

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/329894230>

Use of Principal Components Regression and Time-Series Analysis to Predict the Water Level of the Akosombo Dam Level

Article · December 2018

DOI: 10.5923/j.statistics.20180806.07

CITATION

1

READS

1,395

3 authors, including:



Isaac Ofori Asare

Kwame Nkrumah University Of Science and Technology

4 PUBLICATIONS 12 CITATIONS

[SEE PROFILE](#)



Dorothy Frempong

Ghana Institute of Management and Public Administration

3 PUBLICATIONS 6 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Article [View project](#)



Proximate Breast Cancer Factors Using Data Mining Classification Techniques [View project](#)

Use of Principal Components Regression and Time-Series Analysis to Predict the Water Level of the Akosombo Dam Level

Isaac Ofori Asare^{1*}, Dorothy Anima Frempong², Paul Larbi³

¹Mathematics Department, Kwame Nkrumah University of Science and Technology, Ghana

²Computer science Department, Accra Technical University, Ghana

³Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Abstract Knowing the water level of the Akosombo Dam would help Ghanaian since we depend heavily on hydroelectric power. When the future of the water level is known, society would be able to plan on the usage of electricity for the industries, society, individuals who use some of the water storage for irrigation, water supply purposes. The study employed rainfall from the 12 catchment areas to the River Volta and the daily water level of the dam for a period of 78-years. Principal Component Regression was applied to the input variables for the reduction of its large size to a few principal components to explain the variations in the original dataset. The outcome of the PCR extraction was two principal components. Time Series using Seasonal Autoregressive Integrated Moving Average was used to model the data. The appropriate model that fit the data well was ARIMA (2,1,2) (1,0,0) [12] after comparing other models AICs. The model with the smallest AIC and the least number of parameters was selected as the best model.

Keywords Principal Component Regression, Time series, ARIMA, SARIMA, Measures of Adequacy

1. Introduction

The hydroelectric project is a physical structure constructed for the generation of hydropower. They have been used as a structural mechanism to regulate the flow of water for storages purposes. The structure (hydro project) are made to reduce the fast flow of water to the dam. The hydroelectric projects are capable of storing flow of rainwater to ensure water supply for hydro-power generation and for other economic purposes such as for agro-business, household and industry usage. During raining season, the dam is able to get enough water for it intended purposes, production of enough hydropower. When there is drought, the water level of the dam reduces which limit the production of electricity as Ghanaians depend largely on hydropower for domestic and industrial usage. The dam level depends largely on rainfall from some catchment's areas of Ghana. When it rains in these areas the water flow into the tributaries and then move to the river Volta. There is the need to know the contribution of each of these catchment areas when it

rains and their impact on the water level of the Akosombo dam. Also, as a major source of hydro-power, there is the need to be able to forecast the water level at a time accurately and precisely to determine the amount of energy that might be produced at a particular point in time, daily, weekly, monthly or even yearly. A result of that there was the need to have good forecasting tool that could help forecast the water level of the dam accurately and precisely.

At the beginning of the year 2007, Ghanaians were worried and raised concerns about the limited hydro-power generation (electricity) or supply from the generation plant (dam) due to the low water levels in the Lake Volta reservoir, reports from the dam site indicating that the project was functioning below its capacity due to the problems of drought due to global warming. It implies that when there is minuscule or no rains from the river catchment areas to feed it, the Akosombo dam level would be low and as a result, the dam would not be able to perform up to expectation.

In the year 2010, the country experienced high rains and as a result, the Akosombo dam site recorded the highest water level due to heavy rainfalls in a catchment area contributing to the dam, the reservoir elevation went above it expected a value of 84.73 m (278.0 ft). Due to the raised in the water level of the dam, it caused management of the dam site to let some of the water flow out by allowing some of the water to pass through the floodgates at a basin elevation of 84.45 m (277 ft), and for several days, weeks, water was still

* Corresponding author:

asare.re3@gmail.com (Isaac Ofori Asare)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2018 The Author(s). Published by Scientific & Academic Publishing

This work is licensed under the Creative Commons Attribution International

License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

spattering from the river, creating some flooding in the nearby communities along the dam site.

According to the Ghanaian times on June 7, 2016, at the time when the report has been filed, the dam level of the reservoir was approximately 237.27 ft, just a little below the 240 ft the accepted level that can enhance the operation of the plant a situation the VRA has been threatened with all this while. The Corporate Affairs Manager at the VRA, Mrs Getrude N. Koomson, in a formal interview with one of the new papers; the Ghanaian Times indicated that the situation was causing the machinery to underperform.

