# Desert locust populations, rainfall and climate change: insights from phenomenological models using gridded monthly data

**Jamie A. Tratalos[1],\*, Robert A. Cheke[1], Richard G. Healey[2], Nils Chr. Stenseth[3]**

[1]Natural Resources Institute, University of Greenwich at Medway, Central Avenue, Chatham Maritime, Chatham, Kent ME4 4TB, UK

[2]Department of Geography, University of Portsmouth, Buckingham Building, Lion Terrace, Portsmouth PO1 3HE, UK

[3]Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biology, University of Oslo, PO Box 1066, Blindern 0316, Oslo, Norway

ABSTRACT: Using autocorrelation analysis and autoregressive integrated moving average (ARIMA) modelling, we analysed a time series of the monthly number of 1° grid squares infested with desert locust *Schistocerca gregaria* swarms throughout the geographical range of the species from 1930–1987. Statistically significant first- and higher-order autocorrelations were found in the series. Although endogenous components captured much of the variance, adding rainfall data improved endogenous ARIMA models and resulted in more realistic forecasts. Using a square-root transformation for the locust data improved the fit. The models were only partially successful when accounting for the dramatic changes in abundance which may occur during locust upsurges and declines, in some cases successfully predicting these phenomena but underestimating their severity. Better fitting models were also produced when rainfall data were added to models of an equivalent series for desert locust hoppers (nymphs) that incorporated lagged data for locust swarms as independent variables, representing parent generations. The results are discussed in relation to predicting likely changes in desert locust dynamics with reference to potential effects of climate change.

KEY WORDS: *Schistocerca gregaria* · Time series · Population dynamics · Endogenous · Rainfall

## 1. INTRODUCTION

Predictions of how climate change will affect rainfall in the Sahara Desert are varied and lack consensus. In its most recent report, the Intergovernmental Panel on Climate Change (IPCC) suggested that there are likely to be 18% reductions in the median precipitation response for the 2080–2099 period compared to 1980–1999 for the northern Sahara (18° N, 20° E to 30° N, 65° E) from December to May (Christensen et al. 2007). However, for farther south, the IPCC report stated that it is unclear how rainfall in the Sahel, the Guinean Coast and the southern Sahara will evolve. There are some indications that, for instance in the Sahel, rainfall may increase as a result of increasing

carbon dioxide levels leading to enhanced vegetation growth and moisture levels (Claussen et al. 2003). If the latter proves to be the case, it may lead to increased frequencies of outbreaks of the desert locust *Schistocerca gregaria* (Forskål 1775), since locusts require moist soil for their eggs to incubate, and their migratory behaviour is adapted to take them towards areas where it has recently rained, so that they can feed on green vegetation and breed when mature. Predictions of what is likely to happen, even with a quantitative rainfall link, are difficult and will need to take account of spatiotemporal variation. For instance, desert locust outbreaks are initiated by gregarisation of solitary forms (see below, this section), which can take place in a variety of areas, not only in the Sahara

Desert (in zones where rainfall has been predicted to decrease, e.g. N of 18° N, remain as they are currently, or increase), but also on the Red Sea coast, the Horn of Africa, the Arabian peninsula and east as far as western India (see Fig. 3 of Cheke & Tratalos 2007).

In common with other locusts, the desert locust exhibits 2 distinct behavioural phases, the solitary phase, when individuals actively avoid one another, and the gregarious phase, when they form marching hopper bands (mass aggregations of flightless nymphs) and swarms capable of devastating crops (Uvarov 1966, 1977, Simpson 1999). Changes from the solitary to the gregarious phase are the result of a complex interaction of factors—including high rainfall allowing high survival rates, the type and distribution of the vegetation (Babah & Sword 2004), and behaviour, with the factor that finally leads to the change being an increase in the rate at which hairs on the locusts' back legs are touched by other locusts in a group (Simpson et al. 2001). In the solitary phase, the insects typically occur at low densities across a recession area, which extends for 16 million km$^2$ across the whole of the arid regions of North Africa, across the Middle East to northwest India. The locusts only migrate into the wider domain of the 'invasion area' (29 million km$^2$) frequented by swarms (Fig. 1) once they have begun to gregarise. A sharp increase in the number of swarms over a period of months forms an upsurge, which can lead to plagues lasting several years, before populations decline into periods of recession, when individuals are typically in the solitary phase (see Bennett 1975).

Due to the reliance of locusts on moistened soil for incubation and green vegetation for food in usually arid areas, some researchers have proposed that regional fluctuations in rainfall are the primary influence on locust numbers (see Pedgley 1981, Bennett 1975). However, theoretical studies by Cheke (1978), Blackith & Albrecht (1979), Cheke & Holt (1993, 1996) and Holt & Cheke (1996a) indicated that the population fluctuations of the desert locust may be due to endogenous dynamics. In contrast, Farrow & Longstaff (1986) argued that migration tended to prevent the carrying capacity of the environment ever being exceeded and that endogenous models would therefore do little to explain locust population dynamics. More recently, Magor et al. (2007) proposed that a general reduction in locust activity since 1965 has occurred due to a change in rainfall patterns brought about by a more restricted north–south oscillation of the Inter-Tropical Convergence Zone. Vallebona et al. (2008) showed a link between desert locust upsurges in West Africa, where breeding usually occurs during June to September but can also be from October to January of the following, after a stronger than normal westerly mid-latitude circulation there in March, followed by a weakened African Easterly Jet and a strengthened moisture advection from April to May.

