AN EMPIRICAL MODEL OF SAHEL CLIMATE

by

Andrew Matthews

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at the University of East Anglia

> Climatic Research Unit School of Environmental Sciences University of East Anglia Norwich

> > September 2005

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis, nor any information derived therefrom, may be published without the author's prior, written consent.

Abstract

Since 1968, the semi-arid region of sub-Saharan Africa, known as the Sahel, has seen a series of devastating droughts, which have been the subject of many studies. The majority of these have approached the problem by investigating the interannual variability of rainfall. Few have attempted to study the variability at shorter time scales. In this thesis, an empirical model is developed to study the daily variability of Sahel rainfall.

The model is developed in three parts. Initially, a 1° gridded daily data set of Sahel rainfall is formed for 1958 – 1997 from station data, using smoothing thin plate splines. A suite of atmospheric predictors is created by extracting the main modes of atmospheric variability in the NCEP reanalysis using a three-dimensional Principal Component Analysis. The domain of the PCA covers four atmospheric levels over sub-Saharan North Africa.

The atmospheric predictors are linked to daily rainfall in six regions of the Sahel, via a series of gamma-distributed Generalised Linear Models, with separate models developed for the two halves of the wet season. The predictors used in the model are picked with a forward stepwise selection procedure. Of the fields examined, specific humidity and zonal and meridional wind speed are seen to have the largest influence on rainfall.

June and July rainfall is most associated with the presence of easterly waves, and with the monsoon, providing southerly flow transports it over the Guinea Coast region. August and September are dominated by the strength of the monsoon flow in the Gulf of Guinea, but are also influenced by several factors describing variability in East Africa, which require further study.

Acknowledgements

I wish to express my thanks to all those who have aided me during the writing of this thesis. Firstly, many thanks to my supervisory team: Jean Palutikof, Tom Holt, Phil Jones and Declan Conway, for all their support, advice and encouragement. Secondly, thanks to the Climatic Research Unit and the School of Environmental Sciences at UEA for the friendly environment and technical support. In particular, thanks to Effie, Amanda, Kate, and all others I have shared an office with, for your assistance and understanding, and for making CRU an enjoyable place to work.

I am grateful to those who have generously provided data for use in this project. In particular, thanks to Daouda Diarra of the Mali Meteorological Service; to Peter Lamb and Issa Lele Mouhamadou of the University of Oklahoma; and to David Lister and CRU, for providing station rainfall data. The NCAR / NCEP reanalysis was provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, and is accessible via their Web site at http://www.cdc.noaa.gov. The thin plate splines were fitted using MATLAB routines obtained from Stephen Billings at the University of British Colombia.

Finally, a big thanks to Liz, and rest of my friends and family, for all the support and encouragement you've provided me over the past few years. Without you, this thesis would have remained unwritten.

Contents

ABSTRACT	
ACKNOWLEDGEMENTS	3
CONTENTS	4
LIST OF FIGURES	8
LIST OF TABLES	
CHAPTER 1: INTRODUCTION	25
1.1. Background	25
1.2. Where is the Sahel and what is its climate?	25
1.3. Historical Sahelian Climate	
1.4. A social perspective	29
1.5. Aims and Organisation of the Thesis	
CHAPTER 2: THE CURRENT STATE OF UNDERSTAND	ING OF SAHEL
CLIMATE	
2.1. Introduction - different approaches	
2.2. Elements of Sahel Rainfall Climatology	
2.2.1. Monsoon Flow	
2.2.1.a. A simple monsoon model	
2.2.1.b. The monsoon "jump" model	
2.2.1.c. Theoretical support for the "jump" model	
2.2.2. Other moisture sources	
2.2.3. Jet Streams	

