International Journal of Earth Sciences Knowledge and Applications (2021) 3 (1) 1-11

IJESKA
contribution to science

International Journal of Earth Sciences Knowledge and Applications
www.ijeska.com

Research Article

e-ISSN: 2687-5993

Lake Water Level Prediction Model Based on Autocorrelation Regressive Integrated Moving Average and Kalman Filtering Techniques – An Empirical Study on Lake Volta Basin, Ghana

Michael Stanley Peprah[1]*, Edwin Kojo Larbi[1]

[1]*Department of Cartography, Ghana Cocoa Board, Ghana*

INFORMATION

Article history
Received 23 September 2020
Revised 13 October 2020
Accepted 15 October 2020
Available 15 December 2020

Keywords
Lake Volta
Kalman Filter
Stochastic Models
Time Series Analysis
Water Levels

Contact
*Michael Stanley Peprah
E-mail: mspeprah91@gmail.com
https://orcid.org/0000-0002-3819-8898

ABSTRACT

The continuous decline in lake water levels is not only a major concern but also a daunting challenge to policymakers, demanding a backup technological and policy interventions in context of broader political and socio-economic realities. This study used Lake Volta hydrological system to shed light on the extensive and flexible modelling and simulation capabilities of stochastic models to understand the bigger picture of water level (WL) dynamics. The study used Autocorrelation Regressive Integrated Moving Average (ARIMA) and Kalman Filtering (KF) techniques as the proposed optimal stochastic models for the study area. The first order ARIMA (0, 1, 1) was found suitable for predicting the future monthly Lake Volta WL in the presented study based on expert advice and recommendations from existing studies. The statistical performance indicators used were minimum residual error (r_{min}), maximum residual error (r_{max}), arithmetic mean error (*AME*), arithmetic mean squared error (*AMSE*), arithmetic mean absolute percentage deviation (*AMAPD*), and arithmetic standard deviation (*ASD*). Based on the results achieved in this study, ARIMA (0, 1, 1) achieved *AME*, *AMSE*, *AMAPD* and *ASD* of -0.1268 m, 0.0037 m, 0.5749 m, and 0.0033 m respectively. Ensemble of ARIMA and KF was further used to forecast the upcoming monthly WL trends up to December 2048. ARIMA (0, 1, 1) model is found suitable for forecasting Lake Volta WL which shows positive trend up to December 2048. The study further predicted that Lake Volta WL will increase from the current average level of 0.2272 m to an average of 9.1366 m for the next 28 years. The ensuing conclusions stressed the need for checks against over-release of WL for hydropower production and measures for sustainable land and water management in the entire basin. This study can potentially enhance our understanding of hydrodynamic processes in Lake Volta and support water resource management.

1. Introduction

WL predictions using time series analysis has become one of the major researches focuses among geophysicists, geodesists and oceanographers. WL has been used as a defined stable reference datum representing the earth's actual gravity field (geoid) (Abeho et al., 2014; Turner et al., 2013) for measuring vertical distance. This is because for practical applications such as space research, marine studies, coastal engineering, surveying, higher engineering works, and resource exploitation which require a defined datum to which these systems can be referred, maintained and recovered (Vergos et al., 2004; Rummel and Sanso, 1993), the WL are mostly used. In addition, WL studies help geospatial professionals and governmental authorities in decision making on climate adaptation and mitigating strategies to reduce risk in spatial contest (Grgic et al., 2017). Moreover, WL studies contribute significantly to understanding the nature of processes, patterns and interactions of the changing earth (Pashova and Popova, 2011). Precise determination of the WL can be done using available and relevant monthly meteorological data (El-Shazly, 2005). However, this is a difficult task in many scientific and practical fields of applications (Al-Krargy et al., 2017) due to the variation of the WL known as sea surface topography (SST) and meteorological conditions (Vainu and Terasmaa, 2014; El-Shazly, 2005). In view of that, several scholars have proposed different methodologies to estimate the WL with good precision.

Several studies have suggested that, the WL could be determined either by a relative technique (tide gauges) (Tolkatchev, 1996) that provide data with uneven and sparse spatial distribution (Grgic et al., 2017) with reference to a local geodetic datum (geoid) on land or an absolute technique (satellite altimetry) (Mitchum, 2000) which captures high resolution, uniform accuracy measurements with better coverage in open ocean areas and lower coverage in coastal areas (Grgic et al., 2017) with reference to a global reference datum (ellipsoid). In the absence of tide gauges, the classical levelling techniques can be used to determine WL (El-Shazly, 2005). The presented study considered the latter approach (absolute technique) in determining the WL due to the availability of satellite altimeter data for the study area. The various applications of absolute techniques that have been applied in the past and recent years include tides measurement (Tziavos et al., 2005), precise geoid determination (Andersen and Knudsen, 1998, Cazenave et al., 1996), global mean sea surface models (Lemoine et al., 1998; Wenzel, 1998) and regional mean sea surface models (Arabelos and Tziavos, 1996; Vergos et al., 2003) as well as the recovery of gravity anomalies from the altimetry measurements.

Recent studies reveal that, there are defects in the relative techniques for WL determination such as inadequate and inhomogeneous distribution of the used geodetic data (Amin, 2003) and geophysical process within the earth's system that cause changes in global mean sea level (Makarynskyy et al., 2004; Tziavos et al., 2005). Additionally, the classical levelling techniques does not truly represent WL and it varies from place to place (El-Shazly, 2005). The inconsistency of the relative techniques led to the investigation of how modern techniques and abundant data source may be utilized for geodetic purpose (Turner et al., 2013). Precise prediction of WL can be done using sophisticated mathematical models which include time series and climate data (Pashova and Popova, 2011). The traditional method for WL prediction include the classical regression techniques (Demir and Ulke Keskin, 2020; Kaloop et al., 2020; Thiery et al., 2017; Smith and Semazzi , 2014; Awange et al., 2013; El-Shazly, 2005; Jian-Jun, 2003; Coe and Birkett, 2004), Soil and Water Assessment Model (Muthuwatta, 2004), 2-Dimensional Hydro dynamic model (Qi et al., 2019), 3-Dimensional Hydro dynamic model (Kranenburg et al., 2020), KF (Okwuashi and Olayinka, 2017; Adnan et al., 2012) Kriging method (Hassan et al., 2015), and ARIMA (Makwinja et al., 2017; Farajzadeh et al., 2014; Fernandez et al., 2018; Srivastava et al., 2016; Fernandez et al., 2017). Some of these aforementioned techniques are very viable but have limitations due to their inabilities to model noisy data, nonlinearity problems between the dataset and data availability (Yakubu et al., 2018a) (such as lake discharge and meteorological data which were not available) and for that matter not considered in this study.

In Africa, the WL and WL variations have affected related applications in coastal engineering, geophysics studies and space research (El-Shazly, 2005). The WL modelling was established at Lake Volta Basin, Ghana by using the available altimetry data between the years (October 1992 to September 2020). It is self-evident that, diverse factors interact in space and time in complex dynamics to cause these WL changes. The study for the first time in Ghana, applied and assessed the performance of ARIMA and KF as an effective tool for modelling and predicting WL in the study area. In order to resolve the problems associated with nonlinearity problems and error as a result of noise with the ARIMA techniques and limitations of the classical techniques, KF based on time series for WL forecasting was applied. KF is a mathematical power tool that has been applied in numerous fields of scientific studies (Kim and Bang, 2018; Welch and Bishop, 2001).

