Chapter 6: Summary and Conclusions

6.1. Background

The region of West Africa just south of the Sahara Desert, known as the Sahel, saw a significant decrease in rainfall in the final decades of the twentieth century. These droughts, and their effect on agriculture, have had a devastating effect on the population and economies of the countries of the Sahel. As yet, the scientific community does not fully understand the mechanisms behind the droughts.

This thesis has investigated the relationships between rainfall across the Sahelian band, and atmospheric variability. These relationships were investigated at a daily time scale using a statistical model. The development of this model occurred in four stages, each described by a chapter of the thesis. In Chapter 2, past studies of the Sahel were investigated to establish a suitable approach to the problem, and to identify mechanisms that merited further study. In Chapter 3, the available station rainfall data was collated into a daily gridded rainfall data set using smoothing thin plate splines, creating appropriate response variables for the final model. Chapter 4 focused on the identification of a suitable domain for the atmospheric predictor variables, and the reduction in dimension of those variables via a three-dimensional principal component analysis. Finally, in chapter 5, the empirical model was created. Generalised linear models were used to link the response and predictor variables, revealing common circulation patterns related to Sahelian rainfall.

The following chapter summarises the main findings of this thesis, and recommends some possible avenues for future study.

6.2. A summary of the thesis

6.2.1. The Climate of the Sahel

A review of past studies of the Sahel described a climate system dominated by high spatial and temporal variability. Currently, Sahel rainfall is dominated by a marked low-frequency temporal variability across the whole region. Sahel rainfall since 1968 is estimated to be 20 – 40 % lower than in the period between 1931 and 1960 (Nicholson et al., 2000).

The West African monsoon consists of two stages (Lebel et al., 2003). The first stage, known as the Oceanic phase, is dominated by the progressive onset of rain across West Africa from the tropical Atlantic Ocean. The second stage, known as the Continental phase, is marked by a sudden rise in mean daily rainfall and the number of rainfall events resulting from large convective systems.

The transfer between the phases is marked by an abrupt shift in the ITCZ. In the period between 1968 and 1990, the mean date for this shift was the 24th June, with a standard deviation of 8.0 days (Sultan and Janicot, 2000). This jump is believed to occur when a threshold level in the boundary-layer entropy gradient in reached, moving the circulation from a radiative-convective regime to an angular momentum conserving regime (Eltahir and Gong, 1996; Le Barbe et al., 2002).

This theory is backed up by empirical evidence. Nicholson and Palao (1993) demonstrate there are important differences in the variability in annual June/July and annual August/September rainfall. Whilst these periods do not exactly divide the monsoon into the two phases, clearly June and July will be dominated by the Oceanic phase, and August and September by the Continental phase. Most interannual variability occurs as a result of variability in the August and September / Continental phase.

Interdecadal variability in Sahel can also be separated into two regimes. In the 1950s and the early 1960s, annual rainfall was significantly higher in the Sahel than in the period since 1968. In the earlier period, rainfall over the whole of west Africa tended to be characterised by a dipole pattern, with high rainfall in the Sahel contrasted by low rainfall along the Guinea Coast, and vice versa (Nicholson and Palao, 1993). Annual rainfall was most strongly linked to variations in sea surface temperature in the Atlantic Ocean. In particular, a warm south Atlantic reduced the gradient toward West Africa, reducing the extent of the northward migration of the ITCZ. This resulted in increased rainfall along the Guinea Coast, and drought in the Sahel: the West African rainfall dipole (Camberlin et al., 2001).

Conditions since 1968 have changed significantly. Annual rainfall in the Sahel has been significantly lower, and the West African rainfall dipole is less apparent (Nicholson and Palao, 1993). There have also been marked changes in the atmospheric circulation over the Sahel. The ITCZ has become less intense, and the monsoon flow has not penetrated as far up in the atmosphere as in the dry period. The Tropical easterly jet (TEJ) has become weaker, whereas the African easterly jet (AEJ) has become stronger, but has been located further south, over the monsoon flow (Grist and Nicholson, 2001). The strength of the link between Sahelian rainfall and Atlantic SSTs has reduced. Instead, the role of the Pacific and ENSO has become prominent (Janicot et al., 2001), although the mechanism by which ENSO affects Sahel rainfall is unclear.