Although many non-statistical case studies have been conducted on historical examples of migration and breeding sequences in the desert locust, few studies have applied statistical modelling to the large-scale spatio-temporal dynamics of the species. What work
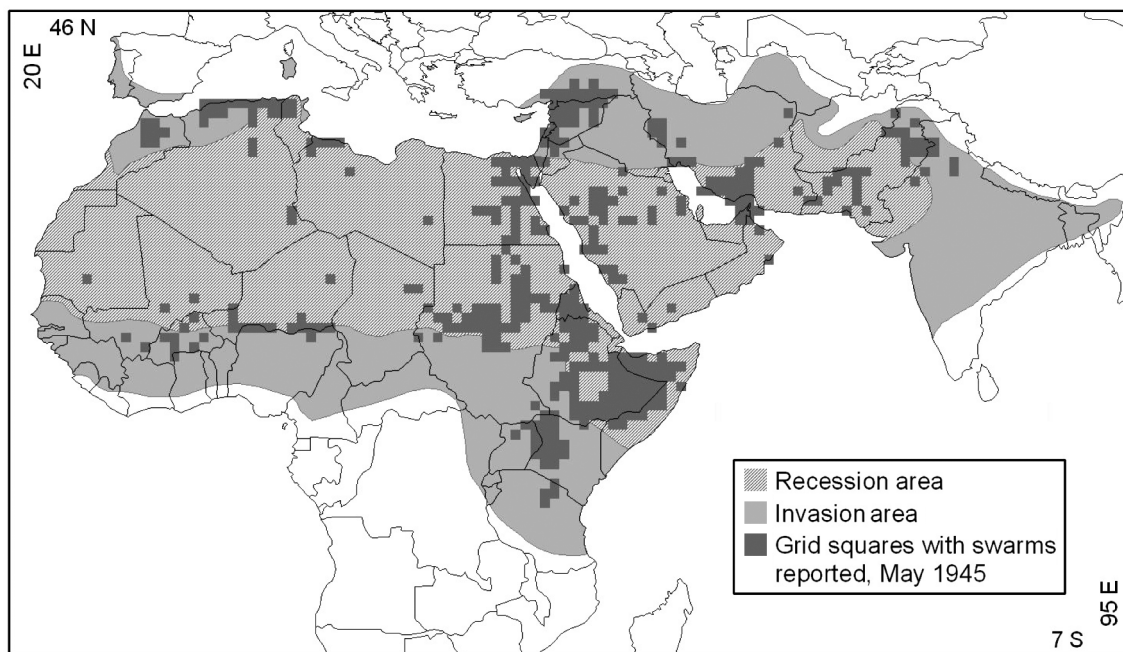


Fig. 1. *Schistocerca gregaria*. Desert locust invasion and recession areas, showing data on 1° grid squares reported as infested with swarms for the month of peak abundance (May 1945)

has been done (Waloff & Green 1975, Waloff 1976, Cheke & Holt 1993, Holt & Cheke 1996b) has used data aggregated at the spatial scale of a country, or large territories within countries, and the temporal scale of 1 yr. Since these publications have appeared, the FAO SWARMS dataset has become available. This database is a compilation of observations recorded by an extensive network of locust control personnel over a 58 yr period (1930–1987), gridded at 1° resolution. As such, it constitutes the most spatio-temporally complete information for any insect pest, with 58 yr of monthly data at 1° resolution covering the entire geographical range of the species (Magor & Pender 1997).

Given the complexity of the biology of the desert locust (see Cheke 1978 and Roffey & Magor 2003 for details of parameter estimates needed for population models), it is difficult to formulate clear-cut theoretical models whose predictions can be tested; therefore, at this stage, we prefer to consider only phenomenological models.

Using autoregressive integrated moving average (ARIMA) models, we analysed monthly data on the number of 1° grid squares with reported swarms and hopper bands of desert locusts during the period 1930–1987. First of all, we examined the dynamics of the locust data alone and then incorporated an index of monthly rainfall into the analyses. Through this we tested whether it is possible to predict locust plagues using endogenous data alone or if there is a need to incorporate rainfall into the modelling process to produce realistic forecasts. If the latter proved to be necessary, then it would form the basis for future tests of how predicted rainfall changes under future climate change scenarios might affect locust abundance.

This study is the first to apply a statistical time series modelling approach to analyse desert locust population dynamics throughout the geographical range of the species. It is also the first to examine desert locust dynamics at a resolution smaller than a national or very large territorial level.

## 2. METHODS

A time series of 696 counts of the monthly number of 1° grid squares reported as infested with desert locust swarms, from 1930 to 1987, throughout the desert locust distribution area, i.e. the recession and invasion areas combined (Fig. 1), was produced from a GIS from the original files used to compile the FAO SWARMS datasets (Healey et al. 1996, J. Magor pers. comm.). An equivalent series for desert locust hopper bands was also generated.

No smoothing or interpolation was applied to the data. An interval of 1 mo, rather than a generation, was chosen for the analysis, as it matched the original data

and as there is no standard period for a locust generation, which varies between ca. 7 wk and several months (Pedgley 1981). Furthermore, phase changes, and hence changes in the number of swarming locusts, can occur over much shorter periods. Data for the whole distribution area were modelled as one time series, as locusts are extremely mobile and are able to travel many thousands of kilometres in a single generation (e.g. Magor et al. 2007 documented migration from Saudi Arabia to Mauritania).

To examine the influence of rainfall on desert locust population dynamics, a time series of monthly rainfall totals for the desert locust recession area was also calculated for 1928–1987.

The rainfall data were derived from a 0.5° global land surface precipitation dataset, acquired from the Climate Research Unit (CRU) of the University of East Anglia (UEA) (see New et al. 2000). These data were taken only from the desert locust recession area (Fig. 1), rather than from the whole of the insect's range, because in recession years locusts are not typically found in the invasion area and therefore would not be able to breed with the arrival of suitable rains in that area. Furthermore, precipitation levels are relatively high in the invasion area, and rainfall is therefore unlikely to be a limiting factor there. Although these data were derived using spatial interpolation techniques from point observations, as we were modelling locust observations amalgamated over the whole range of the insect, we believe that they serve as a suitable proxy for real monthly variations in rainfall affecting overall locust abundance.

There are many grid squares where locusts have never been recorded breeding, and so it could be argued that the true breeding range of the species is mostly restricted to those squares where they have been known to breed. To take account of this, a monthly rainfall time series was produced for recession area grid squares which have at some time been reported as hosting breeding locust populations or locust hoppers, and results obtained for rainfall for the whole recession area checked against results obtained using these data.

The model selection strategy was to derive purely endogenous ARIMA models of the series and to examine the effect of adding lagged rainfall data as exogenous variables. In the case of the hopper bands data, lagged data from the swarms series were also included as exogenous variables, to represent parent generations. ARIMA models of the series were selected on the basis of an examination of autocorrelation and partial autocorrelation functions (ACF and PACF, respectively) using standard techniques (e.g. see Chatfield 1997). The effect of the inclusion of rainfall data, and locust swarms data in the case of the hopper bands

models, was tested using the significance levels of each variable and the effect on the Bayesian information criterion (BIC) value for the model, using backward stepwise techniques.

An ARIMA model consists of a forecasting equation which may include previous lags in the series, or 'autoregressive' terms, and lags of the forecast errors, or 'moving average' terms. A time series which needs to be differenced to be made stationary is said to be an 'integrated' version of a stationary series (Box & Jenkins 1976). The notation used to describe an ARIMA model is of the form (p d q)(P D Q)S. The first set of parentheses represents the non-seasonal part of the model, with p the order of an autoregressive process, d the order of differencing and q the order of a moving average process. The second set of parentheses represents the seasonal component, where P is the order of a seasonal autoregressive process, D the order of seasonal differencing and Q the order of a seasonal moving average process. S represents the length of the seasonal period (Diggle 1990, Chatfield 1997).

