CRD, simulated crop yields were relatively insensitive to the ranges of  $W_{HC}$  for T2, while  $W_{HC}$  was optimal for a broad range of values around 80mm for T1.

*Rainfall attenuation factor:* For CRD, the optimal rainfall attenuation factor  $\varepsilon$ , was about 65% - 70% and about 75% for T1 and T2 respectively. For NBD, optimal  $\varepsilon$  was around 75% and 65% for T1 and T2. For both divisions, the effect on RMSE of predicted yields, due to the different optimal values for the two calibrations of  $\varepsilon$  was found to be small: Assigning  $\varepsilon = 70\%$  for all calibrations resulted in less than a 2% increase in predicted yield RMSE from their respective minimum.

The calibrated attenuation factor for rain for both divisions represents a water loss of 30% through runoff, and other ineffective water usage by the crop. Non-optimal usage of the soil moisture in this case is partly model dependent: Since the amount of extractable water depends on the depth of root penetration (which increases as the crop develops), not all the water within the modelled bucket depth at the early stages of crop growth are utilised. However, relative contributions of runoff loss and the model dependent component to the attenuation factor obtained were difficult to ascertain. From field measurements conducted in the Gambia, Williams (1975) estimated that 10% - 50% of the rainfall could be lost by surface runoff, but it is uncertain how such magnitude of rainfall runoff from the field surveys should be translated to the spatial scale that the crop model currently operates.

Parameters	Values	Reference
$\beta$	0.1 – 1	
$K_{stage}$	[0.4 1.15 1.15 0.6]	Allen <i>et al.</i> (1998)
$L_{stage}$	[24 32 40 24] days	Allen <i>et al.</i> (1998)
$k_y$	0.7	Doorenbos and Kassam (1979)
$Y_m$	2000-3000 [2500] kg/ha	Doorenbos and Kassam (1979)
$W_{HC}$	50 – 200 mm	FAO/UNESCO (1974)
$\varepsilon$	50 – 100 %	

Table 6.1 Parameters for the CSWB model and their values used.  $K_{stage}$  and  $L_{stage}$  are expressed as  $1 \times 4$  column vectors with each column signifying a crop stages (stage = 1 to 4) of Equation 6.15.

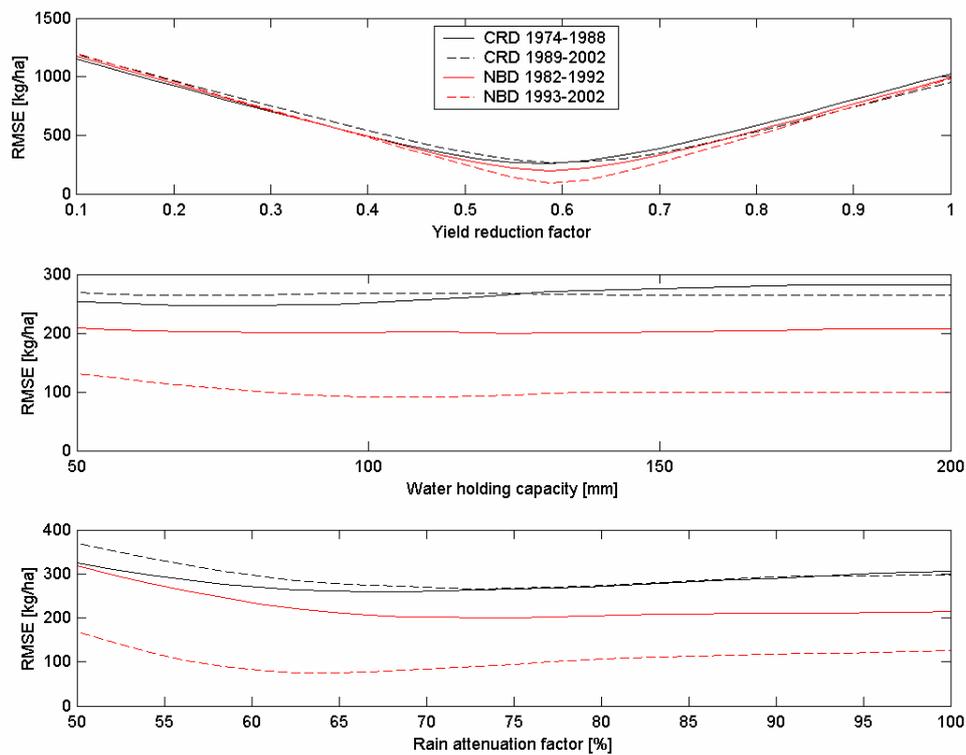


Figure 6.6 RMSE of simulated yield as function of various parameter values for CRD and NBD for different calibrating time periods.

## 6.5 Sensitivity experiments on crop model

Before proceeding to investigate the crop model responses to actual rainfall inputs, idealised experiments were conducted in which both GLAM and CSWB were driven by simple rainfall profiles. The purpose was to gain some basic appreciation of the simulated crop yield responses to simpler forms of variability in rainfall. Observations from these experiments served as references for later discussion when actual rainfall inputs were used to drive the crop models.

### Idealised experiment for GLAM

Four synthetic daily rainfall time series were used as input to GLAM (Figure 6.7). These rainfall time series corresponded to a “rainy season” of 120 days in line with the crop duration of groundnut in Gambia. The control rainfall profile (CTRL), RFD1 and RFD2 has total seasonal rainfall amount about 580 mm, but with maximum rainfall at different points in time. CTRL has its maximum in August, whilst RFD1 and RFD2 attained their maximum at July and September respectively. RFD1 and RFD2 represent water deficiencies at the later and earlier crop developmental stages respectively. RFD3 is used to represent the reduction of 20% of rainfall for CTRL, which is approximately the 20-year standard deviation of seasonal rainfall total for the Gambia.

Planting dates for all simulations were set to 1<sup>st</sup> July at the start of the ‘rainy season’. Since daily water availability for crop depends on the moisture retention characteristics of the soil as well, the sensitivity test is repeated for three different soil hydrologies, Soil Classes 1 to 3, with different drain upper limits ( $\theta_{dul}$ ) of 0.14, 0.20 and 0.24 respectively. This represents the range of possible values for Gambian soil as discussed in Section 6.3. All other agrometeorological inputs are similar for all model runs. The yield gap parameter was set as 0.5, a representative value for both NBD and CRD.

The model yields under RFD1, RFD2 and RFD3 compared with the control for the three different soil classifications are given in Figure 6.8. The following were observed:

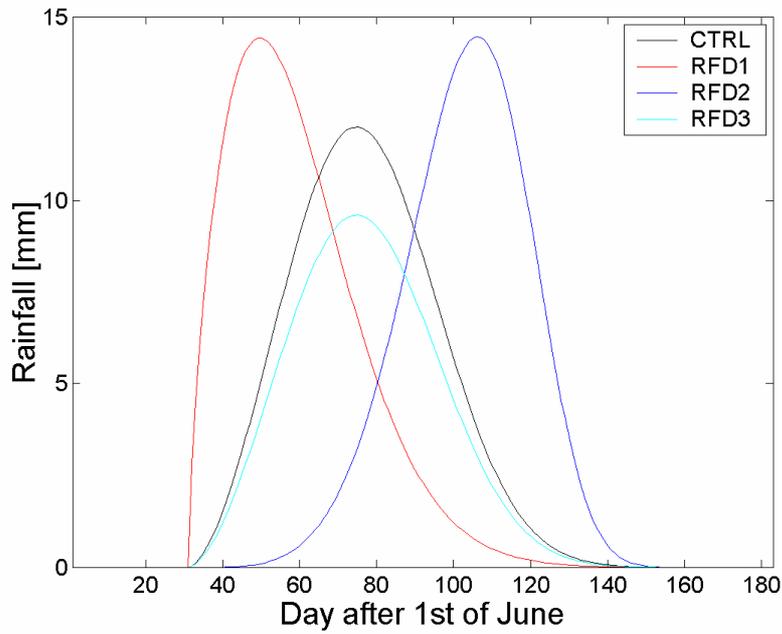


Figure 6.7 Rainfall distribution for 120 days used in the idealised experiment. RFD1 and RFD2 has same seasonal total amount as CTRL. RFD3 has the rainfall maxima as CTRL but with 20% less in total rainfall amount.

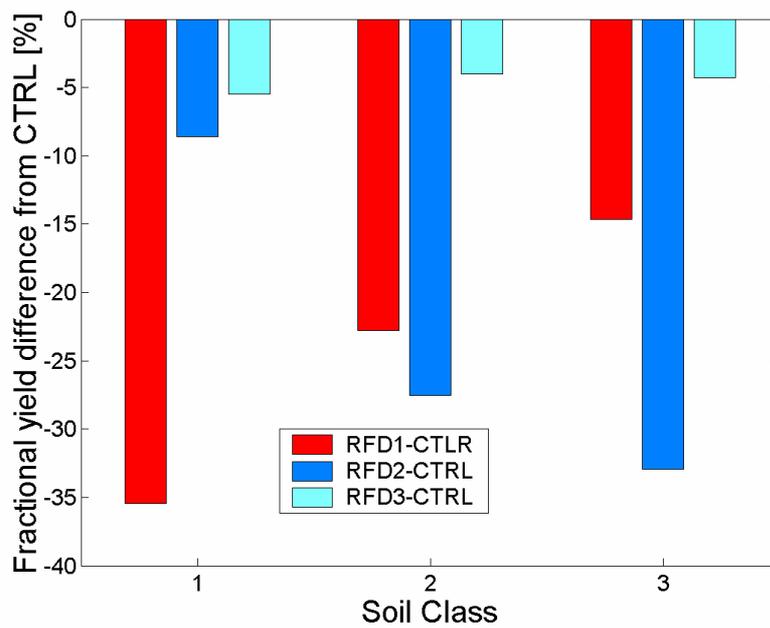


Figure 6.8 Fractional model yield difference from control run with various rainfall input. Soil Class represent soil types with  $\theta_{dul} = 0.14$  (1),  $\theta_{dul} = 0.20$  (2) and  $\theta_{dul} = 0.24$  (3)

1) For all soil classes, the timing of peak rainfall (RFD1 and RFD2) within the crop growing season has larger impact on simulated yields than the reduction of rainfall amounts (RFD3).

2) Rainfall peaking at different time had an unequal impact on the simulated yield, depending on the soil water holding capacity:

a) When rainfall peaked early in the growing season (RFD1), the reduction in the simulated yield was highest when compared to Control (CTRL) for soil with a low water holding capacity (Soil Class 1). The magnitude of the simulated yield reduction decreased with increasing soil water holding capacity.

b) When rainfall peaked late in the growing season (RFD2), the reduction in the simulated yield was the largest when compared to CTRL for soil with a high water holding capacity (Soil Class 3). The magnitude of the yield reduction decreased with decreasing soil water holding capacity.

The reasons for the above observations are best seen from the timeseries of various crop parameters for the experiment shown in Figure 6.9. For soil with the least water holding capacity (Soil Class 1), water deficits during later growing seasons (RFD1) showed the largest reduction in model yield, due to water lost through run-off and deep percolation of the heavy rainfall in the early stages of crop development. Combined with the low rainfall during later crop stages when transpiration demands increased, crop biomass accumulation became severely water limited. As the soil water holding capacity increased, the duration of moisture deficits at the later crop stages decrease. Consequently, the yield difference of RFD1 from CTRL decreased with increasing soil water holding capacity (Figure 6.8). The leaf areas (LAIs) attained were similar to CTRL and the final attainable yield depended on the ability of the soil to retain sufficient moisture for transpiration. Owing to the high transpiration demand with a large LAI, crop rapidly became water-stressed when there was prolonged moisture deficit; hence the yield reduction from CTRL for Soil Class 3 was still substantial.

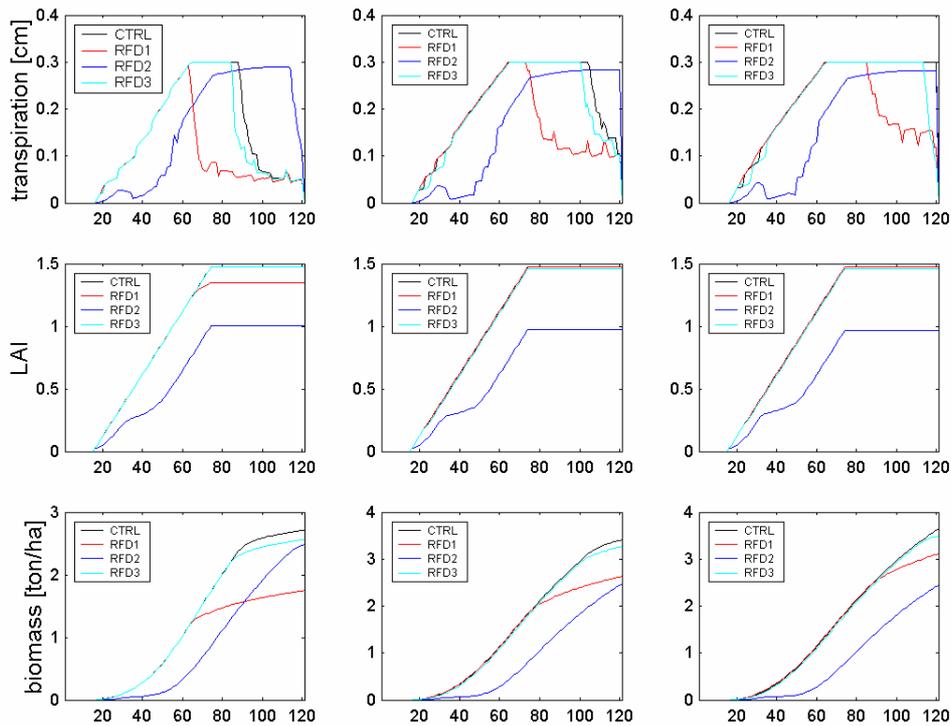


Figure 6.9 Time series of various model crop parameters for the idealised experiments for the three different soil classes (1)  $\theta_{dul} = 0.14$  (1); (2)  $\theta_{dul} = 0.20$ ; (3)  $\theta_{dul} = 0.24$ .

LAI development under RFD2 suffered from water deficits at the earlier crop development stages, resulting in less leaf and root growth when the crop matured. However, rainfall at the later period of crop development enables biomass accumulation to be maintained at its full potential. For soil with a progressively larger water holding capacity (Soil Class 2 and 3), crop development under CTRL became increasingly robust because of sufficient soil moisture. On the other hand, crop development under RFD2 for all soil capacities tested here were similar. The increasing difference in the yield from CTRL as soil capacity increases was a consequence of the increasing yield attained by CTRL as the soil water holding capacity increased (see Figure 6.9).

Two features of the rainfall-yield relationship of GLAM are highlighted from these experiments:

1) When crop water is sufficient in the earlier part of the growing season, crop development is optimal. If rainfall diminishes before the crop matures, the attainable yield depends on the remaining soil moisture to sustain transpiration at the late crop stages. For a given soil hydrology, this implies a certain dependence of yield and total seasonal rainfall.

2) Early water deficits can result in sub-optimal crop development. Therefore, transpiration and hence biomass accumulation, can become physiologically-limited; even when rainfall is abundant at the late growing season, crop yield reduction from its potential non-water limited yield can be substantial.

#### Idealised experiment with CSWB

The idealised experiment was repeated for the calibrated CSWB model, with the rainfall distribution (Figure 6.7) aggregated to their corresponding dekadal (10-day) counterparts. The experiment was conducted with three different soil classes defined by their water holding capacity ( $W_{HC}$ ); Soil Class 1 ( $W_{HC} = 50$  mm), Soil Class 2 ( $W_{HC} = 110$  mm) and Soil Class 3 ( $W_{HC} = 150$  mm), representing uncertainty in the soil water holding capacity in the CSWB model. This is the equivalent of the soil classes used in the experiment for GLAM ( $W_{HC} = \theta_{du} - \theta_{dl}$ ). The impact of the rainfall distribution for the predicted yield is shown in Figure 6.10. Figure 6.11 shows the crop evapotranspiration and the soil moisture when CSWB were driven with the various rainfall profiles.

The observed responses of simulated yields to rainfall distribution for CSWB is summarised as follows:

1) Characteristics of the response to rainfall distribution with various soil classes were different from GLAM:

a) Simulated yield reduction due to rainfall peaking in the early season (RFD1) remained fairly constant for all soil classes.

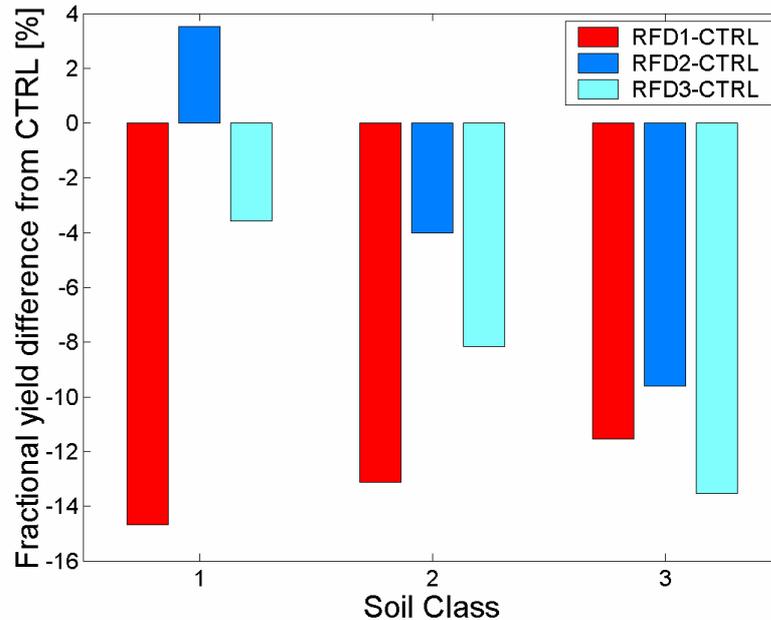


Figure 6.10 Fractional model yield difference from control run with various rainfall input with CSWB model. Soil Class represent soil types with  $W_{HC} = 50$  mm (1),  $W_{HC} = 110$  mm (2) and  $W_{HC} = 150$  mm (3)

b) Simulated yield increased for early water stress conditions (RFD2) for well-drained soil (Soil class 1) when compared to CTRL.

c) Simulated yield reduction due to reduced rainfall amount (RF3) increased with increasing  $W_{HC}$ .

2) The overall effect of rainfall distribution on yields was less than GLAM; the largest yield reduction was about 14%, less than half of the largest yield reduction from the GLAM experiments (Figure 6.8).

The yield-water relationship (Equations 6.17 and 6.19) can be invoked to explain the observed CSWB yield difference to the prescribed rainfall perturbations. Since the total seasonal rainfall amount for RFD1 were the same as CTRL and final soil moisture was zero for both CTRL and RFD1, the difference in crop evapotranspiration (and hence yield) from CTRL is proportional to the difference in total drainage between RFD1 and

CTRL. For all soil classes, early rainfall peak (RFD1) resulted in more water loss through deep percolation than for CTRL, since crop water demand was low (Figure 6.11). This resulted in less water usage by crops under RFD1 compared to CTRL. As soil water retention increases, the drainage loss for both RFD1 and CTRL were proportionally reduced – maintaining a rather constant yield difference.

For RFD2, water demand was better met than RFD1 since most of the rainfall was in the later crop season where crop water demand was higher. Therefore, the yield reduction from CTRL was lesser for all soil classes. For Soil Class 1, water loss through drainage in CTRL was marginally higher than RFD2 resulting in lower overall crop water usage for CTRL than RFD2. Therefore, the simulated yield was actually higher when compared to CTRL. As the soil water holding capacity increased, more soil water was available to meet crop water demand for CTRL as rainfall diminished at the late crop season. On the other hand, benefits to the crop under RFD2 with increased soil water retention are marginal (Figure 6.11). Hence yield reduction from CTRL for RFD2 increases with soil water capacity.

The difference in yield between RFD3 and CTRL depended primarily on the relative magnitude of the seasonal rainfall difference, and the drainage difference since end-of-season crop water was zero for RFD3 (Figure 6.11). As soil water increased, the difference in water usage increased as crops under CTRL were able to better utilise the rainfall to meet evapotranspirative demands, resulting in increasing differences in yields between RFD3 and CTRL.

The reason for the overall lower sensitivity in yield to perturbations in rainfall distribution for CSWB compared to GLAM in the idealised experiments was due to the absence of an environmental-crop feedback in CSWB since the crop coefficient ( $K_c$ ) did not change in response to soil water condition. Potential crop water demand in the CSWB model is therefore independent of crop water-stress. This limits the difference in the total crop evapotranspiration from CTRL (< 25% for RFD1, RFD2 and RFD3). In contrast for GLAM, water deficits can result in non-optimal leaf and root development that limits

transpiration. This amplifies the difference of biomass accumulation when compared to a crop with adequate water. A good illustration of this case is the relative response to RFD2 compared to CTRL for the two models: For GLAM, maximum transpiration attained was lower than CTRL while in CSWB, maximum evapotranspiration for RFD2 was as high as CTRL.

