TIME SERIES MODELLING AND FORECASTING OF MONTHLY TEMPERATURE DATA FOR WESTERN KENYA USING SEASONAL ARIMA METHOD

ALEX BARASA KAKAI

A research project submitted in partial fulfilment of the requirements for Postgraduate Diploma in Applied Statistics of Jomo Kenyatta University of Agriculture and Technology.

2022

Declaration

This project is my original work and has not been presented elsewhere for a degree award.

Signature: .... Date: ....

Alex Barasa Kakai (SCB384-C003-4369/2014).

Declaration by supervisor

This project has been submitted for examination with my approval as the supervisor.

1. Dr. Joseph Mung'atu, Statistics and Actuarial Science Department.

   Signature: .... Date: ....

2. Dr. Herbert Imboga, Statistics and Actuarial Science Department.

   Signature: .... Date: ....

Acknowledgment

I thank the Almighty God for His continued protection and good health. I thank my lecturers and classmates for their support during the entire course, most importantly my supervisors, for their valuable guidance and helpful support during the preparation of this project. Lastly, I thank my family for the love,support and encouragements they gave me during the course.

Contents

List of Figures

List of Tables

Abbreviations and Acronyms

MA	Moving Average
AR	Auto Regressive
ARMA	Auto Regressive Moving Averag
ARIMA	Auto Regressive Integrated Moving Average
SARIMA	Seasonal Auto Regressive Integrated Moving Average
ARCH	Autoregressive Conditional Heteroscedasticity
GARCH	Generalized Auto-Regressive Conditional Heteroscedasticity
PACF	Partial Auto-Correlation Coefficient
ACF	Auto-Correlation Coefficient
AIC	Akaike Information Criterion
IPCC	Intergovernmental Panel on Climate Change
KNBS	Kenya National Bureau Of Statistics
KMD	Kenya Meteorological Department
LDC	Least Developed Countries
GDP	Gross Domestic Product
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MSE	Mean Square Error
MAPE	Mean Absolute Percentage Error

Abstract

Climate change marked by global warming has occurred. Earth's average temperature has gradually increased in the last decades.This has adversely affected farming patterns in Kenya where 60% of households depend on Agriculture for their livelihoods. Time series analysis and forecasting is a versatile and efficient tool with diverse applications, its numerous applications in meteorology and other environmental areas has helped in the understanding of phenomena like rainfall, humidity, temperature, drought etc. To help address this problem, this research project has modelled seasonality of both the average monthly minimum and maximum temperature recordings over a thirty-year period in two agricultural intensive regions in Kenya. Time series SARIMA models were built to help analyse and forecast the maximum and minimum temperature of Kakamega and Eldoret stations. Appropriate orders of the SARIMA models were picked after evaluation of results from ACF and PACF plots and the AIC criterion. The best forecasting SARIMA models in Kakamega Station were SARIMA(2,0,0)$(0,1,1)_{12}$ for the minimum temperature and SARIMA(1,0,0)$(0,1,1)_{12}$ for maximum temperature while for Eldoret Station SARIMA(1,0,1)$(0,1,1)_{12}$ for the minimum temperature and SARIMA(1,0,0)$(0,1,1)_{12}$ for maximum temperature were selected. The results indicate that the minimum temperature is gradually increasing over the years supporting the fact that indeed global warming is happening.

Chapter 1

Introduction

1.1 Background Study

Currently, climate change marked by global warming has occurred (?, ?). Continued analysis has shown that the earth's temperature has increased globally by approximately 0.72% over the past 60 years, and the last three decades have been the worst affected compared to any other since the late 19th century (?, ?). Global temperatures have increased by between 0.4^0C and $0.8\ ^0C$ in the past century and according to ? (?) these temperatures could rise by between $1.4\ ^0C$ and $5.8\ ^0C$ by the end of the 21^{st} century. In Kenya, observed mean annual temperatures have risen by an estimated 1 ^{0}C since 1960 (?, ?). It has been estimated and predicted that this rise will continuously occurs at a rate of $0.2\ ^0C$ per decade. This will continue rise to the level of $2.5\ ^0C$ by the year 2050 and to $4\ ^0C$ by 2100.

Kenya's recent National Climate Change Response Strategy (?, ?) and the National Climate Change Implementation Framework (?, ?) depict that global warming has acquired the status of a key national policy challenge. Kenya was among the first non-LDC countries in Africa to develop government plans for responses to climate change across key economic sectors. The National Climate Change Action Plan(?, ?) was launched in March, 2013. Agriculture is an essential part of both these two frameworks, thus reflecting reliance on agriculture for the national economy and for providing labour locally.The Kenyan government derives more than 45% of its revenue from the agricultural sector which is the largest employer in the economy accounting for about 60% of the country's employment.

The agricultural sector comprises of more than 24% of Kenya's GDP and 19% of the formal wage employment. More than half of Kenyans live in rural areas depend directly or indirectly on agriculture for their livelihoods. An estimated 60% of all households are engaged in farming activities (?, ?). Thus, agriculture is significantly impacted by heat and radiation resources, its productivity may be negatively affected by global warming, especially with changes in precipitation, temperature, and solar radiation (?, ?). Global warming, therefore, is the greatest humanitarian crisis of our time, responsible for rising seas, raging storms, searing heat, ferocious fires, severe drought, and punishing floods (?, ?). It puts at risk our health, communities, economy, and national security. The primary cause of this phenomenon is the release of greenhouse gases by burning of fossil fuels, land cleaning, agriculture, among others, leading to the increase of the so-called greenhouse effect.

This study aims to improve our understanding of the changes in temperature in Kenya, and thus could help decision-makers develop better strategy and measures for adaptation and mitigation of global warming. Different techniques can be applied to evaluate the global warming dynamics, with the best technique being time series analysis of temperature recordings within Kenya.

This analysis allows one to make better predictions increasing our understanding of the phenomenon. Temperature time series data from different sites within Kenya was used. Initially, the approach is verified considering known parts of the time series and afterward, results are extrapolated for future values estimating temperature.

1.2 Statement of the Problem

The emerging dominant narratives in Kenya are focused on the need to protect food security and agricultural resources from the negative impacts of global warming (?, ?), and on the other hand, possible opportunities for capitalizing on carbon funding (?, ?).

Kenya has been ahead of many countries in developing a national climate change strategy, and agriculture is one of the key critical sectors of interest. Various researchers have attempted to come up with the best model for modelling and describing global warming with aim of forecasting.

Little has been done in Kenya to capture the characteristics of global warming and model seasonality with very few research done on this topic. Thus, this research attempts to accomplish this goal through modelling of seasonality in monthly minimum and maximum temperature recordings collected in various stations around Kenya.

1.3 Justification

Seasonal models are an important factor in analysis of data with seasonal trend. Hence this research is justified as its findings will be of interest and relevance to long term planning of various sectors as the results and findings will help government agencies, organizations and analysts to gain in-depth understanding of the dynamics of global warming and climate change.

1.4 Objectives

1.4.1 General Objective

To model and forecast Kenya's monthly minimum and maximum temperature recordings over a certain period of time.

1.4.2 Specific Objectives

1. To analyse variations in the monthly minimum and maximum temperatures.
2. To fit a Seasonal ARIMA model for monthly minimum and maximum temperatures.
3. To carry out short-term forecasts to predict temperature trends using the best fitted model.

1.5 Hypothesis

1. H_0: There is no significant dependence and correlation in the residuals of Temperature Recordings data in Kenya.
2. H_0: The residuals from Temperature Recordings data have zero mean and constant variance.

1.6 Scope of the Study

Kenya Meteorological Department (KMD) has more than twenty four sites distributed across the country. These sites are used to collect various climatic elements including rainfall, temperature, wind, humidity and sunshine. This research project uses secondary data of maximum and minimum temperature from these sites . The data is for a duration of 31 years from 1980 to 2011.

1.7 Limitations

Accessibility of temperature data in Kenya is almost impossible, available data comes with a high price where KMD sells the data. Such data should be available freely both in their offices and on-line for public use. Also data available had many instances for missing data points. This dramatically shrank the sample size. Out of 24 stations, only 11 had all data points from 1980 to 2011.

Chapter 2

Literature review

2.1 Theoretical Review

Time series analysis and forecasting has become a major tool in various meteorological applications to study trends, seasonality and variations in variables like rainfall, humidity, temperature and many other environmental parameters. A number of classical time series studies have been conducted in recent years to assess the nature of climate change, as it has occurred over the world and as it will more in likely do so in the future.

In their early studies ? (?) generalized the ARIMA model to deal with seasonality by forming a multiplicative seasonal ARIMA model (SARIMA).This model is always referred to as the Box Jenkins Methodology. Even though the model is more complicated with both the seasonal and non-seasonal autoregressive components having their PACF and ACF cutting at the seasonal and non-seasonal lags, it allows for randomness in the seasonal pattern from one cycle to the next.

2.2 Empirical Review

A number of studies have been done around the world where the Box Jenkins Methodology has been used. These studies applied seasonal ARIMA model to aid in the analysis and forecasting of time series element.

While modelling and predicting seasonal influenza transmission in warm regions using climatological parameters ? (?), suggested that inclusion of any seasonal component in any data it would be best to use SARIMA methodology to model the given data as this increases the prediction capability.

In their paper: Time Series Analysis and Forecasting of Temperatures in the Sylhet Division of Bangladesh, ? (?) acknowledged that even though there are various methods available for monitoring and evaluating temperature time series, Box-Jenkins ARIMA model stood out as the most effective and popular approach. They employed the Seasonal ARIMA model in modelling seasonality on the temperature recordings recorded at two stations in Sylhet Division between 1977 and 2011. After inspection of the ACF, PACF autocorrelation plots, the most appropriate orders of the ARIMA models were determined and evaluated using the AIC-criterion. For the maximum and minimum temperatures at Sylhet station SARIMA $(1,1,1)(1,1,1)^{12}$ and SARIMA $(1,1,1)(0,1,1)^{12}$, models were selected. They concluded that the SARIMA models if applied could help decision makers to establish better strategies and set up priorities for arming themselves against upcoming weather changes.

? (?) published a paper entitled: Climate Change in Bangladesh: A Historical Analysis of Temperature and Rainfall Data. Here they modelled seasonality in the both the temperature and rainfall recordings in Bangladesh for the period of 1976-2008. They observed that the trend of variation of yearly average maximum temperature has been found to be increasing at a rate of 0.0186 ^{o}C per year whereas the rate was 0.0152 ^{o}C per year for yearly average minimum temperature.

? (?) applied the seasonal ARIMA model to the time series precipitation data from 1961 to 2011, for forecasting monthly precipitation in Yantai, China. They found that the model

SARIMA $(1,0,1)(0,1,1)^{12}$ fitted the past data and could be used successfully for forecasting. Based on this model they predicated that the precipitation in the next three years for the region will decrease.

? (?) forecasted mean temperature by using SARIMA model in Ashanti region of Ghana by analysing past temperature data from 1980 to 2013. The conclusion was that the best model for forecasting was SARIMA $(2,1,1)(1,1,2)^{12}$ as the model recorded the least BIC values. For the forecasts, the ME, RMSE, MAE, MPE, MAPE, MASE values were evaluated.

In the paper: Modelling of Wholesale Prices for Selected Vegetables Using Time Series Models in Kenya, ? (?) incorporated SARIMA model to help exhibit seasonality periodic in-fluctuations that recur with about the same intensity each year. To avoid fitting an over parameterized model, Akaike Information criterion (AIC) was employed in the selection of the best model.

? (?) modelled time series SARIMA model and used it to analyse and forecast the maximum and minimum air temperatures of Nairobi City from 1985 to 2014. Based on the results of the ACF and PACF, SARIMA $(0,0,2)(0,1,1)^{12}$ for maximum temperature and SARIMA $(1,0,0)(0,1,1)^{12}$ for minimum temperature models were picked. On general scale it was realized that the minimum temperature was gradually increasing over the years supporting the fact that global warming was real.

