

Seasonal Prediction of African Rainfall
With a Focus on Kenya

Jennifer Mary Anne Rourke

Mullard Space Science Laboratory
Department of Space and Climate Physics
University College London

*A Thesis submitted to University College London
for the degree of Doctor of Philosophy*

February 2011

Declaration

I, Jennifer Mary Anne Rourke, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. I would like to gratefully acknowledge my supervisor, Prof. Mark Saunders, for his ideas that laid the foundations for the work presented in this thesis.

Jennifer Mary Anne Rourke

Abstract

Africa's climate is prone to extended rainfall deficits. In extreme cases these may lead to droughts and humanitarian disasters. Skilful prediction of seasonal rainfall would bring sound humanitarian and economic benefit to the many African countries that depend on rain-fed agriculture.

Seasonal rainfall hindcast skill from the DEMETER multi-model ensemble system is examined across Africa. Skill at 0-month lead is found to be weak over much of Africa, except for the August-October (ASO) season in the Sahel and the November-January (NDJ) season in equatorial East Africa, Nigeria and South Africa. For the ASO season, correlation values of 0.3-0.8 (p -values < 0.1) are found across the sub-Saharan belt. For the NDJ season, correlation values of 0.5-0.6 (p -values < 0.1) occur in Kenya, Tanzania and Uganda.

Innovative statistical seasonal rainfall hindcast models are developed for six homogeneous rainfall regions in Kenya, using linear regression techniques. Kenya has experienced seven severe droughts over the period 1991-2008 affecting over 35 million people. Lagged sea surface temperature and atmospheric wind predictors are selected based on having a significant and temporally stable correlation with regional rainfall indices, and a clear physical-linking mechanism. Moderate-to-high rainfall hindcast skill is found for most regions at 0- and 1-month leads for the October-December rainy season. In contrast, no robust predictors are found for the March-May rainy season.

In 2009 an improved version of DEMETER, called EUROSIP, was released. This study is the first to assess the skill of the EUROSIP rainfall hindcasts for the Kenyan October-December rainy season and to compare this with the statistical model skill. For the most heavily populated and cultivated West and Southwest regions of Kenya, which are home to 68% of the Kenyan population, the statistical models outperform the EUROSIP model with correlation values ≥ 0.42 (p -values ≤ 0.06) over the common verification period 1987-2005.

Contents

List of Tables	8
List of Figures.....	12
Chapter 1 Introduction.....	19
1.1 Motivation and objectives of the thesis.....	19
1.2 Thesis outline	23
Chapter 2 Background and Literature Review.....	24
2.1 Africa and its climate	24
2.1.1 African droughts.....	24
2.1.2 Selecting a suitable country to study.....	27
2.2 Kenya and its climate	28
2.2.1 Drought impacts in Kenya	30
2.2.2 Kenyan drought case studies.....	33
2.3 Current state of knowledge of Kenyan rainfall variability and predictability	34
2.3.1 Teleconnections between Sea Surface Temperatures (SSTs) and	36
rainfall	36
2.3.2 Pacific Ocean: El Niño Southern Oscillation (ENSO).....	36
2.3.3 Indian Ocean: Indian Ocean Dipole (IOD)	39
2.3.4 The relationship between the IOD and ENSO	41
2.3.5 Atlantic Ocean.....	45
2.4 Long-range forecasting methods.....	46
2.4.1 Statistical forecasting methods.....	46
2.4.2 Dynamical forecasting models.....	46
2.5 Operational seasonal forecasting in Kenya	47
2.5.1 The Greater Horn of Africa Climate Outlook Forum (GHA COF)	49
Chapter 3 Data	51
3.1 Rainfall.....	51
3.1.1 Gridded rainfall data	52
3.1.2 Rain gauge data	53
3.1.3 Quality control	56
3.2 Sea surface temperatures.....	58
3.3 Wind.....	58

3.4	Climate indices and oscillations.....	58
3.4.1	Niño indices (Niño 3.4 and Niño 4).....	58
3.4.2	Southern Oscillation Index (SOI).....	59
3.4.3	Dipole Mode Index (DMI).....	59
3.4.4	Quasi-Biennial Oscillation (QBO) Index.....	59
3.4.5	Madden-Julian Oscillation (MJO).....	60

Chapter 4 Assessment of Current Dynamical Seasonal Rainfall Hindcast

	Skill over Africa	61
4.1	Introduction	61
4.1.1	Aim of the assessment.....	61
4.1.2	Definitions.....	62
4.1.3	Introduction to DEMETER.....	63
4.2	Methodology	65
4.2.1	Bias correction	65
4.2.2	Skill assessment	65
4.2.3	Rank correlation (r_{rank}).....	66
4.2.4	Mean Square Skill Score (MSSS).....	67
4.2.5	Percentage of hindcasts with correct anomaly sign	67
4.2.6	Statistical significance.....	68
4.3	Results.....	69
4.3.1	ASO skill in the Sahel.....	71
4.3.2	NDJ skill in eastern and south-eastern Africa.....	71
4.3.3	Skill extension to lead of 3 months	73
4.4	Conflicting skill results	75
4.5	Verification from the DEMETER website.....	77
4.6	Discussion on lack of skill	80
4.7	Summary	82

Chapter 5 Methodology behind the Development of Statistical Seasonal

	Rainfall Hindcast Models.....	83
5.1	Clustering the rain gauge data into homogeneous rainfall regions	83
5.1.1	Methodology	86
5.1.2	Selected Kenyan rainfall regions	89
5.2	Predictor selection criteria.....	92

5.2.1	Correlation of seasonal rainfall and predictors	94
5.2.2	Stability of predictors over time.....	94
5.2.3	Significance of correlations	95
5.2.4	Potential wind predictor selection method.....	95
5.2.5	Physical link between regional rainfall and potential predictor.....	96
	indices	96
5.3	Formulation of statistical models	96
5.3.1	Test for normality of data.....	98
5.3.2	Simple linear regression.....	102
5.3.3	Multiple linear regression	105
5.4	Skill Assessment	106

Chapter 6 Statistical Seasonal Hindcast Models for the Kenyan ‘Short-Rains’

	Season.....	108
6.1	Overview of known factors contributing to interannual variability in the Kenyan ‘short-rains’	108
6.2	Skill assessment of statistical hindcast models	109
6.2.1	West Kenya	114
6.2.2	Southwest Kenya.....	115
6.2.3	South Kenya	117
6.2.4	Northeast Kenya.....	119
6.2.5	Northwest Kenya.....	121
6.2.6	Southeast Kenya.....	123
6.2.7	Splitting Southeast Kenya into Southeast-east and Southeast-west..	125
6.3	Summary	128

Chapter 7 Statistical Seasonal Hindcast Models for the Kenyan ‘Long-Rains’

	Season.....	129
7.1	Overview of reported factors contributing to the.....	129
	interannual variability of the Kenyan ‘long-rains’	129
7.1.1	The northward passage of the ITCZ	130
7.1.2	Topography, large bodies of water and other mesoscale forcings	130
7.1.3	The stratospheric QBO index.....	133
7.1.4	ENSO and the IOD.....	134
7.1.5	The MJO	134

7.2	Data quality	135
7.3	New methods to investigate ‘long-rains’ predictors	136
7.4	Potential predictors for the ‘long-rains’	138
7.4.1	SST potential predictor regions.....	139
7.4.2	New SST potential predictor indices.....	141
7.4.3	DMI and MJO potential predictor indices	141
7.4.4	QBO potential predictor index	141
7.4.5	Summary	142
7.5	The spring predictability barrier.....	142
7.6	Discussion	143
Chapter 8	Comparison of Statistical and Dynamical Hindcast Models for the Kenyan ‘Short-Rains’	144
8.1	EUROSIP	144
8.2	Methodology for comparing statistical and dynamical hindcast models for the Kenyan ‘short-rains’	147
8.2.1	EUROSIP hindcast skill assessment	147
8.2.2	Statistical hindcast skill assessment	147
8.2.3	Direct comparison methods.....	149
8.3	Comparison results.....	149
8.3.1	Statistical models: Southwest and West Kenya	150
8.3.2	Dynamical models: South and Northeast Kenya	151
8.3.3	Northwest and Southeast Kenya	153
8.4	Discussion	155
Chapter 9	Conclusions and Wider Implications	157
9.1	Summary and Conclusions.....	157
9.2	Recent drought prediction skill	159
9.3	Benefits to operational forecasting in Kenya.....	161
9.4	Future directions.....	162
Appendix A.....		164
Appendix B.....		173
Appendix C.....		182
Acknowledgements.....		185
References.....		186

List of Tables

1.1	Major African droughts over the period 1980-2008 as reported on the Emergency Events Database (EM-DAT, 2009), including the numbers of persons affected. In order to present only the major Africa droughts, this table includes only those with reports of over 1 million people affected or over 100 deaths according to EM-DAT (2009).....	21
3.1	Kenyan rain gauge stations used in the study (data obtained from the Kenya Meteorological Department). Only stations with >80% data available over the period 1959-2006 are included. The positions of these stations are shown on the map in Figure 3.1. The region names are referred to in Chapter 5.....	54
3.2	Average linear trends that are removed from the original regional OND rainfall data over the period 1959-1990 (mm).....	57
4.1	Characteristics of the three DEMETER coupled general circulation models employed in this study (Palmer et al., 2004).....	64
5.1	Population per homogeneous rainfall region. Measured in millions of people and percentage of total Kenyan population for 2000 and estimates for 2015. Population data from the Gridded Population of the World, Version 3 (GPW _{v3}), produced by CIESIN (Centre for International Earth Science Information Network) (CIESIN, 2005).....	90
5.2	Functions used to transform the OND regional rainfall indices to normal distributions. No transformations were required for the potential predictor indices. No transformations are reported for MAM as linear regression forecasts were not developed for this season.....	104
6.1	Locations of the potential predictor regions for each of the OND regional hindcast models. The regions are listed in order of increasing strength of hindcast skill as shown in Table 6.2.....	110
6.2	Table comparing the OND rainfall hindcast and replicated real-time forecast skill scores for the best models for each homogeneous rainfall region in Kenya. Predictor regions and skill score results for the forecasts from each model are shown (r_{rank} and its associated p -value and $MSSS$) over the periods 1959-1990, 1959-1974, 1975-1990 and	

	1991-2006. Red values show when: $r_{rank} \geq 0.4$, $p\text{-value} \leq 0.1$ and $MSSS > 0\%$. The forecast models are presented in order of strength of the hindcast skill, from highest to lowest.....	111
6.3	Table comparing the OND rainfall hindcast and replicated real-time forecast skill scores over the extreme years (<i>upper and lower quartiles of each period only</i>) for the best models for each homogeneous rainfall region in Kenya. Predictor regions and skill score results for the hindcasts from each model are shown (r_{rank} and its associated p -value and $MSSS$ (%)) for the upper and lower quartiles of the periods 1959-1990, 1959-1974, 1975-1990 and 1991-2006. Red values show when: $r_{rank} \geq 0.4$, $p\text{-value} \leq 0.1$ and $MSSS > 0\%$	112
6.4	OND rainfall forecast skill scores for the best hindcast model for the West region of Kenya. The skill score results are shown (r_{rank} and its associated p -value and $MSSS$) over the periods 1959-1990, 1959-1974, 1975-1990 and 1991-2006. Red values show when: $r_{rank} \geq 0.4$, $p\text{-value} \leq 0.1$ and $MSSS > 0\%$	115
6.5	As Table 6.4 but for the southwest region of Kenya.....	116
6.6	As Table 6.4 but for the south region of Kenya.....	118
6.7	As Table 6.4 but for the Northeast region of Kenya.....	120
6.8	As Table 6.4 but for the Northwest region of Kenya.....	122
6.9	As Table 6.4 but for the southeast region of Kenya.....	123
6.10	Table comparing the OND rainfall hindcast skill scores for the best hindcast models for the Southeast-east and Southeast-west sub-regions of Kenya. Predictor regions and skill score results for the forecasts from both models are shown (r_{rank} and its associated p -value and $MSSS$) over the periods 1959-1990, 1959-1974, 1975-1990 and 1991-2006. Red values show: $r_{rank} \geq 0.4$, $p\text{-value} \leq 0.1$ and $MSSS > 0\%$	126
6.11	Locations of the potential predictor regions for the OND regional hindcast models for the Southeast-east and Southeast-west sub-regions of Kenya.....	126
7.1	Results from Indeje (2000) showing contemporaneous and lagged correlations (r_{rank}) between the QBO-30 index and MAM rainfall over the 14-year period 1979-1992. Red values have p -values > 0.1 calculated using a two-sided t -test based on a sample size of 14. Regions, as defined by Indeje and Semazzi (2000), cover (a) the central and western	

	highlands of Kenya, (b) northwest Kenya and (c) western Kenya respectively. This Table is based on Table 3.1 in Indeje (2000).....	133
7.2	r_{rank} values of OND rainfall between the different regions of Kenya over the period 1959-1990. r_{rank} values > 0.4 are shown in bold.....	136
7.3	r_{rank} values of MAM rainfall between the different regions of Kenya over the period 1959-1990. r_{rank} values > 0.4 are shown in bold.....	136
7.4	Potential predictors for the ‘long-rains’ season in Kenya. Potential predictors are shown for MAM, March, April and May for the regions of the Southeast, Southwest, West and Northeast. The South and Northwest regions are not shown due to poor quality of rainfall data. “NA” indicates occasions where no potential predictors have been found. Further details on the correlations between the potential predictor and the rainfall indices can be found in the sections and figures listed in the final column.....	139
8.1	Characteristics of the ECMWF, Météo France and UKMO coupled general circulation models used in the DEMETER and EUROSIP systems (Anderson et al. 2007; Davey and Ferranti 2006; ECMWF 2009; Palmer et al. 2004; Vitart et al. 2007).....	146
8.2	Coordinates of the grid squares with the strongest correlations between OND CMAP rainfall and each regional OND rainfall index. Correlations values are shown over the period 1981-2005.....	148
8.3	Comparison of the Southwest OND rainfall hindcast skill assessments from the individual EUROSIP models, the EUROSIP multi-model and the best statistical model. The EUROSIP results are from the grid square representing the Southwest region of Kenya (35°E, 0.0°N). The model with the best overall skill scores is shown in red. The dynamical model with the strongest skill scores is highlighted green. The statistical model with the strongest skill scores is highlighted blue.....	150
8.4	As Table 8.3 but for the West of Kenya. The EUROSIP results are from the grid square representing the West region of Kenya (35°E, 0.0°N).....	151
8.5	As Table 8.3 but for the South of Kenya. The EUROSIP results are from the grid square representing the South region of Kenya (40°E, 2.5°S).....	152

8.6	As Table 8.3 but for the Northeast of Kenya. The EUROSIP results are from the grid square representing the Northeast region of Kenya (40°E, 2.5°N).....	152
8.7	As Table 8.3 but for the Northwest of Kenya. The EUROSIP results are from the grid square representing the Northwest region of Kenya (37.5°E, 2.5°N).....	153
8.8	As Table 8.3 but for the Southeast of Kenya. The EUROSIP results are from the grid square representing the Southeast region of Kenya (37.5°E,0.0°N).....	154
8.9	Summary of the comparison of the best statistical or dynamical OND rainfall hindcast models for each region of Kenya. The skill scores are calculated over the period 1987-2005, excluding the skill scores for the West of Kenya, which are calculated over the period 1981-2005.....	156
9.1	Table 9.1. OND seasonal rainfall (mm/season) from the rain gauges of the KMD, and hindcasts from the best statistical models for each region and the EUROSIP multi-model ensemble at 0-lead. The values are shown for 1999-2001 and 2005 as examples of recent notable droughts in Kenya. The underlined values highlight the regions that were most badly affected in each drought year. 2005 rain gauge data from the Northwest region of Kenya is not available.....	160

List of Figures

1.1	The proportion of persons killed by droughts between 1970 and 2006 by continent compared to other natural disasters. (Courtesy of the Centre for Research on the Epidemiology of Disasters (EM-DAT, 2006)).....	19
1.2	Annual GPCC rainfall data for Africa between 1958 and 2000. (a) Annual average rainfall totals (0-260 cm/year). (b) Standard deviation of annual rainfall assuming a Gaussian distribution (0-325mm/year).....	20
1.3	Distribution of the highest 30% mortality risk natural disaster hotspots. Geophysical refers to earthquakes, volcanoes and/or landslides; Hydro refers to floods and/or cyclones (Dilley, 2005).....	22
2.1	Relationships between different types of drought and duration of drought events. Based on Figure 1.4 from page 10 of Drought: a Global Assessment (Wilhite, 2000).....	26
2.2	Map of available rain gauge stations which provide input to the GPCC data set. These stations have at least 90% data availability between 1951 and 2000. There are about 500 stations in Africa and 9343 stations worldwide (Courtesy of Beck et al., 2005).....	28
2.3	Map of the position of Kenya on the continent of Africa. Both maps of Africa and Kenya are from the Encyclopaedia Britannica (1998 and 1999 respectively).....	28
2.4	Topographical map of Kenya. The scale is metres above sea level (Courtesy of Richard Anyah, personal communication).....	29
2.5	Annual rainfall time series from a representative sample of rain gauge stations across Kenya averaged over the period 1959-2006. Rainfall data are from the Kenya Meteorological Department, the map of Kenya is from Online Maps (2009).....	31
2.6	Population density (a) and land use (b) maps of Kenya. Population data are from the 1989 population census (UN, 2006).....	32
2.7	The Climate Prediction Center's analysis of the relationships between El Niño (warm episode) and La Niña (cold episode) and global temperature and rainfall patterns (CPC, 2007). The boxed regions highlight the Kenya/Uganda region.....	38

2.8	The Indian Ocean Dipole (IOD) and its associated negative (blue) and positive (red) SST anomalies. White indicates areas of increased convective activity. Low-level wind direction is indicated by the yellow arrows. Significant anomalies appear around May-June, intensify through July to September and peak in October, followed by a rapid demise (Saji et al., 1999).....	40
2.9	African OND rainfall anomaly linked to the IOD independent of ENSO (panels (a) to (c)) and linked to ENSO independent of the IOD (panels (d) to (f)) for 1959-2000. Panel (c) shows the difference in OND rainfall anomaly between those subset years when the OND ENSO is neutral and the OND IOD is (a) above median and (b) below median. Panel (f) shows the difference in OND rainfall anomaly between those subset years when the OND IOD is neutral and the OND ENSO is (d) above median and (e) below median.....	43
2.10	(a) OND 2008 rainfall forecast, issued September 2008 (ICPAC, 2008). Shows probability distributions that indicate above-, near-, or below-normal rainfall for each zone. Green represents near-normal to above-normal rainfall; yellow represents near-normal to below-normal rainfall. (b) OND 2008 rainfall forecast, adapted for Kenya, issued September 2008 (KMD, 2008). Shows rainfall forecasts of near-normal to above-normal rainfall (green) and near-normal to below-normal rainfall (yellow). (c) OND 2008 rainfall performance, issued March 2009 (KMD, 2009). Shows the observed rainfall performance relative to a long-term climatology.....	48
3.1	Map of Kenya showing the rain gauge stations selected for use in this study. The 16 Synoptic rain gauge stations are marked with an asterisk (*) and 6 non-synoptic rain gauge stations are marked by a square (□).....	56
4.1	Seasonal rainfall hindcast skill for ASO at lead 0 from the three individual DEMETER models and the multi-model ensemble. The skill measure is Rank Correlation (r_{rank}) and the verification data are from the GPCC rainfall data set. The panels show r_{rank} for (a) UKMO hindcasts (1959-2000), (b) Météo France hindcasts (1958-2000), (c) ECMWF hindcasts (1958-2000) and (d) Multi-model ensemble hindcasts (1959-2000). White lines denote areas with p -values < 0.05	69

- 4.2 DEMETER multi-model ensemble seasonal rainfall hindcast skill for ASO and NDJ at lead 0; based on *MSSS* (%) and percentage of correct anomaly sign measures over the period 1959-2000. The white lines in panels (a) and (c) denote areas with p -values < 0.05 . The white lines in panels (b) and (d) denote areas where the two-tailed probability of obtaining the displayed percentage by random chance is < 0.05 70
- 4.3 Seasonal rainfall hindcast skill for NDJ seasonal rainfall at lead 0 from the three individual DEMETER models and the multi-model ensemble. The skill measure is Rank correlation (r_{rank}) and the verification data are from the GPCP rainfall data set. The panels show r_{rank} for (a) UKMO hindcasts 1959-2000, (b) Météo France hindcasts 1958-2000, (c) ECMWF hindcasts 1958-2000 and (d) Multi-model ensemble hindcasts 1959-2000. White lines denote areas with p -values < 0.05 72
- 4.4 DEMETER seasonal rainfall hindcast skill for the Sahel (ASO) and East Africa (NDJ) at leads of 0 and 3 months. *MSSS* (%) is the skill measure and the verification data are from the GPCP rainfall data set. Hindcast skill is shown for the three individual DEMETER models and for three multi-models. The latter are distinguished by different weightings as follows: (1) weighted using *MSSS* with negative *MSSS* not contributing, (2) a straight average using no weighting and (3) weighted using *MSE* (this is the weighting used in all other figures and results). Error bars show the 95% confidence intervals on skill..... 74
- 4.5 Investigation of the source of conflicting seasonal rainfall hindcast skill values for NDJ over the period 1959-2000 for a grid square in southeast Chad (centred at 17.5°E, 10°N). Panel (a) shows that the squared difference between the DEMETER multi-model and GPCP anomalies is large in a few years causing *MSSS* to be negative. Panel (b) shows the hindcast anomalies from the individual DEMETER models..... 76
- 4.6 Seasonal rainfall hindcast skill for a) MAM, b) JJA, c) SON and d) DJF at lead 1 from the DEMETER multi-model ensemble over the period 1979-2001. The skill measure is Mean Square Skill Score (%), available from the DEMETER website (<http://www.ecmwf.int/research/demeter/d/charts/verification>)..... 78
- 4.7 Investigation of the source of changing seasonal rainfall hindcast skill with each season from the DEMETER multi-model ensemble. The lead 0

	skill measure shown is Rank correlation (r_{rank}) and the verification data are from the GPCC rainfall data set. The panels show scatter plots for seasonal rainfall for a grid square (centred at 37.5°E, 7.5°S) near Dar es Salaam, Tanzania over the period 1959-2000 for: (a) FMA (b) MJJ (c) ASO and (d) NDJ seasonal averages.....	79
4.8	(a) Map of available rain gauge stations over Africa which provide input to the GPCC verification data set. These stations have at least 90% data availability between 1951 and 2000 (Courtesy of Beck et al., 2005). (b) Rank correlation for ASO seasonal rainfall at lead 0 from the DEMETER multi-model ensemble hindcasts 1959-2000 as shown in Figure 4.1(d). (c) Rank correlation for NDJ seasonal rainfall at lead 0 from the DEMETER multi-model ensemble hindcasts 1959-2000 as shown in Figure 4.3(d). White lines denote areas with p -values < 0.05.....	81
5.1	Part 1: Monthly rainfall climatology (averaged over the period 1959-2006) for each homogeneous rainfall region: (a) South, (b) Southeast and (c) Southwest.....	84
5.1	Part 2: Monthly rainfall climatology (averaged over the period 1959-2006) for each homogeneous rainfall region: (d) West, (e) Northwest and (f) Northeast.....	85
5.2	Homogeneous rainfall regions over Kenya from Indeje (2000) and Indeje et al. (2000). Rainfall was delineated into regions through EOF and correlation analysis using annual (black dotted lines) and seasonal (purple dotted lines) rainfall over the period 1961-1990.....	87
5.3	The black lines show the boundaries of the homogeneous rainfall regions, over Kenya, used in this study. The purple dotted lines show the regional boundaries from Indeje's seasonal rainfall analysis shown in Figure 5.2b (Indeje, 2000 and Indeje et al., 2000). Map of Kenya from Online Maps (2009).....	88
5.4	Example spatial correlation plots: Southwest regional OND seasonal rainfall index correlated with prior August-September (AS) two-month average SST global field over the periods: (a) 1959-1974 and (b) 1975-1990. White areas denote land mass. Red (blue) shows positive (negative) correlations, grey shows low correlations (between -0.2 and 0.2) and the black lines show areas with p -values < 0.15. The AS	

	potential predictor region for this example is highlighted by the purple box.....	93
5.5	OND rainfall index from the Southwest region over the period 1959-1974. (a) Histogram of the rainfall index to highlight the non-normal distribution. (b) Quantile-Quantile (QQ) plot comparing the observed rainfall index to the theoretical normal distribution. The black diagonal line in (b) indicates the theoretical position of the points if the data were normally distributed. The black dashed arrows highlight the high rainfall years that are responsible for the large departure from normality.....	97
5.6	'@Risk' plot showing the gamma distribution (red) that best represents the OND seasonal rainfall index from the Southwest region over the years 1959-1974. The green line shows the normal distribution for comparison..	99
5.7	Transformed (using $\log(x+1)$) OND seasonal rainfall index from the Southwest region over the years 1959-1974. (a) '@Risk' plot showing the Log Logistic (red) and Normal (green) distributions that best represents the transformed OND seasonal rainfall index from the Southwest region over the years 1959-1974. (b) QQ-plot comparing the transformed rainfall index to the theoretical normal distribution.....	101
5.8	Schematic of the fundamentals of least-squares linear regression. Here, x represents the predictor variable, y represents the predictand and \hat{y} specifies a predicted value of y . The regression line ($\hat{y} = c + mx$) minimises the sum of the squared residuals, which are the vertical differences between the points and the regression line $e_i = y_i - \hat{y}(x_i)$	102
6.1	Scatter plots of hindcast vs. observed OND rainfall (mm/season) over the 1991-2006 independent verification period for each region of Kenya: (a) Southeast, (b) northwest, (c) southwest, (d) northeast, (e) south and (f) west. The lead 0 skill measure shown is Rank correlation (r_{rank}) and the observed data are threshold corrected regionally averaged rain gauge data from the KMD.....	113
6.2	The most skilful predictor regions for OND rainfall in West Kenya. Correlation predictor selection plots for these regions are shown in Appendix B.1.....	115

6.3	The most skilful predictor regions for OND rainfall in Southwest Kenya. Correlation predictor selection plots for these regions are shown in Appendix B.2.....	117
6.4	The most skilful predictor regions for OND rainfall in South Kenya. Correlation predictor selection plots for the SST Sep predictor region are shown in Appendix B.3. Composite difference predictor selection plots for the <i>u</i> -wind Aug predictor region are shown in Appendix B.4.....	118
6.5	The most skilful predictor regions for OND rainfall in Northeast Kenya. Composite difference predictor selection plots for this South Kenya predictor region are shown in Appendix B.4.....	120
6.6	The most skilful predictor regions for OND rainfall in Northwest Kenya. Correlation predictor selection plots for this region are shown in Appendix B.5.....	123
6.7	The most skilful predictor regions for OND rainfall in Southeast Kenya. Correlation predictor selection plots for this region are shown in Appendix B.6.....	124
6.8	The most skilful predictor regions for the OND rainfall in (a) Southeast-east and (b) Southeast-west Kenya. Correlation predictor selection plots for these predictor regions are shown in Appendix B.7 and B.8 respectively.....	127
7.1	(a) A west to east topographical cross-section of Kenya taken along the line linking A and B shown in (b). Figures are adapted from Figures 1 and 4 in Hills (1979).....	132
7.2	Composite difference plot showing direction, magnitude and significance of the difference in 950hPa wind anomalies for those subset years in Group 1 and Group 2. Group 1 comprises the 4 years (over the period 1954-1990) with the highest number of days with Tropical Cyclone or Tropical Depression occurrence in the region 30-80°E and 5-35°S. Group 2 comprises the 4 years over the same period with the least occurrence. Shaded areas show <i>p</i> -values < 0.05 in either the <i>u</i> - or <i>v</i> -component of the wind, according to the two-sided <i>t</i> -test. Figure from Shanko and Camberlin (1998).....	132
7.3	March (a) 700hPa and (b) 400hPa and May (c) 700hPa and (d) 400hPa air transport climatologies over Africa (1971-1975) to and from Kenya. Red solid arrows indicate the dominant wind direction. Red dashed arrows	

	indicate the second most dominant wind direction. This Figure is adapted from Figures 5 (a) and (d) and 6 (a) and (d) in Gatebe et al. (1999).....	138
7.4	The best April predictor region for the Southeast region of Kenya. Correlation predictor selection plots for these regions are shown in Appendix C.1.....	140
7.5	The best April predictor region for the West region of Kenya. Correlation predictor selection plots for these regions are shown in Appendix C.2.....	140
8.1	Map of Kenya showing the rain gauge stations selected for use in the study, the homogeneous rainfall region boundaries (purple lines) and the EUROSIP and CMAP grid square boundaries (red lines). The 16 Synoptic rain gauge stations are marked with an asterisk (*) and 6 non-synoptic rain gauge stations are marked by a square (□).....	148

Chapter 1

Introduction

1.1 Motivation and objectives of the thesis

Africa's climate is prone to extended rainfall deficits. In extreme cases these may lead to droughts, economic hardship and humanitarian disasters. Between 1970 and 2006 over 100 million people in Africa were affected by drought, requiring immediate humanitarian assistance (EM-DAT, 2009). Droughts were also responsible for the death of 97% of the persons killed by natural disasters over this period in Africa (Figure 1.1). The impacts of drought are felt in many economic and social sectors including agriculture and food security, livestock development, hydro-electricity production, transport, fire management, water resource management, health and public safety (Heim, 2002; Mukabana, 2008). When any one of these sectors becomes damaged, the effect can spread quickly and a whole country may suffer. Skilful prediction of seasonal rainfall would bring sound humanitarian and economic benefit to the many African countries that depend on rain-fed agriculture. It would enable timely actions to be taken by aid agencies and governments to avert or minimize potential hunger, destitution and famine resulting from drought (Verdin et al. 2005).

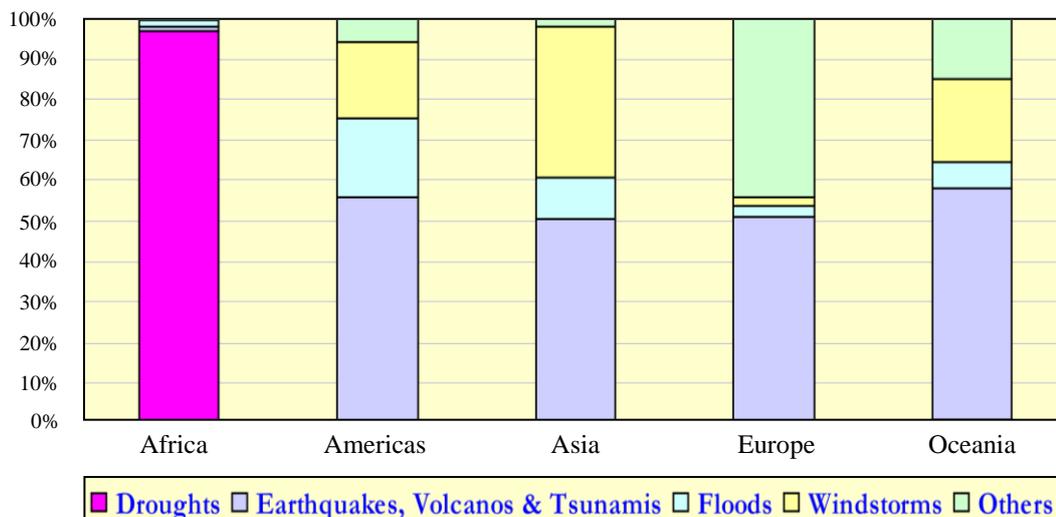


Figure 1.1. The proportion of persons killed by droughts between 1970 and 2006 by continent compared to other natural disasters. (Courtesy of the Centre for Research on the Epidemiology of Disasters (EM-DAT, 2006)).

East Africa is unique as it is the only area on the continent to experience a low annual average rainfall total as well as a high standard deviation of annual rainfall (Figure 1.2). This indicates high climatic variability, leaving East Africa particularly susceptible to climatic extremes such as droughts. Currently, statistical seasonal rainfall forecasts for East Africa are developed for each rainy season by the local Meteorological Agencies (Buizer et al., 2000; Likumana, 2008; Muita, 2008; Ogallo, 2008). Forecasters also refer to outputs from several dynamical seasonal rainfall forecast models, prior to issuing their forecasts. However, the seasonal rainfall forecasting skills of these statistical and dynamical models have not been thoroughly assessed over East Africa. It is therefore uncertain how much confidence one can have in their forecasts.

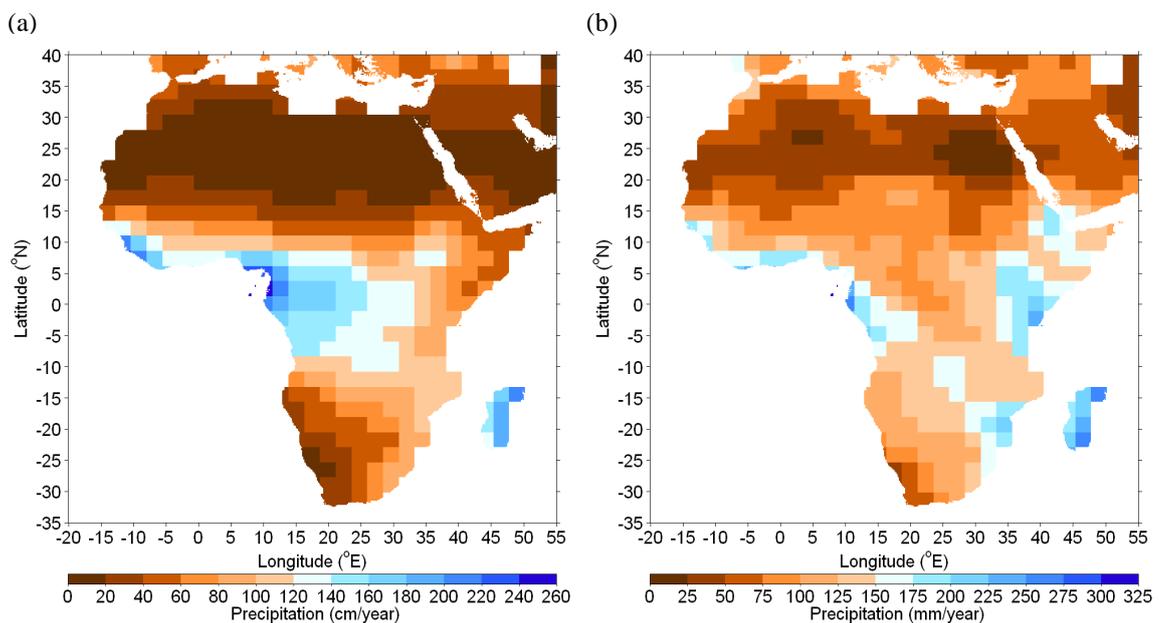


Figure 1.2. Annual GPCC rainfall data (section 3.1.1) for Africa between 1958 and 2000. (a) Annual average rainfall totals (0-260 cm/year). (b) Standard deviation of annual rainfall assuming a Gaussian distribution (0-325mm/year). Note the two plots have different scales.

This study focuses on Kenya, in East Africa. Kenya has experienced 7 severe droughts over the period 1991-2008, affecting over 35 million people who required immediate humanitarian assistance (Table 1.1). Figure 1.3 shows the distribution of the world's highest mortality risk disaster hotspots, by hazard type, across Africa. The location of Kenya is shown by the red box. It can be seen that west and southwest Kenya, which are home to approximately 70% of the country's population (CIESIN, 2005), are amongst the top 30% regions of the world with the highest mortality risk due to droughts and floods.

Years	Country	Affected (nearest 0.1mn)	Years	Country	Affected (nearest 0.1mn)
1989-92	Angola	1.9	1980-85		1.5
1983-85	Benin	2.1	2001-03	Mauritania	1.0
1982-87	Botswana	1.0	1980-85		1.6
1990	Burkina Faso	2.6	2005	Mozambique	1.4
1980-1985		1.3	1991-92		3.3
2005-06	Burundi	2.2	1981-85		4.8
1999-01		1.0	1979-82	6.0	
2008	Eritrea	1.7	2005	Niger	3.6
1999-03		2.3	2001-02		3.6
1993		1.6	1990-91		1.6
2008	Ethiopia	6.4	1988		1.0
2003-04		12.6	1980-85		3.5
1997		1.0	1983-85	Nigeria	3.0
1989-94		6.5	1989	Rwanda	0.1
1987		7.0	1982-85	Senegal	1.2
1983-84		7.8	2008	Somalia	3.3
1983-85	Ghana	12.5	2000-01		1.2
2008	Kenya	1.4	1987		0.5
2005-06		3.5	2004	South Africa	15.0
2004		2.3	1988		1.3
1999-02		23.0	2000-01	Sudan	2.0
1997-98		1.6	1991-92		8.6
1994-95		1.2	1987		3.5
1991-92		2.7	1983-85		8.4
2002-04	Madagascar	1.0	2003-04	Tanzania	1.9
1988-92		1.0	1996-00		3.0
1981		1.0	1984		1.9
2005-06	Malawi	4.5	1999-01	Uganda	0.7
2002		2.8	2005	Zambia	1.2
1992-95		7.0	1995		1.3
1990		2.8	1991-92		1.7
1987-88		1.4	2001-03	Zimbabwe	6.0
2005	Mali	1.0	1991-95		5.0

Table 1.1. Major African droughts over the period 1980-2008 as reported on the Emergency Events Database (EM-DAT, 2009), including the numbers of persons affected (to the nearest million). A person is affected if they require immediate humanitarian assistance including: food, water, shelter, sanitation and immediate medical assistance for their basic survival during an emergency (EM-DAT, 2004). In order to present only the major Africa droughts, this table includes only those with reports of over 1 million people affected or over 100 deaths according to EM-DAT (2009).

The data shown in this table should be regarded as indicative and not absolute. They are collected by EM-DAT from sources including: Governments, insurance companies, press agencies and aid agencies and are retrospectively analysed and cross-referenced. However, there are several problems to consider when using the data: the original information was not necessarily gathered for statistical purposes; collection methodologies and definitions are not standardised across all sources; words such as “affected”, although defined as above by EM-DAT, can be open to interpretation (EM-DAT, 2004).

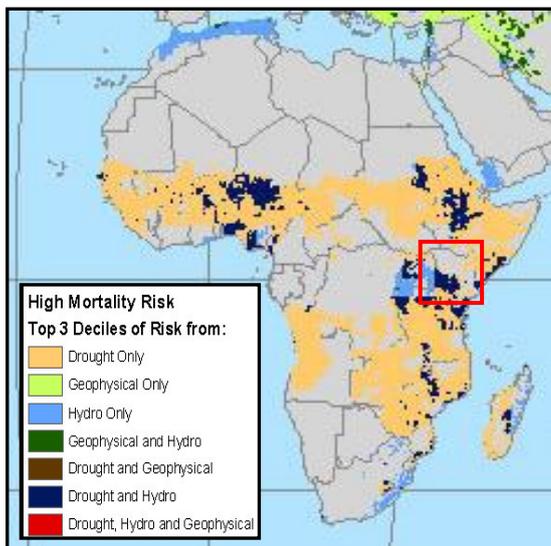


Figure 1.3. Distribution of the highest 30% mortality risk natural disaster hotspots. Geophysical refers to earthquakes, volcanoes and/or landslides; Hydro refers to floods and/or cyclones (Dilley, 2005). The location of Kenya is shown by the red box.

Rain-fed agriculture is the backbone of Kenya's economy. Over 80% of the 38 million population (UN, 2007) are involved in agriculture (Brass and Jolly, 1993; Bowden, 2007), which contributes 24% of the Gross Domestic Product (GDP) (Mwagore, 2002). Agriculture is arguably the most vulnerable sector to drought, with crops affected badly if the drought occurs during the growing season. Any damage to the agricultural sector leaves the country exposed to hunger, famine and increase in disease incidence.

A recent example of a severe drought affecting Kenya was the humanitarian crisis that developed over the period 1999-2002 (EM-DAT, 2009). Over 23 million people were affected by drought in western and central Kenya, the agricultural heartland of the country (Table 1.1). Electricity was rationed due to decreases in hydropower production (CERF, 2008; CIA, 2009) and many crops were damaged by a lack of irrigation. This affected the tea and coffee industry, causing tea production to reduce by 15% over the period 1999-2002 (Rice, 2006). This prolonged drought is estimated to have cost the Kenyan economy around 2.5 billion US dollars (CERF, 2008), which was approximately 20% of Kenya's GDP at the time (IMF, 2008).

The main aims of this thesis are:

- To perform a detailed assessment of the current seasonal rainfall hindcast skill of leading dynamical models over Africa. A *hindcast* is a retrospective forecast.
- To produce skilful statistical hindcast models for seasonal rainfall in Kenya.
- To improve the understanding of which predictors have strong, temporally stable links to Kenyan rainfall and the mechanisms responsible for their influences.
- To compare statistical and dynamical models over Kenya in order to determine which produces the most skilful seasonal rainfall hindcasts.

1.2 Thesis outline

The thesis will begin by providing background information on Africa and its climate. Chapter 2 then progresses to examine the Kenyan climate and the impacts of drought on this country. A literature review follows on the current state of knowledge of Kenyan rainfall variability, predictability and teleconnections between rainfall and global Sea Surface Temperatures (SSTs). The operational seasonal forecasting methods used in Kenya are then discussed. In Chapter 3, the data, quality control techniques and detrending methods employed in this study are described. The results of an assessment of the current dynamical seasonal rainfall hindcast skill over Africa, available from the DEMETER multi-model ensemble hindcast system, are presented in Chapter 4.

Chapter 5 describes the methodology behind the development of statistical seasonal rainfall hindcast models. Monthly rain gauge data are clustered into homogeneous rainfall regions to create new regional, seasonal rainfall indices. The predictor selection criteria are then outlined. These consist of analysis of correlations between Kenyan seasonal rainfall and predictors, their statistical significance and the stability of their relationships over time. The methods used to develop the statistical seasonal rainfall hindcast models and to assess their hindcast skills are presented.

Chapters 6 and 7 present the statistical seasonal rainfall hindcast models for the Kenyan ‘short- and long-rains’ seasons. Both chapters begin with the current state of knowledge of the physical mechanisms for each season. Chapter 6 examines the predictors selected for the ‘short-rains’ season and their physical links to Kenyan rainfall. It then provides the results of a detailed skill assessment of the statistical seasonal rainfall hindcast models over each region of Kenya. An investigation into the predictors for the ‘long-rains’ season is presented in Chapter 7. Issues that arise from this investigation are then discussed.

A comparison of statistical and dynamical OND hindcast models is made in Chapter 8. The seasonal rainfall hindcast skill available from the EUROSIP multi-model ensemble system is assessed over Kenya. A discussion on which type of model produces the most skilful hindcasts for each region of Kenya is then presented. Finally, conclusions are presented in Chapter 9, along with a discussion on the wider implications of this study and suggestions for future directions.

Chapter 2

Background and Literature Review

2.1 Africa and its climate

Africa is the world's second largest continent after Asia. It covers over 30 million km² and has over 50 countries. Africa spans the equator from 37°N to 35°S and comprises a range of extreme climatic zones from tropical rainforests to arid deserts. Most of Africa lies within the tropics and subtropics where temperatures are high throughout the year and where the diurnal temperature range (~10-15°C) exceeds the interannual temperature range (~6-10°C), except in the deserts (Nicholson, 2001). The mean annual rainfall distribution varies from less than 1mm/yr in the Sahara to over 2m/year in the tropical rainforests of the Guinea coast and Congo (Figure 1.2 (a)).

The climate of Africa is dominated by planetary scale features (Glantz, 1988) such as the meridional overturning of the Hadley Circulation, the influences of the Atlantic and Indian Ocean monsoons, the Inter-Tropical Convergence Zone (ITCZ) and teleconnections with sea-surface temperature (SST) and pressure features such as the El-Niño Southern Oscillation. The finer features of the climates of individual countries are determined by localised factors such as topography and the presence of large bodies of water. The latitude of each country is a major factor in the distribution of the rainy seasons across the year, which is linked to the passage of the sun in the Tropics.

2.1.1 African droughts

Africa has a long history of droughts of varying lengths and intensities (Gommes and Petrassi, 1996). Table 1.1 shows that severe droughts have affected over 200 million people across 25 African countries during the period 1980-2008, as reported in the Emergency Events Database (EM-DAT, 2009).

When considering African droughts it is important to first establish a definition of drought. There is considerable disagreement in the literature on this subject. A simple definition is provided by Gibbs (2000) as “a shortage of water to meet essential needs”, although this depends greatly on how water is used in the region. The absence of a precise and universally accepted definition can lead to confusion over determining the onset and cessation of a drought (Wilhite, 2000). The effect of a drought depends on the intensity, duration and spatial extent of the dryness and the vulnerability of the population. Vulnerability depends on the accuracy and availability of any predictions, monitoring and early warning systems, mitigation and the preparedness of the population (Wilhite, 2000).

Figure 2.1 shows the different types of drought that can occur and the impacts over time that these may bring. The four generally accepted forms of drought shown in this figure are: meteorological, agricultural, hydrological and socio-economic (Heim, 2002).

A meteorological drought is defined as an absence or a large reduction in rainfall over an extended period of time relative to the regional climatic norm. There can be no other form of drought without first experiencing a meteorological drought; therefore this is the definition that will be used for the rest of this thesis.

An agricultural drought may follow a meteorological drought, with its lag time depending on the prior moisture of the soil (Heim, 2002). An agricultural drought is caused by short term dryness in the root layer of the soil. If the dry period occurs at a critical time in the growing season then the effect on the crops can be widespread and severe.

After a long period of dryness **a hydrological drought** may develop, when the effect reaches the surface and subsurface water supply. This reduces stream flow and lake levels, which are often the first visible signs of a drought taking place. Hydrological droughts generally persist for longer periods and can have a huge impact on the death toll of a drought in Africa, as the vast majority of the population depend on wells and natural water sources for their drinking water and for crop irrigation.

Finally, **a socio-economic drought** occurs when the situation has reached quite a severe scale and the effects are seen in the supply of economic goods.

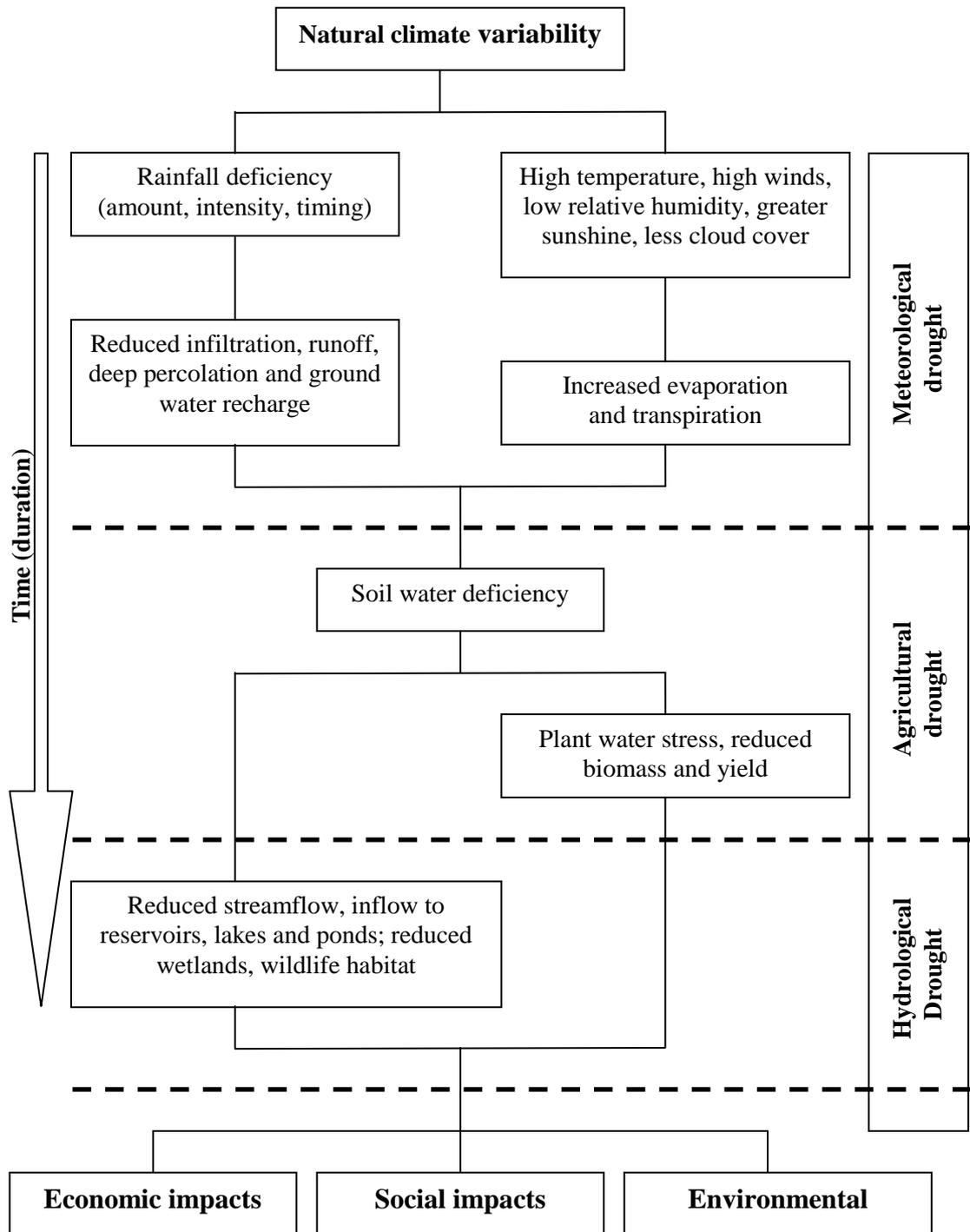


Figure 2.1. Relationships between different types of drought and duration of drought events. Based on Figure 1.4 from page 10 of *Drought: a Global Assessment* (Wilhite, 2000). Reproduced with kind permission of the publishers.

2.1.2 Selecting a suitable country to study

This study starts by examining the seasonal prediction of African rainfall in general and then progresses to examine the seasonal prediction of Kenyan rainfall by region. Why was Kenya chosen?

The first criterion to select a suitable country to study is the number of severe droughts that have affected the country over recent years. Table 1.1 shows that Kenya and Ethiopia have experienced more major droughts than any other African country over the period 1980-2008. Kenya and Ethiopia have experienced drought conditions in approximately 50% and 45% of the years in this period. They have also both had the highest number of people affected by drought during this period, with over 35 and 40 million people respectively. For comparison, this is over 10 million more people than the third highest affected country, Sudan, which experienced drought conditions for 24% of the same period.

The second selection criterion is that the country has an accurate, reliable and dense rain gauge network over an adequately long historical period. This is fundamental to the development and verification of reliable statistical forecast models for seasonal rainfall. Figure 2.2 shows the global availability of rain gauges with a minimum of 90% data availability during the period 1951 to 2000 (Beck et al. 2005). It can be seen that the rainfall data across Africa has a sparse and uneven distribution compared to the more developed areas of the world. In East Africa, Kenya has the highest density of rain gauges with 90% data availability.

Ethiopia and Kenya fulfil both selection criteria. There has been less research published on rainfall prediction and variability in Kenya compared to Ethiopia. Therefore, Kenya was selected as a suitable country to investigate in further detail for this study.

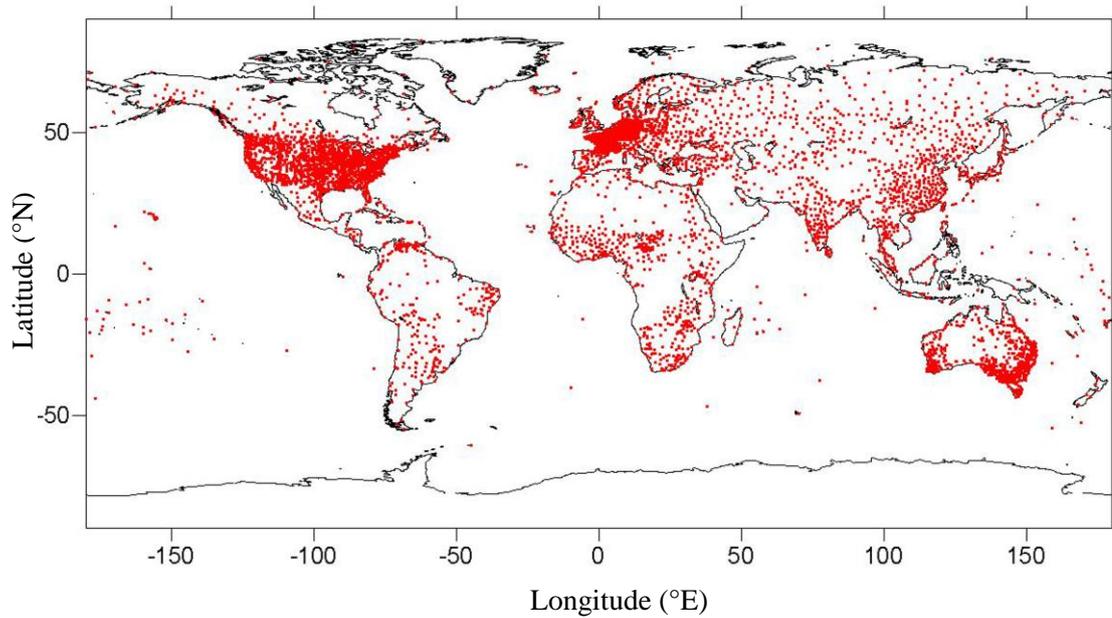


Figure 2.2. Map of available rain gauge stations which provide input to the GPCP data set (section 3.1.1). These stations have at least 90% data availability between 1951 and 2000. There are about 500 stations in Africa and 9343 stations worldwide (Courtesy of Beck et al., 2005).

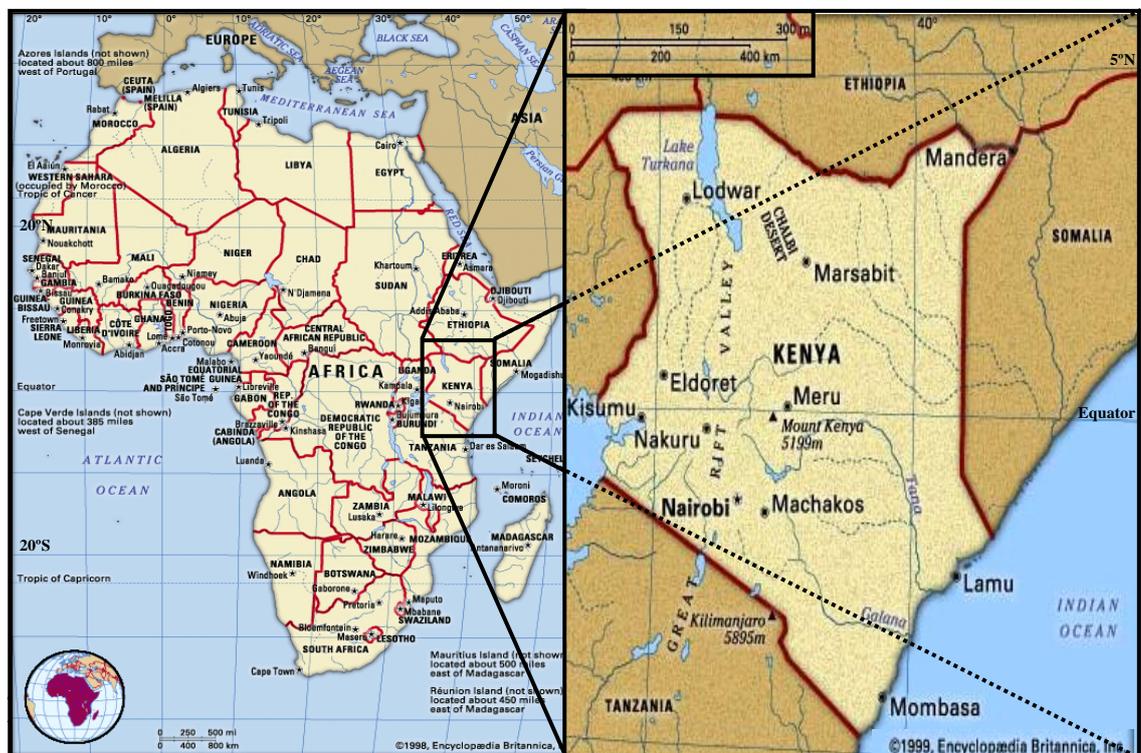


Figure 2.3. Map of the position of Kenya on the continent of Africa. Both maps of Africa and Kenya are from the Encyclopædia Britannica (1998 and 1999 respectively).

2.2 Kenya and its climate

Kenya is located in East Africa, straddling the equator from 5°S-5°N and 34°E-42°E (Figure 2.3). It has a land area of approximately 580,000 km² and a population of over 38 million (UN, 2007). Kenya experiences two main rainy seasons: the ‘long-rains’ in March-May (MAM) and the ‘short-rains’ in October-December (OND). This annual cycle is influenced by the movement of the Inter-Tropical Convergence Zone (ITCZ), which migrates between 15°S and 15°N between January and July respectively. The ITCZ is a surface convergence zone (Nicholson, 2008) of equator-ward moving airmasses from both hemispheres (Okoola, 1998, 1999). Its migration is governed by the overhead passage of the Sun heating the Earth. The Tropical Rain Belt is the zone of maximum cloudiness and rainfall within the ITCZ (Okoola, 1998; Nicholson, 2008). The monsoonal winds of the ITCZ are the major source of Kenya’s moisture flux. Inland, the winds are significantly modified by Kenya’s topography (Ogallo, 1988).

