Afrika Statistika

ISSN 2316-090X

Afrika Statistika
Vol. 13 (1), 2018, pages 1581–1592.
DOI: http://dx.doi.org/10.16929/as/1581.122

Box-Jenkins Analysis of Mean Monthly Temperature in Rwanda

J. D. Maniraguha[1]*, and Emelyse Umunoza Gasana

Department of Mathematics, College of Science and Technology, University of Rwanda, P.O. Box 3900 Kigali, Rwanda.

Selected Papers presented at the East African Conference of Statistical Mathematics and Applications, EACSMA 2017

Received September 10, 2017; Accepted April 9, 2018

Abstract. Generally, temperature is a sensation of coldness and hotness and is affected by air and humidity as climate factors in certain areas. The high altitude of Rwanda provides the country with pleasant tropical highland climate. The recent research done on climate changes and global warming found an increase in temperature of 1 to 2 degrees Celsius between year 2000 and year 2050 due to the natural variability and due to the green house gases that trap the infrared radiation,i.e., these gases will cause the temperature of the earth to increases infrared radiation from the earth surface and reflect it back to the earth. This paper has the general aim of analyze, interpret and forecast the data from Meteo-Rwanda of mean monthly temperature from year 2004-2016 to 2022, from Gitega station in Rwanda. The results of this paper show that there is no true difference in medians of monthly temperature from the taken station. .

Key words: Times series analysis; SARIMA models; temperature forecast
AMS 2010 Mathematics Subject Classification : 62M10; 62M20

*Corresponding author Jean de Dieu Maniraguha : jadomanir99@gmail.com
Emelyne Umunoza Gasana : emygasana@gmail.com

Résumé (French) Généralement, la température est une sensation de froid et de chaleur et, est affectée par l'air et l'humidité en tant que facteurs climatiques dans certaines zones. La haute altitude du Rwanda fournit au pays un climat tropical agréable. Les recherches récentes sur les changements climatiques et le réchauffement climatique ont révélé une augmentation de la température de 1 à 2 degrés Celsius entre 2000 et 2050 en raison de la variabilité naturelle et des gaz à effet de serre qui emprisonnent le rayonnement infrarouge, c-à-d la température de la Terre augmente le rayonnement infrarouge de la surface de la Terre et la renvoie vers la Terre. Cet article a pour objectif général d'analyser, d'interpréter et de prévoir les données de Météo-Rwanda de la température mensuelle moyenne de l'année 2004-2016 à 2022, depuis la station de Gitega au Rwanda. Les résultats de cet article montrent qu'il n'y a pas de différence réelle dans les médianes de la température mensuelle de la station prise.

1. Introduction

Climate change and global warming constitute an important environmental aspect today. Temperature forecasts are very important in assessing such changes. Thus, time series analysis on mean monthly temperature in Rwanda will help the various stakeholder and the government to plan in advance in order to account the climate related disasters and analysis when needed, in order to anticipate or assess the future variations in temperatures in Rwanda. Therefore, in this study, a Box-Jenkins model approach is appropriate and hence used to capture variations within mean monthly temperature series from Gitega station in Rwanda.

To understand the nature and scale of possible climate change in Rwanda using the temperature data as one variable of climate from the meteorology unit in Rwanda has importance in many cases like agriculture production and in human health. Besides, climate change/global warming are currently among the most important points of debate all over the world and the temperature is one of the factors that determine a climate change or global warming. This study will be a very important tool to any person who intends to eradicate the problem of the variation of climate variables in Rwanda. The importance of this research is to forecast temperature with purpose to inform the various sectors concerned by this issue directly or indirectly. Forecasting the variations in temperature is important in different areas, for instance, the very high temperature may predispose plants, live stock and human to heat related disasters.