The ACF measures the correlation between values at each point in a series and values at lags prior to that point (Box & Jenkins 1976, Diggle 1990, Chatfield 1997). This information is further used to calculate the PACF, the correlation remaining between each point and lag in the series after the influences of all closer lags have been removed. The BIC is calculated as $-2\ln(L) + \ln(n)k$, where $L$ is the likelihood function based on the residuals from the model, $n$ is the number of residuals and $k$ is the number of free parameters (Schwarz 1978, Wei 1990). This value therefore takes into account both the fit of the model and its parsimony, and should be as low as possible.

The modelling techniques developed were phenomenological, in the sense that they were neither constrained within a theoretical framework nor analysed mechanistically in relation to a multitude of variables such as vegetation density and condition, soil types, temperature

and hydrology. We were interested in seeing what could be concluded from time series analyses of the locust data alone and then in testing whether inclusion of the main factor in locust survival, i.e. rainfall, improved the predictive power of phenomenological models derived from the time series analyses.

It should be borne in mind that sampling error may confound the results of ARIMA modelling, and that the locust data analysed here are based on coverage of vast areas of territory by locust control personnel, as well as from reports of locust infestations from other sources. Furthermore, the time series analysed are not a direct measure of locust abundance, but rather indicate the area that they covered during a given month. Both factors should be taken into account when interpreting the results of the analysis. However, although there is unlikely to be a strict proportional relationship between this series and actual abundance, a review of the literature suggests a generally good correspondence between the numbers of grid squares occupied and the abundance of swarming locusts (e.g. see Pedgley 1981).

## 3. RESULTS

### 3.1. Time series analyses and ARIMA modelling of locust swarms data

The locust swarms data showed a series of increases from very low to high levels. The latter were often maintained for several years and were followed by declines (Fig. 2), which suggested a high degree of serial correlation. Although it may appear from the time series that locust plagues are becoming a thing of the past, this series does not show the plague which developed between 1988 and 1989, the recent (2003–2005) locust outbreak or the major locust upsurges in 1992–1994 and 1997–1998, data for which were unavailable
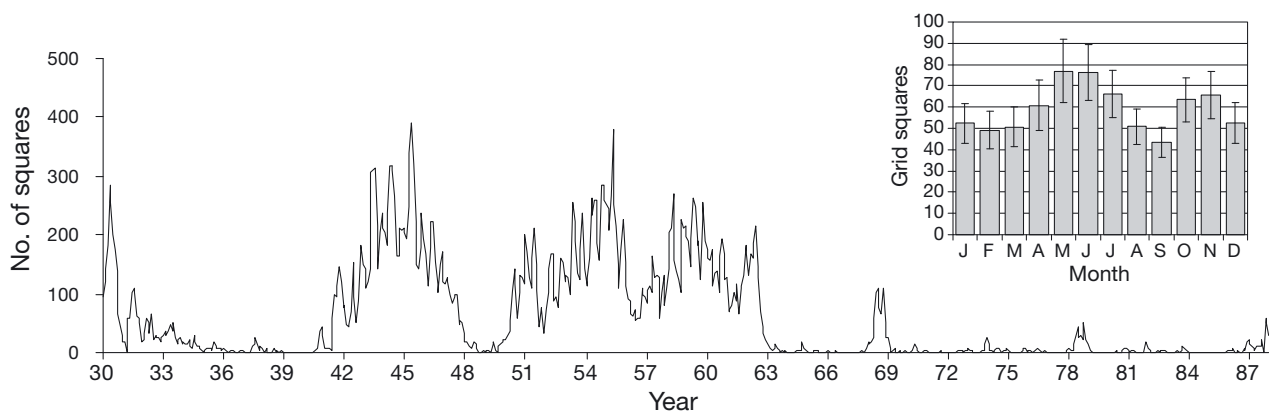


Fig. 2. *Schistocerca gregaria*. Number of 1° grid squares reported as infested with desert locust swarms, 1930–1987, as a monthly time series I, and mean numbers summarised by month (inset, in which error bars show ±2 SE)

in gridded form. For information on recent events, see Magor et al. (2007) or Mullié (2009, their Fig. 114a), who presented time series of non-gridded data of numbers of territories infested with swarms up to and including 2006. Summarising the gridded data by month showed some evidence of seasonality, with higher values for April–July, October and November than for January–March, August, September and December (Fig. 2, inset).

A square-root transformation for the locust data was used for the models to achieve equality of variance and a normal distribution (this was preferred to a logarithmic transformation, as 0 values meant that logarithms could only be calculated after adding an arbitrary constant to the series).

A 25 mo lag ACF of this square-root-transformed time series showed a high degree of autocorrelation, the most significant correlate being at lag 1 (Fig. 3), with additional periodicity revealed by significant positive lags in the PACF at lags 3, 5, 9 and 11 mo, and significant negative lags at 13, 14 and 25 mo (Fig. 3). All of these lags were also found to be significant in PACFs of untransformed data, in addition to many other lags (data not shown). Autocorrelations of both untransformed and transformed series remained significant over more than 100 lags, well in excess of what would be expected given the level of autocorrelation at lag 1.

An ARIMA model based purely upon endogenous factors was developed first. The data were seasonally differenced to take account of seasonal factors and to make the series stationary, and a square-root transformation was applied. An analysis of the ACF of these square-root-transformed data, in which only the first seasonal lag (at 12 mo) was significant, suggested a seasonal moving average process (SMA1). This was confirmed by an ARIMA (0 0 0)(0 1 1)12 model, which
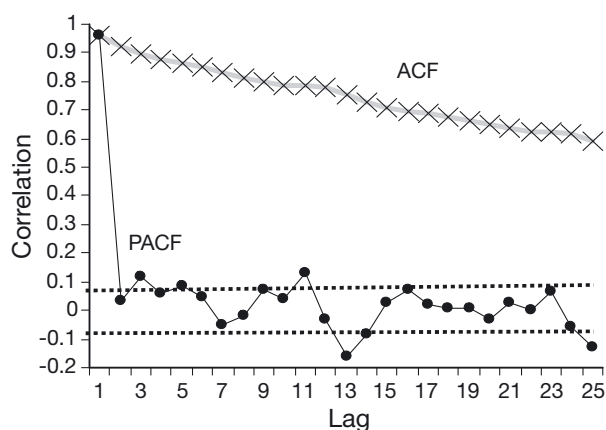


Fig. 3. *Schistocerca gregaria*. Autocorrelation (ACF, crosses) and partial autocorrelation functions (PACF, dots) of the square-root-transformed monthly number of 1° grid squares reported as infested with desert locust swarms, 1930–1987. Dotted horizontal lines: 95% confidence intervals

showed that the SMA1 term was highly significant (p < 0.01). An ACF and a PACF of the residuals from this model suggested the presence of mixed autoregressive (AR) and moving average (MA) processes. AR1 and MA1 terms were therefore incorporated, and both emerged as highly significant (p < 0.01). An AR2 term was also added, as it had been found to be significant in the PACF and was also highly significant in the resulting model. On the basis of these results, the final model selected was of the form (2 0 1)(0 1 1)12, in which all terms were highly statistically significant. This model is shown as Model 1 in Table 1. An ACF of the residuals from this model was similar to what would be expected from random data, and therefore suggested that the model was suitable. Additional endogenous parameters were not found to be statistically significant and also led to an increase in the BIC value.