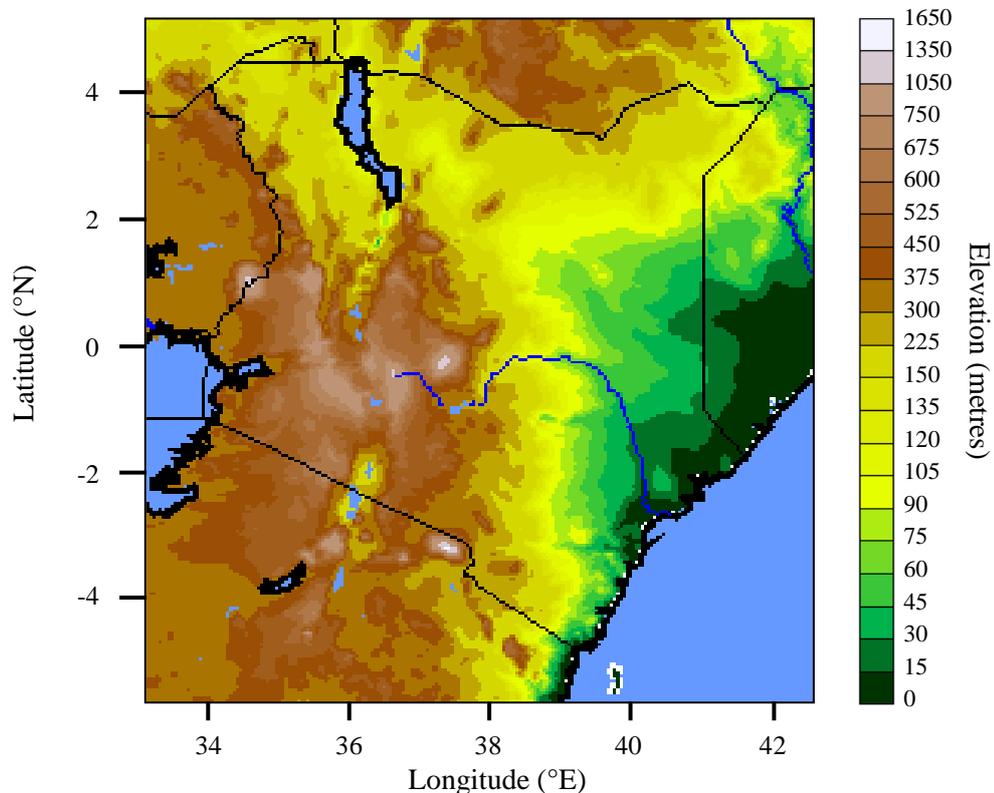


Figure 2.4. Topographical map of Kenya. The scale is metres above sea level (Courtesy of Richard Anyah, personal communication).

Kenya's diverse topography (Figure 2.4) contributes towards the high spatial variance in the seasonal rainfall distribution. Figure 2.5 shows the annual rainfall time series from a representative sample of rain gauge stations across Kenya over the period 1959-2006. The 'long- and short-rain' seasons can be seen at all stations. The stations with the highest annual rainfall totals are found along the Indian Ocean coast and in the western and central highland regions. The high rainfall totals in the coastal region have allowed a narrow band of fertile land to develop along the Indian Ocean coast, extending up to 100km inland. The western and central highland areas are generally the most fertile in Kenya and therefore the most populated and best cultivated (Bowden, 2007). The central highland area is comprised of two mountain belts, divided by the Great Rift Valley, which cuts across the landscape from the north to the south. The width of the Great Rift Valley ranges from 40km at its narrowest to 320km at its widest point (Bowden, 2007). Mount Kenya is the highest point of these highlands at 5,199m.

Kenya's climate is normally dry between the 'long- and short-rain' seasons. The exception to this is the high plateau region to the west of the Great Rift Valley, which has a third rainy season during July-August. This is due to the influx of a moist westerly airstream from the Atlantic Ocean and Congo/Zaire basin (Davies et al. 1985; Ogallo 1988). This third, localised rainy season can be seen in the time series from Kisii Met station in Figure 2.5. Moving towards the north of Kenya, the terrain turns to inhospitable deserts in the northwest and semi-arid conditions in the northeast (Blades, 2000). The driest and least populous region is the northwest, where the Lodwar rain gauge station receives <50mm/month at its peak (Figure 2.5).

2.2.1 Drought impacts in Kenya

The effects of drought on a country depend on three main conditions: the regularity of severe droughts, the country's dependence on rain-fed agriculture and the ability of the population to prepare for and adapt to drought conditions.

The first condition is the regularity of severe droughts. Table 1.1 shows that Kenya experienced severe drought conditions in 50% of the years between 1980 and 2008. Most years Kenya will experience either localised or more widespread droughts (Ogallo, 1989), which affect many sectors including: agriculture, livestock, forests, wildlife, tourism, water resources and hydroelectric power generation.

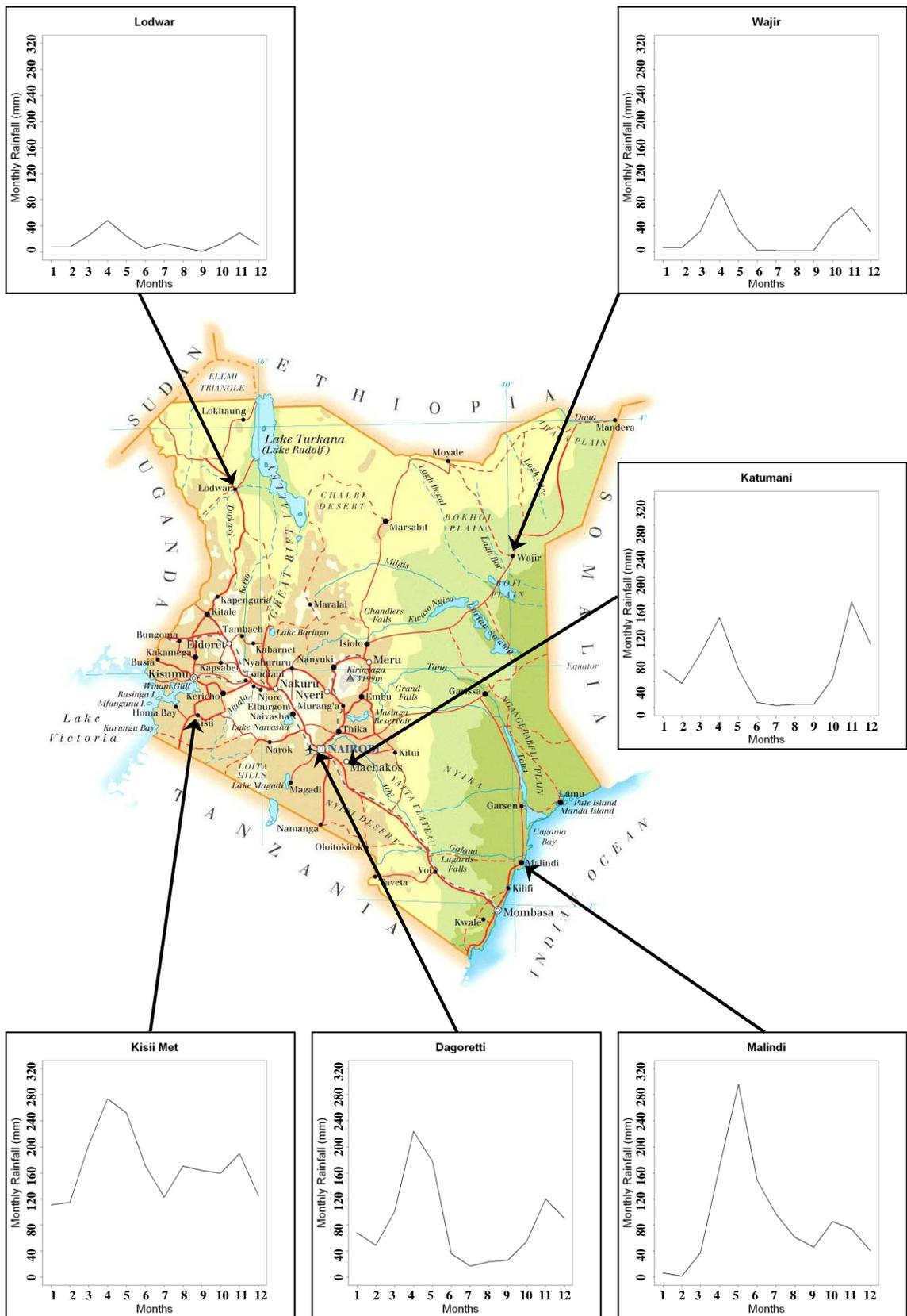


Figure 2.5. Annual rainfall time series from a representative sample of rain gauge stations across Kenya averaged over the period 1959-2006. Rainfall data are from the Kenya Meteorological Department, the map of Kenya is from Online Maps (2009).

The second condition is the country's dependence on rain-fed agriculture. For the case of Kenya, this was established in Chapter 1. The third condition is the ability of the population to prepare for and adapt to drought conditions. Kenya is a developing country with 46% of the rural population living below the poverty line, on less than 1 US \$ per day. 32% of the rural population are also below the food poverty line (Boken et al., 2005), which means they cannot afford enough food to provide the minimum nutritional amount to survive. The Kenyan Government and the population do not have the financial resources to adequately prepare for and adapt to drought conditions.

Therefore, as a result of the above three conditions being met, the impacts of drought are felt very heavily in Kenya. The effects of regional droughts depend on: the population of the region affected and the land use types. Figure 2.6 shows the (a) population density and (b) land use types across Kenya. In areas of greatest population density and agricultural production, which generally are coincident, a drought would cause maximum damage. It can be seen that these areas are located in western and central Kenya, around the rift valley and highlands and in a narrow strip along the Indian Ocean coast.

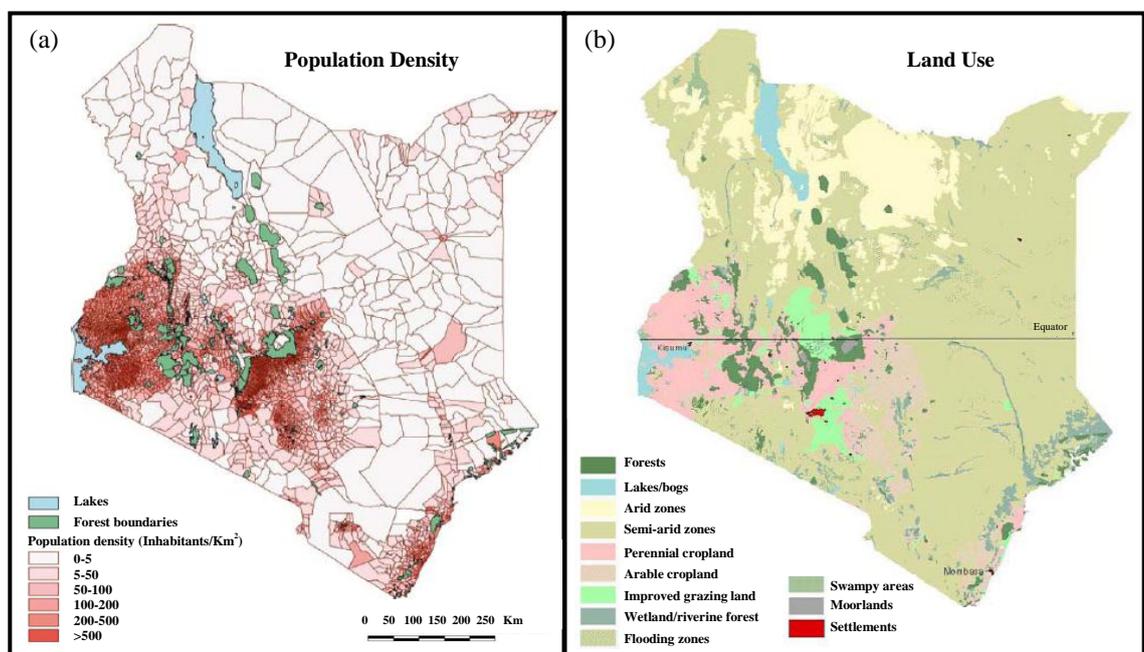


Figure 2.6. Population density (a) and land use (b) maps of Kenya. Population data are from the 1989 population census (UN, 2006).

2.2.2 Kenyan drought case studies

1999-2002: drought in western and central Kenya

A severe drought affected over 37 million people across 9 countries in Africa between 1999 and 2002 (Table 1.1). This drought may have been forced by the La Niña (Chapter 2.3.1) that persisted in the Pacific Ocean over the period 1999-2001. La Niñas are generally linked to reduced rainfall over East Africa in the boreal winter.

The humanitarian effects of this drought were particularly devastating in western and central Kenya. These are the most heavily populated and cultivated regions in the country. Humanitarian assistance was necessary with over 23 million Kenyans being affected (Table 1.1). This was the highest number of people reported to have been affected by drought in Africa for over 100 years (EM-DAT, 2009). At the peak of the drought, in 2000, malnutrition had affected 40% of the population and 3.2 million Kenyans were dependent on food aid (CERF, 2008). The lack of food and water caused severe conflicts between pastoral communities in central Kenya (IRI, 2005).

It is estimated that this prolonged drought cost the Kenyan economy around 2.5 billion US \$ (CERF, 2008). This was approximately 20% of Kenya's GDP at the time (IMF, 2008). The tea and coffee industries, which are the leading contributors to Kenya's GDP, were damaged during this drought as the leaves withered in the dry heat. Tea production fell by 15% between 1999 and 2002 (Rice, 2006). Further damage to Kenya's economy was caused by hydroelectric power rationing (Mutimba, 2005; ISS, 2009), which was implemented throughout the period 1999-2001 (CIA, 2009). This caused a reduction in agricultural output as the amount of available electricity fell by 25% in 2000 (CERF 2008).

2005-2006: drought in northern and eastern Kenya

In 2005-2006 the northern and eastern regions of Kenya were affected by severe drought conditions (CERF, 2009). This drought was initiated by the failure of the 'short-rains' from October to December in 2005.

The northern and eastern regions are the least populous in Kenya, as the land is mainly arid or semi-arid. A drought in these regions, therefore, has a far smaller humanitarian impact than in the western and central regions. Humanitarian assistance was required by 3.5 million people (Table 1.1). The worst affected communities were the nomadic pastoralists in the northeast of Kenya. In some areas of the northeast 70% of the livestock died (CERF, 2009). The London Times (Rice, 2006) reported that this drought also had an adverse affect on the tea industry. The Chief of the Tea Board of Kenya, Sicily Kariuki, reported that black-tea production fell by 20% in the first half of 2006 compared to the same period in 2005 (News 24, 2006).

The Central Emergency Response Fund (CERF) was established by the United Nations in March 2006 to ensure that funds from major international aid organisations (such as the World Health Organization, the World Food Programme and United Nations funding) go to where they are most needed. Between March-December 2006 CERF provided aid costing over 27 million US \$ to Kenya's worst affected sectors including: food, health, agriculture, water and sanitation (CERF, 2009). This aid directly helped over 3 million people (Table 1.1).

2.3 Current state of knowledge of Kenyan rainfall variability and predictability

Although there has been much research into the variability and predictability of monthly and seasonal East African rainfall over the past 25 years (Davies 1985; Nicholson and Entekhabi 1986; Ropelewski and Halpert 1987, 1989; Ogallo et al. 1988; Ogallo 1989; Hastenrath et al. 1993, Gommès and Petrassi 1996; Nicholson and Kim 1997; 2004; Okoola 1999; Saji et al. 1999; Sun et al 1999a, 1999b; Webster et al. 1999; Hastenrath 2000, 2007; Indeje et al. 2000; Nicholson 2000, 2001; Nicholson and Selato 2000; Camberlin et al. 2001; Neng et al. 2002; Black 2003, 2005; Clark et al. 2003; Manpande and Reason 2005; Marchant et al. 2006; McHugh 2006; Zablone and Ogallo 2008), the scientific community is currently unable to consistently issue accurate forecasts of seasonal rainfall for East Africa.

Several publications present statistical models that have been developed to provide *large-scale* seasonal rainfall hindcasts and forecasts at zero-month lead for the whole of

East Africa (Mutai et al., 1998; Mutai and Ward, 2000; Philippon et al., 2002; Colman and Graham, 2005). Mutai et al. (1998) developed statistical models to forecast the large-scale OND rainfall anomaly over East Africa (15°N-15°S and 30°E-41.25°E). Their most skilful forecast for East Africa has a correlation skill value of 0.6 (p -value < 0.05) over an independent verification period of 1981-1994. They note that further work is necessary to enhance the utility of the forecasts and to raise the skill of the forecasts over smaller spatial scales. They suggested that this could be done by training models for specific sub regions of East Africa (Mutai et al. 1998; Mutai and Ward 2000). Philippon et al. (2002) have also developed a skilful hindcast model for OND rainfall over East Africa (4°N-4°S and 30°E-39°E). The hindcast has a cross-validated correlation skill value of 0.8 (p -value < 0.05) over the period 1968-1997. Forecast skill results are not shown over an independent verification period. The Met Office started to issue experimental statistical forecasts for East African (5°N-15°S and 30°E- the Indian Ocean Coast) OND rainfall in 1994 (Colman and Graham, 2005). The Met Office hindcasts have a cross-validated correlation skill value of > 0.5 (p -value < 0.05) over the period 1948-1997.

Statistical hindcast models have also been developed on a smaller spatial scale. These provide *regional* statistical seasonal rainfall hindcast models for the East African coast, Kenya and Ethiopia. Research by Hastenrath et al. (2004) found that correlation was not robust over time for their regional hindcast model, developed for the East African coast. Although correlation values of up to 0.74 were found for OND hindcasts over the period 1958-1977, this dropped to correlation values of <0.32 over the period 1978-1996. A similar result was found by Farmer (1988) when developing a statistical hindcast model for the September-December rainy season in Kenya. Over the period 1901-1942 the hindcast provided only an equivalent skill to using climatology with a correlation of 0.33, whereas over the period 1943-1984 the correlation value increased to 0.52. However, these results may be questionable due to the poor quality of the Kenyan rainfall data in first half of the 20th Century.

There has been more research published about Ethiopian seasonal rainfall prediction compared to that published about Kenya. Skilful regional statistical hindcast models have been developed for the Ethiopian June-September rainy season by Diro et al. (2008), based on the work of Gissila et al. (2004). Diro et al. (2008) divided Ethiopia into homogeneous rainfall regions and developed probabilistic seasonal rainfall hindcast

models for each region. They found that their hindcasts were more skilful than climatology or a random hindcast in each homogeneous rainfall region. Korecha and Barnston (2007) have also developed skilful hindcast and forecast models for the Ethiopian June-September rainy season. Correlation values of 0.64 and 0.51 were found for hindcasts over the periods 1970-1996 and 1997-2004. This latter independent verification period is acknowledged by the authors to be too short to strongly demonstrate that the forecast skill is robust.

2.3.1 Teleconnections between Sea Surface Temperatures (SSTs) and rainfall

A good understanding of the atmospheric dynamics through which seasonal rainfall in East Africa can be forced by large scale coupled ocean-atmosphere systems should accompany the development of statistical forecast models. This knowledge will assist the statistical relationships between rainfall and SST anomalies and their associated pressure and wind anomalies being used with confidence to predict seasonal rainfall (Camberlin et al., 2001). The following sub-sections will outline the teleconnections that have been reported in the literature between SSTs and Kenyan and East African rainfall.

2.3.2 Pacific Ocean: El Niño Southern Oscillation (ENSO)

The El Niño Southern Oscillation (ENSO) is a global climate phenomenon caused by ocean-atmosphere interactions that occur mainly in the tropical-subtropical Pacific and Indian Ocean basins (Diaz and Markgraf, 2000; Chang and Zebiak, 2003). ENSO is an irregular phenomenon that tends to reoccur every 2-7 years (Diaz and Markgraf, 2000) alternating between its two extremes: El Niño and La Niña. An El Niño (La Niña) occurs when warm (cool) SST anomalies and their associated decreased (increased) sea level atmospheric pressure anomalies are observed in the central and/or eastern equatorial Pacific Ocean (Glantz, 2001). An El Niño (La Niña) is defined by Trenberth (1997) to occur if the 5-month running mean of SST anomalies in the Niño 3.4 region (5°N-5°S, 120°-170°W) exceeds (is less than) 0.4°C (-0.4°C) for 6 months or more. Most ENSO events begin in the northern hemisphere spring or summer and reach peak

intensity between November and January. An ‘average’ event tends to last for 18-24 months although all events vary in magnitude, spatial extent, onset, duration and cessation (Diaz and Markgraf, 2000).

The term ‘El Niño’ has been used by Peruvians since before 1892 to describe the current of warm water that moves southwards along their coast every few years (Glantz, 2001). Walker (1924) was the first to document the term ‘Southern-Oscillation’, which he referred to as a ‘seesaw’ of atmospheric pressure between the Pacific and Indian Oceans. Berlage published a paper in 1957 linking the Southern Oscillation to El Niño. Bjerknes (1966 and 1969) was the first to explain the link between the Southern Oscillation and SST changes in the eastern equatorial Pacific Ocean, showing that El Niño was a basin-wide phenomenon.

Although the origins of ENSO are in the Tropical Pacific, its impacts extend to weather patterns around the globe (Chang and Zebiak, 2003). Numerous studies have found a significant correlation between East African rainfall and the El Niño Southern Oscillation, with the sign of the correlation and its variance depending on the precise region and season (Ropelewski and Halpert 1987, 1989; Janowiak 1988; Ogallo et al. 1988; Nicholson and Kim 1997; Mutai and Ward 2000; Nicholson and Selato 2000; Camberlin et al. 2001; Neng et al. 2002; Korecha and Barnston 2007). These studies have shown that El Niño is associated with droughts in Ethiopia with a negative peak in correlation during the July-September rainy season. However, the teleconnection structure is quite different further south in Equatorial East Africa (EEA) during the ‘short rain’ season from October to December (Mutai and Ward, 2000). Here El Niño is associated with abnormally wet conditions with a positive correlation peak during OND (Lau and Sheu 1988; Indeje et al. 2000; McHugh 2006), with a few exceptions such as along the Kenyan coastline (Ogallo 1988; Hastenrath et al. 1993). The opposite situations occur during a La Niña with abnormally wet conditions in Ethiopia during July-September and dry conditions during the ‘short rain’ season in EEA. This link between Kenyan/Ugandan rainfall and ENSO can be seen in Figure 2.7. This Figure shows the Climate Prediction Center’s analysis of the relationships between El Niño and La Niña and global temperature and rainfall patterns during the December-February season (CPC, 2007).

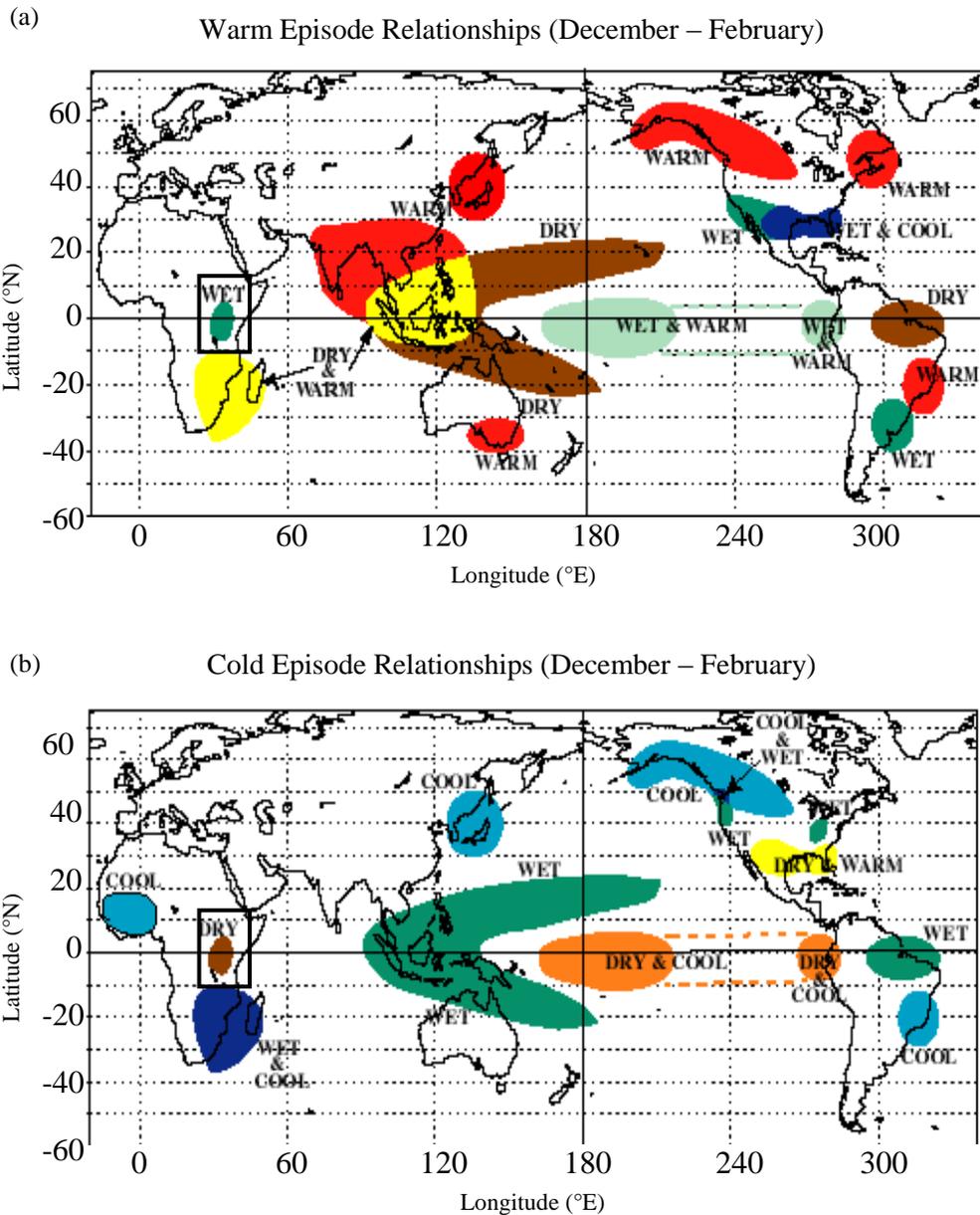


Figure 2.7. The Climate Prediction Center’s analysis of the relationships between El Niño (warm episode) and La Niña (cold episode) and global temperature and rainfall patterns for December-February (CPC, 2007). The boxed regions highlight the Kenya/Uganda region.

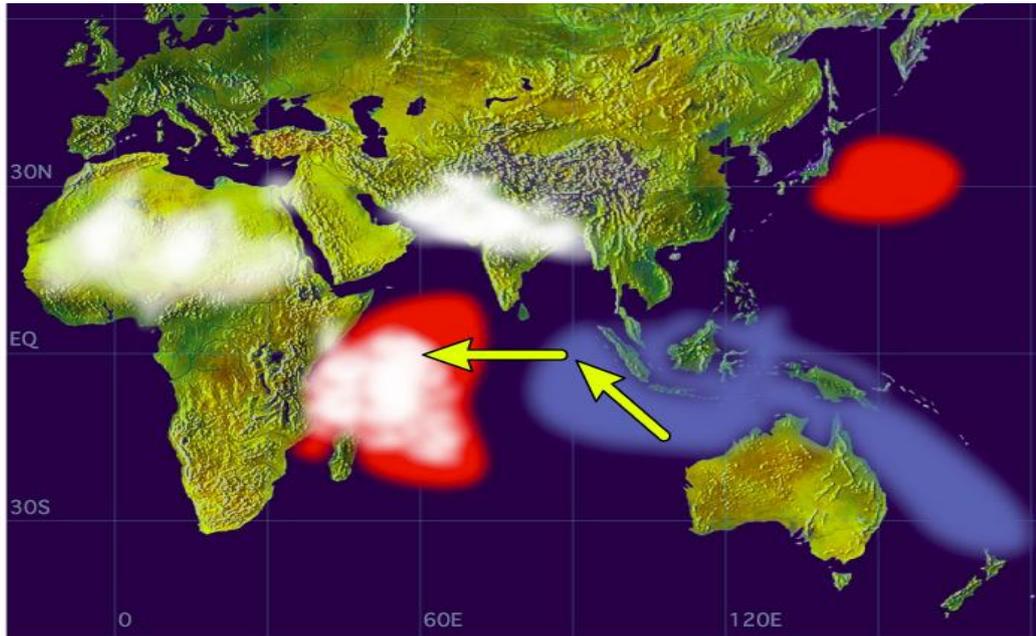
Theories behind the dynamics of this teleconnection are suggested in several papers. Camberlin et al. (2001) suggests that the link between El Niño and East African rainfall is due to El Niño weakening the equatorial Walker cell over the Indian Ocean. This results in low-level convergence and upper-level divergence over East Africa, producing anomalously wet 'short-rains' in most of EEA. Black (2003, 2005) and Mutai and Ward (2000) suggest that El Niño events are associated with a general warming of the Indian Ocean and that anomalous cold SSTs are introduced to the south Indian Ocean around the Maritime Continent via the Indonesian through flow. These cold SST anomalies could be a response to anomalous along-shore southerly winds (Black, 2003, 2005) causing upwelling of colder waters (Xie and Annamalai, 2002). This generates an east-west sea-level pressure gradient that drives moist easterly wind across the Indian Ocean towards East Africa. Black (2003, 2005) expands on this idea by suggesting that the observed link between ENSO and East African rainfall is actually a manifestation of the link between ENSO and the Indian Ocean Dipole (IOD). This theory has gathered support in the literature since the discovery of the IOD in 1999 (Saji et al., 1999; Webster et al., 1999). The following sub-section presents the IOD as a potential predictor for East African rainfall.

2.3.3 Indian Ocean: Indian Ocean Dipole (IOD)

The IOD is a coupled ocean-atmosphere system, with fluctuations in SST anomalies across the Indian Ocean (Saji et al., 1999; Webster et al., 1999). Research into the IOD and its associated climate anomalies is still in its infancy (Luo et al., 2008). Pioneering work on the IOD was first published in 1999 by Saji et al. and Webster et al. with further research being published in recent years supporting the idea of an independent ocean-atmosphere coupled circulation system in the Indian Ocean driving surrounding climate variability. The Dipole Mode Index (DMI) was developed by Saji et al. (1999) as a simple index to measure the IOD, describing the zonal gradient in SST anomalies across the Indian Ocean between the west (50°-70°E, 10°S-10°N) and the east (90°-110°E, 10°S-Equator). These SST anomalies appear around June and generally peak in October. The SST anomalies force changes in atmospheric circulations and rainfall patterns across the Indian and west Pacific Ocean Basins. The SST, wind and rainfall anomalies associated with the IOD are presented in Figure 2.8.

(a)

Positive Dipole Mode



(b)

Negative Dipole Mode

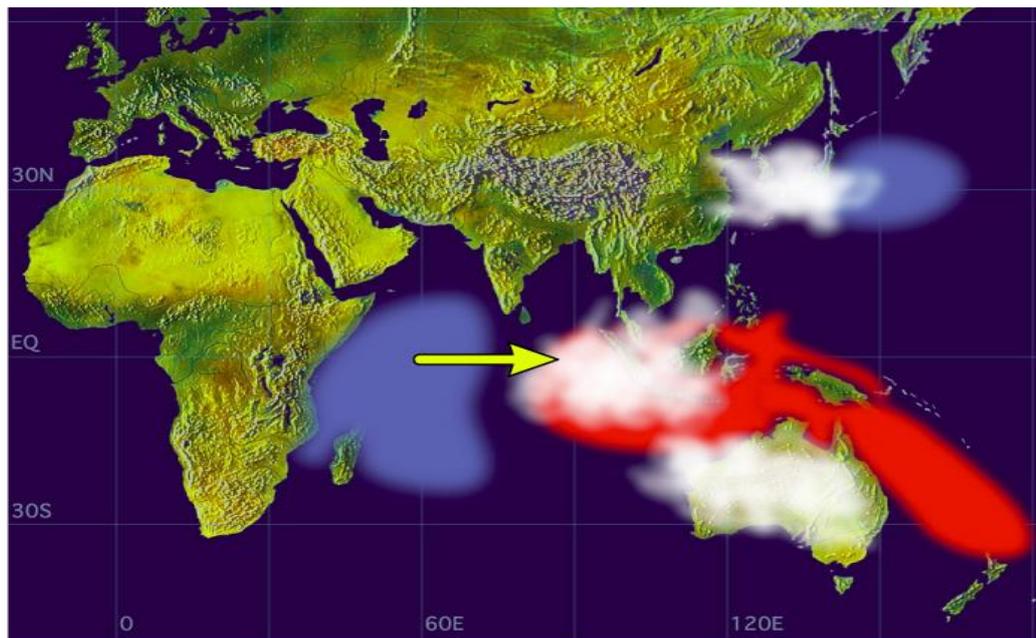


Figure 2.8. The Indian Ocean Dipole (IOD) and its associated negative (blue) and positive (red) SST anomalies. White indicates areas of increased convective activity. Low-level wind direction is indicated by the yellow arrows. Significant anomalies appear around May-June, intensify through July to September and peak in October, followed by a rapid demise (Saji et al., 1999).

The positive mode of the IOD is presented in Figure 2.8 (a). Positive SST anomalies are observed in the equatorial-west Indian Ocean with negative SST anomalies in the equatorial-east Indian Ocean, Indonesia and in the ocean around northern and eastern Australia. This reversed SST gradient, relative to the climatological mean (Webster et al., 1999), drives an atmospheric circulation with easterly low-level wind anomalies. Rising motion is enhanced over the positive SST anomalies in the western Indian Ocean. Convective activity and rainfall amounts have been observed to increase over this area, including over parts of East Africa, during the ‘short-rain’ season, with the relationship strength varying by region (Zablone and Ogallo, 2008).

The opposite SST anomaly patterns occur during the negative mode of the IOD as shown in Figure 2.8 (b). This SST gradient drives an atmospheric circulation with westerly low-level wind anomalies. Rising motion and convective activity is enhanced over the positive SST anomalies in the eastern Indian Ocean. The low-level westerlies lead to enhanced subsidence over East Africa and increased transport of moisture away from the continent, leading to the observed reduction in East African rainfall (Hastenrath et al. 1993; Hastenrath 2000; Manpande and Reason 2005).

2.3.4 The relationship between the IOD and ENSO

The relationship between the IOD and ENSO is a controversial issue in the literature with the scientific community split between the IOD being dependent on ENSO (Reason et al. 2000; Baquero-Bernal et al. 2002; Xie and Annamalai 2002) and independent of ENSO (Saji et al. 1999; Webster et al. 1999; Guan et al. 2003; Saji and Yamagata 2003; Marchant et al. 2006; Luo et al. 2008).

The IOD and ENSO do appear to be linked to some extent. The Spearman’s Rank correlation between the OND Niño 3.4 index and the OND DMI is moderately strong and temporally stable with Rank Correlation values >0.55 for each period 1959-1974, 1975-1990 and 1959-1990. The paragraphs below present the two main suggested theories from the literature for the relationship between the IOD and ENSO:

1. The IOD is forced primarily by ENSO

Baquero-Bernal et al. (2002) conducted a series of coupled ocean-atmosphere model experiments and found that the dipole-like variability in the Indian Ocean can only be explained in the context of ENSO, although they did suggest that the dipole may sometimes be forced stochastically by the atmosphere. Xie and Annamalai (2002) conclude that much of the Indian Ocean variability is due to oceanic Rossby waves that propagate from the east and are forced by ENSO. These Rossby waves are thought to interact with the atmosphere after reaching the western Indian Ocean.

2. The IOD can be forced by different circulations, including ENSO

There are several papers that conclude that although the IOD may sometimes evolve without ENSO external forcing, it does on some occasions interact with ENSO, possibly through the Walker Circulation or via the Indonesian through-flow (Yamagata et al. 2002; Behera and Yamagata 2003; Black 2003, 2005; Behera et al. 2006). Drbohlav et al. (2007) and Behera and Yamagata (2003) add to the debate by suggesting that the IOD can be induced by both ENSO and local Indian Ocean circulations, although the resulting spatial IOD structures may be slightly different. This idea is supported by Black (2003, 2005) and Xie and Annamalai (2002) who have proposed that the occurrence of IOD events with no El Niño, such as in 1961, may suggest that the IOD can be triggered by factors other than ENSO. During the East African floods of 1961, an IOD structure was evident in the SST patterns of the Indian Ocean but there was no El Niño in the Pacific (Saji et al., 1999).

Which is the most dominant forcing: ENSO or IOD?

It is important for this study to explore which SST teleconnections are the most important in forcing East African rainfall variability. Several recent studies provide evidence that the IOD is more dominant than ENSO in forcing East African rainfall variability (Black, 2003; Clark et al., 2003; Ummenhofer et al., 2009). Black (2003) observes that there has been above average rainfall in East Africa during every positive IOD year from 1960 to 2000, compared to only during 4 out of 9 El Niños, and that the top 5 highest rainfall seasons occurred during a positive IOD during this period.

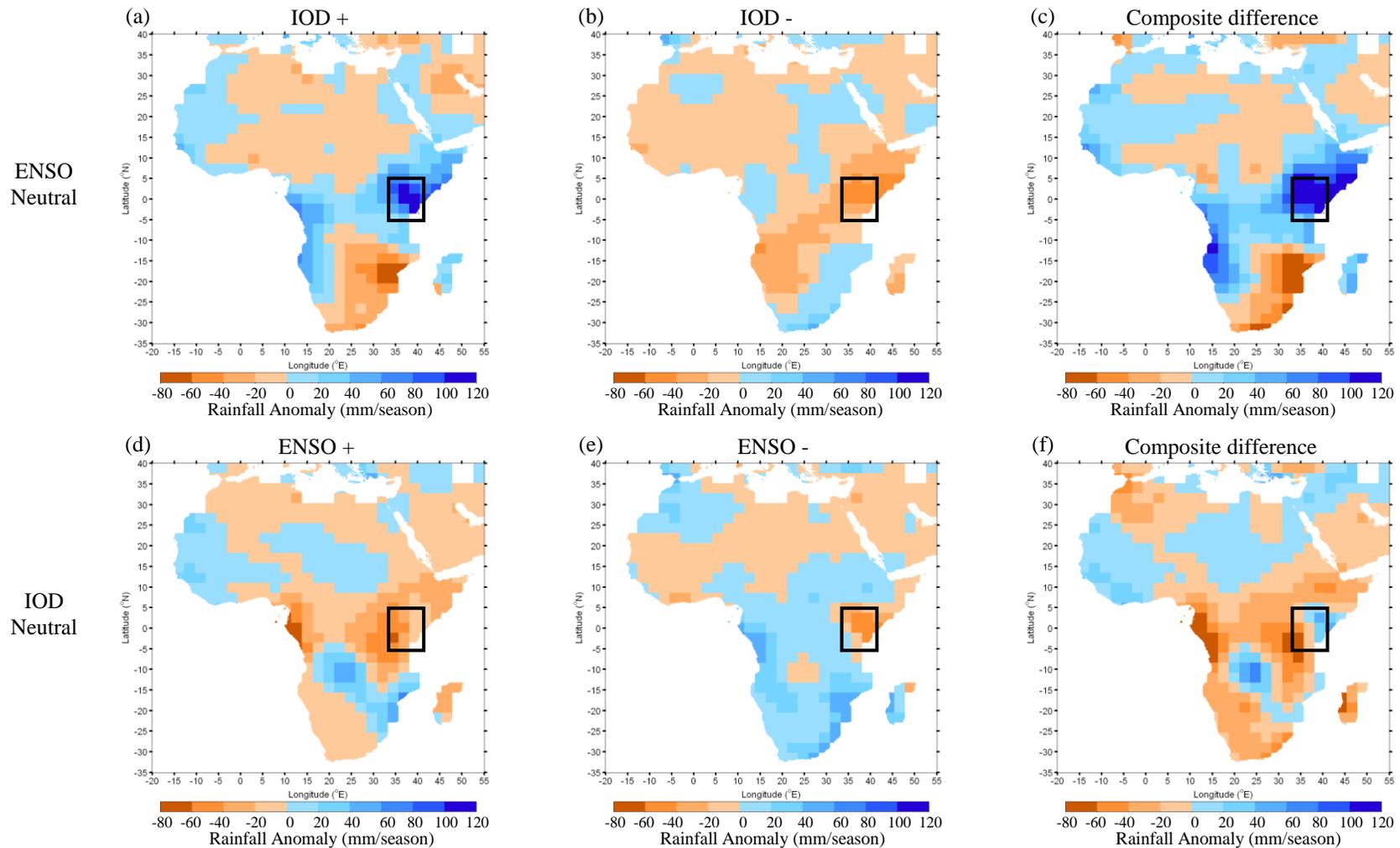


Figure 2.9. African OND rainfall anomaly linked to the IOD independent of ENSO (panels (a) to (c)) and linked to ENSO independent of the IOD (panels (d) to (f)) for 1959-2000. Panel (c) shows the difference in OND rainfall anomaly between those subset years when the OND ENSO is neutral and the OND IOD is (a) above median and (b) below median. Panel (f) shows the difference in OND rainfall anomaly between those subset years when the OND IOD is neutral and the OND ENSO is (d) above median and (e) below median.

Figure 2.9 supports the suggestion that the IOD has a stronger link to East African rainfall variability than ENSO has. OND rainfall anomalies over Africa are presented in each panel for a specified subset of years (see below) selected from the period 1959-2000. The location of Kenya is shown by the black box in each panel.

The link between the IOD and African OND seasonal rainfall anomalies, independent of ENSO, is investigated in panels (a-c). Panel (a) shows the rainfall anomalies for the 5 years when both the OND seasonal average Niño 3.4 index value lies within the middle tercile of the 42-year period ('ENSO neutral phase') and the OND seasonal average DMI value lies above the median of the 42-year period ('IOD positive phase'). Panel (b) shows the rainfall anomalies for the 9 years when both the OND seasonal average Niño 3.4 index value lies within the middle tercile ('ENSO neutral phase') and the OND seasonal average DMI value lies below the median ('IOD negative phase'). Panel (c) shows the difference in rainfall anomalies between panel (a) and panel (b), which represents the overall link between the IOD, independent of ENSO, and OND seasonal rainfall anomalies. The IOD has a strong link to Kenyan OND seasonal rainfall anomalies during ENSO neutral years, producing overall positive rainfall anomalies of 32 to 120mm/season. The results also show that the positive phase of the OND IOD has a stronger affect on Kenyan OND rainfall than the negative OND IOD phase.

The link between ENSO and African OND seasonal rainfall anomalies, independent of the IOD, is investigated in panels (d-f). Panel (d) shows the rainfall anomalies for the 6 years when both the OND seasonal average DMI value lies within the middle tercile of the 42-year period ('IOD neutral phase') and the OND seasonal average Niño 3.4 index value lies above the median of the 42-year period ('ENSO positive phase'). Panel (e) shows the rainfall anomalies for the 8 years when both the OND seasonal average DMI value lies within the middle tercile ('IOD neutral phase') and the OND seasonal average Niño 3.4 index value lies below the median ('ENSO negative phase'). Panel (f) shows the difference in rainfall anomalies between panel (d) and panel (e), which represents the overall link between ENSO, independent of the IOD, on OND seasonal rainfall anomalies. ENSO has a weaker link to Kenyan OND seasonal rainfall anomalies during IOD neutral years, producing overall rainfall anomalies of -71 to 41mm/month.

The IOD therefore has a stronger link to Kenyan OND seasonal rainfall variability than ENSO. The results show that any link between Kenyan OND rainfall and OND ENSO

is not linear. It must be noted that Figure 2.9 considers only contemporaneous relationships between rainfall anomalies and the IOD and ENSO over the OND season.

2.3.5 Atlantic Ocean

There have been several studies suggesting links between positive rainfall anomalies in East Africa and periods of westerly outbreaks, bringing moist Atlantic air into the region (Davies et al. 1985; Nicholson and Entekhabi 1987; McHugh and Rogers 2001; Camberlin and Philippon 2002; McHugh 2004). McHugh and Rogers (2001) demonstrated that this link tended to occur during the positive phase of the North Atlantic Oscillation (NAO), which generally peaks during the boreal winter. The NAO is an atmospheric oscillation in sea-level pressure over the North Atlantic Ocean between the Azores subtropical high and the Icelandic polar low. A positive NAO occurs when these systems have higher and deeper pressures respectively, which can be linked to perturbations in atmospheric circulation away from the North Atlantic.

McHugh (2004) provides a dynamical explanation for the link between westerly outbreaks of moist Atlantic air flow and increased rainfall in East Africa. Atlantic air masses can be advected into central Africa around the northern margins of the climatological heat low over southwest Africa. These are lifted over the highlands of East Africa and meet with the easterly monsoon flows from the Indian Ocean in a complex, meridional convergence zone. This mechanism for increased rainfall is dependent on the convergence of low-level water fluxes into the region, which causes the lower atmosphere to become unstable and produce rainfall. This adds to Ntale et al.'s (2003) idea that low MAM rainfall in East Africa is associated with low SSTs in the Indian Ocean adjacent to East Africa. This negative SST anomaly acts to lower the pressure gradient across the basin and reduce the easterly moisture advection towards East Africa. This would prevent the Congo westerly from depositing moisture in East Africa due to the lack of moisture convergence at low-levels over East Africa.

2.4 Long-range forecasting methods

Seasonal climate forecasting is one of the most promising developments for the early warning of climate hazards (Murphy et al., 2001). Long-range forecasts of seasonal rainfall are made operationally for East Africa using statistical methods, dynamical models and a combination of the two. The National Meteorological Agencies of the countries of East Africa use statistical methods to develop their seasonal forecasts before each rainy season (Section 2.5). They also refer to output from Coupled General Circulation Models (CGCMs), which produce global monthly rainfall forecasts (Chapters 4 and 8). The scientific community has not reached a consensus over which method produces the most consistently skilful forecasts over East Africa. There is a lot of scope for further research and debate on this subject.

2.4.1 Statistical forecasting methods

Reliable statistical seasonal rainfall forecasts would help to provide warnings of droughts in East Africa. Long-range forecasting is possible because climatic perturbations are forced by lower-boundary conditions, such as SSTs. These evolve more slowly than the atmospheric perturbations that they create (Murphy et al., 2001). Due to the coupled nature of the oceanic-atmospheric system first outlined by Bjerknes (1966, 1969), the evolution of SSTs influence the atmosphere through heat, mass and momentum exchanges. The methodology behind the development of statistical rainfall hindcast models is discussed in detail in Chapter 5.

2.4.2 Dynamical forecasting models

Seasonal forecasts can be made using mathematical models of the climate system. Powerful computers are needed in order to run these dynamical climate models by simultaneously solving the fundamental equations of mass, momentum and energy that govern the processes in the atmosphere and the oceans (Doblas-Reyes et al., 2006). Within the computer models, the globe is split into grid boxes, both horizontally and vertically. Within each grid box the fundamental equations are solved as the models are

integrated forward through time, in order to predict the future evolution of the climate system (Murphy et al., 2001). CGCMs are an advanced form of dynamical model, which mathematically model the oceanic and atmospheric processes as well as their interactions and allow SSTs, which usually act as the lower boundary forcing, to evolve over time.

In reality the SSTs often depend on the initial state of the atmosphere. Models are sometimes run several times with slightly perturbed initial conditions. Each individual run is called a *member* and the collective term for all the members is an *ensemble*. The proportion of atmospheric variance that is different in each of these ensemble members is likely to be a result of chaos in the system. Chapters 4 and 8 introduce two ensemble forecasting systems and present their hindcasting skill for seasonal rainfall over Africa.

2.5 Operational seasonal forecasting in Kenya

Seasonal forecasts for Kenyan rainfall are issued twice a year, before each rainy season, for the use of the public, government and businesses. Statistical seasonal rainfall forecasts for each region of the country are originally developed by the Kenya Meteorological Department (KMD), using the methods described in the next paragraph. Their forecasts are then compared against the statistical forecasts from neighbouring countries, as well as against the forecasts from several global dynamical models, issued by Meteorological agencies in Europe and the USA. The forecasts are then amalgamated into a consensus forecast, which gives the tercile probabilistic forecast of rainfall being in a particular category (above-, near- or below-normal rainfall), across the whole of the Greater Horn of Africa. An example OND consensus forecast can be seen in figure 2.10 (a). This 2008 forecast shows areas of near- to above-normal rainfall in green and near- to below-normal rainfall in yellow, with the probabilities for each tercile in the boxes. This forecast is then re-adjusted by each country's Meteorological office. The resultant OND 2008 seasonal forecast for Kenya can be seen in Figure 2.10 (b).

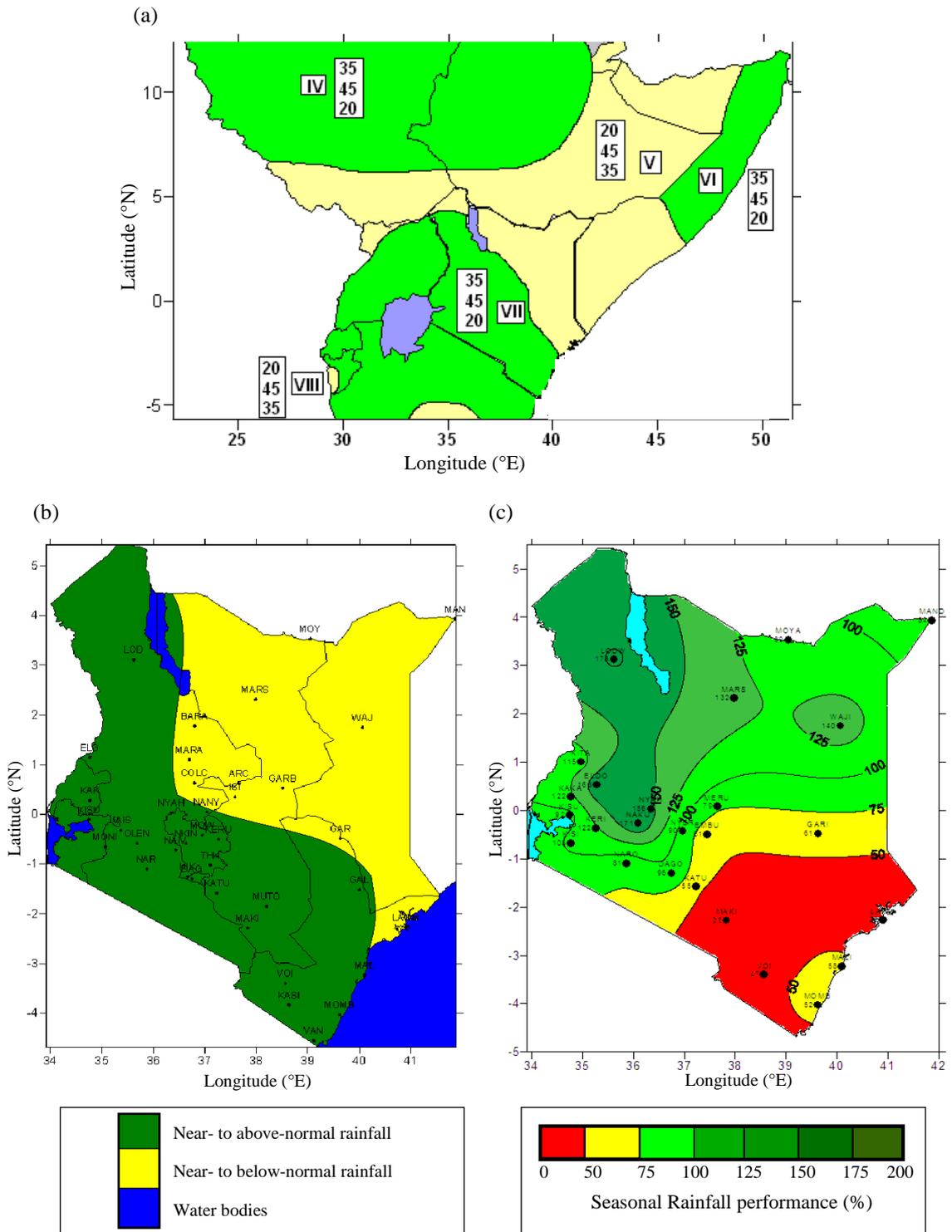


Figure 2.10. (a) OND 2008 rainfall forecast, issued September 2008 (ICPAC, 2008). Shows probability distributions that indicate above-, near-, or below-normal rainfall for each zone. Green represents near-normal to above-normal rainfall; yellow represents near-normal to below-normal rainfall. (b) OND 2008 rainfall forecast, adapted for Kenya, issued September 2008 (KMD, 2008). Shows rainfall forecasts of near-normal to above-normal rainfall (green) and near-normal to below-normal rainfall (yellow). (c) OND 2008 rainfall performance, issued March 2009 (KMD, 2009). Shows the observed rainfall performance relative to a long-term climatology. Regions showing values of > 100% experienced above-normal rainfall and regions showing values of < 100% experienced below-normal rainfall.

The following paragraph details the methods used to develop statistical seasonal rainfall forecasts by the National Meteorological Agencies in the Greater Horn of Africa. The information was gained through personal communication with operational meteorologists from Tanzania and Kenya (Likumana, 2008; Muita, 2008). The statistical seasonal rainfall forecasts are developed using a program called SYSTAT. This program has functions that allow the user to correlate global SSTs, up to 6 months prior to the rainy season, with historical rainfall data from synoptic weather stations. The rainfall data used to represent each of Kenya's homogeneous rainfall regions is taken from a single rain gauge station from each region, with good quality rainfall data. However, it is unclear as to what definition of 'good' is used. When a correlation value of > 0.5 is located, SYSTAT takes the regional historical SSTs over a period from 1961-1991 to develop a forecast model for seasonal rainfall using linear regression. The period from 1992-present day is reserved as an independent verification period to assess the number of years that the model predicts the rainfall to be in the correct tercile. These regional forecasts are then compared against those of the neighbouring countries. If they don't agree, the forecasters look at historical years with similar SST patterns to see which of their forecasts are most similar to the rainfall events in that historical year.

2.5.1 The Greater Horn of Africa Climate Outlook Forum (GHA COF)

The Greater Horn of Africa Climate Outlook Forum (GHA COF) was established in 1997 by ICPAC (the Intergovernmental Authority on Development Climate Prediction and Applications Centre) in Nairobi. The idea behind this bi-annual forum is to bring together meteorologists and climate scientists from the East African National Meteorological Services to develop a consensus forecast before each of the main rainy seasons (Ogallo, 2008). The COF is also attended by Government representatives, Non-Government Organisations, media and other forecast users (Patt et al., 2007). This allows an opportunity for the forecasters to explain the forecasts to the users and to help them interpret the forecasts for their regions of interest.

The primary objectives of the COFs are: (1) to develop a consensus climate outlook, (2) to facilitate research and forecasting cooperation and data exchange between countries and (3) to create a regular dialogue between climate scientists/forecasters and the users of the forecasts (Buizer et al., 2000). During the forum a consensus seasonal climate

forecast is established for the coming rainy season, which is tailored to be useful to climate sensitive sectors. This is made available through an outlook map, with an accompanying description on the internet by the IRI (International Research Institute for Climate and Society), NOAA (National Oceanic and Atmospheric Administration) and ICPAC (Buizer, 2000; Ogallo, 2008). The usefulness of the forecasts is improved in three ways. Firstly they are made to be specific to users needs (e.g. feeding into the crop prediction models). Secondly, forecasters work with users during the forum to help them to interpret the forecasts. Thirdly, inclusive communication is used, as farmers do not understand probabilistic forecasts and need the reliability information to be explained (Patt et al., 2007).

The continuing success and growth of the COF is limited by several factors, some of which were presented by Ogallo (2008) at the COF in February 2008. The primary limitation is finance from the East African Governments. The budget for the GHA COF is very low at US\$20,000 and users have to pay for themselves to attend. Another issue is weak and corrupt Governments and a political mistrust of the forecasts, partially due to a lack of understanding of the science by the politicians. A further problem is in communicating the forecast to the millions of East African people who speak any of the thousands of regional languages within the areas covered by the climate outlook.

One final issue is that the seasonal forecasts are not thoroughly verified following the rainy season. This is primarily because, as the budgets are so low, there is no money left after the COF to spend time on a thorough verification. Figure 2.10 (c) presents the only verification published by the KMD for the 2008 OND rainy season. It shows the observed seasonal rainfall performance relative to a long-term climatology. The regions >100%, which experienced above normal rainfall, cover the majority of the northern half of Kenya. The forecast was therefore not accurate, as Figure 2.10 (b) shows that 2/3 of the northern half of Kenya was forecast to have near- to below-normal rainfall over this season. Also, the south-central area of Kenya was forecast to have near- to above normal rainfall (Figure 2.10 (b)), whereas in fact the majority of this area received only 50% of the average rainfall. This one example shows that a thorough verification of the seasonal forecasts is necessary in order for the KMD to see where their forecast lacked skill so that improvements can be made in the future.

Chapter 3

Data

Chapter 3 describes the data employed in this study. High quality rainfall data are fundamental to the verification of seasonal rainfall hindcasts and to the development of new statistical hindcast or forecast models for seasonal rainfall. Several other forms of data are also used in the study, as predictors to develop the statistical hindcasts and forecasts. These data include global SSTs (Section 3.2), 850hPa winds (Section 3.3) and a selection of climate indices and oscillations (Section 3.4).

3.1 Rainfall

Two forms of historical monthly rainfall data are used in this study. In order to verify the seasonal rainfall hindcasts of leading dynamical models in Chapters 4 and 8, accurate and continuous gridded rainfall data are required with the same spatial and temporal resolution as the hindcasts. Section 3.1.1 describes the gridded rainfall data sets employed in this study. Other gridded rainfall datasets were also tested for potential use, such as the Global Precipitation Climatology Project (GPCP) monthly mean estimates of precipitation dataset (Huffman et al., 1997). Although the GPCP dataset was not used in this study, it provided confidence to see that the gridded rainfall datasets presented in Section 3.1.1 were found to be consistent with the GPCP's monthly mean rainfall values, with correlation values of > 0.9 across the whole of East Africa.

