* For each X/Y point with country ID=Ethiopia, obtain the value of LocPresent for each date.
* For each day, count (i.e. sum the 1s) the number of the grid points in the country (i.e square representing the country) where LocPresent is true.
* So you’ll get something like: (1/1/2000, 10) , (1/2/2000, 0), (1/3/2000, 50) etc. This is your time series.

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

**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.
  + **Exogenous:** Input variables that are not influenced by other variables in the system and on which the output variable depends.
* 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**.

**Introduction:**

* 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

* Recession area :
* Invasion area:
* 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, 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.
* 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.

**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, was produced from a GIS from the original files used to compile the **FAO SWARMS datasets.**
* An equivalent series for **desert locust hopper bands** was also generated.
* 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.
* 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).
* 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.
* The effect of the inclusion of **rainfall data**, and **locust swarms dat**a 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.
  + **ARIMA** Parameters
  + **p:** the number of lag observations in the model; also known as the lag order.
  + **d:** the number of times that the raw observations are differenced; also known as the degree of differencing.
  + **q:** the size of the moving average window; also known as the order of the moving average.
* The **Box-Jenkins Model** is a mathematical model designed to forecast data ranges based on inputs from a specified time series. ... The methodology allows the model to identify trends using **autoregression**, **moving averages**, and **seasonal differencing** to generate forecasts.
* A time series which needs to be differenced to be made **stationary** is said to be an ‘**integrated**’ version of a **stationary series**.
* 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.
* The **ACF** measures the **correlation between values at each point in a series and values at lags prior to that point**.
* 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 **–2ln(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. This value therefore takes into account both the fit of the model and its **parsimony**, and should be as low as possible.
* 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.

**ArchGis API :** [**https://developers.arcgis.com/rest/**](https://developers.arcgis.com/rest/)

**RESULTS**

**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, which suggested a **high degree of serial correlation**.
* Summarising the gridded data by month showed some evidence of **seasonality.**
* 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**, 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**.
* All of these **lags** were also found to be **significant in PACFs** of **untransformed data**, in addition to many other lags. **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).

**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**.