2.3 Research Gaps

Various studies have been done in relation to temperature time series, some have used the linear regression model while others have used Seasonal ARIMA model. In Kenya, all the studies have only used data for a specific station and very few have incorporated more than one station in a single study. This study involved two stations Kakamega and Eldoret which are agricultural intensive regions.

Chapter 3

Methodology

3.1 Research Design

The research design employed in this study is the *empirical* kind. Empirical research is a combination of observational studies, case studies and experimentation. Some of the distinct features of this research design is that it offers more than simply reporting the observed and in the process combines extensive research with detailed case study. It is widely used in scientific methods because of the strict requirements that all evidence must be empirical.

The design fits perfectly in this particular research because the data is observed and applied to a specific case. Moreover, there are hypotheses of causal relationships that are to be tested in the study. The empirical cycle starts with observation, to induction, deduction, testing of hypothesis and finally ends with evaluation.

3.2 Methods

3.2.1 Autoregressive Moving Average Models

An AR model has lagged terms on the time series itself while the MA model has lagged terms on the noise or residuals. If these two models AR(p) and MA(q) are effectively combined they form the ARMA model. Thus ARMA (p,q), where p is the autoregressive order and q the moving average order is written as:

$$X_t = \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + E_t + \beta_1 E_{t-1} + ... + \beta_q E_{t-p} \tag{3.1}$$

Most importantly the ARMA models can only be used when time series data is stationary.

3.2.2 Autoregressive Integrated Moving Average Models

However, in practice, most of the time series data is always non-stationary therefore ARMA models become inadequate to effectively describe non-stationary time series. ? (?) proposed a new model, a generalized version of the ARMA model, the ARIMA model, to include the case of non- Stationarity. With the new model finite differencing is applied to the time series data to remove non-stationarity.

A non-seasonal ARIMA model is classified as an ARIMA (p, d, q) model, where p is the number of autoregressive terms, d is the number of non-seasonal differences, and q is the number of lagged forecast errors (moving average) in the prediction equation. A process,X_t follows an ARIMA (p, d, q) model if the d^{th} order difference of X_t is an ARMA(p,q). If we introduce

$$Y_t = (1 - B)^d X_t \tag{3.2}$$

where B is the back-shift operator, then we can write the ARIMA process using the characteristics polynomials, i.e. $\theta(.)$ that accounts for the MA part , and $\phi(.)$ that stands for the AR part as

$$\phi(B) Y_t = \theta(B) \omega_t \tag{3.3}$$

replacing the Y_t we obtain

$$\phi(B)(1 - B)^d X_t = \theta(B) \omega_t \tag{3.4}$$

3.2.3 SARIMA (p,d,q)(P,D,Q)s model

SARIMA model is an extension of ARIMA model and it is applied when the series contains both seasonal and non-seasonal behaviour. SARIMA model is sometimes called the multiplicative seasonal autoregressive integrated moving average and is denoted by:

$$SARIMA(p,d,q)(P,D,Q)^s \tag{3.5}$$

Often a time series possess a seasonal component that repeats every S observations. For monthly observations $S = 12$ (12 in 1 year) while for quarterly observations $S = 4$ (4 in 1 year). The Seasonal ARIMA models are defined by seven parameters of the form :

$$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)Z_t \tag{3.6}$$

where B denotes the back-shift operator,$\phi_p, \Phi_P, \theta_q, \Theta_Q$ are polynomials of order p,P,q,Q, respectively, Z_t denotes a purely random process and

$$W_t = \nabla^d \nabla_s^D X_t \tag{3.7}$$

and we write

$$X_t \sim ARIMA(p,d,q) \tag{3.8}$$

SARIMA models are ARIMA(p,d,q) models whose residuals t are ARIMA(P,D,Q). With ARIMA(P,D,Q) we intend ARIMA models whose operators are defined on B^s and successive powers. Concepts of admissible regions SARIMA are analogue for the ARIMA processes, they are just expressed in terms of B^s powers.

For climate data which usually follows a seasonal, i.e. an annual cycle, it is more appropriate to use a seasonal ARIMA (p,d,q)(P,D,Q)^S model (?, ?). To identify a perfect ARIMA model for a particular time series, Box and Jenkins proposed a methodology that consists of four phases, namely: model identification, estimation of parameters, diagnostic checks and forecasting.

3.2.4 Model identification

3.2.4.1 Stationarity

A time series is said to be stationary if it has a constant mean, variance and autocorrelation over time. Stationarity is a necessary and sufficient condition for any ARIMA model to be used for analysis. One can check stationarity by plotting the series and its autocorrelation.

For a non stationary series, the autocorrelation function decays slowly. To obtain stationarity from a first order non-stationarity, we difference the data d times, using $\Delta^d y_t$.

3.2.4.2 Autocorrelation and Partial Autocorrelation Functions (ACF and PACF)

A proper time series model can be identified by analysing the both ACF and PACF. First a time series plot is done then trend and seasonality features are examined. The data is observed over time to check if it shows constant seasonal variance where it can be stabilized through applying appropriate transformations on the original time series values.

If the original time series values are non-stationary and seasonal, more complex differencing transformations are required. The ACF of the transformed data is checked both at non-seasonal and seasonal levels for any indication of seasonality.The seasonal behaviours appear at the exact seasonal lags $L, 2L, 3L$, and $4L$.

In general, the transformed time series values are considered stationary if the ACF shows both of the following behaviours:

1. Cuts off or dies down fairly quickly at the non-seasonal level.

2. Cuts off or dies down fairly quickly at the seasonal level (exact seasonal lags or near seasonal lags).

3.2.4.3 Augmented Dickey Fuller test

Augmented Dickey Fuller (ADF) test is used to test for stationarity in a given model. Its fundamental aim is to test the null hypothesis that the series has a unit root i.e $\phi = 1$ in

$$y_t = \phi y_{t-1} + v_t \tag{3.9}$$

The test has two hypothesis:

$$H_0 = \text{that the series has a unit root}$$
$$H_1 = \text{that the series is stationary}$$

3.2.4.4 Kwiatkowski–Phillips–Schmidt–Shin test (KPSS)

The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test help check whether a time series is stationary around its mean or linear trend, or is non-stationary due to a unit root. The null hypothesis tests that the data is stationary while the alternate tests that the data is not stationary i.e

$$H_0 = Y_t \text{ is trend(or level) stationary}$$
$$H_1 = Y_t \text{ is a unit root}$$

3.2.5 Estimation of the Model Parameters

Identification of a seasonal series is much more difficult hence the need for keenness to be able to successfully identify the model order.If the model contains only AR terms, the unknown coefficients can be estimated using ordinary least squares. However, if the model contains MA terms too, the task becomes more complicated, because the lagged values of the innovations are unobservable.

Consequently, it is not possible to derive explicit expressions to estimate the unknown coefficients and therefore one has to use maximum likelihood for estimation purposes. Let

$X_1, X_2, \ldots, X_n$ be drawn from a Gaussian ARMA (p, q) process with mean zero. Then the likelihood parameters $\phi\epsilon\Re\theta\epsilon\Re\sigma_w^2\epsilon\Re_+$ are defined as the density of $X = (X_1, X_2, \ldots, X_n)$' under the Gaussian model with those parameters:

$$L(\phi, \theta, \sigma_\omega^2) = \frac{1}{(2\pi)^{\frac{n}{2}}|A|^{\frac{1}{2}}} exp(-\frac{1}{2}X'\Gamma_n^{-1}X) \tag{3.10}$$

Where $|A|$ denotes the determinant of a matrix A, and Γ_n is the variance/covariance matrix of X with the given parameter values. The maximum likelihood estimator identifies the values of parameters that increase the probability of obtaining fitted values that are close to the observed values.

3.2.5.1 Model selection criterion

One must keep the number of parameters to a minimum, so the values of p, P, q, Q, d, and D that are selected should be less than or equal to two. Values of p and q are selected primarily by examining the values of ACF(r_k) while Q and P by examining values of ACF at $k = 12, 24, ...$ and seasonal model is given as $s = 12$.

According ? (?) even though ACFs and PACFs help in determining the model's order they only hint on where model building can begin. Therefore final model is chosen by penalty function statistics such as Akaike Information Criterion or Bayesian Information Criterion.

3.2.5.2 Akaike Information Criterion (AIC)

The competing models are ranked according to their values of AIC. The model which attains the lowest value of information criterion is considered to be the best. The AIC evaluates a given model depending on the closeness of its fitted values to the observed values.

It selects the simplest model that best explains the given data with minimum number of parameters and penalizes the complex model for having more model parameters. Penalizing the model with more parameters discourages over fitting. It is given by:

$$AIC = -2\ln(max.likelihood) + 2k = 2k + nlog(\frac{RSS}{n}) = 2k - 2log(L) \tag{3.11}$$

Where:

k= the number of parameters in the statistical model, (p+q+P+Q+1)

L= the maximized value of the likelihood function for the estimated model.

RSS= the residual sum of squares of the estimated model.

n= is the number of observation, or equivalently, the sample size

Thus AIC chooses the model with best fit measured by likelihood function. The AIC is biased for small samples and thus one can opt for the bias-corrected version, denoted by AICc

$$AICc = -2\ln(max.likelihood) + 2rN/(N - r - 1) \tag{3.12}$$

3.2.6 Diagnostic Checking of the Models

Here, each selected model is assessed to determine how well it fits the temperature data. For a model that fits the data well, the standardized residuals estimated from it should be independently and identically distributed with zero mean and constant variance.

This is carried out by studying the autocorrelation plots of the residuals to see if further structure (large correlation values) can be found. If all the autocorrelations and partial autocorrelations are small, the model is considered adequate and forecasts are generated. If some of the autocorrelations are large, the values of p and/or q are adjusted and the model is re-estimated.

3.2.6.1 Box-Ljung Test

This test statistic helps tell the adequacy of the overall model. It is defined as:

H_0: the model does not exhibit lack of fit

H_1: the model exhibit lack of fit

$$Ljung - BoxStatistic(Q^*) = n'(n'+2)\sum_{k=1}^{m}(n'-k)^{-1}r_k^2(\hat{a}) \tag{3.13}$$

$n' = n - d$ where n is the number of observations and d is the degree of non-seasonal differencing used to transform the original time series values into stationary while m is the number of lags being tested. Also $r_k^2(\hat{a})$ is the square of the autocorrelation of the residuals at lag k.

3.2.7 Forecasting

Forecasting is a very essential tool in terms decision making therefore any model to be selected has to generate accurate forecasts. However, most models identified rarely give the best forecasting therefore various test statistics such as MAE, MSE and MAPE are used to confirm the forecasting accuracy of the model. Forecasting an ARMA process with mean μ_x, m-step-ahead forecasts can be defined as

$$\tilde{x}_{n+m} = \mu_x + \sum_{j=m}^{\infty}\psi_j\omega_{n+m-j} \tag{3.14}$$

3.2.7.1 Mean Absolute Error (MAE)

Mean Absolute Error is a model evaluation metric applied mostly on regression models. MAE measures the average magnitude of the errors in a set of predictions, without considering their direction.

To get the mean absolute error of a model with respect to a test set, we sum and divide the absolute values of the individual prediction errors on over all instances in the test set. Where the prediction error is the difference between the true value and the predicted value of a given instance.

$$MAE = \frac{1}{n}\sum_{i-1}^{n}|y_i - \bar{y}_i| \tag{3.15}$$

3.2.7.2 Mean Squared Error(MSE)

Just like MAE, MSE is also a model evaluation metric. The mean squared error of a model with respect to a test set is the mean of the squared prediction errors over all instances in the test set. It represents the average of the squared difference between the original and predicted values in the data set.

$$MSE = \frac{1}{n}\sum_{i-1}^{n}(y_i - \bar{y}_i)^2 \tag{3.16}$$

3.2.7.3 Mean Absolute Percentage Error(MAPE)

The mean absolute percentage error (MAPE) is the average of the absolute percentage errors of the forecasts. The errors are obtained by subtracting the forecasted value from the actual/observed value. The smaller the value of MAPE the better the forecast.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}|\frac{A_t - F_t}{A_t}| \tag{3.17}$$

Where:

M = mean absolute percentage error

n = number of times the summation iteration happens

A_t = actual value

F_t = forecast value

3.3 Data Analysis

3.3.1 Population

The population appropriate for this study is monthly temperature recordings for a period of at least thirty years. This may include either the minimum and maximum temperature recordings or the difference of the two from the different stations around the country.