In this paper we will use the Seasonal Autoregressive Integrated Moving Average (SARIMA) models which are adaptation of autoregressive integrated moving average for ARIMA model to specify the fitted seasonal time series data. The purpose of constructing it based on seasonal nature of series to be modelled. The SARIMA can be written as $ARIMA(p, d, q)X(P, D, Q)_s$ where the upper case of the above model represents that there is seasonal parameters and lower case represents the non-seasonal part of the model, where p is the number of non-seasonal autoregressive (AR) term, d the number of non-seasonal differencing, q the number of non-seasonal moving average (MA), s the number of periods per season, P the number of seasonal autoregressive Autoregressive (SAR)term, D the number of seasonal difference and Q the number of seasonal moving average (SMA) (Box et al. C, 1978; Afrifa-Yamoah, 2015).

If the series is seasonal with s periods per years then a seasonal ARIMA (SARIMA) model can be written as

$$\Phi_P(B^s)\Phi_p(B)(1-B)^d \cdot (1-B^s)^D \cdot (Y_t - \mu) = \theta_q(B) \cdot \Theta_Q(B^s)\varepsilon_t, \tag{1}$$

where $\Phi_P(B^s) = 1 - \Phi_1 \cdot B^s - \Phi_2 B^{2s} - \ldots - \Phi_p B^{Ps}$ with $\Phi_p(B) = 1 - \Phi_1 B - \Phi_2 \cdot B^2 - \ldots - \Phi_p \cdot B^p$ for $\Phi_p \neq 0$, and $\theta_q(B) = 1 - \theta_1 \cdot B - \theta_2 \cdot B^2 - \ldots - \theta_q \cdot B^q$, with $\Theta_Q(B^s) = 1 - \Theta_1 \cdot B^s - \Theta_2 \cdot B^{2s} - \ldots - \Theta_Q \cdot B^{Qs}$ for $\theta_q(B) \neq 0$ where Φ and Θ denotes the polynomials B^s of P, Q respectively (Brockwell and Davis , 2013).

One model polynomial useful for seasonal data is SARIMA model of order $(0,1,1)X(0,1,1)_s$ for monthly data $s = 12$. This model can be given as

$$(1-B)(1-B^{12}) = (1+\theta(B))(1+\Theta(B^2)) \tag{2}$$

and has been used by Box et al. C (1978) with the purpose of modelling airline data. For climate data which usually follows seasonal variations, to identify a perfect ARIMA model for a particular time series, Box-Jenkins proposed a methodology that consists of four phases, namely, model identification, estimation of the model parameters, diagnostic checking for the identified model appropriateness and forecasting.

The hypothesis to be tested in this study, for the variations of mean monthly temperature are :

H_0: Average monthly temperature for all months are equal, i.e., $\mu_{2004} = \mu_{2005} = \ldots. = \mu_{2016}$,

versus

H_1: At least two means are different (μ_i, μ_j) where $i \neq j$, starting from 2004 up to 2016.

Furthermore, the general objective of this study is to choose an appropriate time series model that best captures variations within the temperature for the years 2004-2016 and hence forecast mean monthly temperature in Rwanda in the next 6 years. Hence, we will

- Fit the appropriate time series model of mean monthly temperature in Rwanda.
- Based on the fitted model, forecast monthly amount of temperature in Rwanda for the years 2017 to 2022.
- Compare the average monthly temperature variations within the Gitega station.

2. Background

Time series analysis and forecasting has become a major tool in numerous applications in meteorology and other environmental areas to understand phenomena, like rainfall, humidity, temperature, draught etc. In their study, Sultana and Hasan (2015), using the classical Box-Jenkins methodology, stationary, seasonal ARIMA models for the temperatures recorded at two stations in Sylhet division and Moulvibazar districts in north-east

Bangladesh between 1977 and 2011 were set up. Verification of the models has been done for the 2010 to 2012 time period. Based on the 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 $ARIMA(1,1,1)X(1,1,1)_{12}$ and $ARIMA(1,1,1)X(0,1,1)_{12}$, respectively, were obtained, whereas the respective models for the Moulvibazar station were $ARIMA(1,1,0)X(1,1,1)_{12}$ and $ARIMA(0,1,1)X(1,1,1)_{12}$. Using these ARIMA-models one-month ahead forecasts of the temperatures at the two stations for years 2010 and 2011 were carried out (Sultana and Hasan, 2015).