This news indicates that until the water level appreciates to up to certain level to empower all the six turbines run concurrently, then the alternative they would be left with was to relying on or cope with the little inflow and to ensure that the engineers keep at most four of the turbines running subject to the demand of hydro-power in the country (Ghana). This is as a result of low rainfalls in the country. When there are rains from the catchment areas especially the medium or the central belt and the northern zones of the country. The situation the dam is experiencing could change if there are rains from these areas.

The water level of the dam is mainly for hydropower and other purposes such as water storage which has been employed for irrigation needs and water supply for domestic use are some key importance of the dam, therefore there is the need to knowing the water level of the dam for effective and efficient planning and the need to knowing the contributions water from the various catchment areas to the dam for the desired performance. This could be done when we know the main contribution of the various catchment areas to the water level in the dam and based on it the forecasting techniques can be employed to ensure effective and accurate result when we know the contribution of water by the catchment areas can be obtained.

There are factors that influence the water level negatively such as evaporation, soil moisture and human activities along the banks of the river, these are factors that need to be considered in forecasting of the water level. Marino et al (2017) climate change can cause the distribution of rainfall patterns, with potential effect for the water bodies. Changes in water bodies' level are as a result of factors such as like rains and other atmospheric conditions such as temperature, evaporation and humidity. When there is continuous wet and cold condition over a period of time, the volume of water levels rises, on the contrary, warm and dry periods would cause the water levels to decline. The global warming can affect the normal cycle of rainfall, thereby destroying the water supply and demand and having a significant impact on water bodies, agriculture, human health, animals and plant, this condition could prolong drought and water shortages (Brebbia, 2011).

2. Research Problem

The Akosombo Hydroelectric Project requires rainfalls

from its catchment areas to operate effectively for generating of hydroelectric power and production of water for domestic consumption and industrial usage to Ghanaians and other nearby countries. According to the Ghana Metro-logical Agency, the Akosombo Dam takes its volume of water from about 12 catchment areas (stations) when it rains in these areas, therefore contribute significantly to the water level of the dam. They are as follows; Kintampo, Bui, Tamale, Yendi, Akuse, Navrongo, Salaga, Kate-Karachi, Bole, Atebubu, Kpando and Ho. Ghana depends so much on the Akosombo Hydroelectric power for its activities there is the need to get a model that can improve the older models used in predicting the water level. The study adopted Principal Component analysis to reduce these rainfall stations to few stations that could be used to describe the variation explained by all the stations. Though the traditional time series model does not consider nonlinear inputs, hence giving out inconsistency results in its anticipation as indicated by Rani and Parekh (2014). The traditional time series technique is efficient and effective for a long time, but there was a deficiency that they suffer that is the issue of stationary and linearity. Though the ARIMA model does not take into consideration nonlinear data, the process of transformation could be applied to the data to make it linearize. Also differencing could be applied to the data if the dataset is non-stationary to make it stationary. Ghana depends heavily on hydropower (Akosombo) for electricity, there is the need to get model that could be used to forecast the water level of the Akosombo Dam at any given time been daily, monthly or year based for proper planning of the power issues in the country.

Planning for the hydroelectric project is a very essential step for success in the evolution of Ghana since we depend mainly on hydro-power. This progress would be successful if the water level of the dam is determined correctly or checking for accuracy of the dam water level. As a result of this prediction; the study adopted Principle Component Regression (PCR) and variable importance using the random forest technique in the determination of the important rainfall catchment area having an impact on the water level.

3. Specific Objectives of the Study

1. Knowing the impact of rainfall basin stations in terms of percentage to the Akosombo Water level.
2. Reduce the number of rainfall basins stations contributing to the Akosombo water level to a few Principal Components.
3. Get a good forecast technique for the water level of the Akosombo Dam.

4. Research Methodology

The Akosombo hydroelectric project is greatly influenced by several atmospheric conditions or atmospheric factors

such as rainfalls, the flow of water sources, warmth and high temperature, and even heat due to evaporation or humidity. Due to the unavailability of data on the side of the researcher, the study focused on upstream rains and past data on the dam level. The data used for the research were the past rainfalls monthly for a period of 78 years making 948 data point from the 12 major tributaries of the Volta River from 1936-2014. They include; Kintampo, Tamale, Bui, Yendi, Navrongo, Salage, Kata-Krachi, Bole, Atebubu, Kpando, Ho and Akuse stations. Also, monthly data on the water level of the Akosombo Dam was obtained. The principal component was applied to these stations in determining the most significant stations that could be used to explain the variation in the water level when it rained over these stations.



Figure 1. Akosombo Dam water level

5. Model Specification

Principal Component Regression and Random Forest techniques were used to analyse the data gathered. perform principal components analysis (PCA) was performed first on the original data, then perform dimension reduction by selecting the number of principal components (m) using cross-validation or test set error, and finally conduct regression using the first m dimension reduced principal components.

