Applied Geomatics (2024) 16:515–528
https://doi.org/10.1007/s12518-024-00564-9

ORIGINAL PAPER

Estimating future bathymetric surface of Kainji Reservoir using Markov Chains and Cellular Automata algorithms

Pius Onoja Ibrahim[1] · **Harald Sternberg**[2] · **Lazarus Mustapha Ojigi**[3]

Received: 23 October 2023 / Accepted: 25 June 2024 / Published online: 3 July 2024

Abstract

The menace of sedimentation to reservoirs has a significant implication for water quality, storage capacity and reservoir lifetime. Rainfall patterns and other anthropogenic and environmental impacts alter the erosion rate and, by extension, directly affect sedimentation rates if left unchecked. This research focused on using the integration of Markov Chains and Cellular Automata (MC – CA) models to estimate and forecast the future bathymetric surface of the Kainji reservoir in Nigeria for the year 2050. The bathymetric datasets used for this research comprise two different epochs (1990 and 2020). The datasets were acquired using a Single Beam Echosounder at Low and High frequencies of 20 kHz and 200 kHz. The preliminary investigation revealed that sedimentation is exacerbating a greater danger to the reservoir functionality. The results show that the maximum observed depth is 71.2 m, indicating a 7.53% loss in depth from the 1990 archived data and a 16.24% depth loss to sedimentation from 1968 to 2020 and 22.35% depth loss in the year 2050 as shown on the projected surface. Consequently, the integrated model (MC and CA) efficiently predicted the future bathymetric surface of the Kainji reservoir for the year 2050 based on the data characteristics. However, the proven techniques for analysing spatial data, such as the Markov Chain and Cellular Automata, best suited for analysing categorical transition data, show some artefacts (black spots) on the projected generated map which is subject to further investigation.

Keywords Markov Chains · Cellular Automata · Sedimentation and reservoir

Introduction

Bathymetric survey has shown to be a remarkable approach to assessing and estimating the rate of sediment deposits in reservoirs and other water bodies (Ibrahim and Sternberg 2021). The menace of sedimentation and siltation is a continuous process that affects the optimal performance of any reservoir over time if left unchecked (Girish et al. 2014; Estigoni et al. 2014). A greater contributing factor to sediment inflow is the topography gradient which is enhanced by precipitation and the adjoining rivers to a reservoir (Horton 1945; Hansen and Boss 2000); including the hydrology of the catchment and the river basin characteristics. This phenomenon affects the optimal operational capacity of dams and reduces the volumetric capacity, irrigational activities, and to some extent domestic supplies (Dargahi 2012). Consequently, repeated bathymetry has become the most common and generally accepted technique for estimating long-term sediment accumulation in reservoirs (Vahid et al. 2018; Mohammad et al. 2018). This involves using the data from different epochs, such as obtaining the data before impoundment and the bathymetric survey data collected after a period of years (Psilovikos and Margoni 2010; Veli et al. 2019). These two datasets will show the volume lost to sedimentation and erosion. However, a single beam echo sounder is employed to adequately measure the amount of sediment deposited in a specific water body. This is due to the capacity of low-frequency pings to penetrate suspended sediment to a greater degree, which is one of the drawbacks of multibeam echosounder because it uses high frequency for its operations (El-Hattab 2014). The multibeam echosounder is

✉ Pius Onoja Ibrahim
pius.onoja@futminna.edu.ng

[1] Department of Surveying and Geoinformatics, Federal University of Technology, PMB 65, Minna, Niger State, Nigeria

[2] Department of Geodesy and Hydrography, HafenCity University Hamburg, 20457 Hamburg, Germany

[3] African Regional Institute for Geospatial Information Science and Technology (AFRIGIST), Obafemi Awolowo University Campus, Ile-Ife, Nigeria

adequately suitable for complete seafloor mapping because of its multibeam, but lack the capacity to penetrate sediment layers due to its high-frequency calibration (Xavier 2010). Hence, the difference between the high and low-frequency ping readings is the deposited materials (Odhiambo and Boss 2004; Ibrahim and Sternberg 2021). Low-frequency single-beam echo sounder has the tendency to penetrate the sea bottom surface up to 15 m (Chapra 1997; Xavier 2010). In addition, a multibeam echosounder can also be used; however, this will require different epoch datasets of at least five-year intervals and the difference in depth between the datasets is assumed to be the measured suspended sediment (Chapra 1997; Vahid et al. 2018). The information gathered from the bathymetric survey will aid the water body managers in determining the amount of material to be dredged and the section in which the dredging activities should be more concentrated.