On the other hand, Shamsnia et al. (2011) considered 20 years of relative humidity, monthly average temperature and precipitation of Abadeh region in Iran. The results obtained suggested a $ARIMA(0,0,1)X(1,1,1)_{12}$ for precipitation, a $ARIMA(2,1,0)X(2,1,0)_{12}$ and monthly average temperature and $ARIMA(2,1,1)X(1,1,0)_{12}$ for relative humidity.
In another study, Tembo and Chikondi (2015) forecast the climate change pattern for the Bolero agriculture extension area in Malawi, based on temperature data from 1982 to 2013 in order to inform the policy makers and community on the future prospects of climate change and its effects, Tembo and Chikondi (2015) used univariate autoregressive integrated average to model and forecast temperature variation. Based on the ARIMA and its components autocorrelation function, Normalized Bayesian information criteria (BIC), Box-Jenkins statistics and residuals estimate, $ARIMA(1,1,3)$ were selected for maximum temperature data. They conclude that forecasted maximum temperature will increase $1.6°C$ from $27.7°C$ in 1982 to $29.3°C$, in 2030. Therefore, the increase of temperature suggest that climate change could continue to negatively impact on the agriculture.

John et al. (2014) have examined monthly maximum temperature, for year 1977 – 2012, of the South Eastern area of Nigeria. Based on first time series plots and ACF and PACF, and by the principle of parsimony, John et al. (2014) concluded that the $SARIMA(0,0,2)X(2,1,1)_{12}$ model were adequate for the series and monthly forecasts from 2013 to 2017 were performed.

In the research by Ochanga (2016), a time series SARIMA model was built in order to analyse and forecast monthly maximum and minimum air temperature of Nairobi. Based on PACF and ACF plots and using AIC criteria, the best forecasting SARIMA model for maximum temperature were $(0,0,2)X(0,1,1)_{12}$ and that for the minimum temperature were $(1,0,0)X(0,1,1)_{12}$. Ochanga (2016) found that the minimum temperature increase gradually over the years, hence proving that there was fact of global warming.

3. Analysis and results

Data used in this research are secondary data provided by Meteo-Rwanda which is located at Gitega sector in Nyarugenge district. The data constitute mean monthly temperature series from Gitega, from 2009 - 2016. They make 96 observations. They are then analyzed, modeled and interpreted, and forecasted in the next 6 years (2017 to 2022).

First, time series data were plotted without any transformation, to identify whether there is a significant trend in series, then stationary conditions of the series are checked and it is very important to identify whether there are seasonal variations in the data.

