

South Africa drought and wildlife survival

Andrea Corrado 20205529

Contents

1	Introduction	2
2	Technical Background	2
2.1	Generalized Lienar Model	2
2.2	Generalized Additive Model	2
2.3	Weighted Least Square	3
2.4	Local Regression	3
2.5	Ordinary Differential Equation	3
2.6	Levenberg-Marquardt algorithm	3
2.7	Sensitivity Analysis	3
3	Exploratory Data Analysis	3
3.1	Rain fall amount	3
4	Climate Modeling	4
5	Species count Modelling	4
6	Conclusion	4

Contents

Abstract

1 Introduction

During last decades, interest and awareness about the climate change has steeply increased. There have been several studies examining and reporting the potential consequences of these phenomenon. For instance, the temperature and rainfall change has been of particular interest. Since last century, we have witnessed a increasing frequency in the number and severity of droughts in particular territories such as South Africa *SA* [1]. The consequences here can be very dramatic. Due to the fact that area such as *SA* suffer from low economic power, a drought may drive towards dramatic consequences. It can affect either agriculture, for a period that lasts up to 6 months, either hydrological up to 24 months [1].

There have been several studies showing how to reduce the impact of drought on the agriculture field by growing restraint to higher temperature and computational model to predict droughts [2].

However, in this paper, we wish to focus our attention on the drought consequences on wildlife population evolution. For instance, different studies have shown as animals mortality grows during the highest temperature period within the driest areas [3]. We will expand this idea to more recent data about waterbirds [4]. This data set will provide the number of individual for different species in each year within the last 45 years. By doing so, we are able to track the evolution of the species during time. However, the data set will not provide the number of new death, new born and the relative causes. It will be our job to build a model which allows us to investigate the relation other factors. In particular the data we are going to analyse regard the global temperature and rain fall amount [5] from 1901 up to 2020 which would give a foundation to build on the model of interest.

Having access to this information, our goal is to build a model which relates the climate change to the number of individual for each species and predict how, following the current trend, these would evolve over time and the potential consequences these could drive towards.

In section 2 we give a brief introduction to the methodology used within this research project. In section 3 we present the data with an exploratory data analysis. In section 4 we model the climate data and provide prediction on a possible future scenario. In section 5 we model the species count in

relation to the climate data and provide prediction about the future number of individuals. In section 6 we discuss the finding of the project.

2 Technical Background

2.1 Generalized Lienar Model

In order to achieve the goal, we are going to employ different methodology which would allow us to model the data of interest. In particular, first of all we will related the temperature data to the rain fall amount. In order to do so, we will employ the Generalized Linear Model [6] which theory would allow to establish a linear relationship between the quantity of interest (rain) and a set of covariates (temperature) plus a stochastic term which models the uncertainty about the random variable realisation

$$g(\mathbf{Y}) = \mathbf{X}\beta + \epsilon$$

where we define a distribution \mathcal{D} of interest on the stochastic term ϵ

$$\epsilon \sim \mathcal{D}(\theta)$$

and a link function $g(\cdot)$ which is distribution dependent.

The model will be validate by a set of diagnostic on the residuals $\mathbf{Y} - \hat{\mathbf{Y}}$ which provide meaningful insight on the goodness of fitness.

2.2 Generalized Additive Model

We will further extend the *GLM* [6] theory to allow for non-linear relationship between the predictors \mathbf{X} and the response \mathbf{Y} exploiting the *Generalized Additive Model* theory [7] while maintaining the model linear in the parameters

$$g(\mathbf{Y}) = f(\mathbf{X}) + \epsilon$$

where we define a distribution \mathcal{D} of interest on the stochastic term ϵ

$$\epsilon \sim \mathcal{D}(\theta)$$

and a link function $g(\cdot)$ which is distribution dependent. Moreover, we will define a function $f(\cdot)$ which usually is a non linear function such as a *spline* [8].

2.3 Weighted Least Square

During the research project, we perform different model estimation and to remedy to inadequate diagnostics, such as those violating the homoskedasticity assumption, we will employ *Weighted Least Square* estimation theory [9] which allows us to iteratively estimate a weight matrix \mathbf{W} to be employed in the model estimation as a measure of the importance for each unit \mathbf{x}_i $i = \{1, \dots, N\}$ where N is the total number of observations.

2.4 Local Regression

2.5 Ordinary Differential Equation

To track the evolution of a phenomenon over time, it is natural to think about *Ordinary Differential Equation (ODE)* theory [10] which allows us to keep track of the change of a quantity in continuous time

$$\frac{d}{dt}x = f(x)$$

given initial condition $f(x, t = 0) = x_0$

2.6 Levenberg-Marquardt algorithm

Very often, the function defined within an ODE is characterized by a set of parameters β . In empirical studies, we often have access to data realization $\{(y_i, \mathbf{x}_i)\}_{i=1}^n$ we want to find the parameters β so that $y = f(\mathbf{x}, \beta)$ best fits the data of interest. The *Levenberg-Marquardt* [11] provide a useful and efficient solution to the problem by making smart combination of *Gauss-Newton* and *Gradient Descent* theory.

2.7 Sensitivity Analysis

Mathematical modeling is often subject to very strong assumption which may be difficult to evaluate. In order to provide a robust framework, it is necessary to validate the model under perturbations on the findings. This is why, we will exploit *Local Sensitivity Analysis* theory [12] to evaluate the robustness of the estimated models.