KF is a linear, discrete time, finite dimensional time-varying system that evaluates the state estimate to minimize their residuals statistical analysis (Ribeiro, 2004). KF provides optimal estimates (Kim and Bang, 2018; Bekhtaovi et al., 2017), and was applied to smoothen the ARIMA parameters for a better estimate. Upon carefully reviewing of existing literature, this aforementioned technique has not been applied within Ghana for WL forecasting. Therefore, this study will help scientist in Ghana to know the abilities of using ARIMA-KF for future forecasting of WL. This is because, the knowledge of WL is very vital for protection of coastal areas and monitoring changes in marine ecosystem.

2. Study Area

The study area (Figs. 1a to 1c) is one of the highly esteemed projects from the period of Africa's decolonization (Lawson, 1970). Lake Volta is the largest artificial reservoir by surface area in the world, covering an approximated total area of 8500 km^2 which represent 3.2 % of Ghana's total land surface (Ndehedehe et al., 2017; Ni et al., 2017). The lake which lies entirely in Ghana is shared by six West African countries namely, Benin, Burkina Faso, Ivory Coast, Ghana, Mali and Togo.

The aim of construction was to produce hydroelectric power (Owusu et al., 2008; Béné, 2007; Gyau-Boakye, 2001), but the reservoirs fisheries have been a significant socio-economic importance to Ghana (Béné and Obirih-Opareh, 2009; Béné and Russell, 2007; Braimah, 2003; Abban, 1999). The Volta basin is primarily underlain by a Voltarian formation consisting of Sandstones, Shales and Mudstones. Another formation is Precambrian, classified into Birimian, Buem and Tarkwaian rocks (Dickson and Benneh, 1977). The basin stretches over four climate regions.

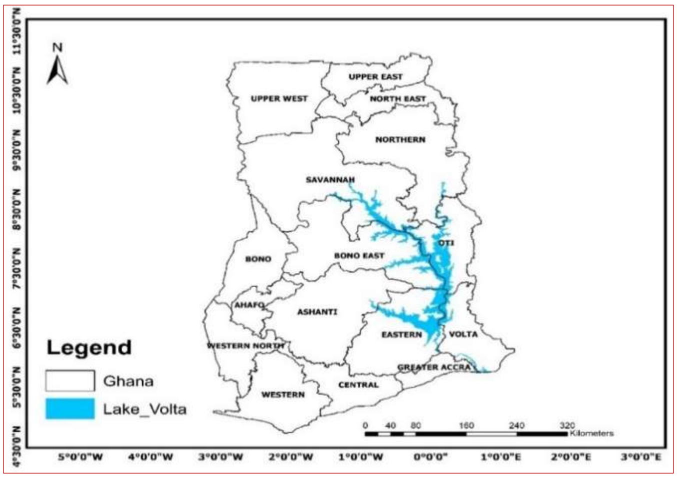

Fig. 1a. Regional map of Ghana showing the study area

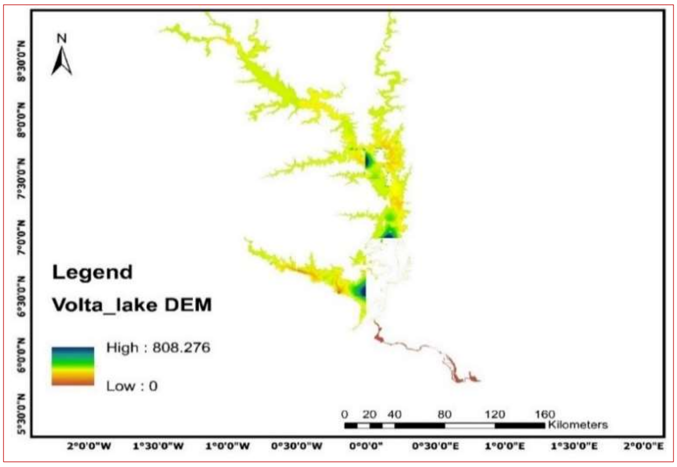

Fig. 1b. Digital Elevation Model of the study area

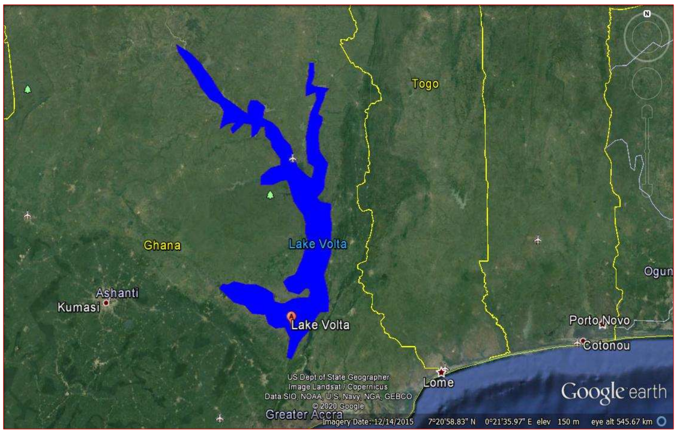

Fig. 1c. Satellite image showing the study area

From lowland rainforest in the South with approximated geographic location of 006° 17′ 20.96" North (N) and 000° 0′ 50.92" West (W) to the Sahel-Sudan desert in the North with geographic location 009° 3′ 8.26" N and 001° 7′ 33.26" W (Rodger et al., 2007). The average topographic elevation of the study is 400 m and the average water level is 4 m. Much of the Volta basin lies in a rainfall region which starts from July to September. The North has only one wet season in September. In the South, there are two rainy seasons which starts from June to July and September to October (Ni et al., 2017).

3. Materials and Methods

3.1. Data

Satellite altimetry data of observed lake height variations computed from (Topex/Poseidon, Jason-1, and Jason-2/OSTM) provided by the United States Department of Agriculture (USDA) were used in this study. The use of altimetry-based measurements for a data deficient region is beneficiary since they are continuous and potentially available few days after measurement (Ndehedehe et al., 2017; Coe and Birkitt, 2004). Thus, time series data of Lake Volta level heights covering the period of 28 years from (October 1992 to September 2020) were downloaded from (*https://www.pecad.fas.usda.gov/cropexplorer/global_reservoir/gr_regional_chart_jason1.aspx?regionid=wafrica&reservoir_name=Volta*) database and used in this study to model and predict Lake Volta WL. As presented in the works of (Ndehedehe et al., 2017; Ni et al., 2017), satellite altimetry data has been successfully validated by comparing altimetric time series and in situ observations. In addition, all classical corrections (polar and solid earth tides, ionospheric and tropospheric delay, and altimeter biases) have been applied to the altimetry data. This was done by using a median type filter to eliminate outliers and reduce high frequency noise.

3.2. Computations of Monthly Mean Lake Water Level

The available time series of Lake Volta WL data for a period of 336 months were used. This period started from October 1992 to September 2020. The monthly Mean Lake Water Level (MLWL) from the given altimetry data were computed according to Eq. 1. However, the monthly meteorological and lake discharge data were not available, hence not included. The geometric technique for estimating geoidal heights using GPS obtained ellipsoidal height *(h)* collocated with orthometric height *(H)* was done according to Eq. 1. The estimated geoidal heights (*N*) derived from GPS ellipsoidal heights and orthometric heights are referred to as NGPS/Levelling.

$$L_{i,j} = \sum_{j=1}^{i} \frac{x_i}{n} \tag{1}$$

where x_i is the time series data, i is an integer ranging from 1 to n, and n is the number of observations. Table 1 is the statistical analysis of the satellite altimeter data; thus, maximum (max), minimum (min), mean value and standard deviation (SD) of the computed monthly MLWL. Figs. 2a and 2b is the monthly distribution and histogram graph analysis of the computed monthly MLWL.