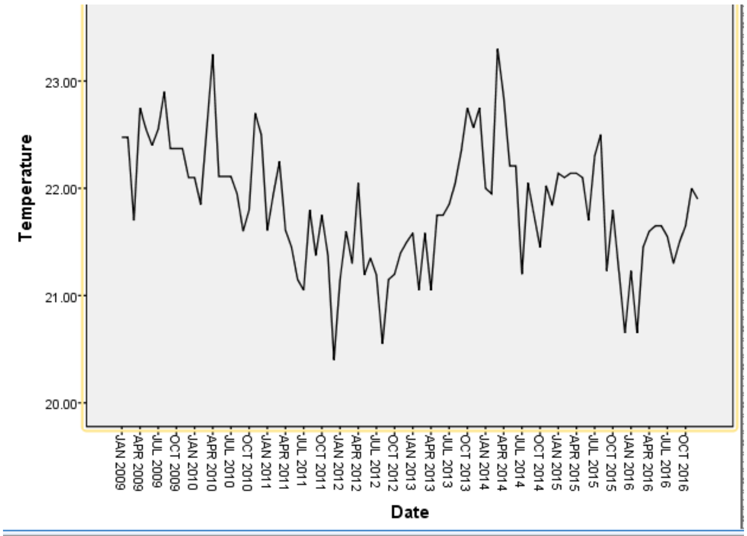

Fig. 1. Time series plots for with transformation

From Figure 1, the series are observed to be non-stationary. They all present a partial trend, i.e., that the mean level is not constant. The data are not stationary in mean and in variance. In addition, it is also necessary to check whether there are seasonal variations or not. If yes, the series need to be transformed into stationary series. The data are said to be stationary in mean and in variance when it fluctuates around mean zero and has no regular pattern, i.e., that mean and variance are taken as constant. Let us now check the correlograms of the original data to confirm non-stationarity.
The correlogram in Figure 2 show that not all spikes at all lags are in between the confidence limits, there are some spikes at some lags. Hence we can conclude that the series are not stationary.

The Box-Jenkins approach requires stationarity of the variables for model identification and selection. The data is hence transformed (see Figure 3) with the purpose of inducing stationarity in mean and variance.

The following are plots obtained by combining the three transformations; nature log, differencing and seasonal differencing transformations.
Figure 3 shows that there is no trend, it is assumed that there is no systematic change in mean and in variance, it means that the mean and variance are approximately constant and no periodic variations are observed.
The next step will be to fit the ARIMA model. The Box and Jenkins in 1970 were first to

approach the test of estimation of ARMA model in systematic manner; model identification, estimation or model fitting, the model diagnostic checking and the forecast (Box et al. C, 1978).

(a) Identification of the model.

After making the data stationary,the model is identified by finding the most number of autocorrelation which are significant at any lags. At the model identification stage, the goal is to identify seasonality, if it exists and identify the orders of AR and MA using the ACF and PACF plots (see Figure 4). The following correlograms are produced:
Figure 4, shows that the transformation of original data, at lag 1, 8, 11, 12 and 13 spikes are significant but more significant at lag 1. The significance at lag 12 represents a seasonal component, and at lag 1, the condition of stationary. By using the AIC of selecting the order of model the following orders are obtained $p = 1$, $d = 1$, $q = 1$, $P = 0$, $D = 1$, and $Q = 1$, then the general model can be expressed as $SARIMA(1, 1, 1)X(0, 1, 1)_{12}$.

(b) Model fitting.

Once potential models have been identified, the second step is to fit the model, but here the model solution is not simple, it requires an iterative process and models must also have a few parameters as much as possible, (Box et al. C, 1978). The slowest, but most accurate method is that termed maximum likelihood as we know that there are other methods of estimating like method of moment, General Least Square (GLS).

Therefore the fitted model for the Gitega station is :

$$\hat{y}_t = \mu + y_{t-12} + \alpha_1 y_{t-1} - \theta \cdot \varepsilon_{t-1} - \Phi \cdot \varepsilon_{t-13} - \theta \cdot \Phi \cdot \varepsilon_{t-1}\varepsilon_{t-13} + \varepsilon_t \tag{3}$$

Table 1: Fitted parameter estimates of Gitega station

As shown in the ARIMA model parameters in Table 1, the model fit of Gitega station series is:

$$\hat{y}_t = 0.005 + y_{t-12} + 0.194 y_{t-1} - 0.695\varepsilon_{t-1} - 0.976\varepsilon_{t-13} - 0.678\varepsilon_{t-1} \cdot \varepsilon_{t-13} + \varepsilon_t \tag{4}$$

(c) Model checking.

The following procedures are used in model checking:

- Plot the multiple plot of original and fitted observations and check the following: If fitted values follow the behaviour of original data, if they are close together, if the fitted capture the variations within the original data.
- Plot the residuals against time, their ACF and PACF, test the noise on the selected model.
- Check if the histogram of residuals are approximately normally distributed.
- Check if there is no autocorrelation between residuals.

