Time Series Modelling and Forecasting of monthly Temperature data for western Kenya using seasonal ARIMA method

Alex Barasa Kakai
SCB384-C003-4369/2014

Jomo Kenyatta University of Agriculture and Technology

February 16, 2022

Supervisor: Prof. Joseph Mung'atu

Background study

- Currently, climate change marked by global warming has occurred. (T.M.Letcher, 2009)
- Global temperatures have increased by between 0.4^0C and $0.8\ ^0C$ in the past century and could rise by between $1.4\ ^0C$ and $5.8\ ^0C$ by the end of the 21^{st} century.(IPCC, 2013)
- Global warming has acquired the status of a key national policy challenge.(IPCC, 2013)
- Agriculture is an essential part of both the National Climate Change Implementation Framework and National Climate Change Action Plan (Government of Kenya, 2013), thus reflecting reliance on agriculture for the national economy.

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 (Government of Kenya, 2010), and on the other hand, possible opportunities for capitalizing on carbon funding (Government of Kenya, 2013)

General Objective and Specific Objectives

To analyse, model and forecast western Kenya's monthly minimum and maximum temperature recordings over a certain period of time to help predict future temperature trends.

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 future temperature trends using the best fitted model.

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.

Theoretical Review

In their early studies Box et al., 1970 generalized the ARIMA model to deal with seasonality by forming a multiplicative seasonal ARIMA model (SARIMA).This model even though 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.

Empirical Review

- Chang et al., 2012 applied seasonal ARIMA model to the time series monthly precipitation data from 1961 to 2011 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 predicted that the precipitation in the next three years for the region will decrease.
- Asamoah-Boaheng, 2014 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, MAPE values were evaluated.
- Ochanda, 2016 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.

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

A time series can possess a seasonal component that repeats every S observations, for monthly observations $S = 12$. The Seasonal ARIMA models are defined by seven parameters of the form :

$$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)Z_t \tag{1}$$

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

$$W_t = \nabla^d \nabla_s^D X_t \tag{2}$$

and we write

$$X_t \sim ARIMA(p, d, q) \tag{3}$$

Proposed Box and Jenkins Methodology

To identify a perfect ARIMA model for a particular time series, Box and Jenkins proposed a methodology that consists of four phases, namely:

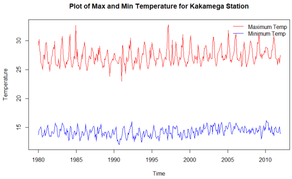
- Model identification
- Estimation of parameters
- Diagnostic checks
- Forecasting

Results and Discussions

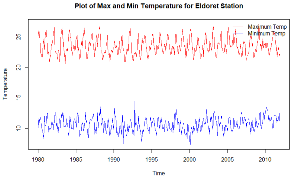
In this chapter a more in-depth analysis of the data is done. The chapter is organised in four sections with respect to the objectives of the study as follows:

- Data exploration using the time plot to check and uncover insights to help identify areas of interest or patterns in the data
- Model identification and selection for the monthly minimum and maximum temperatures for the two stations
- Properties of the selected temperature models identified
- Forecasting using the models selected

Data Exploration



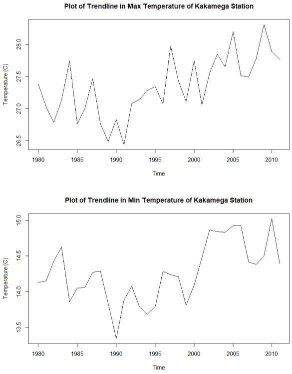
Kakamega Station



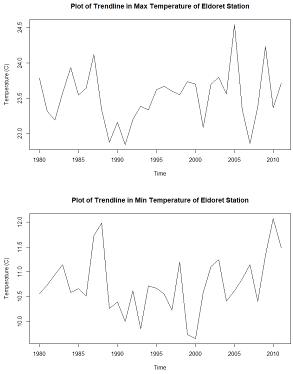
Eldoret Station

- Time plot of both maximum and minimum temperature of both Kakamega and Eldoret Stations.
- Maximum temperature of Kakamega is centred around a mean of 26^0C while that of Eldoret is centred at 24^0C
- Minimum temperature: Kakamega is centred around a mean of 14^0C while for Eldoret is centred around a mean of 10^0C