3.3. ARI MA

The ARIMA model introduced by Box and Jenkins (1976) is a widely used technique for time series forecasting (Boye and Ziggah, 2020; Akyen et al., 2016). This type of model is a hybridized model which consists of autoregressive (*AR*) and moving average (*MA*) respectively (Boye and Ziggah, 2020; Yakubu et al., 2018b). In ARIMA (*p, d, q*) modelling, the first step is to check the stationarity of the time series data. When the used time series data is not stationary, it is transformed into a stationary time series by applying the appropriate order of differencing (*d*) (Yakubu et al., 2018b; Akyen et al., 2016). The desired values of autoregressive order (*p*) and moving average (*q*) is acquired by checking the autocorrelation function and partial autocorrelation function of the time series data (Makwinja et al., 2017; Yusof et al., 2013). The *AR(p)* model is a discrete time linear equation with noise as expressed by Eq. 2 (Yakubu et al., 2018b):

$$\chi_t = \alpha_1 \chi_{t-1} + \cdots + \alpha_p \chi_{t-p} + \xi_t \tag{2}$$

where χ_t is the current forecasted model, p is the order, α_1, $\cdots, \alpha_p$ are the parameters of coefficients of the model formed, χ_{t-1}, χ_{t-p} are the previous observations, and ξ_t is the error of the forecast. The *MA(q)* model is an explicit formula for χ_t in terms of noise as given by Eq. 3:

$$\chi_t = \xi_t - \beta_1 \xi_{t-1} - \cdots - \beta_p \xi_{t-p} \tag{3}$$

The difference operator Δ is given by Eq. 4:

$$\Delta \chi_t = \chi_t - \chi_{t-1} = (1-L)\chi_t \tag{4}$$

The ARIMA model with orders (*p, d, q*) is given by Eq. 5 as:

$$\left(1 - \sum_{j=1}^{p} \alpha_j L^j\right)(1-L)^d \chi_t = \left(1 + \sum_{j=1}^{q} \beta_j L^j\right)\xi_t \tag{5}$$

where; L^j is the time lag operator, ξ_t is an error term, and d_i is the order of integration. In this present study, the first order *AR(1)* and *MA(1)* was adopted due to its simplest non-degenerated time-series process (Boye and Ziggah, 2020) as recommended by Makwinja et al. (2017) for Lake WL forecasting. The basic systematic approach utilizing Box-Jenkins methodology was applied in this present study to build the first order ARIMA model giving by Eq. 5. This includes stationarity checks, model identification, parameter estimation, model selection, and diagnostic checking. A detailed literature review about these aforementioned techniques can be found in the works of (Makwinja et al., 2017; Akyen et al., 2016). Both ARIMA and KF models were implemented and coded in MATLAB environment.

Table 1. Statistical Analysis of the computed monthly MLWL (units in meters)

PCI	min	max	mean	SD
MLWL	-0.0400	7.9967	0.2272	0.0114

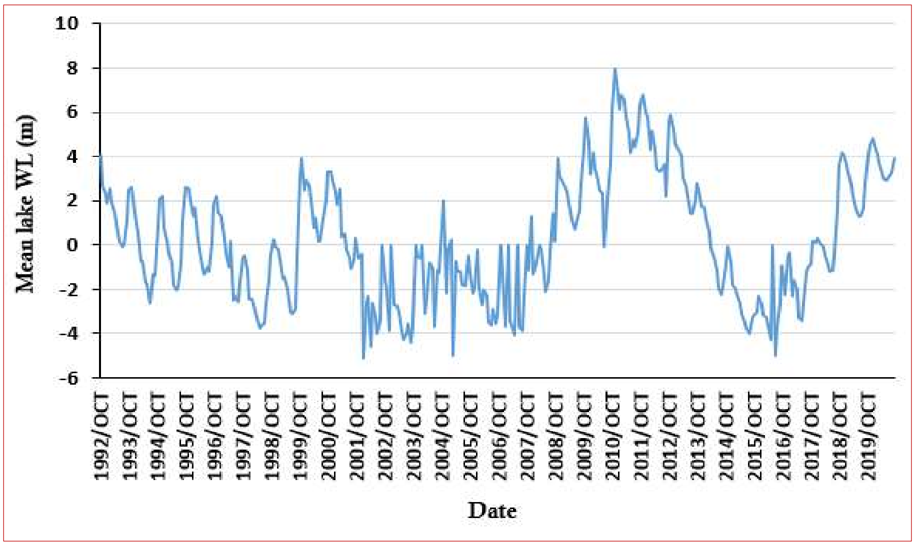

Fig. 2a. Monthly mean distributions of Lake Volta WL

3.4. KF

KF is an estimator for linear-quadratic problem. This technique is widely used for solving problem by estimating the instantaneous state of a linear dynamic system perturbed by white noise, using measurements linearly related to the state but corrupted by white noise. The resulting estimator is statistically optimal with respect to any quadratic function of estimation error (Mohinder and Angus, 2001). The KF model presumes that the state of a system at a time t evolved from the prior state at time $t-1$ according to Eq. 6:

$$\chi_t = \varphi_t \varepsilon_t + \beta_t \sigma_t + w_t \quad (6)$$

where ε_t is the state vector containing the terms of interest, velocity at time t, σ_t is the vector containing any control inputs, φ_t is the state transition matrix which applies the effect of each system state at time *t*. β_t is the control input matrix which applies the effect of each control input parameter in the vector σ_t on the state vector and ω_t is the vector containing the process noise terms for each parameter in the state vector. The process noise is assumed to be drawn from a zero mean multivariate normal distribution with covariance given by the covariance matrix Q. The mathematical representations and detailed computations of the covariance matrix Q can be found in the works of the following existing literatures (Kim and Bang, 2018; Bekhtaovi et al., 2017; Rezaifard and Abbasi, 2017; Ribeiro, 2004; Welch and Bishop, 2001). Fig. 3 shows the mathematical principles used in KF.

3.5. Model performance assessment

To assess the accuracies of the ARIMA model being used, statistical error analysis was carried out. The statistical indicators applied were the mean absolute percentage deviation $(MAPD)$, AME, $AMSE$, r_{min}, r_{max}, and ASD. Their mathematical expressions are given by Eq. 7 to Eq. 12, respectively as:

$$MAPD = \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{a_i - \beta_i}{a_i}\right\| \quad (7)$$

$$AME = \frac{1}{n}\sum_{i=1}^{n}\left(\alpha_i - \beta_i\right) \quad (8)$$

$$AMSE = \frac{1}{n}\sum_{i=1}^{m}\left(\alpha_i - \beta_i\right)^2 \quad (9)$$

$$r_{\max} = \alpha_i - \beta_i \quad (10)$$

$$r_{\min} = \alpha_i - \beta_i \quad (11)$$

$$ASD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\mu - \bar{\mu}\right)^2} \quad (12)$$

where, n is the total number of the observations, α_i and β_i are the measured and predicted lake heights by the ARIMA model, μ denote the residual between the measured and estimated lake level height, $\bar{\mu}$ is the mean of the residual and i is an integer varying from 1 to n.