6. Multicollinearity: Examination of Correlation Matrix

One of the assumption underlings the usage of PCA /PCR is to ensure that, there is independency among the variables. There would be the independence of the variables when there is no multicollinearity among the variables. A high value of the correlation between two variables may indicate that the variables are collinear. This method is easy, but it cannot produce a clear estimate of the degree of multicollinearity. (El-Dereny and Rashwan, 2011). The correlation coefficients are greater than 0.80 or 0.90 then this is an indication of multicollinearity. Variance Inflation Factor (VIF) is one of the techniques that is used to assess the level of collinearity in an ordinary least square regression analysis. if any of the VIF values exceed 5 or 10, it is an indication that

the associated regression coefficients are poorly estimated because of multicollinearity (Montgomery, 2001).

The VIF is calculated as

$$VIF = 1/1 - R_j^2 \quad (1)$$

where R^2 , represent the coefficient of determination when X_j is regressed on all other predictor variables in the model.

7. Eigen Analysis of Correlation Matrix

The eigenvalues can also be used to measure or determine the component the number of components that have to be extracted. It can check for the presence of multicollinearity in the predictor variables, one or more of the eigenvalues will be small (near to zero).

8. Principal Component Regression (PCR)

The PCR is used to handle multicollinearity among variables, it is not usually included in standard regression analysis. The PCA follows from the fact that every linear regression model can be restated in terms of a set of orthogonal explanatory variables. These new variables are obtained as linear combinations of the original explanatory variables. They are referred to as the principal components. The independent variables in the PCR are given as;

$$X_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (2)$$

Where X_{ij} is the i^{th} observation on the j^{th} variable, and \bar{x} and S_j represent the estimated mean and standard deviation respectively. The dependent variable is centered;

$$y_i = Y_j - \bar{y}. \quad (3)$$

The transform matrix $Z = XA$, where X is the $n \times p$ matrix of n observations on p independent variables, Z is the $n \times p$ matrix of transformed data whilst A is the $p \times p$ matrix consisting of eigenvectors.

The regression model is given as;

$$Y = ZB \quad (4)$$

where B is the $p \times 1$ vector of unknown parameters and it is estimated as;

$$\widehat{B} = (ZZ)^{-1}ZY \quad (5)$$

The regression equation for the PCA is given as;

$$DamLevel = \alpha PC1 + \beta PC2 + \gamma PC3 \dots + \delta PCn \quad (6)$$

9. Time Series Statistical Models

The Box -Jenkins methodology was duly followed in the estimating of the parameters in the time series analysis (Box and Jenkins, 1970). The Auto-regressive and the Moving Average (ARMA (p, q)), or Auto-regressive Integrated Moving Average (ARIMA (p, d, q)) was adopted for the

prediction of the time series data. Nevertheless, the application of the ARMA model assumes that the time series data be stationary; which implies, ARMA processes remains in the stability about a constant mean level. However, when data are nonstationary or have obvious trend variability, the ARIMA model built on the differencing algorithm could be adopted (Box et al. 1994). The Augmented Dickey-Fuller (ADF) is used to examine for the stationarity in the dataset test (Elliott et al. 1996).

10. Measures of Adequacy

The following measures of adequacy were used to test for the adequacy of the time series model. The performance of the proposed time series is assessed with these criteria; Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) and Mean Square Error (MSE).

Let x_t represent the it h observations and f_t denote the forecast, where $t = 1, 2, 3 \dots n$ the adequacy measures are as follows;

$$MAE = n^{-1} \sum_{t=1}^n x_t - f_t \quad (7)$$

$$MSE = n^{-1} \sum_{t=1}^n (x_t - f_t)^2 \quad (8)$$

$$MAPE = 100n^{-1} \sum_{t=1}^n x_t - f_t / x_t \quad (9)$$

$$RMSE = \sqrt{n^{-1} \sum_{t=1}^n (y_t - f_t)^2} \quad (10)$$

11. Results

The results in the table below show the descriptive statistics of the monthly rainfall of the 12 tributaries of the Akosombo dam water level of the Akosombo Dam for a period of 78 years, ranging from 1936-2014. The minimum rainfall for all the various tributaries was zero (0) meaning that there were no rains in such area in the case of the water level, the zero value means that the water level record was not captured for a reason not made available to the researcher. The maximum rainfall for all the 12 tributaries is shown in the table. The average rainfall for Kintampo is 108.188 with a deviation of 92.186, Tamale has a mean rainfall of 90.154 and a deviation of 88.979. Bui has an average rainfall of 94.161 and a deviation of 76.757. The results show that a rainfall in Kate-Krachi and Bole were not normally distributed. Since the skewness and the kurtosis for each of them was above ± 1.96 . The rest of the rains from the catchment areas were normally distributed.

