AJAS / SPAS

African Journal of Applied Statistics
Vol. 6 (2), 2019, pages 689 – 710.
DOI: http://dx.doi.org/10.16929/ajas/2019.689.237

ISSN 2316-0861

Time Series Modelling of Monthly average Temperature in Gaborone-Botswana

Keorapetse Sediakgotla *, Wilford Molefe and Kehinde Dahud Shangodoyin

Department of Statistics, University of Botswana

Received September 5, 2019; Accepted on October 10, 2019; Published online October 21, 2019

Abstract. The seasonal series on the average maximum temperatures in Gaborone (Botswana) is used to identify the best time series model that can be used for forecasting. The series was found to be highly seasonal. Seasonally adjusting the series prior to applying the Box and Jenkins procedure did not average out the seasonal effects, despite giving a fairly good ARIMA(1,1,1). We also fitted SARIMA(p,d,q)(P,D,Q) with seasonality effects at lags s=12,24,36. Correlograms, Dickey Fuller tests and other model comparison methods led to an ARIMA(1,1,1)(0,1,1)[12]. The seasonal multiplicative SARIMA was found to be parsimonious as compared to an additive seasonal SARIMA.

Key words: autoregressive integrated moving average (ARIMA); saeasonal auto regressive integrated moving average (SARIMA); aeasonality; aeseasonalization
AMS 2010 Mathematics Subject Classification Objects : 62M10;62P12

Corresponding author : Keorapetse Sediakgotla (sediakgo@mopipi.ub.bw)
Wilford Molefe: molefewb@ub.ac.bw
Kehinde Dahud Shangodoyin: shangodoyink@mopipi.ub.bw

Résumé (Abstract in French) La série temporelle des températures maximales moyennes à Garbonne (Botswanan) est utilisée pour touver le meilleur modèle prévisionnel. La série s'est révelée très saisonnière . La méthode d'ajustage saisonnière avant l'application de la procédure de Box et Jenkis n'a pas pu éliminer les effets saisonniers bien aue le modèle ARIMA(1,1,1) ait été jugé bon. Non plus n'a pas été très bon le modele SARIMA(p,d,q)(P,D,Q) avec des effets saisonniers de temps de retards s=12.24,36. Par contre l'utilisation des correllogrammes, les tests Decker-Feller et autres methodes de comparaison ont abouti à un modèle ARIMA(1,1,1)(0,1,1)[12] satisfaisant. Les effets saisonniers multiplicative SARIMA sont rares par rapport au modèle additif.

1. Introduction

Weather is the atmospheric state measured at a given time and place with regard to temperature, air pressure, wind, humidity and precipitation. The earliest evidence of scientific activity in meteorology especially its relevance to weather forecasting is by Aristotle, Craft (2001): An economic history of weather forecasting (http://eh.net/article/craft.weather.forecasting.history) who wrote what is probably the first treatise on the subject. Craft (2001) mentioned that H.W Brandes attempted to chart weather from reports over a considerable area but it was not until after the invention of telegraph that the rapid collection of weather data from remote stations became possible. Over the years researchers have used computerized models based on mathematical formulations of the dynamics of the atmosphere to produce weather charts that were used as prognostics of future weather patterns. The many simplifying assumptions required in these formulations as well as the incompleteness of weather data limited the accuracy of computer prediction; though as the world achieved advances in mathematical modelling and computer systems, these models are becoming more complete, reliable and accurate.

In this paper, we shall focus on forecasting weather parameters in Botswana using some competitive classical time series models with a view to ascertain their effect on long run climate change . In classical time series modelling we assume that a sequence of series $\{Y_t\}$ could be represented by a definite pattern inclusive of some random component. One major objective of time series analysis is to differentiate the series' pattern from the error component by studying the trend and seasonality; more importantly is forecasting the future values of the series using reliable models Chatfield (1980). In time series literature some methods have been developed to forecast future values of dynamic systems. These include inter alia the generalized exponential weighted moving average process developed by Makridakis *et al.* (2003), the use of univariate time series models discussed by Geweke *et al.* (1983) while Quenouille (1949), Ljung and Box (1978) and Pindyck and Rubinfeld (1981) have emphasized the use of autoregressive moving average process in forecasting. This paper will utilize the autoregressive moving average process and the seasonal autoregressive moving average models in forecasting the weather parameters in Botswana. The choice of these methods is due to their similarity in time series modelling. Their performance will be evaluated using the significance of the coefficients and measures of forecast accuracy.

2. Background

There are many factors that affect the weather and climate of Botswana but the main ones are latitude, position on the continent and cloud cover. These factors often act in combination with each other. Botswana is situated in the centre of Southern Africa, straddling the Tropic of Capricorn, a factor that partly explains why the country is hot. Climatically, the country is continental arid to semi-arid and rainfall is characterised by extreme variability in time and space. Temperatures in the north of the country, which is nearer to the equator, are generally higher than those in the south. For example, Maun, located in the arid region to the north experiences highest mean maximum and minimum temperature in June of 25 degrees Celsius and 6.9 degrees Celsius compared with Tsabong, further south, with a mean maximum and minimum temperatures of 22.0 degrees Celsius and 1.3 degrees Celsius respectively, Selitshena and McLeod (1989). The effect of position on the continent can be seen in the generally wide temperature ranges. Tsabong has a highest monthly mean temperature of 26.7 degrees Celsius in January and a lowest monthly mean temperature of 11.6 degrees Celsius in July. This position also explains why the country receives little rain. The rain-bearing easterly winds - which blow from the Indian Ocean, lose most of their moisture before they reach Botswana. For this reason rainfall tends to be higher in northern and eastern Botswana. The country has also experienced a number of drought periods in 1967 to 1970, 1972, 1981 to 1986, and the early 1990's.

In many places in Botswana, lack of cloud cover is one of the major causes of high daytime and low night-time temperatures. In dry years, when there is less cloud cover, temperatures tend to be higher than in wet years. Also worth noting is the role of man as agents of climatic change in general. Pollution of the atmosphere with gases from power stations and industries, and the removal of vegetation through deforestation and overgrazing, has some effect on climate, though not necessarily in Botswana.