4 Results and Discussions

4.1 Development of the ARIMA model

The stationarity of the used data was visualized in the form of a data plot as shown in Fig. 2a. The identification of the stationarity in time series data is a necessary condition for stochastic process in building a time series model which is useful for future forecasting (Yakubu et al., 2018b; Makwinja et al., 2017; Akyen et al., 2016). From Fig. 2a, it can be observed that, the time series data from Lake Volta WL is non-stationary due to unstable means which increase and decrease at more points throughout (October 1992-September 2020). According to Makwinja et al. (2017) a good autoregressive model of order *p(AR(q))* has to be stationary and a good moving average model of order *q(MA(q))* has to be invertible. Thus, the invertibility and stationarity gives a constant mean, variance and covariance which is a necessary condition for future forecasting. In order to resolve this difficulty in Lake Volta WL time series data, first order differencing (*d*) of the data and stationary test were embarked on the newly constructed series of the data. Since the newly constructed data is stationary in mean, the next step is to know how to select an appropriate model that can precisely predict the future trend based on the description of the historical pattern in the data and how to determine the optimal model order.

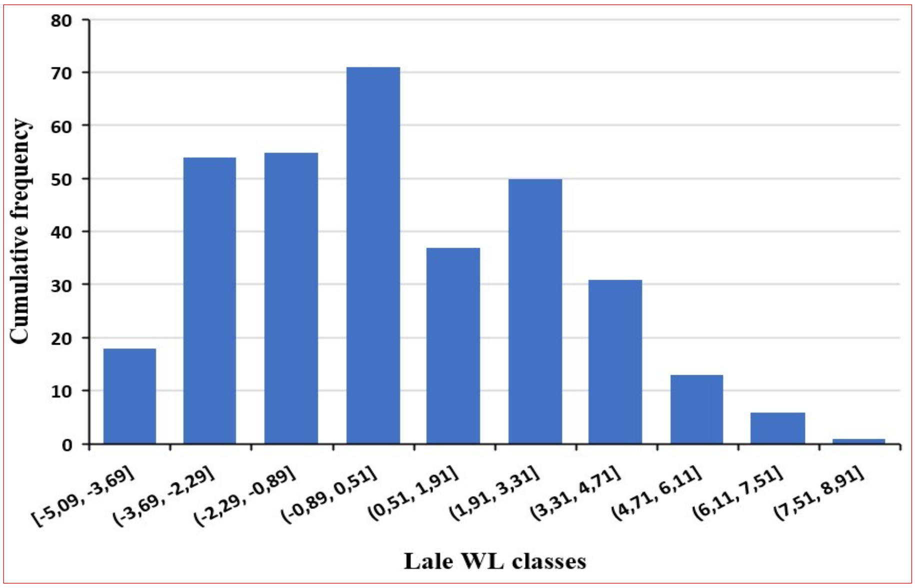

Fig. 2b. Histogram graph analysis of the computed data MLWL

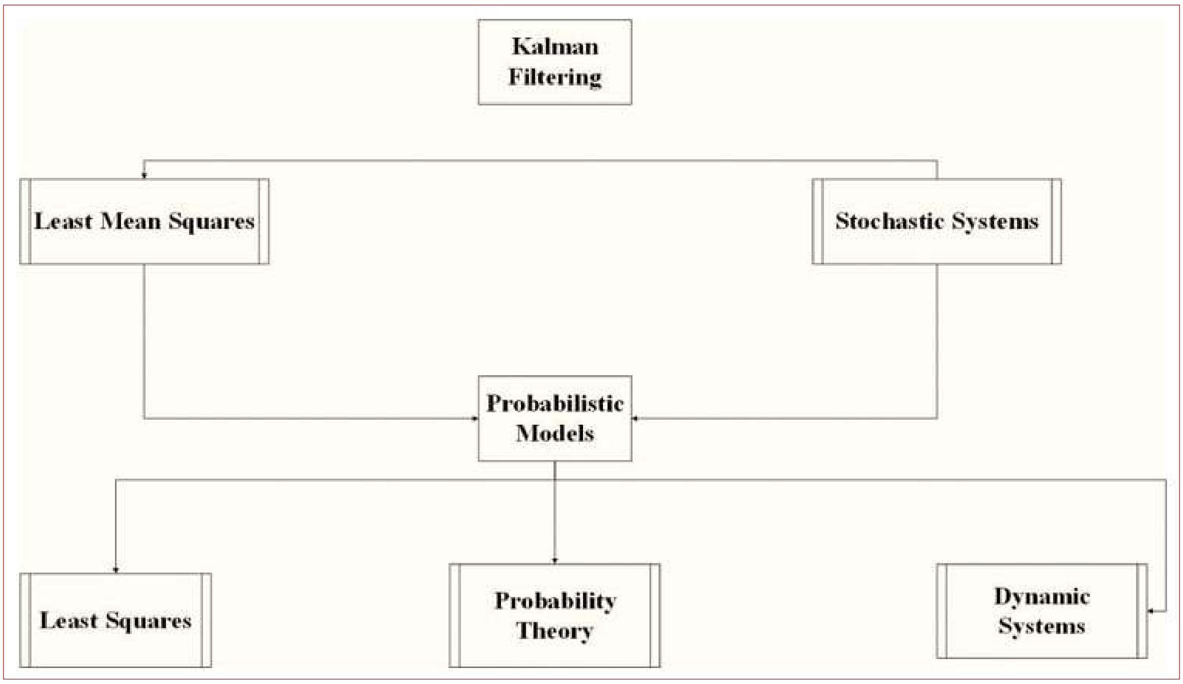

Fig. 3. Mathematical Principles used in KF (Mohinder and Angus, 2001)

However, the first order ARIMA (0, 1, 1) was adopted based on the recommendation by (Boye and Ziggah, 2020; Poku-Gyamfi, 2009; Chen and Hill, 2005; Peprah et al., 2017) and the results achieved in the work of Makwinja et al. (2017). According to the works of (Boye and Ziggah, 2020; Akyen et al., 2016; Makwinja et al., 2017; Poku-Gyamfi, 2009; Chen and Hill, 2005; Peprah et al., 2017), it was observed that, the more complicated the model, the more likely the results deviate from their true mean.

Hence, there was the need to keep the order as low as possible. Table 2 present the model result of ARIMA (0, 1, 1). The *AMSE*, which measured how much dependent series varies from its model predicted level was minimal in ARIMA (0, 1, 1) model which indicated a good forecast of the model. Conversely, *AMAPD* which is also known as *AMAPE*, which is a measure of prediction accuracy of a forecasting method in statistics, for instance in trend estimations was found to be 0.5749. Hence, the opposite signed errors did not offset each other. This implies that ARIMA (0, 1, 1) model indicated a good forecast of the Lake Volta WL. Additionally, *AME* was very small in value of -0.1268 m. It was further noted that, *ASD* was lowest and smallest in ARIMA (0, 1, 1) model. This implies that, the forecasted values do not significantly deviate from their true mean, and the error occurring was very small. Furthermore, r_{min} and r_{max} was found to be 0.0029 m and -1.9870 m respectively. According to Makwinja et al. (2017), the best model for forecasting Lake WL should have adequate accurate statistical performance as low as possible for it to have accurate forecasts. Therefore ARIMA (0, 1, 1) model was selected as the optimal model for WL forecasting for the study area based on the lowest statistical performance and recommendations by researchers. It was further observed that the coefficients of the parameters of ARIMA (0, 1, 1) model was significant ($p<0.05$). According to the works of (Boye and Ziggah, 2020; Makwinja et al., 2017; Akyen et al., 2016), the ARIMA model which indicate lowest normalized Bayesian Information Criterion (NBIC) is significant ($p<0.05$), hence, a better model in terms of forecasting performance than with large NBIC. Fig. 4a shows the model result of the ARIMA (0, 1, 1) model. Based on the results achieved in the present study findings, the most suitable model for forecasting Lake Volta WL for the study area was confirmed to be ARIMA (0, 1, 1).