Section 3.1.2 introduces the Kenyan rain gauge data sets that were obtained from the Kenya Meteorological Department (KMD). These rain gauge data are required to partition Kenya into homogeneous rainfall regions (Chapter 5.1) and to develop skilful seasonal rainfall hindcast models (Chapters 5-7). As rain gauge data are direct measurements, it is necessary to conduct a thorough quality control assessment (Section 3.1.3) in order to assess the data for erroneous values and long-term trends. These then need to be corrected, with erroneous data reduced to a threshold value and any long-term trends removed, so that only the year-to-year climate variability remains. This removes the influence of any multi-decadal variability, as investigation into this aspect of climate is beyond the scope of this thesis.

3.1.1 Gridded rainfall data

Global Precipitation Climatology Center's (GPCC) precipitation climatology

In order to assess the seasonal rainfall hindcast skill available from the DEMETER hindcast system (Chapter 4), a verification rainfall data set is required with the same spatial and temporal resolution as the hindcasts. This allows a direct comparison to be made between the seasonal rainfall hindcasts and the observed rainfall. The verification data requirements for Chapter 4 are a gridded rainfall data set on a $2.5^{\circ} \times 2.5^{\circ}$ latitude/longitude grid with monthly-averaged data spanning the period 1959-2000. The dotted red lines in Figure 3.1 show the $2.5^{\circ} \times 2.5^{\circ}$ grid box boundaries used in this study.

A thorough comparison of all available gridded precipitation data sets is made. The Global Precipitation Climatology Center's (GPCC) VASCLimO (Variability Analyses of Surface **C**limate **O**bservations) 50 year precipitation climatology data set (Beck et al., 2005) best fits these spatial and temporal requirements and is therefore selected for use in this study. The GPCC monthly rainfall data are available on a $2.5^{\circ} \times 2.5^{\circ}$ latitude/longitude grid over the period 1951-2000. The GPCC data are available from <http://www.dwd.de>.

The GPCC data are purely observation-based and are derived from historical rainfall records from three main sources: the precipitation databases of the Food and Agriculture Organization of the United Nations, the Climatic Research Unit at the University of East Anglia in the UK, and the Global Historical Climatological Network. The merging procedure is designed to incorporate data from each information source based on an estimate of its quality. The GPCC apply thorough quality controls to these data to increase the reliability of their gridded monthly data.

The accuracy of an historical rainfall data set is limited by the uneven distribution of rain gauges, incomplete data records, and random errors in the records. This is a particular problem in Africa where the rain gauge network is particularly sparse, comprising about 500 stations. Figure 2.2 shows the distribution of rain gauge stations that contribute to the GPCC gridded data set. Only stations with at least 90% data availability between 1951 and 2000 are included. It is clear that station coverage over half of the continent is either poor or non-existent. The densest coverage is found over sub-Saharan West Africa (between Senegal and Nigeria), parts of southern Africa and

parts of Kenya. This non-uniform station coverage may affect the spatial reliability of hindcast skill assessment. For example, where the station density is low the hindcast skill might be expected to be low, and vice versa.

Climate Prediction Centre's (CPC) Merged Analysis of Precipitation (CMAP)

In Chapter 8 an assessment is made of the seasonal rainfall hindcast skill available from the EUROSIP system. These hindcasts are available on a 2.5°x2.5° latitude/longitude grid over the shorter period of 1981-2005. As this hindcast period is limited to 25 years and the GPCC data set ends in 2001, another gridded rainfall data set is selected in order to verify EUROSIP's hindcasts over its entire hindcast period.

The Climate Prediction Center's Merged Analysis of Precipitation (CMAP) data set (Xie and Arkin, 1997) fulfils the spatial and temporal requirements for the verification of EUROSIP's hindcasts, as the data are available on a 2.5°x2.5° latitude/longitude grid over the period 1979-2008. The CMAP data set is formed by merging rain gauge data from: the GPCC, precipitation estimates from several satellite-based algorithms (infrared and microwave) and precipitation distributions from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP-NCAR) reanalysis. CMAP data are provided by the Physical Sciences Division (PSD) of the Earth System Research Laboratory (ESRL), which is based at the National Oceanic and Atmospheric Administration (NOAA), Boulder, Colorado, USA. The data are available from <http://www.cdc.noaa.gov/data/gridded/data.cmap.html>.

3.1.2 Rain gauge data

Rain gauge data are required to partition Kenya into homogeneous rainfall regions prior to developing statistical seasonal rainfall hindcast models for each region. This is necessary due to the heterogeneous nature of the rainfall distribution over Kenya. Monthly rain gauge data are obtained from the KMD. Long historical rainfall records are vital to assess the long-term stability of statistical relationships. Stations with rainfall records covering the period 1959-2006 are therefore selected for potential use. This 48-year period is split into three 16 year periods. The first two periods are used to find suitable predictors and develop statistical seasonal rainfall hindcast models for each homogeneous rainfall region. The period from 1991 to 2006 is reserved as an independent verification period (Chapter 5).

Station short name	Station full name	Kenyan Region	Latitude (°N)	Longitude (°E)	Data available from	Missing data (%)
Lamu	Lamu Meteorological Station *	South	-2.16	40.54	1908	2.4
Malindi	Malindi Meteorological Station *	South	-3.14	40.06	1961	10.9
Mombasa	Mombasa Port Reitz Airport*	South	-4.02	39.37	1946	2.4
Garissa	Garissa Meteorological Station *	Southeast	-0.29	39.38	1932	2.1
Katamani	Katamani Research Station [□]	Southeast	-1.36	37.14	1953	2.3
Makindu	Makindu Meteorological Station *	Southeast	-2.17	37.50	1904	1.0
Marania Timau	Marania Timau [□]	Southeast	0.05	37.27	1925	5.0
Meru	Meru Forest Station [□]	Southeast	0.03	37.38	1948	6.6
Voi	Voi Meteorological Station *	Southeast	-3.24	38.34	1904	1.2
Dagoretti	Dagoretti Corner Meteorological Station *	Southwest	-1.18	36.45	1954	1.7
Nakuru	Nakuru Railway Station [□]	Southwest	-0.17	36.04	1904	6.1
Narok	Narok Meteorological Station *	Southwest	-1.06	35.52	1913	1.4
Chorlim	Chorlim Agricultural Development Corporation [□]	West	1.02	34.48	1926	14.6
Kakamega	Kakamega Agromet Station*	West	0.17	34.36	1957	1.7
Kisii Met	Kisii Meteorological Station*	West	-0.41	34.47	1963	10.6
Kisumu	Kisumu Meteorological Station *	West	-0.06	34.45	1938	1.2
Baragoi	Baragoi El Barta District Office [□]	Northwest	1.47	36.48	1938	17.0
Lodwar	Lodwar Meteorological Station *	Northwest	3.07	35.37	1919	2.1
Mandera	Mandera Meteorological Station *	Northeast	3.56	41.52	1936	1.9
Marsabit	Marsabit Meteorological Station *	Northeast	2.19	37.59	1920	5.6
Moyale	Moyale Meteorological Station *	Northeast	3.32	39.03	1920	1.9
Wajir	Wajir Meteorological Station *	Northeast	1.45	40.04	1917	3.3

Table 3.1. Kenyan rain gauge stations used in the study (data obtained from the Kenya Meteorological Department). Only stations with >80% data available over the period 1959-2006 are included. The positions of these stations are shown on the map in Figure 3.1. The region names are referred to in Chapter 5.

* Synoptic Stations

[□] Non-Synoptic Stations

The quality of rainfall data varies greatly between the observation stations in terms of time span of available data records and missing data. The first challenge is to select enough well distributed, good quality rain gauge data sets to use in the study. Table 3.1 presents information on the 22 stations that are employed in this study including: the station names, regions (used in Chapter 5), coordinates, years from which data are available and percentages of missing data. Camberlin (2008) recommends obtaining monthly rainfall data from Kenya's synoptic stations, as more confidence can be had in their measurements as they are official meteorological stations. There are 39 official synoptic stations across Kenya. However, upon examination of the data, 23 of these stations cannot be used. This is due to: 3 stations having the same coordinates as other stations, 15 covering too short a time period and 5 missing more than 20% of their monthly data (the acceptance criterion for this study is at least 80% data availability over the period 1959-2006). Details of the remaining 16 synoptic stations that are employed in the study are presented in Table 3.1. The station locations are presented on the Kenyan map in Figure 3.1, marked with an asterisk in both cases.

To provide a wider spread of stations from across the country, data are also obtained from other non-synoptic stations in the southwest and central areas. These consist of over 2000 rain gauges across Kenya with locations ranging from Schools to Police Stations; however the majority have far less than 80% available data. 6 further stations with >80% data availability over the period 1959-2006 are selected for use in the study. Details of these stations are presented in Table 3.1 and their locations can be seen in Figure 3.1. These 6 non-synoptic stations are marked with a square in both cases. All of the rain gauge data from the KMD were compared against the rain gauge data collected by Professor Nicholson across her career (Personal communication, 2007) and were found to be consistent. This provides confidence in the quality of the rain gauge data provided by the KMD. Nicholson's data were also used to verify the seasonal rainfall hindcast skill in Chapter 6. The results were found to be very similar to those found using the KMD rain gauge data. Therefore they were not included in the thesis.

3.1.3 Quality control

The monthly rain gauge data sets obtained from the KMD contain many outlying data points, which can be either erroneous measurements or correct but extreme values. To deal appropriately with these, a quality control method is developed, based on that used by Gonzalez-Rouco et al. (2001) to quality control precipitation data in southwest Europe.

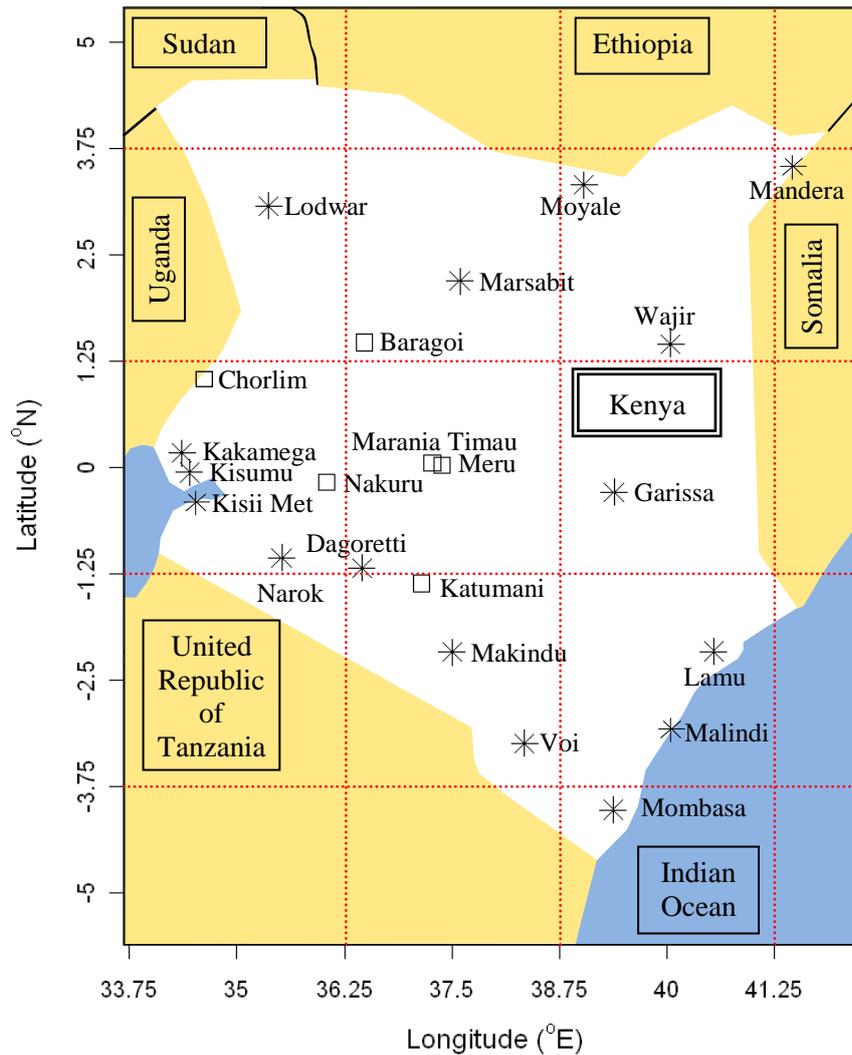


Figure 3.1. Map of Kenya showing the rain gauge stations selected for use in this study. The 16 Synoptic rain gauge stations are marked with an asterisk (*) and 6 non-synoptic rain gauge stations are marked by a square (□). The red dotted lines show the grid box boundaries of the gridded datasets used in this study.

The first stage is to reduce the magnitude of any outliers to a threshold value. This retains the information of an extreme event, but has a smaller influence on non-resistant statistics, such as the mean and standard deviation. This study uses the upper and lower

threshold values (P_{upper} and P_{lower}) given by Gonzalez-Rouco et al. (2001) and Wilks (2006), namely:

$$P_{upper} = q_{0.75} + (3 * IQR) \quad \text{Equation 3.1}$$

$$P_{lower} = q_{0.25} - (3 * IQR) \quad \text{Equation 3.2}$$

where $q_{0.75}$ and $q_{0.25}$ are the upper and lower quartiles and IQR is the inter-quartile range, which is a resistant measure of the spread of data. Any rainfall values above P_{upper} or below P_{lower} are replaced with the respective threshold value.

The final stage in preparing the data involves removing any long-term trends. Removing long-term trends, which could be due to climate change (Jewson et al., 2005), leaves only the natural year-to-year climate variability. It is assumed that the trends are approximately linear. The trend line of the original rainfall can be calculated using linear regression, where the gradient of the line indicates the long-term increase or decreases in rainfall. Each original value is therefore detrended by adding the difference in height of the trend line (change in rainfall) over the period. Therefore, for an overall decreasing trend, the trend line would be negative, and the original value would decrease to compensate for the trend and vice versa for an increasing trend.

The rainfall data for both OND and MAM are detrended over each sub-period used in this study. In Chapters 5-7 the sub-periods used are: 1959-1974, 1975-1990, 1959-1990 and 1991-2006. In Chapter 8 the sub periods used are: 1981-2005 and 1987-2005. Table 3.2 shows a representative example of the average linear trends removed from the OND regional rainfall indices over the period 1959-1990. It can be seen that the linear trends that are removed from the station data are generally small values when compared to the mean rainfall over the OND season. This is the case for all sub-periods in both seasons. The time series were found to have no significant signal of decadal or multi-decadal variability following the linear trend removal.

	South	Southeast	Southwest	West	Northwest	Northeast
Average linear trend in mm/season (1959-1990)	-12.4	-11.0	-1.7	-6.0	-1.7	-9.6
Mean OND rainfall (mm/season)	80	107	73	111	42	49

Table 3.2. Average linear trends that are removed from the original regional OND rainfall data over the period 1959-1990 (mm).

3.2 Sea surface temperatures

SST data are used in several places throughout this study, but arguably their most important function is as predictors for the statistical hindcast models for seasonal rainfall (Chapters 5-7). NCEP/NCAR Global Reanalysis monthly mean skin temperature data covering the period 1948-present are employed in this study (Kalnay et al., 1996). These data are available on the same 2.5°x2.5° latitude/longitude grid as the other gridded data used in this study. These data are available from <http://www.cdc.noaa.gov/data/gridded/data.ncep.reanalysis.derived.surfaceflux.html>.

The SST data are detrended per grid square using the method described in Section 3.1.3.

3.3 Wind

NCEP/NCAR Global Reanalysis monthly mean, 850hPa (roughly 1.5km above sea-level), u - and v -direction wind data over the period 1948-present are used as potential predictors in Chapters 5-7 of this study (Kalnay et al., 1996). These data are available on a 2.5°x2.5° latitude/longitude grid from: <http://www.cdc.noaa.gov/data/gridded/data.ncep.reanalysis.pressure.html>. Winds at 850hPa height are used because these are linked to low-level wind convergence, which directly affects rainfall.

3.4 Climate indices and oscillations

3.4.1 Niño indices (Niño 3.4 and Niño 4)

The strength and sign of the ENSO (Chapter 2.3.1) can be measured by the Niño 3.4 and Niño 4 indices. These monthly indices record the area average SST anomalies in the Niño 3.4 (5°N-5°S, 120°W-170°W) and Niño 4 (5°N-5°S, 160°E-150°W) regions respectively. The Niño 3.4 and Niño 4 indices are available over the period 1950-present from the Climate Prediction Center at: <http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices>.

3.4.2 Southern Oscillation Index (SOI)

The SOI is another ENSO index, which measures the difference in sea-level atmospheric pressure anomalies between Tahiti and Darwin, Australia (Trenberth, 1984). The negative phase of the SOI occurs when below-normal atmospheric pressure at Tahiti (east-central Pacific) and above-normal pressure at Darwin (west Pacific) are observed. Prolonged periods of negative (positive) SOI generally coincide with positive (negative) SST anomalies in the east-central Pacific, associated with an El Niño (a La Niña). These data are available over the period 1951-present from the Climate Prediction Center at: <http://www.cpc.noaa.gov/data/indices/soi>.

3.4.3 Dipole Mode Index (DMI)

The DMI (Chapter 2.3.1) was developed by Saji et al. (1999) to measure the strength of the IOD, describing the zonal gradient in SST anomalies across the Indian Ocean between the west (50°-70°E, 10°S-10°N) and the east (90°-110°E, 10°S-Equator). The DMI time series was kindly extended upon the author's request up to 2008 by Drs. Sasaki and Yamagata at the Japan Agency for Marine-Earth Science and Technology (JAMSTEC). The DMI is available from 1958-2008 from JAMSTEC at: http://www.jamstec.go.jp/frcgc/research/d1/iod/DATA/dmi_HadISST.txt.

3.4.4 Quasi-Biennial Oscillation (QBO) Index

The QBO is a quasi-periodic reversal of the stratospheric and tropospheric zonal winds (Ebdon and Veryard, 1961; Reed et al., 1961; Veryard and Ebdon, 1961; Angell and Korshover, 1964). An 'average' event tends to last for 22-34 months with a mean period of around 27 months (Cariolle et al., 1993), although all events vary in magnitude and duration. Indeje and Semazzi (2000) showed that globally averaged, stratospheric (30hPa) equatorial zonal winds were linked to East African MAM rainfall over the period 1979-1992. Further details on the QBO and its reported teleconnections to East African rainfall are presented in Chapter 7. The QBO monthly mean zonal wind data are produced by combining observations from three radiosonde stations: Canton Island, Gan Maldives Islands and Singapore. QBO-Index (30hPa) data are available from the

Freie Universität, Berlin, over the period 1953-present at: <http://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/qbo.dat>.

3.4.5 Madden-Julian Oscillation (MJO)

The MJO is a tropical atmospheric phenomenon, which develops over the Indian Ocean and progresses east across the tropics with a period of 30-60 days (Madden and Julian, 1994). The active phase of the MJO brings enhanced precipitation followed by suppressed precipitation. Pohl and Camberlin (2006) showed that the MJO was linked to the early onset of the MAM rains in Equatorial East Africa over the period 1979-1995. This study examines the MJO to test the robustness of its relationship with Kenyan seasonal rainfall over different time periods. The CPC have computed MJO indices at a selection of longitudes around the equator. The MJO Index at 70°E is used in this study as this is the closest index available to Kenya, situated in the central Indian Ocean. This MJO index is available over the period 1978-2006 at: http://www.cpc.noaa.gov/products/precip/CWlink/daily_mjo_index/proj_norm_order.ascii.

Chapter 4

Assessment of Current Dynamical Seasonal Rainfall Hindcast Skill over Africa

4.1 Introduction

Dynamical forecast systems, comprised of coupled ocean-atmosphere climate models, are thought to offer greater potential for forecast skill than statistical forecast models and greater skill than from individual coupled climate models (e.g. Krishnamurti et al., 1999; Goddard et al., 2001; Stefanova and Krishnamurti, 2002; Barnston et al., 2003; Palmer et al., 2004; Doblus-Reyes et al., 2005, 2006; Hagedorn et al., 2005). However, few quantitative assessments have been made of the skill of such forecast systems. This chapter presents an in-depth assessment of the seasonal rainfall hindcast skill over Africa available from the DEMETER multi-model ensemble system and from its comprising individual coupled models (Palmer et al., 2004). No such assessment of DEMETER has been made before. However, Thomson et al. (2006) and Jones et al. (2007) used DEMETER multi-model rainfall hindcasts to assess the potential for malaria early warnings in Botswana and Tanzania.

4.1.1 Aim of the assessment

This chapter is the first to examine the seasonal rainfall hindcast skill over Africa available from the DEMETER multi-model ensemble hindcast system. This assessment is made over the period 1959-2000 and employs hindcasts from three of the DEMETER coupled models; those from: the European Centre for Medium-Range Weather Forecasts (ECMWF), the Centre National de Recherches Météorologiques (Météo France) and the UK Met Office (UKMO). These three models are examined individually and as a multi-model.

4.1.2 Definitions

A *hindcast* is a retrospective forecast made by a model employed in real-time forecasting (AMS, 2009). DEMETER's hindcasts are produced by using historical ECMWF 40-year Re-Analysis (ERA-40) data (Uppala et al., 2005) to force the ocean analysis and to provide atmospheric and land surface initial conditions. The model skill is assessed by comparing the output against historical observations.

Hindcast *lead time* is the period of time between when the hindcast is issued and the date that it becomes valid (AMS, 2009). For example, a hindcast issued at the start of May for May is a zero month lead (lead 0) hindcast, while a hindcast issued at the start of May for July is a two months lead (lead 2) hindcast. The results in this chapter are shown mostly for lead 0 to maximise the skill, with the results in Section 4.3.3 showing the decrease in skill when the lead is extended to 3 months.

A *season* refers to a period of 3 months. This study focuses on hindcasts of *seasonal averages* rather than individual months because seasonal (and longer) deficits of rainfall are relevant for droughts. The following (non-standard) seasons receive prime focus in this chapter: February, March, April (FMA); May, June, July (MJJ); August, September, October (ASO); and November, December, January (NDJ). These seasons correspond to the average of the first 3 months of each DEMETER hindcast period. The DEMETER hindcasts are initialised on 1 February, 1 May, 1 August and 1 November.

The term *multi-model ensemble* is defined in this study as an average of several individual coupled models over the period 1959-2000. Unless otherwise stated, this averaging is performed by weighting the contribution from each individual model by the Mean Squared Error (*MSE*) of its hindcasts in order to maximise the hindcast skill.

Coupled ocean-atmosphere climate models are advanced forms of *dynamical models*. These are used to mathematically model oceanic and atmospheric processes and interactions, whilst allowing the sea-surface temperature boundary conditions to evolve over time. The physical equations that govern these processes and interactions are solved for many different variables in each grid box at different levels around the globe as the models are integrated forward through time (Murphy et al., 2001).

4.1.3 Introduction to DEMETER

DEMETER is a European dynamical multi-model ensemble global hindcast system, (Palmer et al. 2004). Its name derives from the project title: **D**evelopment of a **E**uropean **M**ulti-model **E**nsemble system for seasonal-to-in**TE**Rannual prediction.

DEMETER comprises seven state-of-the-art European coupled ocean-atmosphere global climate models. The contributing research centres are: European Centre for Medium-Range Weather Forecasts (ECMWF), Centre National de Recherches Météorologiques (Météo-France), UK Met Office (UKMO), European Centre for Research and Advanced Training in Scientific Computation (CERFACS), Istituto Nazionale di Geofisica e Vulcanologia (INGV), Laboratoire d'Océanographie Dynamique et de climatologie (LODYC) and Max-Planck Institut für Meteorologie (MPI). Each of these seven coupled models comprises nine ensemble members, each initialised with slightly perturbed initial conditions. These produce monthly hindcasts out to a lead of 5 months over the period 1980-2001.

The motivation behind the DEMETER multi-model ensemble system is to attempt to reduce hindcast errors, which are caused by model uncertainties and sensitivity to initial conditions (Palmer et al., 2004). Hindcast errors can occur due to problems with the processing of the model's equations. Although the equations that govern the atmosphere and oceans are well understood at the level of partial differential equations, their representations as ordinary differential equations, for the purpose of integrating on a computer, introduce errors. Such inaccuracies can easily propagate upscale in the model and produce large errors in the final hindcast values. These errors are reduced by combining individual coupled models, developed quasi-independently across Europe, into a multi-model ensemble (Palmer et al. 2004). The potential for errors in DEMETER's hindcasts due to model sensitivity to initial conditions is reduced by running each individual coupled model nine times with slightly perturbed initial conditions. Each of these nine model runs is called a *member*. The hindcasts from each of the nine members are then averaged together to give the resulting hindcast for that model.

A longer hindcast period from 1959 to 2001 is available for three of these models; those from: the ECMWF, Météo France and the UKMO. These three models, with their

extended hindcast availability, are employed in this chapter's skill assessment. Table 4.1 describes the characteristics of these three models. The DEMETER hindcasts are available on a 2.5°x2.5° latitude/longitude global grid. Due to the computational cost of running multi-model ensembles, DEMETER hindcast start dates are quarterly rather than monthly, initialised on: 1 February, 1 May, 1 August and 1 November. Each initialisation is integrated forward to give leads of up to 5 months for all nine members of each model. Hindcasts are available up to 2001 as DEMETER is intended only for research activities, not as an operational system. These hindcasts are accessible from the ECMWF website at http://data-portal.ecmwf.int/data/d/demeter_mnth/1950/.

	DEMETER		
	ECMWF	Météo France	UKMO
Long Title	European Centre for Medium-Range Weather Forecasts	Centre National de Recherches Météorologiques	United Kingdom Met Office
Country of Development	International Organisation (based in UK)	France	UK
System	2	2	2
Atmospheric Component	ECMWF IFS	ARPEGE v4	HadAM3
Atmospheric Horizontal Resolution	1.9° x 1.9°	2.8° x 2.8°	2.5°x3.75°
Atmospheric Vertical Levels	40	31	19
Oceanic Component	HOPE	ORCA	GLOSEA OGCM
Oceanic Horizontal Resolution	1° x 1°	2° x (2° to 0.5 at Equator)	1.25° x (1.25° to 0.3° at Equator)
Oceanic Vertical Levels	29	31	40
Hindcast Ensemble Members	9	9	9
Hindcast Years Available	1958-2001 (44 years)	1958-2001 (44 years)	1959-2001 (43 years)
Hindcast Issue Dates	1 st Feb, 1 st May, 1 st Aug, 1 st Nov	1 st Feb, 1 st May, 1 st Aug, 1 st Nov	1 st Feb, 1 st May, 1 st Aug, 1 st Nov
Hindcast Lead Time* (months)	0-5	0-5	0-5
Grid Resolution	2.5° x 2.5°	2.5° x 2.5°	2.5° x 2.5°

Table 4.1. Characteristics of the three DEMETER coupled general circulation models employed in this study (Palmer et al., 2004).

*Hindcast lead time refers to the period of time between the issue of the hindcast and the start of the hindcast period

4.2 Methodology

4.2.1 Bias correction

Due to inherent systematic errors, DEMETER and other dynamical models tend to drift with time towards a climate that differs to that observed (Vitart et al., 2007). This drift is called a bias and its removal is termed *bias correction*. This correction must be performed before any verification of the hindcasts takes place. The DEMETER hindcasts are bias corrected using a standard cross validation method (Wilks, 2006). This method computes the average difference between the hindcast and observation values for all years excluding the year of interest. The average difference is then removed from the hindcast value for the year of interest. This is repeated for all years, thus yielding a time series of bias-removed DEMETER hindcast values.

In some low rainfall cases this bias removal method produces a negative rainfall amount. If this occurs the rainfall value is forced to zero as negative rainfalls are unphysical. The above method was compared to another bias removal method used by the ECMWF (Doblas-Reyes, 2007), which uses anomalies of observations and hindcasts instead of absolute values. The ECMWF method retains the original error between the hindcast and the observations, which could be artificially reduced when the bias-removed rainfall is corrected from a negative value to zero. Differences between the two bias removal schemes are found to be small and lead to changes in correlation coefficient of <0.1 in over 96% of African grid squares.

4.2.2 Skill assessment

The skill of the deterministic DEMETER hindcasts of seasonal rainfall is examined over the period 1959-2000. Deterministic hindcasts provide a single (i.e. the most likely) output value. Confidence intervals on the deterministic skill are computed (Section 4.2.6). Three standard verification measures are used to assess deterministic hindcast skill (WMO, 2006; Wilks, 2006). These are rank correlation (r_{rank}), mean square skill score ($MSSS$), and percentage of hindcasts with the correct anomaly sign. These measures are described in the following sub-section, together with the methods used to compute the skill significances.

The hindcasts from different ensemble members within each model are sometimes used to create a probability distribution to assess hindcast uncertainty. This method of using probabilistic skill measures is not used in this thesis. There are two main reasons for this. Firstly, the main users of the results of this thesis are the Kenyan Meteorology Department. The only form of probabilistic skills measure that they use operationally is tercile probabilities. This shows the probability of the forecast lying in the below-normal, normal, and above-normal categories (Barnston et al. 2003). It was felt that restricting the results of this study to the form of tercile probabilities would limit the depth of verification of the hindcast results. Also, the author felt that it was important to use different methods to those used by Diro et al. (2008) at The University of Reading. They used probabilistic skill measures to assess the seasonal rainfall hindcast skill of their hindcast models for different regions of Ethiopia. Using deterministic skill measures ensures that the work in this thesis remains original and innovative.

4.2.3 Rank correlation (r_{rank})

Correlation coefficients describe the degree of linear association, or correspondence in phase, between hindcasts and observations (Murphy, 1995). The Spearman rank correlation coefficient (r_{rank}) is used as a robust and resistant alternative to the more common Pearson product-moment correlation coefficient. r_{rank} is robust to deviations from linearity in a relationship, and is resistant to the influence of outliers. Since gridded seasonal rainfall data over Africa often have non-normal distributions and contain outliers, it is more appropriate to use r_{rank} . The rank correlation is bounded by values of -1 and 1, where 1 (-1) represents a perfect positive (negative) linear association between the rankings of the observations and hindcast data. Spearman rank correlation is computed using the following equation from Wilks (2006):

$$r_{rank} = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)} \quad \text{Equation 4.1}$$

where D_i is the difference in ranks between the i^{th} pair of data values for a sample size n . If a particular datum value appears more than once, resulting in tied rankings, all of the equal values are assigned their average rank before computing the D_i values.

4.2.4 Mean Square Skill Score (*MSSS*)

The *MSSS* is the skill metric recommended by the World Meteorological Organisation (WMO) in assessing deterministic seasonal hindcast skill (WMO, 2006). The *MSSS* (Equation 4.2) is the percentage reduction in mean squared error provided by the hindcasts (*MSE*, Equation 4.3) over that provided by using climatology as the hindcast (*MSE_{clim}*, Equation 4.4). *MSSS* is calculated using the following equations from Wilks (2006):

$$MSSS = 1 - \frac{MSE}{MSE_{clim}} \times 100\% \quad \text{Equation 4.2}$$

The *MSE* is calculated using the equation:

$$MSE = \frac{1}{n} \sum_{k=1}^n (y_k - o_k)^2 \quad \text{Equation 4.3}$$

where y_k and o_k are the k^{th} of n pairs of hindcasts and observations. The *MSE* of the climatological reference hindcast (*MSE_{clim}*) is calculated using:

$$MSE_{clim} = \frac{1}{n} \sum_{k=1}^n (\bar{o} - o_k)^2 \quad \text{Equation 4.4}$$

where \bar{o} is the climatological rainfall average over the period 1959-2000.

4.2.5 Percentage of hindcasts with correct anomaly sign

A simple skill measure which is readily understandable by potential users is the percentage of years that the hindcasts correctly predict the sign of the rainfall anomaly. This is computed by noting the number of years that the hindcast rainfall and observed rainfall are both either above median or below median in the same year. The percentage of ‘successful’ hindcast years is then calculated.

4.2.6 Statistical significance

The significances (p -values) of the hindcast seasonal rainfall skills for the period 1959-2000 are computed by grid square. The significances of the three skill measures are computed as follows.

The significances for r_{rank} are computed by randomly shuffling the 42-year time series of hindcast and observed seasonal rainfalls and selecting, with replacement, a number of hindcast and observed rainfall pairs. This number of pairs corresponds to the number of degrees of freedom after correction for serial correlation (Davis, 1976; Chen, 1982) in the unshuffled time series. This process, sometimes known as *bootstrapping* (Efron and Tibshirani, 1993), is repeated 10,000 times per grid square. The r_{rank} from each random set is calculated and the results are displayed in histogram form to give the number of random sets with an r_{rank} greater than the original r_{rank} . This number, divided by 10,000, is the skill p -value. A low p -value therefore indicates a low likelihood that the hindcast skill arose from chance alone.

The significances for $MSSS$ are calculated and defined in a way similar to those of r_{rank} except that a p -value < 0.05 means there is less than a 5% likelihood that the hindcast skill is negative. The p -values are computed by randomly shuffling the 42-year time series for hindcast and observed seasonal rainfall and using the above *bootstrapping* method, selecting a number of pairs corresponding to the total number of years in the time series. The $MSSS$ (%) for each of the 10,000 random sets is then calculated. The p -value is the number of random sets that have a negative $MSSS$ skill, divided by 10,000. A low p -value indicates that there is positive $MSSS$ skill to a high confidence.

The significances for the percentage of years that the DEMETER hindcasts correctly predict the sign of the rainfall anomaly are calculated using *Fisher's Two-Sided Exact Probability Test* (Fisher, 1925). This test involves creating a 2x2 contingency table of the number of years where the hindcast and observation are both above or below median for each grid square. The Fisher probability test gives the p -value that this contingency table could have arisen by random chance. A p -value < 0.05 means there is at least 95% likelihood that the number of correct signed hindcasts did not occur by random chance.

4.3 Results

This section summarises the skill assessment results for the seasonal rainfall hindcasts over Africa from DEMETER. The focus is on the ASO and NDJ seasons where moderate skill is found a lead 0 (Sections 4.3.1 and 4.3.2). Results are not displayed for the FMA and MJJ seasons as little skill was found. The lead 3 skill is assessed in Section 4.3.3, for the two regions and seasons with the highest skill at lead 0.

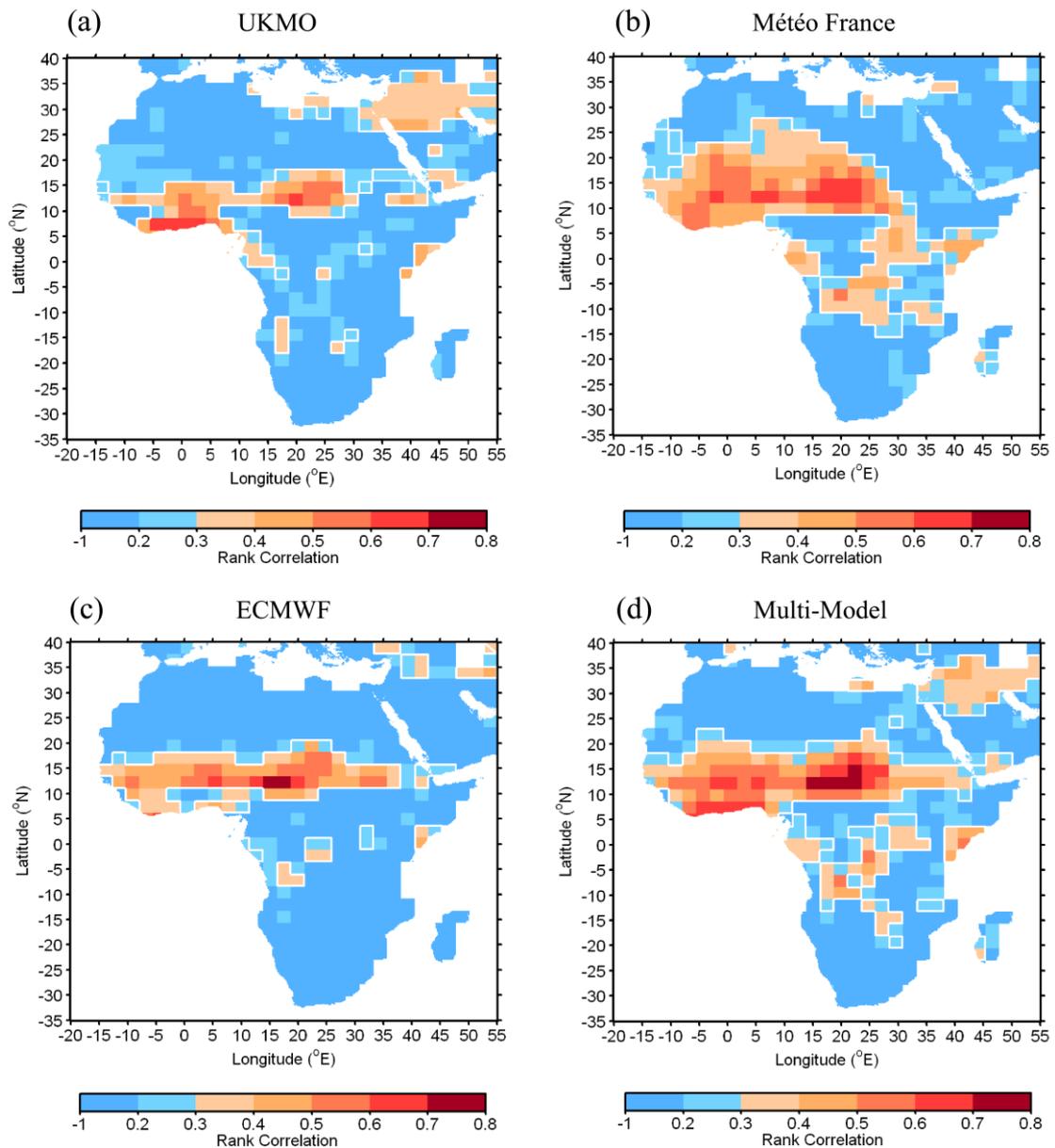


Figure 4.1. Seasonal rainfall hindcast skill for ASO at lead 0 from the three individual DEMETER models and the multi-model ensemble. The skill measure is Rank Correlation (r_{rank}) and the verification data are from the GPCP rainfall data set. The panels show r_{rank} for (a) UKMO hindcasts (1959-2000), (b) Météo France hindcasts (1958-2000), (c) ECMWF hindcasts (1958-2000) and (d) Multi-model ensemble hindcasts (1959-2000). White lines denote areas with p -values < 0.05 .

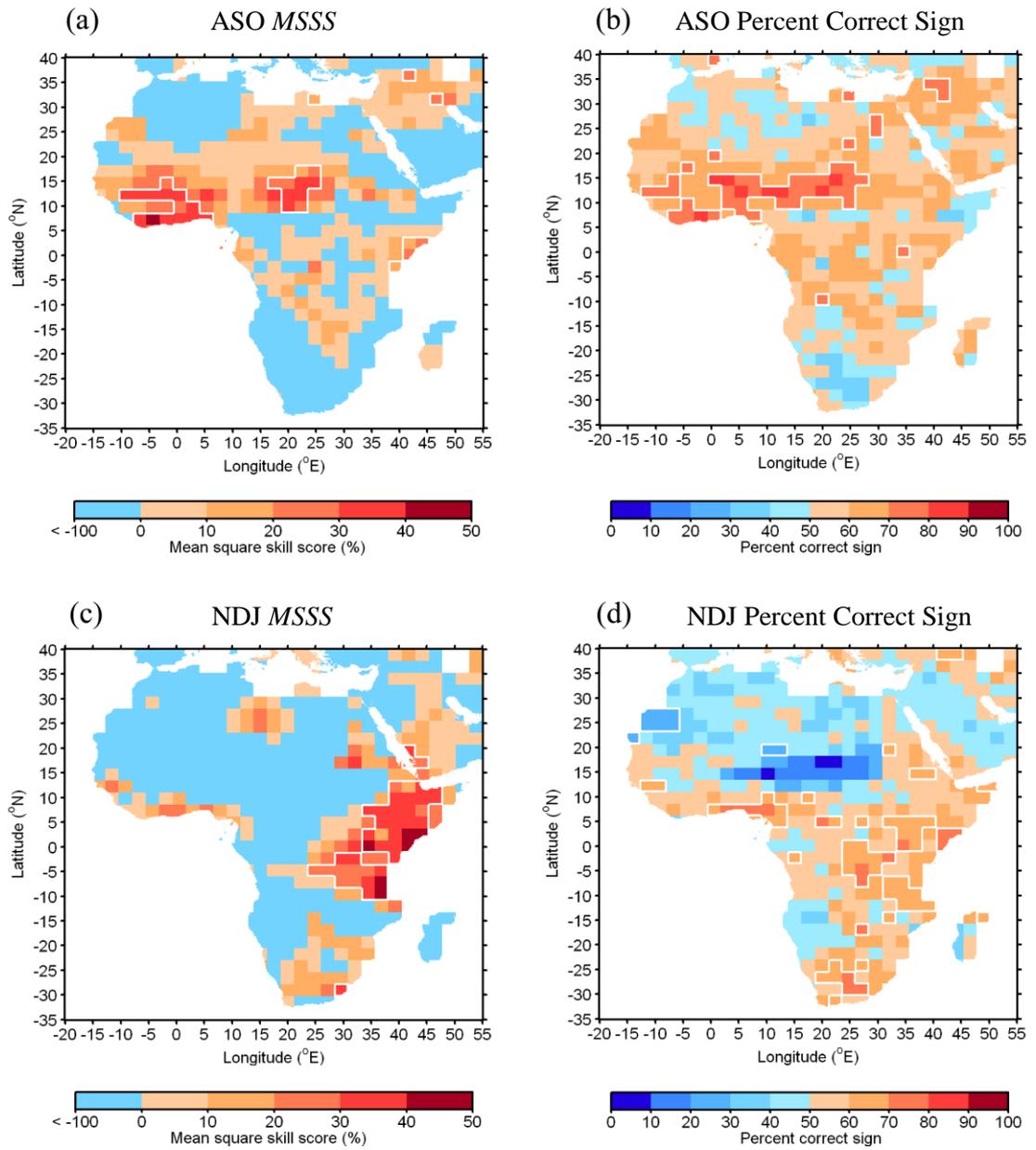


Figure 4.2. DEMETER multi-model ensemble seasonal rainfall hindcast skill for ASO and NDJ at lead 0; based on *MSSS* (%) and percentage of correct anomaly sign measures over the period 1959-2000. The white lines in panels (a) and (c) denote areas with p -values < 0.05 . The white lines in panels (b) and (d) denote areas where the two-tailed probability of obtaining the displayed percentage by random chance is < 0.05 .

4.3.1 ASO skill in the Sahel

The DEMETER seasonal rainfall hindcast skill for ASO over Africa at lead 0 is displayed in Figures 4.1 and 4.2 (a and b). Figure 4.1 shows the hindcast skill based on r_{rank} with panels (a), (b), (c) and (d) displaying r_{rank} for the UKMO, Météo-France, ECMWF and multi-model DEMETER hindcasts respectively. Figures 4.2a and 4.2b show the multi-model seasonal rainfall hindcast skill for ASO based on the *MSSS* and percent of hindcasts with correct anomaly sign measures. White lines denote areas with p -values < 0.05 .

Little or no skill is found over most of Africa, with the exception of the sub-Saharan/Sahel belt (7.5°N-20°N, 30°E-10°W) where significant moderate skill is seen in all skill measures. Visual inspection of Figure 4.1 shows that the multi-model (panel d) has the greatest overall r_{rank} skill, with significant values of 0.3-0.8 extending across the entire sub-Sahara belt. The highest r_{rank} of 0.7-0.8 occurs in central Chad for the multi-model's hindcast. Figures 4.2 (a and b) show a similar zone of elevated ASO rainfall hindcast lead 0 skill across the sub-Sahara. The *MSSS* peaks at values of 30-50% over southern Ghana, the Ivory Coast and along the Chad/Sudan border. The percentages of correct anomaly-signed hindcasts peak at 70-90% in a band stretching across southern Niger/northern Nigeria and central Chad. Thus in these regions the DEMETER multi-model hindcasts for ASO seasonal rainfall outperform a climatology rainfall hindcast in 70-90% of the years. The likelihood of achieving this skill by random chance is $< 5\%$.

4.3.2 NDJ skill in eastern and south-eastern Africa

The DEMETER seasonal rainfall hindcast skill for NDJ over Africa at lead 0 is displayed in Figures 4.3 and 4.2 (c and d). Figure 4.3 shows the skill based on r_{rank} in the same format as in Figure 4.1. Figures 4.2 (c and d) show the multi-model skill based on *MSSS* and percent of hindcasts with correct anomaly sign in the same format as in Figures 4.2 (a and b). As is the case for the ASO season, little or no hindcast skill is also found over most of Africa during the NDJ season. However, exceptions exist for central East Africa (Kenya, Uganda, Tanzania and Somalia), Nigeria and South Africa, where

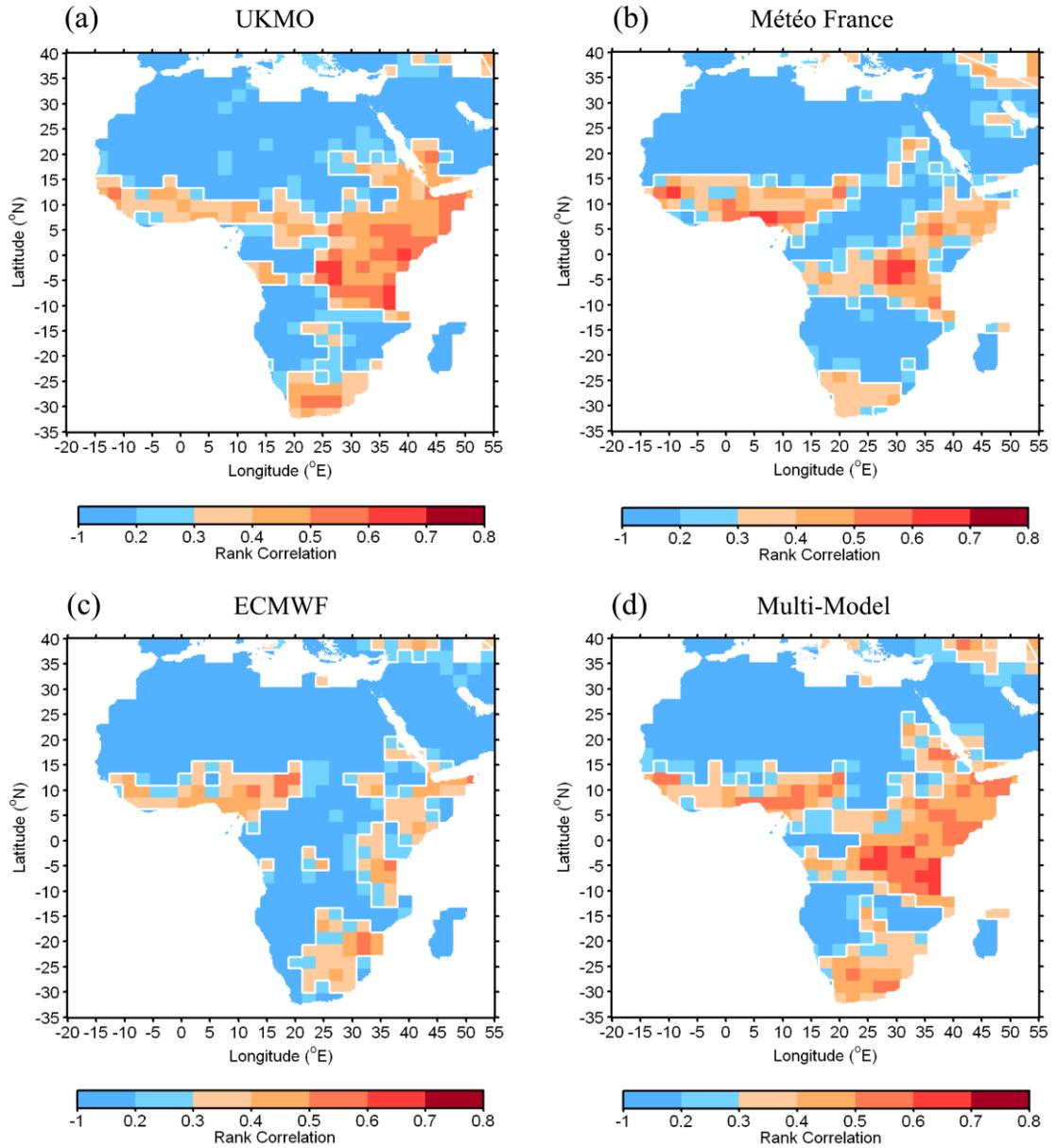


Figure 4.3. Seasonal rainfall hindcast skill for NDJ seasonal rainfall at lead 0 from the three individual DEMETER models and the multi-model ensemble. The skill measure is Rank correlation (r_{rank}) and the verification data are from the GPCP rainfall data set. The panels show r_{rank} for (a) UKMO hindcasts 1959-2000, (b) Météo France hindcasts 1958-2000, (c) ECMWF hindcasts 1958-2000 and (d) Multi-model ensemble hindcasts 1959-2000. White lines denote areas with p -values < 0.05 .

the hindcast skill is significant in all three measures (although this is marginal for *MSSS* in Nigeria). Inspection of Figure 4.3 shows the UKMO (panel a) and DEMETER multi-model (panel d) have the highest r_{rank} skill, with significant values peaking at 0.5-0.6 in Tanzania and the eastern Democratic Republic of Congo. The *MSSS* and percent of correct anomaly-signed hindcast skill measures also have significant peaks in these regions. The *MSSS* peaks at values of 20-40% over Kenya, Uganda, Tanzania and Ethiopia. The percent of correct anomaly-signed hindcasts peak at 60-70% in a zone stretching across central Africa and also in South Africa. Overall the DEMETER peak seasonal rainfall hindcast skill for NDJ is lower than that found for the ASO season.

4.3.3 Skill extension to lead of 3 months

The two regions that were found to have the highest, significant seasonal rainfall hindcast skill for ASO and NDJ at 0-month lead are examined further to see whether this skill extends to a lead of 3-months. A 3-month lead is examined as the quarterly initialisation dates of the DEMETER hindcasts (Section 4.1.2) preclude examination of the ASO and NDJ seasons at leads of 1- and 2-months. During the ASO season the region of greatest lead 0 skill selected for further examination is the Sahel region (10°N-20°N, 10°E-10°W) as defined in Nicholson (2005) and comprising Mali, East Mauritania, West Niger and Burkina Faso. The region of greatest lead 0 skill selected for further examination during the NDJ season is an area of East Africa (10°S-0°N, 30°E-40°E), comprising Tanzania and Southern Kenya. The *MSSS* measure is used for this 3-month lead seasonal rainfall hindcast skill assessment. *MSSS* values are optimised for the two regions by taking a regional average of the hindcast seasonal rainfall then calculating the *MSSS* for each region. Figure 4.4 shows the change in the DEMETER models' seasonal rainfall hindcast skill as the lead increases from 0 to 3 months. It is clear in both ASO and NDJ cases that where moderate hindcast skill exists at lead 0 this skill disappears by lead 3.

In this examination of hindcast skill at 3-month lead, the *MSSS* is computed for the three individual DEMETER models and for three different multi-model ensembles. The multi-models are distinguished by the use of different methodologies to compute the weighting of each individual model. The skills and confidence intervals of the multi-models are sensitive slightly to how the individual models are weighted to produce the

multi-model, as shown by the differences in skill between each multi-model's hindcasts in Figure 4.4. Multi-model 1, marked by unfilled, red circles in Figure 4.4, is formed by weighting the three individual models by *MSSS*. In this case any model with a negative *MSSS* has zero weighting, as a negative *MSSS* indicates no skill. Multi-model 2 uses the simplest method, involving a straight average of the three constituent models with no weighting. This method, used by the ECMWF, consistently produces the lowest multi-model hindcast skill, as shown by the small, filled red circles in Figure 4.4. The final method involves weighting the three constituent models by mean squared error (*MSE*). This is represented by the large, filled red circles in Figure 4.4. Weighting by *MSE* has been selected as the preferred method to compute the weighting for the multi-model ensembles in the other figures and results in this chapter. This weighting method is selected because: it offers improved skill over the straight average; it does not exclude any of the models as the *MSSS* method may do; and it generally produces the tightest 95% confidence intervals, as shown by the error bars in Figure 4.4.

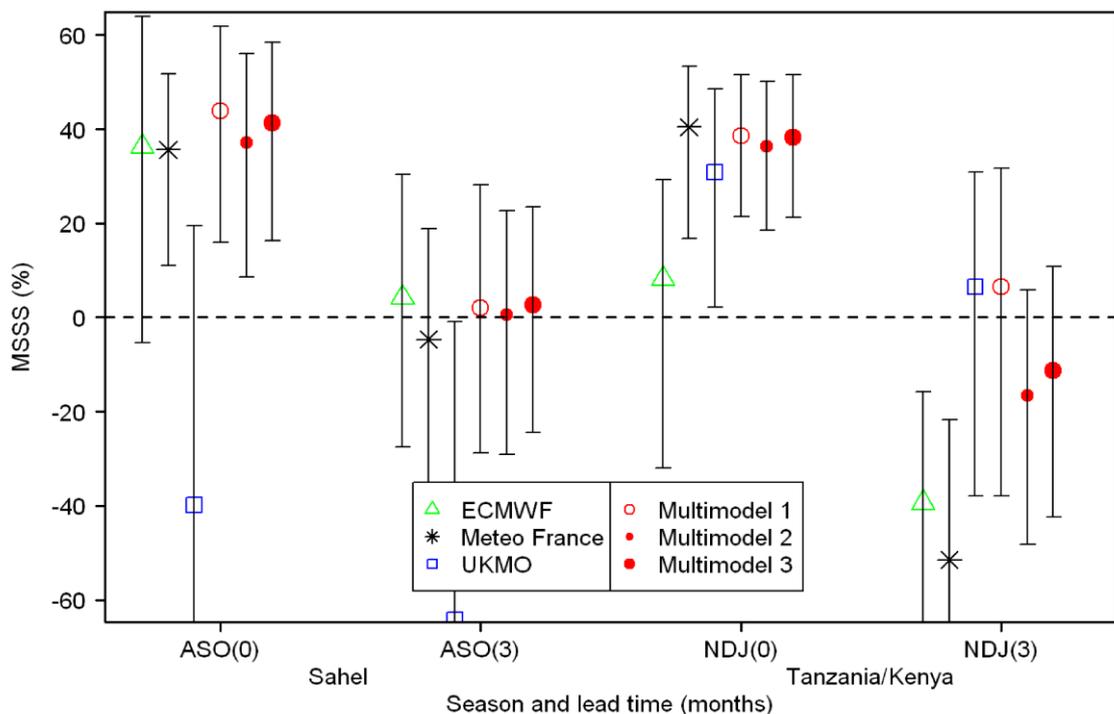


Figure 4.4. DEMETER seasonal rainfall hindcast skill for the Sahel (ASO) and East Africa (NDJ) at leads of 0 and 3 months. *MSSS* (%) is the skill measure and the verification data are from the GPCP rainfall data set. Hindcast skill is shown for the three individual DEMETER models and for three multi-models. The latter are distinguished by different weightings as follows: (1) weighted using *MSSS* with negative *MSSS* not contributing, (2) a straight average using no weighting and (3) weighted using *MSE* (this is the weighting used in all other figures and results). Error bars show the 95% confidence intervals on skill.

The 95% confidence intervals included around the deterministic *MSSS* values in Figure 4.4 are computed using the bootstrap method (Efron and Tibshirani, 1993; Lloyd-Hughes et al., 2004). This involves randomly selecting with replacement, 42 years of observed rainfall data together with the associated hindcast and climatology. The *MSSS* skill is computed and the process is repeated 10,000 times to obtain a distribution of skill values from which a 95% two-tailed confidence interval is readily obtained. This confidence interval means there is 95% probability that the skill computed over the 42 year period will lie within this uncertainty window.

4.4 Conflicting skill results

Conflicting skill results occur sometimes. For example, there are grid squares with a large positive r_{rank} , but a negative *MSSS* skill value. Figure 4.5(a) explains how this is possible using, as an example, NDJ 0-month lead rainfall hindcasts for a grid square in southeast Chad. Here $r_{rank} = 0.55$ and $MSSS = -3\%$. The black dashed line shows the squared difference between the multi-model hindcast rainfall anomalies (red line) and the GPCC observed anomalies (light blue line). Large values of the squared difference in two years (1985 and 1997) give rise to a negative *MSSS*. However, r_{rank} remains moderately positive because the hindcast and observed rainfall time series are generally in phase. Thus the conflicting skills arise due to two poorly hindcast years and the larger impact this has on *MSSS* than on r_{rank} .

The origin of the conflicting skill values in the above example is examined further in Figure 4.5(b). This presents the contribution of each of the individual models to the multi-model ensemble time series shown in figure 4.5(a). Of the three individual models, the UKMO (dark blue line) is the model that over- and under-predicts the magnitude of seasonal rainfall anomalies to the greatest degree. However, when you compare the magnitudes of the single models in figure 4.5 (b) to that of the multi-model ensemble in figure 4.5 (a) the importance of the weighting can easily be seen. The multi-model ensemble, weighted by *MSE*, has given almost zero weighting to the UKMO model and its magnitude is instead similar to the models that are closest to the GPCC verification data set in each year.