Botswana is generally flat to gentle undulations in the east. Deep Kalahari sands cover the centre, south and west. Rock formations are exposed only in the east. Mean annual rainfall varies broadly from 650 millimetres in the north-east to less than 250 millimetres in the south-west. Rates of evaporation and transpiration are high at all times and greatly exceed precipitation. Nationally, average daily maximum temperatures are around 33.0 degrees Celsius in January and 22.0 degrees Celsius in July. Daily minimums are as low as 5.0 degrees Celsius in July and frost is not uncommon. Permanent water sources are scarce; the erratic rainfall, high evaporation and freely draining soils mean rainwater pools are ephemeral.

Contribution of agriculture to GDP in Botswana has declined from 42.7% at independence in 1966 to 2% in 2016, and mining is the largest contributor at 19.9% Statistics Botswana (2016). Rainfall in Botswana is erratic and significantly impacts the country's economy. Water deficits during global El Nino events, which hit Southern Africa, such as in 1983 and 1992, lead to poor food production throughout the region, but not only in Botswana Jury *et al.* (1999). An overwhelming rural population depended mainly on agriculture for a livelihood. Beef production was the mainstay of the economy in terms of the output and export earnings NDP8 (1997). Despite severe ravages on both arable and livestock agriculture by long and severe droughts, farmers in Botswana are still highly dependent on erratic rains, which are highly variable both in time and space. The rains adverse effects may be partially reduced if the occurrence of the events is predicted or known in advance and farmers are suitably advised to take ameliorative measures Parvinder *et al.* (2003). Like rainfall, surface weather parameters like the maximum and minimum temperatures play an important role in agricultural activities. Hence extreme temperatures will have an adverse effect on agricultural operation.

Due to changes in temperature worldwide there is need for a study in the temperature pattern, evaporation and water levels. The scope of this paper is limited to modelling temperature patterns and their long term effects only. Temperature dependent parameters like maximum and minimum temperatures are monitored for seemingly extreme patterns. Simple exploratory models could inform government and other stakeholders on programs and policies that could mitigate the impact of Botswana's fluctuating climate. It is believed that an advance warning of drought risk and seasonal rainfall prospects will improve the economic growth potential of Southern Africa and provide additional security for food and water Jury *et al.* (1997).

The second objective is that this study is first of its nature in Botswana. Several studies have been carried out in Botswana but many of them were on Rainfall. Hence, forecasts of dry periods and extreme temperatures if reliably done can provide an adequate planning tool for weather based agricultural practices in this country. High daytime temperatures lead to high water losses throughout the year from evaporation and transpiration Cooke (1980). Botswana in particular has been experiencing changes in rainfall season which in turn affects planting seasons and livestock rearing.

Implications of climate change for agricultural and economic activities cannot be over emphasized Cline (1992); and effective management of climate change influence on socio-economic activity will depend on the availability of reliable, accurate and cost-effective forecast. According to Romilly (2005), both operational and strategic decision-making in aforementioned activities will have to take into account not just the realized effects of climate variation but also potential causative effects.

To have undaunted effective agricultural and business managerial decision it is highly crucial to have a robust climate change forecast and a leading indicator of climate change is temperature change. This study will focus on developing a parsimonious time series model for the mean temperature change that is effective in forecasting over short term horizon of between 5-10 years in the Gaborone area.

3. Statistical methods

Climate data usually exhibit non-stationary behaviour, this includes long term trends that are mostly associated with independent driving force such as greenhouse gas concentration Brohan *et al.* (2006), this behaviour presents challenges for analysts using traditional methods which assume stationary time series behaviour. The fact that temperature variables are the product of gradually evolving processes makes it appropriate and desirable to calibrate these causal models on data that go back in time as much as possible; of course causal variables may not be reliable and even not available for previous periods under investigation.

To accommodate this challenges a non-seasonal model known as autoregressive integrated moving average (ARIMA) coupled with seasonal unit testing and generalized autoregressive conditional heteroscedasticity (GARCH) models are used by several authors such as (Shangodoyin *et al.* (2010); Mann (2008); Shangodoyin (2008); Baillie and Chung (2002) and Posamentier and Nicolich (1979)) to mention a few.

Specifically therefore in this study we shall utilize monthly data that allows for seasonal unit root testing with or without structural breaks; also more suitable, reliable, accurate and autoregressive integrated seasonal moving average will be utilized to make suitable short term forecasting that can inform climate managerial decision.

4. Analytical approach to time series components and structural breaks in data set

The general features of environmental and climate time series suggest the presence of basic time series components and structural breaks in data set. We briefly present the methodology to be utilized in this study.

4.1. Analyzing Time series components

In the classical analysis of time series data, it has been observed that the general features of environmental and climatic time series suggest the presence of polynomial trend and visible seasonal effect Brohan *et al.* (2006). The intuition behind the concept of trend, seasonal variation and cyclical movement leads to the idea of decomposition of observed series into 'unobserved component' as:

$$y_t = \sum_j y_{jt} + \varepsilon_t \tag{1}$$

where y_{jt} denote the accumulation of unobserved components as trend, seasonal variation and cyclical variation; ε_t is the residual effect which is referred to as the irregular component. Alternatively we could write equation (1) as:

$$y_t = \prod_j y_{jt} + \varepsilon_t \tag{2}$$

In equations (1) and (2) the random shocks are assumed to be mutually and serially uncorrelated with expected value equal to zero and variance σ_ε^2; thus the trend and seasonal effects are assumed to be random variables changing over time Rivera (1990). In this paper, we assume that the observed series is functionally related to both secular trend and seasonal variation.

To estimate the trend component (m_t) we assume the least squares method of estimation to fit the best polynomial model (model with minimum error variance) of the form

$$\hat{m}_t = \hat{a}_0 + \hat{a}_1 t + \hat{a}_2 t^2 + \ldots + \hat{a}_k t^k \tag{3}$$

The presence of trend could be confirmed by testing for significance of the estimates $\hat{a}_i, \quad i = 0, 1, \ldots, k$ in the model specified in equation (3). The seasonal component (s_t) will be estimated using the model

$$\hat{s}_t = y_t - \hat{m}_t \tag{4}$$

Using equation (4), we compute the seasonal monthly index for N years as:

$$SI_{(p)} = \frac{1}{N} \sum_{t=1}^{N} S_{t_{(p)}} \tag{5}$$

For monthly series we define period as $p =$ January, $\ldots$, December for the year in use. To confirm the presence of seasonality in the data set, let the total number of years be denoted as c (column) and r denotes the total number of months. By using the Kendall' test each column represents a permutation of integers 1, 2,3,...,12 and summing across the rows gives the monthly score $M_j (j = 1, 2, \ldots, 12)$.