However, several researchers have conducted investigations to evaluate soil erosion and sediment deposits using different approaches and models which are classified as empirical, physical and hybrid models (Mohammad et al. 2018; Dutta 2016). These models are suitable for modelling soil loss via water erosion and sediment transport in the river channel but cannot accurately account for deposited suspended material due to differences in field and laboratory observations. Psilovikos and Margoni (2010) gave an in-depth description of these models. In addition, Morris and Fan 1998, Martin 2015, and Mohammad et al. 2018 employed some of these models to estimate sediment yield in rivers and reservoirs but could not forecast the surface based on sediment deposit characteristics. Consequently, this research focuses on using Markov Chains and Cellular Automata (MC – CA) models to evaluate and forecast the future bathymetric surface of Kainji reservoir in Nigeria, West Africa, based on prevalent environmental characteristics at the period of this investigation. The Markov process is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event (Gagniuc 2017).

Furthermore, Markov chains are extraordinarily useful to model a discrete-time, discrete space or continuous-time stochastic process/random process of various domains, such as environmental sciences (land use and land cover classification) and meteorology "weather forecasting" (Nurmiaty et al. 2014). A discrete-time markov chain (DTMC) is a countably infinite series in which the chain state changes at discrete time steps, while a continuous-time Markov chain is called a continuous-time Markov chain (CTMC) (Pankin 2017; Serfozo 2009). In addition, Cellular Automata (CA) is a sophisticated mathematical technique employed to manipulate data in rows and columns often represented in extended binary format, where neighbouring pixel information is paramount (Chih-Hung 2014). In addition, researchers such as Nurmiaty et al. 2014; Kumar et al. 2016; Palmate et al. 2017; Tadese et al. 2021; Zouiten et al. 2021 etc., have conducted several investigation employing MC and CA in modelling spatiotemporal land dynamics and forecasting using satellite imageries and their outcome have been impressive. Hence, an effective implementation of MC–CA to assess and forecast future bathymetric surfaces from discrete-time based bathymetric datasets will be an added knowledge to the research community. Additionally, these models were selected as an integration for this research because of their robust nature in manipulating geospatial data, especially the masking attributes of cellular automata.

Study area

Kainji Reservoir was constructed as the largest dam perpendicular to the Niger River in the year 1968. Kainji has a main concrete dam with rockfill embankments and a saddle dam. The saddle dam protects the main dam during flooding. There are four spillways with hydraulically operated gates of 15.2×15.2 m, which are used to control flooding and water supply to the Jebba dam downstream. The dam has a maximum depth of 85.5 m and 4.971 miles in length before joining the lake upstream. The lake and the dam are wholly joined together; the total length is approximately 59.03 miles, as shown in Fig. 1. Figure 1 is the map of the study area describing Kainji Dam. The Kainji dam is located in Niger State and it shares boundaries with Kwara and Kebbi states. The geographical extent of the entire reservoir is $10^0 36^{'} 40.16^{''}$N, $4^0 31^{'} 08.11^{''} E$; $9^0 51^{'} 17.55^{''}$N, $4^0 36^{'} 03.48^{''} E$; $10^0 21^{'} 24.46^{''}$ N, $4^0 40^{'} 46.59^{''} E$; and $10^0 22^{'} 53.66^{''}$N, $4^0 17^{'} 13.64^{''} E$. The major activities of the adjoining communities within the reservoir corridor's are farming and fishing.

Additionally, the image in the lower right part of Fig. 1 shows the location of the hydropower plant, major and minor access roads, and other vital structures, such as offices and security checkpoints. The area has experienced a high influx of dwellers over the years due to the strategic location of the infrastructure. The power station area of the reservoir is considered the administrative section.

Materials and methods

The datasets used for this research are from primary and secondary sources. The spatial datasets reference system is based on the World Geodetic Reference System 1984 (WGS84) with Clark 1880 ellipsoidal parameters to maintain global datum uniformity of the datasets used. The bathymetric datasets are from two different epochs "1990 and 2020", as presented in Table 1. The 1990 dataset is archived data obtained from the reservoir managers. While the data for the