Collinearity diagnostics was performed and the results obtained shows that none of the catchment areas has more than 10 Variance Inflation Factor (VIF). The least VIF value is 1.443 associated with Ho and the maximum VIF is 8.538 also associated with Yendi. This means that there is no problem of multicollinearity.

Once there is no problem of collinearity among the independent variables, the data gathered was standardized first to ensure that each predictor is on the same scale as the

other variable. This is done to prevent the algorithm to be skewed towards predictors that are dominant in absolute scale.

Table 1. Descriptive statistics of the variables

Catchment	Max	Mean	Std	Sks	Kurt
Kintampo(KT)	452.40	108.19	92.19	0.79	0.25
Tamale(T)	495.80	90.15	88.98	0.96	0.49
Bui(BU)	407.50	94.16	76.76	0.77	0.27
Yendi(YD)	577.00	100.53	97.54	0.94	0.83
Navrongo(NV)	662.90	84.93	97.98	1.32	1.81
Salaga(S)	474.70	92.48	94.10	1.06	0.77
KataKrachi(K)	1277.60	111.20	103.20	2.22	16.99
Bole(BL)	1039.40	90.74	94.22	2.41	15.95
Atebubu(AB)	566.20	103.36	95.52	0.84	0.41
Kpando(KD)	446.30	111.20	82.02	0.71	0.29
Ho(HO)	507.70	108.91	79.30	0.96	1.57
Akuse(AK)	530.10	92.59	74.93	1.15	1.97
Dam Water(DW)	277.16	162.31	120.57	-0.59	-1.62

Note sks represent Skewness and kurt, Kurtosis

Table 2. Diagnostics analysis

Catchment area	Collinearity Statistics	
	Tolerance	VIF
Kintampo	0.540	1.852
Tamale	0.244	4.100
Bui	0.619	1.617
Yendi	0.117	8.538
Navrongo	0.386	2.589
Salaga	0.527	1.897
Kata_Krachi	0.549	1.822
Bole	0.471	2.123
Atebubu	0.141	7.111
Kpando	0.462	2.163
Ho	0.693	1.443
Akuse	0.658	1.519

Note: VIF Values greater than 10 indicates collinearity

12. Principal Component Analysis Results

The results in the table below shows the Measures of Sampling Adequacy (MSA), though the MSA does not produce or show the p-values to determine the significance of the results, however MSA value of at least 0.80 is considered acceptable in terms of the sample adequacy used for the study as indicated by Norman & Streiner (2008). The analysis of the results shows as shown in Table 3 below shows the MSA for the variables used for the study. The results in the table show that the least MSA is that the least MSA is 0.87 which is associated with rains from Yendi and

the highest MSA being 0.98 which is also associated with Salaga.

Table 3. Anti-image Correlation

Variable	MSA
Kintampo(KT)	0.96
Tamale(T)	0.95
Bui(BU)	0.96
Yendi(YD)	0.87
Navrongo(NV)	0.94
Salaga(S)	0.98
Kata Krachi(K)	0.95
Bole(BL)	0.97
Atebubu(AB)	0.88
Kpando(KD)	0.95
Ho(HO)	0.94
Akuse(AK)	0.88

The Bartlett Test of Sphericity which was used to compares the correlation matrix with a matrix of zero correlations usually known as an identity matrix, which consists of all zeros except the 1's along the leading diagonal. The results obtained shows that, factor analysis is appropriate to fit the data gathered due to high MSA and also a significant Bartlett's Test of Sphericity value of p-value <0.001 at 5% significance level. However, the overall KMO value of 0.930 indicates that the sample size used for the study is adequate.

Table 4. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.930
Approx. Chi-Square		7545.517
Bartlett's Test of Sphericity	df	66
	Sig.	.000

The results from the principal components analysis show that there are two main components extracted from the from the 12 tributaries that contribute to the Akosombo water level. The Total variance explained which indicates how much of the variability in the data has been explained by the components is shown in table 3 below. From the analysis, the first component has an eigenvalue of 6.538 and a variance of 54.487%. The second components have an eigenvalue of 1.237 and a variance of 10.311%. The results of the components had an eigenvalue less than 1. Cumulatively the two components could explain about 76.798%. This means that the water level from the various tributaries can be clustered into two main groups.

Table 5. Total Variance Explained

Component	Initial Eigenvalues		
	Total	% Variance	Cum %
1	6.538	63.487	63.487
2	1.237	13.311	76.798

The figure 2 below shows the scree plot of the number of principal components that were extracted showing the components eigenvalues. Kaiser (1970) component that has an eigenvalue of at least one was extracted which is what is shown in the figure below.

In this research work, two components were extracted as indicated in the figure below cutting off after component 2. This shows only two components were extracted work. The two principal components could explain about 76.798% as indicated in table 5 above.