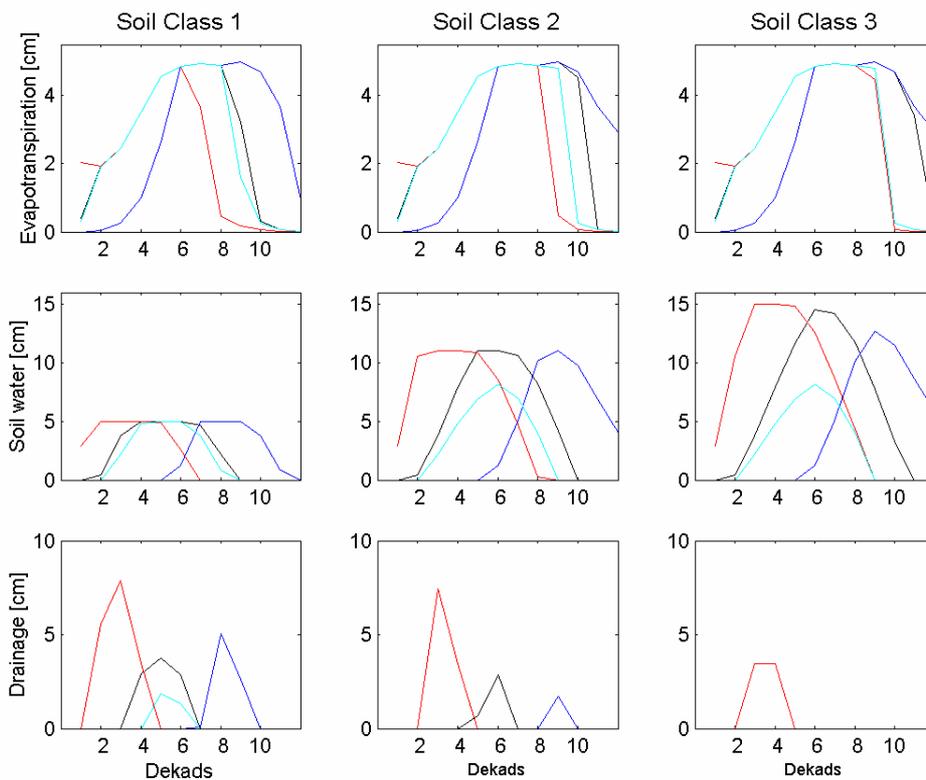


Figure 6.11 Time series of various model crop parameters for the idealised experiments for CSWB. Legend: Black: CTRL; Red: RFD1; Blue: RFD2; Cyan: RFD3. Soil Class represent soil types with  $\theta_{dul} = 0.14$  (1),  $\theta_{dul} = 0.20$  (2) and  $\theta_{dul} = 0.24$  (3).

An analytic form of the fractional yield difference for the CSWB model,  $f$ , can in fact be obtained from Equation 6.16 and 6.17:

$$f(E) = \frac{k_y(E - E')}{E_m - k_y(E_m - E')} \quad 6.20$$

where  $E_m$  is the total seasonal potential evapotranspiration,  $E$  and  $E'$  are the total actual evapotranspiration due to two different rainfall inputs. The function  $f(E)$  for the idealised experiment is shown in Figure 6.12. It is worth noting that a theoretical lower and upper bound exists for the extreme case for small  $E$  or when it approaches its potential ( $E_m$ ). For the idealised experiment,  $E/E_m$  was about 0.6 or greater for all soil classes. To achieve a yield difference of a magnitude comparable with GLAM (~30%),  $E/E_m$  needs to be about 0.5 or lower for all soil classes.

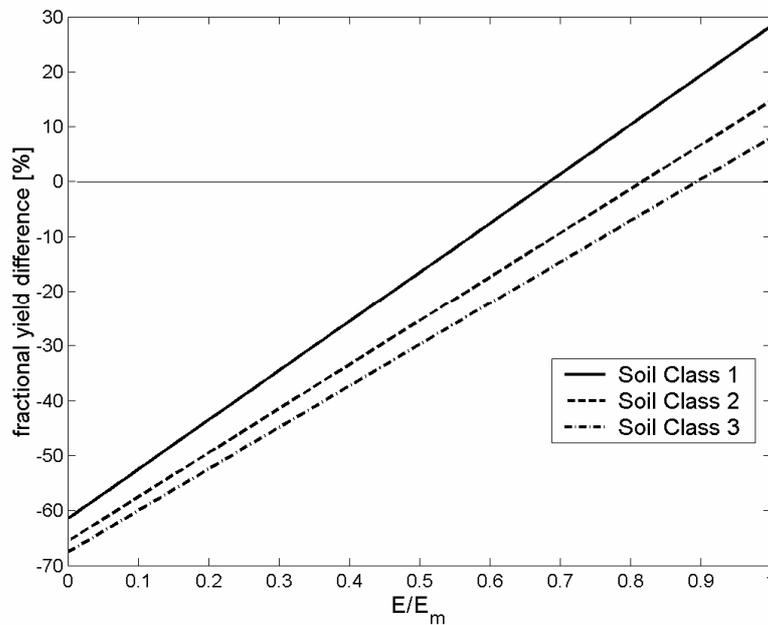


Figure 6.12 Fractional yield difference for CSWB for the various soil classes as a function of the total seasonal crop evapotranspiration.

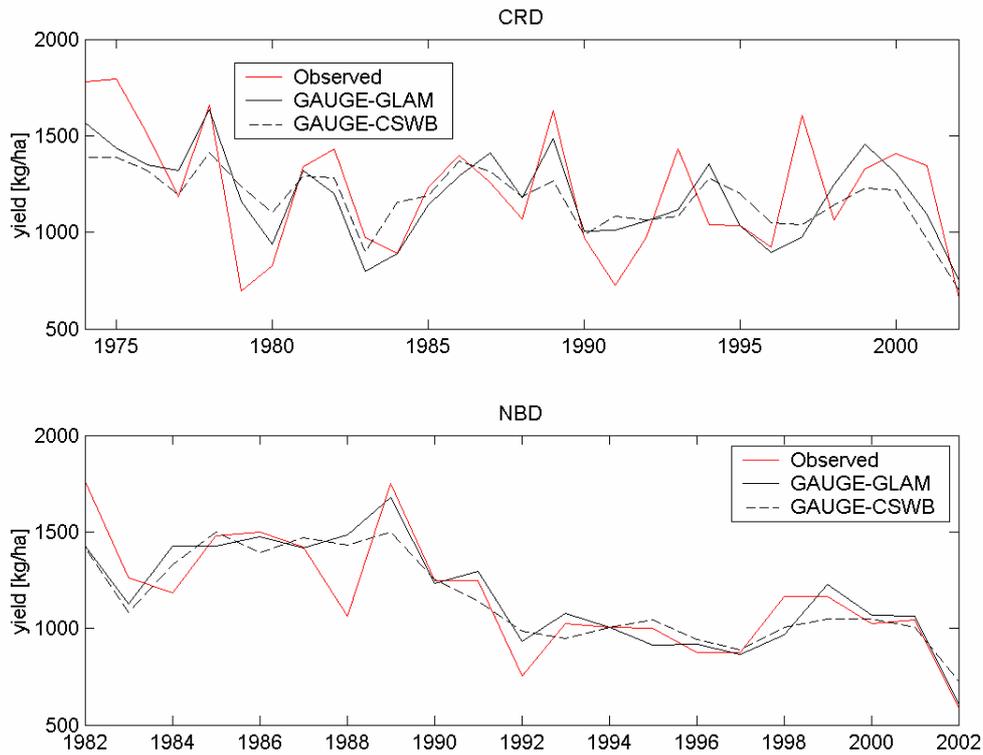


Figure 6.13 GAUGE-GLAM and GAUGE-CWSB model prediction of divisional groundnut yields compared with observed yield. Model yields are formed by adding the respective estimated trends of observed divisional yields (Chapter 2).

		MAE [kg/ha]	RMSE [kg/ha]	$r^2$
CRD	GLAM	171 [130 216]	222 [166 277]	0.47 [0.26 0.68]
	CSWB	204 [157 253]	258 [295 306]	0.32 [0.15 0.51]
NBD	GLAM	98 [60 140]	149 [90 197]	0.41 [0.17 0.67]
	CSWB	117 [81 157]	157 [111 198]	0.33 [0.15 0.55]

Table 6.2 Accuracy measure of GLAM and CSWB model for predicting groundnut yield for CRD and NBD. All computation with respect to detrended yields. Bracketed denotes the 90% confidence interval obtained from  $10^4$  bootstrap iterations.

## 6.6 Crop model prediction using gauge data

In this section, performance of the gauge-rainfall driven GLAM and CSWB in hindcasting the divisional groundnut yield for the Gambia is presented. The results from this section are used for skill comparisons when the crop models are driven by satellite-based rainfall estimates.

The time series of calibrated GLAM and CSWB simulated groundnut yields for the two Gambian divisions, using gauge-estimated rainfall and planting dates as inputs (denoted as GAUGE-GLAM and GAUGE-CSWB respectively hereafter), are shown in Figure 6.13. Three skill measures namely, mean absolute error (MAE), root mean square error (RMSE) and correlation ( $r^2$ ) with respect to the (detrended) yields were used to compare the performance of the model in hindcasting the observed yields, and their values are shown in Table 6.2. Both GLAM and CSWB showed better skill in hindcasting observed yields for NBD than that of CRD in terms of MAE and RMSE. Between models, GLAM outperformed CSWB for both divisions. This was due to GLAM's better ability in replicating the interannual variability of the groundnut yield. In order to test the significance of the difference in the inter-model skills, a re-sampling procedure (Wilks, 1995, Section 5.3.2) using  $10^4$  bootstrap iterations were used to obtain the 90% confidence interval for each skill measure. Comparison of the bootstrapped confidence interval of the various skill measures (Table 6.2) suggested that the difference might not be significant. This was supported by using the Wilcoxon signed rank test (Wilks, 1995, Section 5.3.1) on the differences of the absolute errors, and hypothesis testing on the differences in the correlation (Hinkle *et al.*, 1988), which suggested no significant difference between the model yield predictions for both divisions.

## 6.7 Application of RFE to crop models

Two methods of implementing the satellite-based rainfall estimates to drive crop models were used:

1) *Deterministic Method*: TAMSAT1 was used to drive both GLAM and CSWB (denoted as TAMSAT1-GLAM and TAMSAT1-CSWB respectively hereafter) to hindcast groundnut yields for the two divisions. Daily TAMSAT1 from 1988 to 2002 were available as inputs. Gauge-estimated divisional rainfall (GAUGE) was used for days with missing TAMSAT1. The planting dates for TAMSAT1-GLAM were recomputed from the planting date algorithm discussed in Section 6.3 using TAMSAT1. All other weather inputs as well as model parameters remained the same as GAUGE-GLAM. For TAMSAT1-CSWB, daily TAMSAT1 was aggregated in time to form dekadal TAMSAT1 from 1988-2002. The annual planting dekad was determined from the planting date used in TAMSAT1-GLAM, while all other weather inputs and model parameters in the TAMSAT1-CSWB were similar to that of GAUGE-CSWB.

2) *Ensemble Method*: For each year, 100 realisations of daily rainfall time series conditioned to the METEOSAT CCD were generated using the SIMU algorithm (Chapter 5), to drive the crop models. For each model run, a different planting date was obtained through the planting date algorithm. In this way, an ensemble of 100 simulated groundnut yields was obtained each year from 1996 to 2002 for both CRD and NBD. The period chosen is the longest continuous range of years with more than 95% of daily CCD data available. In order to implement the ensemble method for CSWB, appropriate aggregation to dekadal rainfall and conversion of planting dates to planting dekads was carried out. The ensemble GLAM and CSWB setups are abbreviated as ENS-GLAM and ENS-CSWB respectively hereafter.

#### Uncertainty assessment methods

Three further ensemble experiments similar in setup as the Ensemble Method, for both crop models were conducted: The first two, EXPA and EXPB, were designed to assess the contribution to the variability in the ensemble crop yield obtained from the Ensemble Method due to the uncertainties in 1) in-season rainfall (the randomness of the satellite rainfall estimates from planting to harvest) and 2) the stochastic planting dates. The third experiment, EXPC, assessed the simulated yield variability due to uncertainty in the planting date algorithm. All three experiments focused on simulated yields for the same

time period as the Ensemble Method (1996-2002). The details of these experiments was as follows:

1) In EXPA, the planting dates were the same as those used in the Deterministic Method (fixed for each year). For each year, 100 realisations of the rainfall time series generated from SIMU algorithm were used as crop model inputs as in Ensemble Method.

2) EXPB used TAMSAT1 crop model inputs. Therefore the rainfall amount received at each time step within the crop models was deterministic. The planting date for each year was allowed to vary representing the effect of the uncertainty that TAMSAT1 have on the planting dates. For this experiment, the planting dates used in the Ensemble Method were used to obtain 100 realisations of simulated yields each year.

3) EXPC was similar to EXPB (deterministic rainfall input from TAMSAT1, but varying planting dates). However, in EXPC, simulated crop yields from GLAM (CSWB) were obtained with planting dates vary within a set window of  $\pm 10$  days ( $\pm 1$  dekad) about the planting date used in the Deterministic Method.

A summary of the configurations of the three experiments is given in Table 6.3.

	Planting date	Rainfall inputs	Ensemble size
EXPA	TAMSAT1-X	SIMU-X	$100 \text{ yr}^{-1} \times 7 \text{ yr per division}$
EXPB	SIMU-X	TAMSAT1-X	$100 \text{ yr}^{-1} \times 7 \text{ yr per division}$
EXPC	$\pm 10$ days about TAMSAT1-X	TAMSAT1-X	GLAM: $21 \text{ yr}^{-1} \times 7 \text{ yr per division}$ CSWB: $3 \text{ yr}^{-1} \times 7 \text{ yr per division}$

Table 6.3 Experiments for yield uncertainty assessment due from various error sources. TAMSAT1-X denotes the same data used in either TAMSAT1-GLAM or TAMSAT1-CSWB; SIMU-X means same data as either ENS-GLAM or ENS-CSWB.

### Skill measures

The differences in predicted yields due to different rainfall inputs (TAMSAT1 and GAUGE) for a given crop model assessed using the mean absolute difference between TAMSAT1 and GAUGE driven model yield predictions (MAD) and the correlation ( $r^2$ )

between the predicted yields using the two different rainfall inputs ( $\rho'$ ). The prediction skill of the TAMSAT1-driven models with respect to the observed yield is assessed using the mean absolute difference between predicted and observed yields (MAE) and the correlation between them ( $\rho$ ). The measure selected to quantify the ensemble yield spread for Ensemble Method, EXPA and EXPB was the semi-quartile range (SQR) defined as half of the 75<sup>th</sup> and 25<sup>th</sup> percentile of the yield ensemble. For EXPC, the half-range of ensemble yield (half the difference between the maximum and minimum yield in the ensemble, abbreviated as HR hereafter) was used instead of SQR as the measure, since the entire range of predicted yields within the planting date windows are of interest in this case.

### 6.7.1 GLAM

#### TAMSAT1-GLAM

The predicted divisional groundnut yields for TAMSAT1-GLAM are shown in Figure 6.14 and Table 6.4 summarises the various skills measures when compared to prediction from GAUGE-GLAM, as well as the observed yields. Visual inspection of Figure 6.14 suggest that except for NBD yields for 2002, TAMSAT1-GLAM is as skillful as GAUGE-GLAM in hindcasting the divisional yields; The Wilcoxon signed rank test performed on the absolute prediction errors for the two divisions combined, and excluding years where missing TAMSAT1 was substantial (>10%), showed that the difference in the overall predictive skill for TAMSAT1-GLAM and GAUGE-GLAM is not significant ( $p = 0.25$ ). It was further noted that the correlation between predicted and observed yields (combined divisions, excluding years with >10% missing TAMSAT1 data) is 0.22 for both TAMSAT1-GLAM and GAUGE-GLAM. This is an encouraging result for using TAMSAT1-GLAM for crop yield prediction, in view of the lower correlations between TAMSAT1 and GAUGE for CRD ( $r^2 = 0.47$ ) and NBD ( $r^2 = 0.56$ ).

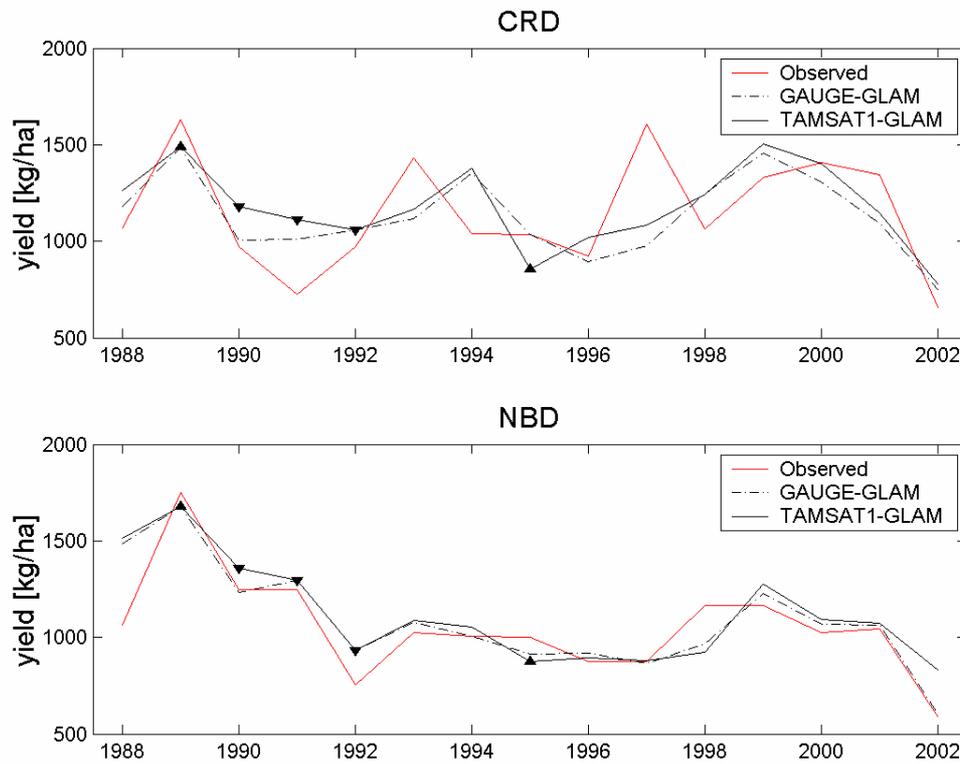


Figure 6.14 Divisional observed and GLAM predicted yield for 1988 to 2002. Model yields are formed by adding the respective estimated trends of observed divisional yields (Chapter 2). Legend: Red line: Observed; Dash line: GAUGE-GLAM; Solid Black line: TAMSAT1-GLAM; '▼': 10-20% missing CCD data; '▲': > 20% missing CCD data.

	CRD	NBD	Combined
MAD [kg/ha]	71	42	57
$\rho$	<b>0.85</b>	<b>0.82</b>	<b>0.83</b>
MAE [kg/ha]	206	120	163
$\rho^2$	<b>0.34</b>	<b>0.37</b>	<b>0.34</b>

Table 6.4 Various skill measures for predicted yield from TAMSAT1-GLAM with respect to predicted yield from GAUGE-GLAM and the observed yields. Bold values for  $\rho$  and  $\rho^2$  are significant at 95% level.

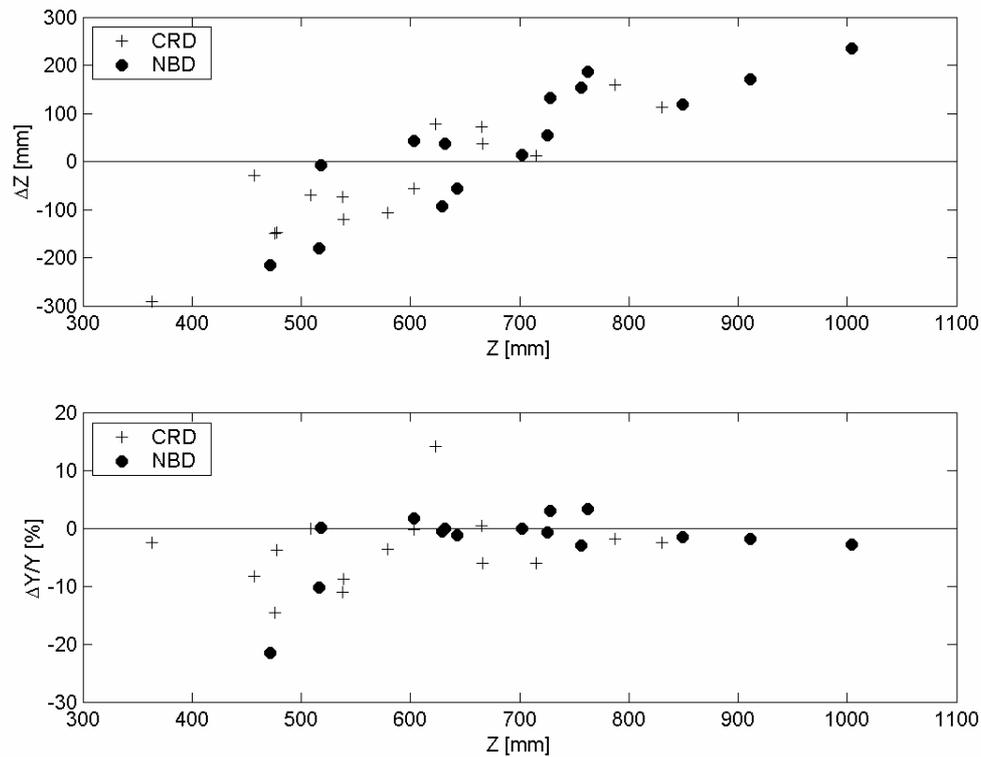


Figure 6.15 Scatterplot of difference in total rainfall received,  $\Delta Z$  (GAUGE-GLAM – TAMSAT1-GLAM) against total rainfall received by GAUGE-GLAM,  $Z$  (top) and fractional difference in predicted yield ( $\Delta Y/Y$ ) against  $Z$  (bottom).  $Y$  is defined as the yield from GAUGE-GLAM and  $\Delta Y$  is the predicted yield difference (GAUGE-GLAM - TAMSAT1-GLAM).