2.2.3.a. The Tropical Easterly Jet Stream (TEJ)	
2.2.3.b. The African Easterly Jet Stream (AEJ)	
2.2.4. Mid-tropsopheric Instability: Easterly Waves and Squall Lines	
2.2.4.a. Easterly Waves	
2.2.4.b. Squall Lines	40
2.2.4.c. Interactions between Squall Lines and Easterly Waves	40
2.3. Changes in Sahel Climate	41
2.3.1. Seasonal Variability – Changes in the mean state	
2.3.1.a. Interannual trends and periodicity in Sahel rainfall	42
2.3.1.b. Spatial patterns in seasonal and annual Sahel rainfall	43
2.3.2. Seasonal Variability – Changes in the annual cycle	48
2.3.3. Intraseasonal Variability	50
2.2.3.a. Daily Variability	50
2.3.3.b. Madden-Julian Oscillation	51
2.4. Sea Surface Temperatures	52
2.4.1. Linking Global Sea Surface Temperatures to Sahel Rainfall	
2.4.2. Changes in the links between SSTs and West African rainfall	58
2.5. The Influence of Land Surface Processes on Sahel Rainfall	61
2.6. Summary and Conclusions	64
CHAPTER 3: CONSTRUCTION OF A RAINFALL DATASET	66
3.1. The Raw Rainfall Datasets	66
3.2. The Creation of a Gridded Dataset	69
3.2.1. The Selection of Suitable Station Data	69
3.2.2. Interpolation using Continuous Global Surfaces	
3.2.3. Smoothing Continuous Global Surfaces	72
3.2.4. The Gridding Procedure: Smoothing Thin-Plate Splines	73
3.3. Evaluation of the Gridded Dataset	77
3.3.1. The Number of Stations Used in the Gridding Procedure	77
3.3.2. Examples of Fitted Surfaces	79
3.3.3. A Comparison of Station and Gridded Data	86
3.3.4. Representation of the Drought in the Gridded Dataset	92

4.1. The NCEP / NCAR reanalysis dataset	94
4.2. An Introduction to Principal Component Analysis	96
4.2.1. The Purpose of Principal Component Analysis	97
4.2.2. Rotation of Principal Components	103
4.2.3. An example of PCA on the Sahelian gridded rainfall data set	111
4.2.4. Complex extensions of PCA	116
4.2.5. Multiway extensions of PCA	117
4.2.6. An Assessment of PCA	119
4.3. The selection of a domain for the PCA – Analysis of the NCEP reanalysis fields	119
4.4. Execution and results of the Principal Component Analysis	128
4.5. A validation of the PCA	134
CHAPTER 5: CREATING THE EMPIRICAL MODEL	136
5.1. Linking atmospheric variability with rainfall	136
5.2. Generalised Linear Models (GLMs)	138
5.2.1. The formulation of GLMs	138
5.2.2 Generalised Linear Model Output	139
5.2.3 Model Selection	143
5.3. Fitting the model to daily Sahelian rainfall	145
5.3.1. Details of the model	146
5.3.2. Is the gamma model valid?	147
5.3.3. The Initial "Full" Model	155
5.3.4. Improving the model	158
5.3.5. Fitting to the gridded rainfall series	167
5.4. Conclusions	170
CHAPTER 6: SUMMARY AND CONCLUSIONS	172
6.1. Background	172
6.2. A summary of the thesis	
6.2.1. The Climate of the Sahel	173
6.3.2. The creation of a rainfall data set	175
6.3.3. The creation of a suite of predictor variables	175
6.3.4. The final empirical model	177

6.3. Recommendations for future studies	
APPENDIX A: PCA PATTERNS	183
APPENDIX B: GLM OUTPUT	246
GLOSSARY	319
Commonly used acronyms and abbreviations:	
Commonly Used Mathematical Symbols:	
REFERENCES	323

List of Figures

Figure 1.1. Map of the Sahel
Figure 1.2. Average annual rainfall across the Sahel, 1958-1997
Figure 1.3. Average seasonal evolution of the West African monsoon at 3°W 27
Figure 1.4. A Sahel rainfall index, 1958-1997
Figure 3.1. Location of the 523 stations
Figure 3.2. Number of complete yearly and wet season (June to September) records for the whole data set
Figure 3.3. Number of complete monthly records per country, January 1958 to December 2000
Figure 3.4. The three interpolation regions used in the gridding of the daily rainfall dataset
Figure 3.5. Number of stations used to create gridded dataset for each gridding region, 1958-1997
Figure 3.6. Number of stations in the vicinity of the centre of each grid box – averaged over the period 1958-1997
Figure 3.7. Surfaces fitted to station rainfall in the central region on 10 th August 1996
Figure 3.8. Smoothing thin-plate splines fitted to daily rainfall for the central region for 17 th and 19 th August 1970
Figure 3.9. Smoothing thin-plate splines fitted to daily rainfall for the central region for 9 th June 1966, with and without the square root transform

Figure 3.10. Daily root-mean-square error of the gridded dataset when compared to original station values, calculated for each station used in the gridding process.