Trend Line Plot

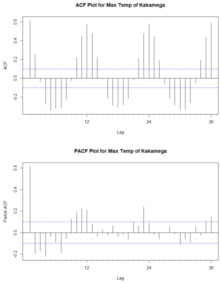


Kakamega Station

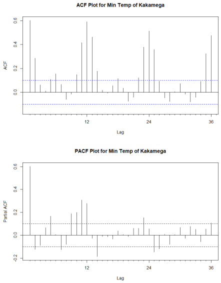


Eldoret Station

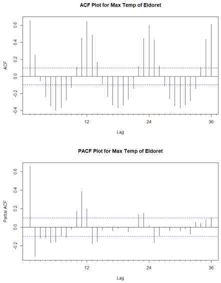
ACF/PACF Kakamega and Eldoret



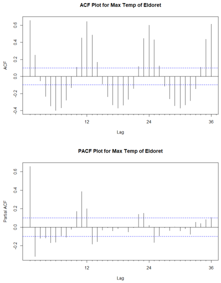
Maximum Temp



Minimum Temp



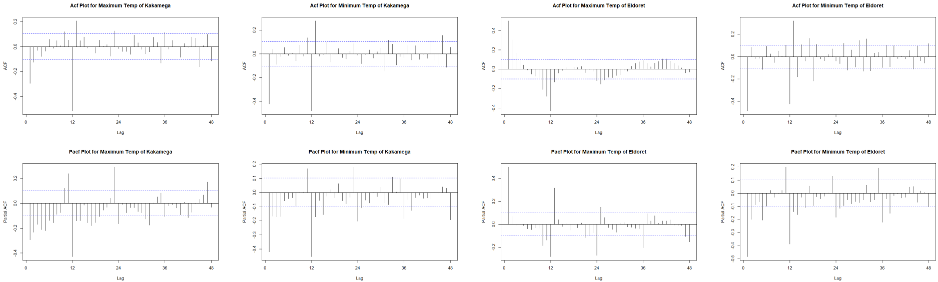
Maximum Temp



Maximum Temp

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 minimum and maximum temperature data respectively.

ACF/PACF of Differenced Data of Kakamega and Eldoret



Maximum Temp | Minimum Temp | Maximum Temp | Minimum Temp

After differencing, a p-value of less than 0.01 was achieved when both the Augmented Dickey-Fuller Test was performed, which is below 0.05 showing stationarity had been achieved.

Model identification and selection

Kakamega

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

Minimum Temperature

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

Maximum Temperature

Eldoret

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

Minimum Temperature

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

Maximum Temperature

Statistics for tentative 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

Kakamega Minimum Temperature

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

Eldoret Minimum Temperature

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

Kakamega Maximum 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

Eldoret Maximum Temperature

Parameter Estimation

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(1,0,1)(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	0.3604	0.0504
sma1	-0.9081	0.0323

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

Parameter	Estimate	Standard Error
ar1	0.5127	0.0446
sma1	-0.9322	0.0365

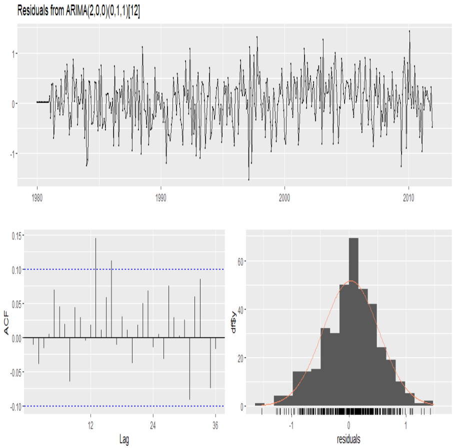
Box-Ljung Test

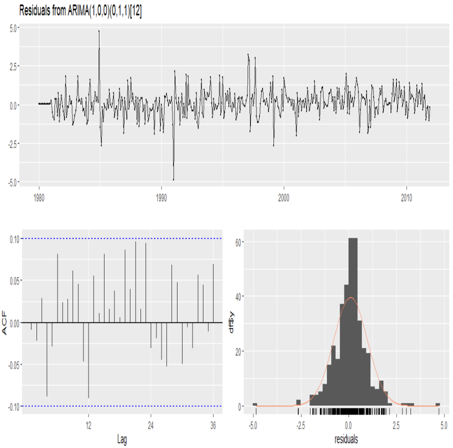
	X-squared	df	P Value
SARIMA(2,0,0)(0,1,1)$_{12}$	25.559	24	0.2238
SARIMA(1,0,0)(0,1,1)$_{12}$	29.873	24	0.1216