Under the null hypothesis of no seasonal variation, the test statistic is

$$T = \frac{12 \sum \left[M_j - \frac{c(r+1)}{2} \right]^2}{cr(r+1)} \tag{6}$$

If the series contains significant seasonal effect it will be desirable to adjust the observed time series by removing the influence of the seasonal component, to have a de-seasonalized series or to fit models that capture seasonality.

4.2. Structural breaks indentification

There are a number of tests that rely on the maximization or minimization of some objective function to determine the time of the structural break in the data set; tests presented by Zivot and Andrews (1992); Perron (1997) and Bai and Perron (2003) are not pure objective since they depend on model specification to conduct the test. In most cases analysts prefer graphical detection methods backed up with Statistical theory Romilly (2005).

We rely on the dummy variable (DV) technique proposed by Perron (1989) to test the existence of structural break in the series; this seems reasonable for the temperature series since it exhibit significant change in level and trend manifest gradually over time. We confirm the presence of structural break at time point in the constant or trend term by estimating the dummy regression equation given below.

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 DVu_t + \alpha_3 DVTB_t + \beta_1 t + \beta_2 DVT_t + \sum w_j \Delta y_{t-j} + \varepsilon_t \quad (7)$$

where t is a time trend,

$$DVu_t = \begin{cases} 0 \text{ if } t \leq T_b \\ 1 \text{ if } t > T - b \end{cases}$$

$$DVT = (DVu_t) * t,$$

$$DVTBu_t = \begin{cases} 0 \text{ if } t \leq T_b + 1 \\ 1 \text{ elsewhere} \end{cases} .$$

and ε_t is independently and identically distributed with mean zero and variance σ^2_ε.

4.3. Modeling temperature data with autoregressive integrated seasonal moving average models

In this section we examine whether the series contains unit roots at seasonal monthly frequencies. The existence of unit roots in a seasonal time series implies that the series is non-stationary. This anomaly creates a number of problems for determining an appropriate forecasting model, and more so model estimation over the sample period determines the parameter values that may not be applicable to forecasting due to changing mean and variance level of the series.

Under the null hypothesis that there is no unit root we shall use the Augmented Dickey Fuller (ADF) procedure to test a data generating process for difference stationary (trend non-stationary) against trend stationary. The requirements for the test are: Given the models

$$\Delta Y_t = \alpha + \beta t + (\theta - 1) Y_{t-1} + \varepsilon_t \quad (8)$$

and if augmented the model becomes:

Model I :

$$\Delta Y_t = \lambda Y_{t-1} + \sum_{j=1}^{p} \tau_j \Delta Y_{t-j} + \varepsilon_t \quad \text{(no intercept or linear trend)} \tag{9}$$

Model II :

$$\Delta Y_t = \alpha + \lambda Y_{t-1} + \sum_{j=1}^{p} \tau_j \Delta Y_{t-j} + \varepsilon_t \quad \text{(no linear trend)} \tag{10}$$

Model III

$$\Delta Y_t = \alpha + \beta t + \lambda Y_{t-1} + \sum_{j=1}^{p} \tau_j \Delta Y_{t-j} + \varepsilon_t \quad \text{(includes intercept and linear trend)} \tag{11}$$

where $\lambda = (\theta - 1)$.
Essentially Dickey-Fuller test assumes that the error terms are an independent and identically distributed process. This assumption is relaxed in the Phillips-Perron test (1988). The Augmented Dickey Fuller (ADF) (1981) test procedure is specified when ε_t is autoregressive to eliminate serial correlation of errors. The parameter of interest in equations (9), (10) and (11) is λ. If $\lambda = 0$, then the series Y_t contains a unit root, hence not stationary. The test adopts a step by step procedure in testing for presence or otherwise of a unit root. The lag order p is selected such that $p \leq n/4$.

The Box and Jenkins model building approach has come to be a major reference procedure in the statistics of constructing a model for series. Adopting this modelling procedure the series being modelled must be stationary with a stable variance. A non-stationary series will have its ACF value dampened out gradually with increasing lag period while it will die out fast for a stationary series. A non-stationary series is made stationary by either de-trending or differencing - regular differencing or seasonal differencing as may be appropriate. Box and Jenkins (1976) established fundamental algorithms for building a model for series. The algorithm is an iterative procedure of model identification, estimation, diagnostic check and forecasting.

After a tentative model has been identified in which the ACF and PACF measures play important roles, estimation of the parameters of the tentatively identified model is performed by minimizing the appropriate conditional sum of squares or evaluating the relationship between ACF or PACF measures and the parameter of the model.

Being a tentative model, diagnostic check is conducted on it through the analysis of the error-terms generated, essentially testing its randomness status and the significance of the parameters estimated for the model. If the check confirm the adequacy of the identified model the last stage of model forecasting is performed, however if the model is found to be inadequate the cycle in the model building algorithms is repeated. The Box and Jenkins model building algorithms is applicable to both the univariate and multivariate modelling procedure. In this paper we will adopt the Box-Jenkins modelling building approach to fit the ARIMA model.

The first model to consider is given by:

$$\vartheta(L)(1-L)^d y_t = \theta_0 + \phi_q(L)\varepsilon_t \tag{12}$$

where $\vartheta(L) = (1 - \vartheta_1 L - \vartheta_2 L^2 - \cdots - \vartheta_p L^p)$ is the non-seasonal AR order model; $\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_q L^q)$ is the non-seasonal MA order model; L is the lag operator, d is the order of differencing; θ_0 , is a constant called the trend parameter. Equation (12) is called the autoregressive integrated moving average (ARIMA) model of order (p, d, q) and is denoted as ARIMA(p, d, q). A general representation of an ARIMA (p,d,q) process is the determination of the three parameters. Typically according to Johnston (1984), d is zero or one, or very occasionally two, and one seeks a parsimonious representation with low values of p and q. The difficulty in choosing the order of p and q may be helped by a numerical procedure suggested by Hatanaka (1996). When d has been found, the ARMA procedure for modelling is applied. Equation (12) will be applied after the necessary transformations and deseasonalization if the series is found to be seasonal.

An alternative modelling procedure will then be adopted. In this case the ARIMA(p,d,q), will be varied by fitting another model that captures seasonality, equation (13). The model does not require deseasonalization of the series before fitting it. The best model amongst the two approaches will be identified and used for forecasting.