To further investigate the difference in model yield response due to the difference in rainfall inputs, HI years and LO years were defined respectively as years where the total rainfall received by GAUGE-GLAM is greater than 700mm and less than 500mm. Annual total rainfall received ( $Z$ ), the differences in rainfall received ( $\Delta Z$ ), and the fractional differences in yield between GAUGE-GLAM and TAMSAT1-GLAM ( $\Delta Y/Y$ ) for both divisions are summarised in Figure 6.15. Two features are prominent:

1) TAMSAT1-GLAM received more (less) rainfall for LO (HI) years. The reason for this is that the satellite rainfall estimates did not reproduce the observed interannual variability of the rainfall. This resulted in a tendency to overestimate (underestimate) rainfall for LO (HI) years.

2) The variability in yield difference between GAUGE-GLAM and TAMSAT1-GLAM was high for LO years and low for HI years. For HI years, simulated yield differences were slight ( $< 6\%$ ) although corresponding rainfall received for TAMSAT1-GLAM was less than that of GAUGE-GLAM (as much as  $24\%$ ). For LO years, although TAMSAT1-GLAM received more rainfall than GAUGE-GLAM, differences in corresponding model yields in these cases had larger variability; for 2002,  $\Delta Z = 80\%$  and  $\Delta Y/Y = 2\%$  for CRD, while  $\Delta Z = 46\%$  and  $\Delta Y/Y = 22\%$  for NBD.

The magnitude of the differences in predicted yields for TAMSAT1-GLAM and GAUGE-GLAM can be explained as follows: For HI years, despite underestimation of rainfall, TAMSAT1 estimates were generally high enough to maintain soil moisture for crop to transpire at a level similar to that of GAUGE-GLAM. Hence, there is little difference between TAMSAT1 and GAUGE driven GLAM (Figure 6.15). For LO years, the difference in amount, as well as distribution between the two rainfall inputs can result in varying level of crop water stress through time, leading to diverging or converging crop development pathways. This is best illustrated using the crop development for 2002 shown in Figure 6.16. For NBD, differences in yields predicted by TAMSAT1-GLAM and GAUGE-GLAM were due solely to differences in the daily rainfall amount, since the planting dates for the division were the same. In this case, overestimation in TAMSAT1 in the earlier crop stages enabled TAMSAT1-GLAM to maintain higher daily transpiration when compared to GAUGE-GLAM resulting in more robust crop growth evident in the leaf and biomass development. For CRD, the planting date estimated by TAMSAT1 was 2 weeks earlier than that computed using GAUGE, effectively shifting the rainfall distribution with respect to the crop development stages. This shift in TAMSAT1-GLAM rainfall induced a prolonged low-rainfall period during the early growing season. Consequently, crop development is similar in the TAMSAT1-GLAM and GAUGE-GLAM, despite a  $\Delta Z$  of 300 mm. When the gauge rainfall estimated planting date was used instead, TAMSAT1-GLAM overestimated the yield. This example serves to highlight GLAM's sensitivity to intraseasonal rainfall variability within the growing season, especially when water deficits are frequent in the early season, as well as the effect on yield due to planting date uncertainties.

	CRD					NBD				
	$\delta Y_A$ [kg/ha]	$\delta Y_B$ [kg/ha]	$\delta d_B$ [days]	$\delta Y_{ENS}$ [kg/ha]	$\Delta Y_C$ [kg/ha]	$\delta Y_A$ [kg/ha]	$\delta Y_B$ [kg/ha]	$\delta d_B$ [days]	$\delta Y_{ENS}$ [kg/ha]	$\Delta Y_C$ [kg/ha]
1996	40	17	1.0	45	144	25	11	1.0	29	156
1997	36	7	1.0	36	61	43	2	1.0	44	16
1998	3	79	4.5	82	109	0	38	2.5	41	162
1999	0	11	2.0	11	89	0	31	2.0	31	61
2000	0	38	3.0	38	64	0	15	2.5	15	87
2001	14	4	0.5	15	123	0	33	1.0	33	120
2002	19	15	2.5	32	52	16	8	4.0	17	29
Mean	16	24	2	37	92	12	20	2	30	90

Table 6.5 Difference between ensemble yields using fix sowing date and variable sowing dates for CRD and NBD.  $\delta Y_A$ : SQR for ensemble yields for EXPA;  $\delta Y_B$ : SQR for ensemble yields EXPB;  $\delta Y_{ENS}$ : SQR for ensemble yields from ENS-GLAM;  $\delta d_B$ : SQR of the planting dates for the ensemble CSWB ( $\delta Y_{ENS}$ );  $\Delta Y_C$ : HR for ensemble yields from EXPC.

#### Effects of random errors in RFE

The summary of the ensemble yields simulated for the two divisions from ENS-GLAM, EXPA, EXPB and EXPC are tabulated in Table 6.5 and illustrated graphically in Figure 6.17. The main results are as follows:

- 1) Uncertainty in simulated yields due to random error in TAMSAT1 was small. The mean SQR for the two divisions (30 kg/ha for NBD and 37 kg/ha for CRD) were about half the mean absolute yield difference (combined divisions from 1996-2002) between TAMSAT1-GLAM and GAUGE-GLAM (61 kg/ha), and was about 20% of the mean absolute error (combined divisions from 1996-2002) between TAMSAT1-GLAM and observed yields (143 kg/ha).
- 2) The impact from the variability in the in-season rainfall (EXPA), and planting dates (EXPB), due to random error in TAMSAT1 varies from year to year.

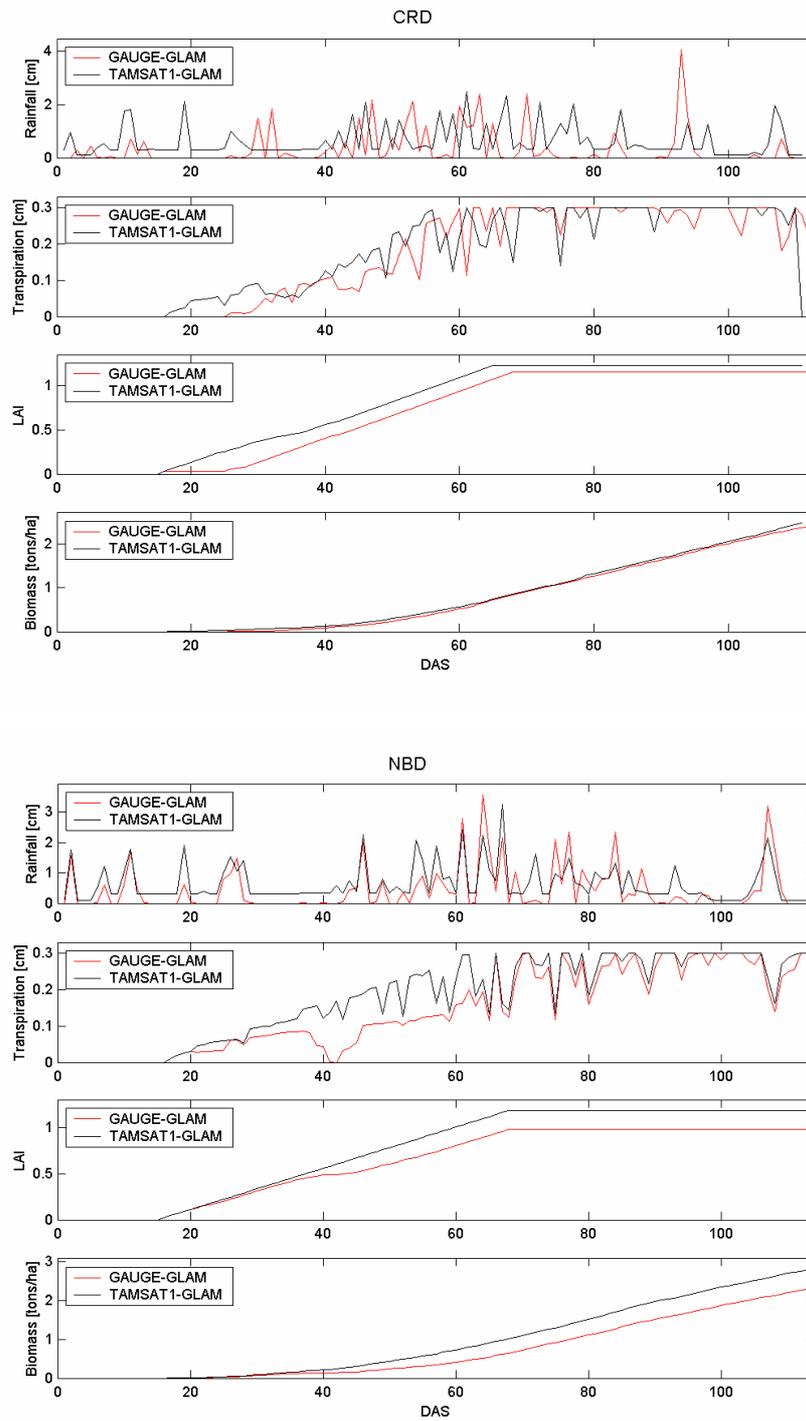


Figure 6.16 Daily rainfall and various model crop variables for CRD and NBD in 2002 from GLAM. DAS means days after sowing.

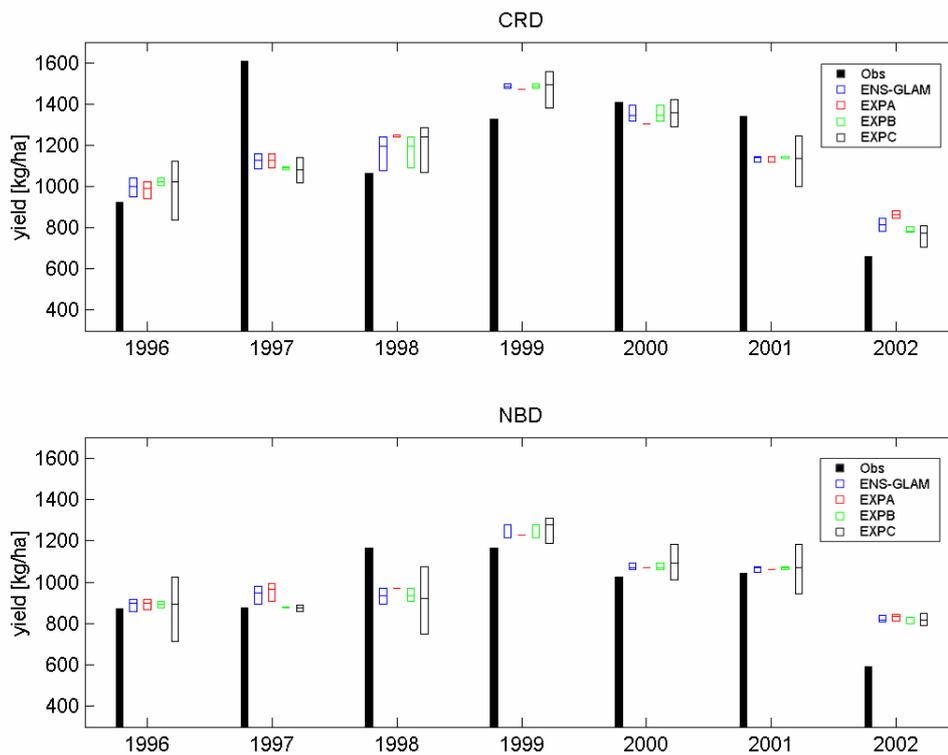


Figure 6.17 Divisional groundnut yield uncertainties from ENS-GLAM, fix date (EXPA), varying date (EXPB), varying date with 10 days windows (EXPC). All open boxes except EXPC show the 25<sup>th</sup> and 75<sup>th</sup> percentile. For EXPC, open box denotes the entire range of yields within the 11 days window (planting day  $\pm$  10 days). The line across the box shows the median.

3) Uncertainty in the planting date algorithm (EXPC) had a larger effect on simulated yields than random error in TAMSAT1 (ENS-GLAM, EXPA, and EXPB). The average HR of the yields for a  $\pm 10$ -day window of uncertainty for both divisions combined was 89 kg/ha, with significant interannual variability (CV= 62%). Although smaller than the mean absolute error of TAMSAT1-GLAM (143 kg/ha), it can be seen from Figure 6.17 and Table 6.5 that the impact on yields due to uncertainty in the planting date algorithm used was greater than those due to stochastic uncertainties in TAMSAT1.

The conclusion that can be drawn from these experiments is that the impact on the simulated yield due to random error in TAMSAT1 is secondary compared to error in capturing the sowing date, in the present case study of Gambian groundnut. This is encouraging from the perspective of applying TAMSAT1-GLAM for crop yield

prediction in Africa. It must be noted that the degree of sensitivity to random errors in TAMSAT1 depends on the degree of non-linearity of the crop model processes, as well as to the spatial scale at which yield is simulated; larger uncertainties in the simulated yield could arise if a more non-linear crop model is used for yield prediction operating at a smaller spatial scale.

### Effects of bias

The uncertainties in planting date algorithm and random error in TAMSAT1 are apparently unable to account for the overestimation of TAMSAT1-GLAM yield for NBD in 2002 (Figure 6.17). The overprediction in this case may be explained by a positive bias in TAMSAT1 for 2002 (see Chapter 4). The impact of such systematic biases was assessed by rerunning GLAM for NBD using TAMSAT1 for 2002 that were bias-corrected for each month within the rainy season. The bias correction was achieved by multiplying a reduction factor of 0.7 to the daily rainfall amount of the targeted month to attain a 30% reduction on the monthly total rainfall. The reduction factor of 0.7 was adopted since such magnitude of reduction in the TAMSAT1 model parameters has been shown to reduce the bias of the seasonal rainfall total for 2002 when applied across July and August (Chapter 4, Section 4.5). The experiment was repeated using bias correction for each month from June to September and the results of these experiments are summarised in Table 6.6. The bias-corrections on the rainfall for all months except for July did not result in significant impact on the predicted yield. Simulated yield using TAMSAT1 with bias-correction for July rainfall was close to that of GAUGE-GLAM (591 kg/ha), achieving a 26% reduction in yield overestimate, with only 7% reduction in total rainfall received by the model.

	Months where bias correction (30%) is applied on TAMSAT1				
	Jun	Jul	Aug	Sep	
Total seasonal rainfall [mm]	720	672	645	670	[567]
Yield [kg/ha]	830	618	805	830	[591]

Table 6.6 Rainfall and GLAM yield due to 'bias correction' in TAMSAT1 for different months. Bracket figures are the GAUGE rainfall estimate and the observed yield for 2002 for NBD respectively.

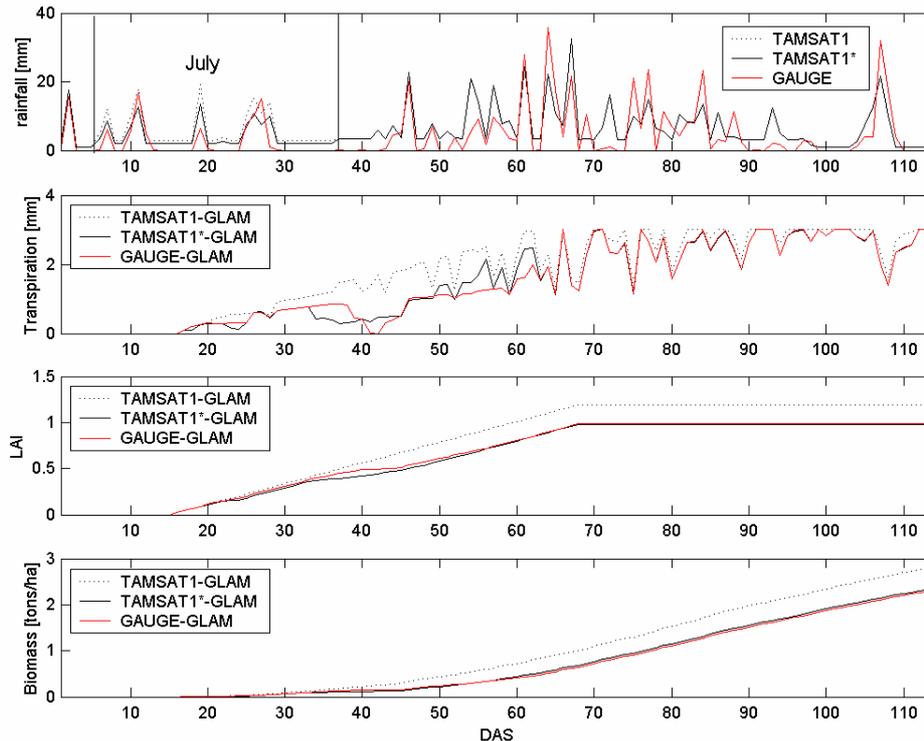


Figure 6.18 Rainfall and crop variables in GLAM for NBD, 2002 using different rainfall inputs. TAMSAT1\* denotes TAMSAT1 with ‘bias-correction’ for July.

The reason for the large reduction in the predicted yield when bias-correction was applied to July rainfall can be seen from Figure 6.18. July for 2002 was characterised by a prolonged period of dryness punctuated by short periods of rainfall. Daily soil moisture available for transpiration became sensitive to the amount of rainfall received and the antecedent soil moisture in the early season. The bias-correction on July reduced the daily rainfall amount especially for several days with significant rainfall. This correction limited the transpiration at an early stage of the crop development, thereby reproducing the sub-optimal crop growth similar to GAUGE-GLAM. Such dramatic effects of yield reduction by inducing moisture deficit at the early crop stages were similar to that observed in the idealised experiments in Section 6.5. On the other hand, the bias-correction in TAMSAT1 for August and September has a limited effect on the crop development (not shown) to cause a yield reduction as significant as those under bias-

corrected July rainfall. It is emphasised that the above observations hold only for the current imposed soil and yield gap parameters; for a low soil water holding capacity, reducing the September rainfall, may result in water deficit in the later crop stages and decelerate biomass accumulation significantly, similar to that observed in the idealised experiments with Soil Class 1 (Section 6.5).

## 6.7.2 CSWB

### TAMSAT1-CSWB

The performance of TAMSAT1-CSWB is summarised in Table 6.7. Agreement with the GAUGE-CSWB yields was poor in terms of absolute difference and correlations. The disparity between TAMSAT1-CSWB and GAUGE-CSWB as well as the observed yield can be seen from Figure 6.19. To ascertain whether the poor agreement with GAUGE-CSWB was due to differences in planting dates used, the divisional groundnut yields were hindcasted with TAMSAT1 driven CSWB using the same planting dates as GAUGE-CSWB. Correlation between predicted yields did not improve significantly (for CRD:  $\rho = 0.30$ ,  $p = 0.03$ ; NBD:  $\rho = 0.05$ ,  $p = 0.25$ ), indicating that the model responses to TAMSAT1 were different from that of GLAM. This can be seen in the scatter plot of differences in yields,  $\Delta Y$ , and total rainfall received,  $\Delta Z$ , between TAMSAT1-CSWB and GAUGE-CSWB (Figure 6.20), where a proportional  $\Delta Y$ - $\Delta Z$  relationship can be discerned when TAMSAT1 overestimated ( $\Delta Z < 0$ ). The correlation ( $r^2$ ) between  $\Delta Y$  and  $\Delta Z$  for all  $\Delta Z < 0$  (combined divisions) is 0.88 ( $p < 0.01$ ). This suggests that yield prediction error for CSWB strongly depends on the accuracy of TAMSAT1 in capturing the total seasonal rainfall, than the intraseasonal rainfall variability.