Now, to start the model adequacy check, let us plot the original data with their corresponding fitted values together.

Figure 5 shows that the fitted values of Gitega station are close enough, move together with the original data, so the fitted values follow the behaviour of original and capture the variations within the original data. This shows that the model fit is good. Next, we are going to check if the histograms of residuals are approximately normally distributed.

From Figure 6, the histogram of residuals of Gitega superimposed by normal curve on mean monthly shows that the curve is normally distributed with mean 0 and variance of $(0.501)^2 \approx 0.25$. Now, we are going to plot the residuals against time with target of checking if residuals are stationary. As shown in Figure 7, the residuals present no regular pattern;

they are stationary since they fluctuate around mean zero and with constant variance. To check clearly this condition of stationarity, let us look at the autocorrelation plot of the residuals, and then make a conclusion about the residuals' autocorrelation.

From Figure 8, all spikes lie between the confidence limits at any lag. In this case we conclude that the errors are uncorrelated.

From Figure 9, the rising trend shows that the temperature is likely to increase in the next 6 years. The forecasted mean monthly temperature in December 2022 will be $21.70°C$ with corresponding interval of $(18.28, 25.13)°C$ at Gitega station. Comparing to the original data at Gitega station, the minimum is $21.63°C$ and the maximum is $22°C$ and the forecast shows the minimum $20.4°C$ and the maximum will be $23.5°C$. This shows that the maximum mean monthly temperature will increase $1.5°C$ for the next 6 years.

References

Afrifa-Yamoah E.(2015). *Application of ARIMA Models in Forecasting Monthly Average Surface Temperature of Brong Ahafo Region of Ghana*. International Journal of Statistics and Applications, (**5**), pp. 237–246.

Black K.(2009). Business statistics: Contemporary decision making. *John Wiley & Sons*

Box G., Hilimer S., and Tiao G.c.(1978). *Analysis and modeling of seasonal time series*. In : Seasonal analysis of economic time series, pp. 309–344. NBER

Brockwell P. J and Davis R. A.(2013). Time series: theory and methods. *Springer Science & Business Media*

Ferdous A., Shahin M. A. and Ali M. A.(2014). *Time Series Model Building and Forecasting on Maximum Temperature Data*. In : Computational Intelligence Techniques in Earth and Environmental Sciences, pp. 79–91. Springer

John C., George U. and Chukwuemeka O. S.(2014). *Time Series Analysis and Forecasting of Monthly Maximum Temperatures in South Eastern Nigeria*. International Journal of Innovative Research and Development—— ISSN 2278–0211, (**1**), pp. 1129–1133.

Joshua M.T.(2010). *Forecasting and Time series notes*. Holden-Day

Maddala G. and Lahiri K.(1992). Introduction to econometrics. *Macmillan New York*

Nagpaul P.S.(2005). *Time series analysis in winidams*. New Delhi India, (**1**), pp. 79–91.

Nanga S.(1976). *Time Series Analysis of Road Accidents in Ghana*. Biometrika, (**1**), pp. 309–344.

Ochanda O. O.(2016). Time Series Analysis and Forecasting of Monthly Air Temperature Changes in Nairobi Kenya. PhD thesis, University of Nairobi.

Shamsnia S. A., Shahidi N., Liaghat A., Sarraf A. and Vahdat S. F.(2011). *Modeling of weather parameters using stochastic methods (ARIMA model) (case study: Abadeh Region Iran)*. In : International Conference on Environment and Industrial Innovation. IPCBEE. Iran, pp. 292–298. Scientific & Academic Publishing

Sultana N. and Hasan M. M.(2015). *Forecasting Temperature in the Coastal Area of Bay of Bengal - An Application of Box-Jenkins Seasonal ARIMA Model*. statistics, (**8**), pp. 157-163.

Tembo M. and Chikondi B.(2015). *Forecasting Climate Change Pattern for Bolero Agriculture Extension Planning Area in Malawi*. Journal of Climatology & Weather Forecasting, (**8**), pp. 1–5.