3.3.2 Data Collection

For this study secondary data from The Kenya Meteorological Department which is the main body mandated to collect and distribute climatic data in Kenya was used. The data incorporated a thirty year duration of minimum and maximum monthly temperature recordings collected from all the stations throughout Kenya.

3.3.3 Data Preparation

The data obtained from The Meteorological Department of Kenya was already in an excel worksheet. The data is arranged in tables format for all the months. Therefore, the data was changed into a single column format in a comma-delimited(CSV) format for easier to exporting to R Programming software for further analysis.

Chapter 4

Results and Discussions

In this chapter a more in-depth analysis of the data is done. The chapter is organised in four sections with respect to the objectives of the study as follows:

(i) Data exploration using the time plot to check and uncover insights to help identify areas of interest or patterns in the data

(ii) Model identification and selection for the monthly minimum and maximum temperatures for the two stations

(iii) Properties of the temperature models identified in (ii) above

(iv) Forecasting using the models selected in (ii) above

R Programming was used for the statistical analysis

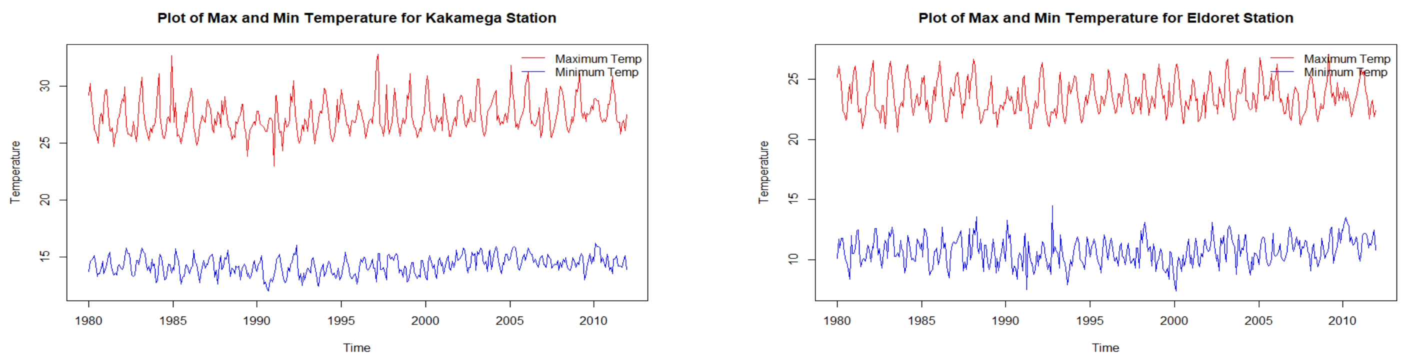
4.1 Data Exploration

4.1.1 Time Plot for Maximum and Minimum Temperature

Figure 4.1 shows a time plot of both maximum and minimum temperature of both Kakamega and Eldoret Stations. It can be observed that the maximum temperatures of

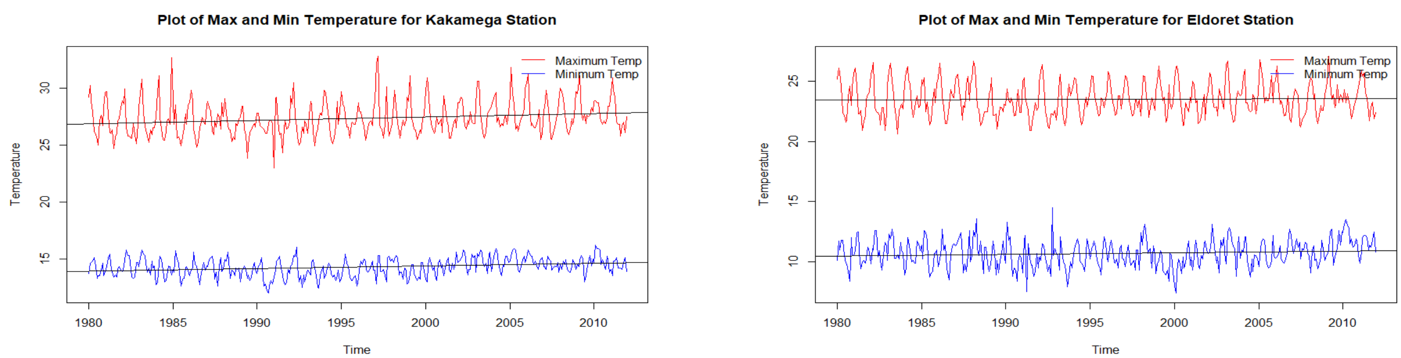
Kakamega station is centred around a mean of 26^0C with possible outlier cases while that of Eldoret is centred at 24^0C and rather smooth.

Both lack visible trend when the trend-line is introduced to the time plots. For the minimum temperature, Kakamega is centred around a mean of 14^0C while for Eldoret is centred around a mean of 10^0C .The first series (maximum temperature) does appear to be stationary in mean and variance, as the level of the series stays roughly constant over time. However, the second series (minimum temperature) does show a slight trend hence a sign of being non-stationary.



(a) Kakamega Station (b) Eldoret Station

Figure 4.1: Time Plot for Maximum and Minimum Temperature

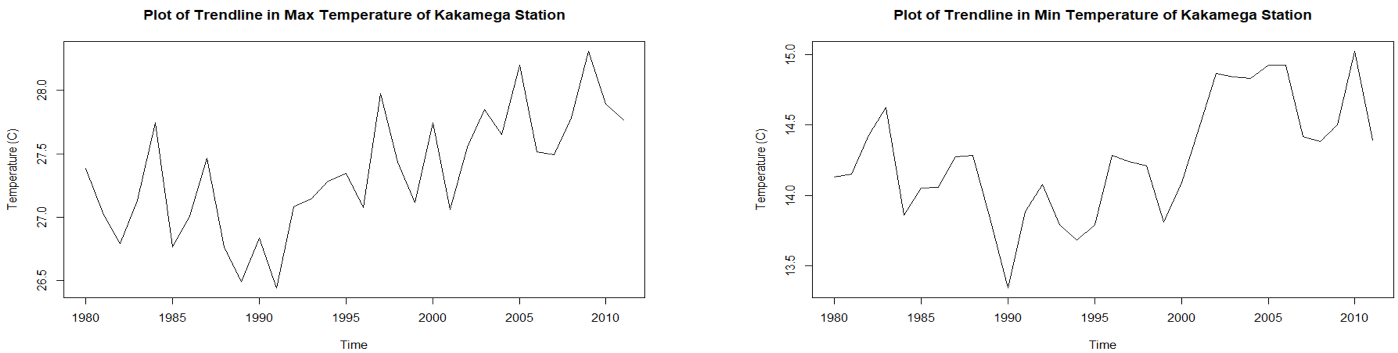


(a) Kakamega Station (b) Eldoret Station

Figure 4.2: Time plot of Maximum and Minimum Temperature with a trend line

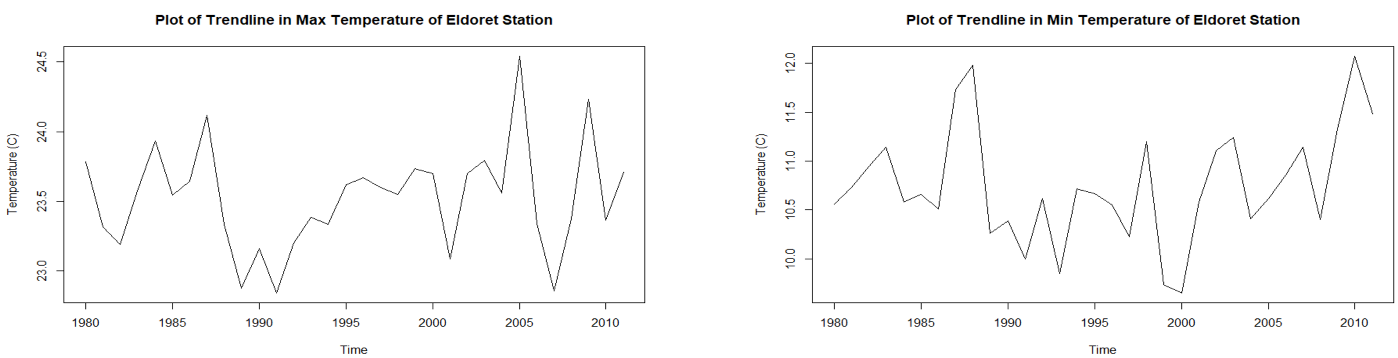
Figure 4.2 shows a trend line introduced in the time plot to try and depict the trend. The trend cannot be clearly depicted from these plots hence a clear plot of the trends was done.

From the trend plots there is a slight upward trend from the year 1990 for both the maximum and minimum temperatures of Kakemega while for Eldoret, the minimum temperature has a slight rise from the year 2000 as observed in Figures 4.3 and 4.4.



(a) Maximum Temperature (b) Minimum Temperature

Figure 4.3: Trend Plot of the Maximum and Minimum Temperatures of Kakamega



(a) Maximum Temperature (b) Minimum Temperature

Figure 4.4: Trend Plot of the Maximum and Minimum Temperatures of Eldoret

Figures 4.5 and 4.6 show box plots of the maximum and minimum temperature data across the months for the two stations giving us a sense of the seasonal effect distribution. Both the stations have got outlier data points showing that for the thirty year period, we have had months where the stations experienced extreme conditions.

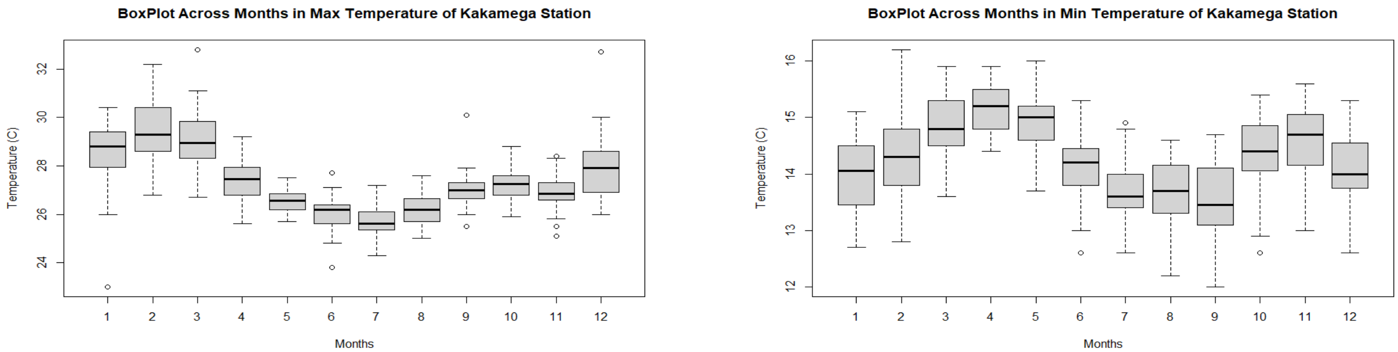


Figure 4.5: Box Plots of Seasonal Trend Across Months for both Maximum and Minimum Temperatures of Kakamega