The reasons for the observed yield-water responses in the CSWB was due to the lack of environment-crop feedback ( $K_c$  invariant to water deficits) and the insensitivity of yield to water satisfaction for different crop growth stages (constant yield coefficient,  $k_y$ ). This implied that  $\Delta Y$  was related to the environmental water budget through Equation 6.19 (ignoring the difference in the initial soil moisture at planting):

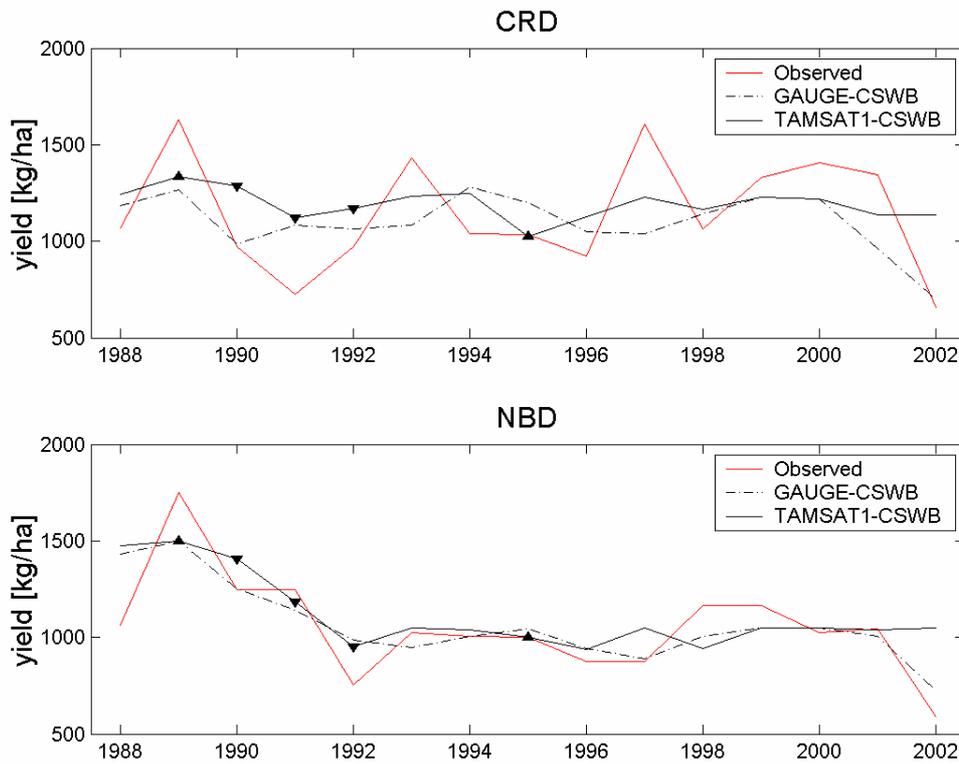


Figure 6.19 Divisional observed and CSWB model predicted yield for 1988 to 2002. Model yields are formed by adding the respective estimated trends of observed divisional yields (Chapter 2). Legend: ‘▼’: 10-20% missing CCD data; ‘▲’: > 20% missing CCD data.

	CRD	NBD	Combined
MAD [kg/ha]	121	69	95
$\rho$	0.03	0.06	0.07
MAE [kg/ha]	230	145	188
$\rho^2$	<b>0.38</b>	0.02	<b>0.20</b>

Table 6.7 Various skill measures for predicted yield from TAMSAT1-CSWB with respect to predicted yield from GAUGE-CSWB and the observed yields. Bold values are correlations that are significant at 95% level.

$$\Delta Y = \alpha [\varepsilon \Delta Z - \Delta D_R - \Delta S]$$

where  $\Delta Z$ ,  $\Delta D_R$  and  $\Delta S$  are respectively the difference in total rainfall received, total drainage of excess rainfall, and the remaining soil moisture at end of harvest between GAUGE-GLAM and TAMSAT1-CSWB. In the above equation  $\varepsilon$  is the rainfall attenuation factor and  $\alpha$  is the water use efficiency which is a function of the total seasonal evapotranspiration ( $E_m$ ). Theoretically,  $E_m$  in Equation 6.19 is identical only when the planting dates for both TAMSAT1-CSWB and GAUGE-CSWB is identical. Although the planting dates in both model runs were generally different, it is a valid approximation to use a common  $E_m$  value in this case as the planting dates computed using TAMSAT1 and GAUGE did not differ greatly to alter  $E_m$  significantly in the CSWB model.

If the dekadal rainfall at any time during the crop season and total rainfall for both rainfall inputs are sufficiently low so that drainage and end-of-season soil moisture are negligible, the computed difference in GAUGE-CSWB and TAMSAT1-CSWB yields,  $\Delta Y^*$ , is approximately linear with respect to the difference in total seasonal rainfall amounts given by

$$\Delta Y^* = \alpha \varepsilon \Delta Z \quad 6.21$$

The scatter plot of  $\Delta Y^*$  against  $\Delta Y$  (Figure 6.21) shows a good agreement in particular for years where satellite rainfall overestimated the seasonal rainfall total ( $\Delta Y < 0$ , see Figure 6.20). For years where rainfall in TAMSAT1-CSWB and GAUGE-CSWB were sufficient to meet the crop water demands (example 1999), simulated yields from both TAMSAT1 and GAUGE driven CSWB were similar despite the difference in rainfall inputs;  $\Delta Y$  in these cases differ from  $\Delta Y^*$ , as differences in soil recharge and drainage were important to account for the differences between the predicted yields.

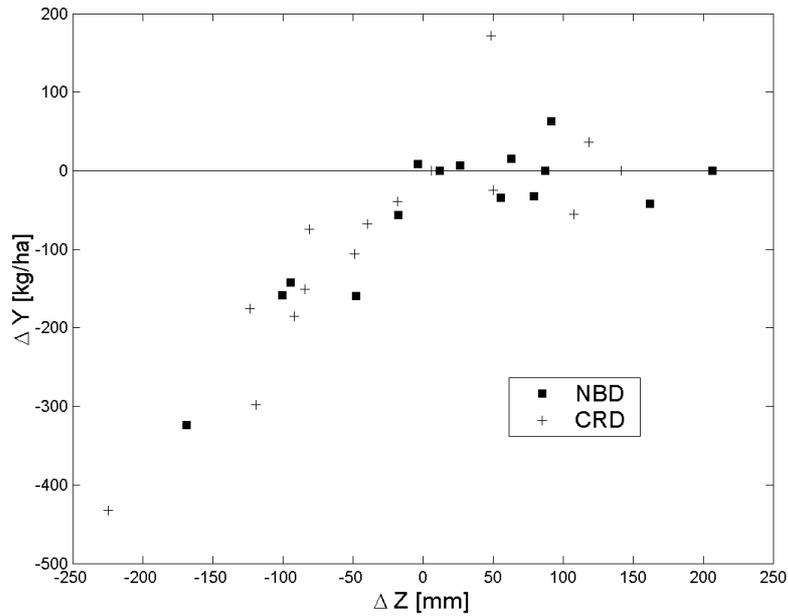


Figure 6.20 Scatter plot of differences between modeled divisional groundnut yield ( $\Delta Y$ , GAUGE-CSWB - TAMSAT1-CSWB) against difference in total rainfall received ( $\Delta Z$ ) due to difference in the TAMSAT1 and gauge rainfall inputs.

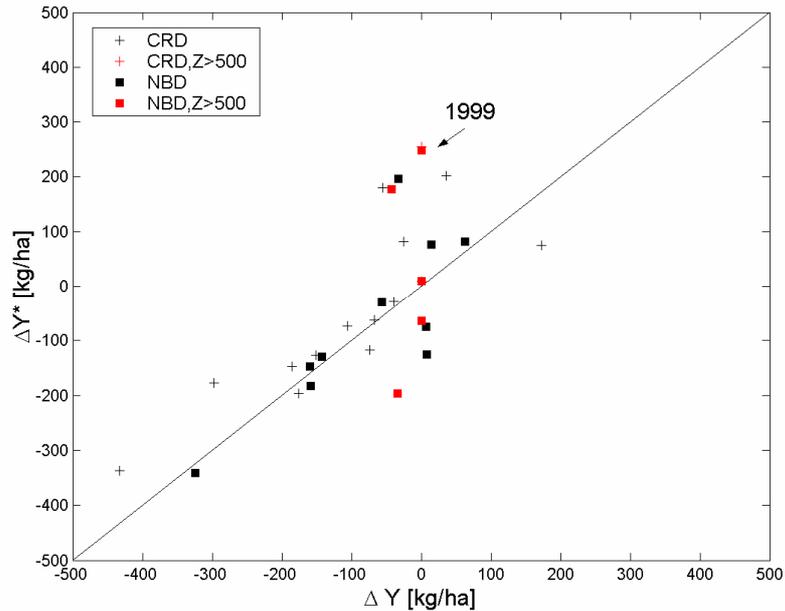


Figure 6.21 Scatterplot of computed yield difference (Equation 6.21) against the observed yield difference (GAUGE-CSWB – TAMSAT1-CSWB) for NBD and CRD. Solid line is the 1:1 line. Red markers are when TAMSAT1-CSWB and GAUGE-CSWB received more than 500mm of rainfall in the growing season.

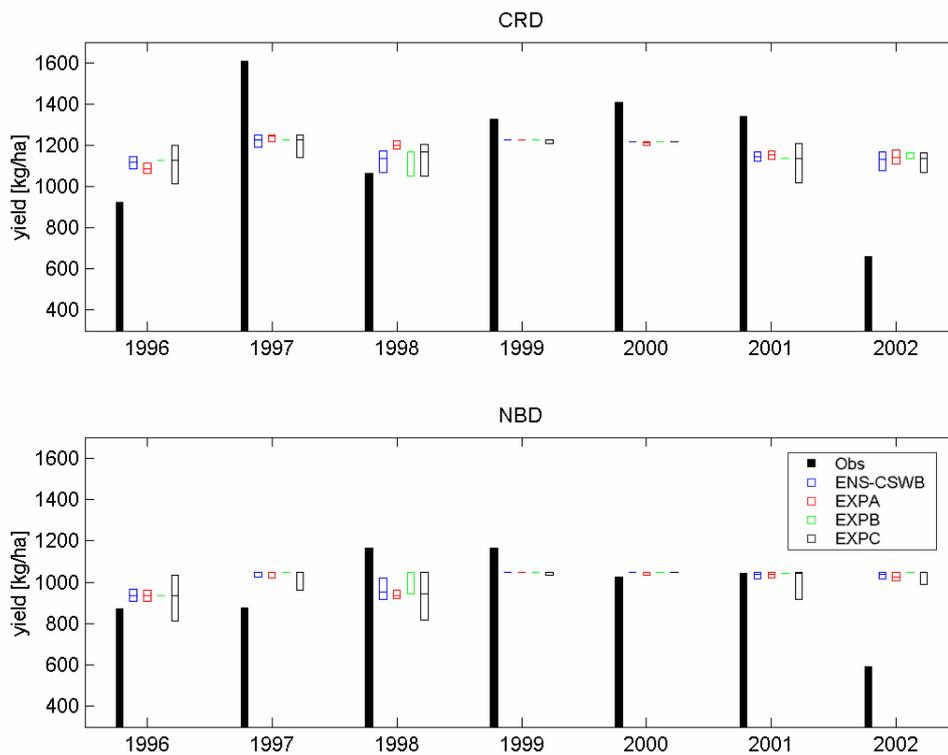


Figure 6.22 Divisional groundnut yield uncertainties from ENS-CSWB fix date (EXPA), varying date (EXPB), varying date with 1-dekad windows (EXPC). All open boxes except EXPC show the 25<sup>th</sup> and 75<sup>th</sup> percentile. For EXPC, open box denotes the entire range of yields within the planting window (planting day  $\pm$  1 dekad). The line across the box shows the median.

### Effects of random errors in RFE

Figure 6.22 shows the ensemble yields spread from ENS-GLAM, EXPA, EXPB and EXPC. The results of these experiments are as follows:

- 1) The simulated yield remained robust under random errors of TAMSAT1 (ENS-CSWB). The average SQR for the ensemble yield from ENS-CSWB was 42 kg/ha (combined divisions), compared to the mean absolute error of TAMSAT1-CSWB simulated yield with observed yield (220 kg/ha). The contribution to the predicted yield variability from random uncertainties in the in-season rainfall (EXPA) and the planting dates (EXPB) was about equal in magnitude. For years when yield was not water-limited (1999 and 2000), no variability in yields was observed.

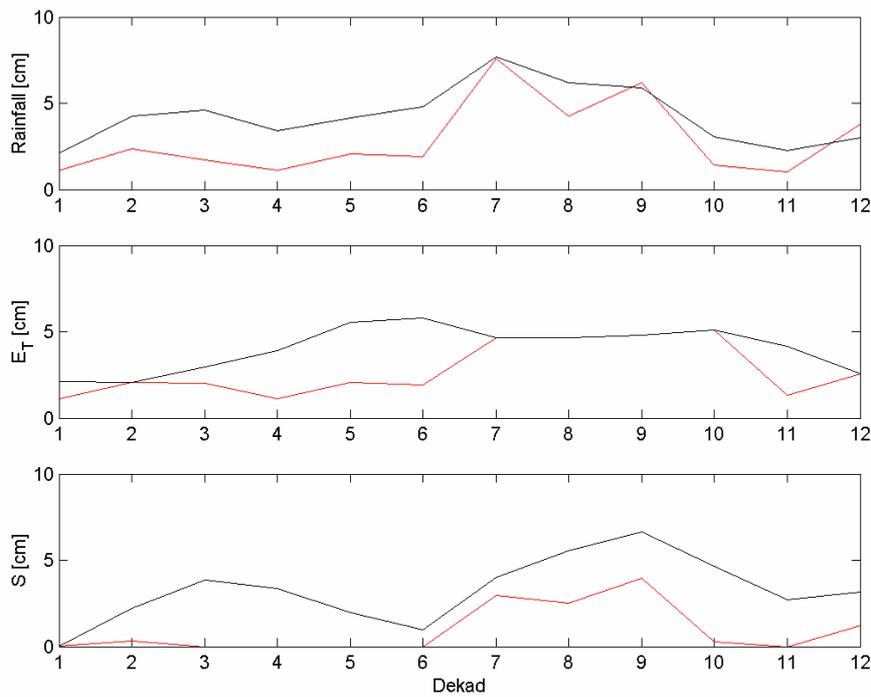


Figure 6.23 Time series of the primary model variables in the CSWB for NBD groundnut crop for the year 2002 using TAMSAT1 and gauge rainfall estimates.  $E_T$  is the crop evapotranspiration and  $S$  is the soil moisture. Legend: Red: GAUGE-CSWB; Black: TAMSAT1-CSWB.

2) Uncertainty in the planting date algorithm (EXPC) had a larger impact than the random errors in TAMSAT1 on the simulated CSWB yields. The magnitude of yield variability was less than that of GLAM on average. Imposing an uncertainty of a one dekad window about the planting date resulted in yield variability with average HR of 52 kg/ha, showing less sensitivity to uncertainty in planting dates compared to GLAM (HR = 89 kg/ha, combined division). For the high rainfall years, 1999 and 2000, simulated yields were robust even under a three-dekad (planting dekad  $\pm$  one dekad) range as there was sufficient rainfall to meet the potential crop water demand for all model runs. Hence, it can be concluded that similar to GLAM, effects on TAMSAT1-CSWB simulated yields due to random errors in TAMSAT1 are secondary compared to the impact from the uncertainty in the planting date algorithm.

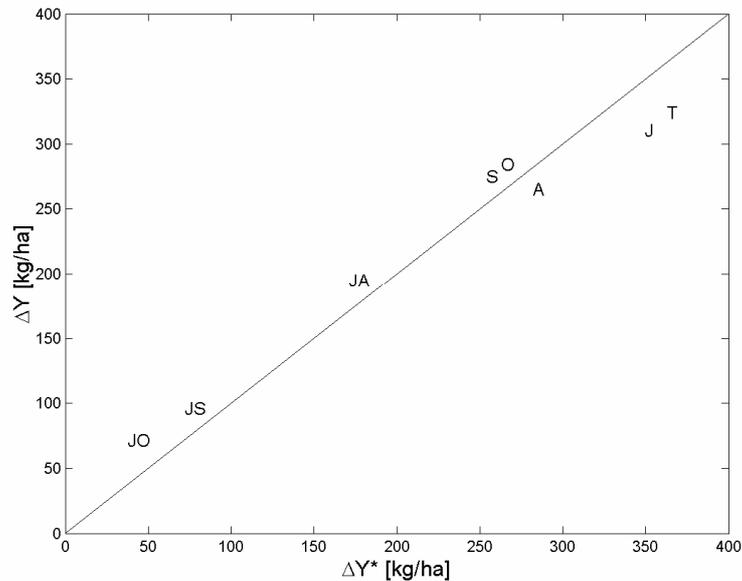


Figure 6.24 Scatter plot of reduction of yield prediction with respect to GAUGE-CSWB ( $\Delta Y$ ) against expected yield reduction computed from the reduction of total rainfall received by CSWB model from ‘bias-correction’ of TAMSAT1 ( $\Delta Y^*$ ; Equation 6.21). Solid line is the 1:1 line. Letters on the scatter plot are indexes represents CSWB runs with ‘bias correction’ applied on TAMSAT1 for the following months: “T” No reduction (TAMSAT1); “J”: July; “A”: August; “S”: September; “O”: October; “JA”: July and August; “JS”: July to September; “JO”: July to October.

### Effects of bias

The bias-correction experiments described in the preceding section for GLAM were repeated with CSWB for NBD for the year 2002. Since CSWB yields have a strong response to the total seasonal rainfall amount, it was expected that any reduction in simulated yields from bias-correcting TAMSAT1 followed Equation 6.21. The water balance profiles for TAMSAT1-CSWB and GAUGE-CSWB (Figure 6.23) supports the validity of this assertion: Drainage (not shown) for both GAUGE and TAMSAT1 driven model was negligible, while differences in the end-of-season soil moisture between TAMSAT1-CSWB and GAUGE-CSWB was small (20 mm) compared to the difference in total rainfall amount received (120 mm); any reduction in TAMSAT1 estimates through bias-correction further reduced the effects of end-of-season soil moisture on yield, reinforcing the dependence of the simulated yields with the total rainfall received.

In addition to bias-correcting individual months in the TAMAST1 inputs, the experiment was extended to include bias reduction (with similar reduction factor of 0.7) on the rainfall time series for combinations of months: July and August, July to September and July to October. Figure 6.24 shows the effect of the various bias-corrections to TAMSAT1-CSWB yields compared to the simulated yield from GAUGE-CSWB; observed and theoretical yield reduction from Equation 6.21 agrees well ( $r^2 = 0.96$ ). Therefore, in contrast with GLAM, significant improvement in CSWB predicted yields could only be achieved when bias-correction on the rainfall is large enough to reduce the bias in the total seasonal rainfall amount.

## 6.8 Conclusion

Two crop models, GLAM and the CSWB were calibrated for divisional groundnut yield prediction for the Gambia. Sensitivity experiments with simple rainfall profiles as inputs to the calibrated crop models showed variability in yield responses to rainfall. GLAM is found to be more sensitive to intraseasonal rainfall distribution due to the non-linear interactions within the crop model; moisture deficits at early and late stages of groundnut development can have different impact on the final crop yield depending on the soil water capacity. CSWB on the other hand, exhibited a more linear yield-water response due to the lack of crop-water stress feedback within the model.

For the case study of divisional Gambian groundnut, GLAM performed better than CSWB in hindcasting of the observed yields when gauge-estimated rainfall was used as an input. But the differences in performance were not statistically significant. However, when satellite rainfall estimates (TAMSAT1) are used to drive the crop models, GLAM showed better reproduction of the crop yields than CSWB. Despite crop model sensitivities to rainfall inputs and the differences between gauge and satellite estimated rainfall, TAMSAT1-GLAM was able to capture the crop development and hence, predicted yield had high correlation with GAUGE-GLAM predictions ( $r^2=0.8$ ). Part of the reason for this was the ability of TAMSAT1 to capture the intraseasonal rainfall distribution. This demonstrates the feasibility of using TAMSAT1 driven GLAM in

operational yield forecasting where the crop model is known to be skillful in yield prediction. The ensemble yield prediction technique showed that GLAM yield was robust under random error in TAMSAT1. A bigger source of error may be the uncertainties in correctly capturing the planting date from planting dates algorithm. Bias in satellite rainfall estimates in the early crop season has been found to be an important source of error: For one particular year where TAMSAT1-GLAM overestimated the yield due to systematic bias in TAMSAT1, bias correction in the satellite estimates for only one month earlier in the crop season resulted in a 26% reduction in simulated yield. This bias-corrected yield was near to the observed yield, even though the reduction of bias in total seasonal rainfall was only 7%.

For CSWB, the difference in simulated yield using TAMSAT1 and gauge-rainfall inputs was large for years when GAUGE-CSWB yields were water limited. This was due to the tendency of TAMSAT1 to overestimate the seasonal rainfall amount for years of anomalously low rainfall. A theoretical linear yield-rainfall response for CSWB has been found to be useful in explaining the observed disparity between TAMSAT1-CSWB and GAUGE-CSWB. Ensemble yield predictions using CSWB with stochastic rainfall inputs indicated that the simulated yield was also fairly robust to the stochastic uncertainties of TAMSAT1. In contrast with GLAM, systematic bias in TAMSAT1-CSWB yield prediction was found to be strongly related to the overall bias in the total season rainfall amount for years especially when actual yields were water limited.