The model is presented as

$$\theta_P(L^S)\vartheta_p(L)(1-L)^d(1-L^S)^D y_t = \mu_Q(L^S)\phi_q(L)\varepsilon_t \tag{13}$$

where $\vartheta_p(L) = (1 - \vartheta_1 L - \vartheta_2 L^2 - \cdots - \vartheta_p L^p)$ is the non-seasonal AR order; $\phi_q(L) = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_q L^q)$ is the non-seasonal MA order; $\theta_P(L^S) = (1 - \theta_S L^S - \theta_{2S} L^{2S} - \cdots - \theta_{PS} L^{PS})$ is the P^{th} order seasonal AR operator; $\mu_Q(L^S) = (1 - \mu_S(L^S) - \mu_{2S}(L^{2S}) - \cdots - \mu_{SQ}(L^{SQ}))$ is the Q^{th} order MA seasonal operator; and ε_t is the random element with $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, P is the multiplicative seasonal AR order, D is the order of seasonal differences and Q is the multiplicative seasonal MA order. The model is denoted as ARIMA(p,d,q)(P,D,Q).

5. Data description and analysis

5.1. Analysis

Table 1 gives the results of the test for seasonality. The series was found to be highly seasonal at the 1% level of significance. Due to the strong seasonal behavior, figure 1 shows that despite seasonally adjusting the series, there are still some inherent seasonal patterns visible. The correlograms of the first difference of the series, figure 2, shows a non-decaying pattern, which is indicative of a non-stationary series. They also show a strong seasonal behavior of the series. The correlogramms of the first difference of the deseasonalized series are given in figure 3. It is observed from the from figure 3, that the auto correlation and the partial autocorrelation depict fair elements of seasonality, however the single prominent spikes on both the ACF and the PACF, and the oscillatory decaying behavior is suggestive of an ARIMA(1,1,1). In a case in which seasonality doesn't die out due to taking the first difference of a non-seasonal difference of the series, it is ideal to then take the seasonal differencing of the series. The series was found not to have structural breaks and the results are excluded for further discussion.

Table 1. Test for the presence of seasonality assuming stability

Source	Sum of Squares	Degrees of freedom	Mean Square	F-value
Between months	77205.4600	11	7018.67818	395.087[a]
Residual	8527.1381	480	17.76487	
Total	85732.5981	491		

[a] Seasonality present at the 0.1 per cent level.

5.2. The 12th seasonal difference

We proceeded and took a 12th seasonal difference of the original series. Figure 4 shows a fairly constant mean, with clearly evident seasonal spikes in the autocorrelation and partial autocorrelation functions. The signatures of the seasonal pattern do appear across the multiple lags of 12 $(s, 2s, 3s, \ldots)$. The null hypothesis of the presence of the unit root is rejected at the 1%, 5% and 10% levels of significance as shown by the Dicky Fuller test (see Table 3), implying that the series is stationary series. The Auto correlation function exhibits an exponential decay starting at lag 1, suggestive of an AR(1) process. Also there are significant spikes at lags at 12, and multiple lags of seasonal order 12 suggestive of seasonal effects interacting with the moving average term at those lags. The Partial Auto Correlation function started to decay at lag 1 indicative of an MA(1) process. We proceed and fit multiplicative auto regressive integrated moving averages due to the inherent oscillatory behavior of the ACF and PACF and the seasonality effect. We fit several processes of an ARIMA(p,d,q)(P,D,Q), as defined in (13), results are given in table 2.

