
Fitting Time-series Models to Kisumu Rainfall Data for the Period 1961-2014

Mawora Thomas Mwakudisa, Edgar Ouko Otumba, Joyce Akinyi Otieno

Department of Statistics and Actuarial Science, School of Mathematics, Statistics and Actuarial Science, Maseno University, Maseno, Kenya

Email address:

tmawora@maseno.ac.ke (M. T. Mwakudisa)

To cite this article:

Mawora Thomas Mwakudisa, Edgar Ouko Otumba, Joyce Akinyi Otieno. Fitting Time-series Models to Kisumu Rainfall Data for the Period 1961-2014. *Mathematical Modelling and Applications*. Vol. 5, No. 1, 2020, pp. 39-46. doi: 10.11648/j.mma.20200501.14

Received: January 16, 2020; **Accepted:** February 6, 2020; **Published:** February 19, 2020

Abstract: Many small-scale farmers require adequate forecasts to help them plan for the rainfall. The National Meteorological Service provides forecasts seasonally, monthly and weekly. The forecasts are qualitative in nature hence inform, but cannot be directly used with decision support models. It is therefore important to consider forecast methods that researchers can use to generate quantitative data that can be applied in the models. In particular, an increasing need for forecasting daily rainfall data. In this study, the ARIMA and VAR models have been used to forecast five time period data for daily, monthly and seasonal rainfall data. The objective was to find the model parameters that best fit the three time periods. Fifty-year data from Kenya Meteorological Station, Kisumu, was used for the analysis. For each time period, five events were used as the test dataset. The ARIMA model was found to be best for forecasting daily rainfall in comparison to the VAR model, while SARIMA was best for monthly and seasonal data. One difference was done for the seasonal rainfall total, but not for monthly and monthly rainfall data. The VAR models included the available daily minimum and maximum temperatures. However, forecasted daily rainfall deviated from the test data, while monthly and seasonal data deviated even more.

Keywords: Arima, Sarima, VAR, Rainfall Data

1. Introduction

In Kenya, weather forecasting is the reserve of the Kenya Meteorological Services [1]. They monitor the oceanic wind movements, the Indian Ocean Dipoles, and predict the weather for seasons, months and weeks. Such forecasts are usually in qualitative form and focus on rainfall, cloud cover and sunrise. They use the common terms like “normal”, “above normal” and “below normal” to represent the intensity of rainfall in comparison to the long-term average. This is particularly useful for the small-scale farmers who depends on rain fed agriculture. In addition to weather forecast, researchers use mathematical Decision Support Tools (DSTs) to help farmers make informed decisions on risk involved. Such tools, most of which are crop models, require detailed information which are not provided in the KMS forecasts.

The Decision Support Tools (DSTs) are used by researchers across disciplines to help make informed decisions. They can be calculations on papers, apps or software. There are many of them and they cover all disciplines. For instance, in 2016, David Rose and colleagues catalogued 395 DSTs specific for

Agriculture [2]. The DSTs used mathematical equations and utilized available information to help make decisions. The models can be simplistic, for example Food and Agricultural Organizations (FAOs) CROPWAT [3]. This is a computer program for irrigation planning and management that utilizes the crop satisfaction index using ten-day (dekadal) data. Others like DSSAT [4] and APSIM [5] are more complex and require inputs like daily weather data and farm management options. The DSTs help researchers and farmers make future decisions using actual quantitative data. Most of the DSTs require a minimum of daily rainfall data in order to calculate water satisfaction indices for crops. Thus, consideration of forecast methods that can provide quantitative data is important. Time series is a main methodology used for forecasting quantitative data.

Time series analysis and forecasting utilizes the time lag to forecast the next few events in similar time gaps. One of the best parametric time series models is the ARIMA model that integrates Auto Regression and the Moving Averages to generate a model of best fit. The ARIMA models have been used for rainfall forecasting, for instance, Somvanshi [6] used

it for Hyderabad region in India using 93 years of annual data as the train data and 10 years as the test data, and Uruya found ARIMA models to be better than ARMA when forecasting rainfall data over Thailand using data from 31 stations [7]. In Africa, Abdul-Aziz used seasonal data for Ashanti Region, Ghana, for the period 1974 to 2010 [8]. He included the seasonal component in the ARIMA model and used it to forecast monthly rainfall totals [8]. In addition, Gorka [9] used ARIMA and the Artificial Neural Networks (ANN) to forecast weekly evapotranspiration for Northern Spain. Aris, [10] found ARIMA (2,1,2) 1,1,2) [6] as best model to forecast daily evapotranspiration over Nile Delta Region, Egypt. The main element was evapotranspiration, as opposed to rainfall which was of interest in this study. None of the studies modeled using daily rainfall data, which is an important component for use in DSTs.