Equivalent modelling techniques applied to untransformed data provided models with poorer fits which were less successful at accounting for the autocorrelation structure of the series. Equivalent models using an ln + 1 transformation gave similar results from autocorrelation analysis but a poorer fit to the data, when measured by the correlation between the real and predicted values (0.94 < 0.96; results not shown).

### 3.2. Inclusion of rainfall data

The lagged rainfall data were seasonally differenced and added to the endogenous (2 0 1)(0 1 1)12 model of square-root-transformed data. In a model incorporating rainfall at lags 0 (the current month) to lag 12, lags 4 to 12 all had highly significant positive coefficients, and lag 1 had a significant negative coefficient (Table 1, Model 2). When added to this model, rainfall lags more remote than lag 12 carried a statistically significant coefficient in only 1 case (lag 22) and in all cases had little effect on the other coefficients and brought about an increase in the BIC value. The removal of rainfall at lags 2 and 3 from Model 2 produced a model in which all terms were significant (Table 1, Model 3).

In this model, rainfall at lags 0 and 1 carried negative coefficients of similar size, and rainfall at lags 4 to 12 carried positive coefficients of similar size. In order to obtain a more parsimonious model, the individual rainfall lags added to the (2 0 1) (0 1 1)12 endogenous model were replaced by averages of rainfall at lags 0 to 1 and at lags 4 to 12 (Table 1, Model 4). This model gave the lowest BIC of all models tested. In comparison with an equivalent model in which the same rainfall lags were modelled separately, the coefficient for the average of the rainfall lags 0 and 1 terms was almost identical to the sum of the coefficients for these terms

Table 1. Summary of ARIMA Models 1 to 4. The endogenous elements in the models were derived from a time series of the square root of the number of 1° grid squares reported anywhere as infested with desert locust swarms in every month, 1930–1987; thus, the time step (*t*) = 1 mo. All rainfall data were derived from a similar time series of monthly rainfall totals for the desert locust recession area, 1928–1987. Coefficients for rainfall variables have been multiplied by $10^8$. Parentheses: 95% confidence intervals, *p < 0.05, **p < 0.01, ***p < 0.001, ns: not significant. AR: autoregressive; MA: moving average; SMA: seasonal moving average; Rain: rainfall; Log: log likelihood; BIC: Bayesian information criterion

| Parameter | MODEL 1 | | MODEL 2 | | MODEL 3 | | MODEL 4 | |
|---|---|---|---|---|---|---|---|---|
| AR 1 | 1.524 | (1.338, 1.711)*** | 1.459 | (1.301, 1.617)*** | 1.449 | (1.287, 1.61)*** | 1.441 | (1.279, 1.603)*** |
| AR 2 | –0.535 | (–0.715, –0.355)** | –0.47 | (–0.622, –0.318)** | –0.459 | (–0.616, –0.303)** | –0.452 | (–0.608, –0.296)** |
| MA 1 | 0.671 | (0.503, 0.839)*** | 0.646 | (0.506, 0.785)*** | 0.634 | (0.491, 0.778)*** | 0.628 | (0.484, 0.771)*** |
| SMA 1 | 0.899 | (0.878, 0.92)*** | 0.906 | (0.884, 0.928)*** | 0.908 | (0.886, 0.93)*** | 0.907 | (0.885, 0.928)*** |
| Rain *t* | | | –683 | (–1025, –341)* | –862 | (–1164, –560)** | | |
| Rain *t*-1 | | | –482 | (–870, –94) ns | –743 | (–1047, –439)* | | |
| Rain *t*-2 | | | 313 | (–111, 737) ns | | | | |
| Rain *t*-3 | | | 628 | (171, 1085) ns | | | | |
| Rain *t*-4 | | | 1393 | (912, 1874)** | 951 | (615, 1287)** | | |
| Rain *t*-5 | | | 1708 | (1205, 2211)*** | 1279 | (904, 1654)*** | | |
| Rain *t*-6 | | | 1721 | (1210, 2232)*** | 1320 | (910, 1730)** | | |
| Rain *t*-7 | | | 1483 | (975, 1991)** | 1140 | (707, 1573)** | | |
| Rain *t*-8 | | | 1340 | (854, 1826)** | 1037 | (605, 1469)* | | |
| Rain *t*-9 | | | 1470 | (1006, 1934)** | 1253 | (817, 1689)** | | |
| Rain *t*-10 | | | 1526 | (1096, 1956)*** | 1363 | (952, 1774)*** | | |
| Rain *t*-11 | | | 1372 | (982, 1762)*** | 1253 | (873, 1633)** | | |
| Rain *t*-12 | | | 1206 | (860, 1552)*** | 1130 | (790, 1470)*** | | |
| Rain *t* to *t*-1 | | | | | | | –1612 | (–2092, –1132)*** |
| Rain *t*-4 to *t*-12 | | | | | | | 10152 | (8063, 12241)*** |
| Log likelihood | –1211.12 | | –1190.56 | | –1191.55 | | –1192.66 | |
| BIC | 2448.36 | | 2492.1 | | 2481.03 | | 2424.5 | |

in the earlier model (–0.0000161 and –0.0000160 respectively), whereas the coefficient for the moving average of lags 4 to 12 was only ca. 6% lower than the sum of their coefficients in the earlier model (0.0001013 < 0.0001072).

There was no temporal trend in the residuals of any of the models incorporating rainfall, and ACFs indicated that there was little significant autocorrelation in the residuals, which was confirmed by the Box-Ljung statistics at each lag (Ljung & Box 1978).

The residuals from the purely endogenous Model 1 were generally larger throughout the length of the series than those from models incorporating rainfall, which can be seen from an examination of 12-monthly and 6-yearly averages of the absolute differences in the residuals (Fig. 4).