4.2. Development of the future forecasting model

Using the selected ARIMA (0, 1, 1) model, the forecast of Lake Volta WL was made from (October 2020-December 2048). In order to resolve the noise issues with the ARIMA model, the KF was applied to denoise, filter, and reconstruct the forecasted data in order to achieve a better estimate for the study area. For precision and accuracy sake, model results discussed in section 4.1 was used to validate the forecast future trends for the study area. Table 3 is the statistical analysis of the forecasted trend up to 2048. Figs. 4b to 4d show the forecast results from (October 2020-December 2048) using the standalone ARIMA (0, 1, 1) model and ensemble ARIMA-KF respectively.

Table 2. Statistical analysis of the ARIMA (0, 1, 1) model (units in meters)

PCI	r_{min}	r_{max}	*AME*	*AMSE*	*AMAPD*	*ASD*
ARIMA	0.0029	-1.9870	-0.1268	0.0037	0.5749	0.0033

Table 3. Statistical analysis of the forecasted Lake WL (units in meters)

PCI	min	max	*AME*	*ASD*
ARIMA-KF	3.0041	15.3742	9.1366	0.0116

According to Makwinja et al., (2017), the forecasted and actual values need to be very close. This implies that, the forecasting error must be very minimal for the model to be considered as optimal. As observed in Table 2, the noise residuals were a combination of the positive and negative errors indicating that, the model had a better performance of forecasting the Lake WL. It was also observed that, the *AME* was very minimal indicating a good forecasting performance. In Figs. 4b to 4d, it is obvious that, Lake Volta WL are fluctuating with a positive trend. Such positive trend will continue up to December 2048. Moreover, it can also be observed from Fig. 2a that, the water levels were inclining towards the ending of 2020.

According to the work of Makwinja et al., (2017), it was revealed that, the basic principle of ARIMA model assumes that, time series data is linear and follows a normal distribution. Therefore, it can be concluded based on the results achieved in this study that, the trend in this present study behaved in a manner consistent with ARIMA principle which is assumed to follow a certain probability model described by joint distribution of random variable.

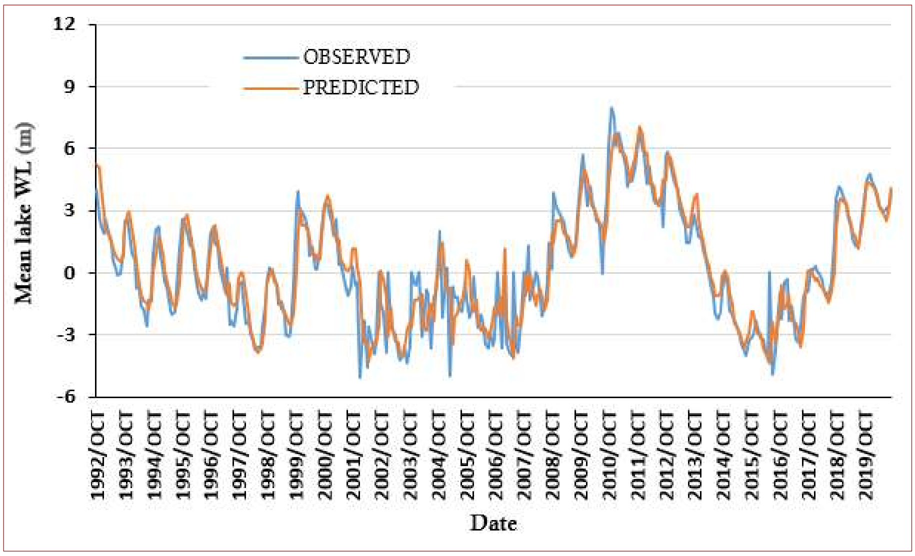

Fig. 4a. Model results of the ARIMA (0, 1, 1)