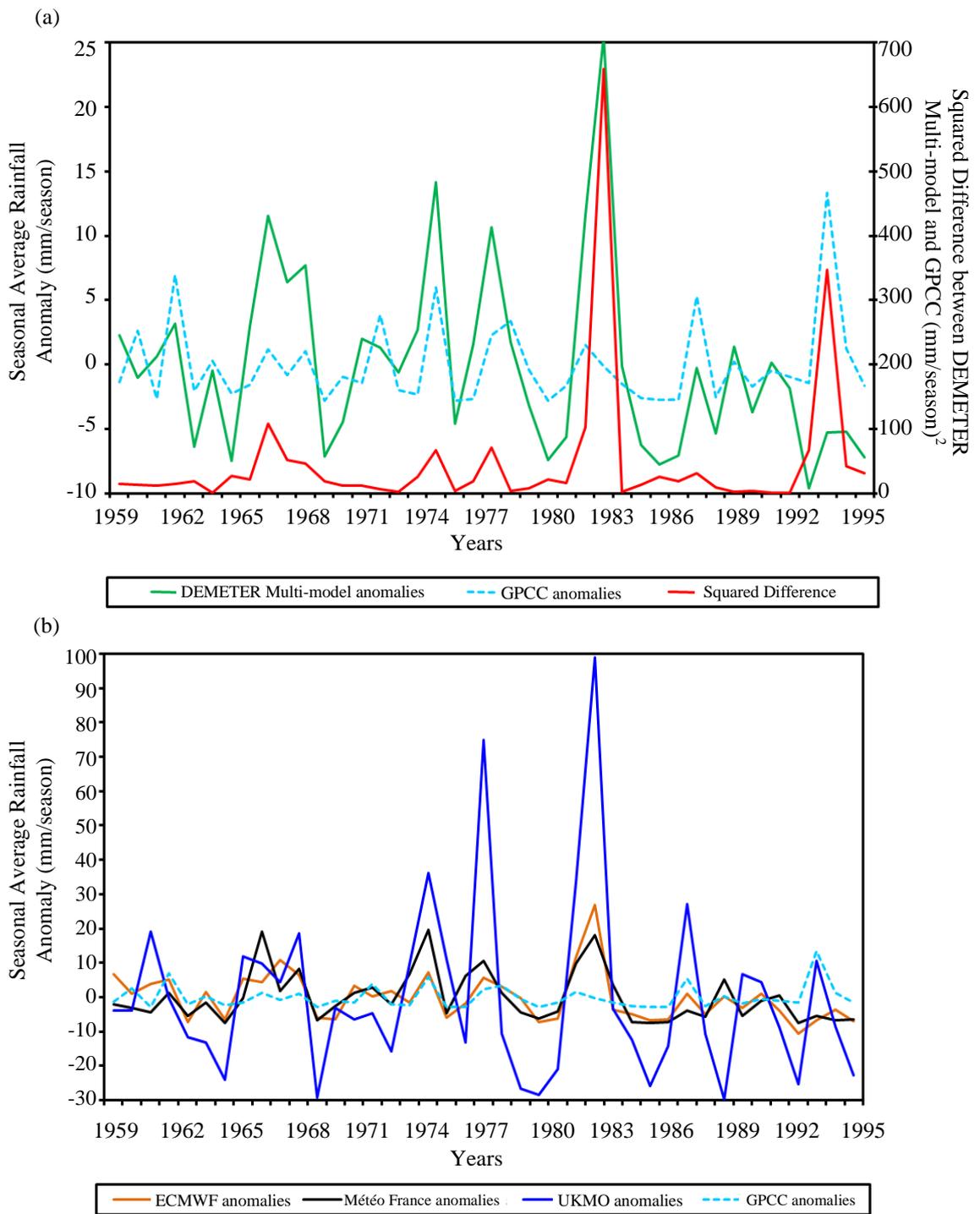


Figure 4.5. Investigation of the source of conflicting seasonal rainfall hindcast skill values for NDJ over the period 1959-2000 for a grid square in southeast Chad (centred at 17.5°E, 10°N). Panel (a) shows that the squared difference between the DEMETER multi-model ensemble and GPCC anomalies is large in a few years thereby causing MSSS to be negative. Panel (b) shows the hindcast anomalies from the three individual DEMETER models.

4.5 Verification from the DEMETER website

The DEMETER website (<http://www.ecmwf.int/research/demeter/d/charts/verification>) displays a general overview of the performance of the DEMETER multimodel and its comprising models. The performance assessment includes a suite of verification scores for deterministic hindcasts of rainfall in the form of global maps and bar charts averaged over a set of regions. The set of regions available over Africa are not coincident with the regions discussed in this Chapter, therefore this section will focus on assessing the global verification maps. It is difficult to perform a direct comparison between the verification assessments available on the DEMETER website and the results from this Chapter. This is mainly because the verification maps on the DEMETER website are produced for the seasons: March-May (MAM), June-August (JJA), September-November (SON) and December-February (DJF). These are all at 1-3 months lead, rather than the 0-2 months lead used in this study.

Figure 4.6 shows the African section of the seasonal rainfall hindcast skill maps available from the DEMETER website, showing the *MSSS* over the period 1979-2001. It can be seen that the majority of Africa has negative *MSSS* over each season (shown in red to yellow). There are however some areas of positive *MSSS* (shown in light grey).

The main areas of positive *MSSS* from the DEMETER website are found in: Figure 4.6 b) Kenya and Tanzania during JJA, Figure 4.6 c) Kenya during SON and Figure 4.6 d) the Sahara during DJF. These broadly agree with the main areas of positive *MSSS* found in this study. Figures 4.2 a) and c) show the *MSSS* across Africa for the ASO and NDJ seasons. Positive *MSSS* values are found over Kenya and Tanzania during both of these seasons, as well as in the sub-Saharan/Sahel belt and parts of the Sahara during ASO. However, in no positive *MSSS* values are found for the FMA and MJJ seasons across the whole of Africa in this study.

The positive *MSSS* values shown in figure 4.6 are much lower than those found in the rest of this Chapter. This is mainly due to the 1-3 month lead time used for the verification on the DEMETER website. Section 4.3.3 showed that the seasonal rainfall hindcast skill reduces dramatically with extended lead time. Using 0-2 month lead time in this thesis has shown the maximum skill available from the DEMETER multi-model

ensemble. It is also important to note that the multi-model ensemble shown on the DEMETER website is a straight average of the UKMO, Météo France and ECMWF models. This differs from the method used in this thesis of weighting the contribution from each individual model by the *MSE* of its hindcasts in order to maximise the skill of the multi-model ensemble.

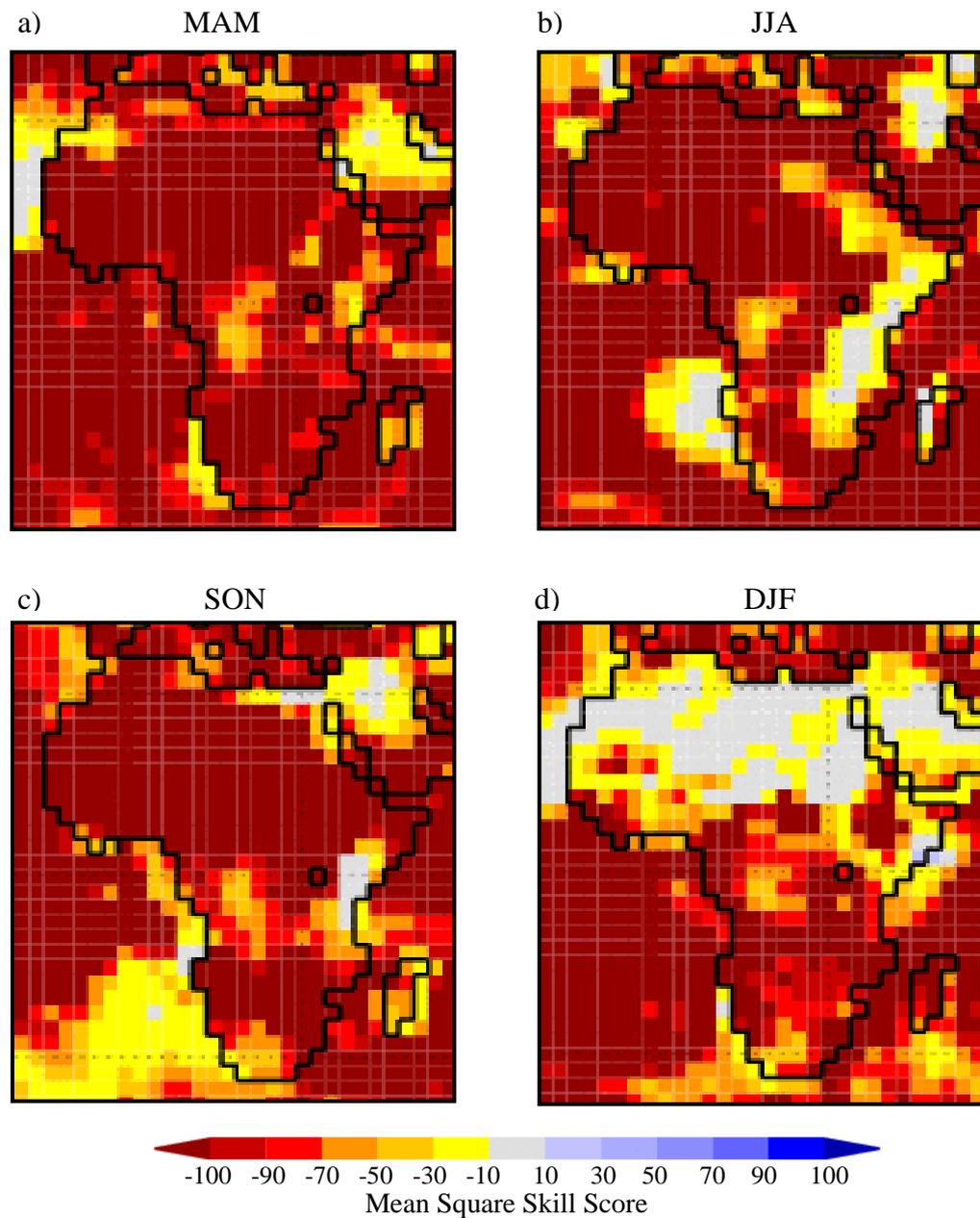


Figure 4.6. Seasonal rainfall hindcast skill for a) MAM, b) JJA, c) SON and d) DJF at lead 1 from the DEMETER multi-model ensemble over the period 1979-2001. The skill measure is Mean Square Skill Score (%), available from the DEMETER website (<http://www.ecmwf.int/research/demeter/d/charts/verification>).

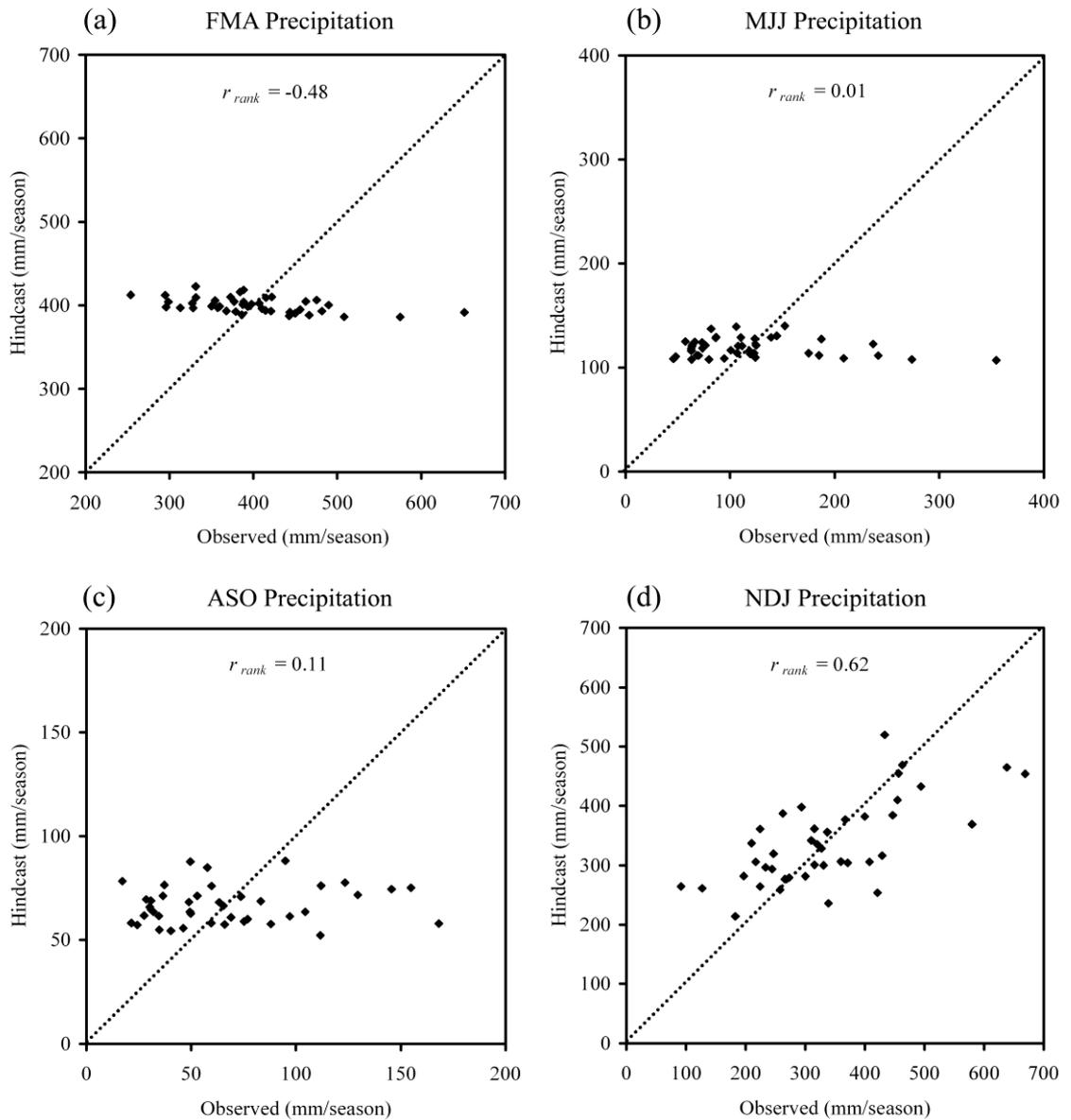


Figure 4.7. Investigation of the source of changing seasonal rainfall hindcast skill with each season from the DEMETER multi-model ensemble. The lead 0 skill measure shown is Rank correlation (r_{rank}) and the verification data are from the GPCC rainfall data set. The panels show scatter plots for seasonal rainfall for a grid square (centred at 37.5°E, 7.5°S) near Dar es Salaam, Tanzania over the period 1959-2000 for: (a) FMA (b) MJJ (c) ASO and (d) NDJ seasonal averages.

4.6 Discussion on lack of skill

With a few notable exceptions there is little seasonal rainfall hindcast skill over Africa offered by the DEMETER hindcasts for the period 1959-2000. Why is this? Many factors may be involved, with these varying with region and season across Africa. Since a full examination of factors is beyond the scope of this study attention is focused instead on one grid square centred at 37.5°E, 7.5°S, near Dar es Salaam, Tanzania to examine why hindcast failures occur at this location through the changing seasons.

Scatter plots of DEMETER multi-model lead 0 seasonal rainfall hindcast data against observed rainfall data for each of the seasons: (a) FMA, (b) MJJ, (c) ASO and (d) NDJ are displayed in Figure 4.7. The greatest r_{rank} is shown in Panel (d) for the NDJ season at 0.62. In this panel the variance of the hindcast values is the greatest at around 300mm/season and the points are positioned the closest to the line dotted line. If the points were positioned exactly along the dotted line, this would represent a perfect positive correlation ($r_{rank} = 1$) between the hindcast and observed data. In the other seasons (Panels a-c) the correlation is either low or negative. In these panels the hindcast data vary by less than 50mm/season, whilst the observed values vary by over 300 mm/season. The lack of variance in the hindcasts compared to that of the observations explains the low or negative correlation skill values in panels (a-c).

Another contributing factor for the lack of skill over Africa is the poor geographical representation of the continent within the models. Not only is the horizontal resolution very coarse, but some important, large geographical features are completely missing from the models. For example, Lake Victoria has a huge influence on the weather of the surrounding countries (Section 7.1.2). However, Météo France and the Met Office do not have a representation of Lake Victoria in their DEMETER models. The area is actually taken to be a land point (Culverwell, 2010 and Doblus-Reyes, 2010). ECMWF on the other hand treat Lake Victoria as part of the HOPE Ocean Model.

One final factor likely related to the poor hindcast seasonal rainfall skill is the sparse density of rain gauge stations over much of Africa. Figure 4.8 shows (a) the distribution of rain gauges over Africa and the r_{rank} for the (b) ASO and (c) NDJ seasonal rainfall at lead 0 from the DEMETER multi-model ensemble hindcasts. Figure 4.8 is a

compilation of Figures 2.4, 4.1(d) and 4.3(d) respectively. It is interesting to note that the regions with the greatest seasonal rainfall hindcast skill correspond to those areas with the densest network of observations; namely the sub-Saharan, South Africa and parts of East Africa. Where the station density is low, the hindcast skill is generally low. One final factor is that the seasons in which the greatest seasonal rainfall hindcast skills are found are often coincident with the rainy season in that region. For example, the ‘short-rains’ of East Africa fall from October-December and this region has the greatest seasonal rainfall hindcast skill in Africa during the NDJ season.

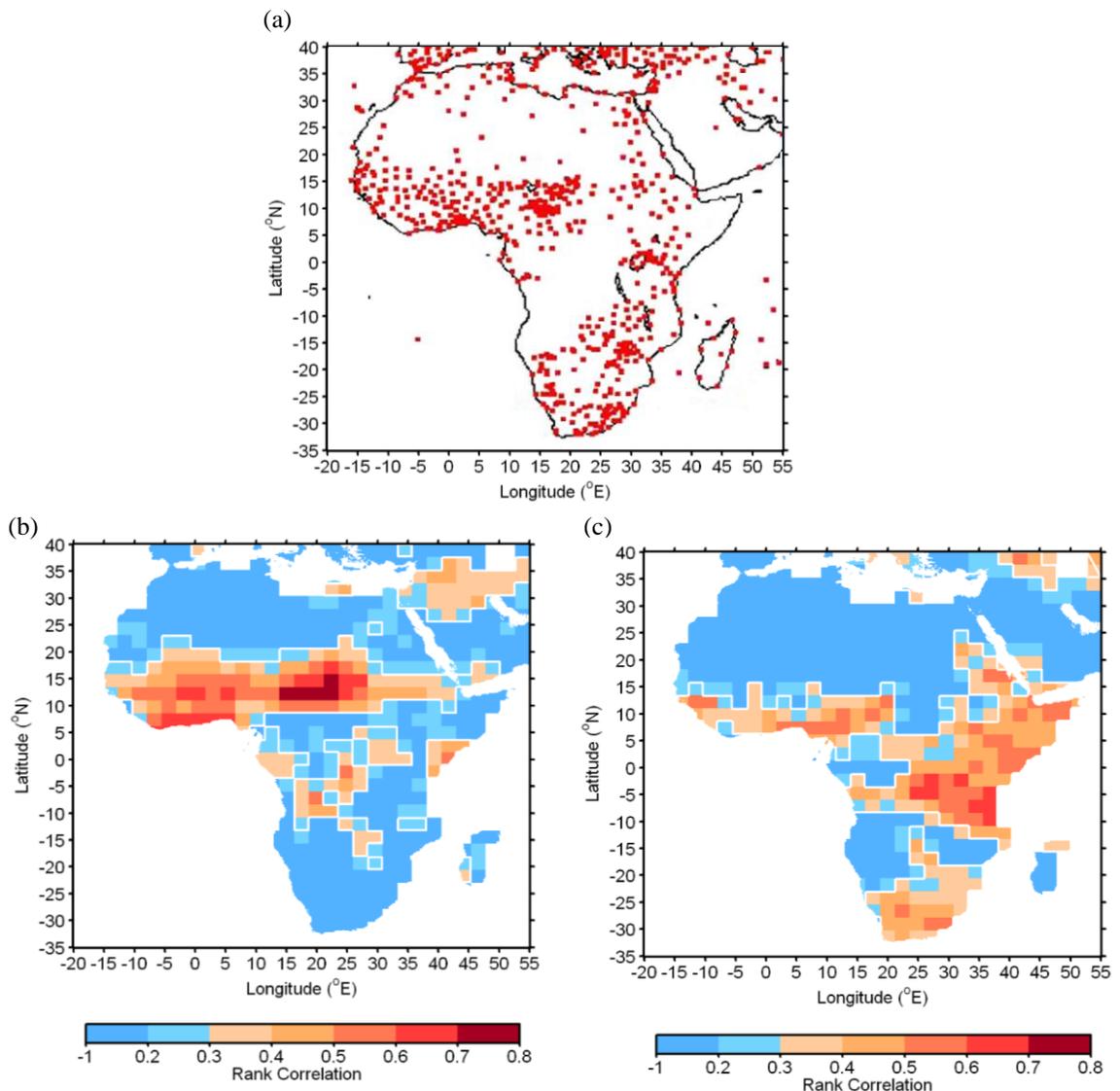


Figure 4.8. (a) Map of available rain gauge stations over Africa which provide input to the GPCP verification data set. These stations have at least 90% data availability between 1951 and 2000 (Courtesy of Beck et al., 2005). (b) Rank correlation for ASO seasonal rainfall at lead 0 from the DEMETER multi-model ensemble hindcasts 1959-2000 as shown in Figure 4.1(d). (c) Rank correlation for NDJ seasonal rainfall at lead 0 from the DEMETER multi-model ensemble hindcasts 1959-2000 as shown in Figure 4.3(d). White lines denote areas with p -values < 0.05.

4.7 Summary

The results of the skill assessment show that the DEMETER individual model and multi-model seasonal rainfall hindcasts have weak correlation with the observed GPCP rainfall data over most of Africa at zero lead for the period 1959-2000. However, moderately high, significant correlations are found during the ASO season at lead 0 over the southern Sahel region and for the NDJ season at lead 0 in the south-eastern Sahel region, and parts of eastern and southern Africa. The regions of high seasonal rainfall hindcast skill found from the DEMETER system cover a small area of Africa. This suggests that there is room for improvement in the development of dynamical forecast models for seasonal rainfall over Africa.

The regions of skill found in the multi-model ensemble hindcasts are coincident with regions and seasons that are affected by severe droughts. Therefore although there were only a small number of skilful areas found, they could be important results due to the humanitarian impacts that droughts cause in those areas. It is important to bear in mind the uneven distribution of rain gauges in Africa. The areas that show the highest DEMETER hindcast skill are coincident with those areas of Africa that have the highest concentrations of rain gauge stations. The DEMETER models seem to be able to best predict the seasonal rainfall during some rainy seasons. This may be because during the rainy seasons the dynamics are better represented over some of these regions of Africa as the dynamical links to global sea-surface temperatures are at their strongest.

Chapter 5

Methodology behind the Development of Statistical Seasonal Rainfall Hindcast Models

This chapter outlines the methodology used to develop regional statistical seasonal rainfall hindcast models for Kenya. The methods used to split Kenya into homogeneous rainfall regions are presented in Section 5.1. The techniques used to select predictors, develop seasonal rainfall hindcast models and to assess their skills are described in Sections 5.2-5.4.

It is important to define the usefulness of a forecast. The Kenyan forecasters need to make use of the forecast models to predict extended rainfall deficits. In this thesis the usefulness is defined by a combination of the skill assessments (Sections 4.2.3, 4.2.4 and 5.4), and the importance of the Kenyan regions in terms of population and the economy (Section 5.1.2). Hindcast models that are found to be skilful in Chapters 6-8 in the most populous and highly cultivated regions of Kenya will only be truly useful if they are used operationally to produce skilful forecasts in Kenya (Section 9.3).

5.1 Clustering the rain gauge data into homogeneous rainfall regions

Uneven rainfall distribution provides a problem when forecasting seasonal rainfall in Kenya, as it is often the case that just a few storms can produce the majority of a season's rainfall (Indeje et al., 2000; Indeje and Semazzi, 2000). The uneven spatial distribution of rainfall (Figure 2.5) is due to Kenya's complex topography (Figure 2.4), the effects of large nearby bodies of water such as the Indian Ocean and Lake Victoria, and to other regional factors (Indeje and Semazzi, 2000). Spatial rainfall averages are therefore necessary in order to reduce the influence of the uneven spatial distribution of the rainfall and to reduce noise in the data (Indeje et al., 2000). For this study Kenya is split into homogeneous rainfall regions. The monthly rainfall data from the rain gauge stations within each homogeneous rainfall region are then averaged together for each rainy season and used as regional seasonal rainfall indices.

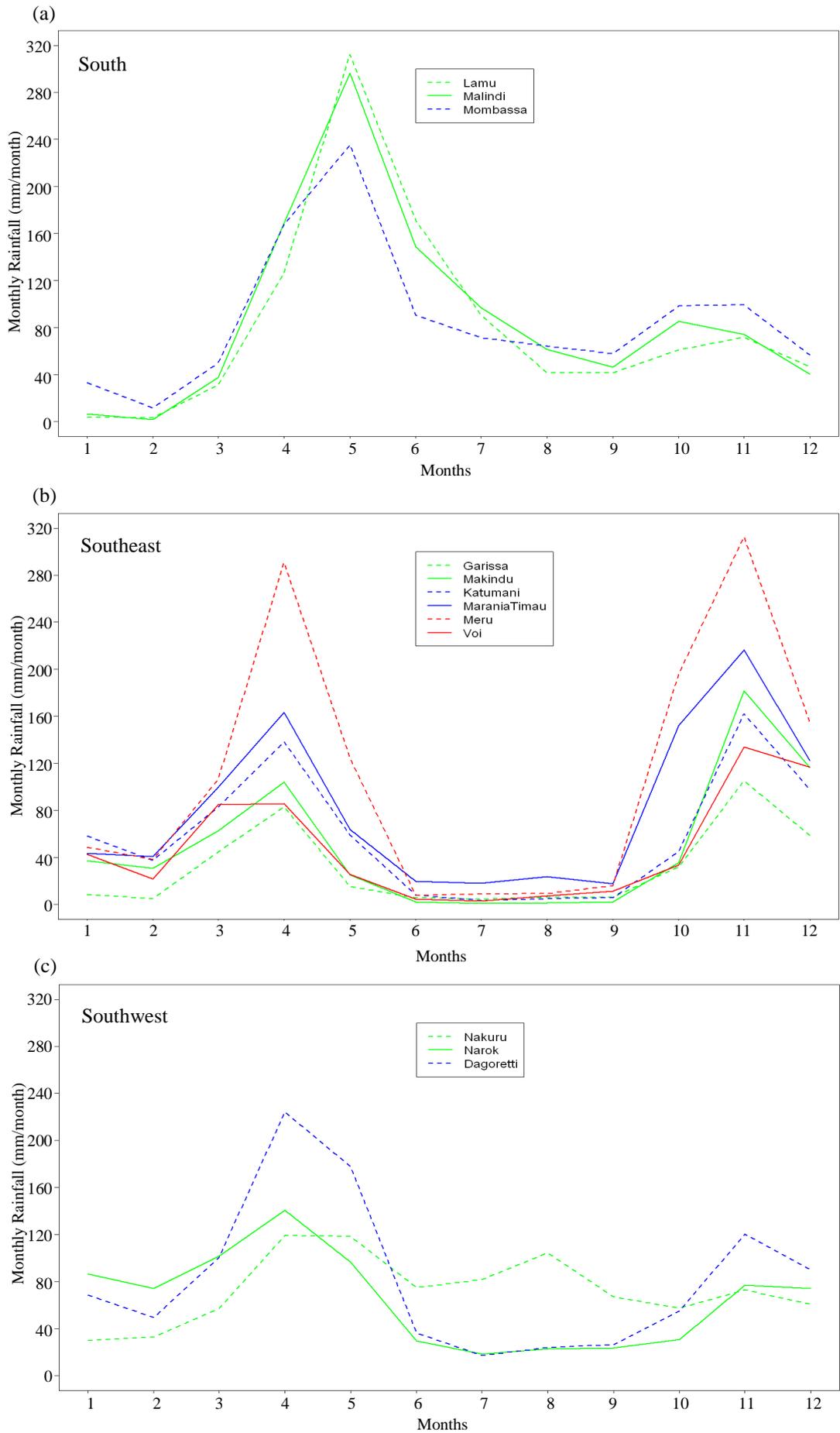


Figure 5.1. Part 1: Monthly rainfall climatology (averaged over the period 1959-2006) for each homogeneous rainfall region: (a) South, (b) Southeast and (c) Southwest.

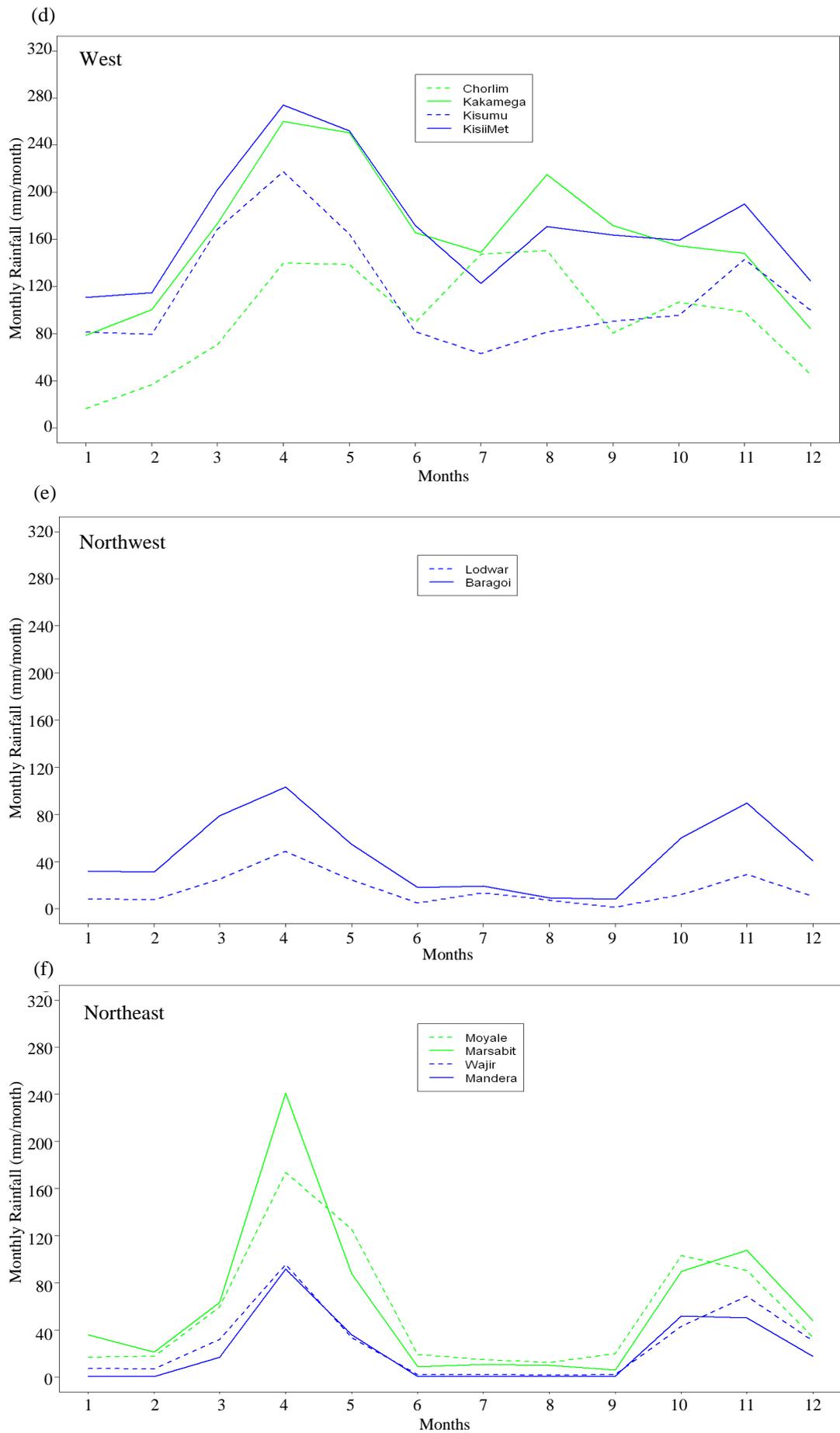


Figure 5.1. Part 2: Monthly rainfall climatology (averaged over the period 1959-2006) for each homogeneous rainfall region: (d) West, (e) Northwest and (f) Northeast.

5.1.1 Methodology

A four stage process is used to cluster Kenya's rain gauge stations into homogeneous rainfall regions.

1. Kenyan topography and geographic proximity of the rain gauge stations

Throughout the clustering process it is important to consider the topography of Kenya (Figure 2.4) and the geographic proximity of the rain gauge stations (Figure 3.1) (Gissila et al, 2004). The escarpments of the Great Rift Valley act as a natural meridional boundary, separating the fertile highlands of the Rift Valley in the southwest from the less fertile lowland regions of the southeast. There is also a natural zonal boundary between the arid region of northern Kenya and the semi-arid and fertile regions of central and southern Kenya. These natural boundaries will help to guide the position of the borders of the homogeneous rainfall regions.

2. Intra-annual rainfall variability

The intra-annual variability is assessed visually at each individual rain gauge station. The rainfall climatology by month (averaged over the period 1959-2006) for each of the 22 rain gauge stations employed in this study are shown in Figure 5.1. The monthly rainfall climatology for those stations located in the South (Panel a), Southeast (Panel b), Northwest (Panel e) and Northeast (Panel f) of Kenya vary in phase with each other. Further clustering analysis is necessary to group the remaining rain gauge stations.

3. Inter-annual rainfall variability

The next stage of the clustering process is a detailed examination of the inter-annual rainfall variability at each rain gauge station. The method is adapted from that used by Gissila et al. (2004), who calculated the cross-correlation between the seasonal average rainfalls from all of the rain gauge stations in Ethiopia. For each station of interest, the three stations with the highest cross-correlations are recorded. The user may then further cluster the stations with the strongest correlations, whilst also taking into account the results of the intra-annual variability assessment.

(Indeje, 2000). Indeje uses the regions formed using seasonal rainfall data (the purple dotted lines in Figure 5.2) throughout his thesis (2000) and published research on East African seasonal rainfall (Indeje et al., 2000; Indeje and Semazzi, 2000). It is important to note that Indeje’s homogeneous rainfall regions are designed to cover the whole of East Africa. Therefore, due to the difference in study area and rain gauge data employed in this study, the regional boundaries will be not be identical to those used by Indeje et al. (2000).

The boundaries of the homogeneous rainfall regions that will be employed throughout the rest of this study are shown in Figure 5.3. Further details on the exact position of the comprising rain gauge stations are presented in Table 3.1. The black lines in Figure 5.3 show the boundaries of the regions developed using the methods outlined above. The purple dotted lines show the boundaries of the regions developed by Indeje (2000) and Indeje et al. (2000) using seasonal rainfall data.



Figure 5.3. The black lines show the boundaries of the homogeneous rainfall regions, over Kenya, used in this study. The purple dotted lines show the regional boundaries from Indeje’s seasonal rainfall analysis shown in Figure 5.2 (Indeje, 2000 and Indeje et al., 2000). Map of Kenya from Online Maps (2009).

It is important to note that stages 1-3 (above) are conducted independently of stage 4. The clustering of the regions is not therefore influenced by the results of Indeje. However, Indeje's boundaries serve to guide the exact location of the new regional boundaries, to take into account local knowledge. Four main changes are made to the regional boundaries developed by Indeje (2000) and Indeje et al. (2000): (1) The West region is expanded to include Chorlim, as this is more strongly linked to the West region than to the Southwest; (2) The Southwest region is expanded to include the greater Nairobi area, so that this heavily populated area is not split into two regions; (3) The Northwest region is expanded to include Baragoi, as this is the most data sparse region of Kenya; (4) The boundary of the South region has been moved due to the low cross-correlations between Voi and the three rain gauge stations in the South region.

The monthly rainfall data from all stations within each homogeneous rainfall region are then investigated over each rainy season. This is to ensure that there is at least 50% monthly rainfall data availability across each region and season per year. If > 50% of the seasonal rainfall data in a given year are not available then the seasonal rainfall average for that region and year is set to NA.

The monthly rainfall data are averaged together over the 'long- and short- rains' seasons for each of the six regions to produce twelve regional seasonal rainfall indices for Kenya. These rainfall indices cover the period 1959-2006. Each rainfall index is split into three sub-periods of 16 years: 1959-1974, 1975-1990 and 1991-2006. The first two sub-periods are used as training periods to find potential predictors and to develop regional seasonal statistical hindcast models. Potential predictors are investigated separately for each rainy season, at time lags of zero to five months. The remaining sub-period is reserved to develop regional statistical forecast models over an independent verification period (Hastenrath, 1995; Mutai et al., 1998; Hastenrath et al., 2004).

5.1.2 Selected Kenyan rainfall regions

The following paragraphs provide details on the importance of each homogeneous rainfall region (Figure 5.3) in terms of: settlement, agriculture and topography. Details on land use and topography are taken from Figures 2.6(b) and 2.4 respectively. Population data from 2000 with estimated population data for 2015 are presented in Table 5.1 (CIESIN, 2005).

Region	Population (millions) 2000	Population (%) 2000	Population (millions) 2015	Population (%) 2015
South	0.95	3.3	1.3	3.4
Southeast	7.6	26.3	9.0	23.8
Southwest	11.1	38.1	15.6	41.4
West	8.6	29.6	10.6	28.2
Northwest	0.37	1.3	0.52	1.4
Northeast	0.42	1.4	0.68	1.8

Table 5.1. Population per homogeneous rainfall region. Measured in millions of people and percentage of total Kenyan population for 2000 and estimates for 2015. Population data from the Gridded Population of the World, Version 3 (GPW_{v3}), produced by CIESIN (Centre for International Earth Science Information Network) (CIESIN, 2005).

Southwest and West regions

The Southwest and West are the most important regions of Kenya in terms of population and economy. These regions produce the majority of Kenya's tea, coffee, cereal, cattle and sugar output. They are the most populous regions of Kenya, comprising 11.1 and 8.6 million people (38.1% and 29.6% of the total population) respectively. The Southwest region includes the 2.3 million people living in the capital city, Nairobi (CIESIN, 2009).

Over half of the land in the Southwest is semi-arid. The second largest land use is perennial cropland, which is cultivated with long term crops that do not need replanting after each harvest. Perennial crops contribute towards Kenya's economy and include: tea, coffee, vine, palm and banana. The next largest land uses are improved grazing (where cattle are raised), forests, arable cropland, arid zones and flooding zones. Arable cropland is cultivated with crops that need replanting after each harvest. Arable crops also contribute towards Kenya's economy and include: cereals, cotton and sugar. Land use in the West region is mainly perennial cropland, with a small area of forest to the

north. Both Southwest and West regions are located at high elevations; mainly over 500m, increasing to over 1.5km along the escarpments of the Rift Valley. The West region borders the fertile shores of Lake Victoria.

Figures 5.1 (c) and (d) show the annual rainfall time series from the Southwest and West regions respectively. Both regions receive moderate to large amounts of rainfall during the ‘long- and short-rains’ seasons, with peaks of 120-280mm/month in April and 80-200mm/month in November. The West region experiences a third rainy season from July to September, with a peak in August of 80-220mm/month.

Southeast and South regions

The Southeast is the third most populous region in Kenya with 7.6 million people (26.3% of total population). The majority of the land is semi-arid, with an exception in the west of the region. The Southeast region slopes down from > 1.5km in the west to < 100m in the east. The main form of livelihood for the population is nomadic pastoralism, where livestock are herded seasonally or continuously in order to find grazing land and water. This subsistence farming does not contribute to the Kenyan economy. In the western third of the Southeast region, around 50% of the land use is arable cropland, which contributes towards Kenya’s economy, with small areas of forests and wetlands. Figure 5.1 (b) shows the annual rainfall time series from the Southeast region. The region receives moderate to large amounts of rainfall during the ‘long- and short-rains’ seasons, with peaks of 80-300mm/month in April and November.

The South region borders the Indian Ocean, and comprises a narrow band of fertile land, extending up to 100km inland. This region has the third smallest population in Kenya with 950,000 people (3.3% of total population). Most of the land use is either semi-arid or wetland, however around a quarter of the land is used as arable cropland, thus contributing towards Kenya’s economy. Figure 5.1 (a) shows the annual rainfall time series from the South region. Large amounts of rainfall are received during the ‘long-rains’ season, with a later peak in May of 220-320mm/month. A lesser amount of rainfall is received during the ‘short-rains’ season, with a peak of 60-100mm/month in October/November.

Northwest and Northeast regions

The Northwest and Northeast are the least important regions of Kenya in terms of population and economy, comprising 370,000 and 420,000 people (1.3% and 1.4% of total population) respectively. They do not significantly contribute to the Kenyan economy, as the main form of livelihood is nomadic pastoralism. The land use is not conducive to farming as it is semi-arid or arid with some lake areas. Elevation varies between 50-300m, sloping down towards the southeast. Figures 5.1 (e) and (f) show the annual rainfall time series from the Northwest and Northeast regions respectively. Both regions receive rainfall during the ‘long- and short-rains’ seasons with peaks of 40-240mm/month in April and 20-110mm/month in November. The amount of rainfall received in these regions is generally far lower than in the rest of Kenya, especially in the Northwest, which receives < 100mm/month at its peak in April.

5.2 Predictor selection criteria

Potential predictors for seasonal rainfall in Kenya are sourced from the monthly data introduced in Chapter 3, namely: global SSTs, global 850hPa winds, Niño indices, the SOI, the DMI, the QBO index and the MJO index. The methods used to select potential SST predictors are based on those used by Diro et al. (2008) and Gissila et al. (2004) in their search for predictors for seasonal rainfall in Ethiopia. Potential wind predictors are selected using methods outlined in Saunders and Lea (2005). Note that in all cases actual values are used unless it is clearly specified that an anomaly is used instead.

The following sections outline the criteria used to identify potential predictor regions, which have strong (Section 5.2.1), temporally stable (Section 5.2.2) and significant (Section 5.2.3) correlations with the Kenyan regional seasonal rainfall indices. Section 5.2.4 describes the methods used to select potential wind predictor regions that have strong, temporally stable and significant links to the Kenyan rainfall indices. The data within these regions are then averaged together to form potential predictor indices.

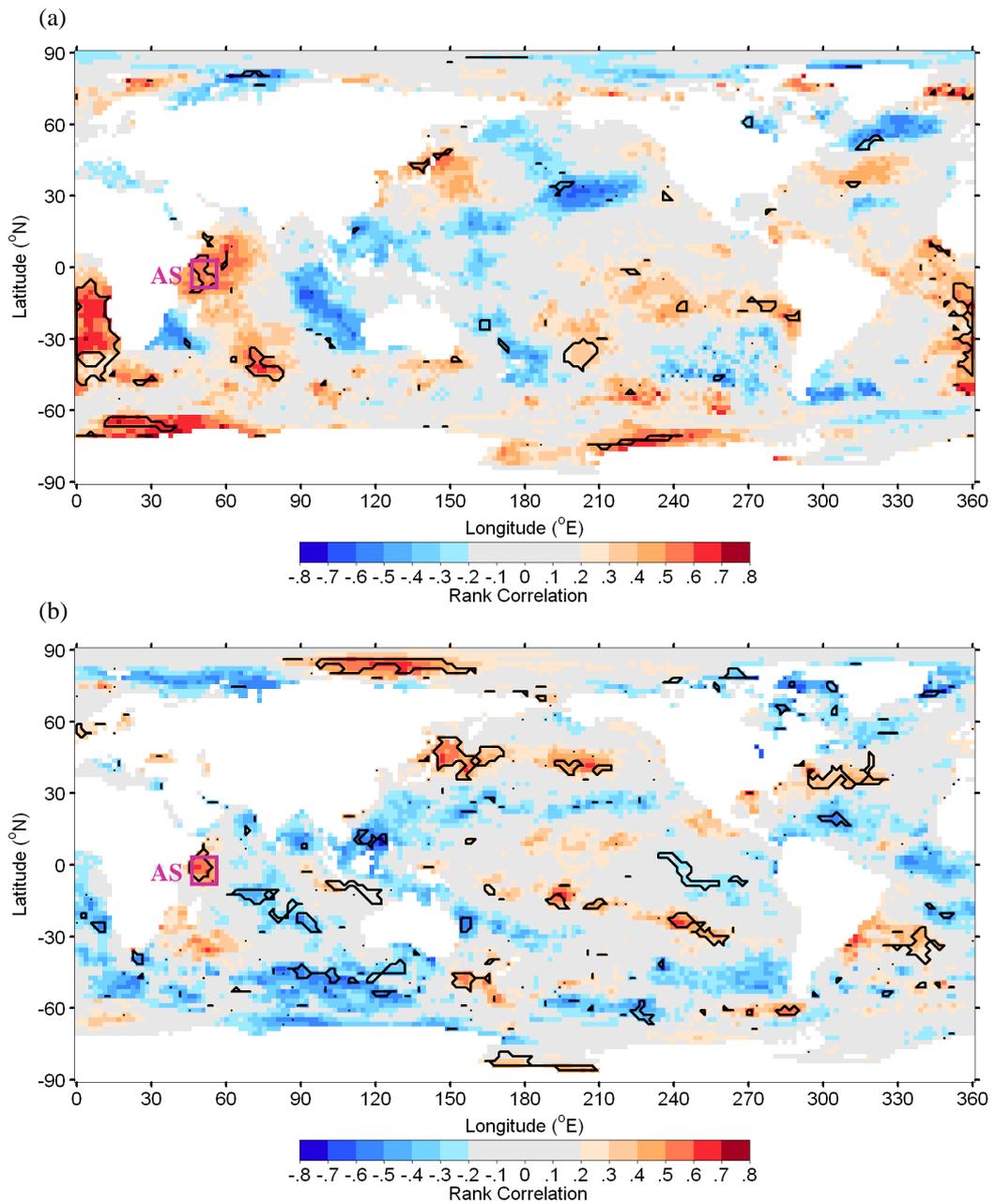


Figure 5.4. Example spatial correlation plots: Southwest regional OND seasonal rainfall index correlated with prior August-September (AS) two-month average SST global field over the periods: (a) 1959-1974 and (b) 1975-1990. White areas denote land mass. Red (blue) shows positive (negative) correlations, grey shows low correlations (between -0.2 and 0.2) and the black lines show areas with p -values < 0.15 . The AS potential predictor region for this example is highlighted by the purple box.

5.2.1 Correlation of seasonal rainfall and predictors

Regional statistical seasonal rainfall hindcast models are developed using linear regression techniques (Section 5.3). There must be a strong linear correlation between predictor and predictand in order for a linear regression model to have skill (Gissila et al., 2004). Potential predictors are selected through assessing their correlations with each of the twelve Kenyan regional seasonal rainfall indices (Section 5.1.1) at lead times of up to 5 months. The correlation assessment is repeated for each homogeneous rainfall region and rainy season. Potential predictors are correlated with the regional seasonal rainfall indices as individual monthly data and as two- and three-month averages in order to select the strongest potential predictors for each rainy season and homogeneous rainfall region.

In the cases when the potential predictors are in the form of indices, such as the Niño 3.4 index, a straight correlation can be calculated between the predictor index and the regional seasonal rainfall index over each training period. In the case of SSTs, when the potential predictor is in the form of a global gridded data field, potential predictor regions are selected through visual analysis of global correlation plots over each training period. These are produced by calculating the correlation between the regional seasonal rainfall index and the potential predictor data within each grid square across the globe.

5.2.2 Stability of predictors over time

Potential predictors need to have strong, temporally stable correlations with the regional seasonal rainfall indices. The training period is therefore split into two 16-year periods: (a) 1959-1974 and (b) 1975-1990. Figure 5.4 shows examples of spatial correlation plots over each of these 16-year training periods. The correlations are between the Southwest OND regional rainfall index and the prior August-September (AS) two-month average global gridded SST field. The purple box shows the region that has been selected as a potential SST predictor region due to the strong correlations that are stable over both half periods. The assessment of the correlation plots is performed visually, rather than by using a program to automatically select grid squares that are stable over time. If a strongly correlated grid square shifts spatially by even 1 grid square over the

two periods then the potential predictor region would be erroneously missed by an automatic selection program.

5.2.3 Significance of correlations

It is important that the potential predictors have a significant correlation with the regional seasonal rainfall indices. The significance threshold required for the correlation assessment has been set at a p -value < 0.15 over both half periods over most of the potential predictor region. To calculate the p -values per grid square for each of the hundreds of spatial correlation plots produced using the bootstrapping method (Chapter 4.2.6) is too computer and time intensive. Fisher's Two-Sided Exact Probability Test (chapter 4.2.6) is a less computer and time intensive method to calculate the p -values for the global spatial correlation plots.

The following compares the p -values calculated using Fisher's Two Sided Exact Probability Test and the bootstrapping method for a random example. The r_{rank} between the South cluster's MAM seasonal average rainfall index and the MAM Niño 4 index is found to be 0.45. The associated p -values are 0.03 and 0.05 calculated using Fisher's Two Sided Exact Probability Test and the bootstrapping method respectively. The similarity between these two results indicates that Fisher's Two Sided Exact Probability Test is a suitably stringent but less computer and time intensive method to calculate p -values. Fisher's Two Sided Exact Probability Test will therefore be used to calculate the p -values for the global spatial correlation plots throughout this study.

5.2.4 Potential wind predictor selection method

Potential 850hPa wind predictors cannot be selected through correlation analysis, as wind data are in vector form. They are therefore selected through visual analysis of global composite difference plots over each training period, based on the method used by Saunders and Lea (2005). Each composite difference plot shows the direction, magnitude and significance of the composite difference in wind anomalies for those subset years when the regional seasonal rainfall index is in its upper and lower quartiles. The statistical significance is calculated using the bootstrapping method as described in

Chapter 4.2.6. Regions of strong, temporally stable and significant wind vectors are linked to above median regional seasonal rainfall in Kenya. The average u -wind is taken over each potential predictor region as a potential predictor index. This is because the u -component of the wind is generally the most dominant within the potential wind predictor regions.

5.2.5 Physical link between regional rainfall and potential predictor indices

A dynamical link must exist between each potential predictor and its associated regional seasonal rainfall index. Potential predictors are therefore only selected from regions with known teleconnections with Kenyan rainfall, as presented in Chapter 2.3. The purple box in Figure 5.4, for example, shows the area of SSTs that have been selected as a potential predictor region for the OND rainfall in the Southwest region of Kenya. This potential SST predictor region, located off the coast of East Africa, has a strong, positive, temporally stable correlation with the Southwest OND rainfall index. Rising motion and convective activity are enhanced over this region due to the positive SST anomalies. Understanding the dynamical link between the regional rainfall variability and the SST anomalies gives us confidence to use this region as a potential predictor.

5.3 Formulation of statistical models

The following sub-sections outline the methodology used to develop regional statistical seasonal rainfall hindcast models for Kenya. For each season and region, hindcast models are first developed using a cross-validation method over three training periods (1959-1974, 1975-1990 and 1959-1990). These three periods are used to assess the temporal stability of the hindcast skill. The hindcast models are then applied in a replicated real-time sense to the independent verification period of 1991-2006. It is then possible to assess the forecast skill available from these models over this independent verification period.

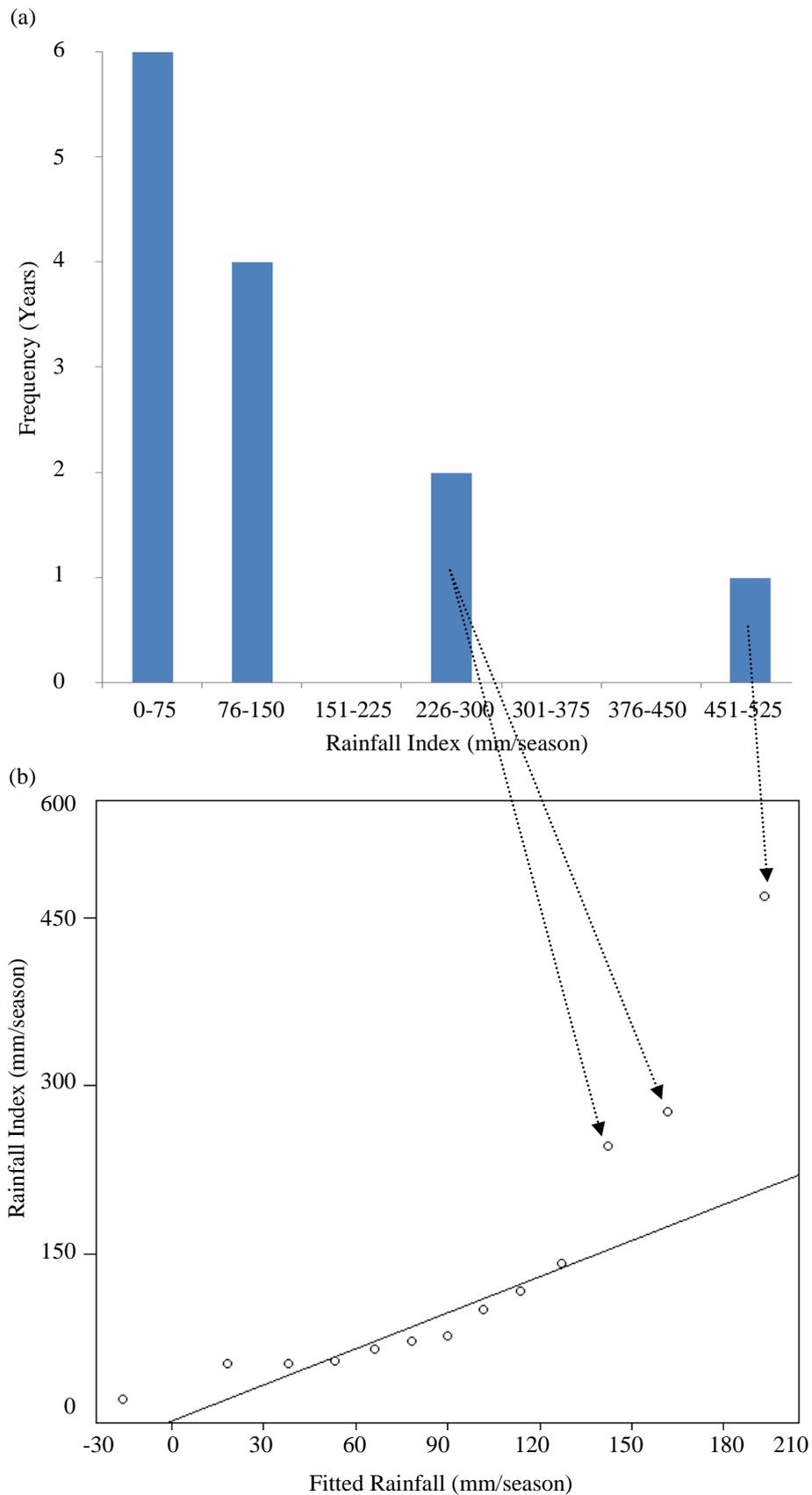


Figure 5.5. OND rainfall index from the Southwest region over the period 1959-1974. (a) Histogram of the rainfall index to highlight the non-normal distribution. (b) Quantile-Quantile (QQ) plot comparing the observed rainfall index to the theoretical normal distribution. The black diagonal line in (b) indicates the theoretical position of the points if the data were normally distributed. The black dashed arrows highlight the high rainfall years that are responsible for the large departure from normality. Note/ Rainfall data from the years 1960, 1964 and 1966 are not available for this example.

5.3.1 Test for normality of data

Seasonal rainfall hindcast models are developed for regions of Kenya using linear regression modelling (Section 5.3.2). In order to satisfy the validity assumptions for using linear regression, the residuals around the regression line (Section 5.3.2) need to follow a normal distribution (Wilks, 2006). To achieve this, the actual potential predictor and predictand indices are tested for normality before the linear regression models are developed. Any indices that do not satisfy the validity assumption are transformed to a normal distribution and re-tested for normality. Linear regression modelling is then performed on the transformed data to produce regional seasonal rainfall hindcasts (Section 5.3.2). The residuals around the regression line are then themselves tested for normality. Finally, the hindcast values are transformed back to their original dimensions before the hindcast skill is computed (Section 5.4).

Rainfall data often follow a Gamma Distribution (Wilks, 2006), which is positively skewed with an elongated tail to the right. This distribution is due to the physical constraint that rainfall must be positive and the fact that most days record low amounts with just a few high rainfall events (Wilks, 2006). Figure 5.5 (a) shows the distribution of the OND rainfall index from the Southwest region over the years 1959-1974. This is a typical example of a gamma distributed regional seasonal rainfall index.

Quantile-Quantile (QQ) scatter plot

The normality of the distributions of the predictor and predictand indices are tested separately using two techniques. Firstly, a Quantile-Quantile (QQ) scatter plot is produced to show how closely the normal distribution represents the distribution of the data (Figure 5.5 (b)). The location of each point is given by a coordinate pair consisting of an observed rainfall data value and the corresponding theoretical estimate for the data value derived from an empirical formula for the normal distribution. Figure 5.5 (b) shows a QQ plot for the example case of the OND Southwest rainfall index. The black diagonal line indicates the theoretical position of the points if the data were normally distributed. The curvature of the points therefore indicates the extent of the departure of the data from normality. It can be seen that the curvature of the data is greatest towards

the right of the plot. The black dashed arrows draw attention to the high rainfall years that contribute towards the departure from normality.

Histogram with fitted theoretical distributions

Further investigation is required for those potential predictor and predictand indices that show departures from the normal distribution. The Excel-based risk analysis software “@Risk” has a function that plots a histogram, superimposed with the theoretical distributions that best represent the data. This software calculates the Goodness-of-Fit of a selection of theoretical distributions to the data using the Kolmogorov-Smirnov (K-S) Test. An example plot is shown in Figure 5.6 for the OND Southwest rainfall index. The theoretical distribution that best fits the data in this example is found to be the gamma distribution (red line). The normal distribution (green line) is also included for comparison.