Box-Ljung Test

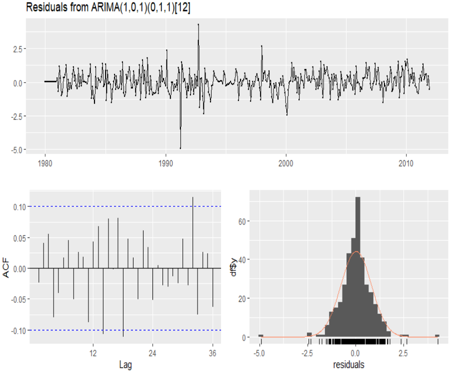
	X-squared	df	P Value
SARIMA(1,0,1)(0,1,1)$_{12}$	33.131	24	0.1013
SARIMA(1,0,0)(0,1,1)$_{12}$	14.824	24	0.9258

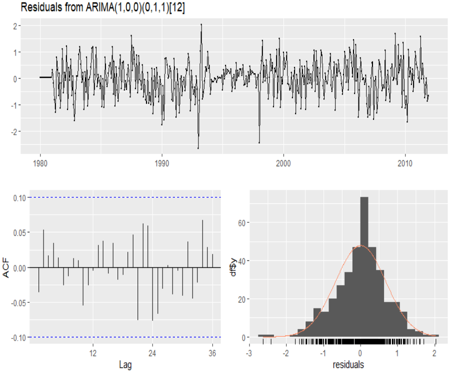
Diagnostic Analysis for Kakamega Station





Diagnostic Analysis for Eldoret Station





There's little 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.

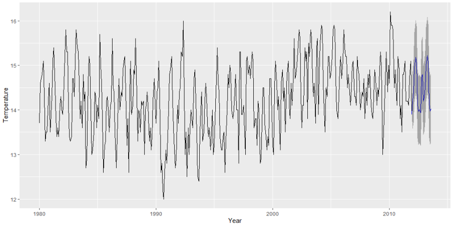
Forecasting: Kakamega Station

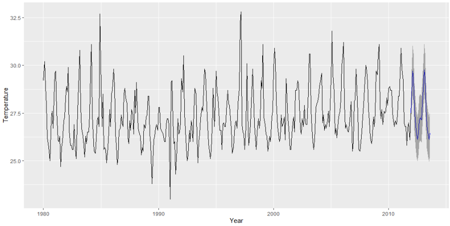
Minimum Temperature

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

Maximum Temperature

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

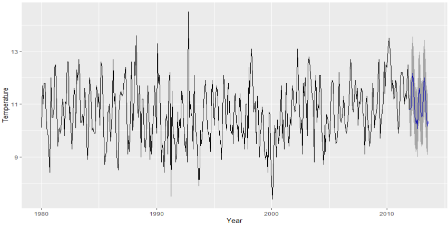


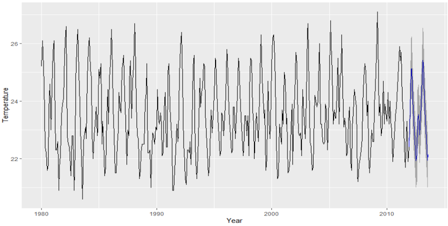


Forecasting: Eldoret Station

Minimum Temperature 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

Maximum Temperature 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





- Summary: The minimum temperature 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.
- 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.
- 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.

Asamoah-Boaheng, M. (2014). Using sarima to forecast monthly mean surface air temperature in the ashanti region of ghana. *4*, 292–298.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1970). *Forecasting, time series, and regression (forecasting, time series, and regression)*. Wiley.

Chang, X., Gao, M., Wang, Y., & Hou, X. (2012). Seasonal autoregressive integrated moving average model for precipitation time series. *8*, 500–505.

Government of Kenya. (2010). National climate change response strategy.

Government of Kenya. (2013). National climate change action plan.

IPCC. (2013). *Intergovernmental Panel on Climate Change*. institution.

Ochanda, O. O. (2016). *Time series analysis and forecasting of monthly air temperature changes in nairobi kenya* (Doctoral dissertation). University of Nairobi.

T.M.Letcher. (2009). *Climate change: Observed impacts on planet earth*. Elsevier Science. https://books.google.co.ke/books?id=x5huZ6SfNwQC

THANK YOU