Zhang Y. and Yang Y.(2015). *Cross-validation for selecting a model selection procedure*. Journal of Econometrics, (**1**), pp. 95–112.

Zucchini W.(2000). *An introduction to model selection*. Journal of mathematical psychology, (**1**), pp. 41–61.

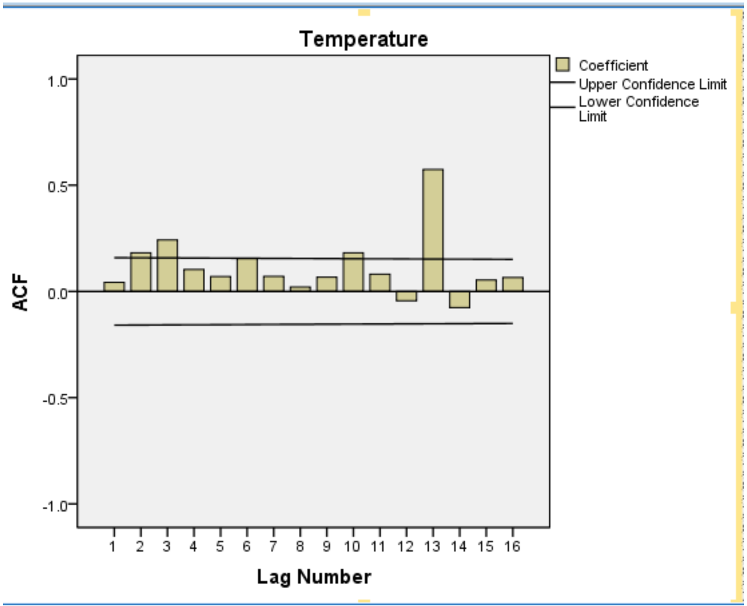

Fig. 2. ACFs of the original series

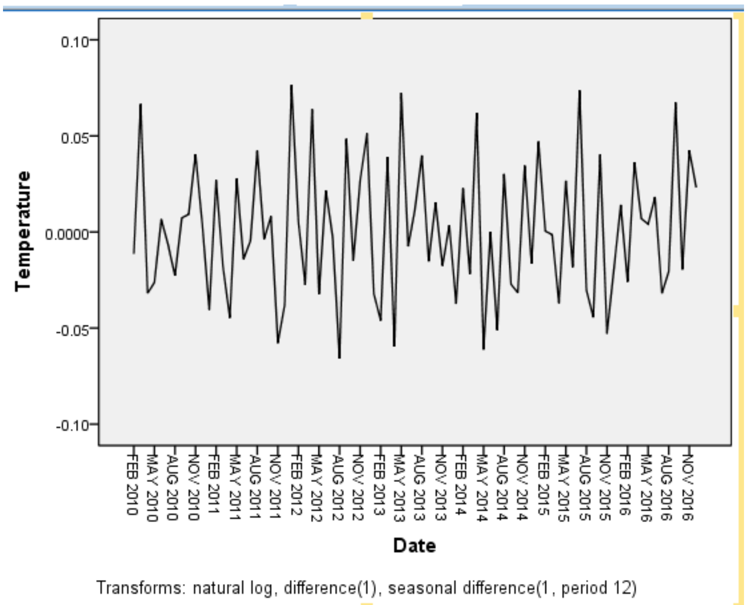

Fig. 3. Transformation of the original data

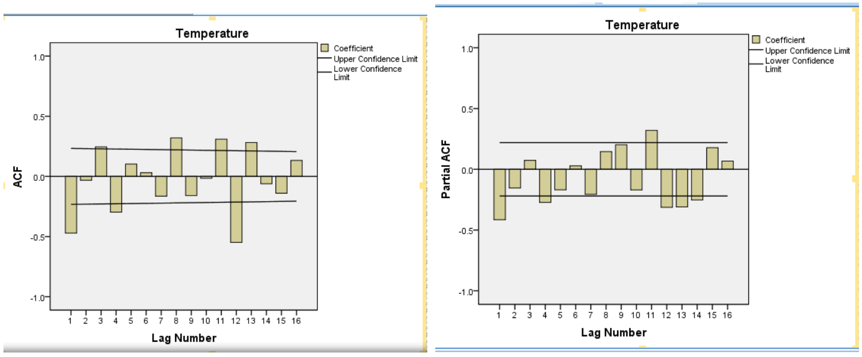

Fig. 4. The ACFs and PACFs of stationary series

ARIMA Model Parameters

					Estimate	SE	t	Sig.
Temperature-Model_1	Temperature	No Transformation	Constant		.005	.008	.603	.548
			AR	Lag 1	.194	.194	.999	.321
			Difference		1			
			MA	Lag 1	.695	.145	4.785	.000
			Seasonal Difference		1			
			MA, Seasonal	Lag 1	.976	2.332	.419	.677