The ARIMA models are univariate and focus on time lags for observed values and the white noise, and differencing in order to make the data stationary. However, statistically speaking, multivariate models would explain more variability in the white noise. Thus, we considered a multivariate time series model for forecasting. The data used had the daily minimum and maximum temperatures. These were used in a Vector Auto Regression (VAR) model. VAR models have been used in macroeconomic studies to provide very useful forecasts. Examples of VAR in use include the study of Australian Economy where 11 variables were used to study the economy for 19 years starting 1980 [11]. The cointegrated VAR models have also been applied to find the direct effects of oil price shocks in the output and price for the G-7 countries [12]. The VAR model was used in Semarang-Central Java Indonesia where the rainfall was coupled with humidity and temperature to forecast with the results showing that it was better than ARIMA [13]. In this study, we used the VAR model to predict the next five events of rainfall and temperature data. The events were in time gaps of days, month and seasons.

Comparison of model performances are not new area of study. For instance, Adamowski [14] compared linear models, the ARIMA model and Artificial Neural Networks (ANN) to forecast peak daily water demands for the summer months residents of Ottawa, Canada. His models showed that ANN was better than the other two, however, he focused on

rainfall occurrence compared to amount, which is of interest in this study. For this project, the main interest is to forecast actual rainfall data in a most precise method.

2. Methodology

This is analysis of rainfall data from Kisumu meteorological station. Trend analysis using Auto Regressive Integrated Moving Average (ARIMA) was conducted for the daily, monthly, seasonal and annual values. An ARIMA (p,d,q) model constitutes AR (p), I (d) and MA (q). The Integrated (I) part considers the number of differences used to make the data stationary. The Auto Regressive (AR) component of the model considers the influence values of previous p terms on the current term. The Moving Average (MA) gives the influence of the previous q value of error terms on the current. The Vectorized Auto Regressive (VAR) model is a multivariate time series analysis model that applies the Auto Regression on a vector, depending on the specified lag. VAR (p) considers the effect of the last p terms in the vector on the events in current time.

All the analysis used data from the Kenya Meteorological Services (Kisumu), for the period 1961-2014. Daily data was used to generate the monthly, seasonal and annual summaries in this paper. All analysis was conducted in R. The models were fit using a train dataset and a test dataset of a five-period time gap used for testing the accuracy of forecasts from the models.

3. Results and Discussions

3.1. General Annual Rainfall

Kisumu experiences a bimodal rainfall pattern (Figure 1) with the long rains experienced between March and May and the short rains experienced between October and December. January and February are usually the driest months, except for some sporadic rainfall. Figure 1 gives the average of the total monthly rainfall using data from Kenya Meteorological Services station in Kisumu using daily data between 1961 and 2014. Sometimes it is hard to distinguish the seasons since the rainfall continue into the next.

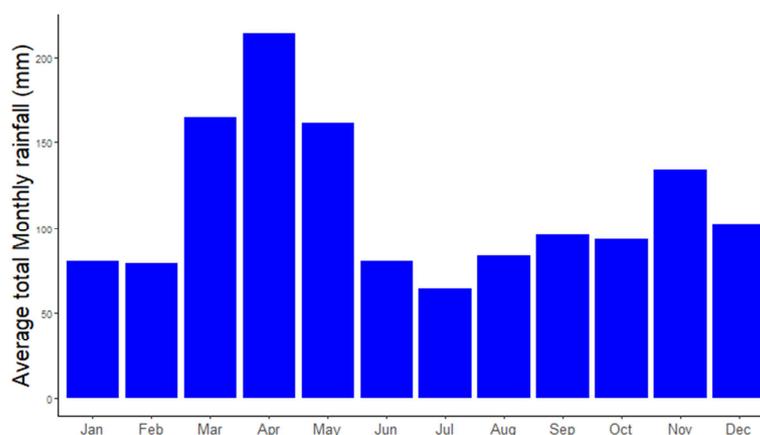


Figure 1. The long-term average cumulative rainfall for different months for Kisumu.

3.2. Fitting ARIMA Models to the Precipitation Data

The data was further plotted in order to check for trend, seasonality and cyclicity. A general plot was used (Figure 2) to explore it. From the figure, the daily and annual total

rainfall exhibited white noise. However, monthly total rainfall had seasonality in them. In order to well differentiate noise from the signals, the Autocorrelation functions (ACFs) were plotted.