When satellite rainfall estimates driven crop models are used for crop yield prediction, the investigation in this chapter highlighted the following: If the RFE can capture the intraseasonal variability of rainfall, it can be used as inputs to GLAM for crop yield prediction even when some systematic bias in the rainfall estimates exist. To use an RFE-CSWB for crop yield forecasting, the satellite rainfall estimates need to capture the seasonal rainfall amount with sufficient accuracy to reliably forecast water-limited crop production.

## Chapter 7 Conclusions

### 7.1 Summary of main results

The main purpose of this thesis was to investigate the feasibility of using satellite based rainfall estimates as inputs into a process-based crop model, GLAM, for crop yield forecasting in Africa. Since process-based crop models are non-linear in their processes, a related issue of interest was the sensitivity of the predicted yield to uncertainties in rainfall inputs. The annual groundnut yields for two major groundnut cultivation divisions for the Gambia were used as case studies. To this end, daily satellite RFEs based on METEOSAT CCD were obtained. As it was found that a single CCD threshold temperature did not appear to delineate the rainy region for the Gambia effectively, a daily satellite RFE, TAMSAT1, based on the TAMSAT algorithm was developed. The novelty of the algorithm is that it comprises a complete statistical model describing the mixed distribution (with a finite probability mass at zero) of pixel rainfall for each CCD value. Under TAMSAT1, the probability of a pixel being rainy and its rainfall amount was then determined separately using historical gauge data.

For the case of Gambian daily rainfall, TAMSAT1 was able to replicate the probability of a pixel being rainy better than the conventional TAMSAT algorithm as well as the pixel level rainfall distribution conditioned to the pixel CCD. However, as TAMSAT1 parameters are calibrated using historical raingauge data (practical for regions where raingauge data is unavailable in real-time) it tends to introduce systematic bias in its rainfall estimates in anomalous rainfall years, resulting in an underestimation or overestimation of seasonal rainfall. This is important when using TAMSAT1 in up-stream applications that are sensitive to the magnitude of the rainfall total received. A case in crop modelling was demonstrated in Chapter 6 (see discussion below). Details of the formulation of the TAMSAT1 algorithm and its validation results for Gambian daily rainfall are in Chapter 4.

To assess the crop model sensitivity to uncertainties in TAMSAT1 inputs, quantification of the uncertainties of TAMSAT1 divisional rainfall estimates (obtain by aggregating pixel rainfall estimates) was necessary. Although the uncertainties in the pixel rainfall estimates were fully defined within the TAMSAT1 algorithm, obtaining the uncertainties of the rainfall estimates at a divisional level was not trivial due to the spatial coherence of the daily rainfall field. No analytical solution exists or even one that is sufficiently tractable for direct deduction of the distribution of the divisional TAMSAT1 estimates from pixel level information. A novel approach used in this thesis was to employ a Monte-Carlo method, SIMU, which stochastically generates rainfall with a two-step process. First, the intermittency of the daily rainfall field was simulated using Sequential Indicator Simulation (SIS) and second, the rainfall amount within the rainy area is simulated using Sequential Gaussian Simulation (SGS). Both SIS and SGS are designed to take account of the pixel CCD through TAMSAT1 CCD-rainfall relationship. In this way, SIMU can be considered a stochastic generalisation of the TAMSAT1 method of rainfall estimation. The SIMU algorithm was validated against gauge-estimated pixel daily rainfall for the Gambia. Results were encouraging; SIMU was able to reproduce the distribution of rainfall at pixel level as well as rainfall-fractional coverage relationship at larger spatial scale for the Gambian rainfall. The latter is significant since it a feature not explicitly defined within the simulation algorithm. The ability to generate such features lends further support to the usefulness of the SIMU algorithm for simulating realistic rainfall fields, at least for the Gambia. In addition, spatial coherence (in terms of variograms) of the daily rainfall amount is well reproduced, though not as well for intermittency of the rainfall field. The reason for the discrepancy in the simulated and observed spatial covariance of the intermittencies field was found to be due to over prediction of rainy areas by the TAMSAT1 algorithm when there are large numbers of pixels with zero CCD. When a correction was applied to take account of such a large spatial extent of zero CCD, the ability to reproduce the intermittency of the rainfall field improved. Details of the formulation of the SIMU algorithm and its validation results for Gambian rainfall are in Chapter 5.

GLAM as well as a simpler CSWB crop model were calibrated for Gambian groundnut and their relative skill in predicting divisional groundnut yields using both gauge-estimated and TAMSAT1 rainfall estimates as inputs have been compared in Chapter 6. The CSWB model was used here as a comparison for two reasons: first, it is commonly used in operational crop yield prediction and second, its algorithm is more linear than GLAM which serves as a good reference in understanding the impact of yield prediction due to non-linearity in GLAM processes when different types of rainfall input are used. Experiments with various simple rainfall distributions showed that for GLAM, model crop development is sensitive to the intraseasonal rainfall variability. Water deficits in the early and later crop stages can have unequal impact on the predicted yields. Further experiments showed that the CSWB is less sensitive than GLAM to rainfall distribution but more sensitive to total rainfall amount.

Gauge-rainfall driven GLAM appeared to perform better than gauge-rainfall driven CSWB in hindcasting the divisional groundnut yields, though the data available was insufficient for the difference to be statistically significant. However, GLAM clearly outperformed CSWB in predicting the divisional groundnut yields, when satellite-based rainfall estimates were used as input. TAMSAT1 driven GLAM reproduced the groundnut yields as well as gauge-rainfall driven GLAM for most years despite the difference between the satellite and gauge rainfall estimates. The main reason is GLAM's higher sensitivity to the intraseasonal rainfall distribution, which TAMSAT1 captures reasonably well, than to systematic bias in rainfall totals. TAMSAT1 driven CSWB shows a lower skill in predicting the divisional groundnut yield when compared to TAMSAT1 driven GLAM. This was due to the combination of the two factors: 1) CSWB yields are sensitive to the total seasonal rainfall, which is shown to be a consequence of the linearity of yield-rainfall responses in the crop model. 2) The TAMSAT1 algorithm is not able to capture the total seasonal rainfall amount accurately in anomalous rainfall years due to systematic bias in the TAMSAT1 calibration.

The above result demonstrates the potential of a RFE driven process-based crop model as a crop yield forecasting system. Such combination amalgamates the strength of both

components in the crop forecast system – the ability for RFE to capture intraseasonal rainfall variability, and the ability of the process-based model to fully utilise this information (intraseasonal rainfall variability) to simulate crop development and, hence crop yield – potentially offers a yield prediction as good as when contemporaneous raingauge observations are available to drive the crop model. As the coverage of RFE is large, such a crop yield forecast system is deployable in many data-sparse regions of Africa especially where agriculture is rainfed and rainfall is marginal.

The effects of random error in TAMSAT1 estimates was investigated using different realisations of the divisional rainfall conditioned to daily CCD data to drive both GLAM and CSWB to obtain an ensemble of yields. The results showed that simulated yields from both crop models are robust to the stochastic nature of the satellite rainfall estimates when compared with the uncertainty in capturing the planting date estimated from a planting date algorithm. More important is the effect of systematic bias in satellite-rainfall estimates to crop model yield prediction. For GLAM, the sensitivity to systematic bias varies with the stage of crop growth: For 2002 when rainfall is anomalously low, bias-correction of the TAMSAT1 rainfall estimates for a month early in the growing season removes most of the bias in predicted yield without much change in the overestimation of the seasonal total rainfall, emphasising the non-linear response of GLAM to rainfall. For the CSWB model, it is the overall bias in the seasonal rainfall that is important; for years when yield is water-limited, the bias in yield estimation is proportional to the bias in the total rainfall received within the growing season.

## 7.2 Issues and recommendations

While the results described above demonstrate the feasibility of using TAMSAT1 rainfall estimates or stochastically generated rainfall from SIMU as inputs to GLAM for crop yield forecasting, several issues remain to be addressed.

One issue that emerges from the investigations reported in this thesis is the systematic bias in TAMSAT1 rainfall estimation in anomalous rainfall years, and their impact on up-

stream applications, demonstrated by the biases in yield predictions by CSWB. When rainfall is anomalously high, the yield is not water-limited; CSWB can be relatively insensitive to underestimation by TAMSAT1 as long as the rainfall estimates is able to satisfy the crop water demand. However, for other applications, such underestimation may be critical: in prediction of river flows in catchments using hydrological models, underestimation of rainfall has been shown to underestimate runoffs (Kobold and Sugelj, 2005). It is therefore important that systematic biases in satellite rainfall estimates are to be minimised. For the TAMSAT1 algorithm, the method of reduction of systematic bias would depend on the available information. If real-time raingauge data is available, a merged gauge and satellite estimation of rainfall has been shown to perform better than either gauge-only or satellite-only estimation (Grimes *et al.*, 1999). Real time recalibration of satellite TIR rainfall estimates using recent raingauge observations or rainfall estimates from microwave soundings as reported in literature (example Todd *et al.*, 1999, 2001) are also viable alternatives that can be incorporated into the TAMSAT1-algorithm. The latter is of special interest for regions in Africa where rain gauge networks are unreliable or non-existent, in view of the proposed implementation of the Global Precipitation Monitoring project in the near future enabling microwave rainfall estimates at 3-hour intervals globally (Smith, 2003). However, more studies need to be conducted over different regions of the continent to validate the accuracy of microwave rainfall measurements in light of the problems of contamination of the microwave measurements from land surface microwave emissions (Morland *et al.*, 2001). Seasonal rainfall prediction in Africa using sea surface temperature (SST) has been a subject of intensive research (example Opokuankomah and Cordery, 1994; Camberlin and Diop, 1999; Black *et al.*, 2003; Paeth and Friederichs, 2004). Hence it is not unreasonable to suggest another possible approach to the bias problem in the TAMSAT1 estimates may be through incorporating SST or any other potential indicators of seasonal rainfall into the TAMSAT1-algorithm for regions where interannual variability of rainfall shows strong correlation with these variables. Work on improving satellite RFE by incorporating NWP model fields have been reported by Grimes and Diop (2003), Grimes *et al.* (2003) and Coppola *et al.* (2006).

One implicit assumption in the SIMU algorithm is the independence of intermittency and rainfall variability within the rainy regions. For the Gambian daily rainfall used in the case study, the independence of the two properties has not been verified. Such dependence may manifest itself in the form of dependency between the mean areal rainfall within the rainy region and the area of rainy region, as well as systematic trends in rainfall amounts, in relation to the shape of the rainy region (Barancourt *et al.*, 1992). In cases where such dependence of rainfall and intermittency is significant, a possible modification to the SIMU algorithm is to replace the Indicator Kriging in the SIS by Indicator Cokriging or Probability Kriging. Likewise, the Simple Kriging algorithm within the SGS when sampling rainfall amount for rainy pixels could be replaced with Cokriging, taking into account the indicator realisation. Details of the Cokriging and Probability Kriging algorithms can be found in Goovaerts (1997). However, implementing such changes would require the inference of various cross-variograms, as well as an increase in the computational overhead due to the larger systems of equation in the new kriging algorithms. A more appealing alternative is to adopt a single SGS algorithm, instead of a double SIS and SGS, which deals with the mixed rainfall distribution directly. Transformations from Gaussian to mixed distribution within the context of stochastic rainfall simulation were used by Bell (1987) and Lanza (2000). In this formulation, the dependency of intermittency and rainfall amount are encapsulated within the two-point covariance of the normal transformed rainfall (including zero) that is used in SGS algorithm.

Finally, the positive results of a combined RFE-GLAM system in providing hindcast of Gambian divisional groundnut yields presented in this thesis provides a motivation to evaluate the applicability of the RFE-GLAM system for operational crop yield forecasts on a wider regional context, as well as for other crops important to the food security on the continent. One additional issue that is not explored in the present work is the capability of such crop forecasting system to provide crop yield prediction in advance of the harvest useful for forward planning by governmental agencies and NGOs (Haile, 2005). One plausible configuration of a comprehensive crop forecasting system would include a initial ensemble crop yield prediction from GLAM driven by rainfall inputs

synthesised from stochastic weather generator, conditioned to seasonal climate predictions (Hansen and Indeje, 2004; Hansen and Ines, 2005) or climatology (Bannayan *et al.*, 2003), before the start of the crop growing season. As the growing season progresses, available satellite-based RFE either from the TAMSAT1 or the SIMU algorithm can be incorporated into the rainfall inputs to reduce the uncertainty in the final crop yield prediction.

As a demonstration of such method for TAMSAT1-GLAM, Figure 7.1 shows the Gambian divisional groundnut yield ‘forecast’ for one year at various time within the current growing season. Following Thornton *et al.* (1997), ‘unobserved’ rainfall is supplemented using historical raingauge data when RFE was not available. Distribution of model end-of-season yield at different times of the forecast shows uncertainties in the predicted yield decreasing, particularly 30 days after sowing (DAS). At 90 DAS, about a month before maturity, the yield uncertainty is sufficiently small (IQR = 69 kg/ha). As there is significant interannual variability in rainfall, using all available historical gauge rainfall to supplement the unobserved rainfall may not be ideal. Better predictability at earlier forecast date may be achieved by using for example, synthetic rainfall generated conditioned to the seasonal rainfall predictions (Hansen and Indeje, 2004).

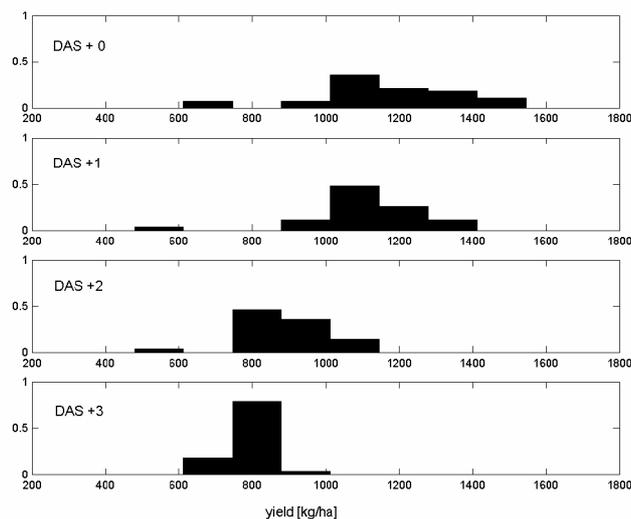


Figure 7.1 Histogram of groundnut yield predicted by TAMSAT1-GLAM at sowing (DAS + 0) and 1, 2 and 3 months after sowing (DAS + 1, DAS + 2, DAS + 3).

## Annex A Determination of various variables in the FAO Penman-Montieth Equation

The computations of  $VPD$ ,  $R_n$  and  $\Delta$  for Equation 2.1 below are adopted from Chapter 3 of Allen *et al.* (1998). In all computation,  $T_{max}$  and  $T_{min}$  are assumed to be in  $^{\circ}C$ ,  $RH_{max}$  and  $RH_{min}$  in % and  $SS$  in hours.

### Saturated vapor deficit and pressure gradient

VDP is defined as the difference between the saturated vapor pressure ( $e_s$ ) and actual vapor pressure ( $e_a$ ). Saturated vapor pressure is determined as the average of the saturated vapor pressure at  $T_{max}$  and  $T_{min}$ :

$$e_s = \frac{e^0(T_{max}) + e^0(T_{min})}{2} \quad A.1$$

where  $e^0(T)$  [kPa] is the saturated vapor pressure at  $T$  [ $^{\circ}C$ ] given by

$$e^0(T) = 0.6108e^{\frac{17.27T}{T+237.3}} \quad A.2$$

Actual pressure  $e_a$  [kPa] is computed using the below equation:

$$e_a = \frac{e^0(T_{min})\frac{RH_{max}}{100} + e^0(T_{max})\frac{RH_{min}}{100}}{2} \quad A.3$$

From Equation A.2, the gradient of the vapor pressure [kPa  $^{\circ}C^{-1}$ ] at a temperature  $T$  is

$$\Delta = \frac{\partial e^0(T)}{\partial T} = \frac{4098e^0(T)}{(T + 237.3)^2} \quad A.4$$

For the purpose of Equation 2.1,  $T = (T_{max} + T_{min})/2$  is used in Equation A.4.

### Net solar radiation

First, the incoming solar radiation  $R_s$  is determined from the sunshine hours:

$$R_s = \left(0.25 + 0.50\frac{SS}{N}\right)R_a \quad A.5$$

where  $N$  is the daily maximum hours of sunshine at the locality and  $R_a$  is the daily extraterrestrial radiation.  $N$  and  $R_a$  given by:

$$R_a = \frac{1440}{\pi} G_{sc} d_r (\omega_s \sin(\varphi) \sin(\delta) + \cos(\varphi) \cos(\delta) \sin(\omega_s))$$

$$N = \frac{24}{\pi} \omega_s$$
A.6

where  $\varphi$  is the latitude of the locality,  $G_{sc}$  [ $\text{MJ m}^{-2}\text{min}^{-1}$ ] is the solar constant,  $d_r$  is the inverse relative distance Earth-Sun and  $\delta$  is the solar declination and  $\omega_s$  is the sunset hour angle.  $d_r$ ,  $\delta$  and  $\omega_s$  is determined by the following:

$$d_r = 1 + 0.033 \cos\left(\frac{2\pi}{365} J\right)$$

$$\delta = 0.409 \sin\left(\frac{2\pi}{365} J - 1.39\right)$$

$$\omega_s = \cos^{-1}(-\tan(\varphi) \tan(\delta))$$
A.7

where  $J$  is the number of days into the year.

From  $R_s$ , the net incoming shortwave radiation,  $R_{ns}$ , and net outgoing longwave radiation,  $R_{nl}$ , can be obtained from the following:

$$R_{ns} = (1 - \alpha) R_s$$
A.8

where  $\alpha$  is the albedo of the surface (0.23 in the case for the reference crop) and

$$R_{nl} = \sigma \left[ \frac{(T_{\max} + 273.16)^4 + (T_{\min} + 273.16)^4}{2} \right] \left[ (0.34 - 0.14 \sqrt{e_a}) \left[ 1.35 \frac{R_s}{R_{so}} - 0.35 \right] \right]$$
A.9

where  $\sigma$  [ $\text{MJ K}^{-4}\text{m}^{-2}\text{day}^{-1}$ ] is the Stefan-Boltzmann Constant,  $e_a$  [kPa] is the actual saturation vapor given by Equation A.3 and  $R_{so}$  is the clear sky radiation, which can be computed from Equation A.5 by setting  $SS$  to  $N$ .

Finally, the daily net solar radiation,  $R_n$ , is the difference between the incoming net shortwave radiation and the net outgoing longwave radiation:

$$R_n = R_{ns} - R_{nl}$$
A.10

Note that all units for the daily radiation fluxes in the above equations are in  $\text{MJ m}^{-2}\text{day}^{-1}$ .

## Annex B Approximation of the mean square prediction error of the pixel rainfall estimation with fractional coverage using Double Kriging

Below gives the argument for the approximation of mean square prediction error of Equation 3.18.