Replacing the rainfall series used in Models 2 to 4 with equivalent data restricted to grid squares where locust breeding or hoppers had been reported produced very little change in the models (BICs for models equivalent to 2, 3 and 4 = 2490.7, 2479.9, 2424, respectively).

The relationship between rainfall and square-rooted locust data was examined for evidence of heteroscedasticity, such as that described by Cheke & Holt (1993). The same moving average of lags 4 to 12 as used in Model 4 was plotted against values of the dependent locust series, for both undifferenced and seasonally differenced data. Little evidence was found

for a heteroscedastic response in either of these plots. R-square values were 0.05 for the undifferenced and 0.17 for the differenced series.

### 3.3. Analyses of the first and second halves of the series

To test the robustness of Models 1 to 4, each model was calculated separately for the first and second halves of the locust series (i.e. 1930–1958, 1959–1987). In all 8 models, all variables carried the same sign as they had in the equivalent model for the whole time series. In these 8 models, the coefficient for the AR1 term ranged from 1.05 to 1.59, the AR2 term from –0.6 to –0.06, the MA1 term from 0.26 to 0.73 and the SMA1 term from 0.82 to 1. For the models of the first half of the series, all endogenous variables were always significant, whereas in models of the second half of the series, only the AR1 term was significant, except in the case of the purely endogenous model, where the SMA term was highly significant. The moving average of rainfall at lags 4 to 12 had a highly significant positive coefficient in both models in which it was included, and the other 4 models including rainfall showed at least 4 rainfall variables at lags 4 to 12, significantly positively correlated (p < 0.05) with the dependent series.
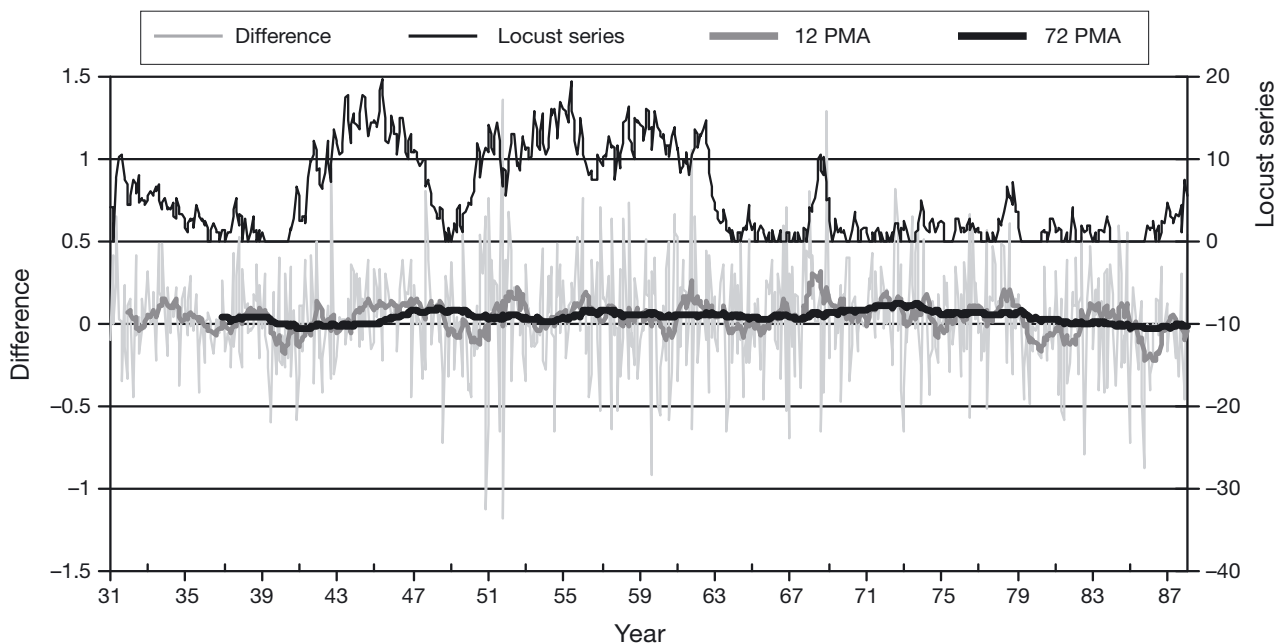
Fig. 4. Time series (1931–1987) of the residuals from a model incorporating rainfall (Model 2, see Table 1) subtracted from the residuals from a purely endogenous model (Model 1), shown as unaveraged (thin grey line), 12-monthly (thick grey line) and 72-monthly (thick black line) prior moving averages (PMA), alongside the square-root-transformed locust series modelled (thin black line)

## 3.4. Predictions of the models

To test the predictive power of the models, forecasts for the square-root-transformed locust series were produced for the period 1959–1987, using each of the models calculated only for the 1930–1958 period. These models did not use any locust data after 1958 but did use the rainfall series from 1959–1987 as an input to the forecast for this period. The same process was conducted using the purely endogenous Model 1, but in this case the forecast did not incorporate any rainfall data.

The above forecasts were also produced using the equivalent models calculated from data for the period from January 1930 to November 1963, the latter being the first month after the halfway point at which the number of grid squares reported as infested with swarms dropped to zero. Models incorporating rainfall data produced more accurate forecasts than those using only half of the previous history of the time series (Fig. 5). However, models based on data up to 1958 predicted much higher abundance than the more realistic models using locust data up to November 1963.

## 3.5. ARIMA modelling of a hopper bands time series, 1930–1987

Using the same techniques used for the swarms series, an endogenous ARIMA model of the square-rooted hopper bands series was selected, with the best model being of the form (1 0 1) (1 1 1)12 (BIC = 2410). To this model the lagged values of the swarms series were added (to represent parent generations) and, in a series of backward stepwise models, lags 1, 2, 4 and 8 were consistently found to be statistically significant. In all models, swarms data at lags 1, 2 and 4 carried positive coefficients and at lag 8 a negative coefficient, and no other lags were statistically significant in models containing swarms data at these lags. Adding these 4 lags as independent variables in the hopper bands ARIMA model brought about an improvement in the BIC value (2203 < 2410).

The effect of adding rainfall at lags 0 to 12 to the model was then tested using backward stepwise techniques. Rainfall at lags 1, 2 and 10 consistently emerged as carrying significant positive coefficients, and the introduction of these 3 variables together reduced the BIC to 2190 (Table 2). No other rainfall lags were found to be significant when these 3 variables were included. Further endogenous or exogenous inputs were not significant when introduced into this model.