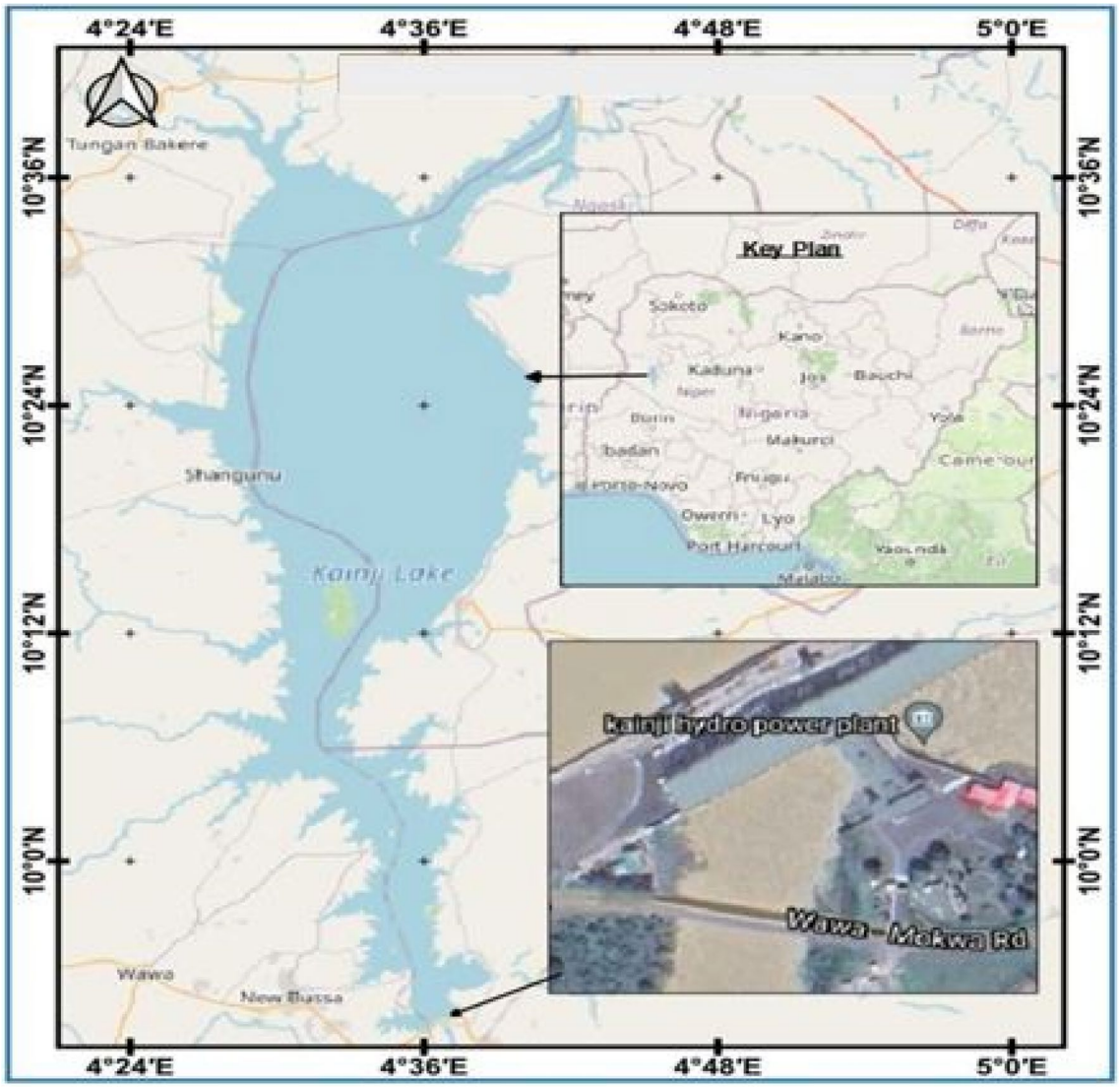

Fig. 1 Map of the Study Area- Kainji Dam. Source: The map was produced using QGIS software and Google Earth Engine

Table 1 Datasets for the study

Data	Data format	Period	Instrument type	Source
Bathymetric	N, E, D (coordinates)/CSV	1990	SBES_200 kHz	Archive/Dam Management
Bathymetric	N, E, D (coordinates)/CSV	2020	SBES_200 and 20 kHz respectively	Field observation
SRTM-OpenDEM	Shape (contour lines) (25 m interval)	2020		https://www.opendem.info/opendem_client.html

year 2020 was observed using a single beam echosounder in low and high frequencies (20 kHz and 200 kHz). The essence of frequency variation is to measure the water depth and the amount of deposited suspended sediment. The high frequency measures the water depth from the water surface to the bottom surface, while the low frequency measures the bottom surface to a certain level below the bottom surface deep.

