Analysis of Rainfall Data in Dodoma, Case study: Choosing the Suitable Model

Clementine Ushizimpumu, Dr. Roger Stern, Dr. David Stern

African Institute for Mathematical Sciences – Ghana

AIMS Ghana
African Institute for Mathematical Sciences

Introduction

- Weather data is collected at many regions all over the world. Rainfall data is a type of meteorological data, which can be analyzed for more information or for making predictions.
- Analysis of rainfall data is very important in our daily life and has many practical applications. This exploration aims to answer some of the questions from farmers in Tanzania (Dodoma region) such as, when the rains will start and when it will end, how long the rainy season will be, the distribution of rain throughout the year and the risk of dry spells.
- A suitable model, is a model which can explain more variabilities with few parameters.
- With some training to the extension officers, the results obtained could be of help to the farmers.

Objectives

To illustrate a suitable model for the site at Dodoma in Tanzania, by discussing graphs obtained from rainfall data using simple and modelling approaches.

Methodology

- We discussed different models of Rainfall data, supplied by Tanzania Meteorological agency (TMA) for 79 years, by using Instat software package.
- Simple approach simply extracts yearly the graphs of characteristics of interest from statistical summaries of the data, like annual rainfall amounts, the risk of dry spells etc. Modelling approach, fits a model which can be used to create graphs of the model or of summaries of simulate data. Therefore, in our discussion we used modelling approach.
- Using modeling approche, involves two phases. First phase, is modelling of the probability of rain on a given day, where it follows binomial distribution because it records only two outcomes (rain or dry). The second phase, is modelling of amount of rainfall on rainy day, where it follows Gamma distribution because amount of rainfall is bounded from zero and toward skewed to positive infinity.
- Our model is made up of two component: Markov chains for the condition of rainfall data and harmonics usually deal with the periodicity and seasonality
- A day with less than 0.85 mm is recorded as dry otherwise as rain. [1]

Results Discussion

- The figures below (Fig. 2A ,2B, 2C) show different fitted models of the probability of rainfall, obtained by using modelling approach, with the following model.

$$\log\left(\frac{P}{1-P}\right)=\alpha_o+\sum_{j=1}^{N}(\alpha_j\cos jt+\beta_j\sin jt)$$

Alpha and Beta are parameters to be estimated.

- The Logit transformation allows us to use any variate different functions, ensuring that the transformed proportion will lie between zero and one.
- Fig. 2A shows the model for the proportion of rain on any given day with eleven parameters.

Results Discussion(cot'nd)

The fitted curve on the proportion of rain is presented by f_r as shown in the below graph. To envelope the entire season clearly, available data will be shifted to get a good representation of the curve, as such our data will start from 1st August to 31st July.