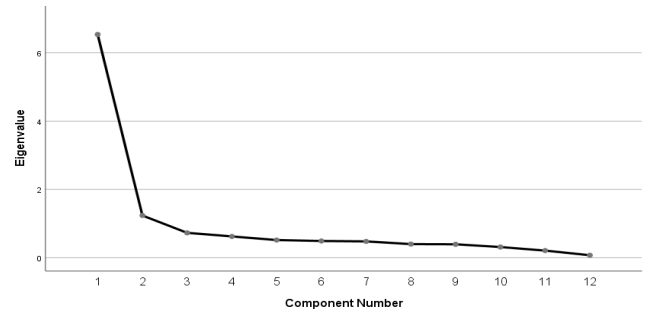


Figure 2. Scree Plot

Table 6. Results of PCA (Varimax Rotated Matrix)

Variable	Eigen vectors		Communalities
	PC1	PC2	
Tamale	0.86		0.82
Yendi	0.88		0.86
Navrongo	0.87		0.76
Salaga	0.68		0.55
Kata Krachi	0.62		0.60
Bole	0.67		0.59
Atebubu	0.85		0.82
Kintampo		0.62	0.58
Bui		0.67	0.59
Kpando		0.58	0.62
Ho		0.73	0.56
Akuse		0.80	0.64

Rotation method: Varimax with Kaiser normalization

The results in Table 6 above shows the Varimax rotation for the components model along with their communalities. The results for the rotation are quite comparable and easy to interpret. Two main components were extracted to represent the 12 components it shows that, seven (7) main variables correlate well with component 1 and among the eight variables, Yendi and Navrongo highly correlate well with component 1. Also, four (4) variables correlate well with the second component. A threshold of 0.60 was used for identifying a reliable factor in this study as indicated by (Stevens, 1996). Hair et.al (2010) indicated that items that load less than 0.50 are not accepted. From the results in table 6, only one variable had a loading of less than 0.60, the rest are having significant loadings above 0.60. There was no problem of cross-loadings between the variables all the

variables have loadings or more than 0.60. This means that more than one-half of the variations in the dataset are accounted for by the loadings on a single factor. Factor loadings (<0.50) have not been shown and the results have been sorted by loadings on a particular component.

The result in table 7 below shows the Pearson bivariate correlation between the standardised of the variables used for the study. Tabachnick and Fidell (2001) indicated that the correlation between variables less than 0.30 is not appropriate for exploratory analysis and from the result obtained as indicated in table 7, below. From the table, there is no such problem. El-Dereny and Rashwan (2011)

indicated that the correlation between variables which results in rho (ρ) of more than 0.80 is also not appropriate for exploratory analysis. The results obtained indicate that there is none of such problem. The minimum correlation value obtained is rho ($\rho = 0.29$ whilst the maximum correlation value rho ($\rho = 0.82$). Those these values were either below or above the accepted range of values their impact is insignificant since it only occurred between two sets of variables, between Akuse and Salaga ($\rho = 0.29$ and between Tamale and Yendi ($\rho = 0.82$). The results suggest that the variables are not much correlated hence there is the independence of the variables in the model.

Table 7. Bivariate Correlation Analysis

	1	2	3	4	5	6	7	8	9	10	11	12
KT	1											
T	.53**	1										
BU	.48**	.48**	1									
YD	.54**	.82**	.48**	1								
NV	.39**	.72**	.37**	.71**	1							
S	.45**	.63**	.37**	.61**	.53**	1						
K	.49**	.55**	.47**	.57**	.55**	.44**	1					
BL	.50**	.67**	.43**	.66**	.54**	.52**	.43**	1				
AB	.54**	.78**	.49**	.79**	.68**	.60**	.57**	.64**	1			
KD	.51**	.60**	.46**	.60**	.47**	.52**	.44**	.54**	.59**	1		
HO	.45**	.35**	.39**	.35**	.31**	.35**	.34**	.37**	.35**	.40**	1	
AK	.40**	.30**	.35**	.32**	.34**	.29**	.36**	.34**	.34**	.50**	.38**	1

**correlation is significant at the 0.01 level. The order of the variables in the table are labelled as 1,2...12

13. The Principal Component Regression (PCR)

After the extraction of the two main components, the eigenvectors for the two components were used as regressors for the regression analysis and the results show that there is at least a significant component in the model. The f-statistics was statistically significant at 5% significance level (F=4.446, p-value=0.012). Also, the R-square was approximately 75% to show how much the components can be explained on the dependent variable. The estimated PCR that fits the data gathered is given as;

$$Waterlevel = 2.256PC1 + 1.003PC2 \quad (11)$$

Table 8. Coefficients of the PCR

	B	Std. Error	Beta	t	Sig
Constant	162.3	3.902	-	41.599	.000
PC1	2.256	3.904	.219	.578	.563
PC2	1.003	0.02.6	0.4987	-2.925	.004
F-statistic	4.446				
P-value	0.012				
R-Square	75%				