Similarly, the main limitation of a higher frequency transducer is that there is high attenuation of the signal with depth, and lacustrine sediments with low specific gravity or lower vegetation will easily reflect the

signal. Similarly, high-frequency transducers are not recommended in areas where layers of lacustrine sediment are common or where the vegetation may mask the required grade of payment. Hence, frequencies ranging from 20 kHz to 50 kHz are commonly used for sediment deposit measurement payment (USACE 2002). The low frequency is used to generate maps showing the accumulated sediment over time after impounding the reservoir (Chapra 1997). Consequently, the change in depth between the high and low frequencies yields the estimated lacustrine sediment thickness of the area being surveyed. The reservoir shoreline was delineated using interactive vectorization. Shoreline delineation using satellite imagery and GIS has received recognition over the years due to the difficult terrain that is associated with waterbodies (Barman et al. 2014). Thus, the spatial attributes were extracted and combined with the bathymetric data to coordinate the data for the effective implementation of MC-CA models. Fig. 2 describes the shoreline in cyan (light blue colour). The reservoir water line was assumed to be the zero level during data modeling analysis. Table 2 is a sample of the spatial sounding data used for the MC-CA data modeling.

Data cross-validation

Descriptive statistics are primarily employed to quantitatively assess interpolation errors such as the minimum, maximum, mean, root mean square error (RMSE), mean error (ME), R – square, and standard deviation (Burrough and McDonnell 1998; Hu et al. 2004). The mathematical expressions for the descriptive statistics used and the result obtained are given in Table 3. Furthermore, the split–sample technique was used to evaluate the dependency of the generated digital depth model surface of the Kainji reservoir from the two epoch datasets. In each case, the data was partitioned into a series of 10,000, 30,000, 50,000, and 70,000; the mean varied from 0.018 to 0.016, and the root mean square (RMS) varied from 0.429 to 0.434 for the 1990 dataset. The 2020 dataset presented a mean of 0.0013 to 0.069, and the RMS was within the neighbourhood of 0.477 to 0.516. Similarly, the final cross-validation was conducted, and the result is presented in Table 3.

From Table 3, as the mean error tends to zero in both datasets, the mean square errors (MSE) also drive toward the same direction, supporting the central limit theorem, which says that as the sample size n increases, the variance of the quantity $\frac{1}{n}\sum_{i=1}^{n}(\widehat{Z}_i - Z_i)^2 = \sum_i {(\epsilon_i)^2}/{n}$ should converge to zero. At the same time, the standard way to measure the error of the model of the predicting quantitative