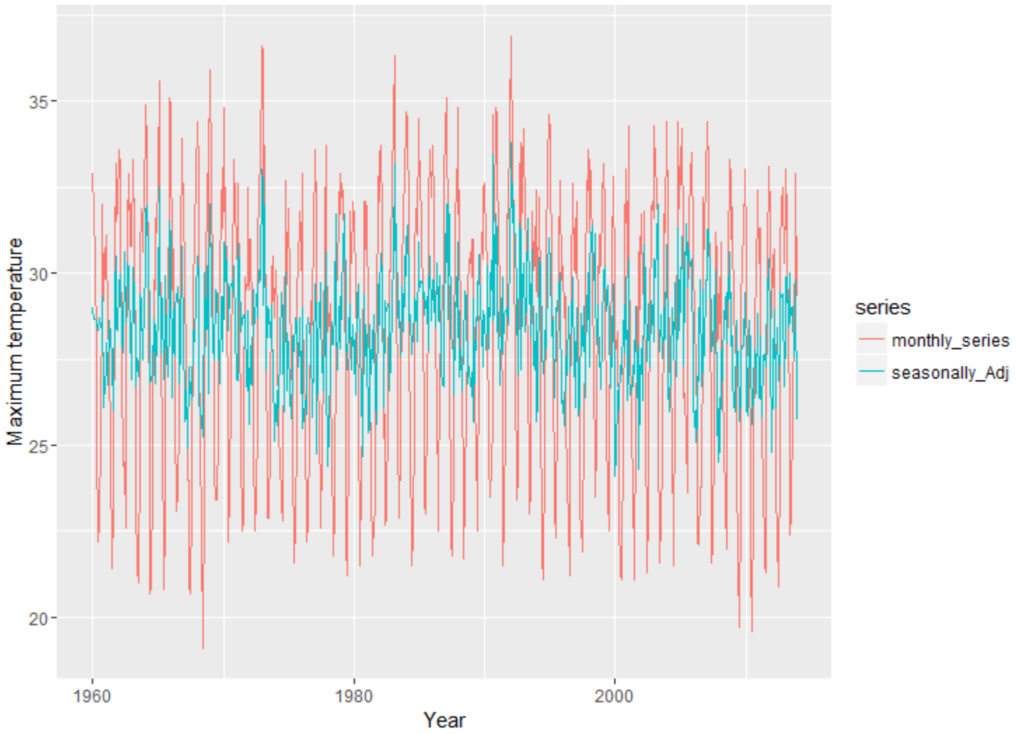


Fig. 1. Time series plot of monthly temperatures

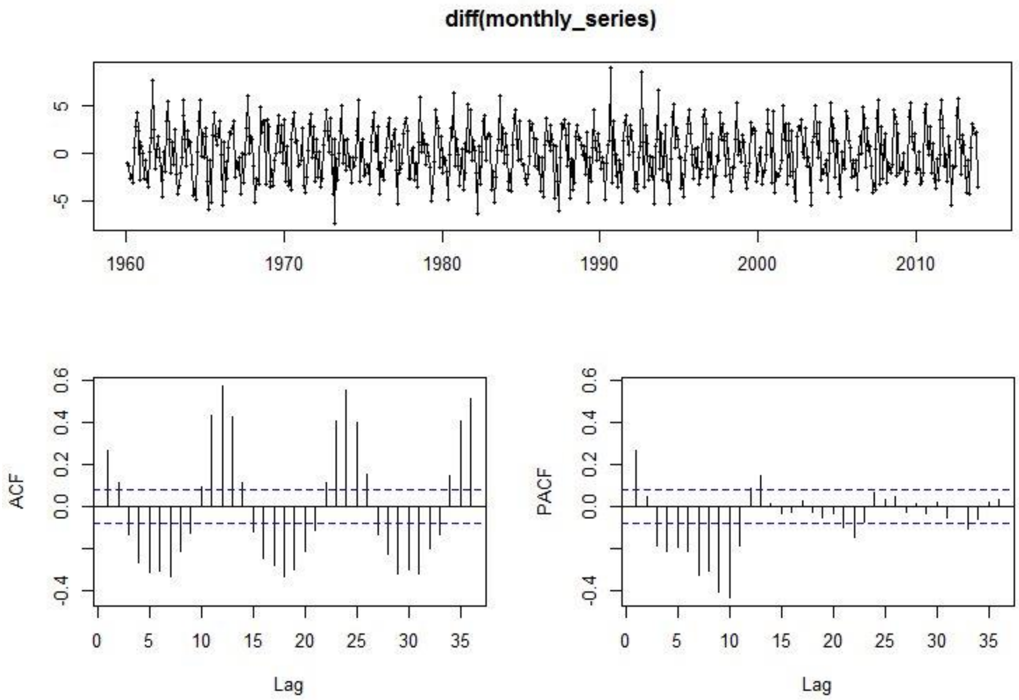
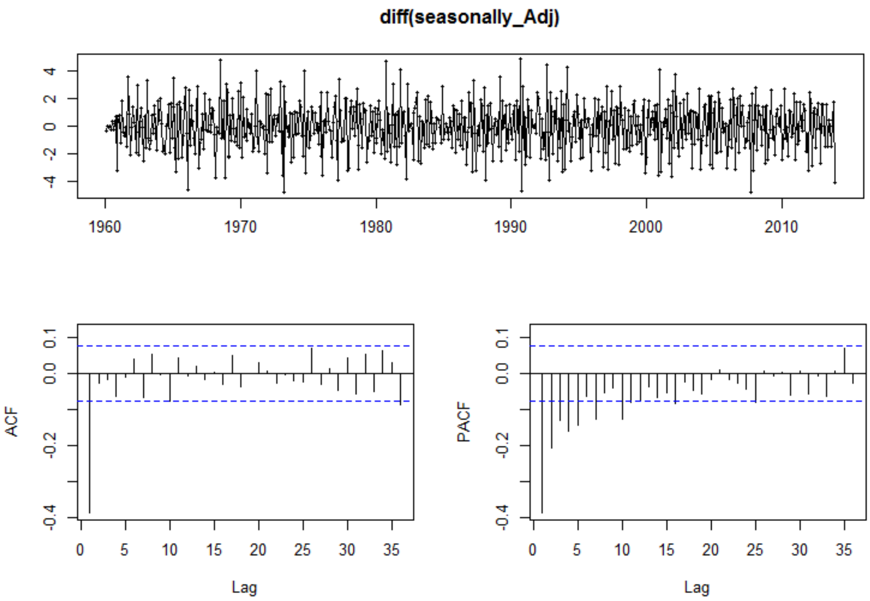