Note B=Unstandardized Estimate, Beta=Standardised value

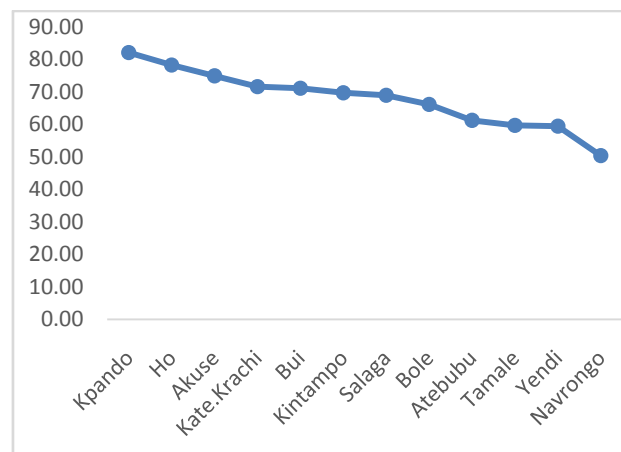


Figure 3. Variable importance (Mean Decrease Gini)

Variable importance test was performed on the data using random forest and the result obtained indicates that rainfall from the catchment area(s) are statistically significant having impact on the water level of the dam. From the figure 3, it could be observed that, among the 12 rainfall catchment areas, rains from Kpando is the most important area having an impact on the Akosombo water level, followed by rains Ho, Akuse, Bui and then Kate-Krachi among others as shown in the figure above. Also, figure 4 shows the multiple

plot of all the rainfall catchment areas. The Geni value for Kpando is estimated to be 81.97, Ho value is 77.23, Akuse is 75.20, Kata Krachi is 70.61 among others as shown in the figure below.

The results in figure 4 below show the multiple plots of the rainfall recorded for a period of 78 years from the twelve (12) catchment areas.

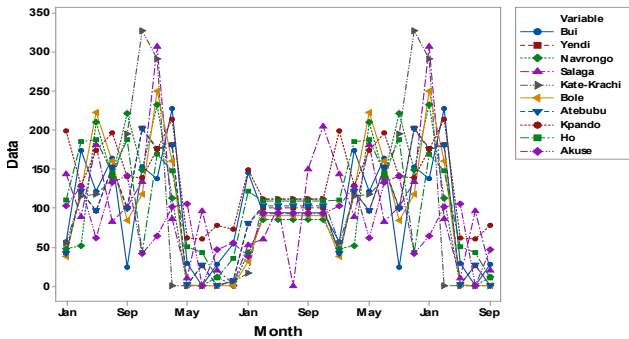


Figure 4. Multiple time plot of the 12 tributaries of the Akosombo water level

14. Seasonal Autoregressive Integrated Moving Average (SARIMA)

Time series analysis was carried on the monthly water level of the dam for the period of 78years. The initial plot of the data shows that there is some level of fluctuation in the dataset. The stationary test was performed and the results show that the dataset is stationary at 5% significance level with Augmented Dickey-Fuller Test value of -5.0468, Lag order=8, p-value=0.01.

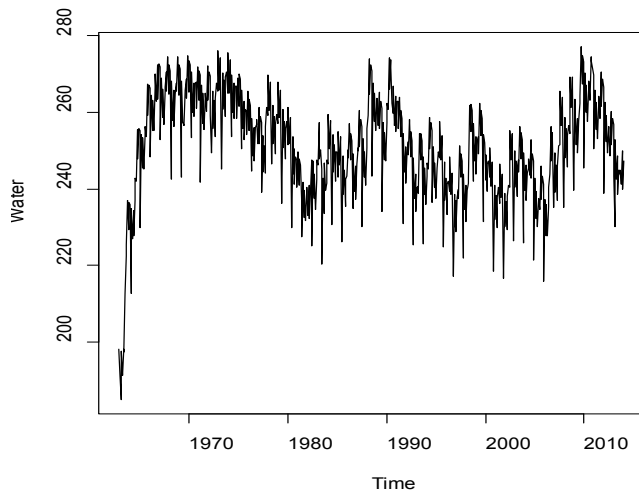


Figure 5. Time series plot of water level

The plot shows some seasonal component and as a result, the appropriate model was selected based on the model with the least AIC. The selected model to fit the data is Seasonal Autoregressive Integrated Moving Average (SARIMA) and from the analysis, the results show that the appropriate model to fit the data gathered is ARIMA (2,1,2) (1,0,0) [12] as indicated in table 9 below.