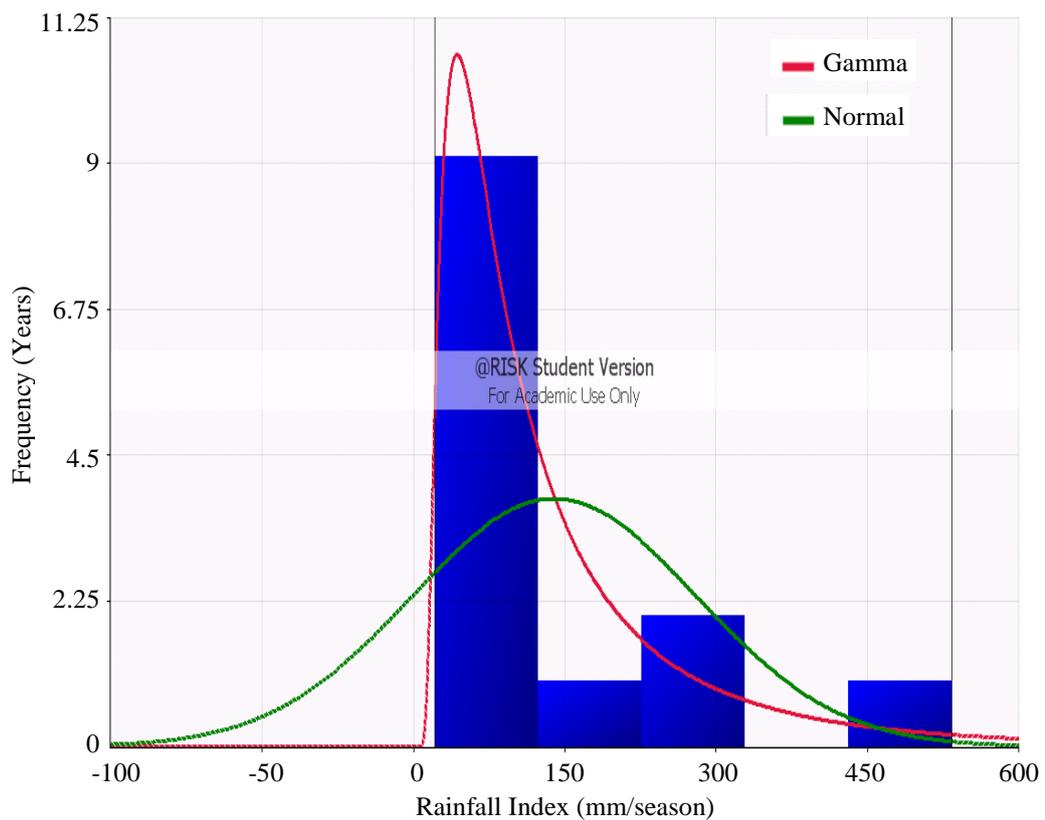


Figure 5.6. ‘@Risk’ plot showing the gamma distribution (red) that best represents the OND seasonal rainfall index from the Southwest region over the years 1959-1974. The green line shows the normal distribution for comparison. Note/ Rainfall data from the years 1960, 1964 and 1966 are not available.

Transformations

Each potential predictor or predictand index that shows departures from the normal distribution needs to be transformed to a normal distribution. In each case several transforms are first tested on the actual data (x) including: $\log(x)$, $\log(1+x)$, $\sqrt[3]{x}$ and the Box-Cox transformation. The Box-Cox transformation, given by:

$$T(x) = x^\lambda \quad \text{Equation 5.1}$$

may be used to transform positively skewed rainfall data to a more normal distribution. The Box-Cox power (λ) varies from 0 to 1 (Box and Cox, 1964) and is calculated using the `box.cox.powers` function in the statistical program “R”. This function calculates the logarithm of the likelihood function (LLF) for λ varying between 0 to 1. The λ that act to maximise the LLF is used as the λ in equation 5.1 (Box and Cox, 1964).

Each transformed data index is then re-tested for normality and the data that have the greatest resemblance to the normal distribution are used to develop the linear regression seasonal rainfall hindcast models. Figure 5.7 shows the results of the transformation for the example case of the OND Southwest rainfall index. The function used here to transform the data to the most normal distribution is given by:

$$T(x) = \log(1 + x) \quad \text{Equation 5.2}$$

where x is the OND Southwest rainfall index over the period 1959-1974. Figure 5.7 (a) shows that the transformed data are jointly best represented by the Log Logistical (red line) and normal (green line) distributions. Figure 5.7 (b) shows the QQ plot associated with these transformed values. The points now lie closer to the black diagonal line than before the transformation (Figure 5.5 (b)) and the large departure from normality has been removed.

Table 5.2 shows the functions that were used in this study to transform the rainfall indices to normal distributions. Normality tests showed that the potential predictor indices were already close to a normal distribution. No transformations were therefore necessary for these data.

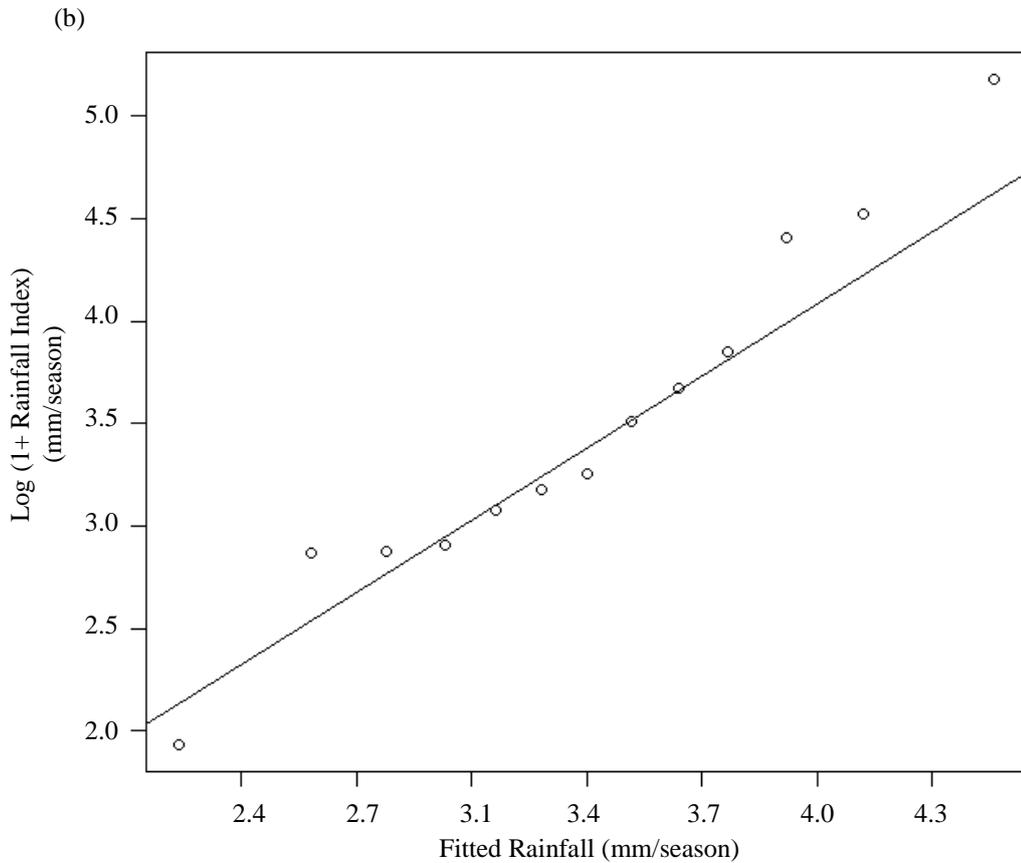
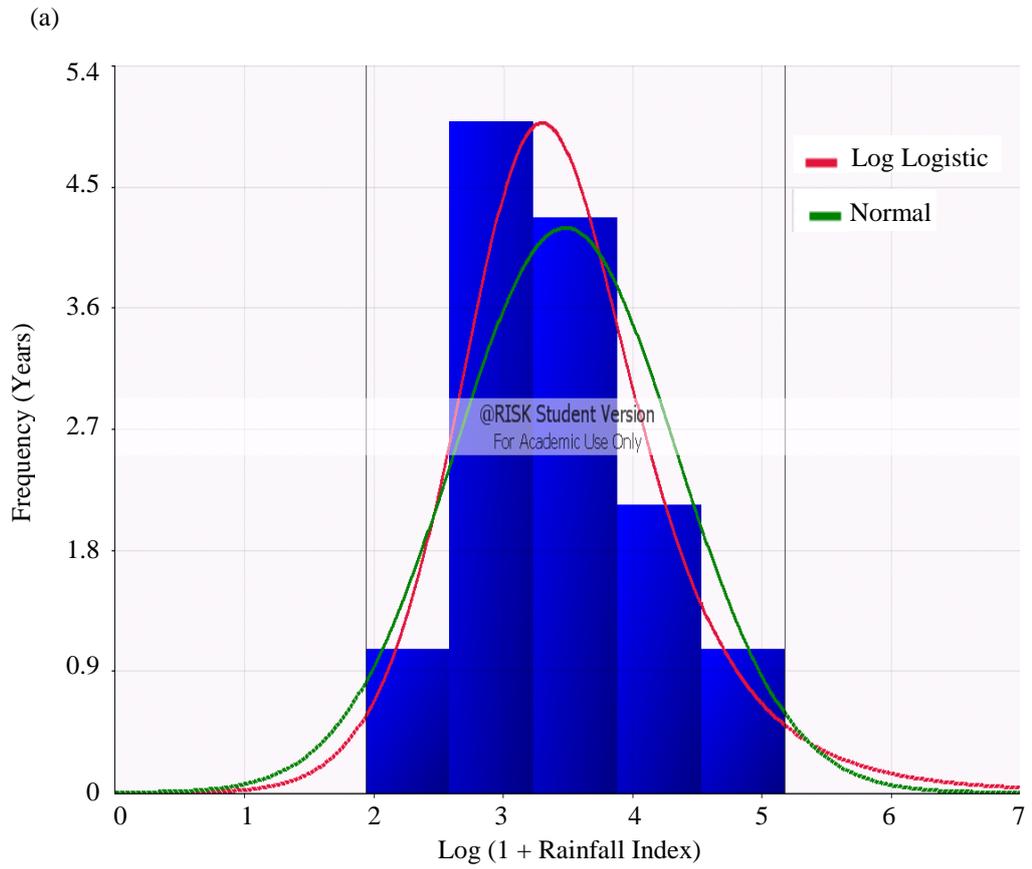


Figure 5.7. Transformed (using $\log(x+1)$) OND seasonal rainfall index from the Southwest region over the years 1959-1974. (a) '@Risk' plot showing the Log Logistic (red) and Normal (green) distributions that best represents the transformed OND seasonal rainfall index from the Southwest region over the years 1959-1974. (b) QQ-plot comparing the transformed rainfall index to the theoretical normal distribution. Note/ Rainfall data from the years 1960, 1964 and 1966 are not available for this example.

5.3.2 Simple linear regression

Regional seasonal rainfall hindcast models are developed using the standard statistical procedure of least-squares linear regression. The fundamentals of least-squares regression are presented below and in Figure 5.8. More details can be found in Wilks (2006).

Simple least-squares linear regression attempts to summarise the relationship between predictor and predictand, by a single straight line, on a scatter plot between the two variables as shown in Figure 5.8. The procedure positions the regression line by minimising the squared error for the prediction of the predictand (y) given the observations of the predictor (x).

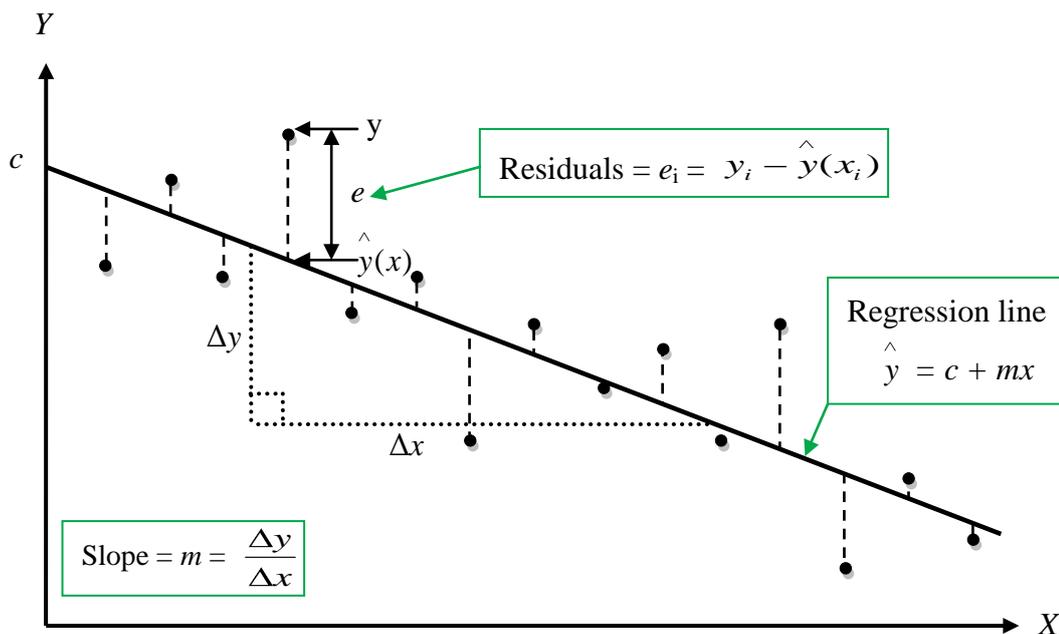


Figure 5.8. Schematic of the fundamentals of least-squares linear regression. Here, x represents the predictor variable, y represents the predictand and \hat{y} specifies a predicted value of y . The regression line ($\hat{y} = c + mx$) minimises the sum of the squared residuals, which are the vertical differences between the points and the regression line $e_i = y_i - \hat{y}(x_i)$.

Minimising the squared error across the line is equivalent to minimising the sum of the squared vertical distances (the dotted lines in Figure 5.8) between the regression line and the data points (the residuals, e_i). Each individual residual is defined as:

$$e_i = y_i - \hat{y}(x_i) \quad \text{Equation 5.3}$$

where \hat{y} specifies a predicted value of y . The resulting prediction equation for the regression line is:

$$\hat{y} = c + mx \quad \text{Equation 5.4}$$

where c is the intercept and m is the gradient of the line as shown in Figure 5.8.

The statistical software “R” includes a function, *lm*, which performs a linear regression between a predictor and a predictand index (R Development Core Team, 2009). The *lm* function calculates the intercept (c) and gradient (m) coefficients of the regression line, allowing the predicted values to be calculated using Equation 5.4.

Examination of the residuals

The residuals in each hindcast model must be normally distributed about the regression line in order to satisfy the validity assumptions of linear regression (Wilks, 2006). It is therefore vital that after each linear regression hindcast model has been developed, the distribution of the residuals (e_i) around the regression line is tested through a QQ plot (Chandler, 2009). Due to the effort of ensuring that the actual predictor and predictand indices follow a normal distribution, the residuals are found in all cases to be normally distributed around the regression line.

After the linear regression hindcast models have been developed, the resulting hindcasts need to be transformed back using the reverse of the transformation functions listed in Table 5.2. This ensures that the hindcast rainfall values are returned to their original dimensions, allowing the seasonal rainfall hindcast skill to be assessed (Section 5.4).

Region of the Rainfall Index	Time period of the Rainfall Index	Transformation Function
South	1959-1990	$\sqrt[3]{x}$
South	1959-1974	x^λ
South	1975-1990	$\log(x)$
Southeast	1959-1990	$\log(x)$
Southeast	1959-1974	$\sqrt[3]{x}$
Southeast	1975-1990	x^λ
Southeast-east	1959-1990	$\sqrt[3]{x}$
Southeast-east	1959-1974	$\sqrt[3]{x}$
Southeast-east	1975-1990	x^λ
Southeast-west	1959-1990	$\log(x)$
Southeast-west	1959-1974	$\log(x)$
Southeast-west	1975-1990	None
Southwest	1959-1990	$\log(1+x)$
Southwest	1959-1974	$\log(1+x)$
Southwest	1975-1990	x^λ
West	1959-1990	$\sqrt[3]{x}$
West	1959-1974	$\log(1+x)$
West	1975-1990	x^λ
Northwest	1959-1990	x^λ
Northwest	1959-1974	x^λ
Northwest	1975-1990	$\sqrt[3]{x}$
Northeast	1959-1990	x^λ
Northeast	1959-1974	$\log(x)$
Northeast	1975-1990	$\sqrt[3]{x}$

Table 5.2. Functions used to transform the OND regional rainfall indices to normal distributions. No transformations were required for the potential predictor indices. No transformations are reported for MAM as linear regression forecasts were not developed for this season.

5.3.3 Multiple linear regression

For those regions and seasons with multiple potential predictors, the rainfall hindcast models are developed using multiple linear regression (MLR). This is the more general form of the simple linear regression described in Section 5.3.2.

The prediction equation for MLR is given by Mutai et al. (1998) and Wilks (2006) as:

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k \quad \text{Equation 5.5}$$

Each of the k predictor variables (x_k) has its own coefficient (β_k), which is analogous to the slope (m) in Equation 5.4. The “regression constant” (β_0) is analogous to the intercept (c) in Equation 5.4. Simple linear regression is the special case where $k=1$.

The *lm* function in the software “R” is used to perform a MLR between a predictand index and various predictor indices. The function calculates the regression constant (β_0) and the regression coefficients (β_k) for each predictor index, allowing the predicted values to be calculated using Equation 5.5.

MLR can only be used when several potential predictor regions are found to have strong, temporally stable and significant correlations with the regional seasonal rainfall index. The potential predictor regions must be independent of one another. Any potential predictors that are used in combination within a MLR are tested for correlation with each other. The predictors should preferably have a low correlation with each other and be separated spatially. The hindcast would not be improved by forming a MLR between two related predictors or through using the same predictor region at different lead times. This is because most of the variance would already be explained by the first predictor. There are occasions when a single (simplifying) index can be created from two, or more, multiple predictors (Sanders and Lea, 2005). An example of this would be combining two potential predictor regions as a dipole. The difference between these two regions of the same predictor field should have a known dynamical link, such as two SST regions that form part of the IOD.

Any potential predictor indices that are used as part of a MLR need to be normalised to give each predictor a similar weighting within the linear regression model. The standardised anomaly of the predictor is used in the MLR, which is the predictor's anomaly normalised by the standard deviation. The standardised anomaly (z), which is dimensionless, is given by Wilks (2006) as:

$$z = \frac{x - \bar{x}}{s_x} \quad \text{Equation 5.6}$$

where \bar{x} is the average of the predictor index (x) over the training period and s_x is the standard deviation of x . The use of the standardised anomalies in the MLR reduces the influence of location and spread from the predictor indices.

5.4 Skill Assessment

The hindcast skills available from the seasonal rainfall hindcast models are examined over the three periods: 1959-1974, 1975-1990 and 1959-1990. The hindcast models are then applied in a replicated real-time sense to the independent verification period of 1991-2006. The forecast skills available over this later period are also examined. The verification measures used to assess the hindcast skills are rank correlation (r_{rank}), mean square skill score ($MSSS$) and the percentage of hindcasts in the correct tercile. R_{rank} and $MSSS$ are introduced in Chapters 4.2.3 and 4.2.4 respectively. The percentage of hindcasts in the correct tercile is computed by noting the number of years that the hindcast and observed rainfall are both in the same tercile in the same year. The percentage of 'successful' hindcast years is then calculated. The statistical significance (p -value) of the hindcast skills are calculated using the methods described in Chapter 4.2.6.

The skills of the hindcasts are also assessed over the extreme rainfall years. The upper and lower quartile rainfall years from each of the training and validation periods are examined separately. The skills found over these extreme years are important to assess as it is these years that are more likely to lead to drought or flooding in Kenya.

Cross validation over the 1959-1990 training periods

The most skilful hindcast model is selected based on the results from the cross validation skill assessment (Chapter 6). A block-removal, cross validation method is used for the development of the regional seasonal rainfall hindcast models and the assessment of their skill over the training periods of 1959-1974, 1975-1990 and 1959-1990. This is based on the method used by Gissila et al. (2004). For each time period of n years, the year of interest is removed along with 2 years either side, leaving a data set of $n-5$ years to use for calibration. The year of interest can then be used to validate the model. Block-removal prevents any artificial skill enhancement from multi-annual persistence.

Replicated real-time verification over the 1991-2006 independent period

Once the most skilful hindcast model has been selected for a given region and season a final verification is necessary for a robust assessment of the forecast skill. The last 16 years of data (1991-2006) are reserved as an independent verification data set. These data have no influence on the development of the models or the selection of the predictors and therefore allow for a stringent, independent assessment of the forecast skill. The hindcast models are assessed using a replicated real-time method over this independent verification period. In this case the regression replicated real-time forecast model is developed over the years proceeding the year of interest. The regression coefficients from this model are substituted into Equation 5.4 along with the predictor values from the year of interest to give the forecast value for that year.

Chapter 6

Statistical Seasonal Hindcast Models for the Kenyan ‘Short-Rains’ Season

Skilful prediction of seasonal rainfall would bring sound humanitarian and economic benefit to Kenya due to its dependence on rain-fed agriculture. This Chapter presents the regional OND rainfall hindcast models for Kenya that have been developed through this study. The hindcast skills of the models are assessed and improvements are made to the knowledge of the predictors that have strong, temporally stable links to regional OND rainfall in Kenya. Mechanisms are suggested that could be responsible for the dynamical links between the regional rainfall and predictor indices.

6.1 Overview of known factors contributing to interannual variability in the Kenyan ‘short-rains’

Previous research into the interannual variability and predictability of Kenyan seasonal rainfall has focused on the ‘short-rains’ season (Chapter 2.3). This is because the ‘short-rains’ are more predictable than the ‘long-rains’, even though they generally produce less rainfall (figure 5.1). The better predictability of the ‘short-rains’ compared to that of the ‘long-rains’ is due to their stronger interannual variability (Camberlin and Philippon, 2002) and spatial coherence of rainfall anomalies (Mutai and Ward, 2000).

Several publications (Lau and Sheu 1988; Indeje et al. 2000; Mutai and Ward 2000; McHugh 2006) have reported teleconnections between ENSO and East African OND rainfall. El Niño (La Niña) is found to have a positive (negative) correlation with Equatorial East African rainfall, with a peak during the OND ‘short-rains’ season. Exceptions to this include along the Kenyan coastline (Ogallo, 1988; Hastenrath et al., 1993). The theories behind the dynamics of these ENSO teleconnections are outlined in Chapter 2.3.1.

Since the discovery of the IOD in 1999 (Saji et al., 1999 and Webster et al., 1999) other publications (Black 2003, 2005; Clark et al 2003; Ummenhofer et al., 2009) have suggested that the IOD may have a stronger influence than ENSO on OND rainfall in EEA. This was further supported by evidence presented in Figure 2.9 (Chapter 2.3.4). The IOD has a positive correlation with Equatorial East African rainfall (Hastenrath et al. 1993; Hastenrath 2000; Manpande and Reason 2005), with a peak during the OND ‘short-rains’ season. The relationship strength varies by region (Zablone and Ogallo, 2008). The theories behind the dynamics of these IOD teleconnections are outlined in Chapter 2.3.1.

During periods of El Niño and the positive phase of the IOD, positive rainfall anomalies are often observed in EEA during the ‘short-rains’ season. During both of these events, similar patterns of SST and wind anomalies are generally observed in the Indian Ocean, with warm SST anomalies in the equatorial-west Indian Ocean and cold SST anomalies in the equatorial-east Indian Ocean (Chapter 2.3.1). These SST anomalies drive an atmospheric circulation of low-level wind anomalies from east to west over the Indian Ocean. The ascending branch of the Walker circulation shifts to a position over EEA as rising motion is enhanced over the positive SST anomalies in the equatorial-west Indian Ocean (Philippon et al., 2002). This leads to the observed increase in convective activity and rainfall amounts in EEA during the ‘short-rains’ season. A shift of the atmospheric Walker circulation is also generally observed over the Pacific Ocean, with the descending branch positioned over Indonesia and westerly winds over the equatorial-west Pacific Ocean (Chapter 2.3.1). The following Sections show that some of these regions of SST and wind anomalies are found to be the strongest predictor regions for regional OND rainfall in Kenya.

6.2 Skill assessment of statistical hindcast models

The most skilful hindcast models for OND rainfall in each Kenyan region are determined using the three cross-validated training periods between 1959 and 1990. Table 6.2 shows the hindcast skill of each model for each period as well as their skill in replicated real-time forecast mode for the period 1991-2006. The locations of the predictor regions, used to develop the best hindcast models, are shown in Figures 6.2-6.7 and their coordinates are given in Table 6.1.

Homogeneous rainfall region	Predictor region	Predictor region longitude range (°E) (min:max)	Predictor region latitude range (°N) (min:max)	Region description
West	SOI Aug	Index defined in Chapter 3.4.2	Index defined in Chapter 3.4.2	Tahiti and Darwin (Australia)
West	DMI Sep	Index defined in Chapter 3.4.3	Index defined in Chapter 3.4.3	West and East Indian Ocean
West*	(SST AS(1)) – (SST AS(2))	SST AS(1): (56:69) SST AS (2): (114:129)	SST AS(1): (2:8) SST AS (2): (-15:-5)	SST AS(1): NW Indian Ocean SST AS(2): East Indian Ocean
Southwest	SST JAS	(48:56)	(-9:4)	Indian Ocean off Kenyan coast
Southwest	SST JA	(46:55)	(-6:2)	Indian Ocean off Kenyan coast
Southwest*	SST AS	(48:58)	(-10:-2)	Indian Ocean off Kenyan coast
South	SST JJA	(223:238)	(11:19)	East Pacific Ocean
South	<i>u</i> -wind Aug	(110:150)	(0:10)	Indonesia
South*	(<i>u</i> -wind Aug) + (SST Sep)	<i>u</i> -wind Aug: (110:150) SST Sep: (46:58)	<i>u</i> -wind Aug: (0:10) SST Sep: (-8:2)	<i>u</i> -wind: Indonesia SST: Indian Ocean off Kenyan coast
Northeast	South region's (<i>u</i> -wind Aug) + (SST Sep)	<i>u</i> -wind Aug: (110:150) SST Sep: (46:58)	<i>u</i> -wind Aug: (0:10) SST Sep: (-8:2)	<i>u</i> -wind: Indonesia SST: Indian Ocean off Kenyan coast
Northeast*	South region's <i>u</i> -wind Aug	(110:150)	(0:10)	Indonesia
Northwest	SST Jun	(47:59)	(-4:2)	Indian Ocean off Kenyan coast
Northwest*	SST JJA	(47:59)	(-4:2)	Indian Ocean off Kenyan coast
Southeast	SST JA	(142:160)	(36:42)	Pacific Ocean off Japanese coast
Southeast*	SST JJA	(142:162)	(36:42)	Pacific Ocean off Japanese coast

Table 6.1. Locations of the potential predictor regions for each of the OND regional hindcast models. The regions are listed in order of increasing strength of forecast skill.

* marks the best hindcast models for each homogeneous rainfall region.

West Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
(SST AS(1)) - (SST AS(2))	0.57	0.01	22.4	0.58	0.02	20.4	0.38	0.10	9.6	0.65	0.01	53.6
Southwest Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
SST AS	0.50	0.01	18.1	0.55	0.03	23.0	0.39	0.09	0.1	0.42	0.07	28.2
South Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
(u -wind Aug) + (SST Sep)	0.46	0.01	22.1	0.37	0.10	0.4	0.41	0.07	7.2	0.53	0.04	30.7
Northeast Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
South region's u -wind Aug	0.40	0.01	19.6	0.37	0.07	21.3	0.36	0.13	0.8	0.46	0.02	5.90
Northwest Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
SST JJA	0.33	0.05	3.9	0.35	0.09	-8.1	0.54	0.02	11.3	0.29	0.23	7.80
Southeast Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
SST JJA	0.25	0.15	0.7	0.37	0.09	12.2	0.49	0.04	14.4	-0.02	0.48	-32.6

Table 6.2. Table comparing the OND rainfall hindcast and replicated real-time forecast skill scores for the best models for each homogeneous rainfall region in Kenya. Predictor regions and skill score results for the hindcasts from each model are shown (r_{rank} and its associated p -value and $MSSS$ (%)) over the periods 1959-1990, 1959-1974, 1975-1990 and 1991-2006. Red values show when: $r_{rank} \geq 0.4$, p -value ≤ 0.1 and $MSSS > 0\%$. The hindcast models are presented in order of strength of skill, from highest to lowest.

West Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
(SST AS(1)) - (SST AS(2))	0.64	0.06	30.2	0.64	0.12	22.5	0.43	0.21	8.3	0.76	0.03	58.8
Southwest Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
SST AS	0.31	0.04	30.7	0.55	0.18	16.0	0.76	0.05	32.4	0.74	0.03	35.0
South Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
(u -wind Aug) + (SST Sep)	0.56	0.03	38.3	0.54	0.09	36.4	0.55	0.08	24.0	0.33	0.10	29.3
Northeast Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
South region's u -wind Aug	0.61	0.08	33.4	0.69	0.07	26.8	0.43	0.15	20.2	0.48	0.29	-5.1
Northwest Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
SST JJA	0.50	0.06	20.7	0.29	0.12	0.4	0.88	0.01	54.7	0.24	0.07	32.0
Southeast Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
SST JJA	0.05	0.14	-2.1	0.48	0.10	11.8	0.31	0.13	20.2	-0.14	0.30	-3.5

Table 6.3. Table comparing the OND rainfall hindcast and replicated real-time forecast skill scores over the **extreme years** (*upper and lower quartiles of each period only*) for the best models for each homogeneous rainfall region in Kenya. Predictor regions and skill score results for the hindcasts from each model are shown (r_{rank} and its associated p -value and $MSSS$ (%)) for the upper and lower quartiles of the periods 1959-1990, 1959-1974, 1975-1990 and 1991-2006. Red values show when: $r_{rank} \geq 0.4$, p -value ≤ 0.1 and $MSSS > 0\%$.

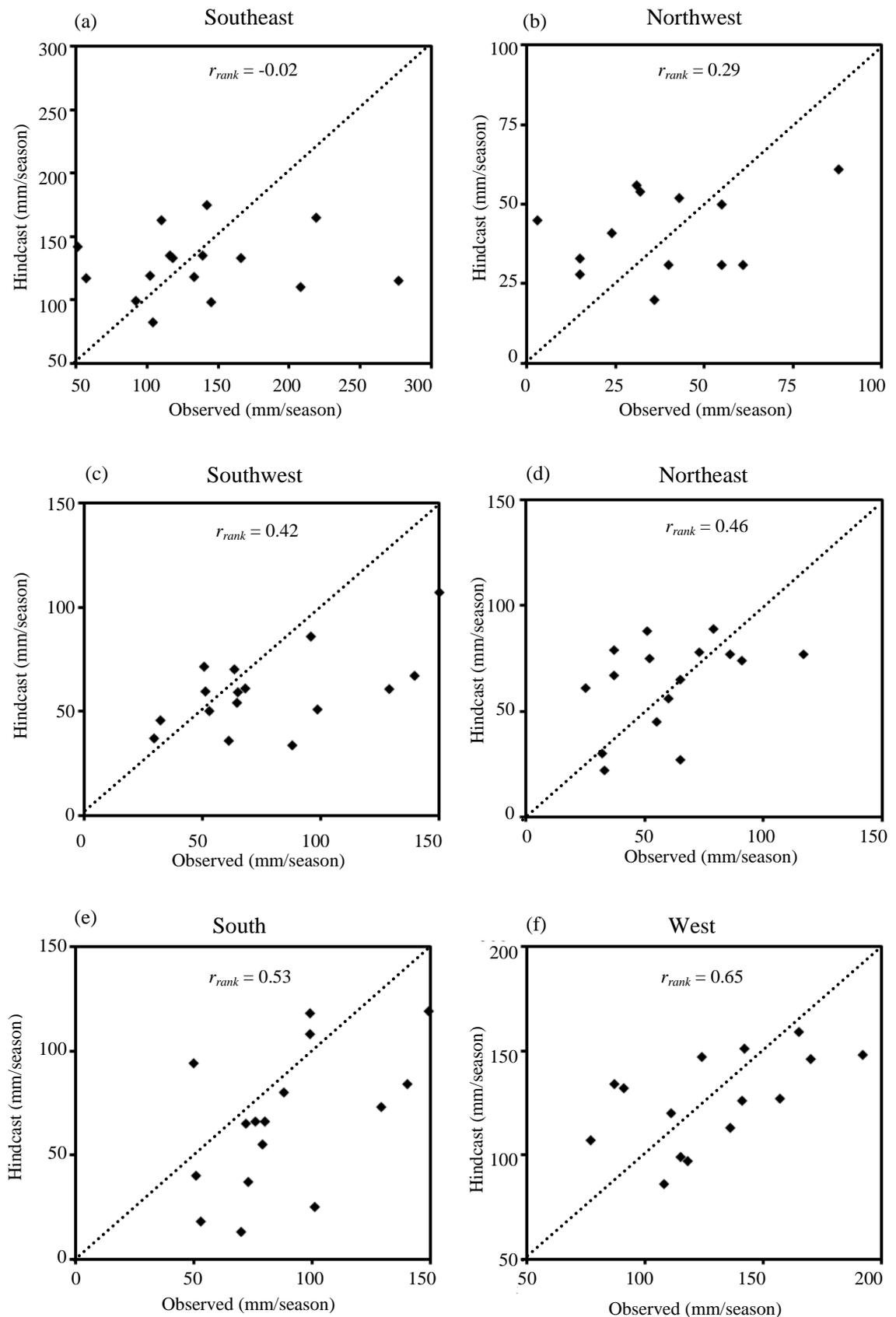


Figure 6.1. Scatter plots of hindcast vs. observed OND rainfall (mm/season) over the 1991-2006 independent verification period for each region of Kenya: (a) Southeast, (b) northwest, (c) southwest, (d) northeast, (e) south and (f) west. The lead 0 skill measure shown is Rank correlation (r_{rank}) and the observed data are threshold corrected regionally averaged rain gauge data from the KMD. The dotted line represents the position of the points if there were a perfect correspondence between the hindcast and observed data.

The hindcast models are listed in order of decreasing strength of OND rainfall hindcast skill. The strongest OND rainfall hindcast skill over all verification periods is from the best model developed for West Kenya (Section 6.2.1). A summary of the OND rainfall hindcast skill results for the best hindcast models for each region of Kenya is presented in Table 6.2. Scatter plots of hindcast vs. observed OND rainfall over the period 1991-2006 are shown for each region of Kenya in Figure 6.1. The dotted line represents the position of the points if there were a perfect correspondence between the hindcast and observed data. The greatest r_{rank} is shown in panel (f) for the West region at 0.65. It can be seen that the points are positioned the closest to the dotted line.

6.2.1 West Kenya

The Kenyan region offering the highest OND rainfall hindcast skill is West Kenya (Table 6.4). This region has an average r_{rank} of 0.55 (p -value ≤ 0.04). The skill scores are consistently high with $r_{rank} \geq 0.38$ (p -values ≤ 0.1) and $MSSS > 0\%$ over all 4 verification periods, peaking with a replicated real-time r_{rank} of 0.65 (p -value ≤ 0.01) over the period 1991-2006 (Figure 6.1 (f)). The skill scores are also consistently high during the extreme years, with $r_{rank} \geq 0.43$ and $MSSS > 0\%$ over all verification periods (Table 6.3). It must be noted, however, that the skill scores for the extreme years are only calculated using the upper and lower quartiles in each period, therefore the results have a lower significance. The hindcast skill is also found to be high for the number of years in the correct tercile, with 16 out of 32 anticipated correctly (Appendix B.1).

The best OND rainfall hindcast model for West Kenya is developed using a MLR based regression on the difference between two predictor indices: SST AS(1) and SST AS (2). These comprise the area average standardised anomalies of the AS two-month average SSTs over the West and East Indian Ocean predictor regions respectively. The predictor regions are shown in 6.2 and their location coordinates are given in Table 6.1.

The ‘short-rains’ of West Kenya are found to have strong links to the IOD (Section 6.1). The SST AS(1) and SST AS(2) predictor regions are coincident with the areas of SST anomalies that form the main part of the IOD, as shown in Figure 2.8. To further demonstrate the strong link to the IOD, the second most skilful hindcast model is developed using a simple linear regression based on the September DMI (Chapter 2.3.1) as the predictor (Appendix B.1).

The West is the second most populous region of Kenya with 29.6% of the total population. The land in the West of Kenya is mainly comprised of perennial cropland, which contributes towards Kenya’s economy (Chapter 5.1.2). It is therefore important for the population of Western Kenya and the economy of the whole country that the hindcasts for OND rainfall are consistently skilful for this region.

Period	r_{rank}	p -value	$MSSS$ (%)
1959-1990	0.57	0.01	22.4
1959-1974	0.58	0.02	20.4
1975-1990	0.38	0.10	9.6
1991-2006	0.65	0.01	53.6

Table 6.4. OND rainfall hindcast skill scores for the best hindcast model for the **West** region of Kenya. The skill score results are shown (r_{rank} and its associated p -value and $MSSS$) over the periods 1959-1990, 1959-1974, 1975-1990 and 1991-2006. Red values show when: $r_{rank} \geq 0.4$, p -value ≤ 0.1 and $MSSS > 0\%$.

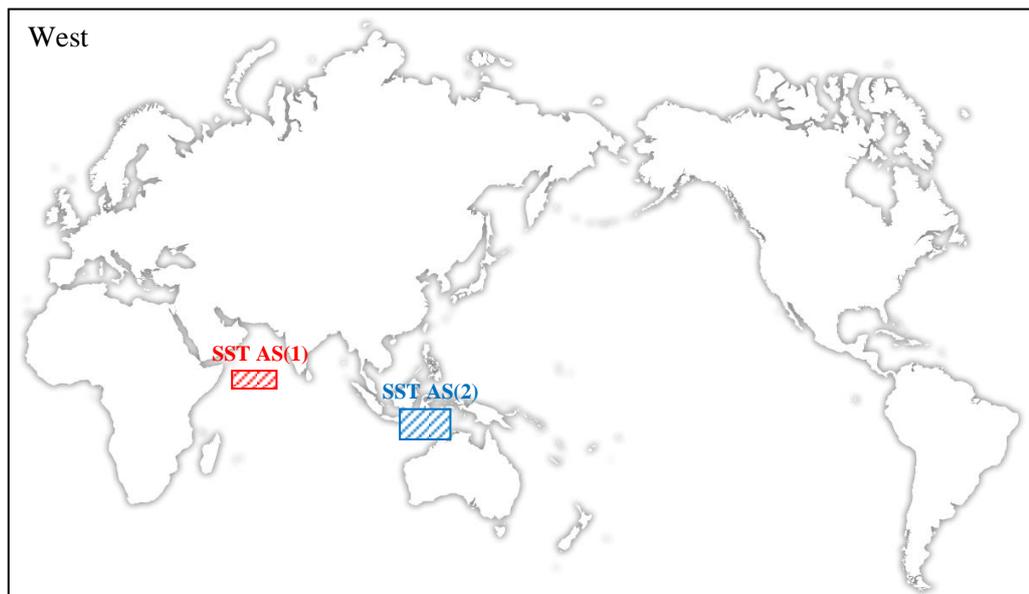


Figure 6.2. The most skilful predictor regions for OND rainfall in West Kenya. Correlation predictor selection plots for these regions are shown in Appendix A.1.

6.2.2 Southwest Kenya

The Kenyan region with the second strongest OND rainfall hindcast skill is the Southwest (Table 6.5), with an average r_{rank} of 0.47 (p -value ≤ 0.05). The skill scores are consistently high with $r_{rank} \geq 0.39$ (p -values < 0.1) and $MSSS > 0\%$ over all 4 verification periods, peaking with a cross-validated r_{rank} of 0.55 (p -value ≤ 0.03) over the period 1959-1974. The skill scores are also high during the extreme years, with r_{rank}

≥ 0.55 (p -values < 0.1) and $MSSS > 0\%$ throughout all verification periods (Table 6.3). Two exceptions to this are $r_{rank} = 0.31$ over the period 1959-1990 and p -value = 0.18 over the period 1959-1974. It must be noted, however, that the skill scores for the extreme years are only calculated using the upper and lower quartiles in each period, therefore the results have a lower significance.

The best OND rainfall hindcast model for Southwest Kenya is developed using a simple linear regression based on the AS two-month average SSTs over a region off the coast of East Africa (SST AS). The predictor region is shown in Figure 6.3 and the location coordinates are given in Table 6.1.

The ‘short-rains’ in Southwest Kenya are found to have strong links to the SSTs off the coast of East Africa, in the West Indian Ocean (Figure 6.3). To further demonstrate this link, the second and third most skilful hindcast models for Southwest Kenya are also developed using SST predictors from this region (Appendix B.2 and Table 6.1).

The SST AS predictor region is coincident with the predictor region used by Philippon et al. (2002) to develop their skilful OND rainfall hindcast models for East Africa (Chapter 2.3). The positive SST anomalies off the coast of East Africa are related to the ascending branch of the Walker circulation (Philippon et al., 2002). The positive correlation with Southwest Kenyan OND rainfall shows that strong ascending motion is favourable to rainfall over this region. The SST AS predictor region is also coincident with the SST AS(1) predictor region for Western Kenya (Section 6.2.1). The SST AS predictor for Southwest Kenya may also, therefore, be linked to the IOD.

The Southwest is the most populous region of Kenya with 38.1% of the total population, including the 2.3 million people who live in the capital city, Nairobi. The land is used to produce perennial and arable crops and improved grazing, all of which contribute towards the Kenyan economy (Chapter 5.1.2). It is therefore important for the population of the Southwest and economy of the whole country that the hindcast for OND rainfall for the Southwest are consistently skilful.

Period	r_{rank}	p-value	$MSSS$ (%)
1959-1990	0.50	0.01	18.1
1959-1974	0.55	0.03	23.0
1975-1990	0.39	0.09	0.1
1991-2006	0.42	0.07	28.2

Table 6.5. As Table 6.4 but for the **southwest** region of Kenya.



Figure 6.3. The most skilful predictor regions for OND rainfall in Southwest Kenya. Correlation predictor selection plots for this region are shown in Appendix A.2.

6.2.3 South Kenya

The Kenyan region with the third highest skill scores for OND rainfall is the South (Table 6.6), with an average r_{rank} of 0.44 (p -value ≤ 0.06). The skill scores are consistently high with $r_{rank} \geq 0.37$ (p -values ≤ 0.1) and $MSSS > 0\%$ over all 4 verification periods, peaking with a replicated real-time r_{rank} of 0.53 (p -value ≤ 0.04) over the period 1991-2006 (Figure 6.1 (e)). During the extreme years the skill scores are also consistently high with $r_{rank} \geq 0.54$ (p -values < 0.1) and $MSSS > 0\%$ over all verification periods (Table 6.3). This excludes the period 1991-2006 where $r_{rank} = 0.33$ (p -value ≤ 0.1). It must be noted, however, that the skill scores for the extreme years are only calculated using the upper and lower quartiles in each period, therefore the results have a lower significance.

The best OND rainfall hindcast model for Southern Kenya is developed using a MLR based regression on a combination of two predictors: an SST predictor index in the West Indian Ocean and a zonal (u) wind predictor index over Indonesia. The SST Sep predictor index comprises the area average standardised anomalies of the September SSTs over the West Indian Ocean as shown in Figure 6.4. The u -wind Aug predictor index comprises the area average standardised anomalies of the August u -winds over Indonesia, also shown in Figure 6.4. The predictor region location coordinates are given in Table 6.1.

Period	r_{rank}	p -value	MSSS (%)
1959-1990	0.46	0.01	22.1
1959-1974	0.37	0.10	0.4
1975-1990	0.41	0.07	7.2
1991-2006	0.53	0.04	30.7

Table 6.6. As Table 6.4 but for the **South** region of Kenya.

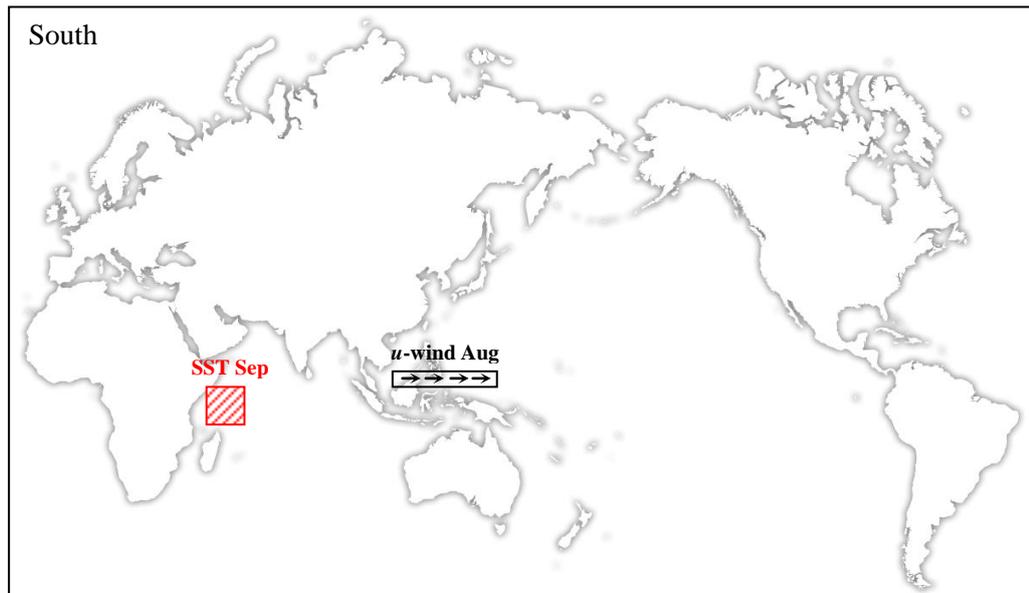


Figure 6.4. The most skilful predictor regions for OND rainfall in South Kenya. Correlation predictor selection plots for the SST Sep predictor region are shown in Appendix A.3. Composite difference predictor selection plots for the u -wind Aug predictor region are shown in Appendix A.4.

To further demonstrate the link between the South Kenyan OND rainfall and these two predictor regions, the second most skilful hindcast model is a simple linear regression model based on the u -wind Aug predictor index. The third most skilful hindcast model is a simple linear regression model based on the average standardised anomalies of the JJA three-month average SSTs off the coast of East Africa (Appendix B.3).

The two predictor regions, SST Sep and u -wind Aug, are related through the dynamics described in Section 6.1. During a positive IOD or an El Niño, positive SST anomalies are observed in the equatorial-west Indian Ocean (SST Sep). The descending branch of the Walker circulation shifts to be positioned over Indonesia, producing westerly low-level winds across the west Pacific Ocean (u -wind Aug). The SST and wind anomalies associated with the shifts in the Walker circulation peak at different lead times, hence the difference in the timings of the predictors.

Clark et al. (2003) also find a strong correlation between the coastal OND rainfall in Kenya and the IOD over the period 1950-1999. The positive (negative) correlations between OND rainfall in coastal Kenya and western (eastern) Indian Ocean SSTs peak in the autumn concurrent with the ‘short-rains’. This is further supported by Hastenrath et al. (2004), who find a positive correlation between OND rainfall in coastal Kenya and SSTs in the West Indian Ocean. The IOD and its associated SST and wind anomalies are, therefore, robust predictors for seasonal rainfall hindcast models for the South region of Kenya.

The South has the third smallest population in Kenya with 3.3% of the total population. Most of these live in the second most populous city in Kenya, Mombasa. Around a quarter of the land use is arable cropland, which contributes towards the Kenyan economy (Chapter 5.1.2). It is therefore important to the 950,000 people that live in the South region and of some importance to the Kenyan economy that the hindcast for OND rainfall for the South are consistently skilful.

6.2.4 Northeast Kenya

The Northeast region OND rainfall index does not have a significant or temporally stable link to any of the predictors introduced in Chapter 3. Thus a different technique is required to develop an OND rainfall hindcast model for the Northeast region. The Northeast OND rainfall index is correlated with each of the other regional OND rainfall indices. The results show that the Northeast rainfall index is most strongly correlated with the South OND rainfall index. The r_{rank} values between the South and Northeast OND rainfall indices are found to be 0.57, 0.81 and 0.70 (p -values < 0.1) for the periods 1959-1974, 1975-1990 and 1959-1990 respectively. This strong correlation is used in combination with the strongest hindcast model for the South region to develop OND rainfall hindcast models for the Northeast of Kenya.

A simple linear regression is formed between the Northeast and South rainfall indices, which have already been transformed to normal distributions using the functions in Table 5.2. The intercept (c) and gradient (m) values (Equation 5.4) from this regression are then used in a second linear regression in which the predictor (x) is the hindcast values produced from one of the strongest hindcast models from the South region of Kenya (Section 6.2.3). These hindcast values are transformed to a normal distribution

Period	r_{rank}	p -value	$MSSS$ (%)
1959-1990	0.40	0.01	19.6
1959-1974	0.37	0.07	21.3
1975-1990	0.36	0.13	0.8
1991-2006	0.46	0.02	5.90

Table 6.7. As Table 6.4 but for the **Northeast** region of Kenya.

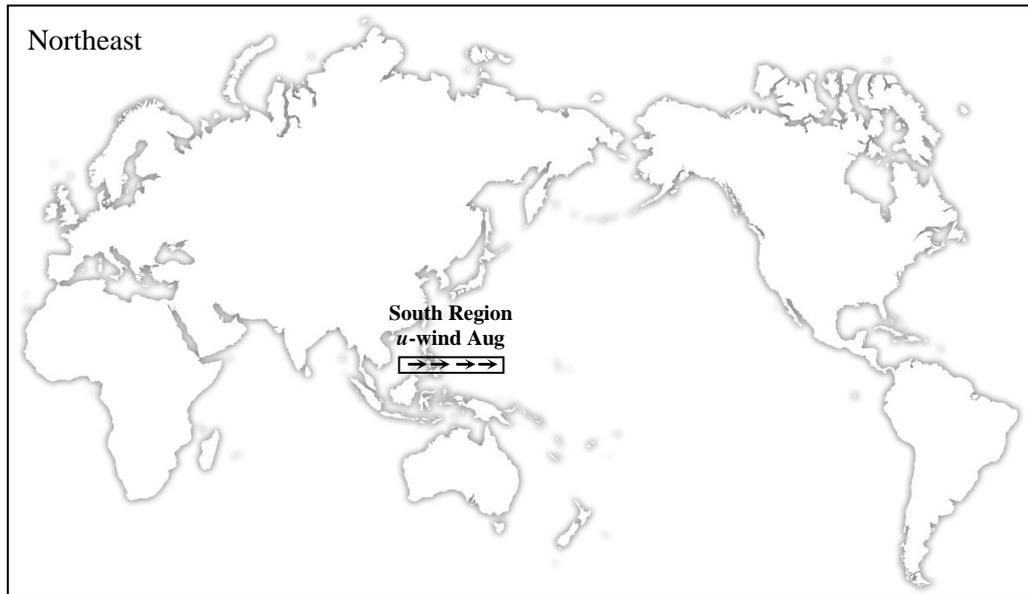


Figure 6.5. The most skilful predictor regions for OND rainfall in Northeast Kenya. Composite difference predictor selection plots for this South Kenya predictor region are shown in Appendix A.4.

using the same function as for the South OND rainfall index (Table 5.2). The predicted values resulting from the regression then need to be returned to the original rainfall dimensions, using the reverse of the transform functions used for the Northeast OND rainfall index (Table 5.2). This produces a new index of hindcast OND rainfall values for the Northeast region. This is repeated for the two other strong hindcast models from the South region. The skill scores of the resulting hindcast values for the Northeast of Kenya are then assessed and the strongest hindcast model is selected. The strongest hindcast model is based on a u -wind Aug predictor, located around Indonesia as shown in Figure 6.5. The location coordinates are given in Table 6.1.

The skill scores for the strongest OND rainfall hindcast model for the Northeast region of Kenya are shown in Table 6.7. These skill scores show an average r_{rank} of 0.40 (p -value ≤ 0.06). The skill scores are moderately high with $r_{rank} \geq 0.36$ (p -values ≤ 0.13) and $MSSS > 0\%$ over all verification periods, peaking with a cross-validated r_{rank} of 0.46 (p -value ≤ 0.02) over the period 1991-2006 (Figure 6.1 (d)). The skill scores are

generally stronger over the extreme years during the periods 1959-1990 and 1959-1974, with $r_{rank} \geq 0.61$ (p -values < 0.1) and $MSSS > 0\%$ (Table 6.3). However, although $r_{rank} > 0.43$ over the extreme years during the periods 1975-1990 and 1991-2006, the skill scores are not significant with p -values ≥ 0.15 and a negative $MSSS$ over the period 1991-2006. It must be noted that the skill scores for the extreme years are only calculated using the upper and lower quartiles in each period, therefore the results have a lower significance.

It remains unclear as to why the Northeast and South OND rainfall indices share such a strong relationship. Their link may be aided by the fact that the two regions follow similar annual rainfall distributions and that they are not separated by any topographical boundaries.

The Northeast has the second lowest population in Kenya with 1.4% of the total population. The main form of livelihood is nomadic pastoralism (Chapter 5.1.2), which means that the Northeast region does not contribute significantly to the Kenyan economy. The effect on the overall economy of Kenya of just moderate prediction skill for Northeast OND rainfall is thus small.

6.2.5 Northwest Kenya

The hindcast skill scores for OND rainfall in the Northwest region of Kenya are the second weakest in the country (Table 6.8), only marginally lower than for the Northeast. The average r_{rank} for the Northwest region is 0.38 (p -value ≤ 0.10), compared to the 0.40 (p -value < 0.06) for the Northeast. The skill scores are generally moderate with $r_{rank} \geq 0.33$ (p -values ≤ 0.09) for all verification periods excluding the replicated real-time period 1991-2006, which has $r_{rank} = 0.29$ (p -value = 0.23). $MSSS$ is positive over all verification periods, excluding over the period 1959-1974. The skill scores are similar during the extreme years, with r_{rank} ranging from 0.24 to 0.88 (p -values ranging from 0.12 to 0.01). $MSSS$ is, however, positive throughout all verification periods during the extreme years (Table 6.3). It must be noted that the skill scores for the extreme years are only calculated using the upper and lower quartiles in each period, therefore the results have a lower significance.

The best OND rainfall hindcast model for Northwest Kenya is developed using a simple linear regression based on the JJA three-month average SSTs over a region off the coast of East Africa (SST JJA). The predictor region is shown in Figure 6.6 and the location coordinates are given in Table 6.1. To further demonstrate the link to this region of SSTs, the second most skilful hindcast models for Northwest Kenya is also based on SSTs in this region (Appendix B.5 and Table 6.1).

The skill scores are not consistently strong over all of the verification periods. Therefore it is not possible to draw conclusions as to the mechanisms that drive rainfall in this region. The poor skill scores may be a result of problems with the quality of the rainfall data in the Northwest of Kenya. Only two stations contribute towards the rainfall data for this region (Baragoi and Lodwar), as no other rain gauge stations have >80% data availability over the period 1959-2006. Table 3.1 shows that Baragoi has the greatest percentage of missing data of all stations employed in the study, with 17% missing months. 93% of the missing months are in the second half of the period. This may explain the reduction in skill scores during the replicated real-time verification over the period 1991-2006. The Northwest OND rainfall index is not strongly correlated with the rainfall indices from any other region. This is indirect evidence that the OND rainfall data from the Northwest are of lower quality than in other regions. The lack of strong correlations to other regional rainfall indices hinders the development of improved hindcasts for this region using the methods employed in Section 6.2.4.

The Northwest is the least populous region of Kenya with 1.3% of the total population. The main form of livelihood is nomadic pastoralism (Chapter 5.1.2) hence the Northwest does not significantly contribute to the Kenyan economy. It is, therefore, of little damage to the economy of Kenya that the hindcasts for OND rainfall for the Northwest produce low skill.

Period	<i>r_{rank}</i>	<i>p</i> -value	<i>MSSS</i> (%)
1959-1990	0.33	0.05	3.9
1959-1974	0.35	0.09	-8.1
1975-1990	0.54	0.02	11.3
1991-2006	0.29	0.23	7.80

Table 6.8. As Table 6.4 but for the **Northwest** region of Kenya.



Figure 6.6. The most skilful predictor regions for OND rainfall in Northwest Kenya. Correlation predictor selection plots for this region are shown in Appendix A.5.

6.2.6 Southeast Kenya

The Kenyan region offering the lowest OND rainfall hindcast skill is Southeast Kenya (Table 6.9), with an average r_{rank} of 0.27 (p -value ≤ 0.19). The skill scores are consistently low with $r_{rank} > 0.40$ in only one verification period (1975-1990). The hindcast values are, however, generally more skilful than climatology over the cross-validated periods between 1959 and 1990, as shown by the positive $MSSS$ results (Table 6.9). The skill scores over the independent verification period (1991-2006) are low with r_{rank} close to zero and $MSSS < 0\%$. Similar results are seen during the extreme years (Table 6.3), with $r_{rank} > 0.4$ in only one verification period (1959-1974) and low skill scores over the period 1991-2006 with r_{rank} close to zero and $MSSS < 0\%$.

Period	r_{rank}	p -value	$MSSS$ (%)
1959-1990	0.25	0.15	0.7
1959-1974	0.37	0.09	12.2
1975-1990	0.49	0.04	14.4
1991-2006	-0.02	0.48	-32.6

Table 6.9. As Table 6.4 but for the **Southeast** region of Kenya.



Figure 6.7. The most skilful predictor regions for OND rainfall in Southeast Kenya. Correlation predictor selection plots for this region are shown in Appendix A.6.

The best OND rainfall hindcast model for Southeast Kenya is developed using a simple linear regression based on the JJA three-month average SSTs over a region off the coast of Japan (SST JJA). The predictor region is shown in Figure 6.7 and the location coordinates are given in Table 6.1.