Let  $Z_V$  be the mean rainfall over a pixel defined as

$$Z_V = \frac{1}{\|V\|} \int_{u \in V} Z(u) du \quad \text{B.1}$$

where,  $\|V\|$  is the area of the pixel. An unbiased estimate for  $Z_V$  is

$$Z_V^* = \frac{1}{\|V\|} \int_{u \in V} Z(u)^* du \quad \text{B.2}$$

where  $Z(u)^*$  is an unbiased estimate for  $Z(u)$ . Then it is easily shown that

$$(Z_V - Z_V^*)^2 = \frac{1}{\|V\|^2} \iint_{u, u' \in V} (Z(u) - Z(u)^*)(Z(u') - Z(u')^*) du du'$$

For the case where  $Z(u) = I(u) \times F(u)$  (Equation 3.6), with no estimation error in  $I(u)$  ( $Z^*(u) = I(u) \times F(u)^*$ ) the above becomes

$$(Z_V - Z_V^*)^2 = \frac{1}{\|V\|^2} \iint_{u, u' \in V} I(u) I(u') (F(u) - F(u)^*)(F(u') - F(u')^*) du du'$$

For fractional coverage, the preceding equation becomes

$$(Z_V - Z_V^*)^2 = \frac{1}{\|V\|^2} \iint_{u, u' \in V1} (F(u) - F(u)^*)(F(u') - F(u')^*) du du'$$

where  $V1$  denotes the region within the pixel that is rainy, where  $\|V1\| = \alpha \|V\|$ ,  $0 < \alpha \leq 1$ , the estimation variance using  $Z_V^*$  is

$$E[(Z_V - Z_V^*)^2] = \alpha^2 \frac{1}{\|V1\|^2} \iint_{u, u' \in V1} E[(F(u) - F(u)^*)(F(u') - F(u')^*)] du du' \quad \text{B.3}$$

Since the last term of RHS of Equation B.3 is the estimation variance of mean positive rainfall over  $V1$ , we have

$$E[(Z_V - Z_V^*)^2] = \alpha^2 E[(F_{V1} - F_{V1}^*)^2] \quad \text{B.4}$$

where  $F_{V1}$  and  $F_{V1}^*$  is the pixel average of  $F(\mathbf{u})$  and its corresponding estimator defined in similar fashion as Equation B.1 and B.2 respectively. If  $F_{V1}$  is estimated using block kriging (Goovaerts, 1997, Section 5.5), Equation B.4 becomes

$$E[(Z_V - Z_V^*)^2] = \alpha^2 \sigma_{BOK}^2(V1)$$

where  $\sigma_{BOK}^2(V1)$  denotes the block kriging variance, with the parenthesis included to emphasize  $\sigma_{BOK}^2$  dependence on the geometry of  $V1$ . If the area of rain coverage within the pixel decrease from full coverage towards zero, it is expected that  $\sigma_{BOK}^2$  would varies from  $\sigma_{BOK}^2(V)$  to the point block kriging estimation variance  $\sigma_{OK}^2(u)$ , for some  $u \in V$ . It is reasonable to expressed this variation as a ratio  $W(\alpha)$  with respect to  $\sigma_{BOK}^2(V)$ :

$$E[(Z_V - Z_V^*)^2] = \alpha^2 W(\alpha) \sigma_{BOK}^2(V) \quad \text{B.5}$$

It follows than that  $W(1) = \sigma_{BOK}^2(V)$  and  $\lim_{\alpha \rightarrow 0} W(\alpha) = \sigma_{OK}^2(u)$  for some  $u \in V$ . In order to get a sense of functional form of  $W(\alpha)$ , it is noted that the block kriging variance for  $V1$  is given by

$$\sigma_{BOK}^2(V1) = \frac{1}{\|V1\|^2} \iint_{\mathbf{u}, \mathbf{u}' \in V1} Cov(\mathbf{u} - \mathbf{u}') d\mathbf{u} d\mathbf{u}' - \frac{1}{\|V1\|} \int_{\mathbf{u} \in V1} \sum_i \lambda_i(V1) Cov(\mathbf{u}_i - \mathbf{u}) d\mathbf{u} - \mu(V1) \quad \text{B.6}$$

where  $\lambda_i$  is the kriging weights with respect to the rainfall observed at location  $\{\mathbf{u}_i\}$  and  $\mu$  is the Lagrange multiplier (Goovaerts, 1997, Section 5.5). Since the kriging variance depends on the location of  $V1$  with respect to  $\{\mathbf{u}_i\}$  and the kriging weights and the Lagrange multiplier are depends on the geometry of  $V1$ , it is impossible to obtain the general solution of  $W(\alpha)$ . Here it is argued that that the functional form of  $\sigma_{BOK}^2(V1)$  follows approximately that of the average of the two point-covariances within  $V1$  (first term of RHS of Equation B.6.). If  $V1$  is a square, and if  $F(\mathbf{u})$  is homogeneous and isotropic, it can be shown (Rodríguez-Itrube and Mejía, 1974)

$$Cov(V1, V1) = \frac{1}{\|V1\|^2} \iint_{\mathbf{u}, \mathbf{u}' \in V1} Cov(\mathbf{u} - \mathbf{u}') d\mathbf{u} d\mathbf{u}' = \int_0^d Cov(r) G(r) dr \quad \text{B.7}$$

where  $r = |u-u'|$  and  $G(r)$  is the probability density of  $r$  between two randomly selected points in  $V1$ , and  $d$  is the diagonal distance of the square area. Further details of the functional form of  $G(r)$  can be found in Bras and Rodriguez-Itrube (1974). Using an exponential form for the two-point covariance with typical values of nugget and scale length of 0.5 and 15km respectively,  $\alpha^2 Cov(V1, V1)$  as a function of  $\alpha$  is shown in Figure B.1, justifying that it is not unreasonable to take  $\alpha^2 W(\alpha) \sim \alpha$  as an approximation. Then Equation B.5 becomes

$$E[(Z_V - Z_V^*)^2] \approx \alpha \sigma_{BOK}^2(V)$$

and finally,

$$E[(Z_V - Z_V^*)^2] \approx I_V \sigma_{BOK}^2(V) \quad B.6$$

where  $I_V$  is the fractional coverage as

$$I_V = \frac{1}{\|V\|} \int_{I \in V} I(u) du$$

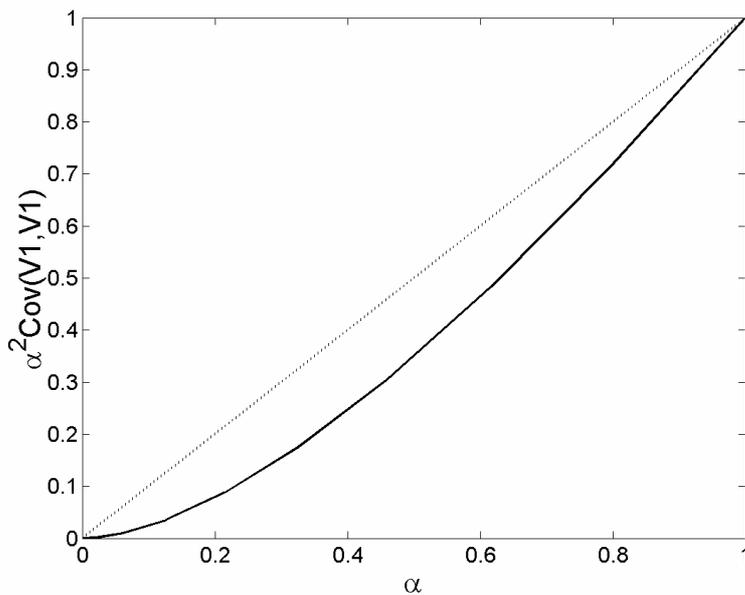


Figure B.1 Functional dependence of  $\alpha^2 Cov(V1, V1)$  with  $\alpha$ . Dotted line is 1-1 line.

## Annex C Parameters for GLAM for divisional Gambian groundnut

Listed are the parameters values used in GLAM for groundnut yield simulation for the Gambia (Chapter 6). “Symbol” refers to the nomenclature used in Challinor *et al.* (2004). Corresponding nomenclature used in this thesis, if different, is in square bracket when applicable.

	Symbol	Values
<b>Growth and development</b>		
rate of leaf growth	$\left(\frac{\partial L}{\partial t}\right)_{\max}$	0.05 per day
critical soil water stress value	$S_{\text{cr}}$	0.7
root growth	$\frac{\partial l_v(z=0)}{\partial L}$	1
extraction front velocity	$V_{\text{EF}}$	1 cm per day
root length density	$l_v$	0.3
normalised transpiration efficiency	$E_T [\hat{E}]$	1.4 Pa
maximum transpiration efficiency	$\hat{E}_{\text{TN,max}}$	3 g kg <sup>-1</sup>
rate of change of harvest index	$\frac{\partial H_I}{\partial t}$	0.007 per day
time of emergences after sowing	$t_{\text{em}}$	8 days
base (cardinal) temperature	$T_b$	10 K
optimal (cardinal) temperature	$T_o$	32 K
maximum (cardinal) temperature	$T_m$	50 K
thermal time from sowing to flowering	$t_{\text{Tto}}$	505 K day
thermal time from flowering to pod filling	$t_{\text{TT1}}$	430 K day
thermal time from pod filling to LAI maximum	$t_{\text{TT2}}$	340 K day
thermal time from LAI maximum to maturity	$t_{\text{TT3}}$	845 K day
<b>Evaporation and transpiration</b>		
Critical threshold of LAI	$L_{\text{cr}}$	0.7
Maximum possible potential transpiration	$T_{\text{Tmax}}$	0.3

	Symbol	Values
pre-correction factor in calculating Priestly-Taylor coefficient	$\alpha_0$	1.26
reference VPD in calculating Priestly-Taylor coefficient	$V_{\text{ref}}$	1 kPa
mean albedo of surface	$A$	0.2
extinction coefficient	$k$	0.5
rainfall amount threshold	$t_R$	0.1 cm
uptake diffusion coefficient	$k_{\text{DIF}}$	0.06 cm <sup>2</sup> per day
<b>Soil submodel and miscellaneous</b>		
depth of soil profile	$z_{\text{max}}$	210 cm
number of soil layer represented	$N_{\text{SL}}$	25
depth of soil from which evaporation occurs	$z_{\text{ed}}$	16.8 cm
yield gap parameter	$C_{\text{YG}}$	0.1-1.0
drained lower limit	$\theta_{\text{dl}}$	0.04-0.14 [0.09]
drained upper limit	$\theta_{\text{dul}}$	0.14-0.24
saturation limit	$\theta_{\text{sat}}$	0.3-0.5

# References

Adler, R.F., (1993). Estimation of monthly rainfall over Japan and surrounding waters from a combination of low-orbit microwave and geosynchronous IR data. *J. Appl. Meteor.*, **32**, 335-356.

Adler, R.F. and Negri, A.J., (1988). A satellite infrared technique to estimate tropical convective and stratiform rainfall. *J. Appl. Meteor.*, **27**, 30-51.

Adler, R.F., Huffman G.J., Bolvin D.T., Curtis S., and Nelkin E.J., (2000): Tropical rainfall distributions determined using TRMM combined with other satellite and gauge information. *J. Appl. Meteor.*, **39**, 2007–2023.

Affholder, F., (1997). Empirically modelling the interaction between intensification and climatic risk in semiarid regions. *Field Crops Res.*, **52**, 79–93.

Ali, A., Amani, A., Diedhiou, A. and Lebel, T., (2005). Rainfall estimation in the Sahel. Part II: Evaluation of rain gauge networks in the CILSS countries and objective intercomparison of rainfall products. *J. Appl. Meteor.*, **44**, 1707-1722.

Allen, R.G, Pereira, L.S., Raes, D. and Smith, M., (1998). Crop evapotranspiration - Guidelines for computing crop water requirements. FAO Irrigation and drainage paper 56, FAO, Rome.

Andersen, J., Dybkjaer, G., Jensen, K.H., Refsgaard, J.C. and Rasmussen, K., (2002). Use of remotely sensed precipitation and leaf area index in a distributed hydrological model. *J. Hydrol.*, **264**, 34-50.

Arkin, P.A. and Meisner, B.N., (1987). The relationship between large-scale convective rainfall and cold cloud cover over the western hemisphere during 1982-84. *Mon. Weather Rev.*, **115**, 51-74.

Astin, I., (1997). A survey of studies into errors in large scale space-time averages of rainfall, cloud cover, sea surface processes and the earth's radiation budget as derived from low orbit satellite instruments because of their incomplete temporal and spatial coverage. *Surv. Geophys.*, **18**, 385-403.

Azimi-Zonooz, A., Krajewski, W.F., Bowels, D.S. and Seo D.J., (1989). Spatial rainfall estimation by linear and non-linear co-kriging of radar-rainfall and raingage data. *Stochastic Hydrol. Hydraul.*, **3**, 51-67.

Baez-Gonzalez, A.D., Chen, P., Tiscareno-Lopez, M. and Srinivasan, R., (2002). Using satellite and field data with crop growth modelling to monitor and estimate corn yield in Mexico. *Crop Sci.*, **42**, 1943–1949.

Bannayan, M., Crout, N.M.J. and Hoogenboom, G., (2003). Application of the CERES-Wheat model for within-season prediction of winter wheat yield in the United Kingdom. *Agron. Jour.*, **95**,114-125.

Barancourt, C., Creutin, J.D. and Roivoirard, J., (1992). A method for delineating and estimating rainfall fields. *Water Resour. Res.*, **28**, 1133-1144.

Barrett, E.C. and Martin, D.W., (1981). The use of satellite data in rainfall monitoring. Academy Press, London, 340 pp.

Barrett, E.C., Dodge, J., Goodman, M., Janowiak, J., Kidd, C., and Smith, E.A., (1994). The First WetNet Precipitation Inter-comparison Project. *Remote Sensing Reviews*, **11**, 49–60.

Bell, T.L., (1987). A space-time stochastic-model of rainfall for satellite remote-sensing studies. *J. Geophys. Res-Atmos.*, **92**, 9631-9643.

- Bell, T.L., Abdullah, A., Martin, R.L. and North, G.R., (1990). Sampling errors for satellite-derived tropical rainfall – Monte-Carlo study using a space-time stochastic-model. *J. Geophys. Res.-Atmos.*, **95**, 2195-2205.
- Bellerby, T.J., (2004). A feature-based approach to satellite precipitation monitoring using geostationary IR imagery. *J. Appl. Meteor.*, **5**, 910-921.
- Bellerby, T.J. and Barrett, E.C., (1993). Progressive refinement : A strategy for the calibration by collateral data of short-period satellite rainfall estimates. *J. Appl. Meteor.*, **32**, 1365-1378.
- Bellerby, T.J., Todd M., Kniveton D., and Kidd, C., (2000). Rainfall estimation from a combination of TRMM precipitation radar and GOES multispectral satellite imagery through the use of an artificial neural network. *J. Appl. Meteor.*, **39**, 2115–2128.
- Bellerby, T.J. and Sun, J., (2005). Probabilistic and ensemble representations of the uncertainty in an IR/Microwave satellite precipitation product. *J. Appl. Meteor.*, **6**, 1032-1044.
- Black, E., Slingo, J. and Sperber, K.R., (2003). An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean SST. *Mon. Wea. Rev.*, **131**, 74–94.
- Bojang, L., (2001). Country Report – The Gambia. Forestry Outlook Study for Africa (FOSA) Working Paper - FOSA/WP/11.
- Brady, P.J., (1978). Optimal sampling and analysis using two variables and modeled cross-covariance functions, *J. Appl. Meteor.*, **17**, 12-21.
- Bras, R.L. and Rodríguez-Iturbe, I., (1976). Rainfall generation – Nonstationary time-varying multidimensional model. *Water Resour. Res.*, **12**, 450-456.

Camberlin, P., (1995). June–September rainfall in North-Eastern Africa and atmospheric signals over the tropics: A zonal perspective. *Int. J. Climatol.*, **15**, 773–783.

Camberlin, P. and Diop, M., (1999). Inter-relationships between groundnut yield in Senegal, interannual rainfall variability and sea-surface temperatures. *Theor. Appl. Climatol.*, **63**, 163-181.

Camberlin, P., Janicot, S. and Pocard, I., (2001). Seasonality and atmospheric dynamics of the teleconnection between African rainfall and tropical sea-surface temperature: Atlantic vs. ENSO. *Int. J. Climatol.*, **21**, 973-1005.

Campling, P., Gobin, A. and Feyen, J., (2001). Temporal and spatial rainfall analysis across a humid tropical catchment. *Hydrol. Process.*, **15**, 359-375.

Camberlin, P. and Diop, M., (1999). Inter-relationships between groundnut yield in Senegal, interannual rainfall variability and sea-surface temperatures. *Theor. Appl. Climatol.*, **63**, 163-181.

Carroll, R.J. and Ruppert, D., (1988). Transformation and weighting regression. London, Chapman and Hall.

Challinor, A.J., Slingo, J.M., Wheeler, T.R., Crauford, P.Q. and Grimes, D.I.F., (2003). Towards a combined seasonal weather and crop productivity forecasting system: determination of the working spatial scale. *J. App. Meteor.*, **42**, 175-192.

Challinor, A.J., Wheeler, T.R., Osborne, T.M. and Slingo, J.M. (2004). Design and optimisation of a large-area process-based model for annual crops. *Agric. Forest Meteorol.*, **124**, 99-120. [doi:10.1016/j.agrformet.2004.01.002].

- Challinor, A.J., Wheeler, T.R., Slingo, J.M., Crauford, P.Q. and Grimes, D.I.F., (2005a). Simulation of crop yields using ERA-40: Limits to skill and nonstationarity in weather-yield relationships. *J. Appl. Meteor.*, **44**, 516-531.
- Challinor, A.J., Slingo, J.M., Wheeler, T.R. and Doblas-Reyes, F.J., (2005b). Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles. *Tellus A*, **57**, 498-512.
- Challinor, A.J., T.R. Wheeler, P.Q. Crauford and J.M. Slingo, (2005c). Simulation of the impact of high temperature stress on annual crop yields. *Agric. Forest Meteorol.*, **135**, 180-189.
- Cheke, R.A. and Holt, J., (1993). Complex Dynamics of desert locust plagues. *Ecol. Entomol.*, **18**, 109-115.
- Cheng, K.S., Yeh, H.C. and Liou, C.Y., (2000). Comparative study of drought prediction techniques for reservoir operation. *J. Am. Water Resour. As.*, **36**, 511-521.
- Chieteka Z.A., Cole, D.L., Freire, M.J. *et al.* (1991), Groundnut Research in the SADCC Region, 1980-1990. In *Groundnut – a global perspective: a proceedings of an international workshop, 25 -29 Nov 1991*. ICRISAT Center, India (Nigam, S.N., Ed) 548 pp.
- Chiu, L. S., North, G.R., Short, D.A. and McConnell, A., (1990). Rain estimation from satellites: Effects of finite field of view. *J. Geophys. Res.*, **95**, 2177–2185.
- Chiu, L.S. and Kendem, B., (1990). Estimating the exceedence probability of rain rate by logistic regression. *J. Geophys. Res.*, **95**, 2217-2227.
- Chua, S.H. and Bras, R.L., (1982). Optimal Estimators of mean areal precipitation in regions of orographic influence. *J. Hydrol.*, **57**, 23-48.

- Clover, J., (2003). Food security in Sub-Saharan Africa, *African Security Review*, **12**, 2003, 17-27.
- Cook, B.G. and Crosthwaite, I.C., (1994). Utilization of *Arachis* species as forage. In “*The Groundnut Crop: A Scientific Basis for Improvement*” (Smart, J., Ed.), 624 – 657.
- Cook, R.D. and Weisberg, S., (1982) *Residuals and Influence in Regression*. New York, Chapman and Hall.
- Commission for Africa (2005). Our common interest. Report of the Commission for Africa. London, UK: Commission for Africa. Published online 11 March 2005. (<http://213.225.140.43/english/report/introduction.html>)
- Coppola, E., Grimes, D.I.F., Verdecchia, M. and Visconti, G., (2006). Validation of improved TAMANN neural network for operational satellite-derived rainfall estimation in Africa. *Submitted to J. App. Met.*
- Cressie, N., (1991). *Statistics for spatial data*. Wiley, New York.
- Creutin, J.D. and Obled, C., (1982). Objective analyses and mapping techniques for rainfall fields – An objective comparison. *Water Resour. Res.*, **18**, 413-431.
- Creutin, J.D., Lacomba, P. and Obled, C., (1987). Spatial relationship between cloud-cover and rainfall fields: A statistical approach combining satellite and ground data. *IAHS Publ.*, **160**, 81-89.
- Creutin, J.D., Delrieu, G., Lebel, T., (1988). Rain measurement by raingage radar combination: A geostatistical approach. *J. Atmos. Oceanic Technol.*, **5**, 102-115.

- Davis, S.A., Leirs, H., Pech R, Zhang, Z.B. and Stenseth, N.C., (2004). On the economic benefit of predicting rodent outbreaks in agricultural systems. *Crop Prot.*, **23**, 305-314.
- Dercon, S., (2002). Income risk, coping strategies, and safety nets. *World Bank Res. Obs.*, **17**, 141–166. doi:10.1093/wbro/17.2.141.
- Dercon, S., (2004). Growth and shocks: evidence from rural Ethiopia. *J. Dev. Econ.*, **74**, 309–329. doi:10.1016/j.jdeveco.2004.01.001.
- Dewhurst, C.F., Page, W.W. and Rose, D.J.W., (2001). The relationship between outbreaks, rainfall and low density populations of the African armyworm, *Spodoptera exempta*, in Kenya. *Entomol. Exp. Appl.*, **98**, 285-294.
- Deutsch, C.V. and Journel, A.G., (1992). *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York.
- Dirks, K.N., Hay, J.E., Stow, C.D. and Harris, D. (1998). High-resolution studies of rainfall on Norfolk Island Part II: Interpolation of rainfall data. *J. Hydrol.*, **208**, 187-193.
- Doneaud, A.A., Smith, P.L., Dennis A.S. and Sengupta, S., (1981). A simple method for estimating convective rain volume over and area. *Water Resour. Res.*, **17**, 1676-1682.
- Doorenbos, J. and Kassam, A.H., (1979). *Yield response to water*. Irrigation and Drainage Paper 33, FAO, Rome.
- Duchon, C.E., Renkevans, T.M. and Crosson, W.L., (1995). Estimation of daily area-average rainfall during the CAPE experiment in Central Florida. *J. Appl. Meteor.*, **34**, 2704-2714.