Table 9. Selection of best model

Model	AIC
ARIMA (2,1,2) (1,0,1) [12]	3901.21
ARIMA (1,1,0) (1,0,0) [12]	3999.20
ARIMA (2,1,2) (2,0,1) [12]	3906.37
ARIMA (2,1,2) (1,0,0) [12]	3926.54
ARIMA (1,1,2) (1,0,0) [12]	3917.91
ARIMA (2,1,1) (1,0,0) [12]	3902.22
ARIMA (1,1,1) (1,0,0) [12]	3925.86
ARIMA (2,1,2) (1,0,0) [12]	3897.48
ARIMA (2,1,2) (2,0,0) [12]	3902.73
ARIMA (1,1,2) (1,0,0) [12]	3916.10

15. Significance of Coefficients

The result of the selected best model has the following coefficient value with their confidence interval at 25% and 97.50%. The results show that the estimated AR1 value is -1.34, with a standard error of 0.08 and a small p-value less than 0.05. AR2 has -0.57 coefficient value and a small p-value less than 0.50. The MA1 component has an estimated value of 0.41, with an error margin of 0.09 and a small p-value of less than 0.5. The MA2 has a coefficient value of -0.13 and a small p-value. However, the seasonal component of the model, SAR1 has an estimated value of 0.72 with an error margin of 0.03 and a small p-value. The result indicates that all coefficients are statistically significant.

Table 10. Coefficient estimate

	Estimate	Std. Error	z	Pr(> z)	25%	97.5%
AR1	-1.34	0.08	-16.78	< 2e-16 ***	-1.49	-1.18
AR2	-0.57	0.05	-11.48	< 2e-16 ***	-0.67	-0.47
MA1	0.41	0.09	4.60	4.2e-06 ***	0.23	0.58
MA2	-0.13	0.06	-2.22	0.02662 *	-0.25	-0.02
SAR1	0.72	0.03	25.07	< 2e-16 ***	0.66	0.78

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

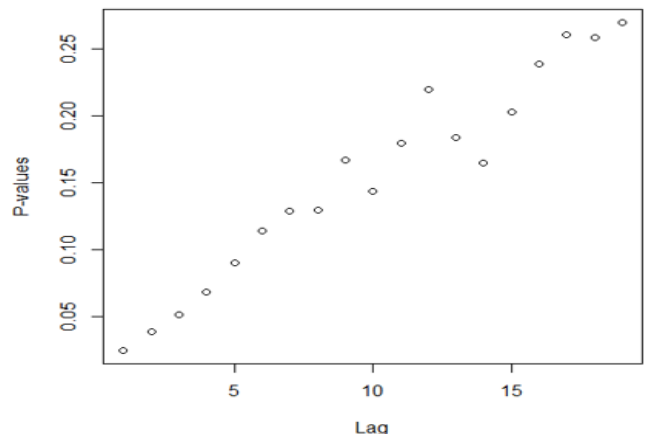


Figure 6. Ljung-Box Q Test

The residual of the model was tested and the results obtained shows the p-value for the Ljung-Box Q test all are well above 0.05, indicating non-significance and this an indication of good and desirable result.

Table 11. Forecast value for the 24months

Month	Year	Point Forecast	Lo 99.5	Hi 99.5
Feb	2014	224.93	208.43	241.43
Mar	2014	247.38	230.84	263.92
Apr	2014	239.06	219.72	258.40
May	2014	241.15	220.60	261.69
Jun	2014	237.52	216.11	258.93
Jul	2014	238.64	215.41	261.86
Aug	2014	240.45	216.57	264.32
Sep	2014	239.72	214.39	265.05
Oct	2014	237.53	211.38	263.69
Nov	2014	244.04	216.84	271.23
Dec	2014	236.34	208.18	264.51
Jan	2015	242.15	213.14	271.17
Feb	2015	225.58	190.68	260.48

Continuation

Mar	2015	242.08	206.31	277.85
Apr	2015	235.93	197.29	274.56
May	2015	237.45	196.95	277.95
Jun	2015	234.9	192.78	277.02
Jul	2015	235.61	191.31	279.91
Aug	2015	237	191.32	282.69
Sep	2015	236.42	188.81	284.02
Oct	2015	234.88	185.81	283.94
Nov	2015	239.55	188.89	290.21
Dec	2015	234	181.82	286.18
Jan	2016	238.2	184.61	291.79

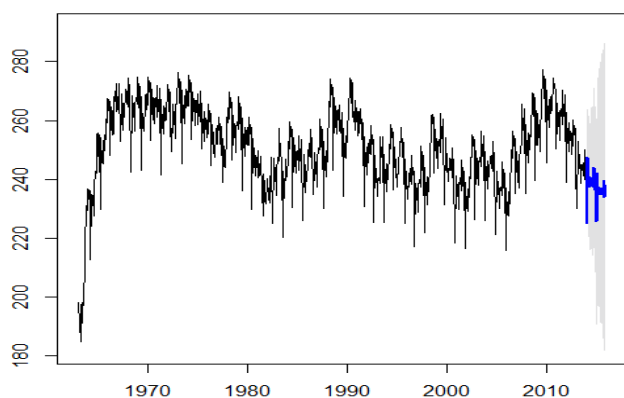


Figure 7. ARIMA (2,1,2) (1,0,0) [12]