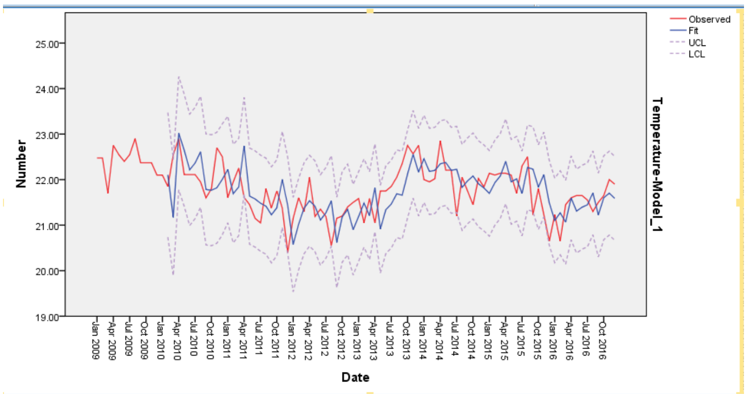

Fig. 5. Multiple plot of fitted and original data

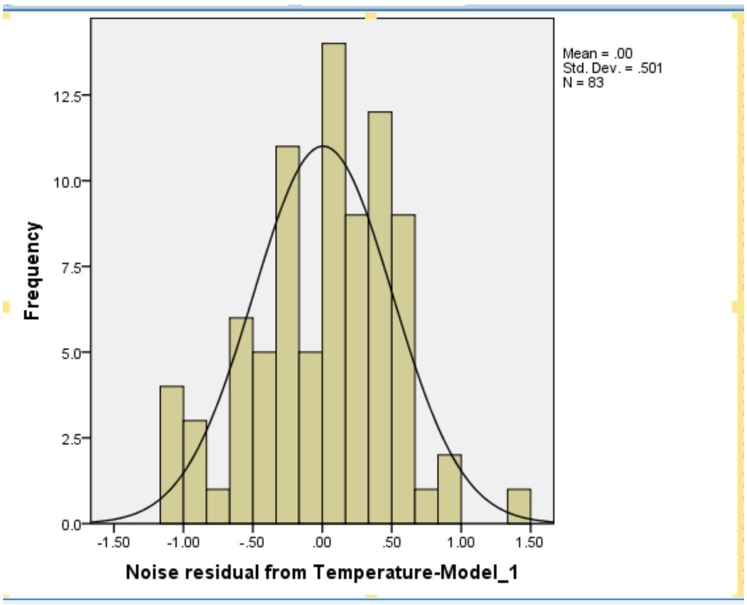

Fig. 6. Histograms of residuals

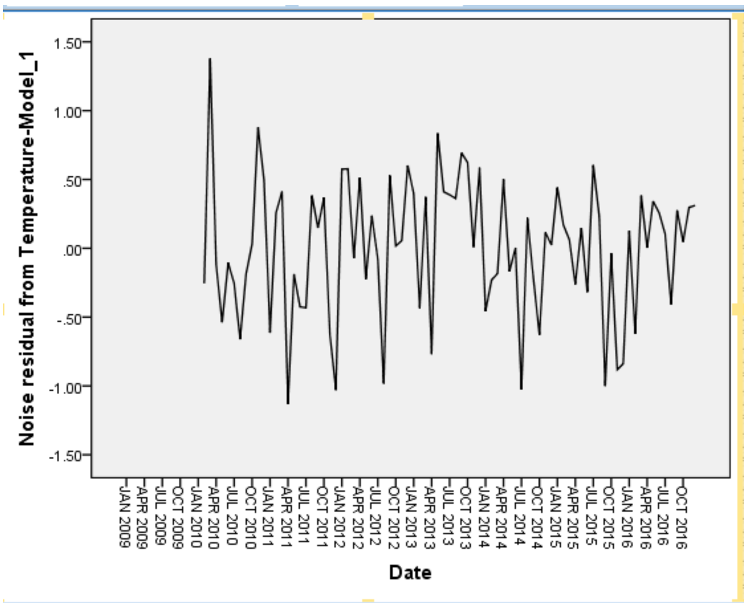