- Figure 3.16. Trend of each grid box in the annual gridded rainfall, 1958-1997. 93
- Figure 4.2. Example of Buell patterns for a 6 × 6 square, after (Buell, 1975)...... 107
- Figure 4.3. The varimax rotation of the first ten components from Figure 4.2..... 108

Figure 4.8. Correlation coefficient between monthly anomalies in geopotential height
and Sahel rainfall, June to September, 1958-1997 123
Figure 4.9. Correlation coefficient between monthly anomalies in specific humidity
and Sahel rainfall, June to September, 1958-1997
Figure 4.10. Correlation coefficient between monthly anomalies in vertical velocity
and Sahel rainfall, June to September, 1958-1997 125
Figure 4.11. Correlation coefficient between monthly anomalies in zonal wind and
Sahel rainfall, June to September, 1958-1997 126
Figure 4.12. Correlation coefficient between monthly anomalies in meridional wind
and Sahel rainfall, June to September, 1958-1997 127
Figure 5.1. The six rainfall regions to be modelled using GLMs 145
Figure 5.2(a). Validation plots for the Far West region
Figure 5.2(b). Validation plots for the Central West region
Figure 5.2(c). Validation plots for the Central North region
Figure 5.2(d). Validation plots for the Central South region
Figure 5.2(e). Validation plots for the Central East region
Figure 5.2(f). Validation plots for the Far East region
Figure 5.3. Evolution of the deviance and R^2 goodness of fit diagnostics during the
stepwise variable selection procedure for the initial GLM fitted to the Central
West region156
Figure 5.4. Bootstrapped coefficients of the specific humidity variables in the
original GLM fitted to the Far East region158
Figure 5.5. Histograms of actual and fitted regional rainfall for the Far West region.

- Figure 5.6. Histograms of deviance residuals for the two models fitted to the Far West region using atmospheric predictors leading rainfall by five days....... 164
- Figure 5.8. Bootstrapped coefficients of variables in the modified GLM fitted to the Far East region, for June-September, using predictors at a lag of five days... 167

- Figure A.10. Loading pattern and time series averages of the fourth rotated principal component carried out on May October air temperature data from the NCEP reanalysis. 200

- Figure A.20. Loading pattern and time series averages of the third rotated principal component carried out on May October specific humidity data from the NCEP reanalysis. 210

- Figure A.26. Loading pattern and time series averages of the third rotated principal component carried out on May October vertical velocity data from the NCEP reanalysis. 216

- Figure A.32. Loading pattern and time series averages of the fifth rotated principal component carried out on May October zonal wind data from the NCEP reanalysis.
- Figure A.34. Loading pattern and time series averages of the first rotated principal component carried out on May October meridional wind data from the NCEP reanalysis. 224

Figure A.47. Loadings of the fourth unrotated principal components of the PCA
carried out on two halves of air temperature data
Figure A 49 Loadings of the first unrotated minsingl components of the DCA
Figure A.48. Loadings of the first unrotated principal components of the PCA
carried out on two halves of geopotential height data
Figure A.49. Loadings of the second unrotated principal components of the PCA
carried out on two halves of geopotential height data
Figure A.50. Loadings of the third unrotated principal components of the PCA
carried out on two halves of geopotential height data
Finne A 51 Leading of the fourth constant desire inclusion and a fithe DCA
Figure A.51. Loadings of the fourth unrotated principal components of the PCA
carried out on two halves of geopotential height data
Figure A.52. Loadings of the first unrotated principal components of the PCA
carried out on two halves of specific humidity data
Figure A.53. Loadings of the second unrotated principal components of the PCA
carried out on two halves of specific humidity data
Figure A 54. Loadings of the third unrotated principal components of the PCA
and out on two holyes of anosific hymidity data
carried out on two narves of specific humidity data
Figure A.55. Loadings of the fourth unrotated principal components of the PCA
carried out on two halves of specific humidity data
Figure A.56. Loadings of the first unrotated principal components of the PCA
carried out on two halves of vertical velocity data
Figure A 57 Loadings of the second unrotated principal components of the PCA
carried out on two halves of vertical velocity data 240
carried out on two narves of vertical velocity data
Figure A.58. Loadings of the third unrotated principal components of the PCA
carried out on two halves of vertical velocity data
Figure A.59. Loadings of the fourth unrotated principal components of the PCA
carried out on two halves of vertical velocity data