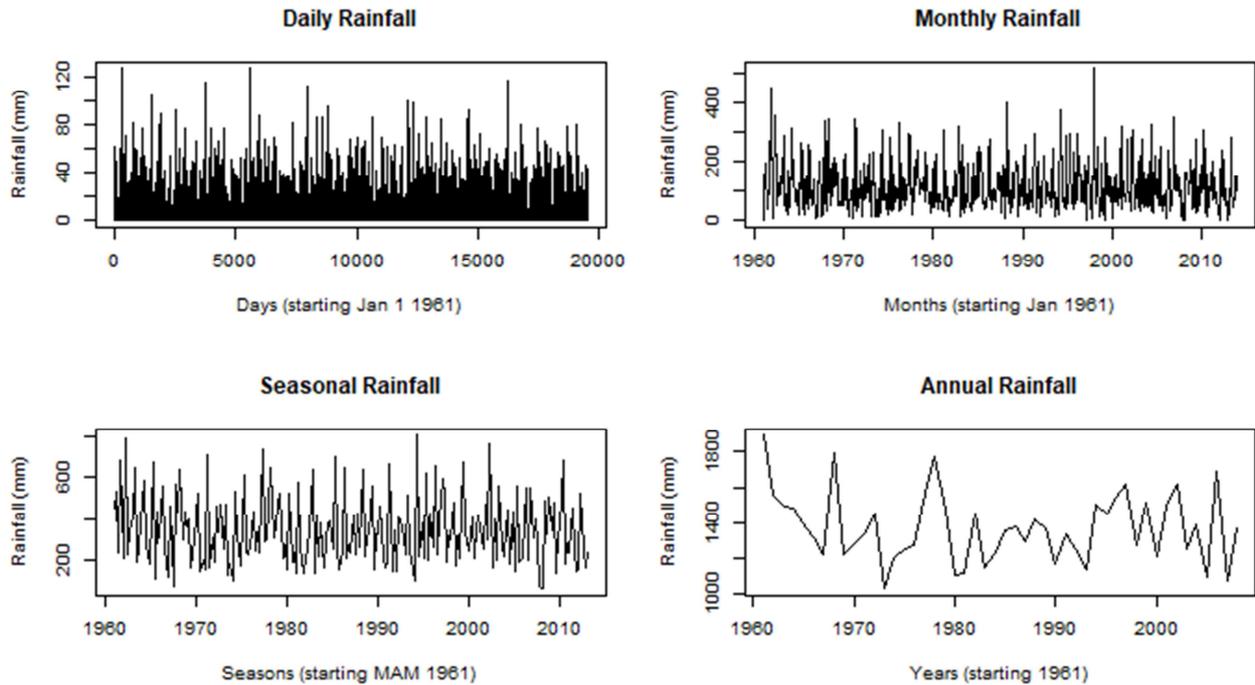


Figure 2. Trend graphs for Kisumu rainfall data (1961 - 2014).

To interpret Figure 3, consider shape of the graph. The first graph, for daily, shows that the ACFs decrease exponentially. This is considered as an interpretation of white noise. The last graph of total annual rainfall has bars not exceeding the blue horizontal lines. This is also indicative of stationarity.