Using rainfall only from grid squares where hopper bands or breeding locusts had been reported resulted in the selection of an equivalent model to that using rainfall from the whole recession area, but with a slightly poorer fit (BIC = 2190.9).
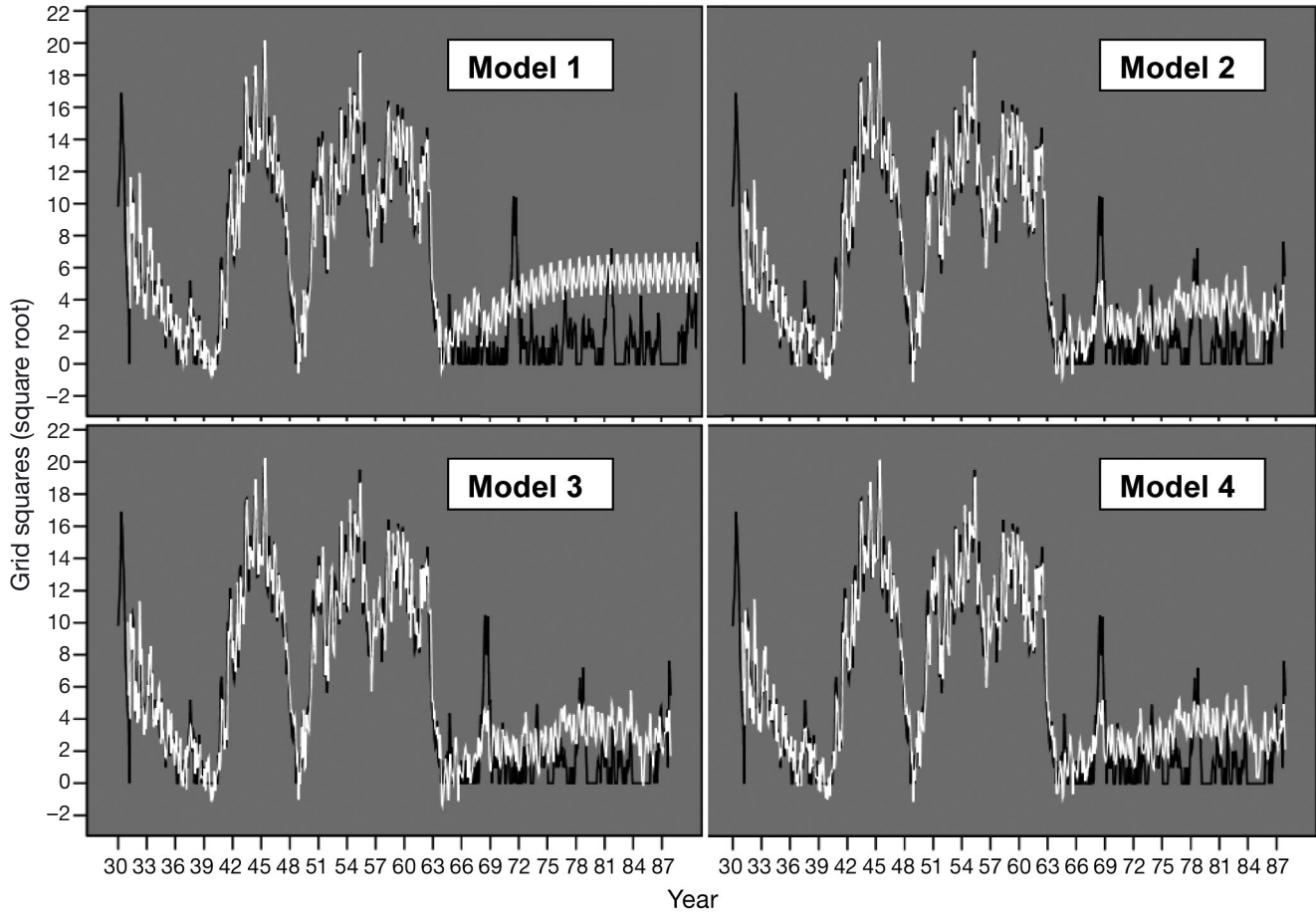
Fig. 5. Fitted values for models incorporating the same independent variables as Models 1 to 4 (white lines; see Table 1), alongside the series modelled (black lines: square-root-transformed monthly number of 1° grid squares reported as infested with swarms). Each model was calculated using data for the period January 1930 to November 1963. Fitted values for December 1963 to December 1987 are predictions from these models; Models 2 to 4 incorporate observed rainfall data for these years

Table 2. ARIMA model of the monthly number of 1° grid squares reported as infested with desert locust hopper bands, 1930–1987. Coefficients for rainfall variables have been multiplied by $10^8$. Time step (t) = 1 mo (see Methods for details). Parentheses: 95% confidence intervals. *p < 0.05, **p < 0.01, ***p < 0.001

| Parameter | Coefficient | |
|---|---|---|
| Autoregressive 1 | 0.252 | (0.177, 0.326)*** |
| Moving average 1 | −0.307 | (−0.38, −0.235)*** |
| Seasonal autoregressive 1 | 0.15 | (0.1, 0.202)** |
| Seasonal moving average 1 | 0.829 | (0.799, 0.861)*** |
| Swarms lag 1 | 0.367 | (0.337, 0.397)*** |
| Swarms lag 2 | 0.185 | (0.153, 0.217)*** |
| Swarms lag 4 | 0.075 | (0.042, 0.108)* |
| Swarms lag 8 | −0.082 | (−0.109, −0.056)** |
| Rainfall at $t$-1 | 1105 | (827, 1382)*** |
| Rainfall at $t$-2 | 1192 | (916, 1466)*** |
| Rainfall at $t$-10 | 672 | (406, 939)* |
| Log likelihood = −1059.14 | | |
| Bayesian information criterion = 2189.96 | | |

## 4. DISCUSSION

Although it did not examine the precise mechanisms driving changes in abundance, the phenomenological approach used here has identified the key roles of both endogenous factors and rainfall in determining the size of the territory occupied by locust swarms and hopper bands across the entire range of the species. As such, it offers hope for the better prediction of the effect of future changes in rainfall patterns on the locust problem and goes beyond approaches such as those by Vallebona et al. (2008), which are limited to examining locust upsurges arising under specific circumstances over relatively small, albeit important, areas (West Africa only) of the breeding range of the species.

The ARIMA models indicated an average generation time of approximately 4 mo, which fits with what is known about the biology of the insect, as swarming desert locusts generally undergo between 2 and 3 generations per year (Waloff 1976). For the swarms series,

rainfall lags more remote than lag 3 appear to be the most important predictors of future swarms, whereas for the hopper bands series, rainfall at lags 1 and 2 and swarms at lags 1, 2, 4 and 8 appeared to be the most important factors. These results suggest a process in which swarms produced in the previous generation breed on rains arriving approximately 2 or 3 mo later, to produce hopper bands after approximately 4 mo. These in turn may give rise to a second swarming generation in subsequent months.