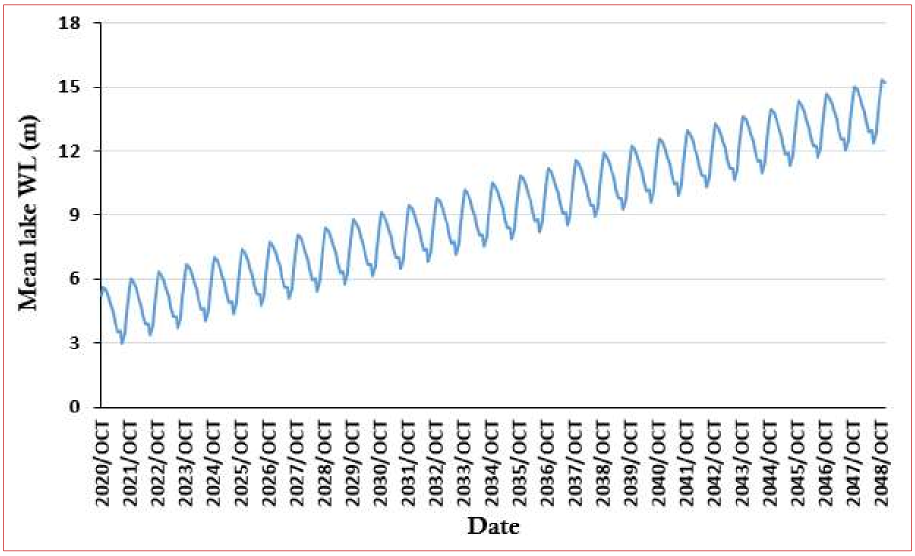

Fig. 4b. Model result of the standalone ARIMA (0, 1, 1) model

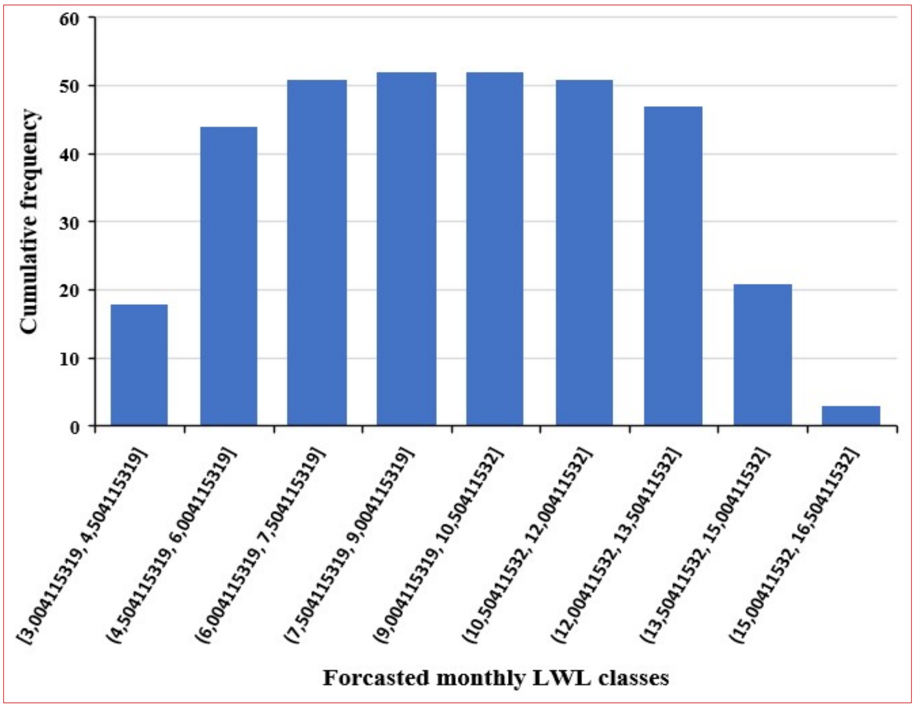

Fig. 4c. Histogram graph analysis of the forecasted water levels

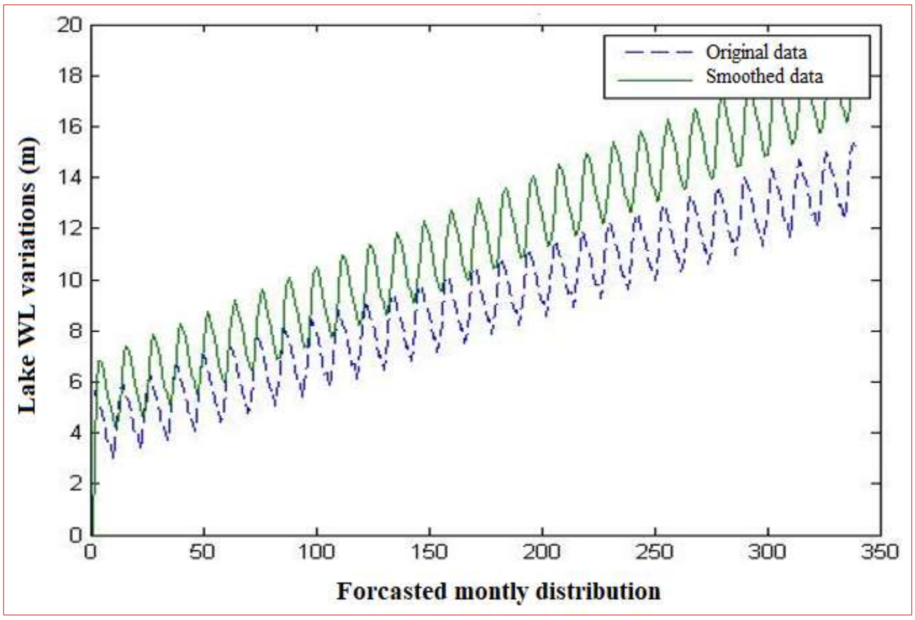

Fig. 4d. Model results of the ensemble ARIMA-KF models

In addition, time series is non-deterministic in nature such that, it cannot predict with certainty that will occur in the future (Makwinja et al., 2017). Based on the achieved results, the study indicated that, there is a high probability that Lake Volta WL will increase as far up to 9.1366 m by 2048. This will affect the socio-economic impacts of agricultural productivity and hydro-power generations as recorded in the works of (Vanderkelen et al., 2018; Hirsch et al., 2014; Béné and Obirih-Opareh, 2009). This future inclining can be attribute to climate change or human anthropogenic activities (Akurut et al., 2014; Vainu and Terasmaa, 2014).

However, the positive trend of Lake Volta WL predicted in the present study is a future risk and needs proper attention in order to mitigate its menace. Moreover, with such future prediction, deliberate effort has to be to done to find appropriate policy options and strategies for sustaining Lake Volta WL.

5. Conclusions and Recommendations

WL studies have become obligatory in establishing a vertical geodetic reference network for measuring vertical distance and proper management decision analysis. This study for the first time in Ghana investigates the theoretical and practical analysis of ARIMA and KF in modelling and predicting the monthly WL of Lake Volta Basin. The WL was found to be varying from time to time due to the change in the weather conditions, temporal variations, and water surface topography. The mathematical model that represents the WL variations have been presented. The available satellite data for twenty-eight years thus, (October 1992- September 2020) was used to model and predict the monthly WL. Performance criteria indices of r_{min}, r_{max}, *AME, AMSE, AMAPD* and *ASD* were adopted to assess the working efficiency and performance of the proposed model. The study selected the first order ARIMA (0, 1, 1) model for forecasting Lake Volta WL based on the recommendations by existing studies. The ARIMA (0, 1, 1) had *AME*, *AMSE* , *AMAPD* and *ASD* of-0.1268 m, 0.0037 m, 0.5749 m, and 0.0033 m respectively which indicated a good forecast of the model. Based on the selected model, it is very apparent that Lake Volta WL fluctuation is showing a positive trend. Such positive trend is predicted to continue up to December 2048. The model further predicted that Lake Volta water levels will increase up to with a *min*, *max, AME*, and *ASD* of 3.0041m, 15.3742 m, 9.1366 m, and 0.0116 m respectively by December 2048. This study provides critical information for future policy making and formulation of intervention strategies for sustaining Lake Volta WL. Based on the results achieved in this study, it is recommended that, ARIMA and KF models be adopted in modelling and predicting MLWL for the study area. In this study, we provide the indications of potential consequences of WL changes for the Lake Volta Basin. The large biases and uncertainties present in the projections stress the need for validation to adequately represent the future trends of WL to be able to make reliable decisions in the basin region. Finally, the evolution of future Lake WL of Lake Volta is primarily determined by the decisions made at the basin. Therefore, the basin management of Lake Volta is of major concern to ensure the future of the people living in the basin, the future hydropower generation and water availability downstream. The major limitation of ARIMA model in this study was that it only captured short-range dependence (SRD). In other words, it belongs to the conventional integer models. In practice, several time series exhibit long range dependence (LRD) in their observations. To overcome this difficulty, it is recommended that a similar study should be conducted using ARFIMA model with ability to capture long range property of the fraction system accordingly and project extended period of more than 28 years. However, more work should be done in Ghana utilizing soft computing techniques which were not considered in this study. Notably among them are deep learning Convolutional Neural Networks (CNN), Least Squares Support Vector Machines (LSSVM), Wavelet Transform (WT) analysis, Extreme Learning Machine (ELM) to classically improve regression techniques such as Gaussian Regression and Kernel Ridge Regression to evaluate its effectiveness for further analysis.

Acknowledgement

We are much grateful to Dr. Christopher E. Ndehedehe, Department of Spatial Sciences, Curtin University, Perth, Australia and Mr. Abubakar Saani Mohammed, Department of Surveying and Mapping Science and Engineering, Hohai University, Nanjing, China for assisting in acquisition of data used for the study area.

Conflict of Interest

The Authors declare that there are no potential conflict of interests in this work. The work does not infringe any copyright, proprietary right or any other right of any third party; and the Authors are the sole owners of the work.

References

Abban, E.K., 1999. Integrated Development of Artisanal Fishes. Integrated Development of Artisanal Fisheries Project, GHA/93/008, 1-43.

Abeho, D.R., Hipkin, R., Tulu, B.B., 2014. Evaluation of EGM08 by means of GPS levelling Uganda. South African Journal of Geomatics 3 (3), 272-284.

Adnan, R., Ruslan, F.A., Samad, A.M., Zain, Z.M., 2012. Artificial Neural Network Modelling and Flood Water Level Prediction Using Extended Kalman Filter. 2012 IEEE International Conference on Control System, Computing and Engineering, 23-25 November 2012, Penang, Malaysia, 535-538.

Akyen, T., Boye, C.B., Ziggah, Y.Y., 2016. Municipal Solid Waste Estimation and Landfill Lifespan Prediction. 4th UMaT Biennial International Mining and Mineral Conference, Tarkwa, Ghana, GG127-133.

Al-Krargy, E.M., Mohamed, H.F., Hosney, M.M., Dawod, G.M., 2017. A High-Precison Geoid for Water Resources Management: A Case Study in Menofia Governorate, Egypt. National Water Research Center (NWRC) Conference on: Research and Technology Development for Sustainable Water Resources Management, Cairo, Egypt, 1-13.