The ‘short-rains’ in Southeast Kenya are found to be linked to a region of SSTs off the coast of Japan. A potential dynamical link between this predictor region and the Southeast OND rainfall index is through the IOD. Figure 2.8 shows the regions of SST anomalies that form the IOD. The SST JJA predictor for the Southeast of Kenya is coincident with the location and timing of the IOD SST anomalies off the coast of Japan in Figure 2.8. Appendix A.6 shows that, in the first training period (1959-1974), the correlation patterns in the Indian Ocean are similar to those that the IOD would produce. However, this Indian Ocean IOD signal is not present in the second period (1975-1990). This is the period with the strongest skill scores in Table 6.9 and therefore where there is greatest predictability. This suggests that the predictor is unlikely to be linked to the IOD. Without a temporally stable IOD signal evident in the Indian Ocean during both periods it may be possible that the potential IOD link is not strong enough to produce strong hindcast skill scores in Southeast Kenya.

The Southeast is the third most populous region of Kenya, with 26.3% of the total population. The majority of the land is semi-arid, where the main form of livelihood is nomadic pastoralism. Around 15% of the region (in the west) is used for arable cropland, which contributes towards the Kenyan economy (Chapter 5.1.2). Therefore, it is of little damage to the economy of Kenya that the hindcast for the OND rainfall for the Southeast produce low skill. It is, however, important for the 7.6 million people that live in this region. Due to the large number of people affected by these poor hindcasts, further work is needed to try to improve the forecast skill for the Southeast region. The Southeast OND rainfall index is not strongly correlated with the rainfall indices from any other region. Therefore, it is not possible to use those methods described in Section 6.2.4 to develop improved forecast models for this region. An alternative method to potentially improve the skill is detailed in the following section.

6.2.7 Splitting Southeast Kenya into Southeast-east and Southeast-west

One potential reason for the low rainfall hindcast skill for the Southeast of Kenya is the diverse topography of this region (Figure 2.4). In order to attempt to improve the OND rainfall hindcast skill for Southeast Kenya, the region is split into two sub-regions: Southeast-east and Southeast-west. The Southeast-east region comprises Garissa, Makindu and Voi, which are located at elevations between 300-1500m on semi-arid plains. The Southeast-west comprises Katumani, Marania Timau and Meru, which are located in the highlands at elevations of over 1500m.

The skill scores for the best OND rainfall hindcast models for the Southeast-east and Southeast-west regions are shown in Table 6.10. The skill scores provide only marginal improvement on the results for the whole Southeast region (Table 6.9) with average r_{rank} values of 0.25 and 0.30 (p -values = 0.15 and 0.16) for the Southeast-east and Southeast-west regions respectively. The skill scores remain consistently low with $r_{rank} > 0.4$ in only one verification period (1975-1990) for each sub-region. The hindcast values are, however, generally more skilful than climatology over the cross-validated periods between 1959 and 1990, as shown by the positive $MSSS$ results (Table 6.10). The skill scores over the independent verification period (1991-2006) are low with r_{rank} close to zero and $MSSS < 0\%$.

Southeast-east Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
SST JA	0.31	0.05	10.2	0.36	0.12	6.5	0.41	0.07	15.3	0.09	0.36	-11.3

Southeast-west Kenya Predictor	1959-1990			1959-1974			1975-1990			1991-2006		
	r_{rank}	p -value	$MSSS$									
SST JA	0.36	0.04	3.0	0.33	0.09	14.8	0.42	0.09	11.2	0.07	0.40	-14.1

Table 6.10. Table comparing the OND rainfall hindcast skill scores for the best forecast models for the Southeast-east and Southeast-west sub-regions of Kenya. Predictor regions and skill score results for the hindcasts from both models are shown (r_{rank} and its associated p -value and $MSSS$ (%)) over the periods 1959-1990, 1959-1974, 1975-1990 and 1991-2006. Red values show when: $r_{rank} \geq 0.4$, p -value ≤ 0.1 and $MSSS > 0\%$.

Homogeneous rainfall region	Predictor region	Predictor region longitude range (°E) (min:max)	Predictor region latitude range (°N) (min:max)	Region description
Southeast-east	SST JJA	(142:158)	(36:42)	Pacific Ocean off Japanese coast
Southeast-east	SST JJ	(142:170)	(36:42)	Pacific Ocean off Japanese coast
Southeast-east*	SST JA	(142:158)	(36:42)	Pacific Ocean off Japanese coast
Southeast-west	SST JJA	(142:156)	(36:42)	Pacific Ocean off Japanese coast
Southeast-west*	SST JA	(143:150)	(38:47)	Pacific Ocean off Japanese coast

Table 6.11. Locations of the potential predictor regions for the OND regional hindcast models for the Southeast-east and Southeast-west sub-regions of Kenya.

* marks the best hindcast models for each homogeneous rainfall region.

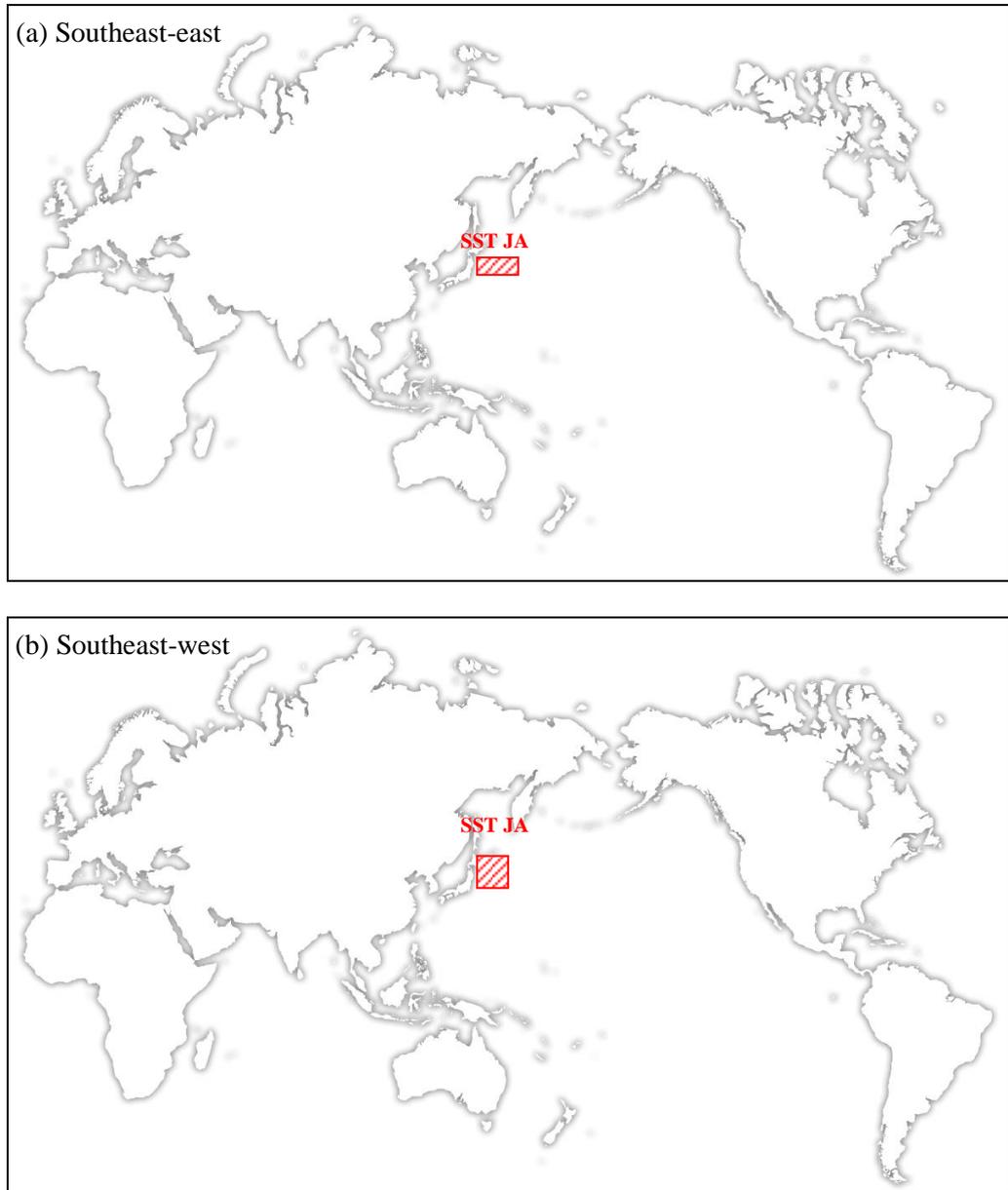


Figure 6.8. The most skilful predictor regions for the OND rainfall in (a) Southeast-east and (b) Southeast-west Kenya. Correlation predictor selection plots for these predictor regions are shown in Appendix A.7 and A.8 respectively.

The best OND rainfall hindcast models for the Southeast-east and Southeast-west regions of Kenya are developed using a simple linear regression based on the JA two-month average SSTs over regions off the coast of Japan (SST JA). The predictor regions are shown in Figure 6.8 and the location coordinates are given in Table 6.11. The predictor regions for the Southeast and its two sub-regions are all independently selected off the coast of Japan. Separating Southeast Kenya into sub-regions therefore demonstrates further the link between the ‘short-rains’ in this region and the region of SSTs off the coast of Japan. However, it does not help to improve the OND rainfall hindcast skill for the Southeast region. Further work is needed to find alternative predictors in order to improve the hindcast skill for this region.

6.3 Summary

This Chapter provides an overview of the OND rainfall hindcast skill for each of the models developed for the six regions of Kenya. Highest hindcast skill scores are achieved for the well populated and cultivated West and Southwest regions of Kenya. These have average r_{rank} values of 0.55 (p -value ≤ 0.04) and 0.47 (p -value ≤ 0.05) respectively. OND rainfall in these regions has strong links to the IOD. The next highest hindcast skill scores are achieved for the South of Kenya, with an average r_{rank} of 0.44 (p -value ≤ 0.06). The model is developed using a MLR based on a regression between the OND rainfall and a combination of SSTs and wind predictors that represent parts of the Walker circulation during positive IOD and El Niño years.

No strong, temporally stable predictor is found for the Northeast region. The correlation between the OND rainfall in the Northeast and South regions does, however, allow for a moderately skilful OND rainfall hindcast for the Northeast to be made. This model has an average r_{rank} of 0.40 (p -value ≤ 0.06). The models for the Northwest and Southeast regions of Kenya produce the lowest OND rainfall hindcast skill scores. These have average r_{rank} values of 0.38 (p -value ≤ 0.10) and 0.27 (p -value ≤ 0.19) respectively. Poor quality rainfall data is thought to be mainly responsible for the lack of Northwest OND rainfall hindcast skill. The OND rainfall in the Southeast region of Kenya is linked to an area of SSTs off the coast of Japan. Saji et al. (1999) found that this region of SSTs was linked to East African rainfall through the IOD. However, it seems that this region of SSTs is not a strong predictor, as the skill results were the lowest of all the regions.

Chapter 7

Statistical Seasonal Hindcast Models for the Kenyan ‘Long-Rains’ Season

Rainfall during the ‘long-rains’ season (MAM) is generally heavier and lasts longer than during the ‘short-rains’ season (Figure 5.1). Thus, rainfall deficits during the ‘long-rains’ season have a larger impact on the Kenyan economy, through water and power shortages, than rainfall deficits during the ‘short-rains’ season (Camberlin and Philippon, 2002). Despite the importance of the ‘long-rains’ season for Kenya and East Africa, little research has been published on its interannual variability and predictability (Camberlin and Philippon, 2002). This chapter presents the current state of knowledge of the mechanisms that may be responsible for the variability and complexity of the ‘long-rains’. The analysis finds that there are few strong, significant and temporally stable potential predictors for the ‘long-rains’ in Kenya. This hinders the development of skilful regional MAM rainfall forecast models for Kenya. Further research into the variability and predictability of the ‘long-rains’ is necessary in order to be able to predict them in the future.

7.1 Overview of reported factors contributing to the interannual variability of the Kenyan ‘long-rains’

Current knowledge is limited on the physical mechanisms which cause variability in the East African ‘long-rains’ season (Camberlin and Philippon, 2002). This is due partly to the ‘long-rains’ exhibiting a lower interannual variability than the ‘short-rains’ (Camberlin and Philippon, 2002; Camberlin and Okoola, 2003). Also, the ‘long-rains’ are associated with complex interactions between regional and large-scale mechanisms, which lead to a heterogeneous spatial and temporal rainfall distribution across Kenya (Beltrando 1990; Semazzi et al. 1996; Okoola 1998; Indeje et al. 2000). As a result, the influence of large-scale climate anomalies on interannual variability of the ‘long-rains’ is found to be low (Camberlin and Philippon, 2002; Camberlin and Okoola, 2003)

(Section 7.2). This is a problem for developing statistical forecast models for the MAM season.

The following sub-sections present a review of the current knowledge of the physical mechanisms that are thought to be linked to the variability of the Kenyan ‘long-rains’.

7.1.1 The northward passage of the ITCZ

The Kenyan ‘long-rains’ are found to be associated with the northward passage of the ITCZ (Hastenrath et al., 1993) as described in Chapter 2.2. The ITCZ favours large scale convergence through deep humidification of the air, giving rise to conditions suitable for rain generation (Mukabana and Pielke, 1996). However, Gatebe et al. (1999) and Okoola (1998) find that although Kenya’s climate is dominated by the migration of the ITCZ, local variations due to topography and large bodies of water introduce significant modifications to the country’s weather, producing mesoscale weather systems (Mukabana and Pielke, 1996). The Kenyan seasonal rainfall cycle cannot, therefore, be said to follow the classical ITCZ pattern. Anyah (2008) finds that MAM rainfall in Kenya is not always located at the exact point of wind convergence, which is the definition of the location of the ITCZ.

7.1.2 Topography, large bodies of water and other mesoscale forcings

It is generally accepted by Kenyan forecasters (Gitau 2008; Likumana 2008; Muita 2008; Mutemi 2008; Ogallo 2008) that Mesoscale processes dominate during the ‘long-rains’ season. Several publications (Mukabana and Pielke 1996; Camberlin and Wairota 1997; Okoola 1998; Gatebe et al. 1999) agree that regional peculiarities in Kenya’s climate are due to the country’s complex topography and the proximity of the Indian Ocean and Lake Victoria. Lake Victoria has a surface area of about 60,000 km². Figure 7.1 shows an east to west cross-section of Kenya. This emphasises the heterogeneous topography across the country and the dominance of the Kenyan highlands, the Indian Ocean and Lake Victoria on the landscape. The South and West regions, which border the large bodies of water, receive the greatest amounts of rainfall out of all the Kenyan regions during the MAM season (Figure 5.1). These regions are affected by land-sea

breeze circulations. These circulations result from the contrast in temperature between the water and the land, due to differential solar heating and radiative cooling (Indeje, 2000). Lake Victoria has an associated semi-permanent trough, which migrates from land to lake and vice versa during the night and day respectively.

Large scale thermal convection is often responsible for clouds and localised storms in Kenya. The country receives strong direct solar insolation due to its equatorial position. Random thermal convective clouds are therefore induced by afternoon heating of the ground (Mukabana and Pielke, 1996). Shanko and Camberlin (1998) provide evidence that the development of above-normal annual numbers of Tropical Cyclones in the Southwest Indian Ocean is associated with significant circulation anomalies. These anomalies reduce moisture advection from the Indian Ocean into EEA and can prevent the northward movement of the ITCZ, thus reducing rainfall over East Africa. Figure 7.2 (Shanko and Camberlin, 1998) presents evidence for this in the form of a composite difference plot showing the difference in 950hPa wind direction and magnitude for those subset years as described in the figure caption. A marked cyclonic circulation can be seen in the Southwest Indian Ocean (SWIO) region during those years with numerous Tropical Cyclones. Enhancement of the southern Hadley Cell is shown to result in reduced low-level moisture advection from the Indian Ocean towards East Africa. This produces unfavourable conditions for convective activity in East Africa.

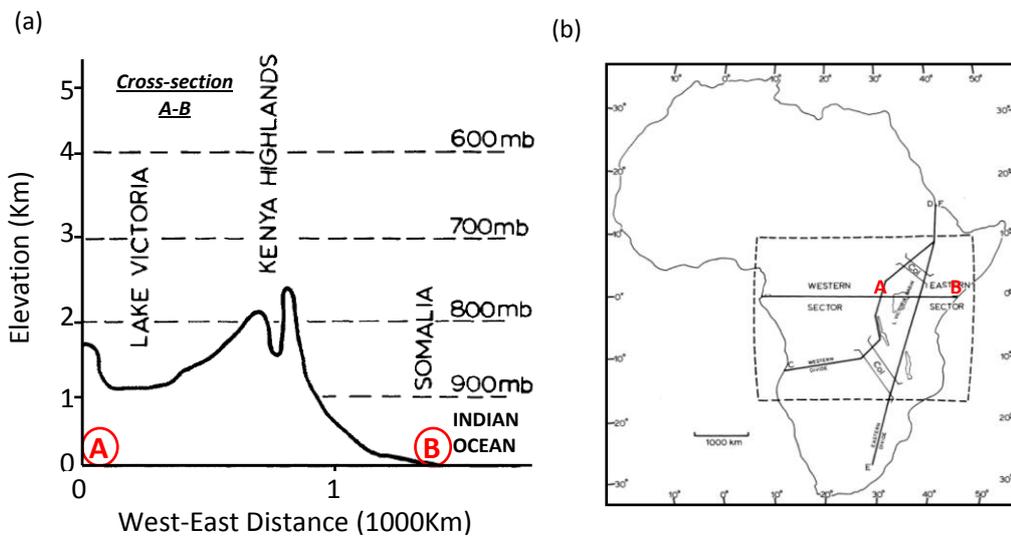


Figure 7.1. (a) A west to east topographical cross-section of Kenya taken along the line linking A and B shown in (b). Figures are adapted from Figures 1 and 4 in Hills (1979).

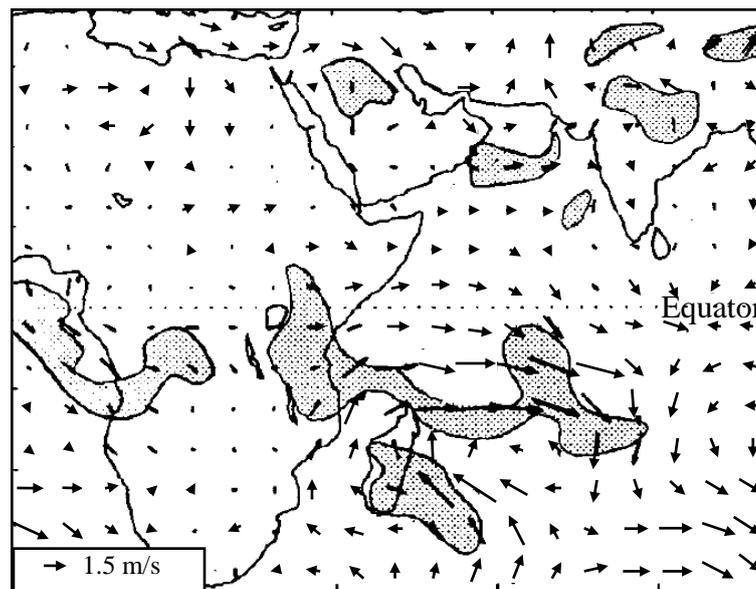


Figure 7.2. Composite difference plot showing direction, magnitude and significance of the difference in 950hPa wind anomalies for those subset years in Group 1 and Group 2. Group 1 comprises the 4 years (over the period 1954-1990) with the highest number of days with Tropical Cyclone or Tropical Depression occurrence in the region 30-80°E and 5-35°S. Group 2 comprises the 4 years over the same period with the least occurrence. Shaded areas show p -values < 0.05 in either the u - or v -component of the wind, according to the two-sided t -test. Figure from Shanko and Camberlin (1998).

7.1.3 The stratospheric QBO index

Indeje and Semazzi (2000) found contemporaneous and lagged significant correlations between MAM rainfall in some regions of Kenya and the stratospheric QBO index (Chapter 3.4.4) at 30hPa (QBO-30) over the period 1979-1992. The correlation results found by Indeje and Semazzi (2000) are presented in Table 7.1. The regions are defined as: (a) the central and western highlands of Kenya, (b) northwest Kenya (c) Western Kenya/the Lake Victoria region. As a result of these strong correlations, Kenyan forecasters assume that the QBO can be used as a predictor for the ‘long-rains’ (Gitau 2008; Muita 2008; Mutemi 2008). However, there is reason to question the strength of the results of Indeje and Semazzi (2000) as the correlations are not tested for temporal stability over two consecutive half-periods. Indeje and Semazzi also admit that the relationship fails during years that experience strong El Niños. Section 7.4.3 presents further analysis of the temporal stability of the link between Kenyan MAM rainfall and the QBO.

QBO-30 Period	Region (a)	Region (b)	Region (c)
JJA	0.84	0.48	0.59
SON	0.63	0.54	0.46
DJF	0.08	0.04	0.27
MAM	0.57	0.15	0.47
June	0.76	0.30	0.48
July	0.84	0.49	0.59
Aug	0.79	0.58	0.61
Sep	0.70	0.57	0.53
Oct	0.62	0.54	0.46
Nov	0.55	0.50	0.37
Dec	0.46	0.46	0.26
Jan	0.12	-0.02	0.28
Feb	0.20	0.00	0.31
Mar	0.39	0.06	0.43
Apr	0.53	0.16	0.47
May	0.67	0.20	0.45

Table 7.1. Results from Indeje (2000) showing correlations (r_{rank}) between the QBO-30 index (over the periods given in the left column) and MAM rainfall over the 14-year period 1979-1992. Red values have p -values < 0.1 calculated using a two-sided t -test based on a sample size of 14. Regions, as defined by Indeje and Semazzi (2000), cover (a) the central and western highlands of Kenya, (b) northwest Kenya and (c) western Kenya respectively. This Table is based on Table 3.1 in Indeje (2000).

7.1.4 ENSO and the IOD

Conflicting results are presented in the literature on the proposed links between ENSO and the ‘long-rains’ in East Africa. Several publications (Ogallo 1988; Ogallo et al. 1988; Hastenrath et al. 1993; Rowell et al. 1995; Philips and McIntyre 2000) show, using different spatial and temporal scales, that there are no significant correlations between East African MAM rainfall and either the atmospheric or oceanic component of ENSO. Other publications (Nicholson and Kim 1997; Indeje et al. 2000) show that the relationship between ENSO and East African MAM rainfall shifts across the season depending on the phase of the El Niño. These publications show weak positive correlations during the onset year of an El Niño, with stronger negative correlations during the decaying phase of an El Niño. Mutai and Ward (2000) show that the May rainfall in East Africa could be linked to the SOI. However, there is no link found between the SOI and the East African rainfall during the months of March and April.

McHugh (2006) finds that circulation variability in the sub-tropical South Pacific Ocean is associated with significant reductions in East African MAM rainfall with $r_{rank} = -0.33$ (p -value < 0.02). McHugh suggests that the north-easterly winds become strengthened, blowing from Sudan, across EEA and on to the Atlantic Ocean. These winds then weaken or prevent the influx of moist westerly airflow, which can be associated with rainfall generation in EEA.

Zablone and Ogallo (2008) find that the IOD is not significantly correlated with MAM rainfall over East Africa. This is attributed to the timing of the IOD, which peaks between July and December. Beltrando (1990) also finds that there are no strong teleconnections between MAM rainfall over East Africa and SSTs in the Indian Ocean.

7.1.5 The MJO

Pohl and Camberlin (2006) find that the MJO is linked to East African MAM rainfall. They propose that, on an interannual timescale, the variability of the ‘long-rains’ is significantly affected by the amplitude of the MJO. A strong MJO is found to be related to the early onset of the ‘long-rains’, leading to a longer rainy season and an increased

likelihood of more extreme events. Pohl and Camberlin do not use the MJO as a predictor, but as a contemporaneous modulator of the large scale rainfall variability over East Africa.

7.2 Data quality

Correlation of MAM rainfall between stations in each region

Previous studies into MAM rainfall predictability in East Africa have found weak correlations between rainfall time series at neighbouring rain gauge stations (Ogallo, 1989; Beltrando, 1999). In this thesis, correlations are made between the MAM rainfall indices from the stations in each homogeneous rainfall region. In contrast to the findings of earlier studies, the correlations between stations in each region are found to be generally strong ($r_{rank} > 0.45$), except between stations in the South and Northwest regions. In the South region, r_{rank} values of ≤ 0.35 are found between the stations of Lamu and Mombasa over the periods 1959-1974 and 1991-2006 and r_{rank} values of ≤ 0.45 are found over the period 1959-1990. Low correlation values are also found in the Northwest region, between Baragoi and Lodwar, with r_{rank} values of ≤ 0.43 and ≤ 0.04 over the periods 1959-1974 and 1991-2006 respectively. The cross-correlation values for these two regions are, therefore, not temporally stable. The MAM rainfall data in the South and Northwest regions of Kenya cannot, therefore, be used with confidence to search for potential predictors and to develop forecast models. Section 7.3 describes the investigation to find potential predictors for the ‘long-rains’ in the remaining **Southeast, Southwest, West** and **Northeast** regions of Kenya.

Correlation of MAM rainfall between regions

Correlations between regional MAM rainfall indices are also investigated. High correlations between MAM rainfall indices from different regions would indicate potential predictability from large-scale climate anomalies. Previous studies have found that MAM rainfall has a heterogeneous spatial distribution across Kenya (Beltrando 1990; Semazzi et al. 1996; Okoola 1998; Indeje et al. 2000). It is therefore suggested that large-scale climate anomalies have a low influence on the interannual variability of the ‘long-rains’ (Camberlin and Philippon, 2002; Camberlin and Okoola, 2003).

Tables 7.2 and 7.3 show the r_{rank} between OND and MAM rainfall from the different regions of Kenya over the period 1959-1990. It can be seen that, in contrast to OND rainfall, MAM rainfall is not as well correlated between regions. The MAM rainfall is therefore not expected to be as predictable as OND rainfall from large-scale climate factors. This is supported by Indeje (2000), who suggests that the high spatial variability across Kenya shows that local factors are more dominant than large-scale factors in modulating the MAM rainfall patterns.

	South	Southeast	Southwest	West	Northwest	Northeast
South	1	0.43	0.58	0.57	0.34	0.70
Southeast	0.43	1	0.40	0.36	0.42	0.37
Southwest	0.58	0.40	1	0.40	0.41	0.45
West	0.57	0.36	0.40	1	0.30	0.49
Northwest	0.34	0.42	0.41	0.30	1	0.27
Northeast	0.70	0.37	0.45	0.49	0.27	1

Table 7.2. r_{rank} values of OND rainfall between the different regions of Kenya over the period 1959-1990. r_{rank} values ≥ 0.4 are shown in bold.

	South	Southeast	Southwest	West	Northwest	Northeast
South	1	0.18	0.13	-0.07	-0.02	0.29
Southeast	0.18	1	0.34	0.41	0.22	0.48
Southwest	0.13	0.34	1	0.43	0.21	0.49
West	-0.07	0.41	0.43	1	0.34	0.44
Northwest	-0.02	0.22	0.21	0.34	1	0.23
Northeast	0.29	0.48	0.49	0.44	0.23	1

Table 7.3. r_{rank} values of MAM rainfall between the different regions of Kenya over the period 1959-1990. r_{rank} values ≥ 0.4 are shown in bold.

7.3 New methods to investigate ‘long-rains’ predictors

Several publications (Ogallo, 1988; Ogallo et al., 1988; Beltrando 1990; Hastenrath et al., 1993; Rowell et al., 1995; Philips and McIntyre, 2000; Camberlin and Philippon 2002; Camberlin and Okoola 2003; Zablon and Ogallo 2008) suggest that the East African ‘long-rains’ do not exhibit strong relationships with large-scale climate or SST anomalies. This study aims to verify whether this is the case for regional MAM rainfall

in Kenya. Potential predictors are sourced from the monthly data introduced in Chapter 3. The strength, significance and temporal stability of the correlations between the potential predictors and the MAM rainfall in Southeast, Southwest, West and Northeast Kenya are assessed, using the methods described in Chapter 5.2.

No strong, significant and temporally stable potential predictor regions are found for the regional MAM rainfalls in Kenya out of the pool of potential predictors introduced in Chapter 3. The methods therefore need to be amended in order to expand the search for more potential predictors for this season. Firstly, two new SST indices are developed (newSST-north and newSST-south). These comprise the average SSTs in the regions off the coast of East Africa bound by the coordinates (i) 50 to 70°E and 0 to 15°N and (ii) 50 to 70°E and 0 to 15°S. It is suggested that these new SST regions could be important to Kenyan moisture availability during the 'long-rains' season. Indeje and Semazzi (2000) suggest links between MAM rainfall in Kenya and the stratospheric QBO (Section 7.1.3). This study tests the relationship between the regional Kenyan MAM rainfall and the QBO indices at both 30hPa and 50hPa (QBO-30 and QBO-50) for an extended period of time.

Finally, the MAM rainy season is divided into its individual months of March, April and May. Beltrando (1990) and Rowell et al. (1994) suggest that, for analysis of the interannual variability of the 'long-rains' over East Africa, each month should be studied separately. Camberlin and Philippon (2002) support this argument as they find that the 'long-rains' season exhibits a low temporal coherence. Gatebe et al. (1999) find that the climatological air transport to and from Kenya switches between March and May. Figure 7.3 shows the air transport climatologies from Gatebe et al. (1999) for March (a) 700hPa and (b) 400hPa, and for May (c) 700hPa and (d) 400hPa. During March, the lower level winds are mainly easterly, from the Indian Ocean. mid level winds are also mainly easterly. During May the lower level wind direction switches to being mainly southerly, from Tanzania and along the Indian Ocean coast. Mid level winds remain mainly easterly, from the Indian Ocean. The switch in low-level winds, from easterly in March to southerly in May, is likely to be linked to the northward passage of the ITCZ and the associated change in the monsoon winds. This strengthens the argument that each month within the MAM season should be assessed independently.

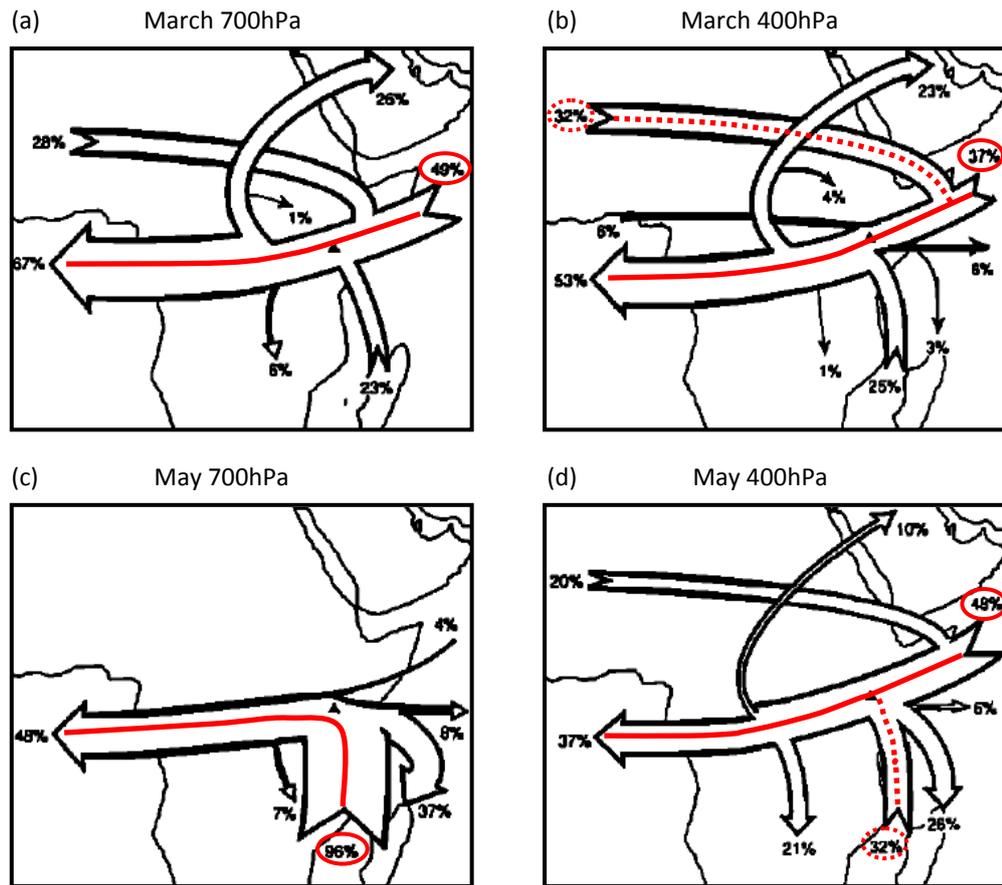


Figure 7.3. March (a) 700hPa and (b) 400hPa and May (c) 700hPa and (d) 400hPa air transport climatologies over Africa (1971-1975) to and from Kenya. Red solid arrows indicate the dominant wind direction. Red dashed arrows indicate the second most dominant wind direction. This Figure is adapted from Figures 5 (a) and (d) and 6 (a) and (d) in Gatebe et al. (1999).

7.4 Potential predictors for the ‘long-rains’

In order to select potential predictors, the methods outlined in Chapter 5 are implemented along with the additions introduced in Section 7.3. A *small* selection of seven potential predictors are identified as having strong, significant and temporally stable correlations with regional March, April or May rainfall in Kenya. A summary of the potential predictors is presented in Table 7.4. Two potential predictors, comprised of DJF SSTs, are found to be linked to the April rainfall indices in the Southeast and West regions of Kenya (Section 7.4.1). The newSST-south index is found to be linked to the March rainfall index in the Southwest of Kenya (Section 7.4.2). Finally, the DMI, MJO and QBO indices are found to be linked to a selection of March, May and MAM rainfall indices in Southwest and Northeast Kenya (Section 7.4.3).

Region	Rainfall Period	Potential Predictors	Further Details
Southeast	MAM	NA	NA
	March	NA	NA
	April	DJF - SSTs East Indian Ocean	Section 7.4.1 and Figure 7.4
	May	NA	NA
Southwest	MAM	NA	NA
	March	February - newSST-South	Section 7.4.2
	April	NA	NA
	May	February - DMI February - MJO	Section 7.4.3 Section 7.4.3
West	MAM	NA	NA
	March	NA	NA
	April	DJF – SSTs East Indian Ocean	Section 7.4.1 and Figure 7.5
	May	NA	NA
Northeast	MAM	NA	NA
	March	October - QBO-30 January - MJO	Section 7.4.3 Section 7.4.3
	April	NA	NA
	May	NA	NA

Table 7.4. Potential predictors for the ‘long-rains’ season in Kenya. Potential predictors are shown for MAM, March, April and May for the Southeast, Southwest, West and Northeast regions. The South and Northwest regions are not shown due to the poor quality of MAM rainfall data. “NA” indicates occasions where no potential predictors have been found. Further details on the correlations between the potential predictor and the rainfall indices can be found in the sections and figures listed in the final column.

7.4.1 SST potential predictor regions

A region of December-February (DJF) SSTs in the eastern Indian Ocean is found to have a strong, significant and temporally stable, negative correlation with the **April** rainfall index in the **Southeast** of Kenya (Figure 7.4). The coordinates of the SST potential predictor region are 98°E to 118°E and 30°S to 42°S. This is the only potential predictor region found for the Southeast region of Kenya.

A second SST potential predictor is a region of DJF SSTs in the eastern Indian Ocean. This has a strong, significant and temporally stable correlation with the **April** rainfall index for the **West** of Kenya (Figure 7.5). The coordinates of this SST potential predictor region are 105°E to 115°E and 8°S to 38°S.

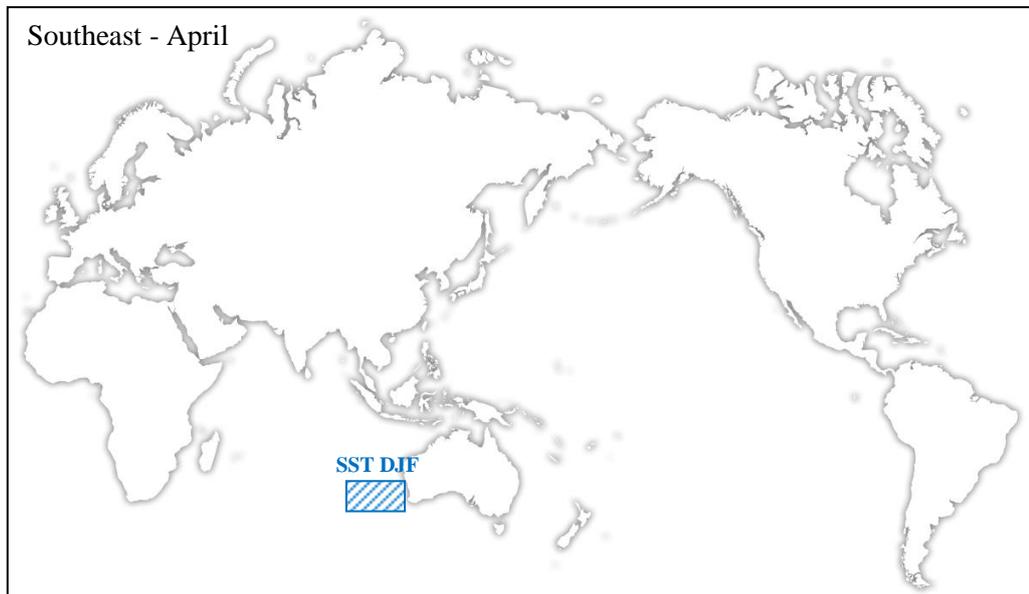


Figure 7.4. The best predictor region for April rainfall in Southeast Kenya. Correlation predictor selection plots for this region are shown in Appendix C.1.

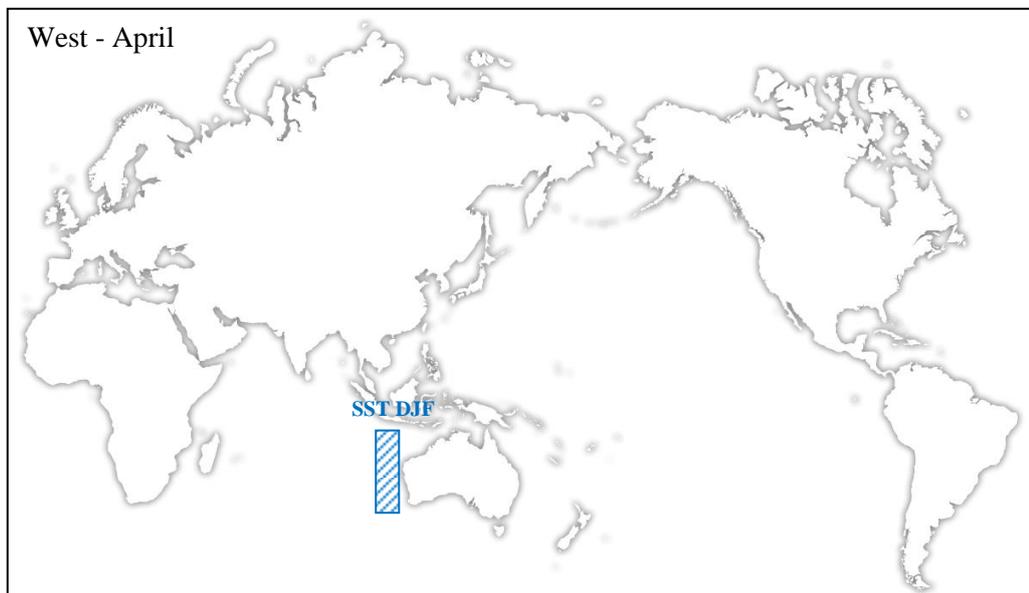


Figure 7.5. The best predictor region for April rainfall in West Kenya. Correlation predictor selection plots for this region are shown in Appendix C.2.

7.4.2 New SST potential predictor indices

One example is found of a strong, temporally stable relationship between one of the new SST indices and a regional rainfall index. The correlations between the February new SST-south index and the **March** rainfall index for the **Southwest** of Kenya are consistently high with $r_{rank} \geq 0.4$ over the periods 1959-1974, 1975-1990 and 1959-1990.

7.4.3 DMI and MJO potential predictor indices

There is one example where the DMI is found to be a potential predictor for a regional rainfall index. The February DMI has a strong, temporally stable correlation with the **May** rainfall in the **Southwest** of Kenya, with $r_{rank} \geq 0.48$ over the periods 1959-1974 and 1975-1990.

There are two examples of a link between the MJO index and regional rainfall indices in Kenya. The February MJO index is found to have strong, temporally stable correlations with the **May** rainfall in **Southwest** Kenya, with $r_{rank} \geq 0.43$ over the periods 1978-1986, 1987-1995 and 1978-1995. Secondly, the January MJO index is found to have strong, temporally stable correlations with the **March** rainfall in **Northeast** Kenya, with $r_{rank} \leq -0.44$ over the same periods.

7.4.4 QBO potential predictor index

The QBO has been accepted by Kenyan forecasters as a skilful operational predictor for the Kenyan 'long-rains' (Indeje and Semazzi 2000; Gitau 2008; Muita 2008; Mutemi 2008). However, this study finds that there is only one example of a strong, temporally stable correlation between a QBO index and a Kenyan regional rainfall index. The correlations between the October QBO-30 index and the **March** rainfall in the **Northeast** of Kenya are consistently ≥ 0.41 over the periods 1959-1974 and 1975-1990. There were no other examples of strong and temporally stable correlations with either QBO index.

The original, strong correlation values found by Indeje and Semazzi (2000) (Table 7.1) have been reproduced in the course of this thesis. However, they are found to be unstable when the period is split in half or tested over the periods used throughout this study (1959-1974 and 1975-1990). The r_{rank} values over these half periods are generally found to be less than 0.3, with only a handful of higher values. It is therefore concluded that the QBO index is not a temporally stable predictor for the ‘long-rains’ in any region of Kenya and should not, therefore, be used as an operational predictor.

7.4.5 Summary

There are few potential predictors identified for the MAM rainfall in Kenya, out of a wide selection of available climate predictors. It is possible that a proportion of the strong correlations between regional MAM rainfall indices and potential predictor indices are statistical artefacts. This is due to the large number of correlations being calculated at different lead times, where one would expect a few strong correlations to occur through random chance. The results therefore need to be viewed with caution. It is not possible, therefore, to use these potential predictors with confidence to develop regional rainfall forecast models for the MAM season in Kenya.

7.5 The spring predictability barrier

Why is it so difficult to find predictors for the ‘long-rains’ in Kenya? The ‘spring predictability barrier’ (Webster, 1995) explains why the ‘long-rains’ cannot be predicted using observations of coupled ocean-atmosphere systems from the Pacific Ocean basin, such as ENSO. Correlations between regional rainfall and Pacific Ocean based ocean-atmosphere systems decrease throughout the boreal spring. This limits the predictability of coupled ocean-atmosphere systems such as ENSO (Webster, 1995). Physically, the pressure gradients across the Pacific Ocean are at their weakest during the spring. Therefore, ENSO events can grow quickly in response to small perturbations. This makes it difficult to make forecast based on ENSO or Pacific Ocean SSTs during this season. Webster (1995) finds that long term coupled numerical model prediction experiments also show this ‘spring predictability barrier’, with a substantial decrease in observation-prediction correlation across the boreal spring. Chapter 2.3.4 showed that

the IOD is contemporaneously linked to ENSO. Hence the ‘spring predictability barrier’ also affects the predictability of the Indian Ocean during this season.

The ‘spring predictability barrier’ also affects the prediction of the main rainy seasons in other East African countries. Korecha and Barnston (2007) show that forecasts for the main rainy season in Ethiopia (named Kiremt), from June-September, are primarily based on ENSO predictors. Skilful forecasts for the Kiremt rainy season can only be produced with at least a 3 month lead (issued on 1st March or before). This lead time is necessary to avoid the spring predictability barrier, which excludes the use of SST predictors from March, April or May. It is possible to develop ENSO-based forecasts for the Kiremt rainy season in Ethiopia due to its later start date at beginning of the boreal summer.

7.6 Discussion

The key question that remains is *what would it take to develop skilful statistical MAM rainfall forecast models for Kenya?* Firstly, it would take a large amount of dedicated research time to search for new potential predictors for the ‘long-rains’. A suggestion for an alternative predictor is the strength of the ITCZ. This could be measured using Outgoing Longwave Radiation (OLR) measured to the south of Kenya. Finally, more research is required into the variability of the Kenyan ‘long-rains’, so that knowledge of the mechanisms that force this variability can be improved.

Chapter 8

Comparison of Statistical and Dynamical Hindcast Models for the Kenyan ‘Short-Rains’

This Chapter presents a comparison of statistical and dynamical hindcast models for the ‘short-rains’ season. The aim is to examine which type of model produces the most skilful regional OND rainfall hindcasts for each homogeneous rainfall region in Kenya.

Statistical seasonal rainfall forecasts are developed operationally before each rainy season in Kenya by the country’s Meteorological Department (Chapter 2.5) (Buizer et al., 2000; Likumana, 2008; Muita, 2008; Ogallo, 2008). Forecasters also refer to outputs from dynamical seasonal rainfall forecast models, prior to issuing their forecasts. This is the first study to compare the OND rainfall hindcast skills available from statistical and dynamical hindcast models for Kenya. Specifically, the hindcasts from the best statistical OND rainfall models for each region of Kenya (Chapter 6) are compared against the hindcasts from the dynamical multi-model ensemble system, EUROSIP (Section 8.1). The results will help to inform the Kenyan forecasters as to which model has the greatest OND rainfall hindcast skill for each region.

8.1 EUROSIP

The EUROSIP (**EURO**pean **S**easonal to **I**nterannual **P**rediction) dynamical multi-model ensemble global forecast system was built on the success of the DEMETER project (Chapter 4) (Vitart et al., 2007). The hindcasts from EUROSIP were first made available to the public in 2009. This is the first study to use the EUROSIP hindcasts independently of the developing institutions.

EUROSIP comprises three state-of-the-art European coupled ocean-atmosphere global climate models. These models are from the ECMWF (European Centre for Medium-

Range Weather Forecasts), Météo-France (Centre National de Recherches Météorologiques) and the UKMO (UK Met Office). The EUROSIP models are updated versions of the models used in the DEMETER system (Chapter 4). Table 8.1 shows the characteristics of these different models in the DEMETER and EUROSIP systems (Palmer et al., 2004; Davey and Ferranti, 2006; Anderson et al., 2007; Vitart et al., 2007; ECMWF, 2009). Improvements have been made to all models, with upgrades between systems 2 and 3. The horizontal and vertical resolutions of the atmospheric components of the ECMWF and Météo France models have been increased, as shown in Table 8.1. A major difference between DEMETER and EUROSIP is that the latter is an operational system. Each forecast is run in real-time with initial conditions from the 1st of each month. The forecasts are then issued on the 15th of the month to allow time for SST field acquisition and to run the forecasts on a reliable operational schedule (Vitart et al., 2007). Stockdale (2007) explains that the operational real-time forecasts are comprised of 41 members per model with forecasts out to 5-6 months lead. The hindcasts are produced from 11 members and are available from 1987 to present (Stockdale, 2007).

It is not possible to directly compare the skills of the hindcasts from the statistical regional OND rainfall hindcast models with those from DEMETER. This is because the hindcasts from DEMETER are issued quarterly and the hindcast issue date closest to the OND rainy season in Kenya is the 1st August, which is a lead of 2 months. Chapter 4.3.3 shows that the DEMETER rainfall hindcast skill is optimised at 0-month lead, decreasing to zero skill at 3-months lead. Therefore, a hindcast issued on the 1st August for OND would have no hindcast skill for rainfall in November and December. In contrast to DEMETER, the EUROSIP hindcasts are issued monthly. Thus it is possible to directly compare a 0-lead OND hindcast from EUROSIP against a statistical OND hindcast.

	DEMETER			EUROSIP		
	ECMWF	Météo France	UKMO	ECMWF	Météo France	UKMO
System	2	2	2	3	3	3
Atmospheric Component	ECMWF IFS	ARPEGE v4	GLOSEA 1	ECMWF IFS	ARPEGE v4.4	GLOSEA 3
Atmospheric Horizontal Resolution	1.9° x 1.9°	2.8° x 2.8°	2.5°x3.75°	1.125° x 1.25°	2.8° x 2.8°	2.5°x3.75°
Atmospheric Vertical Levels	40	31	19	62	91	19
Oceanic Component	HOPE	ORCA	GLOSEA OGCM	HOPE	ORCA	GLOSEA OGCM
Oceanic Horizontal Resolution	1° x 1°	2° x (2° to 0.5 at Equator)	1.25° x (1.25° to 0.3° at Equator)	1° x 1°	2° x (2° to 0.5 at Equator)	1.25° x (1.25° to 0.3° at Equator)
Oceanic Vertical Levels	29	31	40	29	31	40
Hindcast Ensemble Members	9	9	9	11	15	15
Operational Forecast Ensemble Members	Not Operational	Not Operational	Not Operational	41	41	41
Hindcast Years Available	1958-2001 (44 years)	1958-2001 (44 years)	1959-2001 (43 years)	1981-2005 (25 years)	1981-2005 (25 years)	1987-2005 (19 years)
Hindcast/Forecast Issue Dates	1 st Feb, 1 st May, 1 st Aug, 1 st Nov	1 st Feb, 1 st May, 1 st Aug, 1 st Nov	1 st Feb, 1 st May, 1 st Aug, 1 st Nov	15 th of every month	15 th of every month	15 th of every month
Hindcast/Forecast Lead Times* (months)	0-5	0-5	0-5	0-6	0-6	0-5
Grid Resolution	2.5° x 2.5°	2.5° x 2.5°	2.5° x 2.5°	2.5° x 2.5°	2.5° x 2.5°	2.5° x 2.5°

Table 8.1. Characteristics of the ECMWF, Météo France and UKMO coupled general circulation models used in the DEMETER and EUROSIP systems (Anderson et al., 2007; Davey and Ferranti, 2006; ECMWF, 2009; Palmer et al., 2004; Vitart et al., 2007).

*Hindcast lead time refers to the period of time between the issue of the hindcast and the start of the hindcast period

8.2 Methodology for comparing statistical and dynamical hindcast models for the Kenyan ‘short-rains’

8.2.1 EUROSIP hindcast skill assessment

The OND rainfall hindcast skills available from the individual EUROSIP models and the MSE-weighted multi-model ensemble (defined in Chapter 4.1.2 and Equation 4.3) are investigated. Skill assessment methods are based on those described in Chapter 4. Firstly, the EUROSIP hindcasts are bias corrected using the standard cross validation method described in Section 4.2.1. The observed rainfall data are taken from the CMAP data set (Chapter 3.1.1). These data are available on a $2.5^{\circ} \times 2.5^{\circ}$ latitude/longitude grid over the period 1979-2008. The CMAP data are bias corrected over each grid square over the periods 1981-2005 and 1987-2005. These correspond to the periods of available hindcasts from the individual EUROSIP models: 1981-2005 for the ECMWF and Météo France models, and 1987-2005 for the UKMO model and the multi-model ensemble (Table 8.1). The OND hindcasts from each EUROSIP model are then bias corrected per grid square over these available hindcast periods.

The skills of the hindcasts available from the individual EUROSIP models and the MSE-weighted multi-model ensemble are then assessed using the methods described in Chapter 4.2. The assessments are made per grid square over the available hindcast period of each model.

8.2.2 Statistical hindcast skill assessment

In order to conduct a direct comparison between the statistical and dynamical models, the statistical models are re-developed over the periods 1981-2005 and 1987-2005, using the methods described in Chapter 5.3. The original homogeneous rainfall regions and their associated best predictor regions are used (as shown in Table 6.1, marked with an asterisk), with the data detrended over the new verification periods. The skills of the OND rainfall hindcasts for each region and verification period are then assessed using the methods described in Chapters 4.2.3 and 4.2.4. The hindcast skill scores are assessed using a block-removal cross-validation method as described in Chapter 5.4.

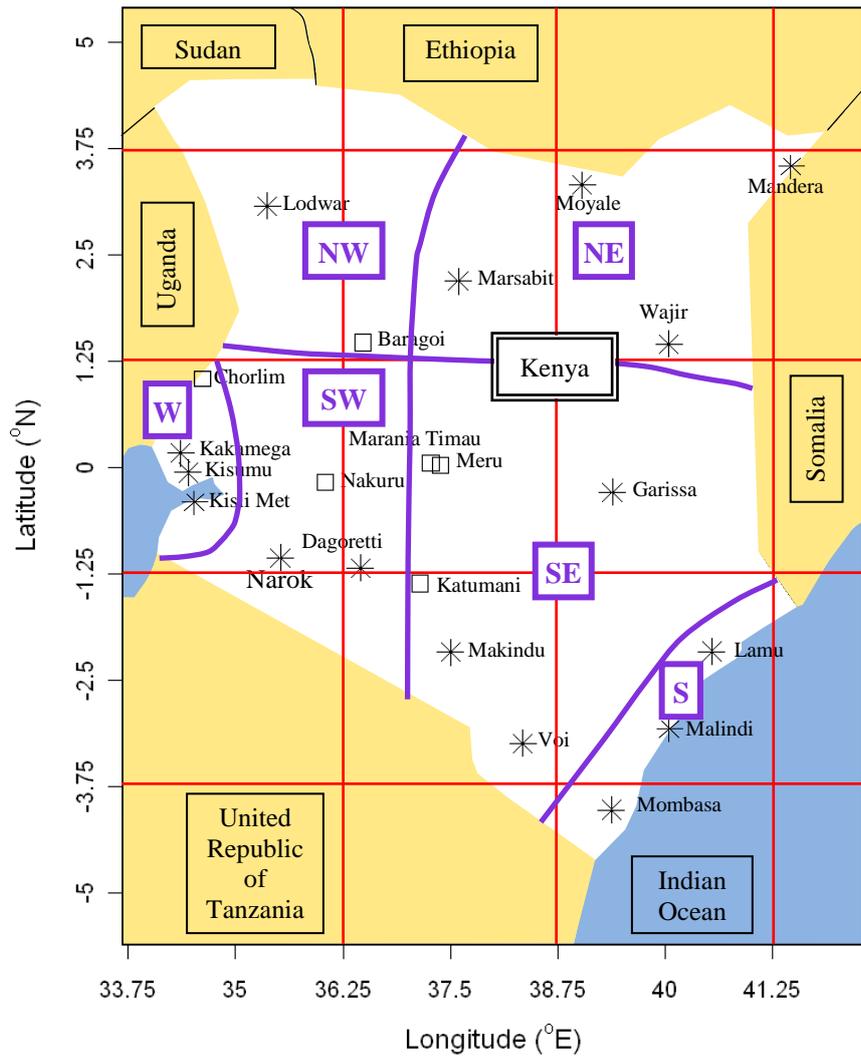


Figure 8.1. Kenyan rain gauge stations selected for use in the study. The homogeneous rainfall region boundaries are shown in dark blue and the EURO-SIP and CMAP grid square boundaries are shown in red. The 16 Synoptic rain gauge stations are marked with an asterisk (*) and 6 non-synoptic rain gauge stations are marked by a square (□). The names of the regions are shown in purple boxes: South (S), Southeast (SE), Southwest (SW), West (W), Northwest (NW) and Northeast (NE).

Region	Southwest	West	South
Coordinates	(35°E, 0.0°S)	(35°E, 0.0°S)	(40°E, 2.5°S)
<i>r_{rank}</i>	0.91	0.79	0.94
Region	Northeast	Northwest	Southeast
Coordinates	(40°E, 2.5°N)	(37.5°E, 2.5°N)	(37.5°E, 0.0°N)
<i>r_{rank}</i>	0.99	0.51	0.94

Table 8.2. Coordinates of the grid squares with the strongest *r_{rank}*. Correlation values are calculated over the period 1981-2005 between OND CMAP rainfall and each regional OND rainfall index.

8.2.3 Direct comparison methods

It is difficult to make a direct comparison between the hindcasts from the statistical and EUROSIP models due to the gridded format of the EUROSIP hindcasts. These are issued on a $2.5^{\circ} \times 2.5^{\circ}$ latitude/longitude grid as presented in Figure 8.1. The red solid lines show the boundaries of the EUROSIP and CMAP grid squares and the dark blue lines show the boundaries of the homogeneous rainfall regions. There are several EUROSIP/CMAP grid squares that cover multiple homogeneous rainfall regions. For example, the grid square centred at 35°E and 0°N covers the West and Southwest regions. There are also several rainfall regions that are covered by multiple grid squares. For example, the Southeast region is covered by four grid squares. In order to select a grid square to represent each homogeneous rainfall region, the regional OND rainfall indices are correlated against the CMAP rainfall data from each overlapping grid square. The coordinates of the grid squares with the strongest correlation between OND CMAP rainfall indices and the regional OND rainfall indices over the period 1981-2005 are shown in Table 8.2. The EUROSIP hindcasts from these grid squares will be used to represent the dynamical hindcasts for each homogeneous rainfall region. This will allow a direct comparison between the dynamical and statistical hindcast skills for each region. Only one model grid square is selected to represent each region in order to be compatible with the methods used by forecaster in East Africa. This should help to make it easier for the Kenyan forecasters to use the results of this study operationally. It must also be noted that the regional seasonal rain gauge rainfall indices from the KMD were also used for comparison in this Chapter, giving very similar results. This gives us confidence in the realistic representation of the CMAP rainfall data over Kenya.

8.3 Comparison results

The following sub-sections compare the OND rainfall hindcast skill available from the statistical and dynamical hindcast models for each region of Kenya. Tables 8.3 to 8.8 show the OND rainfall hindcast skill assessment results, for each region of Kenya, from the individual EUROSIP models, the EUROSIP multi-model and the best regional statistical models over the verification periods 1981-2005 and 1987-2005. In each table, the model with the best overall skill score is shown in red. In the event of several models producing hindcasts with strong skill scores, the greatest weighting is given to

r_{rank} . The dynamical model with the strongest skill score is highlighted in green and the statistical model with the strongest skill score is highlighted in blue.