Since the inclusion of rainfall variables improved the fit of purely endogenous ARIMA models, these analyses have clearly shown that desert locust population dynamics at an intercontinental scale are at least partly driven by rainfall, but have also indicated that endogenous control is important. However, as the time step was only 1 mo, which is less than the generation time of the locusts, autocorrelations at lag 1, and perhaps lags 2 to 3, would be expected. Nevertheless, there was also evidence for a substantial degree of 'memory' in the dynamics, with the ACF of the series showing significant lags up to 100 mo. The PACF is a measure of autocorrelation at lag $k$ that takes account of autocorrelations at lags $<k$, so significance in the PACF shows endogenous control not explained simply by the 1 mo time step. These results agree with findings by Cheke & Holt (1993), who found significant lags when analysing annual data on the number of territories infested by locust swarms.

For the swarms series, adding moving averages of rainfall at lags 0 to 1 and 4 to 12 produced a better model, even allowing for a decrease in parsimony resulting from the inclusion of these variables. Furthermore, incorporation of rainfall terms into predictive models provided more realistic predictions. However, the differences in BIC values between the best fitting endogenous model for swarming locusts and those incorporating rainfall data were small. Furthermore, in the hopper bands model, the addition of data on swarming populations, representing parent generations, reduced the BIC value considerably, and the swarms variables remained significant after the introduction of rainfall data into the model.

The presence of significant negative lags in the PACF of the locust series may indicate delayed density dependence (Holyoak 1994). The negative coefficient for the lag 8 swarms data in ARIMA models of the hopper bands series and the negative AR2 term in the swarms series models may also indicate delayed density dependence. This could be due to factors such as scarcity of food, increasing toxicity of food species exposed to several generations of high locust abundance, the build-up of predators (Smith & Popov 1953, Hudleston 1958, Pradhan 1961, Greathead 1966, Mullié 2009), reduced fecundity of locust populations

with a long history in the gregarious phase (as suggested by Waloff 1976), a tendency for inappropriate migrations in very gregarious populations (e.g. see Rosenberg & Burt 1999) or a time delay in the response of humans in controlling locust populations. Greathead (1966) argued that natural predators may act as a control on low density populations and may be important in bringing about declines, but would be unable to have much impact at the height of plagues. The efficacy of insecticidal control is unclear. Rainey et al. (1979) argued that new developments in control methods since the 1960s were effective in preventing plagues, but Waloff (1976, 1979) pointed out that the termination of only 4 plagues (those between 1950 and 1962) could be ascribed to control measures.

Examination of the residuals from the purely endogenous Model 1 and the endogenous + rainfall Models 2 to 4 showed that they all performed relatively poorly during periods of upsurge and decline, all suggesting more stable dynamics than occur in reality. This was especially noticeable for forecasts calculated using data from 1930–1958, where relatively high populations were predicted throughout the period 1959–1987. This may indicate the possibility of chaotic shifts during these periods, in which the system can shift unpredictably from relatively stable dynamics to periods of relatively rapid increase or decline in numbers. However, an examination of Fig. 5 does show that models incorporating rainfall data can predict upsurges and declines with some degree of success. The 1967–1968 and 1986–1987 upsurges were predicted, and the decline to low abundance during 1984 and 1985 was also forecasted, and especially well by Model 3. High abundance during 1979 was also successfully forecasted. Better models were produced for the swarms series ARIMA models when the locust data were square-root transformed. Using a square-root transformation on the data also tended to increase the degree of first-order autocorrelation and reduced the degree to which more remote lags were found to be significant in autocorrelation analysis. This indicates that at relatively high abundance upward or downward shifts in numbers tend to be greater than at low population levels.

It should be borne in mind that no data on solitary locust populations were included in these models. This was due to the unreliability of the available data, especially during the period up to the late 1950s, when regular surveys for solitary populations were not conducted (Magor & Pender 1997). Furthermore, whereas swarms and hopper bands are conspicuous events, likely to be reported by local people and observed by locust survey and control personnel, solitarious locusts will often have gone unnoticed. It is perhaps due to the absence of data on solitary populations in the models

that rainfall at remote lags (at 12 mo, for example) was found to be significant in models in which more recent data on locust abundance were included. These results may imply that rain falling 12 mo previously, may, through its effect on the abundance of successive generations of solitary locusts, produce an effect on the abundance of gregarious populations in the current month. This may constitute evidence against the swarm continuity hypothesis, the theory that breeding by a few gregarious, swarming populations remaining during recession periods is responsible for locust upsurges, as opposed to a build-up and gregarisation of solitarious populations (Hemming et al. 1979, Rainey & Betts 1979). The relatively poor performance of ARIMA models during periods of upsurge also suggests a role for solitarious populations. However, the fact that the dynamics of swarming populations could be modelled with a fair degree of success using endogenous data only from swarming populations is consistent with the swarm continuity hypothesis, but does not necessarily lend it support.

In the swarms data models, rainfall at lags 0 and 1 was negatively correlated with swarms data for the current month. This may seem to be a surprisingly counter-intuitive result, given that the survival of hopper populations may in some cases be due to the arrival of rainfall. However, what is known about locust ecology suggests that new swarms would most often be produced as a result of rainfall at lags more remote than lag 1 (Bennett 1976, Pedgley 1981). Secondly, it is known that both hopper bands and swarming populations tend to become increasingly solitary when vegetation is dense, as they are no longer crowded into just a few bushes and thus the gregarisation process is disrupted, and therefore an increase in vegetation due to rainfall may reduce gregarious populations. Thirdly, swarming populations are thought to be more sedentary when environmental conditions remain suitable, and would therefore be less likely to be observed. Finally, movements, growth rates and survivorship are likely to be reduced by lower temperatures, which may be brought about by higher rainfall.

Taking into account these 4 points, a negative relationship between rainfall at lags 0 and 1 and the number of grid squares reported as infested with swarms might be expected. Non-significant relationships between rainfall at lags 2 and 3 and swarms in the current month are perhaps due to a combination of these negative effects and the positive effect of rainfall on locust breeding and survivorship.

This study suggests that desert locust dynamics are influenced by endogenous factors and rainfall, and that broad patterns of locust upsurges and declines can be forecast with some degree of success using data on only these factors. Also, once predictions of likely cli-mate changes throughout the recession area can be made with more confidence than at present, our results could be helpful in forecasting whether locust plagues will become more or less frequent. For instance, if the forecasts of increased precipitation over important desert locust habitats in the Sahara along a west–east belt at about 20° N on the basis of the AB1 scenario (Hulme et al. 2001) are realised, then locust upsurges are likely to become more frequent.