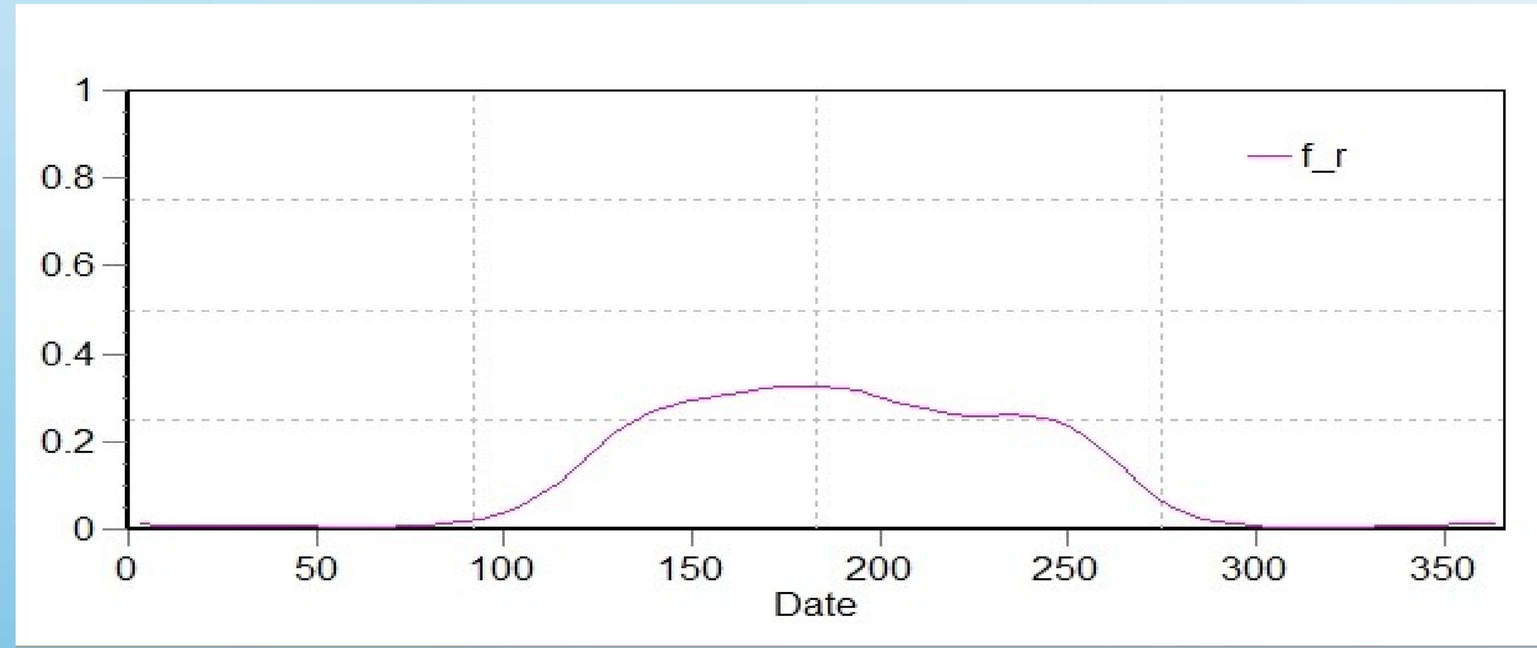

Figure 2A: Fitting harmonic on the proportion of rain

For first order modelling Markov chain, the idea is to split the curve for the probability of rain into two splits, where f_r will be splitted into f_rr and f_rd. The obtained splitted curves explain more variabilities than the original curve, thus The below graph