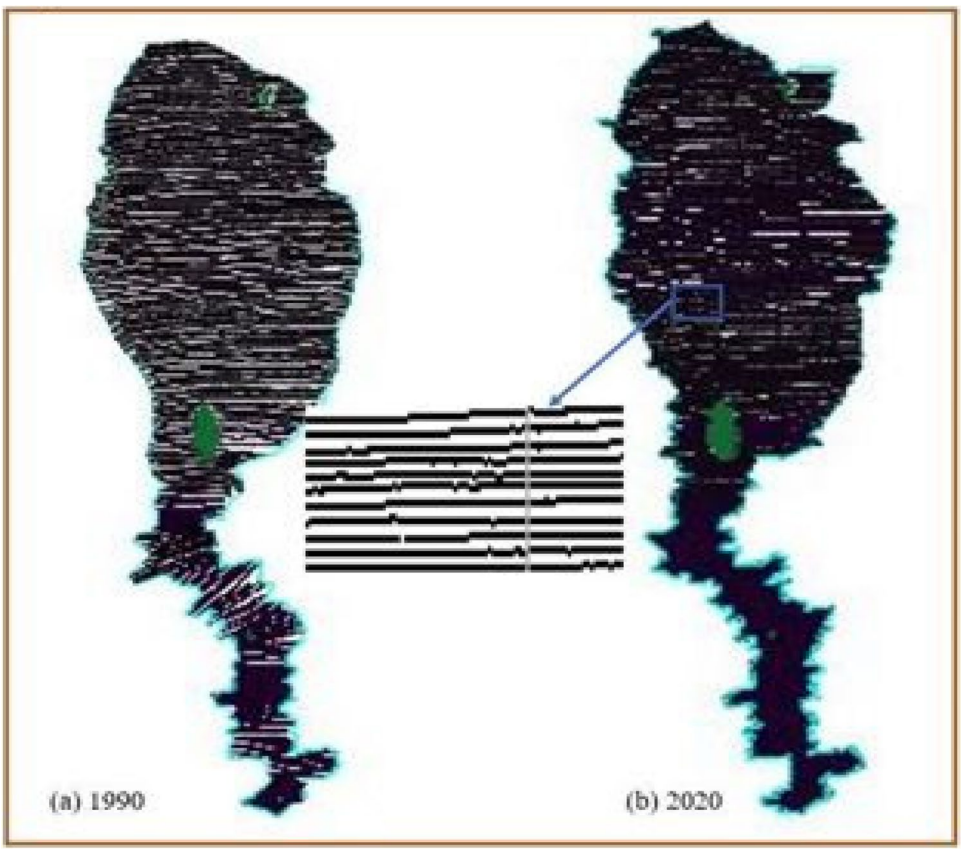

Fig. 2 Point cloud bathymetry data. Source: **a** The 1990 bathymetric data, provided by Mean Stream Dam: Managers of Kainji Dam, Nigeria. **b** 2020 data from field observation. Both data were acquired using SBES at 200 kHz

Table 2 Sample of spatial sounding data

Easting	Northing	Depth	Easting	Northing	Depth
674240.275	1089791.335	8.68	675900.078	1093430.167	43.12
674239.502	1089791.573	8.25	675898.469	1093432.134	39.35
674238.830	1089791.699	7.73	675896.629	1093434.239	39.92
674238.029	1089791.983	7.61	675895.455	1093435.596	39.56
674236.877	1089792.394	7.45	675894.836	1093436.363	41.68
674235.551	1089792.926	7.31	675890.600	1093441.645	42.77
674233.270	1089793.174	7.16	675889.767	1093442.707	41.44
674223.724	1089793.713	6.76	675888.824	1093443.944	41.87
674213.981	1089797.088	6.57	675888.000	1093445.071	39.87
674208.236	1089802.483	6.62	675884.524	1093449.189	39.35
674207.031	1089810.348	6.69	675884.006	1093449.798	42.17
674208.376	1089817.529	6.76	675881.470	1093453.263	39.68

The data presented is a sample of the observed information in Northing, Easting and depth. All measurements are in metres (m)

Table 3 Cross-validation/split–sample technique outcome

Final split – Sample technique		
Statistical model	1990 data	2020 data
$\mathbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Z}_i - Z_i)^2$	0.003	0.00009
$\mathbf{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{Z}_i - Z_i)^2}{n}}$	0.012	0.029
$\mathbf{ME} = \frac{\sum_{i=1}^{n}(Z_i^{PRED} - Z_i^{OBS})}{n}$	0.018	0.069
$R^2 = 1 - \frac{RSS}{TSS}$	0.702	0.804

data (RMSE) suggested that the model error is 0.012 and 0.029 in both cases. The cross-validation outcome indicated that the generated DDMs are replicates of the reservoir bed based on the data trend.

Implementation of Markov Chains – Cellular Automata

Investigators have used Markov Chains and Cellular Automata mathematical techniques in various research on modeling and analyzing spatial data, such as (Samat 2009; Lingling et al. 2011; Jamal et al. 2011; Aqil and Shu 2020). These models have proven to be effective in transients in various data that can be divided into categories, for instance, the type of land use and land cover classification changes in the ratio of variation in development within a region. Figure 3 is the methodology diagram, which illustrates the procedures used to forecast future bathymetry surfaces for the year 2050 due to sediment deposit patterns. The entire process was executed using a Python programming language.

Initial dataset and preprocessing

The initial data for this study are the depths of the studied area of the Kainji reservoir, measuring approximately 24 × 83 km, measured at 2 points in time: 1990 and 2020 (200 kHz). Unfortunately, the coordinates of the measured depths do not correspond to each other; therefore, we have to use interpolation to compare and match the data. The inner rectangle was selected for the interpolation, which is enclosed in all two presented datasets. Then, the rectangle was divided into grids with a step of 100 m. Such a grid step is recognized as probably optimal for spatial data prediction (Samat 2009). After that, the depth values were calculated at each point using linear interpolation of the nearest measured points. After interpolation, contour plots of the study area for each historical moment in time were built (Fig. 7). The time difference between the two datasets is 30 years.

Cellular automata application

Cellular automata represent a discrete model, which is a grid of arbitrary dimensions, each cell of which can assume one of a finite set of states at each moment, and a rule for the transition of cells from one state to another is determined. The rules specify the transition from one state to another, which may include considering the states of the neighbourhood cells. The CA is mathematically expressed as (Hou et al. 2004):

$$S(t, t+1) = f(S(t), N) \quad (1)$$

where S is the set of limited and discrete cellular states, N is the cellular field, t and t + 1 indicate the different times, and f is the transformation rule of the cellular states in local space.