8.3.1 Statistical models: Southwest and West Kenya

The best statistical hindcast model is found to outperform the EUROSIP models in hindcasting OND rainfall for the **Southwest** (Table 8.3) and **West** (Table 8.4) regions of Kenya. These regions are home to 67.7% of the total population of Kenya (CIESIN, 2005) (Chapter 5.1.2).

The best statistical forecast models for the Southwest and West regions are based on the best predictors regions from Chapter 6 (marked with an asterisk in Table 6.1). Tables 8.3 and 8.4 show respectively the skill assessment results for the Southwest and West regions of Kenya. The statistical OND rainfall hindcasts are consistently more skilful than climatology, with $MSSS > 23\%$ over both verification periods. The correlation results are also high with $r_{rank} > 0.4$ (p -values ≤ 0.06) over both verification periods. In comparison, the OND hindcast skill scores from the individual EUROSIP models and multi-model ensemble are consistently lower, with $r_{rank} < 0.4$ (p -values ≥ 0.07).

Model	Period	Lead Time	r_{rank}	p -value	$MSSS$ (%)
UKMO	1987-2005	0	0.29	0.11	15.9
UKMO	1987-2005	1	0.16	0.26	0
UKMO	1987-2005	2	0.11	0.31	0
ECMWF	1981-2005	0	0.04	0.42	0
ECMWF	1981-2005	1	0.10	0.33	0
ECMWF	1981-2005	2	-0.08	0.37	0
Météo France	1981-2005	0	0.21	0.16	0
Météo France	1981-2005	1	0.19	0.18	0
Météo France	1981-2005	2	0.04	0.43	0
Multi-model	1987-2005	0	0.36	0.07	8.3
Multi-model	1987-2005	1	0.22	0.18	0
Multi-model	1987-2005	2	-0.11	0.32	0
Statistical	1981-2005	0	0.42	0.07	23.6
Statistical	1987-2005	0	0.42	0.06	28.5

Table 8.3. Comparison of the **Southwest** OND rainfall hindcast skill assessments from the individual EUROSIP models, the EUROSIP multi-model and the best statistical model. The EUROSIP results are from the grid square representing the Southwest region of Kenya (35°E, 0.0°N). The model with the best overall skill score is shown in red. The dynamical model with the strongest skill score is highlighted in green. The statistical model with the strongest skill score is highlighted in blue.

Model	Period	Lead Time	r_{rank}	p -value	$MSSS$ (%)
UKMO	1987-2005	0	0.29	0.11	15.9
UKMO	1987-2005	1	0.16	0.26	0
UKMO	1987-2005	2	0.11	0.31	0
ECMWF	1981-2005	0	0.04	0.42	0
ECMWF	1981-2005	1	0.1	0.33	0
ECMWF	1981-2005	2	-0.08	0.37	0
Météo France	1981-2005	0	0.21	0.16	0
Météo France	1981-2005	1	0.19	0.18	0
Météo France	1981-2005	2	0.04	0.43	0
Multi-model	1987-2005	0	0.36	0.07	8.3
Multi-model	1987-2005	1	0.22	0.18	0
Multi-model	1987-2005	2	-0.11	0.32	0
Statistical	1981-2005	0	0.59	0.01	33.7
Statistical	1987-2005	0	0.55	0.01	25.4

Table 8.4. As Table 8.3 but for the **West** of Kenya. The EUROSIP results are from the grid square representing Western Kenya (35°E, 0.0°N). Note that the dynamical skill scores are the same as in Table 8.3 as the same EUROSIP grid square covers both the West and the majority of the Southwest regions (as shown in Figure 8.1).

8.3.2 Dynamical models: South and Northeast Kenya

The EUROSIP multi-model ensemble is found to outperform the most skilful statistical models in hindcasting OND rainfall for the **South** (Table 8.5) and **Northeast** (Table 8.6) regions of Kenya. These regions are home to 4.7% of the total population of Kenya (CIESIN, 2005) (Chapter 5.1.2).

The best statistical forecast models for the South and Northeast regions of Kenya are based on the best predictors regions from Chapter 6 (marked with an asterisk in Table 6.1). Table 8.5 shows the skill assessment results for Southern Kenya. The OND rainfall hindcast skill from the EUROSIP multi-model outperforms climatology with $MSSS > 28\%$ between 1987 and 2005. The correlation result is the highest of any region with $r_{rank} = 0.70$ (p -value < 0.01). The skill scores from the best statistical hindcast model are lower with $r_{rank} = 0.53$ (p -value < 0.01) and $MSSS = 13.5\%$ over the period 1981-2005.

The skill assessment results for the Northeast region of Kenya are presented in Table 8.6. The OND rainfall hindcast skills from the EUROSIP multi-model are high with $r_{rank} = 0.60$ (p -value < 0.03) and $MSSS > 9\%$ over the period 1987-2005. In comparison, the skill scores from the best statistical hindcast model are lower, with $r_{rank} = 0.24$ (p -value = 0.11) and $MSSS < 0\%$ over the period 1981-2005.

Model	Period	Lead Time	r_{rank}	p -value	MSSS (%)
UKMO	1987-2005	0	0.67	0.01	35.7
UKMO	1987-2005	1	0.19	0.23	19.1
UKMO	1987-2005	2	0.20	0.21	20.4
ECMWF	1981-2005	0	0.50	0.01	31.9
ECMWF	1981-2005	1	0.19	0.18	18.4
ECMWF	1981-2005	2	0.07	0.36	5.4
Météo France	1981-2005	0	0.24	0.15	6.1
Météo France	1981-2005	1	-0.03	0.45	0
Météo France	1981-2005	2	0.23	0.13	7.1
Multi-model	1987-2005	0	0.70	0.01	28.9
Multi-model	1987-2005	1	0.41	0.04	16.8
Multi-model	1987-2005	2	0.26	0.14	12.7
Statistical	1981-2005	0	0.53	0.01	13.5
Statistical	1987-2005	0	0.47	0.03	6.1

Table 8.5. As Table 8.3 but for the **South** of Kenya. The EUROSIP results are from the grid square representing the South region of Kenya (40°E, 2.5°S).

Model	Period	Lead Time	r_{rank}	p -value	MSSS (%)
UKMO	1987-2005	0	0.51	0.03	22.1
UKMO	1987-2005	1	0.24	0.18	15.2
UKMO	1987-2005	2	-0.02	0.46	7.9
ECMWF	1981-2005	0	0.44	0.02	16.8
ECMWF	1981-2005	1	0.40	0.05	8.6
ECMWF	1981-2005	2	0.27	0.10	6.2
Météo France	1981-2005	0	0.30	0.10	6.5
Météo France	1981-2005	1	0.10	0.32	0
Météo France	1981-2005	2	0.14	0.27	1
Multi-model	1987-2005	0	0.58	0.02	16.1
Multi-model	1987-2005	1	0.60	0.03	9.6
Multi-model	1987-2005	2	0.09	0.35	6
Statistical	1981-2005	1	0.24	0.11	0
Statistical	1987-2005	1	0.17	0.24	0

Table 8.6. As Table 8.3 but for the **Northeast** of Kenya. The EUROSIP results are from the grid square representing the Northeast region of Kenya (40°E, 2.5°N).

8.3.3 Northwest and Southeast Kenya

The skill assessment results for the **Northwest** and **Southeast** regions of Kenya are shown in Tables 8.7 and 8.8 respectively. Both statistical and dynamical hindcast models produce low OND rainfall hindcast skill scores for these regions.

The best statistical forecast models for the Northwest and Southeast regions of Kenya are based on the best predictors regions from Chapter 6 (marked with an asterisk in Table 6.1). Table 8.7 shows the skill assessment results for the Northwest region of Kenya. This region is home to 1.3% of the total population of Kenya (CIESIN, 2005) (Chapter 5.1.2). The statistical OND rainfall hindcast is more skilful than those from the EUROSIP models with $r_{rank} = 0.3$ (p -value = 0.17) and $MSSS = 4.3\%$ over the period 1987-2005. However, this skill does not extend to the period 1981-2005. The skill scores from the dynamical models are lower. The best performing dynamical model is from Météo France, with $r_{rank} = 0.19$ (p -value = 0.17) with $MSSS = 2.1\%$ over the period 1981-2005.

Model	Period	Lead Time	r_{rank}	p -value	$MSSS$ (%)
UKMO	1987-2005	0	-0.02	0.47	0
UKMO	1987-2005	1	-0.08	0.38	0
UKMO	1987-2005	2	-0.20	0.22	0
ECMWF	1981-2005	0	-0.13	0.27	0
ECMWF	1981-2005	1	0.11	0.32	0
ECMWF	1981-2005	2	-0.13	0.27	0
Météo France	1981-2005	0	0.15	0.23	2.2
Météo France	1981-2005	1	0.19	0.17	2.1
Météo France	1981-2005	2	0.09	0.33	0
Multi-model	1987-2005	0	0.16	0.27	0
Multi-model	1987-2005	1	0.10	0.35	2.1
Multi-model	1987-2005	2	-0.23	0.18	0
Statistical	1981-2005	1	0.15	0.29	0
Statistical	1987-2005	1	0.30	0.17	4.3

Table 8.7. As Table 8.3 but for the **Northwest** of Kenya. The EUROSIP results are from the grid square representing the Northwest region of Kenya (37.5°E, 2.5°N).

The skill assessment results from the Southeast region of Kenya are presented in Table 8.8. This region is home to 26.3% of the total population of Kenya (CIESIN, 2005) (Chapter 5.1.2). The EUROSIP multi-model OND rainfall hindcast is more skilful than the hindcast from the statistical model with $r_{rank} = 0.39$ (p -value = 0.07) and $MSSS = 6.5\%$ over the period 1987-2005. The UKMO is the strongest contributing individual model with $r_{rank} = 0.32$ (p -value < 0.09) and $MSSS = 12.4\%$. The skill scores from the best statistical model are lower, with $r_{rank} = 0.10$ (p -value = 0.32) with $MSSS < 0\%$ over the period 1981-2005.

Model	Period	Lead Time	r_{rank}	p -value	$MSSS$ (%)
UKMO	1987-2005	0	0.32	0.09	12.4
UKMO	1987-2005	1	0.22	0.20	1.5
UKMO	1987-2005	2	0.24	0.16	3.6
ECMWF	1981-2005	0	-0.10	0.31	0
ECMWF	1981-2005	1	0.06	0.40	0
ECMWF	1981-2005	2	-0.04	0.43	0
Météo France	1981-2005	0	0.20	0.16	0
Météo France	1981-2005	1	0.18	0.19	0
Météo France	1981-2005	2	0.15	0.24	1.9
Multi-model	1987-2005	0	0.39	0.07	6.5
Multi-model	1987-2005	1	0.31	0.09	0.3
Multi-model	1987-2005	2	0.17	0.24	0
Statistical	1981-2005	1	0.10	0.32	0
Statistical	1987-2005	1	0.05	0.42	0

Table 8.8. As Table 8.3 but for the **Southeast** of Kenya. The EUROSIP results are from the grid square representing the Southeast region of Kenya (37.5°E, 0.0°N).

8.4 Discussion

A summary of the comparison of the best OND rainfall hindcast models for each region of Kenya is presented in Table 8.9. Statistical models are found to produce the best OND rainfall hindcasts for half of the regions in Kenya, comprising 69% of the total population. The EUROSIP dynamical models are found to produce the best OND rainfall hindcasts for the other half of Kenya, which is home to 31% of the total population. Table 8.9 shows that the models offer potentially useful skill (with $r_{rank} \geq 0.39$ and p -values ≤ 0.07) for 98.7% of the Kenyan population (based on 2000 population figures). The available lead time is either 0- or 1-month for the OND season depending on region (Tables 8.3-8.8). Why do all models produce low hindcast skill for the Northwest region? A likely explanation is the poor quality of the rainfall data in this region (as discussed in Chapter 6.2.5).

There is scope for further research in this area. In the cases where there is a large difference in OND hindcast skill, it could be beneficial to examine the causes behind the skill difference. This would provide direction to improve the skill of the less skilful model type. Further improvements to the OND rainfall hindcast skill could be made by combining the statistical and EUROSIP multi-model hindcasts. Finally, it would also be interesting to examine the MAM rainfall hindcast skill of the EUROSIP multi-model.

Region	Best Model	Statistical Model Predictor	Dynamical Model Grid Square	r_{rank}	p -value	$MSSS$ (%)	Population (millions) 2000	Population (%) 2000	Population (millions) 2015	Population (%) 2015
Southwest	Statistical	AS SSTs off East African Coast	(35°E, 0.0°S)	0.42	0.06	28.5	11.1	38.1	15.6	41.4
West	Statistical	AS SSTs off East African Coast	(35°E, 0.0°S)	0.59	0.01	33.7	8.6	29.6	10.6	28.2
South	Multi-model	NA	(40°E,2.5°S)	0.70	0.01	28.9	0.95	3.3	1.3	3.4
Northeast	Multi-model	NA	(40°E,2.5°N)	0.60	0.03	9.6	0.42	1.4	0.68	1.8
Northwest	Statistical	JJA SSTs off East African Coast	(37.5°E,2.5°N)	0.30	0.17	4.3	0.37	1.3	0.52	1.4
Southeast	Multi-model	NA	(37.5°E,0°N)	0.39	0.07	6.5	7.6	26.3	9.0	23.8

Table 8.9. Summary of the best OND rainfall hindcast models for each region of Kenya. The skill scores are shown for the period 1987-2005 (except for the West region of Kenya, where skill scores are shown for the period 1981-2005).

Regional population figures are shown (in millions of people and percentage of total Kenyan population) for 2000 with future estimates for 2015. Population data are from the Gridded Population of the World, Version 3, produced by CIESIN (Centre for International Earth Science Information Network) (CIESIN, 2005).

Chapter 9

Conclusions and Wider Implications

9.1 Summary and Conclusions

The aim of this thesis has been to investigate the seasonal prediction of African rainfall, with a focus on Kenya. Africa's climate is prone to extended rainfall deficits. In extreme cases these can lead to droughts and humanitarian disasters. Skilful prediction of seasonal rainfall would therefore bring sound humanitarian and economic benefits to countries, such as Kenya, that are highly dependent on rain-fed agriculture.

The thesis first performed a detailed assessment of the current seasonal rainfall hindcast skill available from leading dynamical models over Africa. The DEMETER individual model and multi-model seasonal rainfall hindcasts are found to have weak correlation with the observed GPCP rainfall data over most of Africa at zero lead for the period 1959-2000 (Chapter 4) with the following two exceptions. In the sub-Sahara/Sahel belt moderate ASO rainfall hindcast skill is found from the DEMETER multi-model ensemble hindcasts with r_{rank} values of 0.3-0.8 (p -values < 0.1) and positive $MSSS$ values. In Equatorial East Africa and around Nigeria and South Africa moderate NDJ rainfall hindcast skill is found from the DEMETER multi-model ensemble hindcasts with r_{rank} values of 0.5-0.6 (p -values < 0.1) and positive $MSSS$ values. In both cases this skill disappears as the lead increases from 0- to 3-months. Many factors may be responsible for the overall low hindcast skill across Africa. One contributing factor is the lower variance in the hindcasts, compared to the observations. Another factor is the uneven distribution of rain gauges across the continent. The areas that show the highest DEMETER hindcast skill are found to be coincident with those areas of Africa that have the highest concentrations of rain gauge stations.

The thesis then progressed to focus on Kenya, as it has experienced 7 severe droughts over the period 1991-2008, affecting over 35 million people. Kenya has a highly variable spatial rainfall distribution. Therefore, it is necessary to split the country into

six homogeneous rainfall regions (Chapter 5). Statistical OND rainfall hindcast models are developed for each region, using linear regression techniques as described in Chapter 5. Predictors are selected from lagged SST and atmospheric wind fields (Chapter 3) based on having significant, temporally-stable correlations with regional rainfall indices and clear physical linking mechanisms.

Chapter 6 shows that temporally stable, moderate-to-high OND rainfall hindcast skills are found for the West, Southwest, South and Northeast regions of Kenya with average $r_{rank} \geq 0.40$ (p -value ≤ 0.06) and $MSSS > 0$. These regions comprise 72.4% of the total population of Kenya. The strongest hindcast skill scores in Kenya are produced by the best OND rainfall hindcast model for the West and Southwest regions of Kenya with average r_{rank} values of 0.55 (p -value < 0.04) and 0.47 (p -value < 0.05) respectively. These models are based on SST predictors located in the Indian Ocean.

The statistical OND rainfall hindcast models for the Northwest and Southeast regions of Kenya produce the lowest OND rainfall hindcast skill scores. These have average r_{rank} values of 0.38 (p -value ≤ 0.10) and 0.27 (p -value ≤ 0.19) respectively. It is suggested that the lack of Northwest OND rainfall hindcast skill is mainly a result of poor quality rainfall data. The OND rainfall in the Southeast region of Kenya is found to be linked to an area of SST off the coast of Japan. Saji et al. (1999) found that this region of SST was linked to East African rainfall through the IOD. However, it seems that this region of SST is not a strong predictor, as the skill results are the lowest of all the regions.

Few strong, significant and temporally stable potential predictors are found for the MAM rainy season or for its individual comprising months (Chapter 7). It is possible that a proportion of the strong correlations between regional MAM rainfall indices and potential predictor indices are statistical artefacts. The results of Chapter 7 therefore need to be viewed with caution. Thus it is not possible to use the potential predictors with confidence to develop regional rainfall forecast models for the 'long-rains' season in Kenya. Previous research has suggested that the QBO could be a strong predictor for the Kenyan 'long-rains' (Indeje and Semazzi, 2000). The QBO has been accepted by Kenyan forecasters as a skilful operational predictor for this season. However, this study shows that the QBO does not have strong, temporally stable correlations with the regional MAM rainfall indices in Kenya. The QBO should, therefore, not be used operationally as a predictor for the Kenyan 'long-rains'.

The thesis concluded by comparing statistical and dynamical hindcast models over Kenya, in order to determine which produces the most skilful OND rainfall hindcasts (Chapter 8). The results of this comparison could help to inform Kenyan forecasters as to which method has the greatest OND rainfall forecast skill for each region. This is the first independent study to use the hindcasts from the EUROSIP multi-model ensemble system, which is an improved version of DEMETER. The statistical hindcast models are found to produce the most skilful OND rainfall hindcast, compared to those from EUROSIP, in the most populous and heavily cultivated regions in Kenya. These regions of the west and southwest are home to 67.7% of Kenya's population. The 19.7 million people living in these regions could directly benefit from operational use of these statistical forecasts, rather than dynamical forecasts. The EUROSIP multi-model ensemble is shown to produce more skilful OND rainfall hindcasts than the statistical models in the South and Northeast regions of Kenya. These less populous regions comprise 4.7% of the total population of Kenya. Both statistical and dynamical hindcast models produce low OND rainfall hindcast skill scores for the Northwest and Southeast regions of Kenya. The Northwest region is sparsely populated with 0.37 million people (1.3% of the total population). However, the Southeast region comprises 7.6 million people (26.3% of the total population). The poor hindcasts from both the dynamical and statistical models in the southeast region could leave this large population exposed to drought and famine with no prior warning.

9.2 Recent drought prediction skill

The prime motivation for this study was the high number of severe droughts that affect African countries such as Kenya, and the humanitarian impacts that these often lead to. Skilful prediction of seasonal rainfall would bring sound humanitarian and economic benefit to Kenya, but only if these predictions are skilful in years of rainfall deficit that lead to severe drought. This section investigates how well the statistical and dynamical hindcast models discussed and developed in this thesis performed over two of the most recent and notable droughts in Kenya.

The most recent and notable droughts to affect Kenya occurred in 1999-2002 and 2005 affecting 23 and 3.5 million people respectively (Table 1.1 and Section 2.2.2). The 1999-2002 drought affected people in the west and central areas of Kenya and is

Year	OND seasonal rainfall (mm/season)								
	Southeast			Southwest			Northeast		
	Rain gauge	Statistical	EUROSIP	Rain gauge	Statistical	EUROSIP	Rain gauge	Statistical	EUROSIP
1999	99	122	137	<u>68</u>	<u>61</u>	<u>98</u>	55	45	74
2000	77	114	134	<u>96</u>	<u>86</u>	<u>99</u>	37	67	75
2001	83	93	119	<u>64</u>	<u>54</u>	<u>88</u>	51	88	70
2005	<u>57</u>	<u>82</u>	<u>112</u>	29	37	74	<u>25</u>	<u>61</u>	<u>69</u>
Year	OND seasonal rainfall (mm/season)								
	Northwest			West			South		
	Rain gauge	Statistical	EUROSIP	Rain gauge	Statistical	EUROSIP	Rain gauge	Statistical	EUROSIP
1999	<u>32</u>	<u>54</u>	<u>90</u>	<u>78</u>	<u>84</u>	<u>98</u>	73	37	81
2000	<u>31</u>	<u>56</u>	<u>84</u>	<u>99</u>	<u>88</u>	<u>99</u>	76	66	79
2001	<u>3</u>	<u>45</u>	<u>77</u>	<u>95</u>	<u>79</u>	<u>88</u>	50	94	75
2005	<u>N/A</u>	<u>26</u>	<u>76</u>	77	75	74	<u>51</u>	<u>40</u>	<u>76</u>

Table 9.1. OND seasonal rainfall (mm/season) from the rain gauges of the KMD, and hindcasts from the best statistical models for each region (Chapter 6) and the EUROSIP multi-model ensemble at 0-lead. The values are shown for 1999-2001 and 2005 as examples of recent notable droughts in Kenya. The underlined values highlight the regions that were most badly affected in each drought year. 2005 rain gauge data from the Northwest region of Kenya is not available.

thought to have been forced by the 1999-2001 La Niña. The drought in 2005 affected the north and eastern areas of Kenya and was forced by the failure of the OND rains in 2005. Table 9.1 displays the OND seasonal regional rainfall totals recorded by the KMD over the years 1999-2001 and 2005 as well as the corresponding hindcast values from the best statistical model for each region and from the EUROSIP multi-model ensemble at 0-lead. This is a clear way of showing which model produced the most accurate seasonal regional hindcast values in each case.

It can be seen from Table 9.1 that in 70% of cases the statistical models produced a rainfall hindcast value that was closer to the observed value than that produced from EUROSIP. The regions that were reported to be most badly affected in each drought are underlined in Table 9.1. In 75% of these cases the statistical models produced a more accurate hindcast than those from the EUROSIP multi-model ensemble at 0-lead. In the majority of cases the seasonal rainfall hindcast values from the statistical models were very close to the regional rain gauge values from KMD. This shows that the rainfall deficits responsible for these two high impact droughts would have been predictable for most of the regions of Kenya. These results give added confidence in using these statistical forecast models operationally over Kenya.

9.3 Benefits to operational forecasting in Kenya

Chapter 6 has shown that it is possible to develop consistently skilful statistical OND rainfall forecast models for the West, Southwest, South and Northeast regions of Kenya. These regions are home to 72.4% of the total population of Kenya. Further research into improving the forecast skill for the remaining regions is necessary, especially the Southeast, before these forecast models could be used as a real time operational forecast system for the whole country.

This study has highlighted several points that may be useful to forecasters in Kenya. Firstly, The Kenyan Meteorological Department issues seasonal rainfall forecasts before both the 'short-' and 'long-rains' seasons. The forecasts for the 'long-rains' season are based on weak correlations between regional rainfall indices and SSTs and the QBO. The skill of these operational MAM forecasts may be questionable due to the poor quality of the predictors available for this season. This study recommends that the

Kenyan Meteorological Department should make the users aware of the reliability of the MAM forecasts. Users will then be able to make an informed decision on how much confidence they can have in the forecasts.

Secondly, the most consistently skilful regional OND rainfall hindcasts found in this study are from forecast models based on Indian Ocean SST predictors off the coast of East Africa. Convective activity and rainfall are enhanced over this region by positive SST anomalies. Similarly, negative SST anomalies here tend to suppress East African rainfall. Currently, operational forecasters in Kenya use SYSTAT to automatically select predictor regions with strong correlations to Kenyan rainfall. These predictor regions are currently accepted for operational use with no consideration of whether a viable physical mechanism exists to link them to the ‘short-rains’. Improvements to forecast skill could be made by only accepting predictor regions that have established physical links to the Kenyan ‘short-rains’. The predictor regions found in this study could also be used by the Kenyan forecasters when developing their OND rainfall forecast models in addition to those found through the SYSTAT program. Chapter 8 shows the optimal combination of statistical and dynamical forecast models to maximise regional OND rainfall hindcast skill across Kenya. This could assist forecasters in deciding which type of forecast model to trust for each region of Kenya.

9.4 Future directions

This thesis has focused on the seasonal prediction of African rainfall, with a focus on Kenya. The skilful OND rainfall forecasts presented in this thesis could be used to explore the development of useful regional Kenyan drought prediction models. This work would need to be conducted in partnership with the end users, in order to optimise the usefulness of the forecasts. The first step would be for the user to determine which definition of drought the user is most interested in (Chapter 2.1.1) and to quantifying this. Thresholds of seasonal rainfall deficit would then need to be set to establish when a drought is likely to occur for each region. The results of this study should be taken into account when deciding whether or not to act on the seasonal rainfall forecasts for each region. Drought prediction models are very complex and need to consider variables such as: duration of rainfall deficit, surface land use, soil moisture capacity,

groundwater storage, streamflow, evapotranspiration and water demand in order to determine the extent of a drought based on the seasonal rainfall deficit predicted.

Scope exists for further research into the variability of the Kenyan 'long-rains'. Further research is required to try to identify and understand the physical factors which affect the year-to-year variability in Kenyan regional MAM rainfall. This understanding would be a pre-requisite for improving Kenyan MAM rainfall predictions.

This was the first study to assess the skill of the EUROSIP seasonal rainfall hindcasts (independently of the developing institutions) as the hindcasts were only made available to the public in 2009. It would be interesting to directly compare the hindcasts over Africa available from the DEMETER and EUROSIP systems in order to quantify the improvement in skill that has resulted from the model upgrades. A more detailed skill assessment of the EUROSIP seasonal rainfall hindcasts over the whole of Africa would allow users to know how much confidence they can have in the EUROSIP forecasts for their region of Africa. Further work is also needed to examine whether merging the statistical and EUROSIP models would lead to an improved multi-model prediction of OND Kenyan rainfall.

It would be useful to examine the potential for using satellite-derived rainfall data to improve the quality of the Kenyan historical rain gauge data. Satellite-derived rainfall data have a shorter available climatology than rain gauge data. However, they have other benefits, such as higher spatial and temporal resolutions. This could lead to improvements to the seasonal rainfall forecast skill, particularly in regions of sparse rain gauge density.

Finally, it is hoped that future studies may be able to apply the techniques developed in this thesis for the seasonal prediction of regional rainfall in other East African countries that are also prone to rainfall deficits and droughts.

Appendix A
Predictor Selection Plots for OND Rainfall
for each Region of Kenya

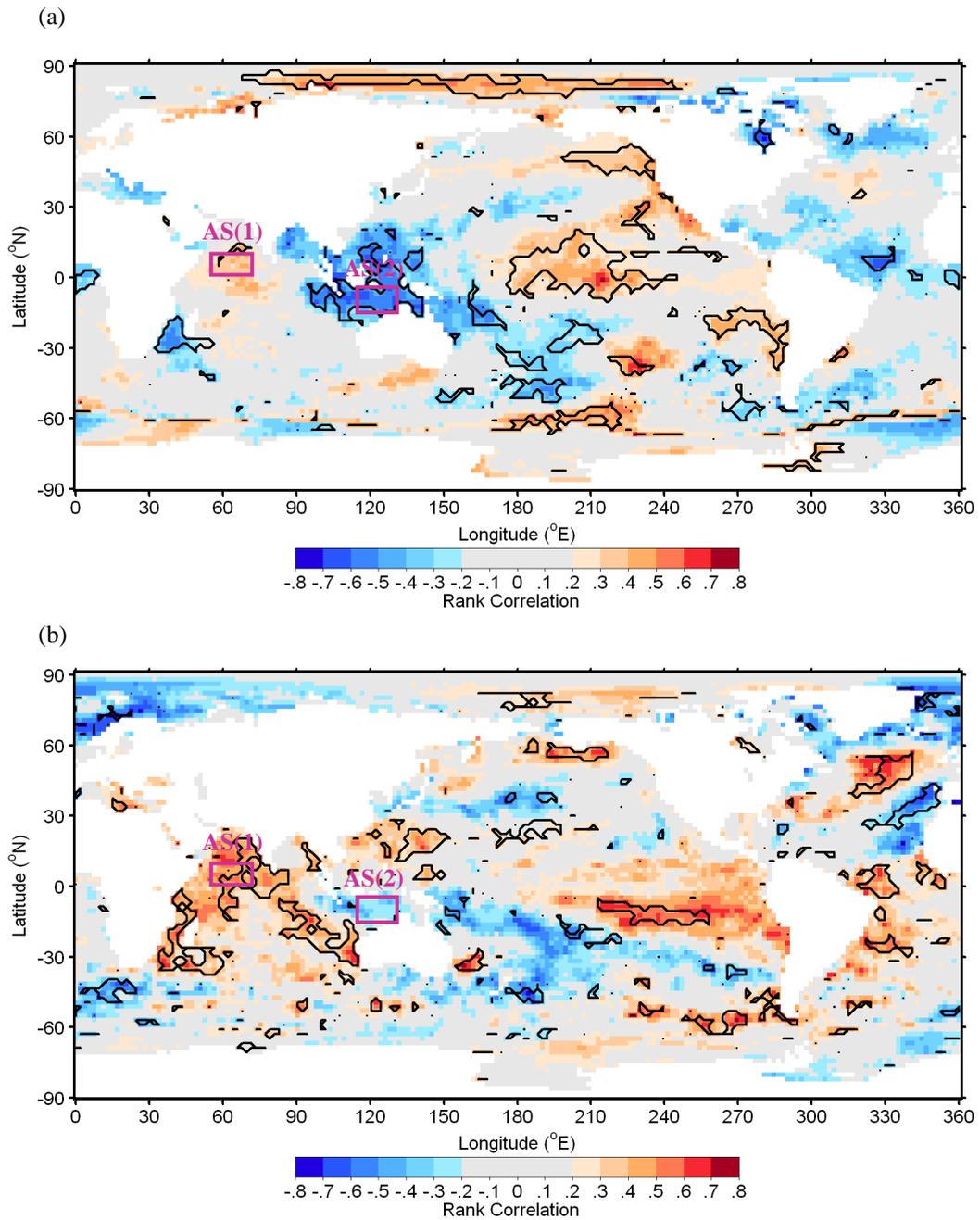


Figure A.1. West October-December (OND) rainfall index correlated with prior August-September (AS) two-month average SST global field over the periods: (a) 1959-1974 and (b) 1975-1990. White areas denote land mass. Red (blue) shows positive (negative) correlations, grey shows low correlations (between -0.2 and 0.2) and the black lines show areas with p -values < 0.15 . The SST AS(1) and SST AS(2) predictor regions are highlighted by the purple boxes.

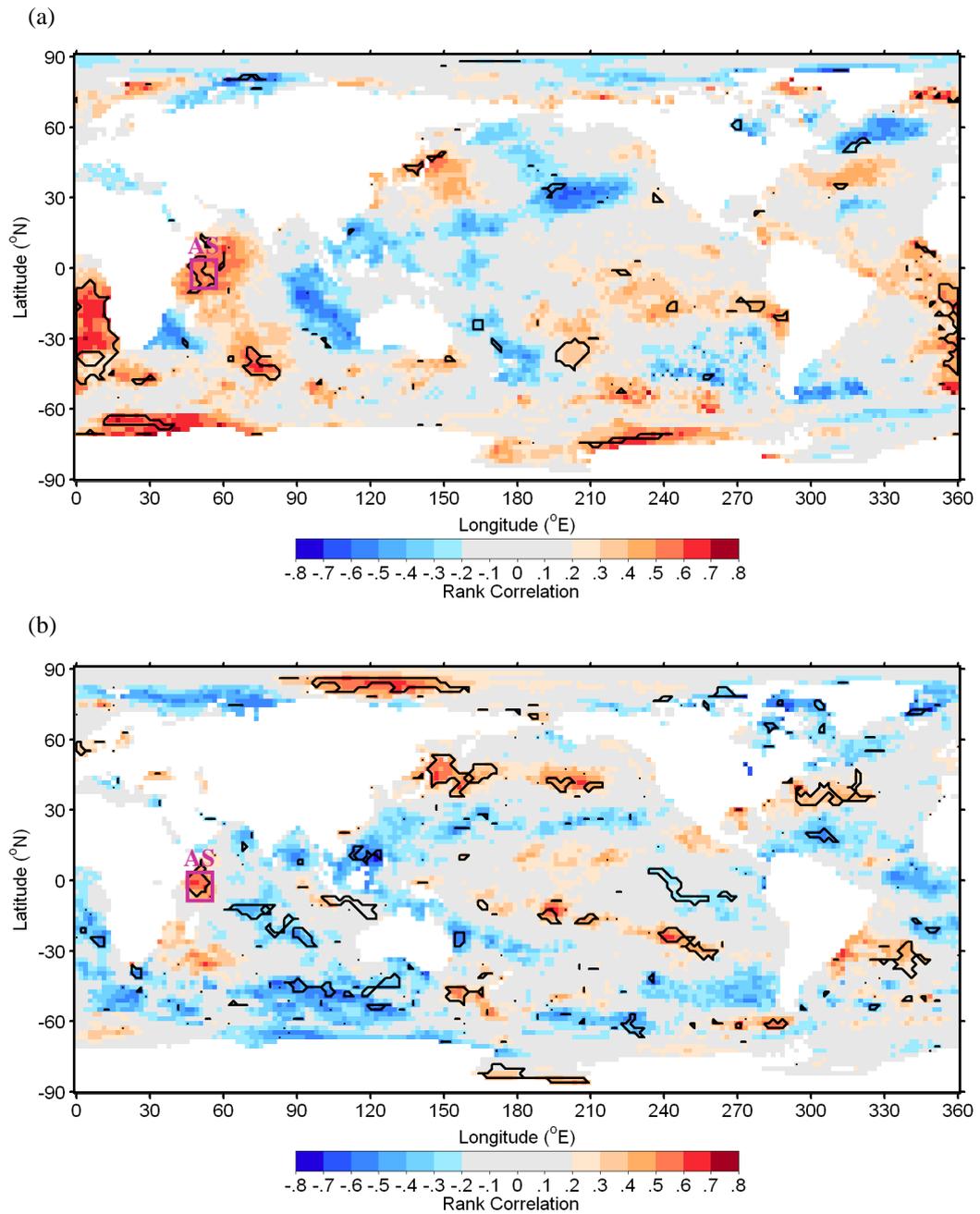


Figure A.2. Southwest OND rainfall index correlated with prior August-September (AS) two-month average SST global field over the periods: (a) 1959-1974 and (b) 1975-1990. As Figure B.1 but the SST AS predictor region is highlighted by the purple box.

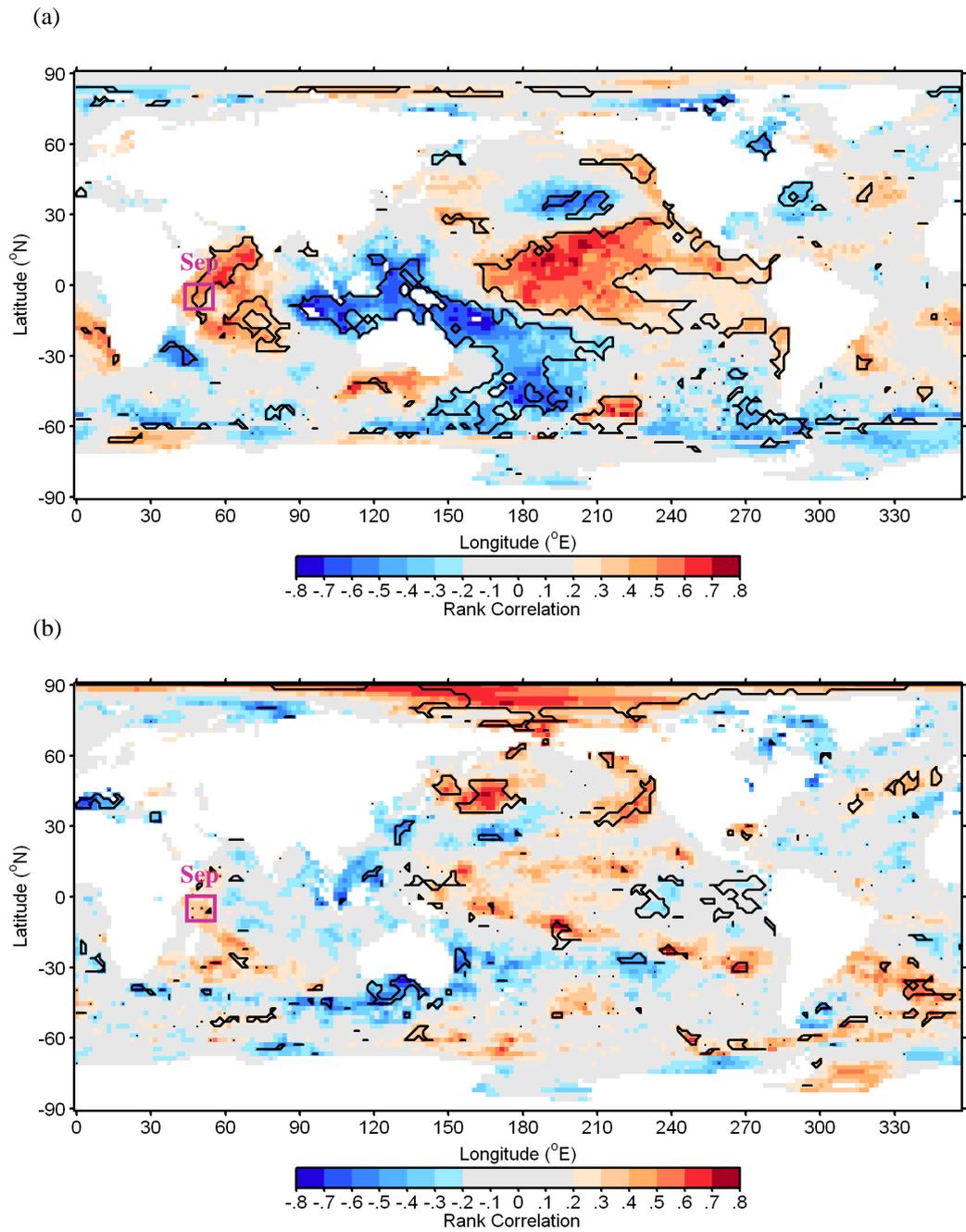


Figure A.3. South OND rainfall index correlated with the prior September SST global field over the periods: (a) 1959-1974 and (b) 1975-1990. As Figure B.1 but the SST Sep predictor region is highlighted by the purple box.

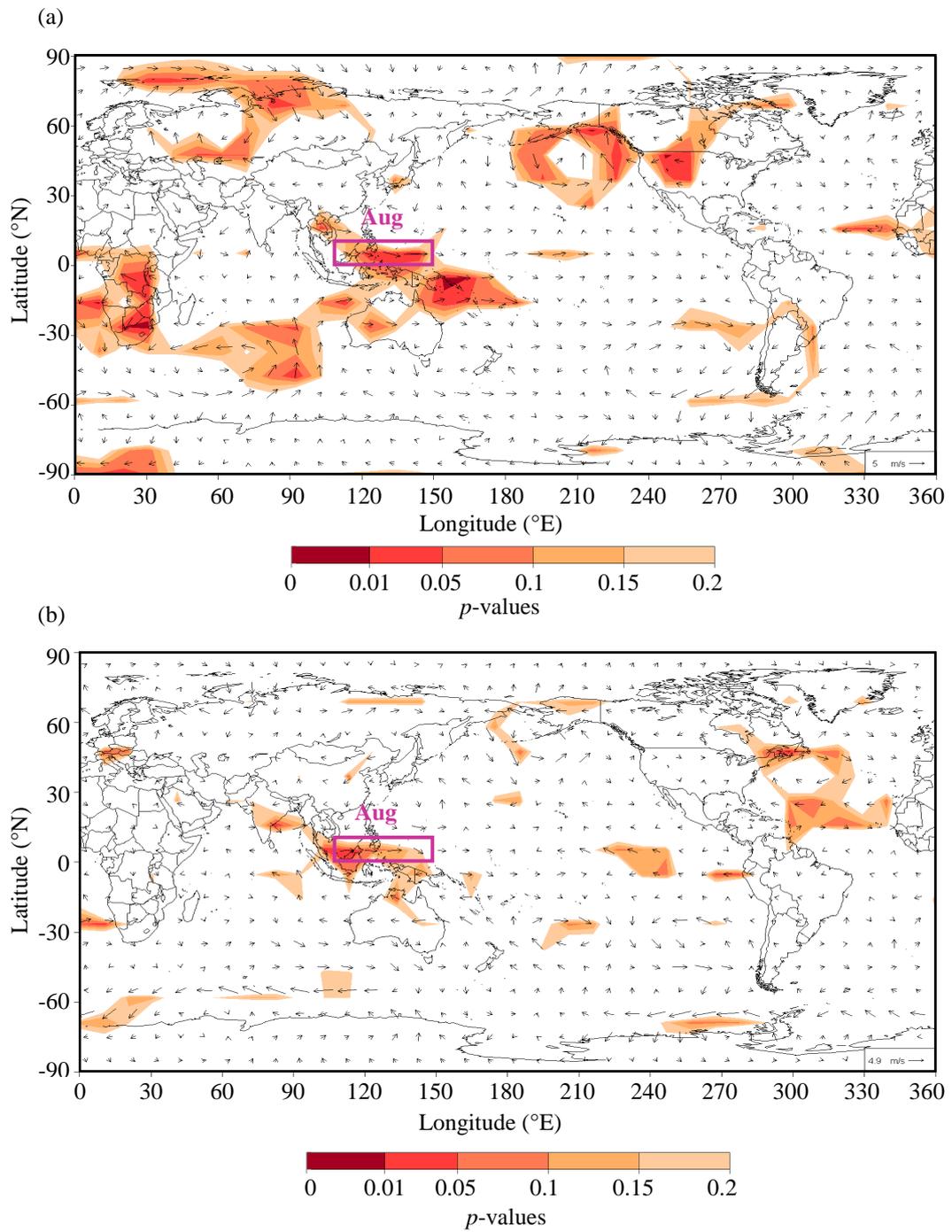


Figure A.4. Composite difference plot showing the direction, magnitude and significance of the composite difference in 850 hPa August wind anomalies for those subset years when the OND South rainfall index is in its upper and lower quartiles over the periods: (a) 1959-1974 and (b) 1975-1990. Colours show p -values with the darker shades of red indicating the greatest significance. The u -wind Aug predictor region is highlighted by the purple box.

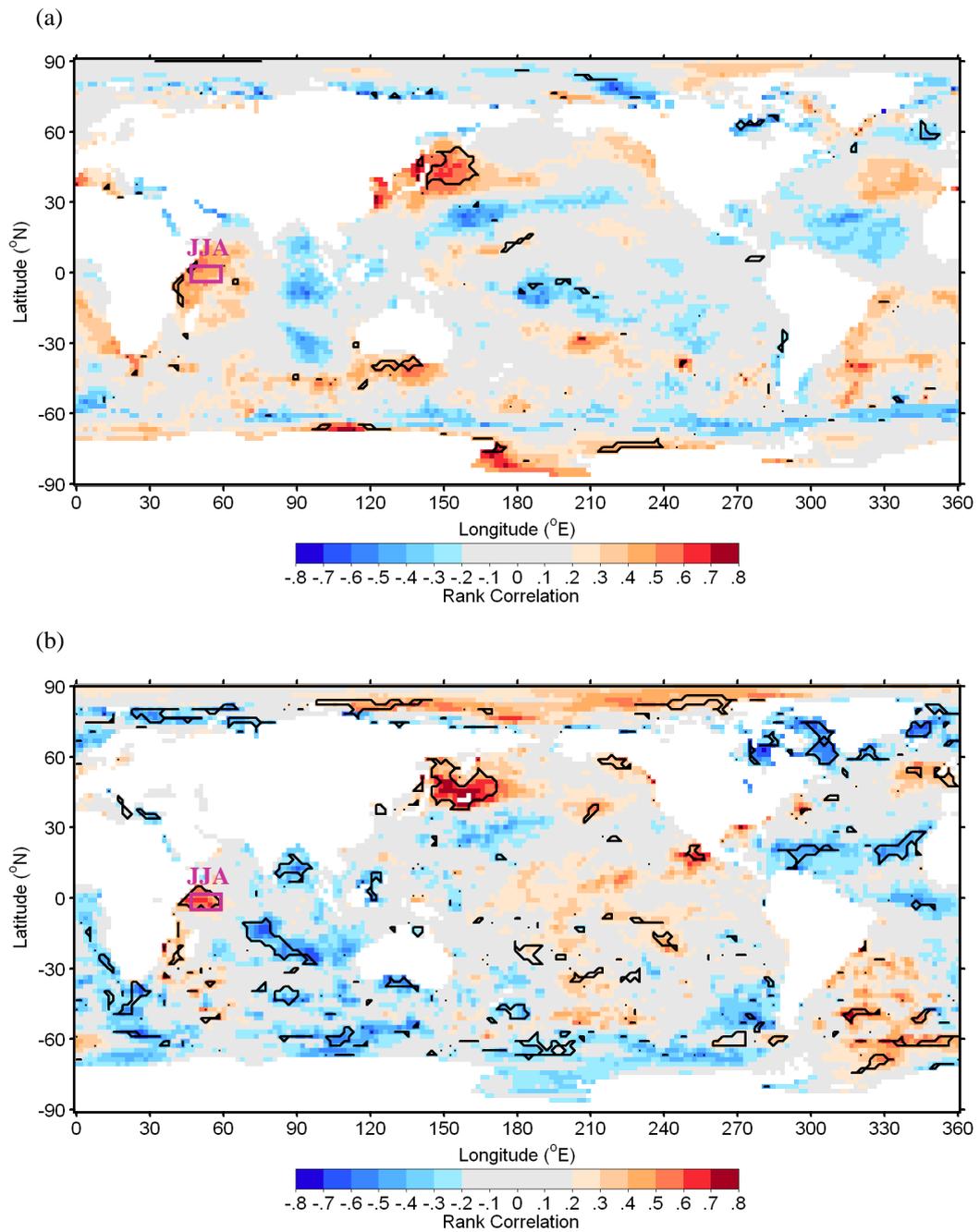


Figure A.5. Northwest OND rainfall index correlated with the prior June-August (JJA) SST global field over the periods: (a) 1959-1974 and (b) 1975-1990. As Figure B.1 but the SST JJA predictor region is highlighted by the purple box.

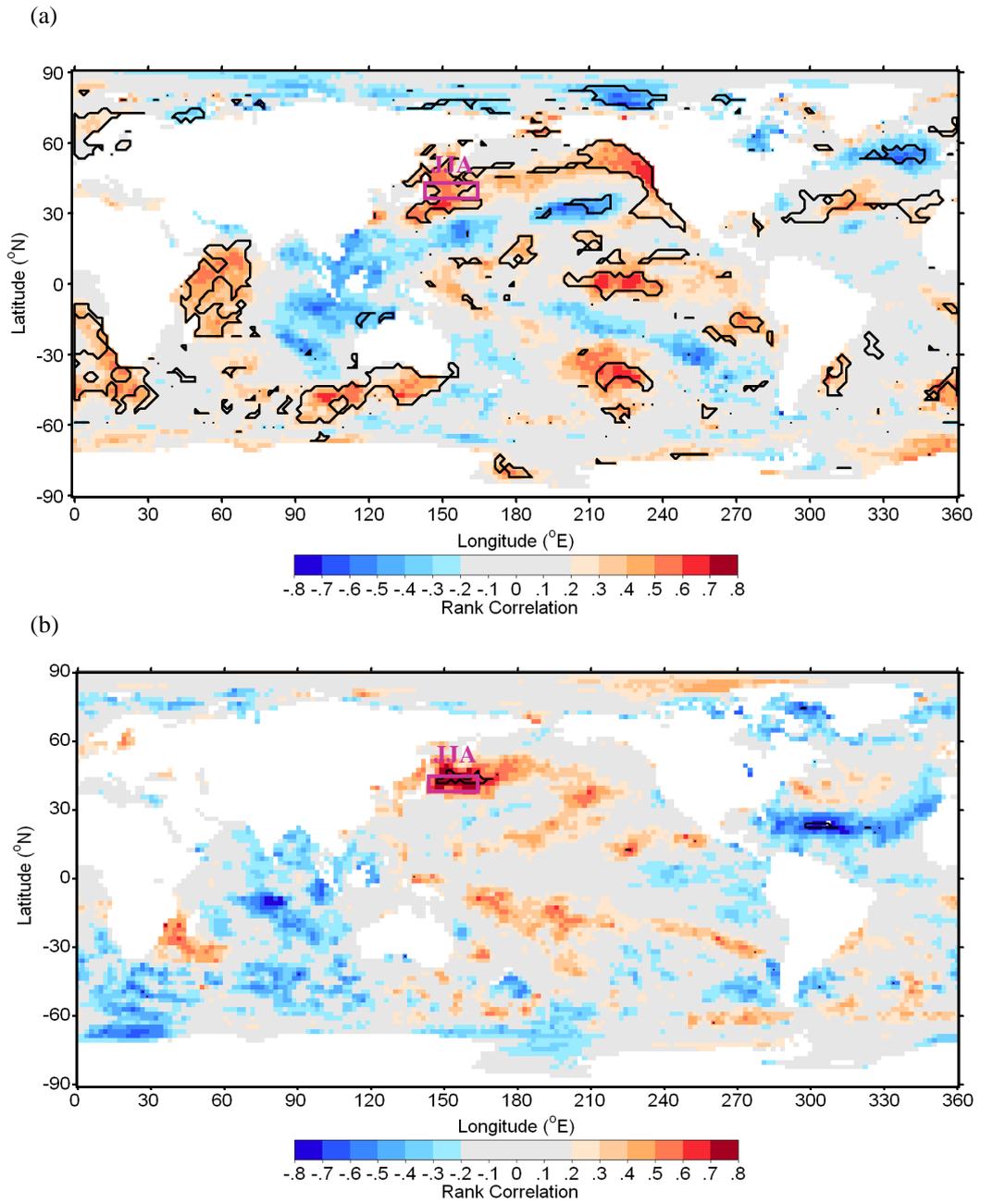


Figure A.6. Southeast OND rainfall index correlated with the prior June-August (JJA) SST global field over the periods: (a) 1959-1974 and (b) 1975-1990. As Figure B.1 but the SST JJA predictor region is highlighted by the purple box.

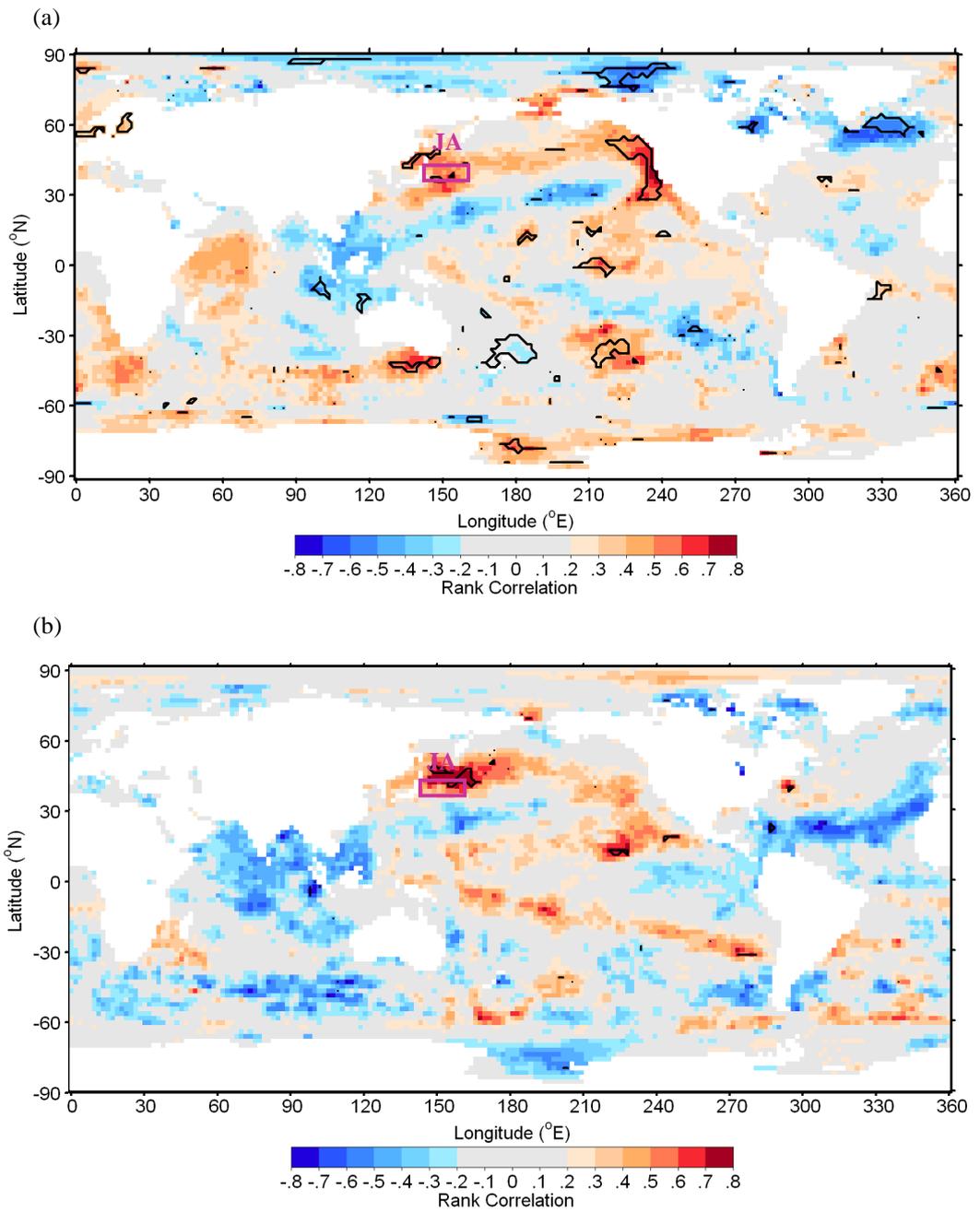


Figure A.7. Southeast-east OND rainfall index correlated with the prior July-August (JA) SST global field over the periods: (a) 1959-1974 and (b) 1975-1990. As Figure B.1 but the SST JA predictor region is highlighted by the purple box.

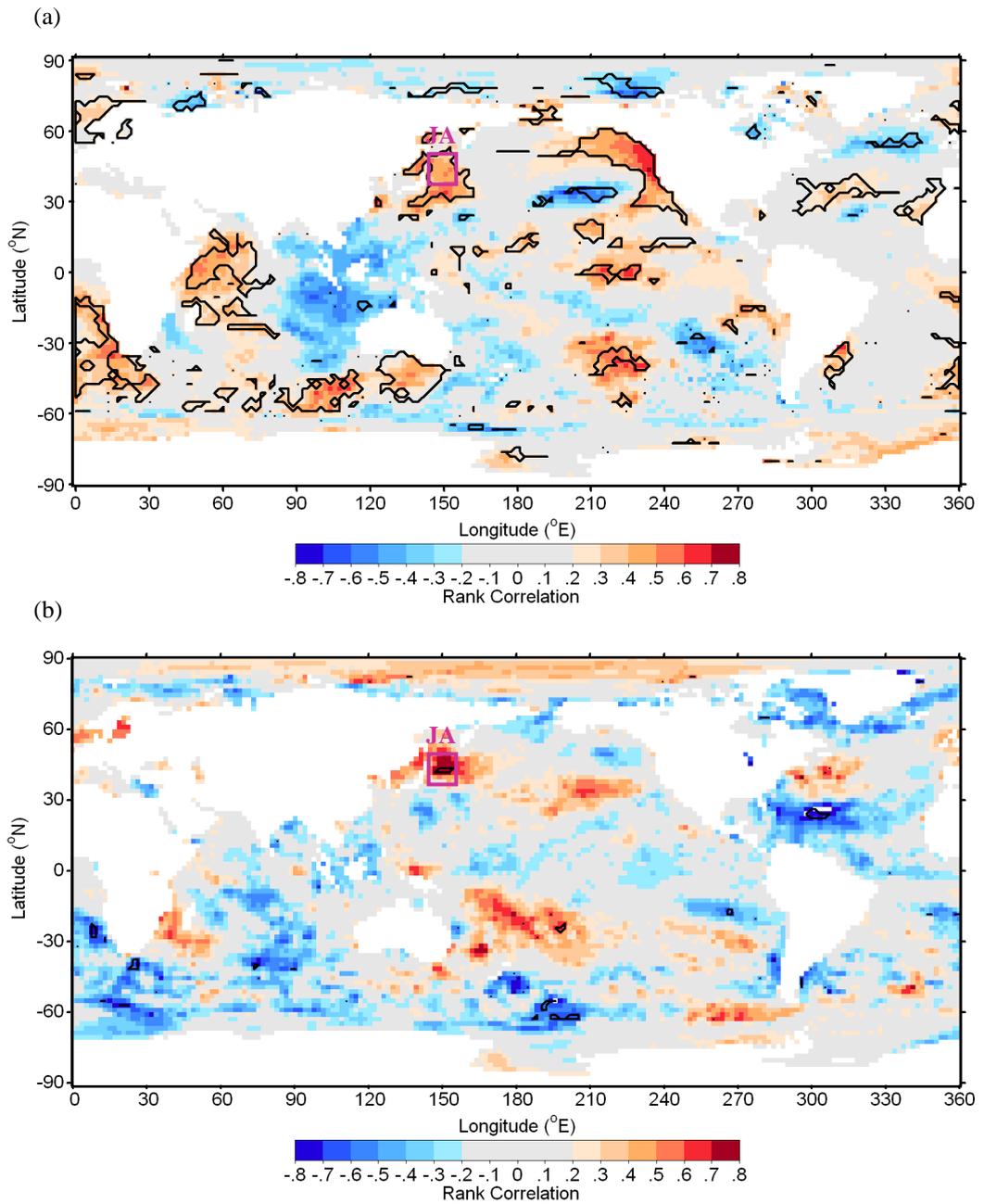


Figure A.8. Southeast-west OND rainfall index correlated with the prior July-August (JA) SST global field over the periods: (a) 1959-1974 and (b) 1975-1990. As Figure B.1 but the SST JA predictor region is highlighted by the purple box.