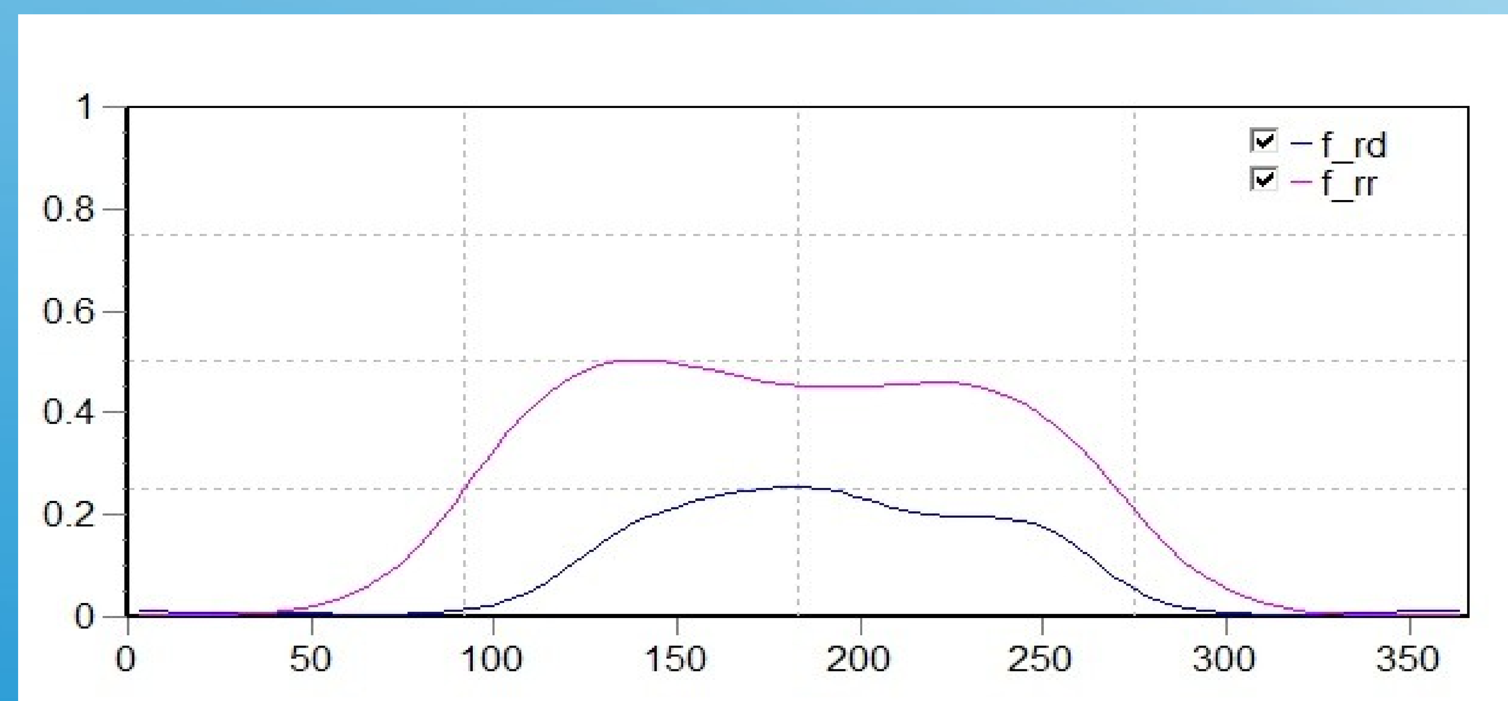

Figure 2. B: Fitted probability, first order modelling Markov chain

Fig. 2B represents a fitted model of first order modelling Markov chain with 16 parameters, Where f_rd means the fitted curve on the probability of rain given one day before was dry and f_rr is the fitted curve on the probability of rain given one day before was rainy.

For second order modelling Markov chain, the idea is to split curves for first order into four curves. Here f_rr splits into f_rrr and f_rrd, f_rd splits into f_rdr and f_rdd. To avoid to complicate the model we used only f_rr, because it can explain what f_rrr and f_rrd can explain, and we keep using f_rdr and f_rdd because they explain more variabilities than f_rd as shown in Fig. 2 C.

Fig. 2C represents a fitted model of second order modelling Markov chain with 24 parameters. Here f_rdd is the fitted curve on the probability of rain given two days before was rainy, f_rdr is the fitted curve on the probability of rain given the day before was dry and the day before yesterday was rainy and f_rr is the fitted curve on the probability of rain given one day before was rain.