The atmospheric factors influencing interannual variability also influence intraseasonal variability. Sultan et al. (2003) discovered a quasi-periodic cycle in rainfall averaging a period of 15 days, which was linked to dynamic changes in the ITCZ, AEJ and TEJ, very similar to those seen by Grist and Nicholson (2001) in their interannual study. However, the intraseasonal studies have stressed the important role of other factors, such as the Madden-Julian Oscillation (Matthews, 2004) and African easterly waves and squall lines (Fink and Reiner, 2003). Furthermore, the role of events to the east of the Sahel should not be ignored, as the area provides an important source of moisture for Sahelian rainfall (Long et al., 2000; Fontaine et al., 2003).

Finally, the conditions of the West African land have an impact on Sahelian rainfall. Charney (1975) suggested that a large-scale change in albedo could be the cause of the Sahelian drought, but observational evidence suggests that such a large-scale change has not occurred (Nicholson et al., 1998). However, changes in surface variables such as evapotranspiration and soil moisture have been shown to have an effect on circulation and rainfall (e.g. Cook, 1999; Zheng et al., 1999). The reason for the recent abrupt changes in Sahel rainfall is still unclear, but sea surface temperatures and land conditions have clearly been important factors.

6.3.2. The creation of a rainfall data set

The main analysis of this thesis began with the development of a daily gridded rainfall data set using smoothing thin plate splines. The data set was generated at a 1° resolution for three domains, covering an area from Senegal to western Niger, using data from 416 daily rainfall stations. The time series for each grid point represented rainfall in the vicinity of that point, for each day of the period 1958-1997.

A study of the distributions of the raw and gridded data sets revealed that the gridded data set under-represents rainfall variability at the daily scale, but represents monthly and yearly variability well. A possible alternative data set, formed using a square-root transform, could not represent the longer-term variability, and so was rejected.

Analysis of the spatial distribution of stations, and the degree of smoothing in the three domains, suggested that the central domain was the best represented. Areas on the edge of the domain, particularly in the east, are seriously affected by lack of data. However, only grid boxes that contained station data were used in later analyses. These grid boxes were grouped into six regions using a rotated principal component analysis. A brief regression analysis on the gridded data set illustrated a significant drying trend in all but the most southeasterly boxes.

6.3.3. The creation of a suite of predictor variables

The second stage of the analysis was the identification of a set of atmospheric predictors suitable for use in the empirical model. The atmospheric predictors were based on data taken from the NCEP reanalysis, due to the lack of observational data over West Africa. An analysis of the correlations between the NCEP data and Sahel rainfall at a monthly rainfall suggested a domain of 0 – 20 °N, 30 °W – 60 °E contains most of the important atmospheric factors linked to rainfall. Four atmospheric levels were considered: 1000, 850, 600 and 200 hPa. Daily NCEP data for six atmospheric fields were obtained over this domain: air temperature, geopotential height, specific humidity, vertical velocity, zonal wind and meridional wind.

The dimensionality of this data set was far too great to be incorporated into an empirical model, so was reduced using Principal Component Analysis (PCA). A PCA was performed on each of the six fields, reducing the data set to 37 predictor variables representing the main modes of 3-dimensional variability of the fields over the domain. The proportion of total variability extracted varied considerably, from 19.7 % for vertical velocity to 83.8 % for geopotential height. The components were rotated to increase interpretability.

Most rotated components generally represent coherent features, but vary considerably in scale and complexity. For example, 'shum5' represents a comparably simple pattern, measuring specific humidity over East Africa at all levels. Conversely, 'uwind1', containing 29 % of zonal wind variability, is a complex pattern incorporating the whole domain. It contrasts westerlies across the whole domain at the top of the atmosphere, and over the Gulf of Guinea in the middle atmosphere, against easterlies elsewhere, but particularly over the Indian Ocean at lower altitudes. The amount of variability represented indicates this is clearly an important factor in African atmospheric flow, and indeed it proves a good predictor at a lag of several days in the Chapter 5. However, the complexity of the pattern makes it hard to interpret.

The rotated components for vertical velocity produced very 'noisy' patterns, with multiple positive and negative areas in each plot. This is a disappointing, if unsurprising result: the underlying field is less coherent than the others used. Nevertheless, each factor is dominated by one coherent feature. For example, the most noticeable feature in 'omega1' is an area of ascent in the low atmosphere over the Horn of Africa. Therefore, they were retained for the next stage of the model. If, as suspected, they were too chaotic to be effective predictors, then the procedure used would ensure they were not selected.

6.3.4. The final empirical model

The study concluded by linking the two data sets via a Generalised Linear Model (GLM). The use of GLMs permits the response variable to come from a member of the exponential family of distributions. This makes them particularly useful for modelling daily rainfall, which is typically gamma-distributed.