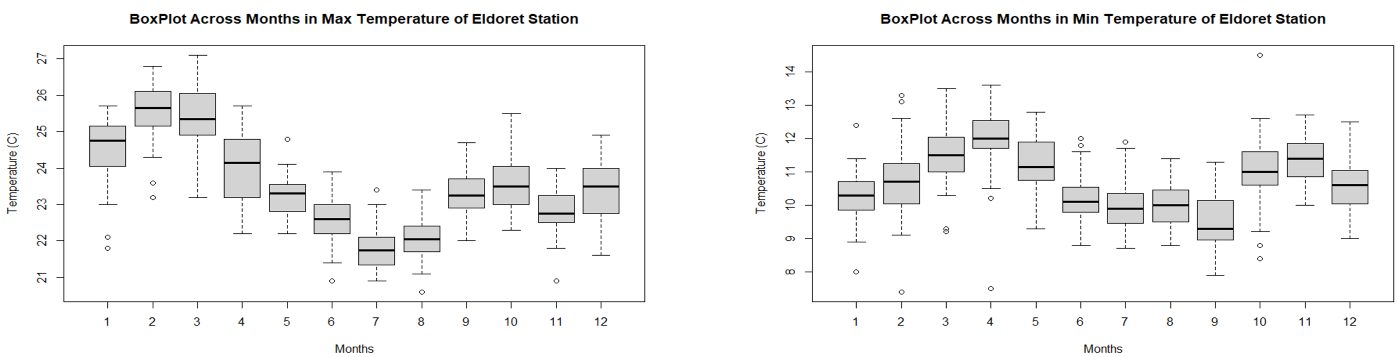


Figure 4.6: Box Plots of Seasonal Trend Across Months for both Maximum and Minimum Temperatures of Eldoret

4.1.2 Descriptive Statistics

	Min	Max	Mean	Median	SE	SD	Variance	Range	Sen's slope
Max	23	32.8	27.35	27	0.08	1.49	2.21	9.8	0.002631579
Min	12	16.2	14.26	14.3	0.04	0.8	0.64	4.2	0.001851852

Table 4.1: Summary Statistics of Maximum and Minimum Temperatures of Kakamega

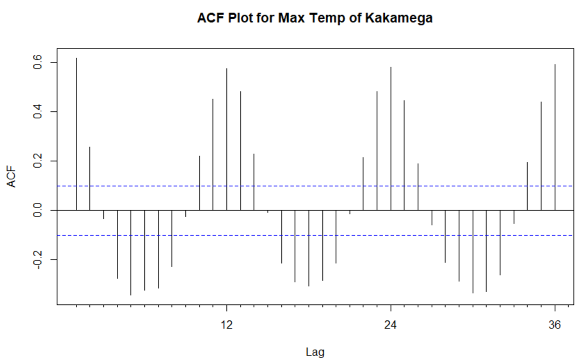
Summary of descriptive statistics for both maximum and minimum temperature recordings (January, 1980 - December, 2011) for the two Stations are provided in Tables 4.1 and 4.2. From the Sen's Slope it's observed that for Kakamega station, both the minimum and maximum temperature have been increasing on averagely by 0.0019 ^{0} C and 0.0026 ^{0} C

	Min	Max	Median	Mean	SE Mean	SD	Variance	Range	Sen's slope
Max	20.6	27.1	23.3	23.52	0.07	1.36	1.84	6.5	0.0004878049
Min	7.4	14.5	10.7	10.7	0.06	1.13	1.27	7.1	0.0012738853

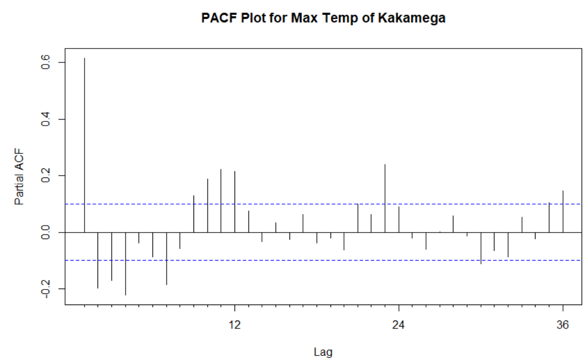
Table 4.2: Summary Statistics of Maximum and Minimum Temperatures of Eldoret

yearly since 1980 respectively. Likewise for Eldoret station, both the minimum and maximum temperatures have been increasing on average yearly by 0.0013 ^{0} C and 0.0005 ^{0} C respectively.

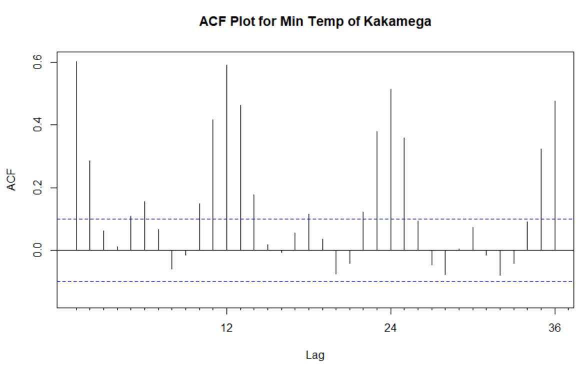
4.1.3 ACF and PACF Plots



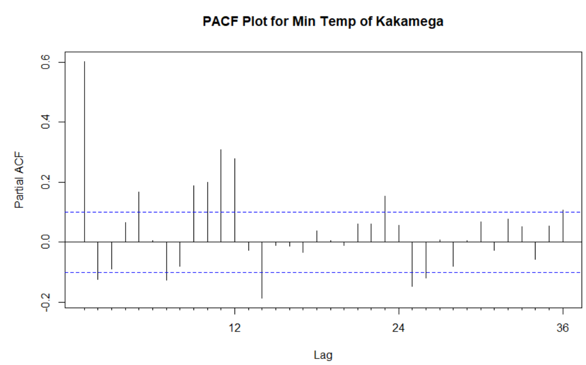
(a) Autocorrelation Function



(b) Partial Autocorrelation Fuction



(c) Autocorrelation Fuction

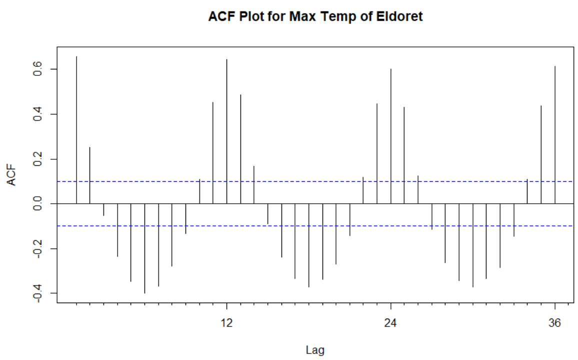


(d) Partial Autocorrelation Fuction

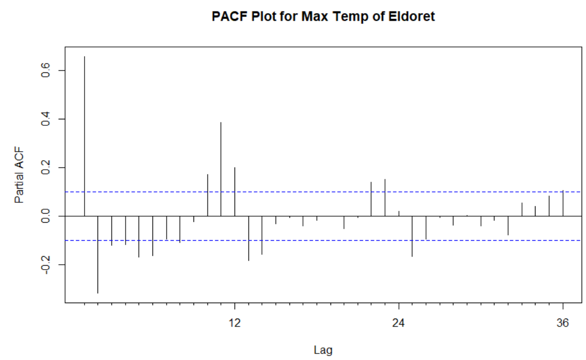
Figure 4.7: Autocorrelation and Partial Autocorrelation functions Plots of both Maximum and Minimum Temperature of Kakamega

Figures 4.7 and 4.8 shows the ACF and PACF for both maximum and minimum temperature of the two stations. From both ACF plots it can be observed that the ACF is cyclic owing to a strong seasonal wave patterns that decline moderately, likewise the envelope decays very slowly. This confirms the presence of seasonality behaviour and thus the time series is non stationary.

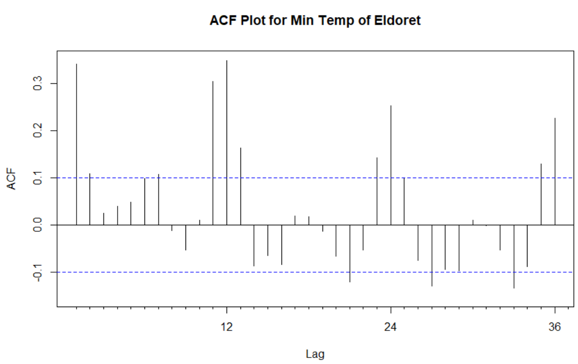
Several tests of stationarity are thus performed to verify the visual observation, p-values of 0.4368 and 0.4627 for Kakamega and 0.4037 and 0.3488 for Eldoret were obtained when the Augmented Dickey-Fuller Test is performed on the data of the two stations. Both the two values are greater than 0.05 hence we fail to reject null hypothesis that the series has a unit root hence supporting earlier observation that the datasets are not stationary.



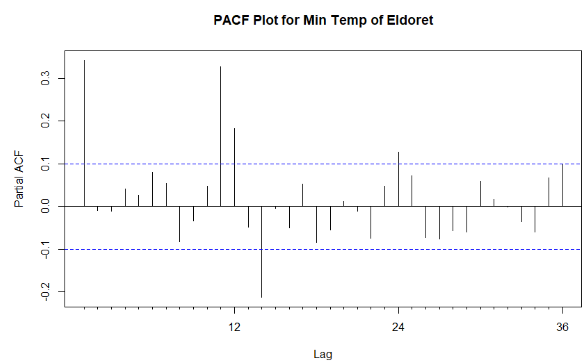
(a) Autocorrelation Function



(b) Partial Autocorrelation Function



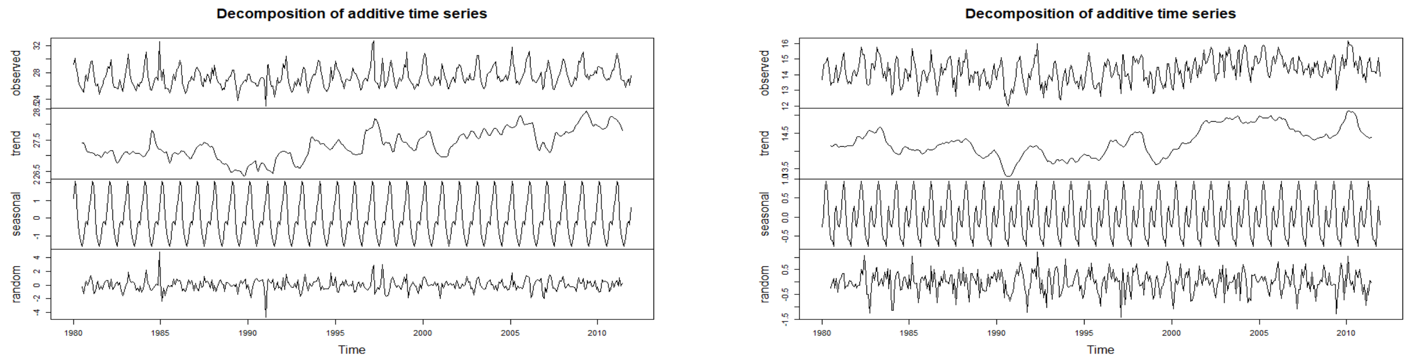
(c) Autocorrelation Function



(d) Partial Autocorrelation Function

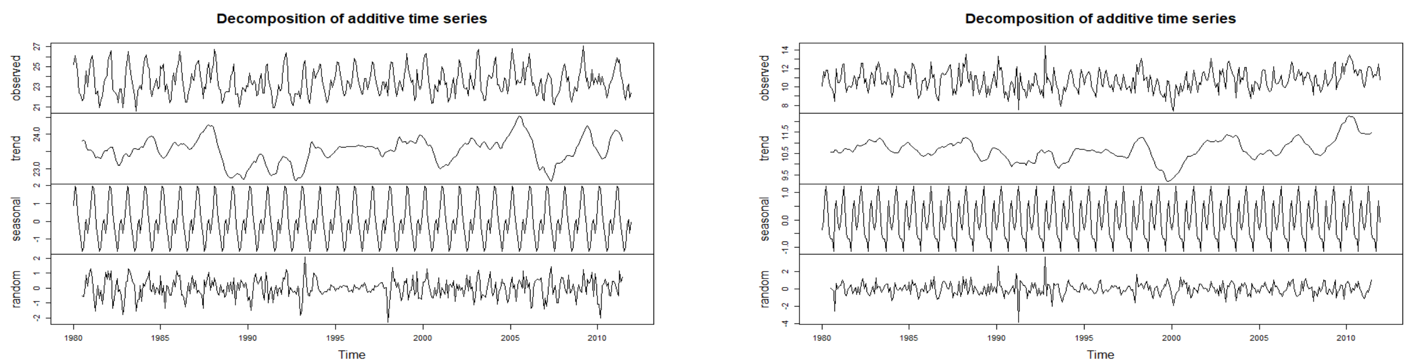
Figure 4.8: Autocorrelation and Partial Autocorrelation functions Plots of both Maximum and Minimum Temperature of Eldoret

4.1.4 Time Series Decomposition



(a) Maximum Temperature of Kakamega (b) Minimum Temperature of Kakamega

Figure 4.9: Decomposed Maximum and Minimum Temperatures of Kakamega



(a) Maximum Temperature of Eldoret (b) Minimum Temperature of Eldoret

Figure 4.10: Decomposed Maximum and Minimum Temperatures of Eldoret

Figure 4.9 and 4.10 show both Maximum and Minimum Temperature decomposed into the various components. It can be observed that there is a systematic upward trend on both the minimum and maximum temperature of Kakamega Station as compared to only the minimum temperature for Eldoret station.