Fig. 2. Correlograms of the first difference of the series



plot.jpg

Fig. 3. Correlograms of the seasonally adjusted series

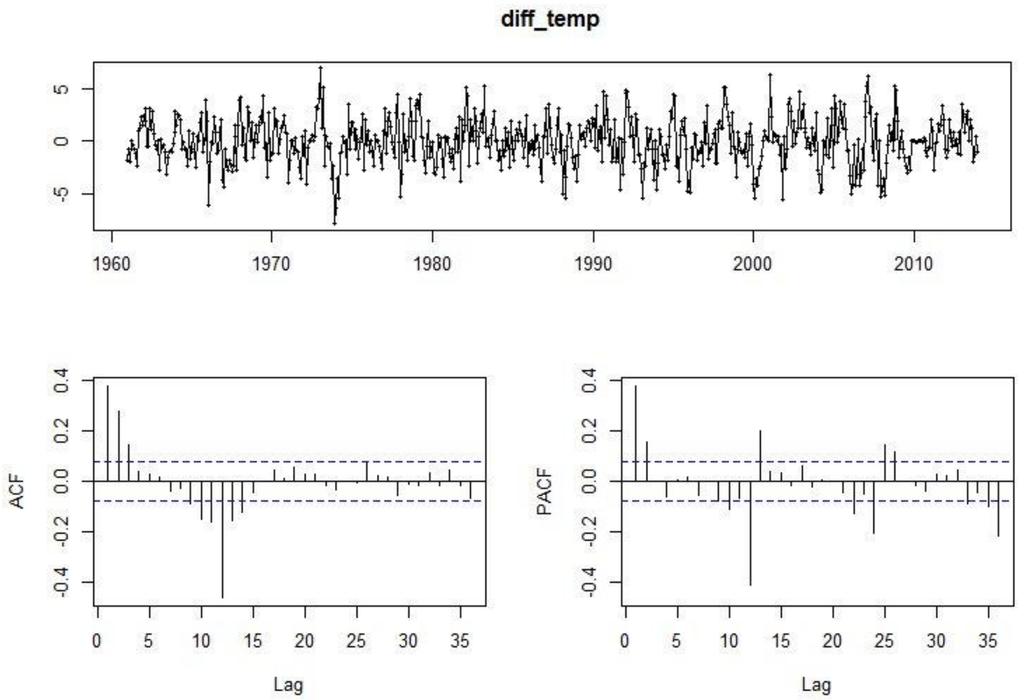


Fig. 4. 12^th seasonal differences of the series

Table 2. Summary of results

coefficients	ARIMA(1,1,1)(1,1,1)	ARIMA(1,0,1)(0,1,1)	ARIMA(1,0,0)(1,1,1)	ARIMA(1,1,1)	ARIMA(1,1,1)(0,1,1)
C		-0.0060(0.4297)	-0.00587(0.3706)		
AR1	0.4046(2.2e-16)	0.6648(2.2e-16)	0.4122(2e-16)	0.389(2.2e-16)	0.378(2.2e-16)
MA1	-0.99(2.2e-16)	-0.3065(0.0002)		-0.977(2.2e-16)	-0.968(2.2e-16)
SAR1	-0.4705(2.2e-16)		0.0379(0.3464)		
SMA1	-0.99(2.2e-16)	-0.9933(2.2e-16)	-0.999(2e-16)		-0.999(2.2e-16)
AIC	2597.33	2350.33	2361.22	2342.72	2363.52
BIC	2619.5	2372.61	2383.5	2356.14	2381.33
σ^2	3.343	2.179	2.206	2.167	2.213
MAE	1.39	1.898	1.2	1.18468	1.1938
RMSE	1.8037	1.4715	1.481	1.468614	1.4829
MASE	0.4717	0.4017	0.04053	0.6941102	0.403

Table 3. Testing for unit root for the 12th differenced Series: Augmented Dickey-Fuller Test

data: monthly series		
Dickey-Fuller = -8.4789	Lag order = 8	p-value = 0.01
alternative hypothesis: stationary		

An ARIMA(1,1,1)(0,1,1), is identified amongst the models that capture seasonality. It has the lowest values of the AIC and BIC. It also has the lowest value of the sums of squares error. The model has captured the seasonal effects at lag 12 for both the Auto regressive and the Moving Average components, and has highly significant coefficients. For the de-seasonalised series, we identified an ARIMA(1,1,1), amongst all the compared models with the series adjusted for seasonality. This model has highly significant coefficients and fares better over ARIMA(1,1,1)(0,1,1) in terms of its forecast power as shown by the marginally lower mean absolute error and the residual mean squared error.

Bell and Hillmer (1984) state that seasonal adjustment is done to simplify data so that they may be more easily interpreted by statistically unsophisticated users without a significant loss of information. Grether and Marlove (1970), further argue that consumers of seasonally adjusted series are not clear about the use of such and that those who give it extensive thought often finish by becoming hopelessly confused by using such a series. Also too many people fall in to the trap of ignoring seasonality if they are working with de-seasonalized or seasonally adjusted series, Hatanaka (1996); Enders (1995). These findings imply that it is wise to avoid using a seasonally adjusted data, but instead models that capture seasonality such as SARIMA models. We therefore present ARIMA(1,1,1)(0,1,1)[12] as our preferred model and we further asses its suitability by the residual diagnostics.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The expanded form of the identified ARIMA(1,1,1)(0,1,1)[12] is as follows:

$$\theta_P(L^S)\vartheta_p(L)(1-L)^d(1-L^S)^D y_t = \mu_q(L^S)\phi_q(L)\varepsilon_t$$

by substituting for $p = 1$, $d = 1$, $q = 1$, $P = 0$, $D = 1$ and $Q = 1$ in to (13), gives

$$\theta_1(L)(1-L)^1(1-L^{12})y_t = \mu_1(L^{12})\phi_1(L)\varepsilon_t$$

where $\vartheta_1(L) = (1 - \vartheta_1 L)$ is the non-seasonal AR order 1, $\phi_1(L) = (1 - \phi_1 L)$ is the non-seasonal MA order 1, $\mu_1(L^{12}) = (1 - \mu_{12}(L^{12}))$ is the MA seasonal operator at seasonal period 12.

Upon substituting for the lag operators and the coefficients, (13) can be expanded into

$$\begin{aligned} y_t &= \vartheta_1 y_{t-1} - \mu_{12}\varepsilon_{t-12} - \phi_1\varepsilon_{t-1} + \phi_1\mu_{12}\varepsilon_{t-13} + \varepsilon_t \quad \varepsilon_t \sim N(0,\sigma^2) \\ fit(y_t) &= 0.378047y_{t-1} + 0.999\varepsilon_{t-12} + 0.9685\varepsilon_{t-1} + 0.9675\varepsilon_{t-13} \end{aligned}$$

Table 4. Fitted Model - ARIMA(1,1,1)

	ar1	ma1		
Coefficients	0.3890	-0.9773		
s.e.	0.0419	0.0163		
$(\hat{\sigma}^2 = 2.167)$	$\log(L) = -1168.36$	AIC=2342.72	AICc=2342.76	BIC=2356.14
z test of coefficients				
	Estimate	Std. Error	z-value	$Pr(>\lvert z\rvert)$
ar1	0.389027	0.041932	9.2775	$< 2.2e-16$ [a]
ma1	-0.977250	0.016261	-60.0993	$< 2.2e-16$ [b]

[a] significant code at 0.000
[b] significant code at 0.000

Table 5. Fitted Model ARIMA(1,1,1)(0,1,1)[12]

	ar1	ma1	sma1	
Coefficients	0.378	-0.9685	-1.0000	
s.e.	0.044	0.0201	0.0481	
$\hat{\sigma}^2 = 2.213$	log(L)=-1177.76	AIC=2363.52	AICc=2363.58	BIC=2381.33
z-test of coefficients				
	Estimate	Std. Error	z-value	$Pr(>\lvert z\rvert)$
ar1	0.378047	0.044009	8.5902	$< 2.2e-16$ [a]
ma1	-0.968455	0.020113	-48.1509	$< 2.2e-16$ [b]
sma1	-0.999998	0.048084	-20.7970	$< 2.2e-16$ [c]

[a] significant code at 0.000
[b] significant code at 0.000
[c] significant code at 0.000

The interaction term $\phi_1\mu_{12}$ allows the model to be parsimonious. Had we fitted an additive model, we could have had an unconstraint, coefficient of MA at lag 13, and this would have led to an over parametrized model.

5.3. Residual Diagnostics

Figure 5 gives the results of the diagnostic analysis. The histogram of residuals resembles a bell shaped curve, indicating that the normality assumption on the error terms is not violated. The QQ plot shows that our sample data plotted against the quantiles computed from a theoretical standard normal distribution follows its behavior of a fairly straight line. The plot of the residuals against the fitted values show a random scatter around the zero line, safe for the concentration of the fitted values at 22, 25 and 32 points, and it also shows a fairly horizontal band about that line showing that the variance of the errors is constant. The auto correlation function shows an immediate decay to zero, a condition consistent with stationary processes. Overall our plots point to a good model fit.