The results in Table 11 depict the forecast value of the next 2years (24 months). The estimation was done with a 99.5% confidence level (lower and the upper confidence level) and figure 7 above shows the diagrammatical representation of the forecast values for the period of 24 months. The forecasts are shown as a blue line, with the 80%

prediction intervals as a dark shaded area, and the 95% prediction intervals as a lightly shaded area as indicated in the table 6 above.

16. Conclusions

In this work, the Principal component analysis was used to classify the tributaries of the Akosombo water level into two main components (PC1 & PC2). The variances that were explained by the two components was 76.798% as indicated in the table above. The approach presented here is efficient and appropriate for classification of water level that makes up the Akosombo water level. After the classification, the eigenvectors were regressed on the water level and the results show that PC2 has a significant impact on the water level. This means that, when there are more rains in these catchment areas such as Kintampo, Bui, Kpando, Ho and Akuse, it will have a significant impact on the Akosombo water level. The impact of rains from these catchment areas is approximately 50% and whilst the PC2 has about 22% impact on the dependent variables as indicated in the table above. The estimated RMSE value of the model was 7.07, MAE=4.89, MPE=0.11, MAPE=1.97. The best model was estimated to be; ARIMA (2,1,2) (1,0,0) [12] with an overall AIC value of 3926.54.

17. Recommendations

Further studies must be done to improve upon the results of this study, by considering factors such as accurate prediction of the water level of the dam at any time using other modelling technique such as the Artificial Neural Networks, since most researchers have used it in modelling water resources data. Prediction for a period of 2years (24 months) was estimated at a 99.5% confidence level.

18. Limitation of the Study

This study made use of the rainfall from the river basins areas for its prediction, due to the lack of data on other variables for predictions. For accurate and precise prediction of the water level of the dam, factors such as temperature, humidity and the flow rate of the water from the main tributaries of the river Volta must be included, because these variables have got much impact on the water level. Also, further studies could be done to improve upon the prediction level of the ARIMA (2,1,2) (1,0,0) [12] by the use of Artificial neural networks.

ACKNOWLEDGEMENTS

We are very grateful to the staff of the Akosombo Engineering Department (Akosombo Dam Site) for their contribution towards this study.

REFERENCES

- [1] Box, G.E.P. and G.M. Jenkins (1970). *Time Series Analysis, Forecasting, and Control*. Oakland, CA: Holden-Day.
- [2] El-Dereny, M., & Rashwan, N. I. (2011). Solving multicollinearity problem using ridge regression models. *Int. J. Contemp. Math. Sciences*, 6(12), 585-600.
- [3] Elliott, G., Rothenberg, T. J. and Stock, J. H. (1996). 'Efficient Tests for an Autoregressive Unit Root', *Econometrica*, 64, 813-836.
- [4] Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis with readings* (4th ed). New Jersey: Prentice-Hall International.
- [5] Hair, J. F., Black, B., Babin, B., Anderson, R. E. and Tatham, R. L. (2010). *Multivariate Data Analysis: A Global Perspective*. Pearson Education Inc., NJ.
- [6] Kaiser, H. F. (1970), "A Second-Generation *Little Jiffy*," *Psychometrika*, 35, 401-415.
- [7] Marino, N. A., Srivastava, D. S., MacDonald, A. A. M., Leal, J. S., Campos, A. B., & Farjalla, V. F. (2017). Rainfall and hydrological stability alter the impact of top predators on food web structure and function. *Global change biology*, 23(2), 673-685.
- [8] Montgomery DC, Runger GC (1999). *Applied statistics and probability for engineers*. Wiley, New York.
- [9] Norman, G. R., & Streiner, D. L. (2008). *Biostatistics: the bare essentials*. PMPH-USA.
- [10] Rani, S., & Parekh, F. (2014). Predicting reservoir water level using artificial neural network. *International Journal of Innovative Research in Science, Engineering and Technology*, 3(7), 14489-14496.
- [11] Stevens, J. (2001). *Applied Multivariate Statistics for the Social Sciences*; Lawrence Erlbaum: Mahwah, NJ, USA, 1996.
- [12] Tabachnick, B. G., & Fidell, L. S. (2001). Cleaning up your act: Screening data prior to analysis. *Using multivariate statistics*, 5, 61-116.
- https://en.wikipedia.org/wiki/Volta_River
- <http://www.ghanaiantimes.com.gh/akosombo-dam-level-drops/>