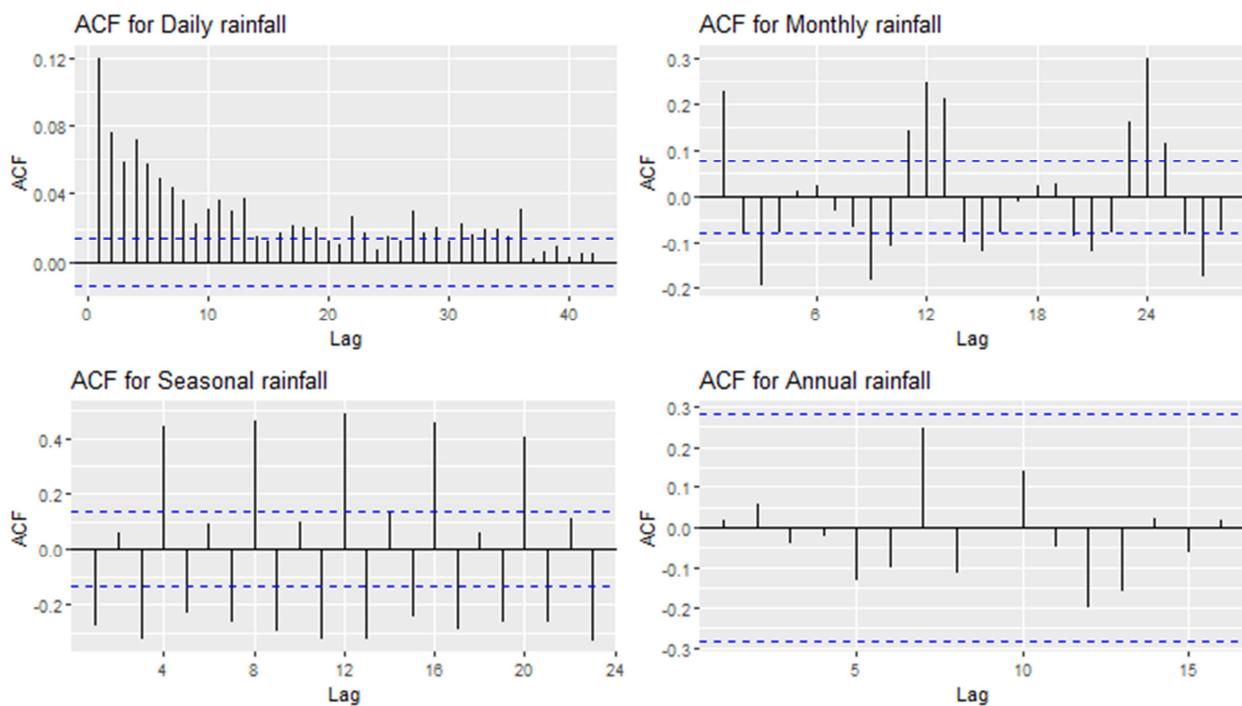


Figure 3. ACF plots for the KMS Kisumu rainfall data.

The ACF plots in Figure 3 show that the total annual rainfall was random while daily, monthly and seasonal totals had significant ACF values at lags greater or equal to 1. This justified the use of ARIMA models for the daily, monthly and seasonal since at least one difference will be required.

The models were generated using the *auto.arima* function of the *forecast* library in R 3.6.0. The function used the Bayesian Information Criterion (BIC) and the Akaike’s Information Criterion (AIC) to test for model of best fit. The model whose BIC and AIC values were lowest were selected. As expected, the more the parameters, the lesser the AIC value. Thus, BIC was particularly useful since it added a

penalty term for every coefficient included in the model. The models were generated for the daily, monthly, seasonal and annual data. Differencing was not done to any of the data, except seasonal differencing conducted on the seasonal data (Table 1). Daily data exhibited an ARMA model, monthly exhibited ARMA model with monthly AR while annual data was purely white noise since the bars do not cross the blue horizontal lines. One seasonal difference was conducted on the seasonal data to make the seasonal stationary, but in both cases the AR component explained more than the MA component.

Table 1. Results of the best fitting ARIMA model for daily, monthly, seasonal and annual rainfall data using KMS rainfl data for kisumu (1961 - 2014).

parameter	Estimate	se	z	p	LB	UB
ARIMA (1,0,2) with non-zero mean for the daily rainfall data AIC=142539.5 AICc=142539.5 BIC=142578.9						
ar1	0.918	0.014	66.804	< 2.2e-16	0.891	0.945
ma1	-0.817	0.016	-52.138	< 2.2e-16	-0.848	-0.786
ma2	-0.040	0.008	-4.769	< 2.2e-16	-0.057	-0.024
intercept	3.717	0.116	32.016	< 2.2e-16	3.489	3.944
ARIMA (1,0,3) (2,0,0) [12] with non-zero mean for the total monthly rainfall data AIC=7238.99 AICc=7239.22 BIC=7274.63						
ar1	0.527	0.419	1.256	0.209	-0.295	1.348
ma1	-0.366	0.419	-0.875	0.382	-1.187	0.454
ma2	-0.099	0.076	-1.312	0.190	-0.247	0.049
ma3	-0.107	0.052	-2.072	0.038	-0.209	-0.006
sar1	0.137	0.041	3.306	0.001	0.056	0.218
sar2	0.243	0.041	5.991	<2e-16	0.163	0.322
intercept	113.756	4.008	28.381	<2e-16	105.900	121.612
ARIMA (1,0,0) (1,1,0) [4] with drift for the total seasonal rainfall AIC=2605.91 AICc=2606.11 BIC=2619.2						
ar1	0.001	0.070	-0.001	0.999	-0.137	0.137
sar1	-0.522	0.060	-8.671	<2e-16	-0.640	-0.404
ARIMA (0,0,0) with non-zero mean for the total annual rainfall AIC=645.73 AICc=645.99 BIC=649.47						
intercept	1372.075	27.942	49.104	< 2.2e-16	1317.31	1426.84

The time gap and the unit summed differed for the different scenarios. Thus we let $Y^d_{t^d}$ be the amount of rainfall experienced on day t^d , $Y^m_{t^m}$ be the total monthly rainfall for month t^m and $Y^s_{t^s}$ be the total seasonal rainfall for season t^s . The ARIMA models above can be represented as below.

3.2.1. ARIMA for the Daily and Annual Data

The total annual rainfall was completely random hence did not fit any ARIMA model. The average total annual rainfall

$$Y^d_{t^d} = c + \phi_1^d Y^d_{(t^d-1)} + e_{t^d} + \theta_1^d w^d_{(t^d-1)} + \theta_2^d w^d_{(t^d-2)} \tag{1}$$

And when the coefficients are fit, the model is written as

$$Y^d_{t^d} = 3.717 + 0.918Y^d_{(t^d-1)} - 0.817w^d_{(t^d-1)} - 0.040w^d_{(t^d-2)} \tag{2}$$

The model shows that the amount of rainfall experienced on the previous day contributed positively to the amount recorded for the next day. However, the error terms experienced on the preceding two days had reducing effect on the amount of rainfall experienced on the day of interest.

experience at KMS Kisumu was 1372.075 mm with a 95% Confidence interval of (1317.31, 1426.84) mm. Because of this, annual data was not used for forecasting.