Appendix B
Skill Score Results for Potential OND Rainfall
Hindcast Models for each Region of Kenya

(a)

Skill scores for West Kenya - OND

		Predictor Models		
		SOI Aug	DMI Sep	(SST AS(1)) - (SST AS(2))
1959-1990	r_{rank}	0.48	0.51	0.57
	p -value	0.01	0.01	0.01
	MSSS	17.2	37.2	22.4
1959-1974	r_{rank}	0.41	0.38	0.58
	p -value	0.10	0.08	0.02
	MSSS	9.7	23.2	20.4
1975-1990	r_{rank}	0.39	0.42	0.38
	p -value	0.09	0.13	0.10
	MSSS	13.5	25.7	9.6
1991-2006	r_{rank}	0.49	0.65	0.65
	p -value	0.03	0.01	0.01
	MSSS	31.5	36.9	53.6

(b)

Extreme Years only: Upper and lower quartiles

		Predictor Models		
		SOI Aug	DMI Sep	SST AS(1)) - (SST AS(2))
1959-1990	r_{rank}	0.62	0.63	0.64
	p -value	0.05	0.07	0.06
	MSSS	25.8	46.6	30.2
1959-1974	r_{rank}	0.50	0.57	0.64
	p -value	0.15	0.10	0.12
	MSSS	13.1	28.5	22.5
1975-1990	r_{rank}	0.36	0.52	0.43
	p -value	0.17	0.11	0.21
	MSSS	11.9	23.3	8.3
1991-2006	r_{rank}	0.79	0.79	0.76
	p -value	0.05	0.02	0.03
	MSSS	43.9	67.30	58.8

(c)

Number of years in correct tercile

		Predictor Models		
		SOI Aug	DMI Sep	SST AS(1)) - (SST AS(2))
Num/32		15	21	16
p -value		0.29	0.01	0.11
Num/1st16		6	8	10
p -value		0.27	0.30	0.27
Num/2nd16		8	9	9
p -value		0.23	0.19	0.20

Figure B.1. Skill score results for the hindcasts models based on the best potential predictor regions for the West region of Kenya. (a) r_{rank} with associated p -values and MSSS results for each of the periods: 1959-1990, 1959-1974, 1975-1990 and 1991-2006. (b) Same as (a) for the extreme years only (upper and lower quartile of rainfall years). (c) Shows the Number of years in the correct tercile for the first 32 years (1959-1990), the 1st 16 years (1959-1974) and the 2nd 16 years (1975-1990) and associated p -values. The best predictor is highlighted (yellow for training periods, green for independent verification period). Red values show when: $r_{rank} \geq 0.4$, p -value ≤ 0.1 , MSSS > 0 , (Num/32) ≥ 16 and (Num/16) ≥ 8 .

(a)

Skill scores for Southwest Kenya - OND

		Predictor Models		
		SST JAS	SST JA	SST AS
1959-1990	<i>r_{rank}</i>	0.34	0.24	0.50
	<i>p</i> -value	0.05	0.12	0.01
	MSSS	12.2	7.7	18.1
1959-1974	<i>r_{rank}</i>	0.51	0.52	0.55
	<i>p</i> -value	0.04	0.04	0.03
	MSSS	35.9	42.7	23.0
1975-1990	<i>r_{rank}</i>	0.41	0.42	0.39
	<i>p</i> -value	0.09	0.06	0.09
	MSSS	-2.4	-7.2	0.1
1991-2006	<i>r_{rank}</i>	0.52	0.40	0.42
	<i>p</i> -value	0.03	0.08	0.07
	MSSS	11.3	-2.2	28.2

(b)

Extreme Years only: Upper and lower quartiles

		Predictor Models		
		SST JAS	SST JA	SST AS
1959-1990	<i>r_{rank}</i>	0.51	0.47	0.31
	<i>p</i> -value	0.03	0.02	0.04
	MSSS	36.2	35.2	30.7
1959-1974	<i>r_{rank}</i>	0.55	0.55	0.55
	<i>p</i> -value	0.13	0.12	0.18
	MSSS	30.1	38.0	16.0
1975-1990	<i>r_{rank}</i>	0.76	0.76	0.76
	<i>p</i> -value	0.04	0.05	0.05
	MSSS	33.6	32.4	32.4
1991-2006	<i>r_{rank}</i>	0.93	0.67	0.74
	<i>p</i> -value	0.05	0.05	0.03
	MSSS	18.0	5.1	35.0

(c)

Number of years in correct tercile

		Predictor Models		
		SST JAS	SST JA	SST AS
Num/32		11	11	15
<i>p</i> -value		0.05	0.21	0.05
Num/1st16		7	6	7
<i>p</i> -value		0.23	0.32	0.22
Num/2nd16		7	5	7
<i>p</i> -value		0.19	0.13	0.20

Figure B.2. As Figure B.1 but for the Southwest region of Kenya.

(a)

Skill scores for South Kenya - OND

		Predictor Models		
		SST JJA	<i>u</i> -wind Aug	(<i>u</i> -wind Aug) + (SST Sep)
1959-1990	<i>r</i> _{rank}	0.38	0.50	0.46
	<i>p</i> -value	0.02	0.01	0.01
	MSSS	-1.1	17.1	22.1
1959-1974	<i>r</i> _{rank}	0.32	0.39	0.37
	<i>p</i> -value	0.07	0.12	0.10
	MSSS	-18.2	25.0	0.4
1975-1990	<i>r</i> _{rank}	0.36	0.51	0.41
	<i>p</i> -value	0.10	0.06	0.07
	MSSS	-1.3	-19.7	7.2
1991-2006	<i>r</i> _{rank}	0.36	0.37	0.53
	<i>p</i> -value	0.10	0.10	0.04
	MSSS	-24.1	4.5	30.7

(b)

Extreme Years only: Upper and lower quartiles

		Predictor Models		
		SST JJA	<i>u</i> -wind Aug	(<i>u</i> -wind Aug) + (SST Sep)
1959-1990	<i>r</i> _{rank}	0.58	0.56	0.56
	<i>p</i> -value	0.02	0.04	0.03
	MSSS	16.1	28.9	38.3
1959-1974	<i>r</i> _{rank}	0.54	0.77	0.54
	<i>p</i> -value	0.27	0.10	0.09
	MSSS	-36.8	56.0	36.4
1975-1990	<i>r</i> _{rank}	0.67	0.64	0.55
	<i>p</i> -value	0.12	0.11	0.08
	MSSS	14.7	2.8	24.0
1991-2006	<i>r</i> _{rank}	0.26	0.12	0.33
	<i>p</i> -value	0.23	0.23	0.10
	MSSS	-14.2	2.4	29.3

(c)

Number of years in correct tercile

	Predictor Models		
	SST JJA	<i>u</i> -wind Aug	(<i>u</i> -wind Aug) + (SST Sep)
Num/32	13	12	11
<i>p</i> -value	0.07	0.02	0.02
Num/1st16	6	6	6
<i>p</i> -value	0.13	0.12	0.12
Num/2nd16	9	9	7
<i>p</i> -value	0.28	0.20	0.19

Figure B.3. As Figure B.1 but for the South region of Kenya.

(a)

Skill scores for Northeast Kenya - OND

		Predictor Models	
		South region's (<i>u</i> -wind Aug) + (SST Sep)	South region's <i>u</i> -wind Aug
1959-1990	<i>r</i> _{rank}	0.36	0.40
	<i>p</i> -value	0.03	0.01
	MSSS	21.6	19.6
1959-1974	<i>r</i> _{rank}	0.17	0.37
	<i>p</i> -value	0.27	0.07
	MSSS	20.3	21.3
1975-1990	<i>r</i> _{rank}	0.22	0.36
	<i>p</i> -value	0.20	0.13
	MSSS	-0.3	0.8
1991-2006	<i>r</i> _{rank}	0.31	0.46
	<i>p</i> -value	0.08	0.02
	MSSS	22.10	5.90

(b)

Extreme Years only: Upper and lower quartiles

		Predictor Models	
		South region's (<i>u</i> -wind Aug) + (SST Sep)	South region's <i>u</i> -wind Aug
1959-1990	<i>r</i> _{rank}	0.52	0.61
	<i>p</i> -value	0.12	0.08
	MSSS	33.4	33.4
1959-1974	<i>r</i> _{rank}	0.45	0.69
	<i>p</i> -value	0.06	0.07
	MSSS	26.2	26.8
1975-1990	<i>r</i> _{rank}	0.33	0.43
	<i>p</i> -value	0.14	0.15
	MSSS	23.2	20.2
1991-2006	<i>r</i> _{rank}	0.69	0.48
	<i>p</i> -value	0.11	0.29
	MSSS	13.9	-5.1

(c)

Number of years in correct tercile

	Predictor Models	
	South region's (<i>u</i> -wind Aug) + (SST Sep)	South region's <i>u</i> -wind Aug
Num/32	12	13
<i>p</i> -value	0.02	0.05
Num/1st16	7	8
<i>p</i> -value	0.55	0.55
Num/2nd16	6	6
<i>p</i> -value	0.35	0.35

Figure B.4. As Figure B.1 but for the Northeast region of Kenya.

(a) Skill scores for Northwest Kenya - OND

		Predictor Models	
		SST Jun	SST JJA
1959-1990	r_{rank}	0.26	0.33
	p -value	0.08	0.05
	MSSS	4.6	3.9
1959-1974	r_{rank}	0.27	0.35
	p -value	0.16	0.09
	MSSS	-3.6	-8.1
1975-1990	r_{rank}	0.51	0.54
	p -value	0.03	0.02
	MSSS	15.3	11.3
1991-2006	r_{rank}	0.18	0.29
	p -value	0.28	0.23
	MSSS	-11.30	7.80

(b) Extreme Years only: Upper and lower quartiles

		Predictor Models	
		SST Jun	SST JJA
1959-1990	r_{rank}	0.43	0.50
	p -value	0.11	0.06
	MSSS	17.5	20.7
1959-1974	r_{rank}	0.35	0.29
	p -value	0.08	0.12
	MSSS	11.0	0.4
1975-1990	r_{rank}	0.90	0.88
	p -value	0.01	0.01
	MSSS	57.1	54.7
1991-2006	r_{rank}	0.10	0.24
	p -value	0.30	0.07
	MSSS	-5.50	32.00

(c) Number of years in correct tercile

	Predictor Models	
	SST Jun	SST JJA
Num/32	13	17
p -value	0.34	0.12
Num/1st16	5	7
p -value	0.08	0.16
Num/2nd16	8	7
p -value	0.14	0.27

Figure B.5. As Figure B.1 but for the Northwest region of Kenya.

(a)

Skill scores for Southeast Kenya - OND

		Predictor Models	
		SST JA	SST JJA
1959-1990	<i>r</i> _{rank}	0.21	0.15
	<i>p</i> -value	0.16	0.23
	MSSS	1.4	0.7
1959-1974	<i>r</i> _{rank}	0.34	0.37
	<i>p</i> -value	0.11	0.09
	MSSS	11.5	12.2
1975-1990	<i>r</i> _{rank}	0.42	0.49
	<i>p</i> -value	0.07	0.04
	MSSS	6.7	14.4
1991-2006	<i>r</i> _{rank}	-0.09	-0.02
	<i>p</i> -value	0.37	0.48
	MSSS	-24.6	-32.6

(b)

Extreme Years only: Upper and lower quartiles

		Predictor Models	
		SST JA	SST JJA
1959-1990	<i>r</i> _{rank}	0.15	0.05
	<i>p</i> -value	0.30	0.14
	MSSS	-2.6	-2.1
1959-1974	<i>r</i> _{rank}	0.48	0.48
	<i>p</i> -value	0.09	0.10
	MSSS	10.7	11.8
1975-1990	<i>r</i> _{rank}	0.14	0.31
	<i>p</i> -value	0.22	0.13
	MSSS	6.4	20.2
1991-2006	<i>r</i> _{rank}	-0.21	-0.14
	<i>p</i> -value	0.39	0.30
	MSSS	7.4	-3.5

(c)

Number of years in correct tercile

	Predictor Models	
	SST JA	SST JJA
Num/32	11	12
<i>p</i> -value	0.82	0.82
Num/1st16	6	6
<i>p</i> -value	0.35	0.35
Num/2nd16	11	6
<i>p</i> -value	0.02	0.13

Figure B.6. As Figure B.1 but for the Southeast region of Kenya.

(a)

Skill scores for Southeast-east Kenya - OND

		Predictor Models	
		SST JJA	SST JA
1959-1990	<i>r</i> _{rank}	0.27	0.31
	<i>p</i> -value	0.07	0.05
	MSSS	9.0	10.2
1959-1974	<i>r</i> _{rank}	0.40	0.36
	<i>p</i> -value	0.09	0.12
	MSSS	8.7	6.5
1975-1990	<i>r</i> _{rank}	0.51	0.41
	<i>p</i> -value	0.03	0.07
	MSSS	30.7	15.3
1991-2006	<i>r</i> _{rank}	0.03	0.09
	<i>p</i> -value	0.46	0.36
	MSSS	-22.5	-11.3

(b)

Extreme Years only: Upper and lower quartiles

		Predictor Models	
		SST JJA	SST JA
1959-1990	<i>r</i> _{rank}	0.41	0.38
	<i>p</i> -value	0.06	0.10
	MSSS	16.0	13.3
1959-1974	<i>r</i> _{rank}	0.38	0.33
	<i>p</i> -value	0.10	0.16
	MSSS	9.6	4.6
1975-1990	<i>r</i> _{rank}	0.43	0.29
	<i>p</i> -value	0.05	0.12
	MSSS	38.0	16.1
1991-2006	<i>r</i> _{rank}	-0.12	0.01
	<i>p</i> -value	0.41	0.39
	MSSS	-6.0	9.7

(c)

Number of years in correct tercile

	Predictor Models	
	SST JJA	SST JA
Num/32	14	11
<i>p</i> -value	0.54	0.32
Num/1st16	8	7
<i>p</i> -value	0.32	0.17
Num/2nd16	6	5
<i>p</i> -value	0.36	0.06

Figure B.7. As Figure B.1 but for the Southeast-east sub-region of Kenya.

(a)

Skill scores for Southeast-west Kenya - OND

		Predictor Models	
		SST JJA	SST JA
1959-1990	<i>r</i> _{rank}	0.11	0.36
	<i>p</i> -value	0.30	0.04
	MSSS	-3.8	3.0
1959-1974	<i>r</i> _{rank}	0.26	0.33
	<i>p</i> -value	0.17	0.09
	MSSS	15.2	14.8
1975-1990	<i>r</i> _{rank}	0.03	0.42
	<i>p</i> -value	0.47	0.09
	MSSS	-12.3	11.2
1991-2006	<i>r</i> _{rank}	-0.01	0.07
	<i>p</i> -value	0.49	0.40
	MSSS	-5.4	-14.1

(b)

Extreme Years only: Upper and lower quartiles

		Predictor Models	
		SST JJA	SST JA
1959-1990	<i>r</i> _{rank}	0.06	0.43
	<i>p</i> -value	0.42	0.05
	MSSS	-7.2	6.2
1959-1974	<i>r</i> _{rank}	0.12	0.38
	<i>p</i> -value	0.10	0.10
	MSSS	8.6	9.4
1975-1990	<i>r</i> _{rank}	0.01	0.17
	<i>p</i> -value	0.47	0.13
	MSSS	-3.0	19.7
1991-2006	<i>r</i> _{rank}	0.01	0.17
	<i>p</i> -value	0.50	0.36
	MSSS	-5.4	16.2

(c)

Number of years in correct tercile

		Predictor Models	
		SST JJA	SST JA
Num/32		11	16
<i>p</i> -value		0.43	0.40
Num/1st16		6	5
<i>p</i> -value		0.27	0.02
Num/2nd16		3	6
<i>p</i> -value		0.36	0.71

Figure B.8. As Figure B.1 but for the Southeast-west sub-region of Kenya.

Appendix C
Predictor Selection Plots: Correlation Plots Between
Regional Seasonal Rainfall Indices and Global SSTs

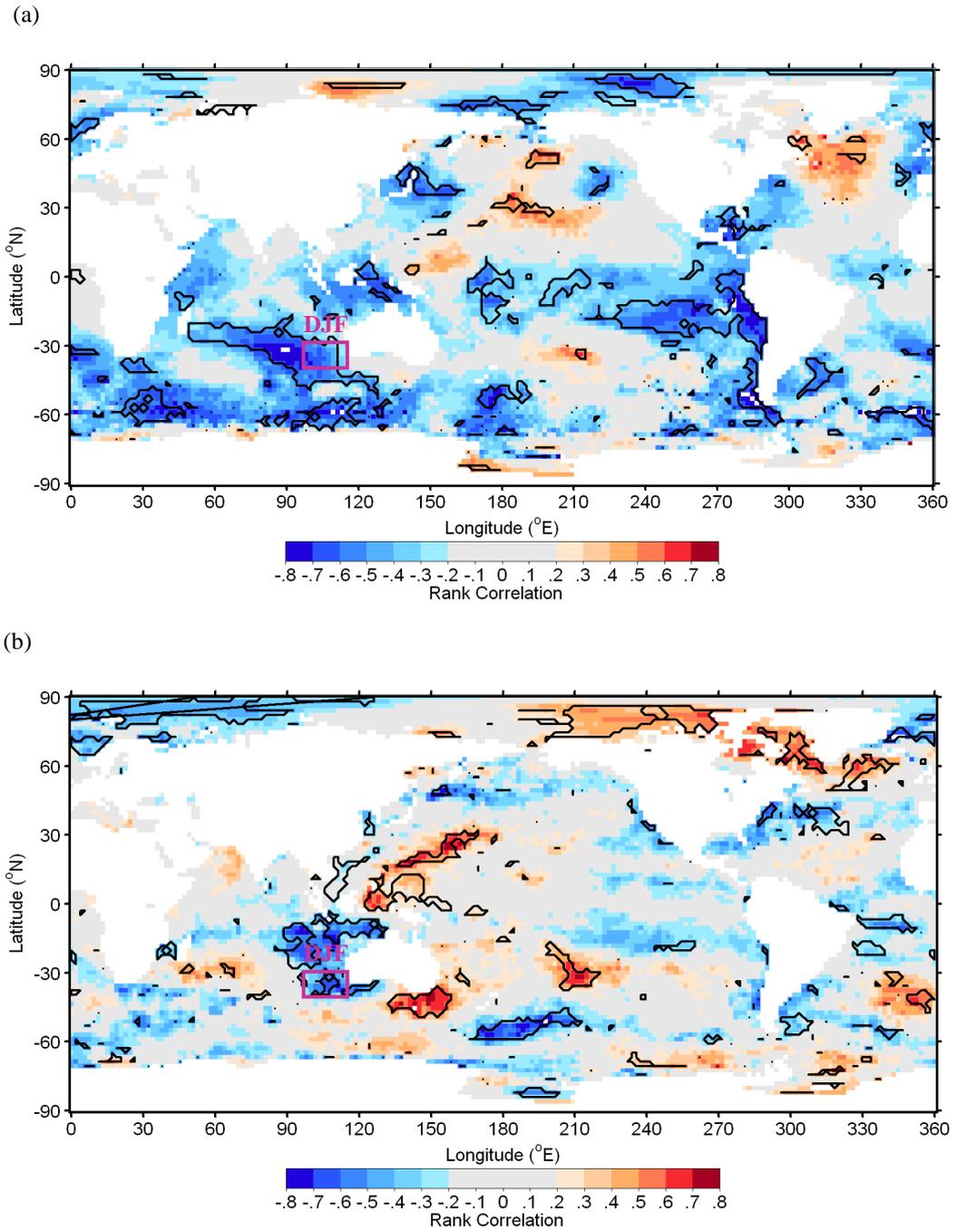
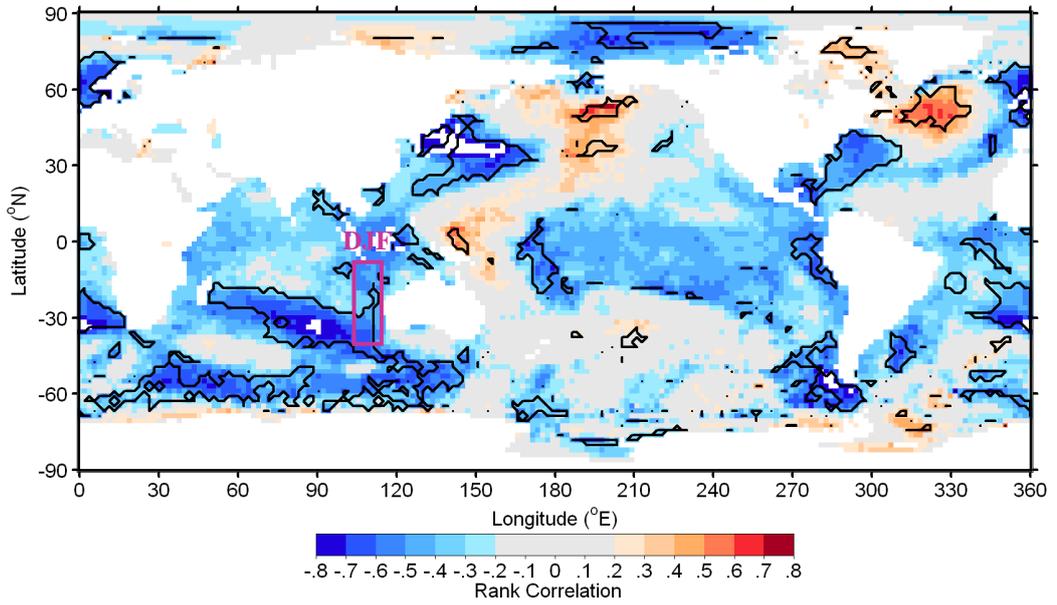


Figure C.1. Southeast April rainfall index correlated with December-February (DJF) three-month average SST global field over the periods: (a) 1959-1974 and (b) 1975-1990. White areas denote land mass. Red (blue) shows positive (negative) correlations, grey shows low correlations (between -0.2 and 0.2) and the black lines show areas with p -values < 0.15 . The SST DJF predictor region is highlighted by the purple box.

(a)



(b)

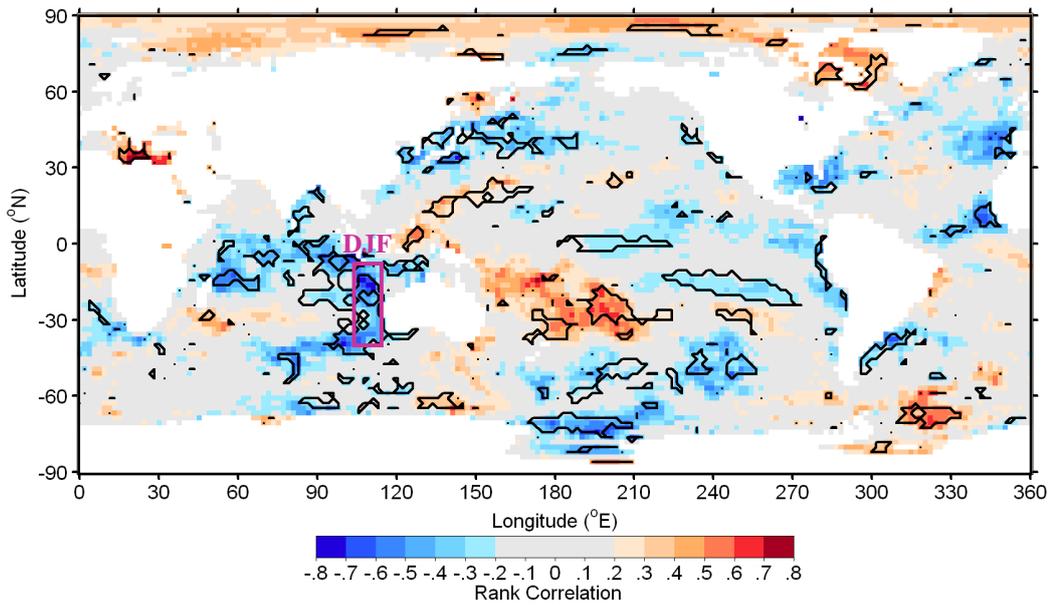


Figure C.2. West April rainfall index correlated with December-February (DJF) three-month average SST global field over the periods: (a) 1959-1974 and (b) 1975-1990. White areas denote land mass. Red (blue) shows positive (negative) correlations, grey shows low correlations (between -0.2 and 0.2) and the black lines show areas with p -values < 0.15 . The SST DJF predictor region is highlighted by the purple box.

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Mark Saunders. Thank you for your constant support, encouragement and wisdom over the past 3 years. I greatly appreciate all of the time and effort that you have dedicated to my PhD. I would also like to thank Dr. Benjamin Lloyd-Hughes for his support, especially during the first year of my PhD. I could not have learnt how to programme in “R” without your help. Thanks also to Dr. Celine Herweijer for her support during my PhD. Thank you also to Dr. Adam Lea for always making time for me and answering my many questions.

This thesis is dedicated to my wonderful family and especially to my husband, Gerard, who has basically made this PhD possible. Gerard, thank you for your love, encouragement, unwavering confidence in my ability and support every step of the way. This PhD would definitely not have been possible without you. I love you with all of my heart! My family have made me who I am today and I can't thank you enough. Mum and Dad, you are the best parents in the world and I love you so much. Thanks also to my extended family, my in-laws and my close friends for your constant encouragement and support. Thanks also to everyone at MSSL for being so kind and welcoming. Thanks to my fellow PhD students and my other friends at MSSL for all the support and fun times along the way.

Thanks to the ECMWF, Météo France and the UK Met Office for providing access to their hindcast data, especially to Paco Doblus-Rayes for his helpful discussions. Thanks also to the GPCC and CPC for allowing access to their precipitation data sets. Thanks to NCEP/NCAR for allowing access to their potential predictor indices and to JAMSTEC for providing access to their DMI data set and for extending the period available to help with my research. Thanks also to Dr. David Grimes and Dr. Gulilat Diro for their advice on seasonal rainfall prediction in East Africa. I owe a huge debt of thanks to the Kenya Meteorological Department for providing the monthly rain gauge data. I have been very fortunate to make contact with many Kenyan and East African scientists and forecasters through the GHACOF in Uganda (February 2008) and the Conference on African Drought in Trieste (June 2008) – thanks to all of you for your time, advice and local knowledge. Finally, a very big thank you to the UK Natural Environment Research Council for funding my research.

References

- AMS (American Meteorological Society) (2009). Glossary of Meteorology. <http://amsglossary.allenpress.com/glossary>, last accessed 05/01/10.
- Anderson, D., Stockdale, T., Balmaseda, M., Ferranti, L., Vitart, F., Molteni, F., Doblas-Reyes, F. J., Mogenson, K., and Vidard, A. (2007). Development of the ECMWF seasonal forecast system 3. *ECMWF Technical Memorandum*, No. **503**.
- Angell, J. K. and Korshover, J. (1964). Quasi-biennial variations in temperature, total ozone, and tropopause height. *Journal of the Atmospheric Sciences*, **21**, 479-492.
- Anyah, R. O. (2008). *Personal communication*, The University of Connecticut, USA
- Anyah, R. O. (2009). *Personal communication*, The University of Connecticut, USA
- Baquero-Bernal, A., Latif, M., and Legutke, S. (2002). Notes and correspondence on dipolelike variability of sea surface temperature in the tropical Indian Ocean. *Journal of Climate*, **15**, 1358-1368.
- Barnston, A. G., Mason, S. J., Goddard, L., DeWitt, D. G., and Zebiak, S. E. (2003). Multimodel ensembling in seasonal climate forecasting at IRI. *Bulletin of the American Meteorological Society*, **84**, 1783-1796.
- Beck, C., Grieser, J., and Rudolf, B. (2005). *A new monthly precipitation climatology for the global land areas for the period 1951 to 2000*. Climate status report, German Weather Service, Germany.
- Behera, S. K. and Yamagata, T. (2003). Influence of the Indian Ocean Dipole on the Southern Oscillation. *Journal of the Meteorological Society of Japan*, **81**, 169-177.
- Behera, S. K., Luo, J. J., Masson, S., Rao, S. A., and Sakuma, H. (2006). A CGCM study on the interaction between IOD and ENSO. *Journal of Climate*, **19**, 1688-1705.
- Beltrando, G. (1990). Space-time variability of rainfall in April and October-November over East Africa during the period 1932-1983. *International Journal of Climatology*, **10**, 691-702.

- Bjerknes, J. (1966). A possible response of the atmospheric Hadley Circulation to Equatorial anomalies of ocean temperature. *Tellus*, **18**, 820-829.
- Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial Pacific. *Monthly Weather Review*, **97**, 163-172.
- Black, E. (2003). The impact of Indian and Pacific Ocean processes on the East African short rains. *CLIVAR Exchanges*, **8**, 40-42.
- Black, E. (2005). The relationship between Indian Ocean sea-surface temperature and East African rainfall. *Philosophical Transactions of The Royal Society A*, **363**, 43-47.
- Blades, H. (2000). *Kenya*. Heinemann, Oxford, UK.
- Boken, V. K., Cracknell, A. P., and Heathcote, R. L. (2005). *Monitoring and predicting agricultural drought: A global study*. Oxford University Press, UK.
- Bowden, R. (2007). *Kenya, 2nd edition*. Evan Brothers, London, UK.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformation. *Journal of the Royal Statistical Society: Series B*, **26**, 211-246.
- Brass, W., Jolly, C. L. e., and Working Group on Kenya, N. R. C. (1993). *Population dynamics of Kenya*. National Academy Press, Washington D.C., USA.
- Buizer, J. L., Foster, J., and Lund, D. (2000). Global impacts and regional actions: preparing for the 1997-98 El Nino. *Bulletin of the American Meteorological Society*, **81**, 2121-2139.
- Camberlin, P. and Wairoto, J. G. (1997). Intraseasonal wind anomalies related to wet and dry spells during the "long" and "short" rainy seasons in Kenya. *Theoretical and Applied Climatology*, **58**, 57-69.
- 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. *International Journal of Climatology*, **21**, 973-1005.
- Camberlin, P. and Philippon, N. (2002). The East African March–May rainy season: associated atmospheric dynamics and predictability over the 1968–97 period. *Journal of Climate*, **15**, 1002-1019.

- Camberlin, P. and Okoola, R. E. (2003). The onset and cessation of the "long rains" in eastern Africa and their interannual variability. *Theoretical and Applied Climatology*, **75**, 43-54.
- Camberlin, P. (2008). *Personal communication*, Centre de Recherche de Climatologie, University of Dijon, France.
- Cariolle, D., Amodei, M., Deque, M., Mahfouf, J. F., Simon, P., and Teysseire, H. (1993). A quasi-biennial oscillation signal in general circulation model simulations. *Science*, **261**, 1313-1316.
- CERF (Central Emergency Response Fund) (2008). *African review report on drought and desertification*. United Nations Economic Commission for Africa, Addis Ababa, Ethiopia.
- CERF (Central Emergency Response Fund) (2009). *CERF funds jump start emergency aid operations in Kenya*. United Nations Economic Commission for Africa, Addis Ababa, Ethiopia.
- Chandler, R. (2009). *Personal communication*, University College London, UK.
- Chang, P. and Zebiak, S. E. (2003). El Nino and the Southern Oscillation: theory. *Encyclopedia of Atmospheric Sciences: Volume 2*, edited by J. R. Holton, J. A. Curry, and J. A. Pyle. Elsevier Science Ltd., London, UK.
- Chen, W. Y. (1982). Fluctuations in northern hemisphere 700mb height field associated with the Southern Oscillation. *Monthly Weather Review*, **110**, 808-823.
- CIA (Central Intelligence Agency) (2009). The World Fact Book. <https://www.cia.gov/library/publications/the-world-factbook/geos/ke.html>, last accessed 05/01/10.
- CIESIN (Center for International Earth Science Information Network) (2005). Gridded population of the world, version 3 (GPWv3) data collection. <http://sedac.ciesin.columbia.edu/gpw/global.jsp>, last accessed 05/01/10.
- CIESIN (Center for International Earth Science Information Network) (2009). Africa - Kenya Summary. <http://sedac.ciesin.columbia.edu/gpw/country.jsp?iso=KEN>, last accessed 05/01/10.

- Clark, C. O., Webster, P. J., and Cole, J. E. (2003). Interdecadal variability of the relationship between the Indian Ocean Zonal Mode and East African coastal rainfall anomalies. *Journal of Climate*, **16**, 548-554.
- Colman, A. and Graham, R. (2005). Forecast of East African rainfall for October-December 2005 using dynamical and statistical methods. <http://www.iges.org/ellfb/Sep05/Colman/colman.htm>, last accessed 05/01/10.
- CPC (Climate Prediction Center) (2007). El Nino relationship between global SSTs and precipitation Dec-Feb. <http://www.cpc.noaa.gov>, last accessed 05/01/10.
- Culverwell, I. (2010). *Personal communication*, Met Office, Exeter, UK.
- Davey, M. and Ferranti, L. (2006). Towards a multi-model real-time ensemble system for seasonal prediction: the EURO-SIP consortium. [http://www.wmo.int/pages/prog/www/DPFS/Meetings/ET-LRF_ECMWF2006/Doc3-2\(1\).doc](http://www.wmo.int/pages/prog/www/DPFS/Meetings/ET-LRF_ECMWF2006/Doc3-2(1).doc), last accessed 05/01/10.
- Davies, R. E. (1976). Predictability of sea-surface temperature and sea level pressure anomalies over the north Pacific Ocean. *Journal of Physical Oceanography*, **6**, 249-266.
- Davies, T. D., Vincent, C. E., and Beresford, A. K. C. (1985). July-August rainfall in West-Central Kenya. *Journal of Climatology*, **5**, 17-33.
- Diaz, H. F. and Markgraf, V. (2000). *El Nino and the Southern Oscillation: multiscale variability and global and regional impacts*, Cambridge University Press, UK.
- Dilley, M., Chen, R. S., Deichmann, U., Lerner-Lam, A. L., Arnold, M., Agwe, J., Buys, P., Kjekstad, O., Lyon, B., and Yetman, G. (2005). *Natural disaster hotspots: A global risk analysis*. World Bank, Washington D.C., USA.
- Diro, G., Black, E., and Grimes, D. (2008). Seasonal forecasting of Ethiopian spring rains. *Meteorological Applications*, **15**, 73-83.
- Doblas-Reyes, F. J., Hagedorn, R., and Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting - II. Calibration and combination. *Tellus*, **57A**, 234-252.

- Doblas-Reyes, F. J., Hagedorn, R., and Palmer, T. N. (2006). Developments in dynamical seasonal forecasting relevant to agricultural management. *Climate Research*, **33**, 19-26.
- Doblas-Reyes, F. J. (2007). *Personal communication*, European Centre for Medium-Range Weather Forecasts, Reading, UK.
- Doblas-Reyes, F. J. (2010). *Personal communication*, Catalan Institute of Climate Sciences, Barcelona, Spain.
- Drbohlav, L. H. K., Gualdi, S., and Navarra, A. (2007). A diagnostic study of the Indian Ocean Dipole Mode in El Niño and non-El Niño years. *Journal of Climate*, **20**, 2961-2977.
- Ebdon, R. A. and Veryard, R. G. (1961). Fluctuations in equatorial stratospheric winds. *Nature*, **189**, 791-793.
- ECMWF (2009). EUROSIP users guide. <http://www.ecmwf.int/products/forecasts/seasonal/documentation/eurosip/index.html>, last accessed 05/01/10.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the bootstrap. Monographs on statistics and applied probability*. Chapman and Hall, London, UK.
- EM-DAT (2004). Thirty years of natural disasters 1974-2003: The numbers. http://www.em-dat.net/documents/Publication/publication_2004_emdat.pdf, last accessed 05/01/10.
- EM-DAT (2006). CRED Crunch Newsletter, December issue. <http://www.cred.be/sites/default/files/CredCrunch07.pdf>, last accessed 05/01/10.
- EM-DAT (2009). EM-DAT: The OFDA/CRED International Disaster Database Université catholique de Louvain, Brussels, Belgium. www.emdat.be, last accessed 05/01/10.
- Encyclopedia Britannica (1998). *Encyclopedia Britannica*, Encyclopedia Britannica Inc., Chicago, Illinois, USA.
- Encyclopedia Britannica (1999). *Encyclopedia Britannica*, Encyclopedia Britannica Inc., Chicago, Illinois, USA.

- Farmer, G. (1988). Seasonal forecasting of the Kenya coast short rains. *Journal of Climatology*, **8**, 489-497.
- Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd, Edinburgh.
- Gatebe, C. K., Tyson, P. D., Annegarn, H., Piketh, S., and Helas, G. (1999). A seasonal air transport climatology for Kenya. *Journal of Geophysical Research*, **104**, 14237-14244.
- Gibbs, W. J. (2000). *Foreword, Drought: a global assessment*, edited by D. A. Wilhite, Routledge, New York, New York, USA.
- Gissila, T., Black, E., Grimes, D., and Slingo, J. M. (2004). Seasonal forecasting of the Ethiopian summer rains. *International Journal of Climatology*, **24**, 1345-1358.
- Gitau, W. (2008). *Personal communication*, Centre de Recherche de Climatologie, University of Dijon, France.
- Glantz, M. H. (2001). *Currents of change: Impacts of El Nino and La Nina on climate and society, second edition*. Cambridge University Press, Cambridge, UK.
- Glantz, M. H. ed. (1988). *Drought and hunger in Africa*. Cambridge University Press, Cambridge, UK.
- Goddard, L., Mason, S. J., Zebiak, S. E., Ropelewski, C. F., Basher, R., and Cane, M. A. (2001). Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology*, **21**, 1111-1152.
- Gommes, R. and Petrassi, F. (1996). *Rainfall Variability and drought in sub-saharan Africa since 1960*. Food and Agriculture Organisation of the United Nations, Rome, Italy.
- Gonzalez-Rouco, J. F., Jimenez, J. L., Quesada, V., and Valero, F. (2001). Quality control and homogeneity of precipitation data in the southwest of Europe. *Journal of Climate*, **14**, 964-978.
- Guan, Z., Ashok, K., and Yamagata, T. (2003). Summertime response of the tropical atmosphere to the Indian Ocean Dipole sea surface temperature anomalies. *Journal of the Meteorological Society of Japan*, **81**, 533-561.

- Hagedorn, R., Doblas-Reyes, F. J., and Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. *Tellus*, **57A**, 219-233.
- Hastenrath, S., Nicklis, A., and Greischar, L. (1993). Atmospheric-hydrospheric mechanisms of climate anomalies in the western equatorial Indian Ocean. *Journal of Geophysical Research*, **98**, 20219-20235.
- Hastenrath, S. (1995). Recent advances in tropical climate prediction. *Journal of Climate*, **8**, 1519-1532.
- Hastenrath, S. (2000). Zonal circulations over the equatorial Indian Ocean. *Journal of Climate*, **13**, 2746-2756.
- Hastenrath, S., Polzin, D., and Camberlin, P. (2004). Exploring the predictability of the short rains at the coast of East Africa. *International Journal of Climatology*, **24**, 1333-1343.
- Hastenrath, S. (2007). Circulation mechanisms of climate anomalies in East Africa and the equatorial Indian Ocean. *Dynamics of Atmospheres and Oceans*, **43**, 25-35.
- Heim Jr, R. R. (2002). A review of twentieth century drought indices used in the United States. *The Bulletin of the American Meteorological Society*, **83**, 1149-1165.
- Hills, R. C. (1979). The structure of the Inter-Tropical Convergence Zone in Equatorial Africa and its relationship to East African rainfall. *Transactions of the Institute of British Geographers, New Series* **4**, 329-352.
- Huffman, G. J., Adler, R. F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B., and Schneider, U. (1997). The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bulletin of the American Meteorological Society*, **78**, 5-20.
- ICPAC (2008). *Statement from the twenty second Greater Horn of Africa Climate Outlook Forum, 28-29 August 2008*. ICPAC, Nairobi, Kenya.
- IMF (International Monetary Fund) (2008). World Economic Outlook Database 2008. <http://www.imf.org/external/pubs/ft/weo/2008/01/weodata/index.aspx>, last accessed 25/09/09.

- Indeje, M., Semazzi, F. H. M., and Ogallo, L. A. (2000). ENSO signals in East African rainfall seasons. *International Journal of Climatology*, **20**, 19-46.
- Indeje, M. and Semazzi, F. H. M. (2000). Relationships between QBO in the lower equatorial stratospheric zonal winds and east African seasonal rainfall. *Meteorology and Atmospheric Physics*, **73**, 227-244.
- Indeje, M. (2000). *Prediction and numerical simulation of the regional climate of equatorial eastern Africa*. PhD Thesis, North Carolina State University, USA.
- IRI (2005). Regional Climate prediction and risk reduction in the Greater Horn of Africa. http://unfccc.int/files/adaptation/sbsta_agenda_item_adaptation/application/pdf/200711_nwp_iri_prediction_africa.pdf, last accessed 05/01/10.
- ISS (Institute for Security Studies) (2009). Kenya, Economy. <http://www.iss.co.za/AF/profiles/kenya/Economy.html>, last accessed 05/01/10.
- Janowiak, J. E. (1988). An investigation of interannual rainfall variability in Africa. *Journal of Climate*, **1**, 240-255.
- Jewson, S., Brix, A., and Ziehmann, C. (2005). *Weather derivative valuation: the meteorological, statistical, financial and mathematical foundations*. Cambridge University Press, Cambridge, UK.
- Jones, A. E., Uddenfeldt Wort, U., Morse, A. P., Hastings, I. M., and Gagnon, A. S. (2007). Climate prediction of El Nino malaria epidemics in north-west Tanzania. *Malaria Journal*, **6**, 162.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woolen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J. E., Mo, K. C., Ropelewski, C. F., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437-470.
- KMD (Kenya Meteorological Department) (2008). *Review of the weather in June-July-August (JJA) 2008 seasons and the outlook for the October-November-December 2008 "short rains" season*. KMD, Nairobi, Kenya.

- KMD (Kenya Meteorological Department) (2009). *Climate outlook for March-April-May 2009 ("long-rains" season)*. KMD, Nairobi, Kenya.
- Korecha, D. and Barnston, A. G. (2007). Predictability of June-September rainfall in Ethiopia. *Monthly Weather Review*, **135**, 628-650.
- Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, C. R., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S. (1999). Improved weather and seasonal climate forecasts from multi-model superensemble. *Science*, **285**, 1548-1550.
- Lau, K. M. and Sheu, P. J. (1988). Annual cycle, Quasi-Biennial Oscillation and Southern Oscillation in Global precipitation. *Journal of Geophysical Research*, **93**, 10975-10988.
- Likumana, M. (2008). *Personal communication*, Tanzania Meteorological Agency, Tanzania.
- Lloyd-Hughes, B., Saunders, M. A., and Rockett, P. (2004). A consolidated CLIPER model for improving August-September ENSO prediction skill. *Weather and Forecasting*, **19**, 1089-1105.
- Luo, J., Behera, S., Masumoto, Y., and Sakuma, H. (2008). Successful prediction of the consecutive IOD in 2006 and 2007. *Geophysical Research Letters*, **35**, L14502.
- Madden, R. A. and Julian, P. R. (1994). Observations of the 40-50-day tropical oscillation - A review. *Monthly Weather Review*, **122**, 814-837.
- Mapande, A. T. and Reason, C. J. C. (2005). Interannual rainfall variability over western Tanzania. *International Journal of Climatology*, **25**, 1355-1368.
- Marchant, R., Mumbi, C., Behera, S., and Yamagata, T. (2006). The Indian Ocean dipole - the unsung driver of climatic variability in East Africa. *African Journal of Ecology*, **45**, 4-16.
- McHugh, M. J. and Rogers, J. C. (2001). North Atlantic Oscillation Influence on Precipitation Variability around the Southeast African Convergence Zone. *Journal of Climate*, **14**, 3631-3642.

- McHugh, M. J. (2004). Near-surface zonal flow and East African precipitation receipt during austral summer. *Journal of Climate*, **17**, 4070-4079.
- McHugh, M. J. (2006). Impact of south Pacific circulation variability on East African rainfall. *International Journal of Climatology*, **26**, 505-521.
- Muita, R. (2008). *Personal communication*, Kenya Meteorological Department, Nairobi, Kenya.
- Mukabana, J. R. and Pielke, R. A. (1996). Investigating the influence of synoptic-scale monsoonal winds and mesoscale circulations on diurnal weather patterns over Kenya using a mesoscale numerical model. *Monthly Weather Review*, **124**, 224-242.
- Mukabana, J. R. (2008). Press release: review of the performance of the "long rains" (March-May) 2008 season and outlook for June-August 2008. www.meteo.go.ke/pws/wx/jja2008.doc, last accessed 05/01/10.
- Murphy, A. H. (1995). The coefficients of correlation and determination as measures of performance in forecast verification. *Weather and Forecasting*, **10**, 681-688.
- Murphy, S. J., Washington, R., Downing, T. E., Martin, R. V., Ziervogel, G., Preston, A., Todd, M., Butterfield, R., and Briden, J. (2001). Seasonal forecasting for climate hazards: prospects and responses. *Natural Hazards*, **23**, 171-196.
- Mutai, C. C., Ward, M. N., and Colman, A. W. (1998). Towards the prediction of the East Africa short rains based on sea-surface temperature-atmosphere coupling. *International Journal of Climatology*, **18**, 975-997.
- Mutai, C. C. and Ward, M. N. (2000). East African rainfall and the tropical circulation/convection on interseasonal to interannual timescales. *Journal of Climate*, **13**, 3915-3939.
- Mutemi, J. (2009). *Personal communication*, University of Nairobi, Nairobi, Kenya.
- Mutimba, S. N. (2005). *Kenya energy atlas*. United Nations Development Program, Global Village Energy Project, Nairobi, Kenya.
- Mwagore, D. (2002). *Land use in Kenya: The case for a national land use policy*. Printfast Kenya Ltd., Nairobi, Kenya.

- Neng, S., Luwen, C., and Dongdong, X. (2002). A preliminary study on the global land annual precipitation associated with ENSO during 1948-2000. *Advances in Atmospheric Sciences*, **19**, 993-1002.
- News 24 (2006). Kenya's tea production declines. http://www.afdevinfo.com/htmlreports/org/org_20174.html, last accessed 05/01/10.
- Nicholson, S. E. and Entekhabi, D. (1986). The quasi-periodic behaviour of rainfall variability in Africa and its relationship to the Southern Oscillation. *Archives for Meteorology, Geophysics, and Bioclimatology, Series A*, 311-348.
- Nicholson, S. E. and Entekhabi, D. (1987). Rainfall variability in Equatorial and Southern Africa: relationships with sea surface temperatures along the Southwestern Coast of Africa. *Journal of Climate and Applied Meteorology*, **26**, 561-578.
- Nicholson, S. E. and Kim, J. (1997). The relationship of the El Nino-Southern Oscillation to African rainfall. *International Journal of Climatology*, **17**, 117-135.
- Nicholson, S. E. and Selato, J. C. (2000). The influence of La Nina on African rainfall. *International Journal of Climatology*, **20**, 1761-1776.
- Nicholson, S. E. (2000). The nature of rainfall variability over Africa on time scales of decades to millennia. *Global and Planetary Change*, **26**, 137-158.
- Nicholson, S. E. (2001). Climatic and environmental change in Africa during the last two centuries. *Climate Research*, **17**, 123-144.
- Nicholson, S. E. (2005). On the question of the "recovery" of the rains in the West African Sahel. *Journal of Arid Environments*, **63**, 615-641.
- Nicholson, S.E. (2007). *Personal communication*. Florida State University, Tallahassee, Florida, USA.
- Nicholson, S. E. (2008). *A revised view of the West African monsoon: in the ITCZ really necessary?* Presentation at the Conference on African Droughts, 2nd-6th June 2008, ICTP, Trieste, Italy.
- Ntale, H. K., Gan, T. Y., and Mwale, D. (2003). Prediction of East African seasonal rainfall using simplex canonical correlation analysis. *Journal of Climate*, **16**, 2105-2112.

- Ogallo, L., Janowiak, J. E., and Halpert, M. S. (1988). Teleconnections between seasonal rainfall over East Africa and global sea surface temperature anomalies. *Journal of the Meteorological Society of Japan*, **66**, 807-821.
- Ogallo, L. (1989). The spatial and temporal patterns of the East African seasonal rainfall derived from principal component analysis. *International Journal of Climatology*, **9**, 145-167.
- Ogallo, L. (2008). *Celebrating 10 years of COFs*. Presentation at the 21st Greater Horn of Africa Climate Outlook Forum, 27-29th February 2008, Entebbe, Uganda.
- Ogallo, L. (2008). *Personal communication*, ICPAC, Nairobi, Kenya.
- Ogallo, L. (1988). Relationships between seasonal rainfall in East Africa and the Southern Oscillation. *Journal of Climatology*, **8**, 31-43.
- Okoola, R. E. (1998). Spatial evolutions of the active convective patterns across the equatorial eastern Africa region during northern hemisphere spring season using Outgoing Longwave Radiation records. *Meteorology and Atmospheric Physics*, **66**, 51-63.
- Okoola, R. E. (1999). A diagnostic study of the Eastern Africa monsoon circulation during the northern hemisphere spring season. *International Journal of Climatology*, **19**, 143-168.
- Online Maps (2009). Map of Kenya. <http://en.18dao.net/images/8/84/Map-Kenya.jpg>, last accessed 05/01/10.
- Palisade (2009). *@Risk*. Palisade Europe, Middlesex, UK.
- Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., and Davey, M. e. al. (2004). Development of a European multi-model ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, **85**, 853-872.
- Patt, A. G., Ogallo, L., and Hellmuth, M. (2007). Learning from 10 years of climate outlook forums in Africa. *Science*, **318**, 49-50.

- Philippon, N., Camberlin, P., and Fauchereau, N. (2002). Empirical predictability study of October-December East Africa rainfall. *Quarterly Journal of the Royal Meteorological Society*, **128**, 2239-2256.
- Phillips, J. G. and McIntyre, B. (2000). ENSO and interannual rainfall variability in Uganda: implications for agricultural management. *International Journal of Climatology*, **20**, 171-182.
- Pohl, B. and Camberlin, P. (2006). Influence of the Madden-Julian Oscillation on East African rainfall. Part II: March-May season extremes and interannual variability. *Quarterly Journal of the Royal Meteorological Society*, **132**, 2541-2558.
- R Development Core Team (2009). *R: A language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria.
- Reason, C. J. C., Allan, R. J., Lindesay, J. A., and Ansell, T. J. (2000). ENSO and climatic signals across the Indian Ocean basin in the global context: part 1, interannual composite patterns. *International Journal of Climatology*, **20**, 1285-1327.
- Reed, R. J., Campbell, W. J., Rasmussen, L. A., and Rogers, R. G. (1961). Evidence of a downward propagating annual wind reversal in the equatorial stratosphere. *Journal of Geophysical Research*, **66**, 813-818.
- Rice, X. (2006). Kenya's tea industry wilts in the sun. <http://business.timesonline.co.uk/tol/business/markets/africa/article735324.ece>, last accessed 05/01/10.
- Ropelewski, C. F. and Halpert, M. S. (1987). Global and regional scale precipitation patterns associated with the El Nino/Southern Oscillation. *Monthly Weather Review*, **115**, 1606-1626.
- Ropelewski, C. F. and Halpert, M. S. (1989). Precipitation patterns associated with the high index phase of the Southern Oscillation. *Journal of Climate*, **2**, 268-284.
- Rowell, D. P., Folland, C. K., Maskell, K., and Ward, M. N. (1995). Variability of summer rainfall over tropical north Africa (1906-92): observations and modelling. *Quarterly Journal of the Royal Meteorological Society*, **121**, 669-704.
- 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.

- Saji, N. H. and Yamagata, T. (2003). Structure of SST and surface wind variability during Indian Ocean Dipole Mode events: COADS observations. *Journal of Climate*, **16**, 2735-2751.
- Saunders, M. A. and Lea, A. S. (2005). Seasonal prediction of hurricane activity reaching the coast of the United States. *Nature*, **434**, 1005-1008.
- Semazzi, F. H. M., Burns, B., Lin, N. H., and Schemm, J. K. (1996). A GCM study of the teleconnections between the continental climate of Africa and global sea surface temperature anomalies. *Journal of Climate*, **9**, 2480-2497.
- Shanko, D. and Camberlin, P. (1998). The effects of the southwest Indian Ocean tropical cyclones on Ethiopian drought. *International Journal of Climatology*, **18**, 1373-1388.
- Stefanova, L. and Krishnamurti, T. N. (2002). Interpretation of seasonal climate forecast using Brier skill score, the Florida State University superensemble and the AMIP-1 dataset. *Journal of Climate*, **15**, 537-544.
- Stockdale, T. N. (2007). *EUROSIP: seasonal forecasting with ocean-atmosphere models*. Presentation at the COSMOS General Assembly, 23rd-25th May 2007, Mainz, Germany.
- Sun, L., Semazzi, F. H., Giorgi, F., and Ogallo, L. (1999). Application of the NCAR regional climate model to eastern Africa 1. Simulation of the short rains of 1988. *Journal of Geophysical Research*, **104**, 6529-6548.
- Sun, L., Semazzi, F. H., Giorgi, F., and Ogallo, L. (1999). Application of the NCAR regional climate model to eastern Africa 2. Simulation of interannual variability of short rains. *Journal of Geophysical Research*, **104**, 6549-6562.
- Thompson, M. C., Doblus-Reyes, F. J., Mason, S. J., Hagedorn, R., Connor, S. J., Phindela, T., Morse, A. P., and Palmer, T. N. (2006). Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, **439**, 576-579.
- Trenberth, K. E. (1984). Signal versus noise in the Southern Oscillation. *Monthly Weather Review*, **112**, 326-332.

- Trenberth, K. E. (1997). The definition of El Nino. *Bulletin of the American Meteorological Society*, **78**, 2771-2777.
- Ummenhofer, C. C., Gupta, A. S., England, M. H., and Reason, C. J. C. (2009). Contributions of Indian Ocean sea surface temperatures to enhanced East African rainfall. *Journal of Climate*, **22**, 993-1013.
- UN (2006). Kenya national water development report. <http://unesdoc.unesco.org/images/0014/001488/148866E.pdf>, last accessed 05/01/10.
- UN and Department of Economic and Social Affairs, P. D. (2007). *World population prospects: The 2006 revision highlights*. Working paper No. ESA/P/WP.202 ESA/P/WP.202, New York, New York, USA.
- Uppala, S. M., Kallberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Anderson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfoof, J. F., Morcrette, J. J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woolen, J. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131**, 2961-3012.
- Verdin, J., Funk, C., Senay, G., and Choularton, R. (2005). Climate science and early warning. *Philosophical Transactions of the Royal Society B: Biological sciences*, **360**, 2155-2168.
- Veryard, R. G. and Ebdon, R. A. (1961). Fluctuations in tropical stratospheric winds. *Meteorological Magazine*, **90**, 125-143.
- Vitart, F., Huddleston, M. R., Deque, M., Peake, D., Palmer, T. N., Stockdale, T. N., Davey, M. K., Ineson, S., and Weisheimer, A. (2007). Dynamically-based seasonal forecasts for Atlantic tropical storm activity issued in June by EUROSIP. *Geophysical Research Letters*, **34**, L16915.

- Walker, G. T. (1924). Correlation in seasonal variations of weather, IX: A further study of world weather. *Memoirs of the Indian Meteorological Department XIV (part IX)*, 275-332.
- Webster, P. J. (1995). The annual cycle and the predictability of the tropical coupled ocean-atmosphere system. *Meteorology and Atmospheric Physics*, **56**, 33-55.
- Webster, P. J., Moore, A. M., Loschnigg, J. P., and Leben, R. R. (1999). Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997-98. *Nature*, **401**, 356-360.
- Wilhite, D. A. (2000). *Drought: a global assessment*. Routledge, New York, New York, USA.
- Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences (2nd edition)*. Academic Press, San Diego, California, USA.
- WMO (World Meteorological Organization) (2006). Standardised verification system (SVS) for long-range forecasts. http://www.bom.gov.au/wmo/lrfvs/Attachment_II-8.doc, last accessed 05/01/10.
- Xie, P. and Arkin, P. A. (1997). Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs. *Bulletin of the American Meteorological Society*, **78**, 2539-2558.
- Xie, S. P. and Annamalai, H. (2002). Structure and mechanisms of south Indian Ocean climate variability. *Journal of Climate*, **15**, 864-878.
- Yamagata, T., Behera, S. K., Rao, S. A., Guan, Z., Ashok, K., and Saji, N. H. (2002). The Indian Ocean Dipole: a physical entity. *CLIVAR Exchanges*, **7**, 15-17.
- Zablone, O. and Ogallo, L. A. (2008). Linkages between the Indian Ocean Dipole index and East African seasonal rainfall anomalies. *Journal of the Kenya Meteorological Society*, **2**, 3-17.