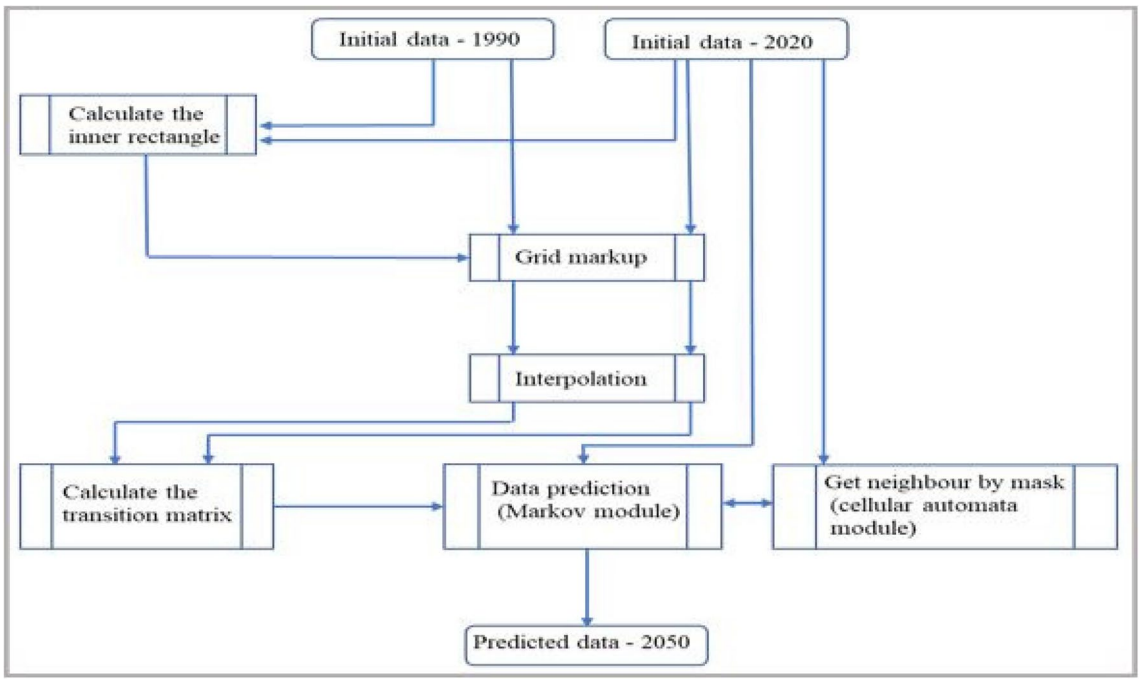

Fig. 3 Flow diagram of the CA-Markov chain model for predicting future depth changes due to sediment deposits in the Kainji reservoir

Thus, according to the following template, this research defines a significant neighbourhood (Fig. 4).

Furthermore, the behaviour of Cellular Automata models is affected by uncertainties arising from interactions between model elements, structures, and the quality of the data sources used as input to the model (Batty et al. 1999; Peterson et al. 2009). Hence, the Cellular Automata takes into account the data trend for effective output.

Markov chains procedures

Markov chains imply that we have a particular graph of states and transition probabilities from one state to another. The Markov process explains the rate of changes in states between the loss in depths due to sedimentation over time and uncovers the degree of depth losses in the depth classes. Hence, to predict the surface's future state, the formula based on the conditional Bayes probability formula is also used (Hou et al. 2004; Yang et al. 2007; Jiang et al. 2009).

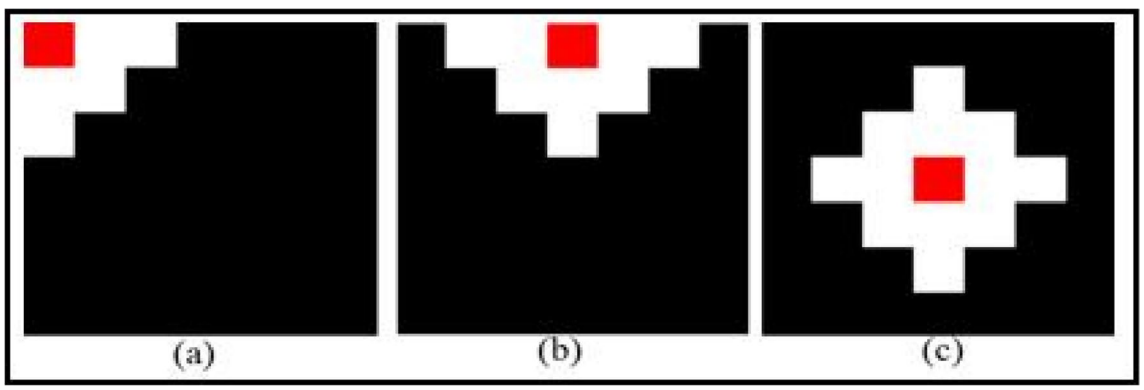

Fig. 4 Cellular Automata significant neighbourhood template. The cells to be defined are highlighted in red, significant neighbours are highlighted in white, and insignificant neighbours are black. In this case, the state of the determined cell at the previous time also affects the cell's future state. For instances where the cell is close to the edge of the study area (a, b), some of the neighbours are not considered in the calculations