The daily rainfall was stationary hence no differencing was done, with the best model being ARIMA (1,0,2) model. This ARMA model had statistically significant coefficients (p-value<0.05) for the single AR and the two MA terms. The mean was none zero, hence the intercept was 3.7171 mm of rainfall. The model can be represented using the notations as below.

3.2.2. SARIMA for the Monthly and Seasonal Data

The monthly data and the seasonal data both exhibited seasonality of frequencies twelve (12) and four (4) respectively. The model selected for the monthly data was

ARIMA (1,0,3) (2,0,0) [12] with non-zero mean, while for the seasonal data was ARIMA (0,0,1) (2,1,0) [4] with drift.

The ARIMA (1,0,3) (2,0,0) [12] for the monthly total rainfall indicated that the amount of rainfall experienced in

$$Y^m_{t^m} = c + \phi_1^m Y^m_{(t^m-1)} + w_{t^m} + \theta_1^m w^m_{(t^m-1)} + \theta_2^m w^m_{(t^m-2)} + \theta_3^m w^m_{(t^m-3)} + \phi_2^m Y^m_{(t^m-12)} + \phi_3^m Y^m_{(t^m-24)} \quad (3)$$

And as below after using the modelled coefficients

$$Y^m_{t^m} = 113.756 + 0.527Y^m_{(t^m-1)} - 0.366w^m_{(t^m-1)} - 0.099w^m_{(t^m-2)} - 0.107w^m_{(t^m-3)} + 0.137Y^m_{(t^m-12)} + 0.243Y^m_{(t^m-24)} \quad (4)$$

However, in the above model, only the seasonal AR terms and the non-zero mean had statistical significance at $\alpha = 0.05$. The other AR and MA terms all had p-values greater than 0.05 hence there was not enough evidence to reject the null hypothesis that the coefficients are different from zero (0).

The ARIMA (1,0,0) (1,1,0) [4] for the seasonal total rainfall showed that the amount of rainfall experienced in the same season in the previous year contributed significantly to

$$Y^s_{t^s} - Y^s_{(t^s-4)} = C + \phi_1^s (Y^s_{(t^s-1)} - Y^s_{(t^s-2)}) + \phi_2^s (Y^s_{(t^s-1)} - Y^s_{(t^s-5)}) \quad (5)$$

When the coefficients have been included, the model becomes:

$$Y^s_{t^s} - Y^s_{(t^s-4)} = C + 0.001(Y^s_{(t^s-1)} - Y^s_{(t^s-2)}) - 0.522(Y^s_{(t^s-1)} - Y^s_{(t^s-5)}) \quad (6)$$

3.3. Forecasting Precipitation Data

For the above analysis, the train data was used to fit models appropriate for the region. A test dataset was left out for forecasting purpose and to test the strength of the models used. A test data of 5 values were used. Although this figure was selected arbitrarily, a lot of importance is attached to the next day, month or season for general forecasts. The annual data was excluded from the forecasts since the model was simply an average.

the same months for two preceding years contributed significantly to the amount of cumulative rainfall for the month of interest. The model above is be represented as:

the amount of cumulative rainfall for the current. The Seasonal Auto Regression lag had a decreasing component, an indication of long-term decrease in rainfall amount. One difference was used to make the seasons component stationary. Only the seasonal AR terms had statistical significance at $\alpha = 0.05$. The AR term had p-value greater than 0.05 hence the we fail to reject the null hypothesis that the coefficients are different from zero (0). The equation for the model may be represented as below:

Table 2 has several measures of forecast accuracy, they include the Mean error (ME), the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE). The MASE is the quotient of MAE and a scaling constant. The ARIMA functions generated were useful since they are assumed to provide the lowest forecasting errors for the data.

Table 2. Testing for accuracy of forecasted data using KMS Kisumu rainfall data.

		ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Daily	Training	0.000	9.285	5.175	-Inf	Inf	0.907	0.000	NA
	Test set	-2.637	2.651	2.637	-Inf	Inf	0.462	0.264	0
Monthly	Training	-0.488	70.650	53.818	-Inf	Inf	0.774	-0.001	NA
	Test set	17.911	61.436	56.487	-1907.207	1950.424	0.813	-0.099	0.328
Seasonal	Training	-5.460	135.007	105.472	-16.462	39.683	0.846	0.005	NA
	Test set	104.919	286.668	237.370	2.852	61.454	1.904	-0.526	1.065

The forecasted values are provided in Table 3 for the next five time periods, that is days, months and seasons. The monthly and seasonal forecasts did not deviate very much from the long-term averages, except for the fourth quarter. Of importance is the forecast for the days. The forecasted values ran for the period 1st July 2014 to 5th July 2014. The forecasted values were all above 2 mm of rain. When you compare the forecasted value to the long-term average value

in the last column of Table 3, there was some disparity as expected. Further analysis of the exact periods for the preceding 53 years showed that there was no rain for the said days for at least 60% of the years. A major reason may be assumed to be because of one event in which an extreme value of 128 mm was experienced in one day. We therefore generated a VAR model by adding the daily minimum and maximum temperatures to create the vector.

Table 3. Forecasted rainfall for the following five time periods.

Date	Forecast	Lo 80	Hi 80	Lo 95	Hi 95	Long term average
26-Jun-2014	2.479	-9.420	14.378	-15.719	20.676	1.983
27-Jun-2014	2.682	-9.278	14.642	-15.609	20.973	3.396
28-Jun-2014	2.767	-9.210	14.743	-15.550	21.083	1.594
29-Jun-2014	2.844	-9.146	14.835	-15.493	21.182	1.272
30-Jun-2014	2.916	-9.086	14.918	-15.440	21.271	2.862

Date	Forecast	Lo 80	Hi 80	Lo 95	Hi 95	Long term average
Month						
Jan 2014	97.441	6.899	187.983	-41.032	235.913	81.268
Feb 2014	102.247	10.547	193.946	-37.995	242.489	80.432
Mar 2014	140.636	48.927	232.345	0.3794	280.893	164.936
Apr 2014	123.779	31.480	216.077	-17.380	264.938	217.183
May 2014	131.745	39.283	224.207	-9.663	273.153	162.379
Season						
2013 Q2	483.890	309.192	658.5883	216.7123	751.068	544.498
2013 Q3	249.306	74.608	424.0046	-17.8714	516.484	228.443
2013 Q4	153.773	-20.925	328.4714	-113.4046	420.951	323.608
2014 Q1	193.297	18.598	367.9948	-73.8812	460.474	264.217
2014 Q2	505.036	311.406	698.6665	208.9044	801.168	544.498

3.4. Including Temperature Data in the Models

The Vector Auto Regressive (VAR) models were used to conduct a multiple time series analysis. The VAR models are Auto Regressive models for a vector. In this case the vector constituted rainfall, minimum and maximum temperatures. Given that $y_t = [y_{1t}, y_{2t}, y_{3t}]^t$, then the VAR (p) model is represented as:

$$Y_{t,d} = G_0 + \sum_{i=1}^p G_i Y_{t,d-p} + e_{t,d} \quad (7)$$

Where $E[e_{t,d}] = 0$ and $E[e_{t,d}, e_{\tau}] = \begin{cases} \Omega, & \text{if } t = \tau \\ 0, & \text{otherwise} \end{cases}$. The model was run with a maximum of 5 lags to produce the output in Table 4. In the VAR (5) model, the five lags of rainfall contributed significantly to the rainfall at time t. Minimum temperature contributed significantly at the first three lags while only the first two lags of maximum temperature contributed significantly.

Table 4. VAR (5) model for the predicting rainfall considering lagged rainfall, minimum and maximum temperatures.

	Estimate	Std. Error	t value	Pr (> t)
Rain. I1	0.099	0.007	13.839	< 2e-16
MaxT. I1	-0.109	0.050	-2.189	0.02864
MinT. I1	0.233	0.061	3.827	0.00013
Rain. I2	0.046	0.007	6.426	1.34e-10
MaxT. I2	-0.112	0.057	-1.950	0.05120
MinT. I2	0.133	0.067	1.981	0.04758
Rain. I3	0.028	0.007	3.910	9.25e-05
MaxT. I3	0.066	0.057	1.158	0.24675
MinT. I3	0.193	0.067	2.856	0.00430
Rain. I4	0.049	0.007	6.763	1.39e-11
MaxT. I4	-0.049	0.057	-0.851	0.39479
MinT. I4	-0.041	0.067	-0.614	0.53901
Rain. I5	0.031	0.007	4.347	1.39e-05
MaxT. I5	-0.002	0.050	-0.042	0.96651
MinT. I5	0.101	0.060	1.682	0.09258
const	-1.847	1.492	-1.238	0.21575

Residual standard error: 9.278 on 19518 degrees of freedom
 Multiple R-Squared: 0.02975, Adjusted R-squared: 0.029
 F-statistic: 39.9 on 15 and 19518 DF, p-value: < 2.2e-16