4.1.5 Differencing

Having observed the seasonal pattern in both the two data sets and an additional midway upward trend for Kakamega station, appropriate numbers of seasonal and first difference are required. Its observed that both maximum and minimum temperatures require a seasonal difference due to the presence of the seasonal pattern, however, we might require an additional first difference due to the presence of what seems like an upward trend midway through the data.

After differencing, a p-value of less than 0.01 was achieved when both the Augmented Dickey-Fuller Test and KPSS test were performed, which is below 0.05 showing stationarity had been achieved. Figure 4.11 and 4.12 show time plots of differenced data while Figure 4.13 and 4.14 is the ACF and PACF plots of the differenced minimum and maximum temperature.

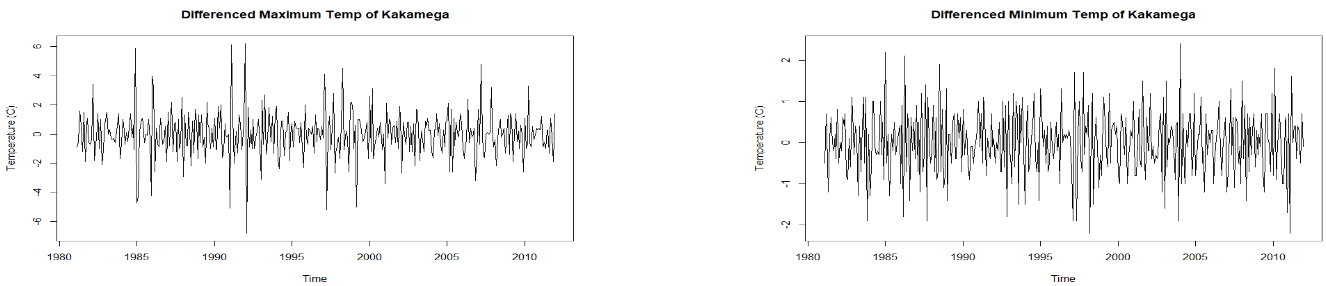


Figure 4.11: Differenced Max and Min Kakamega Temperature Data

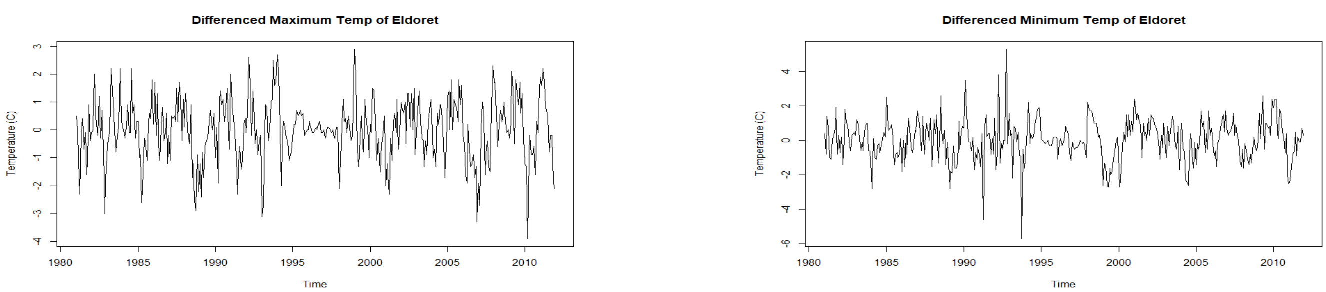
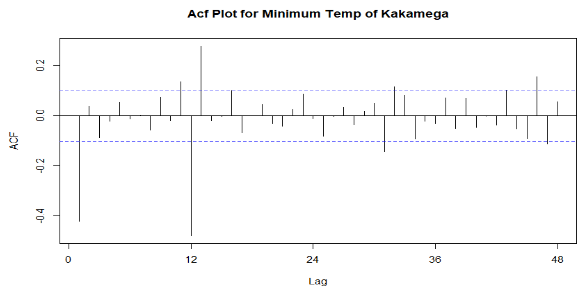
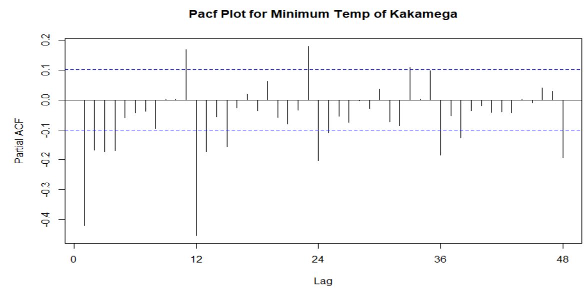


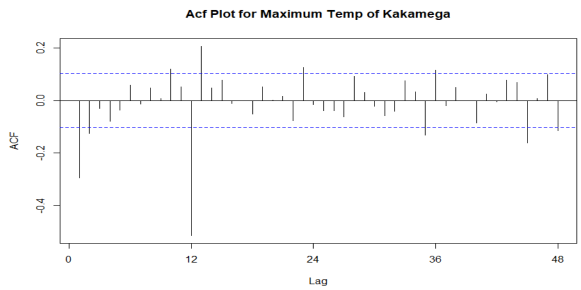
Figure 4.12: Differenced Max and Min Eldoret Temperature Data



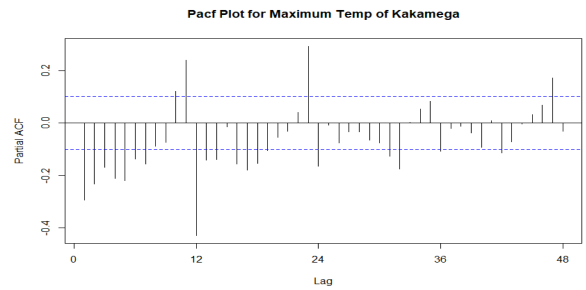
(a) Acf of Diff Min Temp of Kakamega Station



(b) Pacf of Diff Min Temp of Kakamega Station

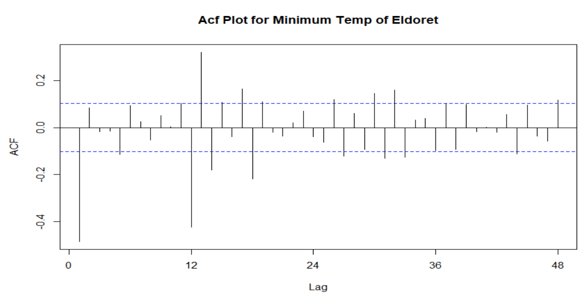


(c) Acf of Diff Max Temp of Kakamega Station

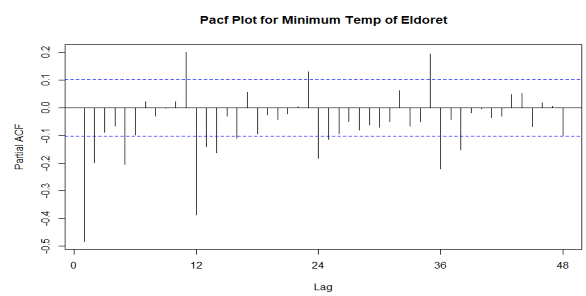


(d) Pacf of Diff Max Temp of Kakamega Station

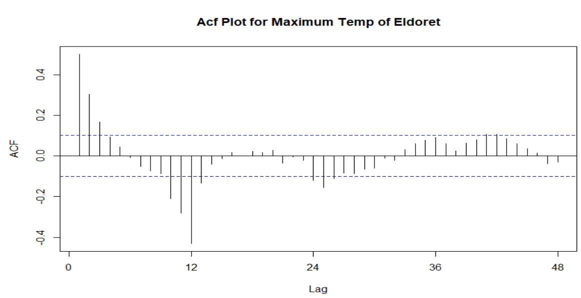
Figure 4.13: Diferenced Minimum and Maximum Temperature for Kakamega Station



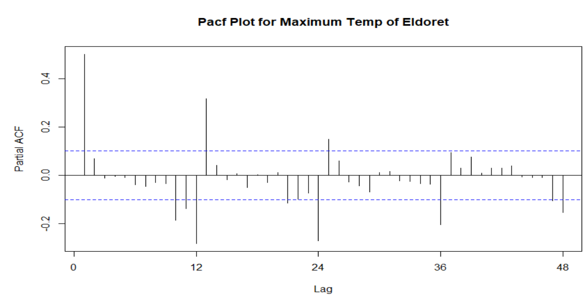
(a) Acf of Diff Min Temp of Eldoret Station



(b) Pacf of Diff Min Temp of Eldoret Station



(c) Acf of Diff Max Temp of Eldoret Station



(d) Pacf of Diff Max Temp of Eldoret Station

Figure 4.14: Differenced Minimum and Maximum Temperature for Eldoret Station

4.2 Models for Monthly Minimum and Maximum Temperature

The ACF and PACF plots are used to identify the best model through studying the dependency structure. These two help determine the values p, q, P and Q. An observation of the coefficients cutting off at low lags indicate direct, short-term dependency and determine orders p and q while coefficients/cutting off at multiples of the period s, imply seasonal dependency and determine P and Q.

Spikes of the ACF at low lags identify the value of q while spikes at low lags of the PACF identify the value of p. Likewise, spikes at lags that are multiples of s observed from the ACF give the value of Q while for PACF give the value of P.

4.2.1 Model Identification for Kakamega Station

From Figure 4.13 (a),(b),(c) and (d) we observe that the autocorrelation plot of minimum temperature at lag $1, 11, 12$ and maximum temperature at lag $1, 2, 10, 12, 13$ exceeds the significance bounds between lags 1-20. Also the partial correlogram shows that the partial autocorrelations at lags $1, 2, 3, 4, 11, 12, 13, 15$ for minimum temperature and at lag $1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 16, 17, 18$ exceed the significance bounds between lags 1-20. It appears that either:

(a) for the minimum temperature the ACF displays a sharp cut-off at lag 2 with a positive autocorrelation that is slightly higher than -0.4. We know that if the lag 1 ACF falls below -0.5, then the series might be over-differenced. Positive spikes in ACF have become negative, another sign of possible over differencing.

    Likewise, the PACF seems to die off rather slowly but irregularly signifying presence of an AR term of order 2. This model might be suffering from slight over-differencing which can be compensated by dropping the 1st order difference.

Both the ACF and PACF have lags at seasonal points exceeding the significance bounds so either:

(i) the ACF cuts off after $1s$ while the PACF also cuts off after lag $1s$ signifying presence of MA term of order 1 and AR of order 1

(ii) the ACF cuts off after lag $1s$ while the PACF is tailing off in the seasonal lags signifying presence of an MA of order 1.

(b) for the maximum temperature the ACF seems to tail off with a negative autocorrelation. Just like the minimum temperature, positive spikes in ACF have become negative, a possible sign of over differencing.

The PACF also is negative and dies off slowly signifying an AR term of order 1. Both the ACF and PACF have lags at seasonal points exceeding the significance bounds so either:

(i) the ACF cuts off after lag $1s$ while the PACF is tailing off in the seasonal lags signifying an MA of order 1

(ii) the ACF cuts off after lag $1s$ while the PACF also cuts off at lag $3s$ signifying an MA of order 1 and AR of order 2.