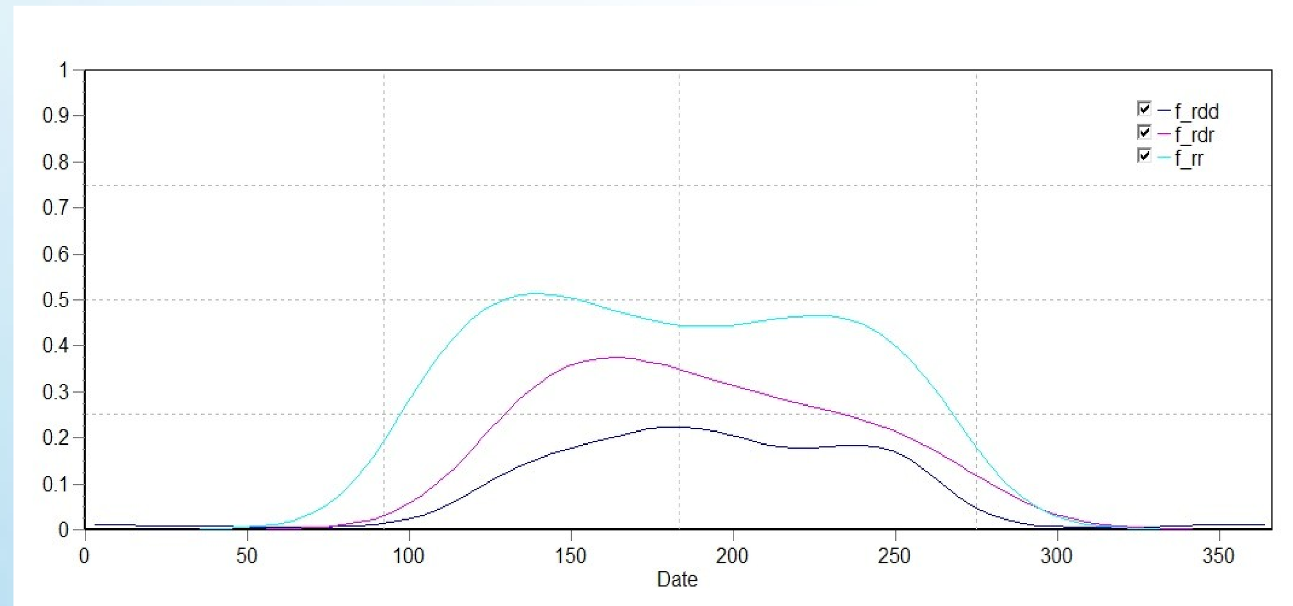

Figure 2C: Fitted probability second order modelling Markov chain

Fig. 3 represents the graph of the model of mean amount of rainfall per rainy day obtained by using the following model.

$$\log(\mu)=\alpha_o+\sum_{j=1}^{N}(\alpha_j\cos jt+\beta_j\sin jt)$$

Here fm represents the fitted curve on mean amount of rainfall per rainy day.

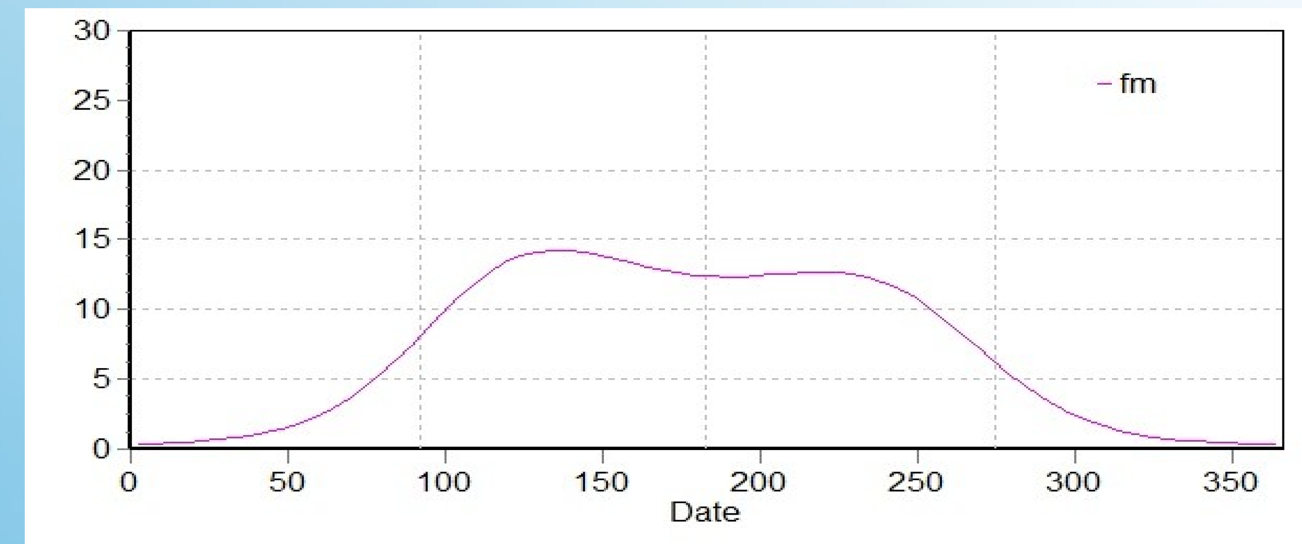

Figure 3: Fitted mean per rain day amount of rain, zero order modelling Markov chain

Conclusions

- We found that we needed one harmonic curve (first order model) for the chance of rain following rain and two harmonic curves (second order model) for the chance of rain after dry days because the results obtained show that, Fig. 2C explains *more variability* than others.
- We have been able also to model the amount of rain on rainy days, with a single curve.
- The model obtained starts from August to July (day 0 to day 366), thus unimodal.
- The obtained suitable model is fitted by three harmonic curves with 21 parameters. And the rain season runs from November to April.
- Generally this study will help the farmers from Dodoma to be aware of climate information and to know when and what they can plant to get good yields.
- For further work, we will use R software package to improve our model.

References

[1] Rachel Waters. African Meteorology: Making the best use of available data. Phd, University of Reading, 2005.