Using y_{1t}^d, y_{2t}^d and y_{3t}^d as the rainfall, minimum and maximum temperatures respectively, then model can be represented as:

$$\begin{bmatrix} y_{1t}^d \\ y_{2t}^d \\ y_{3t}^d \end{bmatrix} = \begin{bmatrix} -1.847 \\ 6.492 \\ 3.320 \end{bmatrix} + \begin{bmatrix} 0.099 & -0.109 & 0.233 \\ -1.310e-2 & 0.547 & 1.683e-2 \\ -3.48e-2 & 0.125 & 0.474 \end{bmatrix} \begin{bmatrix} y_{1,t^d-1} \\ y_{2,t^d-1} \\ y_{3,t^d-1} \end{bmatrix} + \begin{bmatrix} 0.046 & -0.112 & 0.133 \\ -1.297e-03 & 9.243e-02 & -2.366e-2 \\ 5.87e-3 & -1.054e-2 & 8.634e-2 \end{bmatrix} \begin{bmatrix} y_{1,t^d-2} \\ y_{2,t^d-2} \\ y_{3,t^d-2} \end{bmatrix} \\ + \begin{bmatrix} 0.028 & 0.066 & 0.193 \\ 3.560e-4 & 5.522e-2 & 1.730e-4 \\ 3.299e-3 & -5.075e-3 & 3.008e-2 \end{bmatrix} \begin{bmatrix} y_{1,t^d-3} \\ y_{2,t^d-3} \\ y_{3,t^d-3} \end{bmatrix} + \begin{bmatrix} 0.049 & -0.049 & -0.041 \\ -2.240e-5 & 4.467e-2 & -2.829e-2 \\ 2.152e-3 & -8.246e-3 & 3.208e-2 \end{bmatrix} \begin{bmatrix} y_{1,t^d-4} \\ y_{2,t^d-4} \\ y_{3,t^d-4} \end{bmatrix} \\ + \begin{bmatrix} 0.031 & 0.101 & -1.847 \\ -6.271e-7 & 6.253e-2 & 4.843e-3 \\ 3.183e-3 & -1.787e-2 & 4.14e-2 \end{bmatrix} \begin{bmatrix} y_{1,t^d-5} \\ y_{2,t^d-5} \\ y_{3,t^d-5} \end{bmatrix} \quad (8)$$

With the residual covariance matrix:

$$\begin{bmatrix} 86.083 & -0.699 & 0.407 \\ -0.699 & 1.772 & 0.097 \\ 0.407 & 0.097 & 1.196 \end{bmatrix}$$

And the residual correlation matrix

$$\begin{bmatrix} 1.000 & -0.057 & 0.040 \\ -0.057 & 1.000 & 0.067 \\ 0.040 & 0.067 & 1.000 \end{bmatrix}$$

The VAR (5) model had a p-value of $< 2.2e-16$ for rainfall, minimum and maximum temperatures hence statistically significant. The model explained 3%, 51% and 38% of the variability in rainfall, minimum and maximum temperatures. According to the model, the forecast for time t depended a lot on the rainfall and temperature values for the first lag, with reducing dependencies as the lag increased.

The model was used to forecast for the same five days as used in Table 3. The forecast results are provided in the last three columns of Table 5. The observed values were used as the five lags used in the model. The model resulted in accurate prediction of the temperature values. However, the forecast errors were bigger compared to forecast errors when the ARIMA (1,0,2) was used (Table 3).

Table 5. Forecast results using VAR (5) model for daily rainfall and temperature data.

Date	Observed			Forecasted		
	Rainfall	Max T	Min T	Rainfall	Max T	Min T
21-Jun-2014	0.6	26.3	20.3			
22-Jun-2014	0.0	28.5	19.5			
23-Jun-2014	0.0	29.5	18.2			
24-Jun-2014	0.0	29.0	18.3			
35-Jun-2014	0.0	29.5	16.8			
26-Jun-2014	0.0	28.2	17.1	6.285	29.261	17.401
27-Jun-2014	0.5	30.2	17.5	5.756	28.824	17.127
28-Jun-2014	0.0	29.3	16.3	5.869	29.850	17.493
29-Jun-2014	0.0	30.0	18.8	5.582	29.481	16.806
30-Jun-2014	0.0	28.8	15.6	5.708	29.920	17.934

When the same VAR (5) model was constructed for the monthly and seasonal data. The average of the monthly minimum and maximum temperatures were used together with the total rainfall. The same summaries were used with the seasonal VAR (5) model. The lag of five was selected arbitrarily to tally with the lags used in the model for daily data. Model results showed that the first and second lags of the minimum temperatures contributed to the rainfall data with statistical significance, and the maximum temperatures of the second and fifth. The model explained 23.24% of the total variability in the rainfall data, and the forecast was closer to the long-term average, rather than the observed values for that year. A similar view was noticed for the seasonal forecasts.