Figure A.60. Loadings of the first unrotated principal components of the PCA
carried out on two halves of zonal wind speed data
Figure A.61. Loadings of the second unrotated principal components of the PCA
carried out on two halves of zonal wind speed data
Figure A 62 Loadings of the third unrotated principal components of the PCA
carried out on two halves of zonal wind speed data
carried out on two narves of zonar whild speed data
Figure A.63. Loadings of the fourth unrotated principal components of the PCA
carried out on two halves of zonal wind speed data
Figure A.64. Loadings of the first unrotated principal components of the PCA
carried out on two halves of meridional wind speed data
Figure A.65. Loadings of the second unrotated principal components of the PCA
carried out on two halves of meridional wind speed data
Figure A 66, Loadings of the third unrotated principal components of the PCA
carried out on two halves of meridional wind speed data 245
curred out on two harves of mendional while speed data
Figure A.67. Loadings of the fourth unrotated principal components of the PCA
carried out on two halves of meridional wind speed data
Figure B.1. The number of atmospheric factors included in the GLMs fitted to each
of the six rainfall regions
\mathbf{F} and $\mathbf{D} = \mathbf{A} \mathbf{A}$ is a standard \mathbf{D}^2 for the CLM. Standard standard state of the size
Figure B.2. Adjusted R for the GLMs fitted to each of the six rainfall regions 247
Figure B.3. Deviance of the GLMs fitted to each of the six rainfall regions
Figure B.4. Intercept coefficient in the GLMs fitted to each of the six rainfall
regions
Figure B.5. 'Prior1' coefficient in the GLMs fitted to each of the six rainfall regions.
The 'Prior1' variable uses quantity of rainfall on the previous day as a predictor
and is included in the model to account for autocorrelation in the rainfall series.

- Figure B.12. Coefficient relating to the first rotated principal component of geopotential height in the GLMs fitted to each of the six rainfall regions..... 257
- Figure B.13. Coefficient relating to the second rotated principal component of geopotential height in the GLMs fitted to each of the six rainfall regions..... 258
- Figure B.14. Coefficient relating to the third rotated principal component of geopotential height in the GLMs fitted to each of the six rainfall regions..... 259
- Figure B.15. Coefficient relating to the fourth rotated principal component of geopotential height in the GLMs fitted to each of the six rainfall regions..... 260
- Figure B.16. Coefficient relating to the fifth rotated principal component of geopotential height in the GLMs fitted to each of the six rainfall regions..... 261
- Figure B.18. Coefficient relating to the second rotated principal component of specific humidity in the GLMs fitted to each of the six rainfall regions....... 263

- Figure B.20. Coefficient relating to the fourth rotated principal component of specific humidity in the GLMs fitted to each of the six rainfall regions....... 265

- Figure B.33. Coefficient relating to the first rotated principal component of meridional wind speed in the GLMs fitted to each of the six rainfall regions. 278
- Figure B.34. Coefficient relating to the second rotated principal component of meridional wind speed in the GLMs fitted to each of the six rainfall regions. 279
- Figure B.35. Coefficient relating to the third rotated principal component of meridional wind speed in the GLMs fitted to each of the six rainfall regions. 280
- Figure B.36. Coefficient relating to the fourth rotated principal component of meridional wind speed in the GLMs fitted to each of the six rainfall regions. 281
- Figure B.37. Coefficient relating to the fifth rotated principal component of meridional wind speed in the GLMs fitted to each of the six rainfall regions. 282
- Figure B.38. Coefficient relating to the sixth rotated principal component of meridional wind speed in the GLMs fitted to each of the six rainfall regions. 283
- Figure B.39. Coefficient relating to the seventh rotated principal component of meridional wind speed in the GLMs fitted to each of the six rainfall regions. 284
- Figure B.40. Coefficient relating to the eighth rotated principal component of meridional wind speed in the GLMs fitted to each of the six rainfall regions. 285
- Figure B.41. Coefficient relating to the ninth rotated principal component of meridional wind speed in the GLMs fitted to each of the six rainfall regions. 286
- Figure B.42. Coefficient relating to the tenth rotated principal component of meridional wind speed in the GLMs fitted to each of the six rainfall regions. 287

Figure B.45. Deviance of the GLMs fitted to individual grid boxes...... 290

Figure B.46. Intercept coefficient in the GLMs fitted to individual grid boxes..... 291

- Figure B.47. 'Prior1' coefficient in the GLMs fitted to individual grid boxes. The 'Prior1' variable uses quantity of rainfall on the previous day as a predictor and is included in the model to account for autocorrelation in the rainfall series.. 292

List of Tables

Table 3.1. Percentiles of ν for surfaces fitted to June to September days for each of
the six gridded datasets
Table 4.1. The six modes of decomposition for climatological studies. Adapted from
(Richman, 1986)
Table 4.2. Statistics for the first ten eigenvalues of the PCA carried out on the daily
gridded Sahel rainfall data set. See text for further explanation 112
Table 4.3. Details of eigenvalues and percentage of total percentage of variance
explained for each of the six PCAs. The bottom three rows indicate the number
of eigenvalues greater than or equal to one, the number of components
eventually extracted, and the percentage of total variance they explained
collectively 129
Table 4.4. Main physical processes involved in each of the principal component
patterns
Table 5.1. Size of model and goodness of fit statistics for the initial six models 156