Akurut, M., Williams, P., Niwagaba, C.B., 2014. Potential Impacts of Climate Change on Precipitation Over Lake Victoria, East Africa, in the 21st Century. Water 6, 2634-2659.

Amin, M.M., 2003. An up to Date Precise 5′ * 5′ Geoid Grid for Egypt by Collocation Tecnique. Port-Said Engineering Research Journal (PSERT), Published by faculty of Engineering, Suez Canal University, Port-Said, Egypt, 1-22.

Andersen, O.B., Knudsen, P., 1998. Global marine gravity field from the ERS1 and Geosat geodetic mission. J Geophys Res, 103, 8129–8137.

Arabelos, D., Tziavos, I.N., 1996. Combination of ERS1 and TOPEX altimetry for precise geoid and gravity recovery in the Mediterranean Sea. Geophysical Journal International 125, 285-302.

Awange, J.L., Forootan, E., Kusche, J., Kiema, J.B.K., Omondi, P.A., Heck, B., Fleming, K., Ohanga, S O., Goncalves, R.M. 2013. Understanding the Decline of Water Storage Across the Ramsev-Lake Naivasha Using Satellite-Based Methods. Advances in Water Resources, 60, 7-23.

Bekhtaovi, Z., Abdelkrim, M., Mohammed, D., Abed-Meraim, K., 2017. Maneuvering Target Tracking Using q-Learning Based Kalman Filter. The 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B) October 29-31 2017, Boumerdes, Algeria, 1-5.

Béné, C., 2007. Diagnostic study of the Volta Basin fisheries Part 1 - Overview of the fisheries resources. Volta Basin Focal Project Report No 6. World Fish Center Regional Offices for Africa and West Asia, Cairo Egypt, and CPWF, Colombo, Sri Lanka, 1-31.

Béné, C., Russell, A.J.M., 2007. Diagnostic study of the Volta Basin fisheries. Part 1 - Livelihoods and poverty analysis, current trends and projections. *Volta* Basin Focal Project Report No 7. World Fish Center Regional Offices for Africa and West Asia, Cairo Egypt, and CPWF, Colombo, Sri Lanka, 1-67.

Béné, C., Obirih-Opareh, N., 2009. Social and Economic Impacts of Agricultural Productivity Intensification: The Case of Brush Park Fisheries in Lake Volta. Agricultural Systems, 102, 1-10.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control, Holden-Day, Boca Raton, Fla, USA.

Boye, P., Ziggah, Y.Y., 2020. A Short-Term Stock Exchange Prediction Model Using Box-Jenkins Approach. Journal of Applied Mathematics and Physics 8, 766-779.

Braimah, L.I., 2003. Fisheries Management Plan for the Volta Lake, Ministry of Food and Agriculture, Directorate of Fisheries, Accra, Ghana, 1-77.

Cazenave, A., Schaeffer, P., Berge, M., Brosier, C., Dominh, K., Genero, M.C. 1996. High-resolution mean sea surface computed with altimeter data of ERS1 (Geodetic Mission) and TOPEX/POSEION. Geophysical Journal International 125, 696-704.

Chen, W., Hill, C. 2005. Evaluation Procedure for Coordinate Transformation. Journal of Surveying Engineering, 43-49.

Coe, M.T., Birkett, C.M., 2004. Calculation of River Discharge and Prediction of Lake Height from Satellite Radar Altimetry: Example for the lake Chad Basin. Water Resources Research 40 (W10205), 1-11.

Demir, V., Ulke Keskin, A., 2020. Height Modelling with Artificial Neural Networks (Samsun_Mert River Basin). Journal of Engineering Sciences, 6 (1), 54-61.

Dickson, K. B., Benneh, G., 1977. A New Geography of Ghana. Longman Group, London, UK, 1-173.

El-Shazly, A.H., 2005. The Efficiency of Neural Networks to Model and Predict Monthly Mean Sea Level from Short Spans Applied to Alexandria Tide Gauge. From Pharaohs to Geoinformatics, FIG Working Week 2005 and GSDI-8, Cairo, Egypt, April 16-21, 2005, 1-13.

Farajzadeh, J., Fard, A.F., Lotfi, S., 2014. Modelling of Monthly Rainfall and Runoff of Urmia Lake Basin Using Feed-Forward Neural Network and Time Series Analysis Model. Water Resources and Industry 7 (8), 38-48.

Fernandez, F.R.Q., Montero, N.B., III, R.B.P., Addawe, R.C., Diza, H.M.R., 2018. Forecasting Manila South Harbour Mean Sea Level Using Seasonal ARIMA Models. Journal of Technology Management and Business 5 (1), 1-7.

Fernandez, F.R., III, R.P., Montero, N., Addawe, R., 2017. Prediction of South China Sea Level Using Seasonal ARIMA Models. Proceedings of the 13th IMT-GT International Conference on Mathematics, Statistics and their Applications (ICMSA 2017), 050018-1-050018-6.

Grgic, S.M., Jukic, S., Nerem, R.S., Basic, S.T. 2017. The Assessment of an Impact of Mean Sea Level change in the Mid-Adriatic Region Based on Satellite Altimeter Records. 17th International Multidisciplinary Scientific Geoconference SGEM 2017, Section Photogrammetry and Remote Sensing, 263-270.

Gyau-Boakye, P., 2001. Environmental Impacts of the Akosombo Dam and Effects of Climate Change on the Lake Levels. Environment, Development and Sustainability 3 (1), 17-29.

Hassan, O., Elnazeer, E., Zomrawi, N., 2015. Application of Artificial Neural Network for Height Modelling. International Journal of Recent and Innovation Trends in Computing and Communication 3 (3), 1374-1377.

Hirsch, P.E., Schillinger, S., Weigt, H., Burkhardt-Holm, P., 2014. A Hydro-Economic Model for Water Level Fluctuations: Combining Limnology with Economics for Sustainable Development of Hydropower. PLoS ONE, 9 (12), 1-26.

Jian-Jun, S., 2003. Prediction and Analysis of Tides and Tidal Currents. International Hydrographic Review, 4 (2), 24-29.

Kaloop, M.R., Beshr, A.A.A., Zarzoura, F., Ban, W.H., Hu, J.W., 2020. Predicting Lake Wave Height Based on Regression and Multi Input-Single Output Soft Computing Models. Arabian Journal of Geosciences 13 (591), 1-14.

Kranenburg, W., Tiessen, M., Veenstra, J., de Graaff, R., Uittenbogaard, R., Bouffard, D., Sakindi, G., Umutoni, A., Van de Walle, J., Thiery, W., 2020. 3D-Modelling of Lake Kivu: Horizontal and Vertical Flow and Temperature Structure Under Spatially Variable Atmospheric Forcing. Journal of Great Lakes Research 46 (4), 947-960.

Kim, Y., Bang, H., 2018. Introduction to Kalman Filter and Its Application. Published Book, Korea Advanced Institute of Science and Technology, Daejeon, South Korea, 1-19.

Lawson, G.W., 1970. Lessons of the Volta - A New Man-made Lake in Tropical Africa. Africa Biological Conservation 2 (2), 91-96.

Lemoine, F.G., Kenyon, S.C., Factor, J.K., Trimmer, R.G., Pavlis, N.K., Chinn, D.S., Cox, C., Klosko, S.M., Luthcke, S.B., Torrence, M.H., Wang, Y.M., Williamson, R.G., Pavlis, E.C., Rapp, R.H., Olson, T.R., 1998. The development of the join NASA GSFC and NIMA geopotential model EGM96. NASA Technical Paper, 1-4.