3 Exploratory Data Analysis

3.1 Rain fall amount

During this exploratory data analysis context, we apply some pre-processing to the data do not impose any distribution on it. However, we make use of non parametric model, such as local regression presented in section 2, to get an idea of the overall trend.

In the first place, we investigate the rainfall data. We have access to the information between January 1901 up to December 2020. For simplicity, we decide to compute an annual average and keep track of it over time.

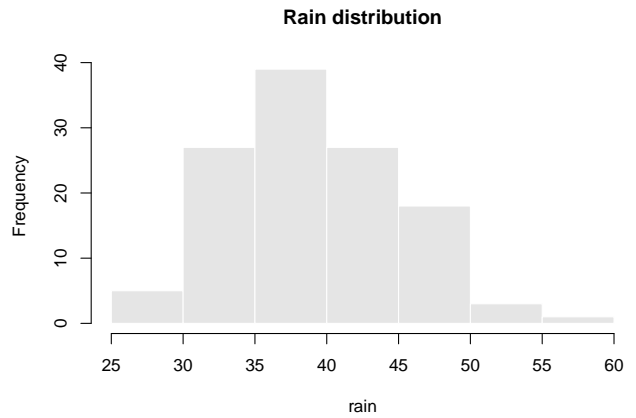


Figure 1: Yearly rainfall distribution

From figure 1 we can see that the data distribution looks approximately *Normal*. However, we can easily notice that the distribution is right skewed meaning that we can rarely observe very extreme event such as very heavy rain. We can quantify the amount of skewness by computing the data moments ratio known as the *Skewness* index which result to be 0.42.

To track the evolution of the phenomenon we fit a local regression models with different values of the span. In particular:

- 100% of the neighbor data to take into account a global context
- 50% of the neighbor data to allow for a more local structure

In despite of the fact that different values of the span were used, in figure ?? we can observe a similar trend, indicating a constant trend for the first half of the last century. Afterwards, depending on the model fitted,

we can observe a linearly decreasing trend which result in nowadays lower amount of rainfall

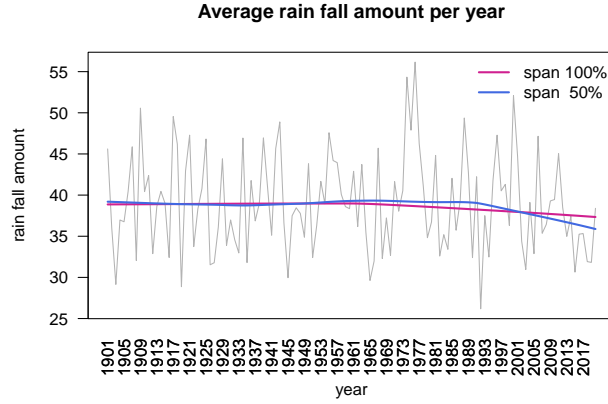
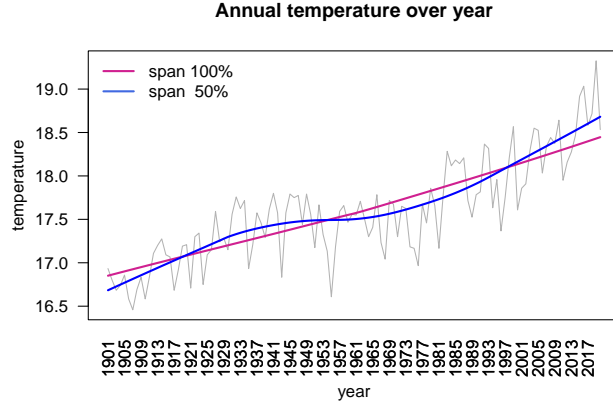


Figure 2: Annual rainfall



	year	rain	temp
year	1.00	-0.06	0.82
rain	-0.06	1.00	-0.33
temp	0.82	-0.33	1.00

4 Climate Modeling

5 Species count Modelling

6 Conclusion

References

[1] M. R. . Y. Richard, "Intensity and spatial extension of drought in South Africa at different time

scales," *African Journals Online*, vol. 29, no. 4, 2003.

- [2] M. Grayson, "Agriculture and drought," *Nature Outlook*, vol. 501, no. 7468, 2013.
- [3] J. C. H. et. al, "Mortality of wildlife in Nairobi National Park, during the drought of 1973-1974," *African Journal of Ecology*, vol. 15, no. 1-18, 1977.
- [4] G. G. B. I. Facility, "Free and open access to biodiversity data." www.gbif.org/occurrence/.
- [5] W. B. Group, "Climate change knowledge portal." <https://climateknowledgeportal.worldbank.org/download-data>.
- [6] P. McCullagh and J. Nelder, *Generalized Linear Models*. International series of monographs on physics, Chapman and Hall, 1989.
- [7] R. T. Trevor Hastie, *Generalized Additive Models*. 1990.
- [8] M. W. M. Aertsa, G. Claeskensb, "Some theory for penalized spline generalized additive models," *Journal of Statistical Planning and Inference*, vol. 103, no. 455-470, 2002.
- [9] P. M. Hooper, "Iterative Weighted Least Squares Estimation in Heteroscedastic Linear Models," *Journal of the American Statistical Association*, vol. 103, no. 88:421, 2012.
- [10] G. Gbur, *Mathematical methods for optical physics and engineering*. 2011.
- [11] M. J.J., *The Levenberg-Marquardt algorithm: Implementation and theory*. Lecture Notes in Mathematics, Springer, 1978.
- [12] J. Morio, "Global and local sensitivity analysis methods for a physical system," *European Journal of Physics*, vol. 32, no. 6, 2011.