- Dwivedi, S.L., Crouch, J.H., Nigam, S.N., Ferguson, M.E., and Paterson, A.H., (2003). Molecular breeding of groundnut for enhanced productivity and food security in the semi-arid tropics: Opportunities and challenges *Adv. Agron.*, **80**, 153-221.
- Faivre, R., Leenhardt, D., Voltz, M., Benoit, M., Papy, F., Dedieu, G. and Wallach, D., (2004) Spatialising crop models. *Agronomie*, **24**, 205-217.
- FAO, (1995). Digital soil map of the world. Land and Water Development Division, FAO, Rome.
- FAO, (1978). Report on the agro-ecological zones project. Vol.1: Methodology and results for Africa. FAO World Soil Resources Report 48/1. FAO, Rome.
- FAO, (2003). The state of food insecurity in the world 2003. *Monitoring progress towards the World Food Summit and Millennium Development Goals*. FAO, Rome, Italy.
- FAO Statistical Database, (2006). <http://faostat.fao.org/faostat/>. Last updated: 24<sup>th</sup> April 2006.
- Farmer, G., (1988). Seasonal forecasting of the Kenya Coast short rains 1901–84. *Int. J. Climatol.*, **8**, 489–497.
- Ferris, R., Ellis, R.H., Wheeler T.R. and Hadley, P., (1998) Effect of high temperature stress at anthesis on grain yield and biomass of field-grown crops of wheat. *Ann. Bot.* **82**, 631–639.
- Frère, M. and Popov, G.F., (1979). Agrometeorological crop monitoring and forecasting. FAO Plant Production and Protection Paper No. 17, FAO, Rome, Italy.
- Frère, M. and Popov, G.F., (1986). Early agrometeorological crop yield assessment. FAO Plant Production and Protection Paper No. 73, FAO, Rome, Italy.

Fyhri, T., (1998). The Gambia: The Complexity of Modernising the Agricultural Sector in Africa. Unpublished thesis in Geography of Resources, Department of Geography, University of Oslo (Norway).

Fontaine, B. and Janicot, S., (1996). Sea surface temperature fields associated with West African rainfall anomaly types. *J. Climate*, **9**, 2935–2940.

Fiorucci, P., La Barbera, P, Lanza, L.G. and Minciardi, R., (2001). A geostatistical approach to multisensor rain field reconstruction and downscaling. *Hydrol. Earth Syst. Sc.*, **5**, 201-213.

Funk, C., Michaelsen, J., Verdin, J., Artan, G., Husak, G., Senay, G., Gadain, H. and Magadzire, T., (2003). The collaborative historical African rainfall model: Description and evaluation. *Int. J. Climatol.*, **23**, 47-66.

Funk, C. and Michaelsen, J.A., (2004). Simplified diagnostic model of orographic rainfall for enhancing satellite-based rainfall estimates in data-poor regions. *J. Appl. Meteor.*, **43**, 1366-1378.

Gebremichael, M. and Krajewski, W., (2004). Characterization of temporal sampling error in space-time-averaged rainfall estimates from satellites. *J. Geophys. Res.-Atmos.*, **109**, D11110, doi:10.1029/2004JD004509.

Graef, F. and Haigis, J., (2001). Spatial and temporal rainfall variability in the Sahel and its effects on farmers' management strategies. *J. Arid Environ.*, **48**, 221-231.

Granjon, L., Cosson, J.F., Quesseveur, E. and Sicard, B. (2005). Population dynamics of the multimammate rat *Mastomys huberti* in an annually flooded agricultural region of central Mali. *J. Mammal.*, **86**, 997-1008.

Griffith, C.G. and Woodley, W.L., (1973). On the variation with height of top brightness of precipitating convective clouds, *J. Appl. Meteor.*, **12**, 1086-1089.

Griffith, C.G., Woodley, W.L., Grube, P.G., Martin, D.W., Stout, J. and Sikdar, D.N., (1978). Rain estimates from geosynchronous satellite imagery: Visible and Infrared studies. *Mon. Wea. Rev.*, **106**, 1153-1171.

Grimes, D.I.F., Pardo-Igúzquiza E., and Bonifacio, R., (1999). Optimal areal rainfall estimation using raingauges and satellite data. *J. Hydrol.*, **222**, 93-108.

Grimes, D.I.F. and Diop, M., (2003). Satellite-based rainfall estimation for river flow forecasting in Africa. I: Rainfall estimates and hydrological forecasts. *Hydrolog Sci. J.*, **48**, 567-584.

Grimes, D.I.F., Coppola, E., Verdecchia, M. and Visconti, G., (2003). A neural network approach to real-time rainfall estimation for Africa using satellite data. *J. Hydrometeorol.*, **4**, 1119-1133.

Gommes, R., (1993). FAOINDEX, Version 2.1m, Agrometeorology Group, FAO, Rome, Italy.

Goovaerts, P., (1994). Comparative performance of indicator kriging algorithms for modeling conditional probability distribution functions. *Math. Geol.*, **26**, 389-411.

Goovaerts, P., (1997). Geostatistics for Natural Resources Evaluation. Oxford Univ. Press, New York.

Goovaerts, P., (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrometeorol.*, **228**, 113-129.

Grover-Kopec, E., Kawano, M., Klaver, R.W., Blumenthal, B., Ceccato, P. and Connor, S.J. (2005). An online operational rainfall-monitoring resource for epidemic malaria early warning systems in Africa. *Malaria J.*, **4**, Art. No. 6.

Groten, S.M.E., (1993). NDVI-Crop monitoring and early yield assessment of Burkina Faso. *Int. J. Remote Sens.* **14**, 1495–1515.

Guetter, A.K., Georgakakos, K.P., Tsonis, A.A., (1996). Hydrologic applications of satellite data .2. Flow simulation and soil water estimates *J. Geophys. Res.-Atmos.*, **101**, 26527-26538.

Gupta, V.K. and Waymire, E.C., (1979). Stochastic Kinematic study of subsynoptic space-time rainfall. *Water Resour. Res.*, **15**, 637-644.

Gutman, G.G., (1991). Vegetation indices from AVHRR: an update and future prospects. *Remote Sens. Environ.*, **35**,121–136.

Gutman, G., (1999). On the long0tem global data of land reflectances and vegetation indices derived from the advanced very high resolution radiometer. *J. Geophys. Res.*,**194**, 6241-6255.

Haile, M., (2005). Weather patterns, food security and humanitarian response in sub-Saharan Africa. *Philos. T. Roy. Soc. B.*, **360** (1463), 2169-2182.

Hansen, J.W. and Jones, J.W., (2000). Scaling-up crop models for climatic variability applications. *Agric. Syst.*, **65**, 43–72.

Hansen, J.W. and Indeje, M., (2004). Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid Kenya. *Agric. Forest Meteorol.*, **125**, 143-157.

Hansen, J. W., Potgieter, A. and Tippett, M.K., (2004). Using a general circulation model to forecast regional wheat yields in northeast Australia. *Agric. Forest Meteorol.*, **127**, 77–92. doi:10.1016/j.agrformet.2004.07.005.

Hansen, J.W., (2005). Integrating seasonal climate prediction and agricultural models for insights into agricultural practice. *Philos. T. Roy. Soc. B*, **360**, 2037-2047.

Hansen, J.W. and Ines, A.V.M., (2005). Stochastic disaggregation of monthly rainfall data for crop simulation studies *Agric. Forest Meteorol.*, **131**, 233-246.

Hastenrath, S., (1984). Interannual variability and annual cycle: Mechanisms of circulation and climate in the Tropical Atlantic sector. *Mon. Wea. Rev.*, **112**, 1097–1107.

Hastenrath, S., Nicklis, A. and Greischar, L., (1993). Atmospheric-hydrospheric mechanisms of climate anomalies in the western equatorial Indian Ocean. *J. Geophys. Res.*, **98**, 20219–20235.

Hay, S., Were, E., Renshaw, M., Noor., A.M., Ochola, S. and Olusanmi, I., (2003). Forecasting, warning, and detection of malaria epidemics: a case study. *Lancet*, **361**, 1705-1706.

Hayes, M. J. and Decker, W. L., (1996). Using NOAA AVHRR data to estimate maize production in the United States Corn Belt. *Int. J. Remote Sens.*, **17**, 3189–3200.

Harvey, A.W. and Mallya, G.A., (1995). Predicting the severity of *Spodoptera exempta* (Lepidoptera: Noctuidae) outbreak seasons in Tanzania. *Bull. Entomol. Res.*, **85**, 479-487.

Heinemann, A.B., Hoogenboom, G. and Chojnicki, B., (2002). The impact of potential errors in rainfall observation on the simulation of crop growth, development and yield. *Ecol. Model.*, **157**, 1-21.

- Herman, A., Kumar, V.B., Arkin, P.A. and Kousky, J.V., (1997). Objectively determined 10-day African rainfall estimates created for famine early warning systems. *Int. J. Remote Sens.*, **18**, 2147-2159.
- Hinkle, D.E., Wiersma, W. and Jurs, S.G., (1988). Applied statistics for the behavioral sciences. 2nd ed. Boston: Houghton Mifflin Company.
- Holben, B.N., (1986). Characteristics of maximum-values composite images from temporal AVHRR data. *Int. J. Remote Sens.*, **7**, 1417– 1434.
- Hossain, F. and Anagnostou, E.N., (2004). Assessment of current passive-microwave- and infrared-based satellite rainfall remote sensing for flood prediction. *J. Geophys. Res.- Atmos.*, **109**, Art. No. D07102.
- Hossain, F., Anagnostou, E.N. and Dinku, T., (2004a). Sensitivity analyses of satellite rainfall retrieval and sampling error on flood prediction uncertainty. *IEEE T. Geosci. Remote*, **42**, 130-139.
- Hossain, F., Anagnostou, E.N., Dinku, T., and Borga, M., (2004b). Hydrological model sensitivity to parameter and radar rainfall estimation uncertainty. *Hydrol. Process.*, **18**, 3277-3291.
- Hossain, F. and Anagnostou, E.N., (2005). A two-dimensional satellite rainfall error model. In press *IEEE T. Geosci. Remote*.
- Hsu, K.L., Gupta, H.V., Gao, X.G., Sorooshian, S., (1999). Estimation of physical variables from multichannel remotely sensed imagery using a neural network: Application to rainfall estimation. *Water Resour. Res.*, **35**, 1605-1618.
- Huffman, G.J., (1997). Estimates of root-mean-square random error for finite samples of estimated precipitation. *J. Appl. Meteor.*, **36**, 1191-1201.

Huffman, G.J., Adler, R.F., Stocker, E.F., Bolvin, D.T. and Nelkin E.J., (2003). Analysis of TRMM 3-hourly multi-satellite precipitation estimates computed in both real and post-real time. Preprints, *12th Conf. on Satellite Meteorology and Oceanography*, Long Beach, CA, Amer. Meteor. Soc., CD-ROM, P4.11.

Hutchinson, P., (1992). The Southern Oscillation and prediction of 'Der' season rainfall in Somalia. *J. Climate*, **5**, 525–531.

Ingram, K.T., Roncoli, M.C. and Kirshen, P.H., (2002). Opportunities and constraints for farmers of West Africa to use seasonal precipitation forecasts with Burkina Faso as a case study. *Agric. Syst.*, **74**, 331–349.

Ininda, J., Desalegne, B. and Befekadu, A., (1987). The characteristics of rainfall in Ethiopia and its relationship to El Nino–Southern Oscillation. In *Proceedings of the First Technical Conference on Meteorology Research in Eastern and Southern Africa*, Kenya Meteorology Department, Nairobi, 6–9 January 1987; 133–135.

Ismail, A.M. and Hall, A.E., (1999). Reproductive stage heat tolerance, leaf membrane thermostability and plant morphology in cowpea. *Crop Sci.*, **39**, 1762–1768.

Janicot, S., (1992). Spatiotemporal variability of West African rainfall. Part 2: Associated surface and airmass characteristics. *J. Climate*, **5**, 499–511.

Janicot, S., Moron, V. and Fontaine, B., (1996). Sahel droughts and ENSO dynamics. *Geophys. Res. Lett.*, **23**, 515–518.

Jeffrey, S.J., Carter, J.O., Moodie, K.B. and Beswick, A.R., (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environ. Modell. Softw.*, **16**, 309–330.

Joyce, R.J., Janowiak, J.E., Arkin, P.A., and Xie, P., (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.*, **5**, 487–503.

Jury, M.R., McQueen, C. and Levey, K., (1994). SOI and QBO signals in the African region. *Theor. Appl. Climatol.*, **50**, 103–115.

Justice, C.O., Eck, T.F., Tandr , D. and Holben, B.N., (1991). The effect of water vapour on the normalized difference vegetation derived for the Sahelian region from NOAA AVHRR data. *Int. J. Remote Sens.*, **12**, 1165–1187.

Kavvas, M.L., Saquib, M.N. and Puri, P.S., (1987). On a stochastic description of the time-space behavior of extratropical cyclonic precipitation fields. *Stoch. Hydrol. Hydraul.*, **1**, 37-52.

Kedem, B. and Pavlopoulos, H., (1991). On the threshold method for rainfall estimation: Choosing the optimal threshold. *J. Am. Stat. Assoc.*, **86**, 626-633.

Kidd, C., (2001). Satellite rainfall climatology: A review. *Int. J. Climatol.*, **21**, 1041-1066.

Kidd, C., Kniveton, D.R., Todd, M.C. and Bellerby, T.J., (2003). Satellite rainfall estimation using combined passive microwave and infrared algorithms. *J. Hydrometeor.*, **4**, 1088–1104.

Kilian, A.H., Langi, P., Talisuna, A. and Kabagambe G., (1999). Rainfall pattern, El Nino and malaria in Uganda. *Trans. R. Soc. Trop. Med. Hyg.*, **93**, 22-23.

Kobold, M. and Sugelj, K., (2005). Precipitation forecasts and their uncertainty as input into hydrological models. *Hydrol. Earth Syst. Sc.*, **9**, 322-332.

Krajewski, W.F., Lakshmi, V., Georgakakos, K.P. and Jain, S.C., (1991). A Monte-Carlo study of rainfall sampling effect on a distributed catchment model. *Water Resour. Res.*, **27**, 119-128.

Krige, D.G., (1994). An analysis of some essential basic tenets of geostatistics not always practised in ore valuations, in *Proceedings Regional APCOM: Computer Applications and Operations Research in the Minerals Industries, Slovenia*, 15-18.

Krige, D.G., (1996a). A basic perspective on the roles of classical statistics, data search routines, conditional biases and information and smoothing effects in ore block valuations, in *Proceedings Conference on Mining Geostatistics, Kruger National Park, South Africa*, 1-10 (Geostatistical Association of South Africa).

Krige, D.G., (1996b). A practical analysis of the effects of spatial structure and data available and accessed, on conditional biases in ordinary kriging, in *Proceedings Fifth International Geostatistical Congress, Geostatistics Wollongong '96* (Eds: E Y Baafi and N A Schofield), 799-810.

Kyriakidis, P.C., Miller, N.L., and Kim, J. (2004). A spatial time series framework for simulating daily precipitation at regional scales. *J. Hydrol.*, **297**, 236-255.

Lamb, P.J., (1978a). Large-scale Tropical Atlantic surface circulation patterns associated with sub-saharan weather anomalies. *Tellus A*, **30**, 240–251.

Lamb, P.J., (1978b). Case-studies of tropical Atlantic surface circulation pattern during recent sub-saharan weather anomalies: 1967 and 1968. *Mon. Wea. Rev.*, **106**, 482–491.

Lanza, L.G., (2000). A conditional simulation model of intermittent rain fields. *Hydrol. Earth Syst. Sc.*, **4**, 173-183.

- Lare, A.R. and Nicholson, S.E., (1994). Contrasting conditions of surface water balance in wet years and dry years as a possible land surface-atmosphere feedback mechanism in the West African Sahel. *J. Climate*, **7**, 653–668.
- Laurent, H., Jobard, I. and Toma, A., (1998). Validation of satellite and ground-based estimates of precipitation over the Sahel. *Atmos. Res.*, **48**, 651-670.
- LeCam, L., (1961). A stochastic description of precipitation, In J. Neyman (ed.), *Proc. 4th Berkeley Symp. Math., Stats. and Prob.*, Vol. 3, Univ. California, Berkeley, 165-186.
- Levizzani, V., Amorati, R. and Meneguzzo, F., (2002). A review of satellite-based rainfall estimation methods. *European Commission Project MUSIC Report (EVK1-CT-2000-00058)*, 66 pp.
- Lebel, T., Bastin, G., Obled, C. and Creutin, J.D., (1987). On the accuracy of areal rainfall estimation – A case-study. *Water Resour. Res.*, **23**, 2123-2134.
- Lindesay, J.A., Harrison, M., Haffner, T. (1986). The Southern Oscillation and South African rainfall. *S. Afr. J. Sci.*, **82**, 196–198.
- Loevinsohn, M.E., (1994). Climatic warming and increased malaria incidence in Rwanda. *Lancet*, **343**, 714-718.
- Linthicum, K.J., Anyamba, A., Tucker, C.J., Kelley, P.W., Myers, M.F., Peters, C.J., (1999). Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science*, **285**: 397-400.
- Lough, J., (1986). Tropical Atlantic sea surface temperatures and rainfall variations in Sub-saharan Africa. *Mon. Wea. Rev.*, **114**, 561–570.

- Lovejoy, S. and Austin, G.L., (1979). The delineation of rain areas from visible and IR satellite data from GATE and mid-latitudes. *Atmos.-Ocean*, **17**, 77-92.
- Lovejoy, S. and Mandelbrot, B.B., (1985). Fractal Properties of Rain, and a Fractal Model. *Tellus*, **37A**, 209-232.
- Marzano, F. S., Palmacci, M., Cimini, D., Giuliani, G. and Turk, F. J., (2004). Multivariate statistical integration of satellite infrared and microwave radiometric measurements for rainfall retrieval at the geostationary scale. *IEEE Trans. Geosci. Remote Sens.*, **42**, 1018–1032.
- Matheron, G., (1976) A simple substitute for conditional expectation: The disjunctive kriging. *NATO ASI Ser., Ser. C*, **24**, 221-236.
- Matsui, T., Omasa K. and Horie, T., (2001). The difference in sterility due to high temperatures during theowering period among japonica rice varieties. *Plant Prod. Sci.* **4**, 90–93.
- Matthews, R., Stephens, W., Hess, T., Mason, T. and Graves, A., (2000). Crop-soil simulation models: applications in developing countries. (Matthews R. and Stephens, W., Ed.), CABI, Wallingford.
- Martin, D.W. and Howland, M.R., (1986). Grid History: A geostationary satellite technique for estimating daily rainfall in the tropics. *J. Appl. Meteor.*, **25**, 184-195
- Martin, D. and Suomi, V. E., (1972). A satellite study of cloud clusters over the tropical north Atlantic Ocean. *Bull. Amer. Meteor. Soc.*, **53**, 135-156.
- Maselli, F., Conese, C., Petkov, L., and Gilabert, M.A., (1993). Environmental monitoring and crop forecasting in the Sahel through the use of NOAA NDVI data. A case study: Niger 1986–89. *Int. J. Remote Sens.*, **14**, 3471–348.

Maselli, F., Romanelli, S., Bottai, L. and Maracchi, G., (2000). Processing of GAC NDVI data for yield forecasting in the Sahelian region. *Int. J. Remote Sens.*, **21**, 3509-3523.

McKee, T.B. and Cox, S.K., (1974). Scattering of Visible Radiation by Finite Clouds. *J. Atmos. Sci.*, **31**, 1885-1892.

Mearns, L.O., Rosenzweig, C., Goldberg, R., (1996). The effect of changes in daily and interannual climatic variability on cereals-wheat: A sensitivity study. *Climatic Change*, **32**, 257-292.

Mellor, D., (1996). The Modified Turning Bands (MTB) model for space-time rainfall.1. Model definition and properties. *J. Hydrol.*, **175**, 113-127.

Mellor, D. and O'Connell, P.E., (1996). The Modified Turning Bands (MTB) model for space-time rainfall .2. Estimation of raincell parameters. *J. Hydrol.*, **175**, 129-159.

Mellor, D. and Metcalfe, A.V., (1996). The Modified Turning Bands (MTB) model for space-time rainfall .3. Estimation of the storm rainband profile and a discussion of future model prospects. *J. Hydrol.*, **175**, 161-180.

Mejía, J.M. and Rodríguez-Iturbe, I., (1974). Synthesis of random field sampling from spectrum – application to generation of hydrologic spatial processes. *Water Resour. Res.*, **10**, 705-711.

Mkhabela, M.S., Mkhabela, M.S. and Mashinini, N.N., (2005). Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR. *Agric. Forest Meteorol.*, **129**, 1-9.