Therefore, the following SARIMA models are possible for the minimum and maximum temperature time series:

$$SARIMA(2,1,0)(1,1,1)_{12}$$
$$SARIMA(2,1,0)(0,1,1)_{12}$$
$$SARIMA(2,0,0)(1,1,1)_{12}$$
$$SARIMA(2,0,0)(0,1,1)_{12}$$

Table 4.3: Minimum Temperature

$$SARIMA(1,1,0)(0,1,1)_{12}$$
$$SARIMA(1,1,0)(2,1,1)_{12}$$
$$SARIMA(1,0,0)(0,1,1)_{12}$$
$$SARIMA(1,0,0)(2,1,1)_{12}$$

Table 4.4: Maximum Temperature

4.2.2 Model Identification for Eldoret Station

From Figure 4.14 (a),(b),(c) and (d) we observe significant bounds at lag $1, 2, 3, 5, 10, 11, 12, 14, 15$ and lag $1, 2, 11, 12, 13, 18$ for the ACF and PACF respectively of the minimum temperature for first 20 lags. For the maximum temperature we observe significant bounds at lag $1, 2, 3, 10, 11, 12, 13$ and lag $1, 10, 11, 12, 13$ for the ACF and PACF respectively for first 20 lags. It appears that:

(a) for the minimum temperature positive spikes in ACF have become negative, sign of possible over differencing.The ACF cuts off after lag 1 while the PACF tails off signifying presence of an MA term of order 1, we can also introduce an AR of order 1 to counter over-differencing. Both the ACF and PACF have lags at seasonal points exceeding the significance bounds so either:

 (i) the ACF cuts off after $1s$ while the PACF tails off in the seasonal lags signifying presence of MA term of order 1.

 (ii) the ACF cuts off after lag $1s$ while the PACF cuts off signifying presence of an MA of order 1 and AR of order 1

(b) for the maximum temperature the ACF seems to die off rather slowly with a positive autocorrelation while the PACF also is positive and cuts off sharply at lag 1 signifying an AR term of order 1. Both the ACF and PACF have lags at seasonal points exceeding the significance bounds so either:

 (i) the ACF cuts off after lag $1s$ or lag $2s$ while the PACF is tailing off in the seasonal lags signifying an MA of order 1 or 2

 (ii) the ACF cuts off after lag $1s$ or $2s$ while the PACF also cuts off at lag $1s$ signifying an MA of order 1 or 2 and AR of order 1.

Therefore, the following SARIMA models are possible for both the minimum and maximum temperature time series of Eldoret station.

Table 4.5	Table 4.6
$SARIMA(1,1,1)(1,1,1)_{12}$	$SARIMA(1,0,0)(0,1,1)_{12}$
$SARIMA(1,0,1)(1,1,1)_{12}$	$SARIMA(1,0,0)(0,1,2)_{12}$
$SARIMA(1,1,1)(0,1,1)_{12}$	$SARIMA(1,0,0)(1,1,1)_{12}$
$SARIMA(1,0,1)(0,1,1)_{12}$	$SARIMA(1,0,0)(1,1,2)_{12}$

Table 4.5: Minimum Temperature

Table 4.6: Maximum Temperature

4.3 Properties of the minimum and Maximum Temperature Models Selected

4.3.1 Statistics for tentative models

Both AIC and BIC methods were used to the assess the best model to fit. Both the methods penalize depending on the number of estimated free parameters in a model to combat over-fitting and under-fitting. Whereas AIC is less tolerant to free parameters therefore it might lead over-fitting, BIC is more tolerant to the free parameter which might also lead to under-fitting. Therefore both the methods were used to fit the best model.

4.3.1.1 Statistics for tentative models in Kakamega Station

As observed from the tables 4.7 and 4.8, $SARIMA(2,0,0)(0,1,1)_{12}$ has an AIC of 561.06 and a BIC of 580.66 for the minimum temperature while $SARIMA(1,0,0)(0,1,1)_{12}$ has an AIC of 1005.17 and BIC of 1016.93 . Both have the lowest values in the two tables hence they are selected as the best fitted models.

Models	AIC	AICc	BIC
SARIMA(2,1,0)(1,1,1)$_{12}$	600.11	600.27	619.69
SARIMA(2,1,0)(0,1,1)$_{12}$	598.31	598.42	613.97
SARIMA(2,0,0)(1,1,1)$_{12}$	561.06	561.23	580.66
SARIMA(2,0,0)(0,1,1)$_{12}$	559.67	559.78	575.35

Table 4.7: Statistics for tentative models of the Minimum Temperature

From table 4.10 on Box-Ljung test, p-values of 0.2238 and 0.1216 for residuals of the minimum and maximum temperature respectively were obtained. The values are greater than 0.05 meaning

Models	AIC	BIC	AICc
SARIMA(1,1,0)(0,1,1)$_{12}$	1100.8	1100.87	1112.55
SARIMA(1,1,0)(2,1,1)$_{12}$	1101.39	1101.55	1120.97
SARIMA(1,0,0)(0,1,1)$_{12}$	1005.17	1005.24	1016.93
SARIMA(1,0,0)(2,1,1)$_{12}$	1006.03	1006.19	1025.62

Table 4.8: Statistics for tentative models of the Maximum Temperature

SARIMA(2,0,0)(0,1,1)$_{12}$

Parameter	Estimate	Standard Error
ar1	0.4279	0.0523
ar2	0.2135	0.0510
sma1	-0.8839	0.0374

SARIMA(1,0,0)(0,1,1)$_{12}$

Parameter	Estimate	Standard Error
ar1	0.3604	0.0504
sma1	-0.9081	0.0323

Table 4.9: Parameter Estimation

the residuals are independent and therefore we fail to reject the null hypothesis: H_0, that our model does not show lack of fit and conclude that the data values are independent hence the model is adequate for forecasting.

Box-Ljung Test

	X-squared	df	P Value
SARIMA(2,0,0)(0,1,1)$_{12}$	25.559	21	0.2238
SARIMA(1,0,0)(0,1,1)$_{12}$	29.873	22	0.1216

Table 4.10: Box-Ljung test

4.3.1.2 Statistics for tentative models in Eldoret Station

Likewise, from the tables 4.11 and 4.8, SARIMA(1,0,1)(0,1,1)$_{12}$ has an AIC of 926.3 and a BIC of 941.98 for the minimum temperature while SARIMA(1,0,0)(0,1,1)$_{12}$ has an AIC of 808.87 and BIC of 808.94 for the maximum temperature. These are the models with the lowest values therefore they are selected as the best fitted models.

Table 4.14 shows the results of the Box-Ljung test, p-values of 0.1013 and 0.9258 for residuals of the minimum and maximum temperature respectively were obtained. The residuals were found to be independents as the obtained p-values are greater than 0.05 and therefore we fail to reject the null hypothesis: H_0, that our model does not exhibit lack of fit and conclude that the data values

Models	AIC	AICc	BIC
SARIMA$(1,0,1)(1,1,1)_{12}$	927.54	927.70	947.13
SARIMA$(1,1,1)(1,1,1)_{12}$	946.42	946.58	966
SARIMA$(1,1,1)(0,1,1)_{12}$	944.45	944.56	960.12
SARIMA$(1,0,1)(0,1,1)_{12}$	926.30	926.41	941.98

Table 4.11: Statistics for tentative models of the Minimum Temperature

Models	AIC	BIC	AICc
SARIMA$(1,0,0)(0,1,1)_{12}$	808.87	820.63	808.94
SARIMA$(1,0,0)(0,1,2)_{12}$	810.87	826.55	810.98
SARIMA$(1,0,0)(1,1,1)_{12}$	810.87	826.55	810.98
SARIMA$(1,0,0)(1,1,2)_{12}$	809.77	829.37	809.94

Table 4.12: Statistics for tentative models of the Maximum Temperature

SARIMA$(1,0,1)(0,1,1)_{12}$

Parameter	Estimate	Standard Error
ar1	0.8405	0.0506
ma1	-0.5386	0.0768
sma1	-0.9123	0.0351

SARIMA$(1,0,0)(0,1,1)_{12}$

Parameter	Estimate	Standard Error
ar1	0.5127	0.0446
sma1	-0.9322	0.0365

Table 4.13: Parameter Estimation

are independent hence the model is adequate for forecasting.

Box-Ljung Test

	X-squared	df	P Value
SARIMA$(1,0,1)(0,1,1)_{12}$	33.131	24	0.1013
SARIMA$(1,0,0)(0,1,1)_{12}$	14.824	24	0.9258

Table 4.14: Box-Ljung test

4.3.2 Diagnostic Analysis

A well fitted model has estimated standardized residuals that behave as an independent identically distributed sequence with zero mean and constant variance i.e the residuals reflect the main properties of the innovations. Namely, they should be stationary and free of any dependency, as well as approximately Gaussian. There are no Acfs or Pacfs that exceed the confidence bounds, hence we can safely conjecture that the residuals are not correlated and hence, all the dependency signal has been captured by the models.

From the normal probability plot of the residuals against time on figure 4.15 and 4.16 for Kakamega Station and figure 4.17 and 4.18 for Eldoret Station, we can see that there is no obvious pattern in the plot with no presence of outliers, it can be observed the residuals are approximately normal. Both the plots of the ACF and PACF of the residuals lack enough evidence of significant spikes that are outside the confidence bounds hence clearly shows that the residuals are white noise.

Therefore SARIMA(2,0,0)$(0,1,1)_{12}$ and SARIMA(1,0,0)$(0,1,1)_{12}$ for Kakamega station are adequate for forecasting, same to SARIMA(1,0,1)$(0,1,1)_{12}$ and SARIMA(1,0,0)$(0,1,1)_{12}$ for Eldoret station.