Previous studies (such as Coe and Stern, 1982) use a two-step process to model rainfall: first fit a model establishing whether rain occurs on a given day, then model rainfall on the days when it occurs. In this study, the gridding and regionalising procedures meant some rainfall occurred almost every day, so only the second stage was necessary. Rainfall amounts were modelled using a gamma-distributed GLM with a log link function. Predictor variables were allowed to lead rainfall by up to five days, giving a total of 37 × 6 = 222 possible predictors. A suitable subset was selected using a forward-stepwise variable selection method, with a 5 % significance level required to change the model. The loadings of the fitted models represent the strength of the link between the factors and rainfall.

Initial results were not encouraging, despite the models explaining up to 25 % of the variability in regional rainfall. All models included 23 to 45 atmospheric variables, far too many to allow for a coherent physical explanation. This large number of variables suggests overfitting has occurred. This suspicion was investigated, and confirmed when a bootstrap analysis indicated large instability in the fit of some coefficients.

Several changes were implemented to improve the model. First, in order to reduce the number of terms and the interpretability of the output, each lag would be considered separately. Second, the significance level required to make a change in the model was increased to 1 %. Also, an additional predictor was included to account for any autocorrelation in the rainfall series: rainfall on the previous day (Chandler and Wheater, 2002). Finally, models were also fitted to June / July only, and August / September only, to allow for different influences on the two halves of the wet season (Nicholson and Palao, 1993).

These new models ('regional GLMs') explain up to 20 % of variability, using between 2 and 13 atmospheric predictors. Central regions tend to use less predictors than the outlying Far East and Far West regions. Similarly, models fitted with a longer lag time tend to be smaller. The central, short lead models tend to be the best fitting. These improvements had successfully reduced model size, making the results much more interpretable. Bootstrap tests also indicated the results were much more stable. Further models, referred to as 'gridded GLMs' were fitted to the gridded data points, increasing the resolution of results.

The 'rainfall on a previous day' predictor, referred to as 'Prior1', has a noticeable stronger influence in the Far West region, covering the Senegal coastline, due to the comparatively high autocorrelation in the daily rainfall time series of the area. Unfortunately, inclusion of 'Prior1' meant that modelled rainfall in this region was vastly overpredicted on days after major rainfall events, particularly in the early part of the wet season. Therefore, results for the Far West region for June and July models were rejected. Similar problems led to the rejection of all gridded GLMs in the Far West region over the whole wet season, with the exception of a few boxes in the south.

The fitted models indicate the complexity of rainfall across the Sahel, demonstrating that the most influential factors vary widely, depending on geographical location and the time of year. However, by considering the results as a whole, we can build up a general picture. In the first half of the wet season, there are five particularly important factors:

- 'shum1': High humidity over the Gulf of Guinea and Guinea Coast throughout the atmosphere. The pattern is linked to rainfall in the east immediately, and the west a few days later.

- 'vwind2': Dominated by northerlies at the surface, overlaid by southerlies at 600 hPa, centred at about 20 °E. 'vwind2' is linked to rainfall in the western half of the domain over the next three days. The link is not obvious; perhaps the increased vertical shear influences wave activity.

- 'vwind3': Low-level southerlies, centred over western Niger. This pattern suggests enhanced rainfall in the east of the domain, and reduced in the west (to the west of the southerlies). This indicates enhanced moisture transport across West Africa, determining whether the monsoon flow is reaching the inland parts of the Sahel.

- 'vwind5' and 'vwind6': Two very similar patterns, marked by a band of northerlies to the west of a band of southerlies in the low and middle troposphere. These patterns are linked to rainfall via patterns of positive and negative anomalies in rainfall, which over the course of five days move eastward, reminiscent of African Easterly Waves.

These factors are all also influential in the second half of the season. However, a number of additional factors become more important:

- 'uwind4': Describes a strengthening of low-level westerlies across the oceanic Gulf of Guinea. In the second half of the monsoon season, the shift of the ITCZ (Sultan and Janicot, 2000) means moisture from the Gulf of Guinea reaches Sahel latitudes more easily, hence this pattern can be physically linked to Sahelian rainfall.

- 'uwind5' and 'vwind4': Increased rainfall in the Sahel is associated with changes in atmospheric dynamics to the east. In particular, rainfall is associated with increased low-level easterly flow, and low and mid-level northerly flow, over Sudan. Unsurprisingly, this affects eastern regions before western. The link between motion and rainfall is not entirely clear; it may be related to changes in moisture transport or the effect on easterly waves.