- Milford, J.R. and Dugdale, G., (1990). Estimation of rainfall using geostationary satellite data. *Applications of Remote Sensing in Agriculture, Butterworth, London Proceedings of the 48<sup>th</sup> Easter School in Agricultural Science*. University of Nottingham. April 1989.
- Morland, J.C., Grimes, D.I.F. and Hewison, T.J., (2001). Satellite observations of the microwave emissivity of a semi-arid land surface. *Remote Sens. Environ.*, **77**, 149-164.
- Moulin, S., Bondeau, A. and Delecolle, R., (1998). Combining agricultural crop models and satellite observations: from field to regional scales. *Int. J. Remote Sens.*, **19**, 1021-1036.
- Nageswara Rao, R.C., Sardar Singh, Sivakumar, M.V.K., Srivastava, K.L., (1985). Effect of water deficit at different growth phases of peanut. I. Yield responses. *Agron. J.*, **77**, 782-786.
- Nicholson, S.E., (1993). An overview of African Rainfall fluctuations of the last decade. *J. Climate*, **6**, 1463-1466.
- Nicholson, S.E., (2001). Climatic and environmental change in Africa during the last two centuries. *Climate Res.*, **17**, 123-144.
- Nicholson, S.E., (2005). On the question of the "recovery" of the rains in the West African Sahel. *J. Arid Environ.*, **63**, 615-641.
- Nicholson, S.E., Some, B. and Kone, B., (2000). An analysis of recent rainfall conditions in West Africa, including the rainy seasons of the 1997 El Nino and the 1998 La Nina years *J. Climate*, **13**, 2628-2640.
- Nicholson, S. E., Some, B., McCollum, J., Nelkin, E., Klotter, D., Berte, Y., Diallo, B.M., Gaye, I., Kpabeba, G., Ndiaye, O., Noukpozoukou, J.N., Tanu, M.M., Thiam, A., Toure, A.A. and Traore, A.K., (2003a) Validation of TRMM and other rainfall estimates

with a high-density gauge dataset for West Africa. Part I: Validation of GPCP rainfall product and pre-TRMM satellite and blended products. *J. Appl. Meteor.*, **42**, 1337–1354.

Nicholson, S. E., Some, B., McCollum, J., Nelkin, E., Klotter, D., Berte, Y., Diallo, B.M., Gaye, I., Kpabebe, G., Ndiaye, O., Noukpozoukou, J.N., Tanu, M.M., Thiam, A., Toure, A.A. and Traore, A.K., (2003b). Validation of TRMM and other rainfall estimates with a high-density gauge dataset for West Africa. Part II: Validation of TRMM rainfall products. *J. Appl. Meteor.*, **42**, 1355–1368.

Nonhebel, S, (1994a). Inaccuracies in weather data and their effects on crop growth simulation result s. II. Water-limited production. *Climate Res.*, **4**, 61-74.

Nonhebel, S, (1994b). The effects of use of average instead of daily weather data in crop growth simulation-models. *Agric. Syst.* **44**, 377-396.

Ogallo, L.J., Janowiak, J.E. and Halpert, M.S., (1988). Teleconnection between seasonal rainfall over East Africa and global sea-surface temperature anomalies. *J. Meteorol. Soc. Jpn.*, **66**, 807–822.

Onibon, H., Lebel, T., Afouda, A. and Guillot, G., (2004). Gibbs sampling for conditional spatial disaggregation of rain fields. *Water Resour. Res.*, **40**, Art. No. W08401.

Opokuankomah, Y. and Cordery, I., (1994). Atlantic sea-surface temperatures and rainfall variability in Ghana. *J. Climate*, **7**, 551-558.

Paeth, H. and Friederichs, P., (2004). Seasonality and time scales in the relationship between global SST and African rainfall. *Clim. Dynam.*, **23**, 815-837.

Paeth, H. and Hense, A., (2004). SST versus climate change signals in West African rainfall: 20th-century variations and future projections. *Climatic Change*, **65**, 179-208.

Pallas, J.E., Jr., Stansell, J.R., and Koske, T.J., (1979). Effects of drought on Florunner peanuts. *Agron. Jour.*, **71**, 853-858.

Pansofsky, H.A. and Brier, G.W., (1965). Some Applications of Statistics to Meteorology. The Pennsylvania State University Press.

Pardo-Igúzquiza, E., (1998). Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *J. Hydrol.*, **210**, 206-220.

Pardo-Igúzquiza, E., (1999). VARFIT: a Fortran-77 program for fitting variogram models by weighted least squares. *Comput. Geosci.*, **25**, 251-261.

Pardo-Igúzquiza, E., Grimes, D.I.F., and Teo, C.K., (2006), Assessing the uncertainty associated with intermittent rainfall fields. *Water Resour. Res.*, **42**, W01412, doi:10.1029/2004WR003740.

Pellenq, J. and Boulet, G., (2003). A methodology to test the pertinence of remote-sensing data assimilation into vegetation models for water and energy exchange at the land surface. *Agronomie*, **24**, 197–204.

Perica, S. and Foufoula-Georgiou, E., (1996). Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions. *J. Geophys. Res.-Atmos.*, **101**, 26347-26361.

Prasad, P.V.V., Craufurd, P.Q., Summerfield, R.J. and Wheeler, T.R., (2000). Effects of short episodes of heat stress on over production and fruit-set of groundnut *Arachis hypogaea* L. *J. Exp. Bot.* **51**, 777–784.

Priestly, C.H.B. and Taylor, R.J., (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Wea. Rev.*, **100**, 81–92.

Rasmussen, M.S., (1992), Assessment of millet yields and production in northern Burkina-Faso using integrated NDVI from the AVHRR. *Int. J. Remote Sens.*, **13**, 3431-3442.

Rao, K.N., Gadgil, S., Rao, P.R.S. and Savithri, K., (2000). Tailoring strategies to rainfall variability - The choice of the sowing window. *Curr. Sci. India*, **78**, 1216-1230.

Reynolds, D.W. and Vonder Haar T.H., (1973). A comparison of radar determined cloud height and reflected solar radiance measured from geosynchronous satellite ATS-3. *J. Appl. Meteor.*, **12**, 1082-1085.

Reynolds, C.A., Yitayew, M., Slack, D.C., Hutchinson, C.F., Huete, A. and Petersen, M.S. (2000). Estimating crop yields and production by integrating the FAO Crop specific Water Balance model with real-time satellite data and ground-based ancillary data. *Int. J. Remote Sens.*, **21**, 3487-3508.

Rhenals-Figueroa, A.E., Rodríguez-Iturbe, I. and Shaake, Jr. J.C., (1974). Bidimensional spectral analysis of rainfall events. Cambridge, Mass.: Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics, M.I.T. Technical Report 193.

Riha, S.J., Wilks, D.S. and Simoens, P., (1996). Impact of temperature and precipitation variability on crop model predictions. *Climatic Change*, **32**, 293-311.

Rijks, D., Rembold, F., Nègre, T., Gommès, R. and Cherlet, M. (Eds), (2003). Crop and Rangeland Monitoring in Eastern Africa - for Early Warning and Food Security, *Proceedings of the International Workshop on Crop and Rangeland Monitoring in East Africa*, Nairobi, Jan., 2003; European Commission (pub).

Rodríguez-Iturbe, I. and Mejía, J.M., (1974). The design of rainfall networks in time and space. *Water Resour. Res.*, **10**, 713-729.

- Rodríguez-Iturbe, I., Wang, Q., Jacobs, B.L. and Eagleson, P.S., (1986), Spatial modeling of total storm rainfall. *Proc. R. Soc. London, Ser. A*, **403**, 27-50.
- Roebeling, R.A., Rosema, A., Kashasha, D.A., Masamvu, K. and van der Harten C.A.J., (1999). ACMP Agromet and Crop Monitoring Project in the SADC region- Phase 3. BCRS rapport, Delft, the Netherlands.
- Ropelewski, C.F. and Halpert, M.S., (1987). Global and regional scale precipitation and temperature patterns associated with El Nino/Southern Oscillation. *Mon. Wea. Rev.*, **115**, 1606–1626.
- Rosenzweig, M.R. and Binswanger, H.P., (1993). Wealth, weather risk and the composition and profitability of agricultural investments. *Econ. J.*, **103**, 56–78.
- Rowell, D.P., Folland, C.K., Maskell, K. and Ward M.N., (1995). Variability of summer rainfall over tropical North Africa. *Q. J. Roy. Meteor. Soc.*, **121**, 669–704.
- Sanchez, P.A., (2002). Soil fertility and hunger in Africa. *Science*, **295**, 2019–2020, doi:10.1126/science.1065256.
- Saji, N.H., Goswami, B.N., Vinayachandran, P.N. and Yamagata T., (1999). A dipole mode in the tropical Indian Ocean. *Nature*, **401**, 360–363.
- Sandholt, I., Andersen, J., Dybkjaer, G., Nyborg, L., Lo, M., Rasmussen, K., Refsgaard, J.C., Jensen, K.H. and Toure, A., (2003). Integration of earth observation data in distributed hydrological models: the Senegal River basin. *Can. J. Remote Sens.*, **29**, 701-710.

- Savage, G.P. and Keenan, J.I., (1994). The composition and nutritive value of groundnut kernels. In “*The Groundnut Crop: A Scientific Basis for Improvement*” (J. Smart, Ed.), 173 – 213.
- Scofield, R.A. and Oliver, V.J., (1977). A scheme for estimating convective rainfall from satellite imagery. *NOAA Tech. Memo. NESS*, **86**, 47 pp.
- Schilling, R., (1996). Groundnut in *The Tropical Agriculturalist series*. (Rene C., Ed.). Macmillian Education, London, 146 pp.
- Seleshi, Y., Demarèe, G.R., (1995). Rainfall variability in the Ethiopian and Eritrean highlands and its links with the Southern Oscillation Index. *J. Biogeogr.*, **22**, 945–952.
- Semenov, M.A. and Porter, J.R., (1995). Climatic variability and the modeling of crop yields. *Agric. Forest. Meteorol.*, **73**, 265-283.
- Senay, G.B. and Verdin, J.P., (2003). Characterization of yield reduction in Ethiopia using a GIS-based crop water balance model. *Can. J. Remote Sens.*, **29**, 687-692.
- Senay, G.B. and Verdin, J.P., (2004). Developing index maps of water-harvest potential in Africa. *Appl. Eng. Agric.*, **20**, 789-799.
- Seo, D.J., (1998). Real-time estimation of rainfall fields using rain gage data under fractional coverage conditions. *J. Hydrol.*, **208**, 25-36.
- Servain, J., (1991). Simple climatic indices for the tropical Atlantic ocean and some applications. *J. Geophys. Res.*, **96**,15137–15146.
- Shah, S.M.S., O’Connell, P.E. and Hosking, J.R.M., (1996). Modelling the effects of spatial variability in rainfall on catchment response. 2. Experiments with distributed and lumped models. *J. Hydrol.*, **175**, 89-111.

- Sivapalan, M. and Wood, E.F., (1987). A multidimensional model of nonstationary space-time rainfall at the catchment scale. *Water Resour. Res.*, **23**, 1289-1299.
- Sivapalan, M. Woods, R.A. and Kalma, J.D., (1997). Variable bucket representation of TOPMODEL and investigation of the effects of rainfall heterogeneity. *Hydrol. Process.*, **11**, 1307-1330.
- Smith, E., (2003). NASA's scientific agenda for GPM mission. *Proc. Third GPM Workshop on Consolidating the Concept*, Noordwijk, Netherlands, ESTEC.
- Smith, J.A. and Karr, A.F., (1985). Parameter-estimation for a model of space-time rainfall. *Water Resour. Res.*, **21**, 1251-1257.
- Smith, J.A. and Krajewski, W.F., (1987). Statistical modeling of space-time rainfall using radar and rain-gauge observations. *Water Resour. Res.*, **23**, 1893-1900.
- Sorooshian, S., Hsu K., Gao, X., Gupta H., Imam, B. and Braithwaite, D., (2000). Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046.
- Spencer, R.W., (1993). Global oceanic precipitation from the MSU during 1979-91 and comparisons to other climatologies. *J. Climate*, **6**, 1301-1326.
- Steiner, M., Bell, T.L., Zhang, Y. and Wood, E.F., (2003). Comparison of two methods for estimating the sampling-related uncertainty of satellite rainfall averages based on a large radar dataset. *J. Climate*, **16**, 3759-3778.
- Stone, R.C., Smith, I. and McIntosh, P., (2000). Statistical methods for deriving seasonal climate forecasts from GCMs. In *Applications of seasonal climate forecasting in agriculture and natural ecosystems: the Australian experience* (ed. G. L. Hammer, N.

Nicholls & C. Mitchell), 135–147. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Stout, J.E., Martin, D.W. and Sikdar D.N., (1979). Estimating GATE rainfall with geosynchronous satellite images. *Mon. Wea. Rev.*, **107**, 585–598.

Suleiman, A.A., (1999). Assessing and modeling the spatial variability of soil water redistribution and wheat yield along a sloping landscape. Ph.D. Thesis. Crop and Soil Sciences. Department, Michigan State University.

Sultan, B., Baron, C., Dingkuhn, M., Sarr, B. and Janicot, S. (2005). Agricultural impacts of large-scale variability of the West African monsoon. *Agric. Forest Meteorol.*, **128**, 93–110.

Sutton, R.T., Jewson, S.P. and Rowell, D.P., (2000). The elements of climate variability in the tropical Atlantic region. *J. Climate*, **13**, 3261–3284.

Sun, X., Mein, R.G., Keenan, T.D. and Elliott, J.F., (2000). Flood estimation using radar and raingauge data. *J. Hydrol.*, **239**, 4-18.

Syed, K.H., Goodrich, D.C., Myers, D.E., Sorooshian, S., (2003). Spatial characteristics of thunderstorm rainfall fields and their relation to runoff. *J. Hydrol.*, **271**, 1-21.

Tadesse, T., (1994). Summer monsoon seasonal rainfall of Ethiopia in ENSO episodic years. In *Proceedings of the International Conference on Monsoon Variability and Prediction*, Trieste, Italy, 9–13 May 1994, WMO/TD no. 619, 48–55.

Taylor, C.M., (2005). Soil-moisture-rainfall feedbacks in the Sahel (abstract). In AMMA Dakar: 1<sup>st</sup> International Conference, 28<sup>th</sup> Nov – 2<sup>nd</sup> Dec 2005, Dakar, Senegal.

Taylor, C.M., Saïd F. and Lebel, T., (1997). Interactions between the land surface and mesoscale rainfall variability during HAPEX-Sahel. *Mon. Wea. Rev.*, **125**, 2211–2227.

Taylor, C.M. and Lebel, T., (1998) Observational evidence of persistent convective scale rainfall patterns. *Mon. Wea. Rev.*, **126**, 1597–1607.

Taylor, C.M., Ellis, R.J., Parker, D.J., Burton, R.R. and Thorncroft, C.D., (2003). Linking boundary-layer variability with convection: A case-study from JET2000. *Q. J. Roy. Meteor. Soc.*, **129**, 2233-2253.

Teklehaimanot, H.D., Schwartz, J., Teklehaimanot, A. and Lipsitch, M., (2004). Weather-based prediction of Plasmodium falciparum malaria in epidemic-prone regions of Ethiopia II. Weatherbased prediction systems perform comparably to early detection systems in identifying times for interventions. *Malar. J.*, **3**, Art. No. 44.

Thomson, M.C., Mason, S.J., Phindela, T. and Connor, S.J., (2005). Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana. *Am. J. Trop. Med. Hyg.*, **73**, 214-221.

Thorne, V., Coakeley, P., Grimes D.I.F. and Dugdale, G., (2001). Comparison of TAMSAT and CPC rainfall estimates with raingauges, for southern Africa. *Int. J. Remote Sens.* **22**, 1951-1974.

Thornton, P.K., Bowen, W.T., Ravelo, A.C., Wilkens, P.W., Farmer, G., Brock, J. and Brink, J.E., (1997). Estimating millet production for famine early warning: An application of crop simulation modelling using satellite and ground-based data in Burkina Faso. *Agric. Forest Meteorol.*, **83**, 95-112.

Todd, M.C., Barrett, E.C., Beaumont, M.J. and Bellerby, T.J., (1999). Estimation of daily rainfall over the upper Nile river basin using a continuously calibrated satellite infrared technique. *Meteorol. Appl.*, **6**, 201-210.

- Todd, M.C., Kidd, C., Bellerby, T.J. and Kniveton, D.R., (2001). A combined satellite infrared and passive microwave technique for estimation of small-scale rainfall. *J. Atmos Ocean. Tech.*, **18**, 742-755.
- Todd, M.C., Washington, R., Cheke, R.A. and Kniveton, D.R., (2002). Brown locust outbreaks and climate variability in southern Africa. *J. Appl. Ecol.*, **39**, 31-42.
- Tsintikidis, D., Georgakakos, K.P., Artan, G.A. and Tsonis, A.A., (1999). A feasibility study on mean areal rainfall estimation and hydrologic response in the Blue Nile region using METEOSAT images. *J. Hydrol.*, **221**, 97-116.
- Tsonis, A.A., Triantafyllou, G.N. and Georgakakos, K.P. (1996). Hydrological applications of satellite data .1. Rainfall estimation. *J. Geophys. Res.-Atmos.*, **101**, 26517-26525.
- Tucker, M.R., (1984). Forecasting the severity of armyworm seasons in East Africa from early season rainfall. *Insect Sci. Appl.*, **5**, 51-55.
- UN Millennium Project, (2005). Halving Hunger: It Can Be Done. Task Force on Hunger. (Available from: [http://www.unmillenniumproject.org/reports/tf\\_hunger.htm](http://www.unmillenniumproject.org/reports/tf_hunger.htm))
- USDA-SCS, (1964). National Engineering Handbook, Section 4, Part I, Watershed Planning. Soil Conservation Service, US Department of Agriculture, Washington, DC.
- van Heerden, J., Terblanche, D.E. and Schulze, G.C., (1988). The Southern Oscillation and South African summer rainfall. *J. Climatol.*, **8**, 577-597.
- Verdin, J. and Klaver, R., (2002). Grid-cell-based crop water accounting for the famine early warning system. *Hydrol. Process.*, **16**, 1617-1630.

Verdin, J., Funk, C., Senay, G. and Choularton, R., (2005). Climate science and famine early warning. *Philos. T. Roy. Soc. B.*, **360**, 2155-2168.

Vordzorgbe, S.D., (2003). Regional report on early warning of natural disasters in Africa. Second International Conference on Early Warning 16 – 18 October, Bonn, Germany.

Walker, T.S., (1991). Improved technology in relation to other options for dealing with climatic risk to crop production in the semiarid tropics. In *Climate risk in crop production: models and management for the semiarid tropics and subtropics* (ed. Muchow R.C. and Bellamy J.A.), 511–525. Walingford: CAB International.

Ward, M.N., (1998). Diagnosis and short-lead time prediction of summer rainfall in tropical north Africa at interannual and multidecadal timescales. *J. Climate*, **11**, 3167–3191.

Washington, R., Harrison, M., Conway, D., Black, E.C.L., Challinor, A.J., Grimes, D.I.F., Jones, R., Morse, A. and Todd, M., (2006) African Climate - Forecast to Change? In press *Bull. Amer. Met. Soc.*

Waymire, E., Gupta, V.K. and Rodríguez-Iturbe, I., (1984). A spectral theory of rainfall intensity at the meso-beta scale. *Water Resour. Res.*, **20**, 1453-1465.

Webster, P.J., Moore, A.M., Loschnigg, J.P. and Leben, R.R., (1999). Coupled ocean–temperature dynamics in the Indian Ocean during 1997–98. *Nature*, **401**, 356–360.

Weiss, E.A., (1983). Oilseed Crops. Longman, London, 660 pp.

Wheeler, T.R., Craufurd, P.Q., Ellis, R.H., Porter, J.R. and Prasad, P.V.V., (2000). Temperature variability and the annual yield of crops. *Agric. Ecosyst. Environ.* **82**, 159–167.

- Williams, J.B., (1975). Soil water investigation in The Gambia. Technical Bulletin 3, Land Resource Development Centre, Ministry of Overseas Development, UK.
- Wilks, D.S., (1989). Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resour. Res.*, **25**, 1429-1439.
- Wilks, S.D., (1995). Statistical methods in the atmospheric sciences. Academic Press.
- Williams, J.H., Nageswara Rao, R.C., Matthews R. and Harris D., (1985). Responses of Groundnut Genotypes to Drought. Agrometeorology of Groundnut. In *Proceedings of an International Symposium*. ICRISAT Sahelian Center. Niamey, Niger.
- Wright G.C. and Nageswara Rao R.C., (1994). Groundnut water relations. In “*The Groundnut Crop: A Scientific Basis for Improvement*” (J. Smart, Ed.), 281 – 325.
- Woodley, W.L. and Sancho, B., (1971). A first step towards rainfall estimation from satellite cloud photographs, *Weather*, July, 279-289.
- Xie, P.P. and Arkin, P.A., (1996). Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *J. Climate*, **9**, 840-858.
- Xu, L.M., Gao, X.G., Sorooshian, S., Arkin, P.A. and Imam, B., (1999). A microwave infrared threshold technique to improve the GOES precipitation index. *J. Appl. Meteor.*, **38**, 569–579.
- Yuan, M. and Duchon, C.E., (2001). Estimation of daily area-average rainfall in central Florida using arithmetic averaging and kriging. *Phys. Geogr.*, **22**, 42-58.
- Zhou, G.F., Minakawa, N., Githeko, A.K., Snow, R.W., Randolph, S.E. and Rogers, D.J., (2005). Climate variability and malaria epidemics in the highlands of East Africa. *Trends Parasitol.*, **21**, 54-56.

---

Zimmerman, F.J. and Carter, M.R., (2003). Asset smoothing, consumption smoothing and the reproduction of inequality under risk and subsistence constraints. *J. Dev. Econ.*, **71**, 233–260, doi:10.1016/S0304-3878(03)00028-2.