Makwinja, R., Phiri, T., Kosamu, I.B.M., Kaonga, C.C., 2017. Application of Stochastic Models in Predicting Lake Malawi Water Levels. International Journal of Water Resources and Environmental Engineering 9 (9), 191-200.

Makarynskyy, O., Kuhn, M., Makarynska, D., Featherstone, W.E., 2004. The Use of Artificial Neural Networks to Retrieve Sea Level Information from Remote Data Sources. IAG International Symposium: Gravity, Geoid and Space Missions, Porto, Portugal, August 30-September 3, 2004, Springer Verlag, Berlin, Germany, 1-4.

Mitchum, G.T., 2000. An Improved Calibration of Satellite Altimetric Heights using Tide Gauge Sea Levels with Adjustment for land Motion. Marine Geodesy 23, 145-166.

Mohinder, S.G. Angus, P.A., 2001. Kalman Filtering: Theory and Practice Using MATLAB. Wiley-Interscience Publication, New York, 1-410.

Muthuwatta, L.P., 2004. Long Term Rainfall-Runoff-Lake Level Modelling of the Lake Naivasha Basin, Kenya. Published MSc Dissertation, Water Resources Survey and Environmental Systems Analysis and Management, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands, 1-89.

Ndehedehe, C.E., Awange, J.L., Kuhn, M., Agutu, N.O., Fukuda, Y., 2017. Analysis of Hydrological Variability over the Volta River Basin using in-situ Data and Satellite Observations. Journal of Hydrology: Regional Studies, 12, 88-110.

Ni, S., Cehn, J., Wilson, C.R., Hu, X., 2017. Long-Term Water Storage Changes of Lake Volta from GRACE and Satellite Altimetry and Connections with Regional Climate, Remote Sensing 9 (842), 1-15.

Okwuashi, O., Olayinka, D.N., 2017. Tide Modelling Using the Kalman Filter. Journal of Spatial Science, 62(2), 353-365.

Owusu, K., Waylen, P., Qiu, Y., 2008. Changing Rainfall Inputs in the Volta Basin: Implications for Water Sharing in Ghana. GeoJournal, 71(4), 201-210.

Pashova, L., Popova, S., 2011. Daily Sea Level Forecast at Tide Gauge Burgas, Bulgaria Using Artificial Neural Networks. Journal of Sea Research 66, 154-161.

Peprah, S.M., Yevenyo, Y.Y., Issaka, I., 2017. Performance Evaluation of the Earth Gravitational Model (EGM2008) – A Case Study. South African Journal of Geomatics 6 (1), 47-72.

Poku-Gyamfi, Y., 2009. Establishment of GPS Reference Network in Ghana. Published MPhil Dissertation, Universitat Der Bundeswehr Munchen Werner Heisenberg-Weg 39, 85577, Germany, 1-218.

Qi, L., Huang, J., Gao, J., Cui, Z., 2019. Modelling the Impacts of Bathymetric Changes on Water Level in China's Largest Freshwater Lake. Water 11 (1469), 1-16.

Rezaifard, E., Abbasi, P., 2017. Inertial Navigation System Calibration Using GPS Based on Extended Kalman Filter. 25th Iranian Conference on Electrical Engineering (ICEE), May 2017, Iran, 1-5.

Ribeiro, M.I., 2004. Kalman and Extended Kalman Filters: Concept, Derivation and Properties. Institute for Systems and Robotics, Av. Rovisco Pais, 1, 14049-001 Lisboa Portugal, 1-44.

Rodgers, C., Van de Giesen, N., Laube, W., Vick, P.L.G., Youkhana, E., 2007. The GLOWA Volta Project: A Framework for Water Resources Decision-making and Scientific Capacity Building in a Transnational West African Basin. In Integrated Assessment of Water Resources and Global Change: A North-South Analysis, Springer: Dordrecht, The Netherlands, 295-313.

Rummel, R., Sanso, F., 1993. Satellite Altimetry in Geodesy and Oceanography. Springer-Verlag, Berlin Heidelberg, 1-469.

Smith, K. A., Semazzi, F.H.M., 2014. The Role of the Dominant Modes of Precipitation Variability Over Eastern Africa in Modelling the Hydrology of Lake Victoria. Advances in Meteorology 2014 (516762), 1-11.

Srivastava, P.K., Islam, T., Singh, S.K., Petropoulos, G.P., Gupta, M., Dai, Q., 2016. Forecasting Arabian Sea Level Rise Using Exponential Smoothing State Space Models and ARIMA from Topex and Jason Satellite Radar Altimeter Data. Meteorological Applications 23, 633-639.

Thiery, W., Gudmundison, L., Bedka, K., Semazzi, F.H.M., Lhermitte, S., Willems, P., Van Lipzig, N. P. M., Seneviratne, S. I., 2017. Early Warnings of Hazardous Thunderstorms Over Lake Victoria. Environmental Research Letters 12, 1-9.

Tolkatchev, A., 1996. Global Sea Level Observing System. Marine Geodesy, 19, 21-62.

Turner, J.F., Iliffe, J.C., Ziebart, M.K., Jones, C., 2013. Global Ocean Tide Models: Assessment and use Within a Surface Model of Lowest Astronomical Tide. Marine Geodesy 36, 123-137.

Tziavos, I.N., Vergos, G.S., Kotzev, V., Pashova, L., 2005. Mean Sea Level and Sea Level Variation Studies in the Black Sea and the Aegean. Gravity, Geoid and Space Missions, 254-259.

Vainu, M., Terasmaa, J., 2014. Changes in Climate, Catchment Vegetation and Hydrogeology as the Causes of Dramatic Lake-Level Fluctuations in the Kurtna Lake District, NE Estonia. Estonia Journal of Earth Sciences 63 (1), 45-61.

Vanderkelen, I., Van Lipzig, N.P.M., Thiery, W., 2018. Modelling the Water Balance of Lake Victoria (East Africa) – Part 2: Future Projections. Hydrology and Earth System Sciences, 22, 5527-5549.

Vergos, G.S., Tziavos, I.N., Andritsanos, V.D., 2004. On the Determination of Marine Models by Least Squares Collocation and Spectral Methods using Heterogeneous Data. International Association of Geodesy Symposia 128, 1-382.

Vergos, G.S., Tziavos, I.N., Andritsanos, V.D., 2003. On the Determination of Marine Geoid Models by Least Squares Collocation and Spectral Methods using Heterogeneous Data. Presented at Session G03 of the 2004 IUGG General Assembly, Sapporo, Japan, July 2-8, 2003, 1-5.

Welch, G., Bishop, G., 2001. An Introduction to the Kalman Filter. Published Lecture Book, Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3175, 1-50.

Wenzel, H.G., 1998. Ultra-high degree geopotential model GPM3E97 to degree and order 1800 tailored to Europe. Presented at the 2nd Continental Workshop on the geoid in Europe, Budapest, Hungary, 1-4.

Yakubu, I., Ziggah, Y.Y., Peprah, M.S. 2018a. Adjustment of DGPS Data using artificial intelligence and classical least square techniques. Journal of Geomatics 12 (1), 13-20.

Yakubu, I., Ziggah, Y.Y., Baafi, K.A., 2018. Prediction of Tidal Effect in Crustal Deformation Monitoring: A Geodetic Perspective. Ghana Journal of Technology 2 (2), 63-69.

Yusof, F., Kane, I.L., Yusof, Z., 2013. Hybrid of ARIMA-GARCH Modelling in Rainfall Time Series. Journal Teknologi (Sciences and Engineering) 63 (2), 27-34.