5.4. Forecasts

Figure 5 shows the forecasts together with the forecast confidence band. It shows that our forecasts mimic the past behavior of the series. Table 6 shows that our forecasted values are highly significant with an 80 to 90% confidence level.

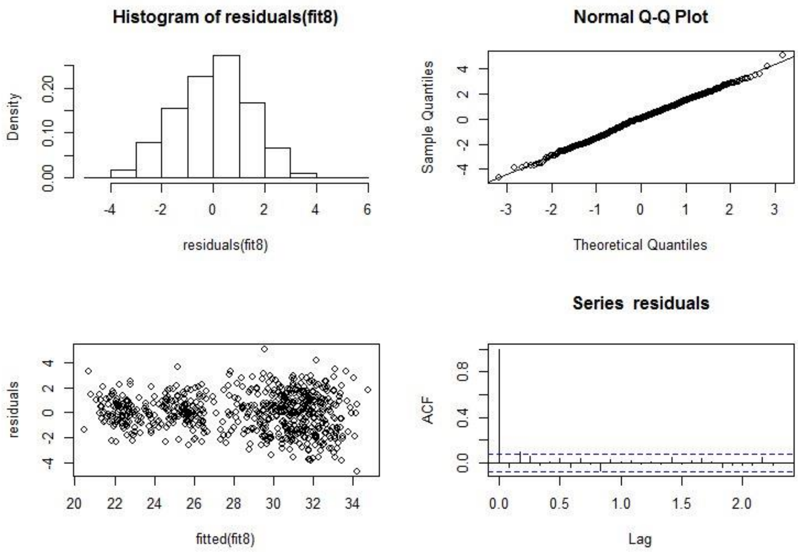


Fig. 5. Residual Diagnostics

Table 6. Forecast

Jan 2014	31.02735	29.10302	32.95167	28.08435	33.97034
Feb 2014	30.78203	28.70248	32.86157	27.60163	33.96242
Mar 2014	29.60049	27.49019	31.71079	26.37306	32.82791
Apr 2014	27.03743	24.91800	29.15686	23.79604	30.27882
May 2014	24.67792	22.55419	26.80165	21.42995	27.92588
Jun 2014	21.85014	19.72347	23.97680	18.59768	25.10259
Jul 2014	21.98954	19.86038	24.11870	18.73327	25.24580
Aug 2014	25.05969	22.92820	27.19118	21.79986	28.31952
Sep 2014	29.10115	26.96739	31.23491	25.83784	32.36445
Oct 2014	30.68132	28.54531	32.81733	27.41457	33.94807
Nov 2014	31.08225	28.94398	33.22051	27.81206	34.35244
Dec 2014	31.56043	29.41991	33.70095	28.28679	34.83407
Jan 2015	31.87609	29.73152	34.02066	28.59625	35.15593
Feb 2015	31.09709	28.94960	33.24458	27.81278	34.38140
Mar 2015	29.71379	27.56378	31.86381	26.42563	33.00196
Apr 2015	27.07447	24.92207	29.22686	23.78266	30.36627
May 2015	24.68611	22.53140	26.84083	21.39076	27.98147
Jun 2015	21.84743	19.69041	24.00445	18.54856	25.14631
Jul 2015	21.98271	19.82340	24.14202	18.68033	25.28509
Aug 2015	25.05131	22.88971	27.21290	21.74543	28.35718
Sep 2015	29.09218	26.92829	31.25606	25.78280	32.40155
Oct 2015	30.67213	28.50596	32.83830	27.35925	33.98500
Nov 2015	31.07297	28.90450	33.24144	27.75658	34.38936
Dec 2015	31.55112	29.38034	33.72190	28.23119	34.87105

6. Conclusion

The average maximum temperatures in Gaborone observed at the meteorology Department from January 1960 to December 2013 were studied with the intention of identifying the best time series model that can be used for forecasting. The series was found to be highly seasonal with peaks observed around the periods of December to January. It was found that seasonally adjusting the series prior to applying the Box and Jenkins procedure for identifying the best model did not lead to the seasonal effects dying out, despite giving a fairly good ARIMA(1,1,1). It is argued by many authors that it is wise to avoid using a seasonally adjusted data, but instead models that capture seasonality such as SARIMA models. This prompted fitting a seasonal auto regressive moving average process, ARIMA(p,d.q)(P,D,Q). Seasonality was captured at lags s=12, and multiple lags $s, 2s, 3s, \ldots$. The autocorrelation function and the partial autocorrelation function, Dickey Fuller tests of stationarity and other model comparison methods led to an ARIMA(1,1,1)(0,1,1)[12].The seasonal multiplicative ARIMA was found to be parsimonious as compared to an additive seasonal ARIMA which would have been over parametrized. The identified model gave a good fit and can be used for short term forecasts.

References

Bai J. and Perron P.(2003) Critical values for multiple structural change tests. *The Econometrics Journal*. Vol 6 (1), pp.72-78.

Baillie R. T. and Chung S. K.(2002) Modeling and forecasting from trend stationary long memory models with applications to climatology. *International Journal of Forecasting*. Vol 18, pp.215-226.

Bell R. W. and Hillmer C. S.(1984) Issues Involved With the Seasonal Adjustment of Economic Time Series. *Journal of Business & Economic Statistics*. Vol 2, pp.291-320.

Box G. E. P. and Jenkins G. M.(1976) The Analysis of Time Series: An Introduction. Forecasting and Control. *Holden-Day*. San Francisco. Revised Edition.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., and Jones, P. D., 2006. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysics*, 93, 34-38, Res.111 , D12106, doi:10.1029/2005JD006548

Chatfield C. (1980) The Analysis of Time Series-An Introduction. *Chapman and Hall CRC, London*. Fifth edition.

Cline W. R.(1992) The Economics of Global warning. *Institute for International Economics*.

Cooke.(1980) Settlements in Botswana. *Heinemann Educational Books Ltd*. Published in collaboration with Botswana Society.

Cooke R. D.(1980) Influential observations in linear regression. *Journal of American Statistical Association*, 74, pp.169-174.

Craft.(2001) An economic history of weather forecasting. *Journal of the Royal Statistical Society Series B*. URL: http://eh.net/article/craft.weather.forecasting.history.