## LITERATURE CITED

Babah MAO, Sword GA (2004) Linking locust gregarization to local resource distribution patterns across a large spatial scale. Environ Entomol 33:1577–1583

Bennett LV (1975) Development of a desert locust plague. Nature 256:486–487

Bennett LV (1976) The development and termination of the 1968 plague of the desert locust, *Schistocerca gregaria* (Forskål) (Orthoptera: Acrididae). Bull Entomol Res 66: 511–552

Blackith RE, Albrecht FO (1979) Locust plagues, the interplay of endogenous and exogenous control. Acrida 8:83–94

Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control. Holden-Day, San Francisco, CA

Chatfield C (1997) The analysis of time series. Chapman & Hall, London

Cheke RA (1978) Theoretical rates of increase of gregarious and solitarious populations of the desert locust. Oecologia 35:161–171

Cheke RA, Holt J (1993) Complex dynamics of desert locust plagues. Ecol Entomol 18:109–115

Cheke RA, Holt J (1996) Chaos in desert locust plague dynamics? In: Walter K, Mills N, Watt A (eds) Populations and patterns in biology. Intercept Press, Andover, p 42–45

Cheke RA, Tratalos JA (2007) Migration, patchiness and population processes illustrated by two migrant pests. Bioscience 57:145–154

Christensen JH, Hewitson B, Busuioc A, Chen A and others (2007) Regional climate projections. In: Solomon S, Qin D, Manning M, Chen Z and others (eds) Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, p 847–940

Claussen M, Brovkin V, Ganopolski A, Kubatzki C, Petoukhov V (2003) Climate change in northern Africa: tThe past is not the future. Clim Change 57:99–118

Diggle PJ (1990) Time series: a biostatistical introduction. Oxford University Press, Oxford

Farrow RA, Longstaff BC (1986) Comparison of the annual rates of increase of locusts in relation to the incidence of plagues. Oikos 46:207–222

Greathead DJ (1966) A brief survey of the effects of biotic factors on populations of the desert locust. J Appl Ecol

3:239–250

Healey RG, Robertson SG, Magor JI, Pender J, Cressman K (1996) A GIS for desert locust forecasting and monitoring. Int J Geogr Inf Syst 10:117–136

Hemming CF, Popov GB, Roffey J, Waloff Z (1979) Characteristics of desert locust plague upsurges. Philos Trans R Soc Lond B Biol Sci 287:375–386

Holt J, Cheke RA (1996a) Models of desert locust phase changes. Ecol Model 91:131–137

Holt J, Cheke RA (1996b) Réussites et échecs d'une règle simple de prévision des infestations du criquet pèlerin au Sahel. Sécheresse 7:151–154

Holyoak M (1994) Identifying delayed density dependence in time-series data. Oikos 70:296–304

Hudleston JA (1958) Some notes on the effect of bird predators on hopper bands of the desert locust (*Schistocerca gregaria* Forskål). Entomol Mon Mag 94:210–214

Hulme M, Doherty R, Ngara T, New M, Lister D (2001) African climate change: 1900–2100. Clim Res 17:145–168

Ljung GM, Box GEP (1978) On a measure of lack of fit in time series models. Biometrika 65:297–303

Magor JI, Pender J (1997) Desert locust forecasters' GIS: a researcher's view. In: Krall, S, Peveling R, Ba Diallo D (eds) New strategies in locust control. Birkhäuser Verlag, Basel, p 21–27

Magor JI, Ceccato P, Dobson HM, Pender J, Ritchie L (2007). Preparedness to prevent desert locust plagues in the central region, an historical review. Part 1: Text & Part 2: Appendices. Desert Locust Tech Ser No AGP/DL/TS/35. Food and Agriculture Organization of the United Nations, Rome

Mullié WC (2009) Birds, locusts and grasshoppers. In: Zwarts L, Bijlsma RG, van der Kamp J, Wymenga E (eds) Living on the edge: wetlands and birds in a changing Sahel. KNNV Publishing, Zeist, p 202–223

New M, Hulme M, Jones P (2000) Representing twentieth-century space–time climate variability. II. Development of 1901–96 monthly grids of terrestrial surface climate. J Clim 13:2217–2238

Pedgley DE (1981) (ed) The desert locust forecasting manual. Centre for Overseas Pest Research, London

Pradhan S (1961) Probable role of biotic factors in the periodicity of locust cycles. Indian J Entomol 1:1–6

Rainey RC, Betts E (1979) Continuity in major populations of migrant pests: the desert locust and African armyworm. Philos Trans R Soc Lond B Biol Sci 287:359–374

Rainey RC, Betts E, Lumley A (1979) The decline of the desert locust plague in the 1960s: control operations or natural causes? Philos Trans R Soc Lond B Biol Sci 287:315–344

Roffey J, Magor JI (2003) Desert locust population parameters. FAO Desert Locust Tech Ser No AGP/DL/TS/30. Food and Agriculture Organization of the United Nations, Rome

Rosenberg J, Burt PJA (1999) Windborne displacements of desert locusts from Africa to the Caribbean and South America. Aerobiologia 15:167–175

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Simpson SJ (1999) A behavioural analysis of phase change in the desert locust. Biol Rev Camb Philos Soc 74:461–480

Simpson SJ, Despland E, Hägele BF, Dodgson T (2001) Gregarious behavior in desert locusts is evoked by touching their back legs. Proc Natl Acad Sci USA 98:3895–3897

Smith KD, Popov GB (1953) On birds attacking desert locust swarms in Eritrea. Entomologist 86:3–8

Uvarov B (1966) Grasshoppers and locusts. A handbook of general acridology, Vol 1. Cambridge University Press, Cambridge

Uvarov B (1977) Grasshoppers and locusts. A handbook of general acridology, Vol 2. Centre for Overseas Pest Research, London

Vallebona C, Genesio L, Crisci A, Pasqui M, Di Vecchia A, Maracchi G (2008) Large-scale climatic patterns forcing desert locust upsurges in West Africa. Clim Res 37:35–41

Waloff Z (1976) Some temporal characteristics of desert locust plagues, with a statistical analysis by S.M. Green. Anti-locust Memoir No 13. Anti-Locust Research Centre, London

Waloff Z (1979) Discussion. In: Rainey RC, Betts E, Lumley A (1979) The decline of the desert locust plague in the 1960s: control operations or natural causes? Philos Trans R Soc Lond B Biol Sci 287:342

Waloff Z, Green SM (1975) Regularities in the duration of regional desert locust plagues. Nature 256:484–485

Wei WS (1990) Time series analysis: univariate and multivariate methods. Addison-Wesley, Harlow