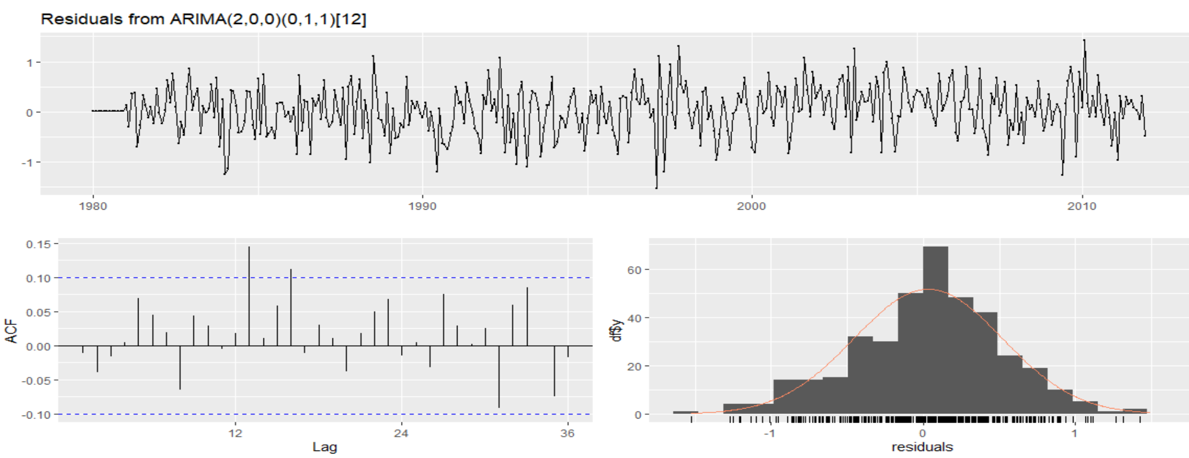


Figure 4.15: Residuals form ARIMA(2,0,0)$(0,1,1)_{12}$

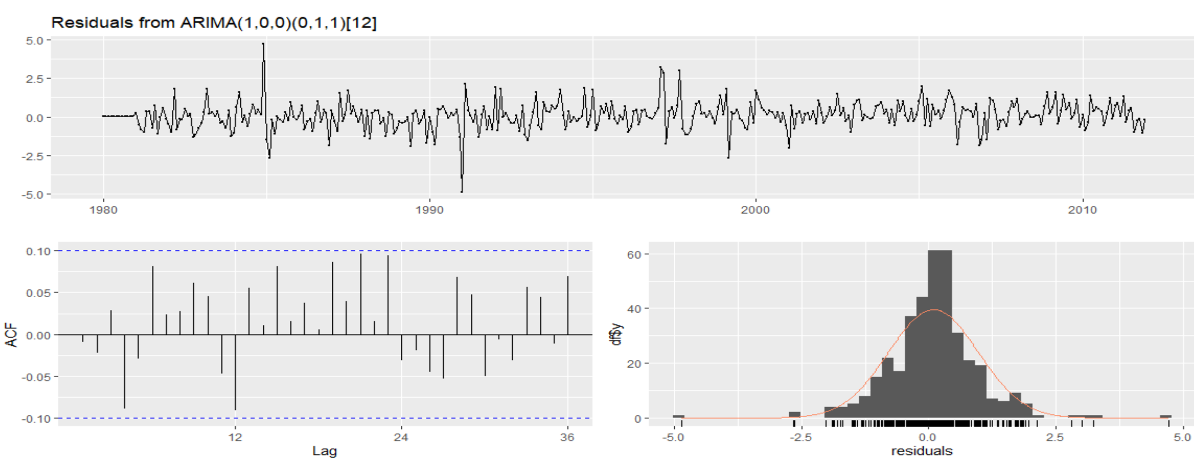


Figure 4.16: Residuals form ARIMA(1,0,0)$(0,1,1)_{12}$

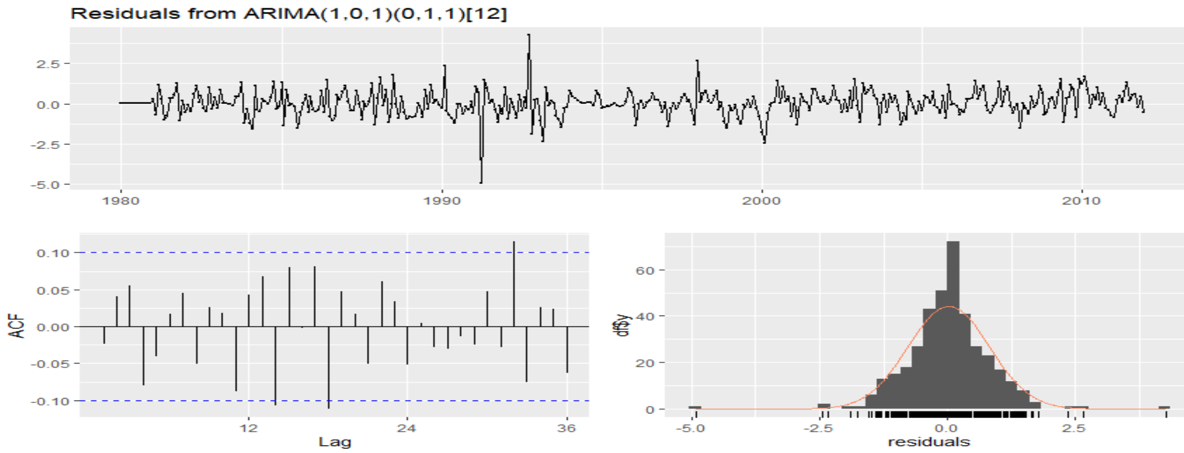


Figure 4.17: Residuals form ARIMA(1,0,1)$(0,1,1)_{12}$

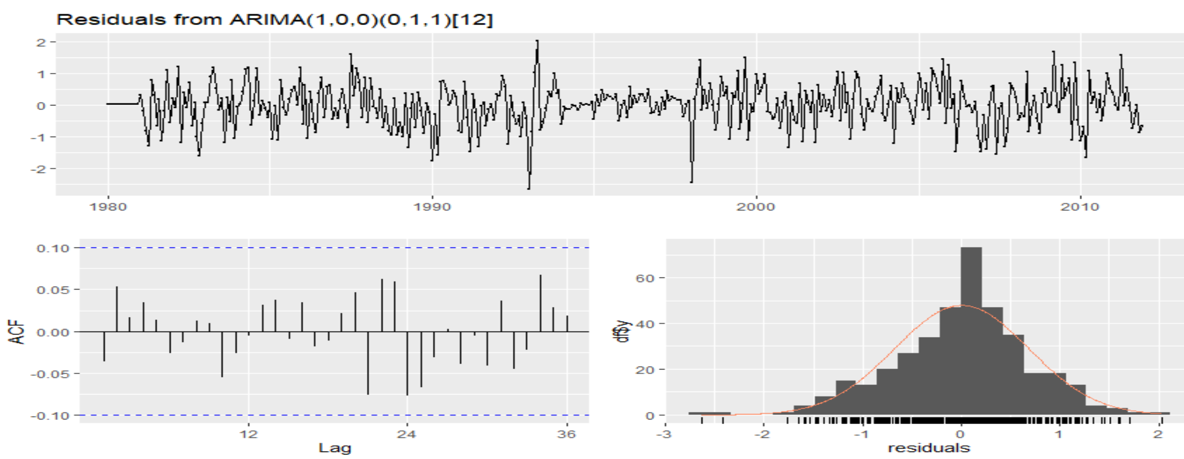


Figure 4.18: Residuals form ARIMA(1,0,0)$(0,1,1)_{12}$

4.4 Forecasting

Forecasting gives an overview of the future observations or occurrences using the current and past behaviour of given observations thereby helping in planning and decision making process. The selected models were subjected to various tests such as MAE, MAPE and RMSE to test their adequacy and predictive behavior.

Table 4.15 and 4.16 below shows a summary of the error statistics: ME, MAPE, MPE, RMSE and MAE for Kakamega and Eldoret respectively. We observe small values for the MAPE statistic, the smaller the value the better the forecast while bigger numbers suggest bad forecast. Model

Accuracy is obtained by subtracting MAPE from 100 ($100-MAPE = ModelAccuracy$) which gives the accuracy of the model. All the four models have an accuracy of more than 95%.