- 'uwind1' and 'vwind10': These factors are good predictors of rainfall at higher lead times, but very hard to interpret. 'vwind10' suggests increased northerly flow at 200 hPa over Sudan is linked to an increase in rainfall in the eastern half of the domain 2-5 days later. 'uwind1' is a very complex large-scale pattern, relating increased rainfall drifting in from the east to increased strength of the westerlies at the top of the troposphere and over the Gulf of Guinea at 600 hPa, and to increased strength of the easterlies elsewhere, particularly over East Africa at low levels. Neither pattern can be easily linked to rainfall through a physical mechanism.

To put this picture in the context of Lebel et al.'s (2003) two-stage monsoon, in the early half of the rainfall season, easterly waves dominate, as the ITCZ has not yet 'jumped' to its more northerly position. However, at some point in late July, the jump occurs, and the increased southerly flow brings extra moisture into the Sahel. The 'vwind3' factor may indicate whether this jump has occurred. In the second half of the season, the ITCZ lies to the north of the region, so the strength of the monsoon flow over the Atlantic ('uwind4') becomes more important. Curiously, so do wind speeds in several areas to the east of the Sahel.

Thus, the three statistical techniques used in this thesis have allowed the construction of an empirical model that represents the factors enhancing and suppressing daily rainfall in the Sahel. The results confirm the two-phase nature of the monsoon season, the importance of the position of the ITCZ, and the critical nature of easterly waves. Finally, several modes of variability over East Africa are associated with Sahel rainfall, but the mechanism behind the links remains unclear.

6.3. Recommendations for future studies

This study has succeeded in creating a statistical model of the Sahelian climate, but there are many areas where the approach could be improved and extended. The ability of Generalised Linear Models to model daily rainfall has been demonstrated, but the results are rather unclear and, given extra time, would benefit from some extra investigation.

The use of a stepwise selection method has enabled the identification of several factors that influence rainfall. However, this method has also made interpretation difficult, as the importance of factors left out of the model is not investigated. These factors may still influence rainfall; this influence is better explained by the more dominant factors.

This problem could be resolved by the selection of a handful of variables, using these to fit models to all regions, periods and lags. The variables would need to be selected carefully; highly correlated variables would have to be avoided, to prevent problems with multicollinearity. However, the fitted models would allow a direct comparison of the changing nature of influences through time and space, and the comparative importance of the factors. Furthermore, incorporating interaction terms (defined as the product of two factors, see Chandler and Wheater, 2002) would allow investigation of the combined impact of the factors.

The understanding of West Africa as a whole could be improved by extending the range of this study. The current understanding notes that Sahel climate exhibits several strong links to the climate of the Guinea Coast, for example via the Sahel – Guinea rainfall dipole. If daily station data could be obtained for the coastal countries, the study could be extended southward. This would help verify some results, for example those related to movement of the ITCZ. Similarly, different processes are clearly at work in the Senegal region, as demonstrated by the higher autocorrelation in rainfall series. Further work to solve the deficiencies in the models fitted there would allow these processes to be investigated.

The model would also benefit from being extended in time. Up to date rainfall data could easily be incorporated, increasing the statistical power of the results. However, the greatest benefit would come from extending the analyses back to the wet period of the 1950s. This was not an option in this thesis due to doubts over the quality of the NCEP reanalysis prior to 1968 (Camberlin et al., 2001; Janicot et al., 2001). Other reanalyses, such as ERA40 (produced by the ECMWF), could be used. However, the NCEP deficiencies are related to a lack of observations before 1968; it seems unlikely that other data sources would not suffer from the same problem.

This study has tried to incorporate variability across the Sahel by investigating a large number of stations, and has modelled the 'spatial average' for several areas. However, the high level of spatial variability in daily rainfall in the area means that a large amount of smoothing has occurred. An alternative approach, taken by Chandler and Wheater (2002), is to model the rainfall at the station locations. This approach is more complicated, as it would necessitate a two-step GLM, and would risk missing the variability from the stations not studied. However, this approach might allow for a better modelling of individual rain events.

The fourth chapter of this thesis hinted at some possible statistical extensions to Principal Component Analysis that could benefit this, and other studies.

Unfortunately, they are not as well researched as PCA itself, and hence not well understood. Further studies into the application of these methods to climatological studies could assist in utilising these novel approaches.

Finally, the most mysterious aspect of the results is the importance of events to the east of the Sahel. A closer inspection of the variability of East African climate, and its relation to West Africa, could greatly enhance the understanding of what factors influence daily Sahelian rainfall.