Thus, the results showed a divergent behavior for univariate and multivariate time series model to the models by Jones [15]. Jones used the VAR model to find demands for key resources in the emergency department over time. However, given the variability in weather, in particular rainfall, this makes the VAR models not very suitable for the modelling, in comparison to the univariate ARIMA models.

4. Conclusion

In this paper, 50-year historical data has been to generate ARIMA, SARIMA and VAR models for forecasting rainfall values for Kisumu. SARIMA models were fit and forecasted monthly and seasonal rainfall totals, while ARIMA and VAR models were fit and forecasted the daily rainfall amount. The best fit models for the daily, total monthly and total seasonal rainfall were ARIMA (1,0,2), ARIMA (1,0,3) (2,0,0) [12] and ARIMA (0,0,1) (2,1,0) [4]. The latter two predicted total monthly and total seasonal rainfall more accurately. In addition, ARIMA (1,0,2) had smaller forecast errors than the VAR (5) model which considered the minimum and maximum daily temperatures.

Appreciation

The data used in this paper was collected under a CCAFS funded project in Nyando region. The project “Enhancing Climate Smart Agriculture in Nyando through innovative partnerships, Decision Support Tools and Climate Information Services” ran between 2014 and 2015. We are grateful to CCAFS for the support. We are also very grateful to the KMS Kisumu and Kericho offices who worked with us in analysis of their climate data to produce the outputs used in this paper.

References

- [1] Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., Huth, N. I., Hargreaves, J. N., Meinke, H., Hochman, Z. and McLean, G., 2003. *An overview of APSIM, a model designed for farming systems simulation*. European journal of agronomy, 18 (3-4), pp. 267-288.
- [2] Somvanshi, V. K., Pandey, O. P., Agrawal, P. K., Kalanker, N. V., Prakash, M. R. and Chand, R., 2006. *Modeling and prediction of rainfall using artificial neural network and ARIMA techniques*. J. Ind. Geophys. Union, 10 (2), pp. 141-151.
- [3] Rose, D. C., Sutherland, W. J., Parker, C., Lobley, M., Winter, M., Morris, C., Twining, S., Ffoulkes, C., Amano, T. and Dicks, L. V., 2016. *Decision support tools for agriculture: Towards effective design and delivery*. Agricultural systems, 149, pp. 165-174.
- [4] Cologni, A. and Manera, M., 2008. *Oil prices, inflation and interest rates in a structural cointegrated VAR model for the G-7 countries*. Energy economics, 30 (3), pp. 856-888.
- [5] Psilovikos, A. and Elhag, M., 2013. *Forecasting of remotely sensed daily evapotranspiration data over Nile Delta region, Egypt*. Water resources management, 27 (12), pp. 4115-4130.

- [6] Cropwat, F. A. O., 1992. *A computer program for irrigation Planning and Management*, By M. Smith. FAO Irrigation and Drainage Paper, (46).]
- [7] Weesakul, U. and Lowanichchai, S., 2005. *Rainfall forecast for agricultural water allocation planning in Thailand*. Science & Technology Asia, pp. 18-27.
- [8] Dungey, M. and Pagan, A., 2000. *A structural VAR model of the Australian economy*. Economic record, 76 (235), pp. 321-342.
- [9] Landeras, G., Ortiz-Barredo, A. and López, J. J., 2009. *Forecasting weekly evapotranspiration with ARIMA and artificial neural network models*. Journal of irrigation and drainage engineering, 135 (3), pp. 323-334.
- [10] Shilenje, Z. W. and Ogwang, B. A., 2015. *The role of Kenya meteorological service in weather early warning in Kenya*. International Journal of Atmospheric Sciences, 2015.
- [11] Nugroho, A., Hartati, S. and Mustofa, K., 2014. *Vector Autoregression (Var) Model for Rainfall Forecast and Isohyet Mapping in Semarang–Central Java–Indonesia*. Editorial Preface, 5 (11).
- [12] Adamowski, J. F., 2008. *Peak daily water demand forecast modeling using artificial neural networks*. Journal of Water Resources Planning and Management, 134 (2), pp. 119-128.
- [13] Abdul-Aziz, A. R., Anokye, M., Kwame, A., Munyakazi, L. and Nsowah-Nuamah, N. N. N., 2013. *Modeling and forecasting rainfall pattern in Ghana as a seasonal ARIMA process: The case of Ashanti region*. International Journal of Humanities and Social Science, 3 (3), pp. 224-233.
- [14] Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J. and Ritchie, J. T., 2003. *The DSSAT cropping system model*. European journal of agronomy, 18 (3-4), pp. 235-265.
- [15] Jones, S. S., Evans, R. S., Allen, T. L., Thomas, A., Haug, P. J., Welch, S. J. and Snow, G. L., 2009. *A multivariate time series approach to modeling and forecasting demand in the emergency department*. Journal of biomedical informatics, 42 (1), pp. 123-139.