$$S(t+1) = P\_ij \times S(t) \tag{2}$$

where S are the system states at time t or $t+1$ and P is the transition probability matrix in a state, which is calculated as follows (Tadese et al. 2021):

$$P_{ij} = \begin{bmatrix} p_{11} & p_{11} & \cdots & p_{1n} \\ p_{21} & p_{22} & & p_{2n} \\ \cdots & \cdots & \cdots & \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \tag{3}$$

$$0 \le Pij < 1 \; and \; \sum_{j=i}^{N} Pij = 1, (i,j = 1,2,\ldots,n)$$

Furthermore, in this research, the states are the depth intervals of the studied area 'Kainji reservoir'. Therefore, the entire depth of the study area was divided into depth ranges of ten states from zero to nine (Table 4).

Consequently, the initial datasets were transformed into datasets of Markov chain states. Fig. 8 describes the historical data's contour plots as often transformed into states. In addition, based on the historical data of transitions, the transition probability matrix for the chain was calculated. In this case, first, each cell was considered a separate experiment. Then, the research considers its state at the next point of time as a result of the experiment. Table 5 presents the results generated from the calculated transition probability matrix.

The CA-Markov technique integrates the theories of Markov and CA, which are about the time series and space for an advantage of forecasting. Thus, to predict the future state of each cell, the researcher used the transition matrix for the current state of the cell and all significant neighbours determined by the mask as described above. The majority principle determined the final state of the desired cell. The resulting maps for the estimated state of the study area in 2050 compared to historical data are shown in Fig. 9.

Table 4 Depth classified according to Markov states

Depth span, meters	Markov state
0.00–8.50	0
8.50–17.00	1
17.00–25.20	2
25.50–34.00	3
34.00–42.50	4
42.50–51.00	5
51.00–59.50	6
59.50–68.00	7
68.00–76.50	8
76.50 +	9

Evaluation of forecasted bathymetric surface

Basically, two methods of accuracy evaluation were employed to test the dependency and reliability of the forecasted bathymetric surface. The methods are the matching spatial coordinates and quantitative approaches of precision testing. The former involves interactive performance check by digitally matching the spatial coordinates of the 2020 and 2050 bathymetric surfaces to examine the differences between the two datasets; the results is as shown in Table 6. Similarly, some selected quantitative precision testing employed are the mean square error (MSE), the root mean square error (RMSE), the coefficient of multiple determination (R^2), and the coefficient in depth axis. RMSE is used to determine the rate of error size; in contrast, it is sensitive to outliers due to the magnitude of weight it allocates to large errors (Hernandez-Stefanoni et al. 2006). Mean error (ME) determine the level of bias in the output (Isaaks and Srivastava 1989); however, negative and positive predicted value should be considered (Nalder and Wein 1998). Meanwhile, the mathematical expression is given in Eqs. 4 and 5 (Ahmed and Marsily 1987; Vicente-Serrano et al. 2003):

Table 5 Transition probability matrix of the Markov chain states in percentage

	0	1	2	3	4	5	6	7	8	9
0	*97.543075*	2.360739	0.096186	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	13.910420	*65.007047*	20.010129	1.070202	0.002202	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.195477	16.368107	*76.101465*	7.273866	0.047342	0.013744	0.000000	0.000000	0.000000	0.000000
3	0.074890	0.773861	*76.920090*	21.695837	0.418828	0.036058	0.049926	0.030511	0.000000	0.000000
4	0.507068	0.476337	35.172096	*52.596804*	9.864782	0.645360	0.199754	0.460971	0.076829	0.000000
5	3.176471	1.529412	5.529412	23.058824	*41.411765*	15.411765	4.470588	2.352941	3.058824	0.000000
6	0.374532	0.561798	4.494382	9.176030	15.355805	*47.191011*	14.981273	5.992509	1.872659	0.000000
7	0.000000	0.000000	2.736318	4.975124	7.462687	16.915423	*51.741294*	15.174129	0.995025	0.000000
8	0.000000	0.000000	4.838710	9.139785	12.365591	11.290323	*33.870968*	24.731183	3.763441	0.000000
9	0.000000	0.000000	6.896552	*37.931034*	24.137931	6.896552	6.896552	5.172414	12.068966	0.000000