Fig. 7. Time sequence plot of residuals

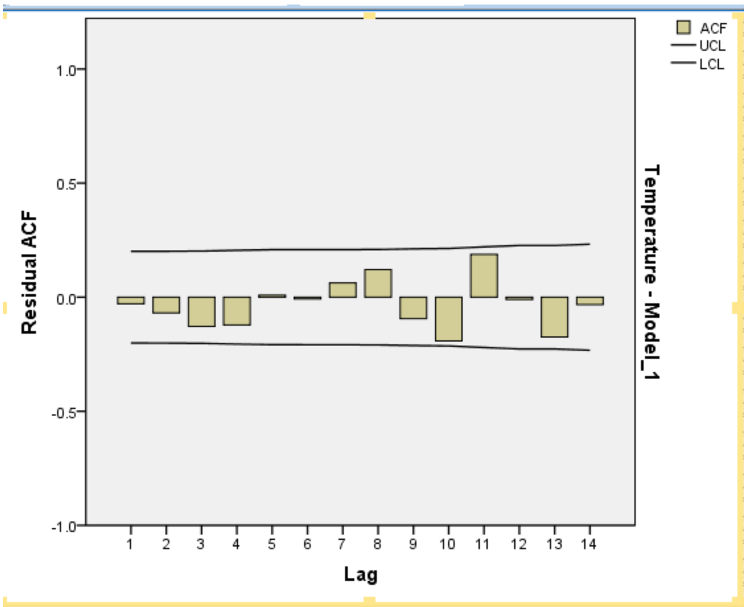

Fig. 8. The correlogram of ACF the residuals in Gitega station

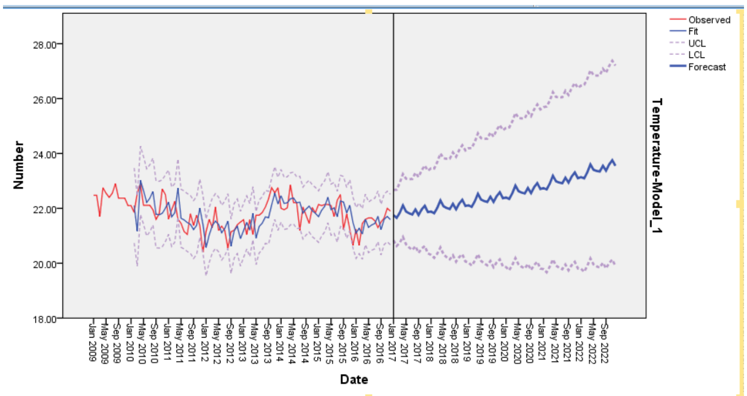

Fig. 9. Mean monthly temperature forecasts at Gitega station