SARIMA$(2,0,0)(0,1,1)_{12}$	
MAE	0.1144612
MAPE	2.687615
RMSE	0.3807247
MPE	0.1144612
MASE	0.03123463
ME	0.4875491

SARIMA$(1,0,0)(0,1,1)_{12}$	
MAE	0.6277194
MAPE	2.267641
RMSE	0.8868727
MPE	0.2550877
MASE	0.6747479
ME	0.09383642

Table 4.15: Forecasting Accuracy Statistic for Kakamega Station

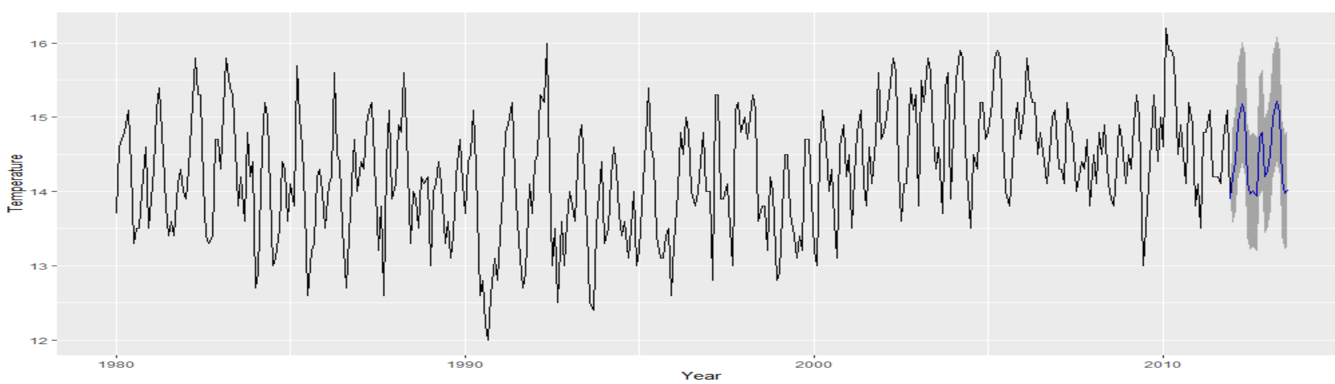


Figure 4.19: Two Year Forecast for Minimum Temperature of Kakamega Station

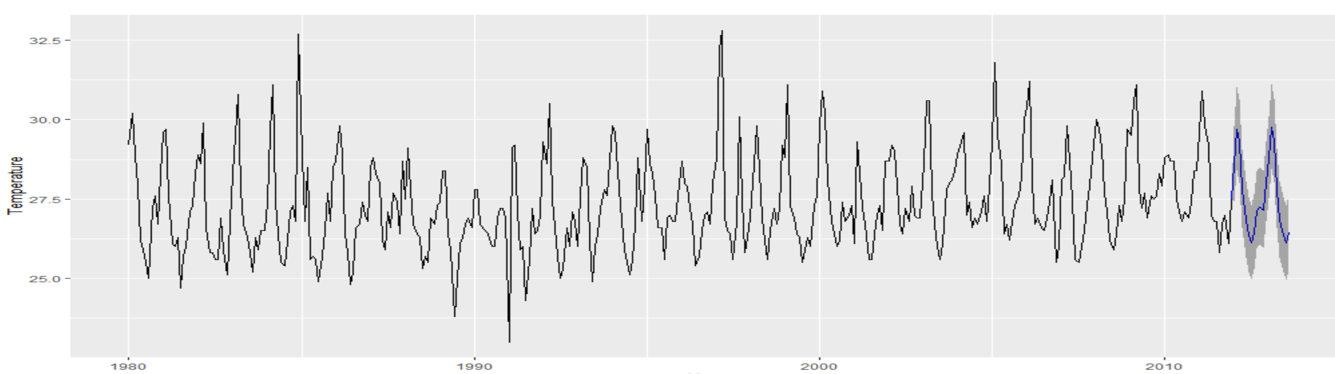


Figure 4.20: Two Year Forecast for Maximum Temperature of Kakamega Station

SARIMA(1,0,1)$(0,1,1)_{12}$	
MAE	0.5646188
MAPE	5.360194
RMSE	0.794963
MPE	-0.2591158
MASE	0.6355165
ME	0.02420571

SARIMA(1,0,0)$(0,1,1)_{12}$	
MAE	0.5122463
MAPE	2.183582
RMSE	0.6779582
MPE	-0.05379152
MASE	0.6176843
ME	0.003334289

Table 4.16: Forecasting Accuracy Statistic for Eldoret Station

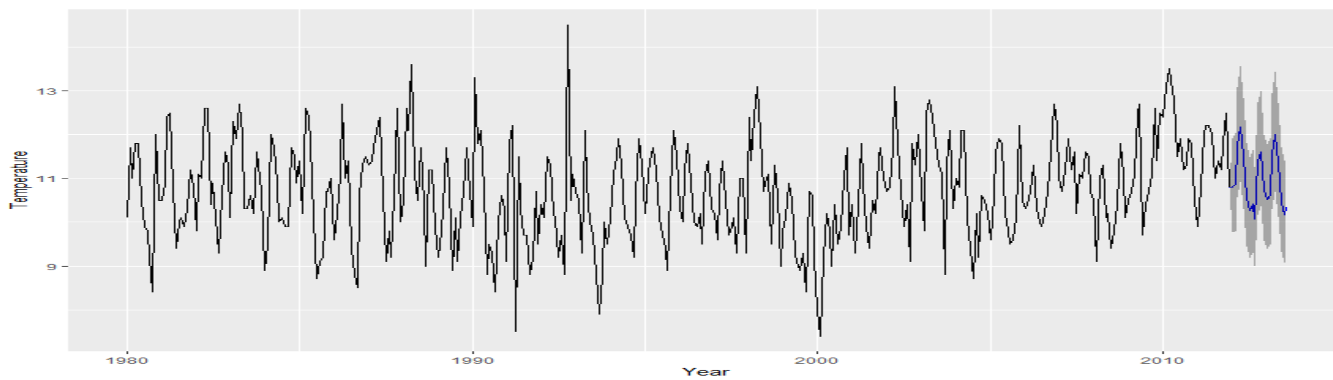


Figure 4.21: Two Year Forecast for Minimum Temperature of Eldoret Station

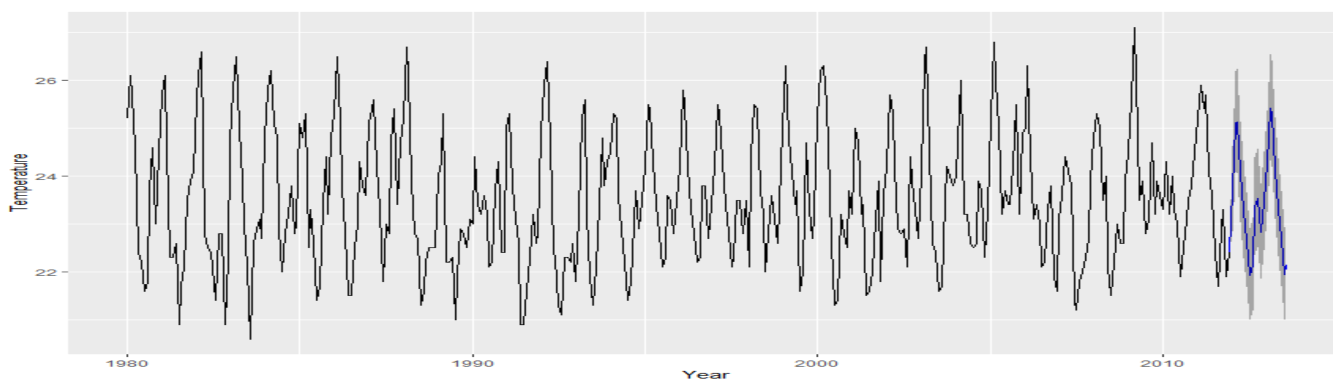


Figure 4.22: Two Year Forecast for Maximum Temperature of Eldoret Station

Chapter 5

Summary, Conclusion and Recommendation

5.0.1 Summary

The first quarter of the year is the hottest time of the year in Kenya, from the data of the two stations February and March are the most hottest months in the year. Likewise, coldest months range between June and September, however, September is the coldest month having the most lowest numbers.

Both maximum and minimum temperatures from the two stations are unstable if observed from month to month, the maximum temperature has the most variance, 2.21 and 1.84 for Kakakega and Eldoret respectively. Likewise, the maximum temperature had the largest standard deviation, values of 1.49 in Kakakega and 1.36 in Eldoret.

The minimum temperature seems steady as compared to maximum temperature however, it has a slight upward trend, increasing over time with Kakamega station having an average of 0.18^0C while Eldoret station has 0.12^0C yearly increment. High seasonality was observed in both maximum and minimum temperature series data which was differenced to be stationary.

Using Box Jenkin's model building procedures, several SARIMA models were suggested and developed with the best models being selected based on AIC criterion.

5.0.2 Conclusion

The best models selected with the aid of AIC criterion in Kakamega Stations were: SARIMA(2,0,0)(0,1,1)$_{12}$ and SARIMA(1,0,0)(0,1,1)$_{12}$ while for Edoret station were: SARIMA(1,0,1)(0,1,1)$_{12}$ and SARIMA(1,0,0)(0,1,1)$_{12}$ for minimum and maximum temperature respectively.

It was found that the model fitted the data well as the residuals for both series are near normality with most points falling on the straight line with a few close to it while the stochastic seasonal fluctuation was successfully modelled except for some extreme values.

The selected SARIMA models were used to forecast a two year monthly period. From the forecasts, its was observed that the selected models from Kakamega station best predicted the data. Both models (maximum and minimum) had a small MAPE value of less than three. This shows that the two models for the station had at least a 97% accuracy as compared to Eldoret station whose maximum temperature model had a MAPE value of 5.36 translating to a 94% model accuracy.

5.0.3 Recommendation

The predictions based on the models indicated that the minimum temperature will continue to rise in the coming years. This shows a distinct trend proving that indeed globe warming is a fact and is happening.

It is recommended that multiple univariate Box-Jenkins time series models should be developed for the various stations across the country for more representation. Likewise, more time series models should also be developed for other weather variables like rainfall and humidity for a better comparison with temperature models for insightful recommendations to be made.

Appendices

Two Year Forecast

Month	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2012	10.83583	9.796293	11.87537	9.245993	12.42567
Feb 2012	10.94203	9.856160	12.02790	9.281336	12.60272
Mar 2012	11.77361	10.656174	12.89105	10.064638	13.48258
Apr 2012	12.19556	11.056353	13.33477	10.453292	13.93783
May 2012	11.53695	10.382606	12.69128	9.771536	13.30235
Jun 2012	10.59287	9.427960	11.75778	8.811295	12.37444
Jul 2012	10.29149	9.119179	11.46381	8.498593	12.08440
Aug 2012	10.42278	9.245258	11.60030	8.621916	12.22364
Sep 2012	10.07734	8.896154	11.25852	8.270875	11.88380
Oct 2012	11.38658	10.202820	12.57034	9.576175	13.19698
Nov 2012	11.61473	10.429157	12.80031	9.801551	13.42791
Dec 2012	10.75586	9.569007	11.94272	8.940724	12.57100
Jan 2013	10.54216	9.347349	11.73697	8.714855	12.36946
Feb 2013	10.69521	9.498545	11.89187	8.865068	12.52535
Mar 2013	11.56617	10.368196	12.76414	9.734027	13.39831
Apr 2013	12.02122	10.822319	13.22011	10.187661	13.85477
May 2013	11.39042	10.190867	12.58997	9.555865	13.22497
Jun 2013	10.46972	9.269709	11.66973	8.634462	12.30497
Jul 2013	10.18799	8.987659	11.38833	8.352241	12.02374
Aug 2013	10.33579	9.135227	11.53635	8.499687	12.17189
Sep 2013	10.00422	8.803502	11.20495	8.167878	11.84057
Oct 2013	11.32513	10.124298	12.52597	9.488615	13.16165
Nov 2013	11.56309	10.362177	12.76400	9.726452	13.39973
Dec 2013	10.71246	9.511493	11.91342	8.875740	12.54918

Table 5.1: Two Year Forecast for Minimum Temperature of Eldoret Station

Month	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2012	23.74806	22.86225	24.63386	22.39334	25.10278
Feb 2012	25.15350	24.15809	26.14891	23.63115	26.67585
Mar 2012	25.16078	24.13851	26.18305	23.59735	26.72421
Apr 2012	24.09835	23.06913	25.12756	22.52430	25.67240
May 2012	23.29805	22.26702	24.32909	21.72122	24.87488
Jun 2012	22.67253	21.64102	23.70404	21.09497	24.25009
Jul 2012	21.94205	20.91041	22.97368	20.36430	23.51980
Aug 2012	22.15000	21.11833	23.18167	20.57220	23.72781
Sep 2012	23.36528	22.33361	24.39696	21.78747	24.94310
Oct 2012	23.54215	22.51047	24.57383	21.96433	25.11996
Nov 2012	22.83453	21.80286	23.86620	21.25672	24.41234
Dec 2012	23.47232	22.44066	24.50397	21.89454	25.05010
Jan 2013	24.29782	23.26440	25.33124	22.71734	25.87829
Feb 2013	25.43535	24.40145	26.46925	23.85414	27.01656
Mar 2013	25.30528	24.27125	26.33931	23.72387	26.88669
Apr 2013	24.17243	23.13837	25.20649	22.59097	25.75389
May 2013	23.33603	22.30197	24.37010	21.75456	24.91750
Jun 2013	22.69200	21.65793	23.72607	21.11053	24.27347
Jul 2013	21.95203	20.91796	22.98610	20.37056	23.53351
Aug 2013	22.15512	21.12105	23.18919	20.57365	23.73660
Sep 2013	23.36791	22.33384	24.40198	21.78643	24.94938
Oct 2013	23.54349	22.50942	24.57756	21.96202	25.12496
Nov 2013	22.83522	21.80116	23.86928	21.25376	24.41668
Dec 2013	23.47267	22.43863	24.50672	21.89124	25.05411

Table 5.2: Two Year Forecast for Maximum Temperature of Eldoret Station

Month	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2012	28.59636	27.43826	29.75445	26.82520	30.36751
Feb 2012	29.71064	28.47962	30.94166	27.82796	31.59332
Mar 2012	29.32033	28.08015	30.56051	27.42364	31.21702
Apr 2012	27.79868	26.55732	29.04004	25.90018	29.69718
May 2012	26.82797	25.58645	28.06948	24.92924	28.72670
Jun 2012	26.34703	25.10550	27.58857	24.44827	28.24579
Jul 2012	26.10304	24.86150	27.34458	24.20427	28.00181
Aug 2012	26.43546	25.19392	27.67700	24.53669	28.33423
Sep 2012	27.13056	25.88902	28.37210	25.23179	29.02933
Oct 2012	27.25262	26.01108	28.49416	25.35385	29.15139
Nov 2012	27.15816	25.91662	28.39970	25.25939	29.05693
Dec 2012	28.04275	26.80121	29.28428	26.14399	29.94151
Jan 2013	28.79198	27.54589	30.03807	26.88626	30.69770
Feb 2013	29.78115	28.53447	31.02783	27.87451	31.68778
Mar 2013	29.34574	28.09899	30.59250	27.43899	31.25249
Apr 2013	27.80784	26.56107	29.05461	25.90107	29.71461
May 2013	26.83127	25.58450	28.07804	24.92450	28.73804
Jun 2013	26.34822	25.10145	27.59499	24.44145	28.25499
Jul 2013	26.10347	24.85670	27.35024	24.19670	28.01024
Aug 2013	26.43561	25.18884	27.68238	24.52884	28.34238
Sep 2013	27.13062	25.88385	28.37739	25.22385	29.03739
Oct 2013	27.25264	26.00587	28.49941	25.34587	29.15941
Nov 2013	27.15817	25.91140	28.40494	25.25140	29.06493
Dec 2013	28.04275	26.79598	29.28951	26.13599	29.94951

Table 5.3: Two Year Forecast for Maximum Temperature of Kakamega Station

Month	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2012	14.20800	13.57058	14.84542	13.23315	15.18285
Feb 2012	14.42518	13.73186	15.11849	13.36485	15.48551
Mar 2012	14.95007	14.21211	15.68803	13.82146	16.07868
Apr 2012	15.18621	14.42972	15.94269	14.02927	16.34315
May 2012	15.04341	14.27664	15.81018	13.87074	16.21608
Jun 2012	14.12630	13.35438	14.89823	12.94575	15.30686
Jul 2012	13.95800	13.18336	14.73264	12.77329	15.14271
Aug 2012	14.00418	13.22813	14.78024	12.81732	15.19105
Sep 2012	13.93323	13.15645	14.71002	12.74524	15.12123
Oct 2012	14.71830	13.94112	15.49547	13.52971	15.90688
Nov 2012	14.80108	14.02370	15.57845	13.61219	15.98997
Dec 2012	14.20351	13.42603	14.98098	13.01445	15.39256
Jan 2013	14.27404	13.49211	15.05596	13.07819	15.46988
Feb 2013	14.51823	13.73537	15.30110	13.32095	15.71552
Mar 2013	15.00399	14.22037	15.78760	13.80556	16.20241
Apr 2013	15.22914	14.44520	16.01309	14.03021	16.42808
May 2013	15.07329	14.28916	15.85742	13.87407	16.27251
Jun 2013	14.14826	13.36403	14.93248	12.94889	15.34762
Jul 2013	13.97378	13.18950	14.75805	12.77433	15.17322
Aug 2013	14.01562	13.23132	14.79992	12.81614	15.21510
Sep 2013	13.94149	13.15718	14.72581	12.74199	15.14100
Oct 2013	14.72427	13.93995	15.50859	13.52476	15.92379
Nov 2013	14.80540	14.02108	15.58972	13.60588	16.00492
Dec 2013	14.20663	13.42231	14.99095	13.00711	15.40615

Table 5.4: Two Year Forecast for Minimum Temperature of Kakamega Station