Ender, W.(1995) Applied Econometric Time Series. *John Wiley and Sons, INC.*

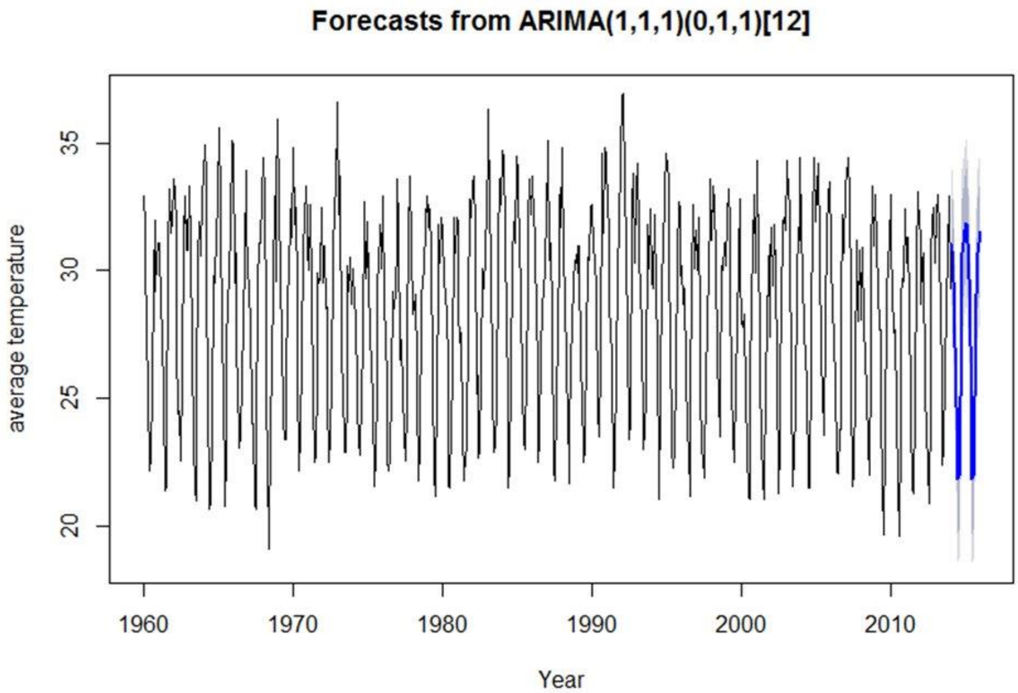


Fig. 6. Time series forecasting plot of Temperature

Fako T. T. and Molamu L.(1995) The Seven-Year Drought, Household Food Security and Vulnerable Groups in Botswana. *Pula Journal*. Vol 9 (2), pp.48-70.

Geweke, J., Messe, R., and Dent, W., 1983. Comparing alternative tests of causality in temporal systems, Analytic results and experimental evidence. *Journal of Econometrics*, 21, 161-194.

Goldhaber, D., Grout, C., and Huntington-Klein, N., 2014. Screen Twice, Cut Once: Assessing the Predictive Validity of Teacher Selection Tools. Working Paper 120. *National Center for Analysis of Longitudinal Data in Education Research (CALDER).*

Grether D. M. and Marlove M.(1970) Some properties of optimal seasonal adjustment. *Econometrica*. Vol 38, pp.683-703.

Hatanaka M. (1996) Time-series-based econometrics, unit roots and cointegration. *Oxford University Press*.

Johnston J. (1984) Econometric Methods. *International edition, McGraw-Hill, Inc. Singapore*. Third edition.

Jury, M. R., Mulenga, H. M., and Mason, S. J., 1999. Exploratory Long-Range Models to Estimate Summer Climate Variability over Southern Africa. *Journal of Climate*, 12, 1892-189.

Jury, M. R., Weeks, S., Gondwe, M. P., 1997. Satellite-observed vegetation as an indicator of climate variability over southern Africa. *South. Afr. J Science*, 93, 34-38.

Ljung G. M. and Box G. E. P.(1987) On a measure of lack of fit in time series models. *Biometrika*. Vol 65.

Makridakis, C., Wheelwright, S. C., and Hyndman, R. J., 2003. Forecasting: Methods and Applications. *John Wiley & Sons*.

Mann M. E.(2008) Smoothing of Climate time series revisited. *Geophysical Research Letters*. Vol 35, pp.1-5.

NDP8.(1997 - 2003) National Development Plan 8. *Government of Botswana Printer*.

Parvinder, M., Kumar, A., Rathore, L.S., and Singh, S.V., 2003. Forecasting Maximum and Minimum Temperatures by Statistical Interpretation of Numerical Weather Prediction Model Output. *Weather and Forecasting*, 18, 938-952.

Perron P.(1997) Further Evidence on Breaking Trend Functions in Macroeconomic Variables. *Journal of Econometrics*. Vol 80 (2), pp.355-385.

Perron P.(1989) The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis. *Econometrica, Econometric Society*. Vol 57 (6), pp.1361-1401.

Pindyck R. S. and Rubinfeld D. L.(1981) Econometric Models and Economic Forecasts. *McGraw-Hill, New York*.

Posamentier H. W. and Nicolich M. J.(1979) Time series transfer model relating temperature to behavior of Glacier Termini. *Mathematical geology*. Vol 11, pp.691-699.

Quenouille M. H.(1949) Approximate Tests of Correlation in Time Series. *Journal of the Royal Statistical Society Series B*, 11, pp.68-84.

Rivera, P. M.(1990) Analysis of a cross section of time series using structural time series models. *Unpublished PHD thesis*. London School of Economics.

Romilly P.(2005) Time series modeling of global mean temperature for managerial decision making. *Journal of environmental management*, 76, pp.61-70.

Selitshena R. M. K. and McLeod G.(1989) Botswna - A Physical, Social and Economic Geography. *Longman Botswana*.

Shangodoyin D. K.(2008) On the selection of best subset autoregressive models in prediction with application to real life data. *Statistics in Transition*. Vol 9 (3), pp.471-484.

Shangodoyin, D. K., Setlhare, K., Sediakgotla, K.,and Moseki, K., 2010. On the appropriate transformation technique and model selection in forecasting economic time series: An application to Botswana GDP data. *Journal of modern applied statistical methods*. Vol.9 (1), pp.287-295.

Zivot E. and Andrews W. K.(1992) Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business & Economic Statistics*. Vol 10 (3), pp.251-270.