Source: Research lab

Table 6 Quality and accuracy evaluation: Spatial coordinates matching technique

Matched coordinates (E, *N*)	1990	2020	CA-Markov 2050	Change
676182.42, 1093090.31	75.42	70.27	68.20	0.07
675839.24, 1091575.45	68.08	62.36	59.41	0.09
674489.16, 1106413.62	44.92	41.78	39.01	0.21
664890.56, 1150213.41	23.41	20.82	18.92	0.51
670145.08, 1162347.31	29.42	24.62	21.29	0.65
670056.56, 1113540.71	21.01	18.83	15.97	0.03
669957.34, 1159313.12	25.21	23.73	20.51	0.09

The measurement unit in meter (m)

Table 7 Quality and accuracy evaluation: quantitative precision testing techniques

Some quantitative model used	Data considered: 2020 HF and 2050 MC_CA
ME	0.046
MSE	0.214
RMSE	0.463
$\boldsymbol{R}^2$	0.012
Coef. of depth variations	0.629

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Z}_i - Z_i\right)^2 \tag{4}$$

While the root mean square error is given as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{Z}_i - Z_i\right)^2}{n}} \tag{5}$$

where Z_i = the observed value, $\hat{Z}_i$ = the predicted value and n = the total number of points considered in Eqs. 4 and 5, respectively. Additionally, the mean error is computed as:

$$ME = \frac{\sum_{i=1}^{n}\left(Z_i^{PRED} - Z_i^{OBS}\right)}{n} \tag{6}$$

where Z_i^{PRED} and Z_i^{OBS} are the predicted and observed depths, and n is the total number of observations. The results from the quantitative methods of precision testing is as presented in Table 7.

Results and analysis

The negative impact of sedimentation on Kainji reservoir is evident based on the measured data. The echo sounding operation of 2020 shows that the maximum depth of the reservoir is 71.2 m, while the 1990 bathymetric data present a maximum depth of 77.3 m. This accounts for a 7.49% (5.8 m) loss in depth over the past 30 years. The construction highest depth of the reservoir as of 1968 is 85 m. This implies that the reservoir has lost an estimated 13.4 m within the period of 52 years due to sedimentation and siltation. This geomorphological region is located at the dam axis area. This region is termed the hydrological powerhead of the reservoir. The total area of 1990 data is 1007.780 km^2, as recorded, while as of 2020, the estimated area is 1009.110 km^2. Fig. 5 describes the comparison of the computed reservoir storage capacity curve from HF_1990, HF_2020 and LF_2020 data.

Similarly, the output of the computed volume from the 200 kHz data of 2020 shows that the volume of the reservoir at the time of the survey was $10.5 \times 10^9 m^3$. The 1990 200 kHz data presented an estimated reservoir volume of $12.9 \times 10^9 m^3$. This implies that the reservoir has experienced an 18.6% loss in storage capacity within the period of 30 years. However, the estimated volume for 2020 does not include some of the reservoir tributaries, incredibly inaccessible areas, due to the terrain's hazardous nature. Fig. 6 shows the 100% stacked column chart showing the percentage loss in volume over time relative to the observed depths. According to survey data from 2020, a larger percentage has been lost to sedimentation, and this loss is more noticeable at depth ranges of 76.1 m to 64.1 m. Hence, Depths of 76.1 m to 72 m are not in existance, indicating that if sediment inflow patterns continue unchecked, depth values from 71 m to 64 m will soon go extinct in a few years. This loss of depth poses a significant threat to marine life and ecosystems in the area, as it disrupts habitats and can lead to a decrease in biodiversity. Efforts must be made to address sedimentation issues in order to preserve the remaining depth ranges for future generations.

Markov chains and Cellular Automata output

The computation outcome of transition probabilities using Markov chain analysis is presented in Table 5 below; the rows reflect from which state the transition occurs and the columns to which state. For example, the probability of going from state 1 to state 0 is 13.9%. It is evident that the Table is not symmetrical about the main diagonal. For example, the probability of going from state 0 to state 1 is much less than the probability of going back, which is 2.4%. The research printed the zero transition probabilities in light gray for adequate presentation. The maximum values bylines are underlined. Analysis of Table 5 shows that states with small ordinal numbers (shallow depths) tend to preserve the status quo. Their most significant value lies on the main diagonal. It is also evident that the values under the main diagonal are more significant than those above the main diagonal, which means there is a tendency for a decrease

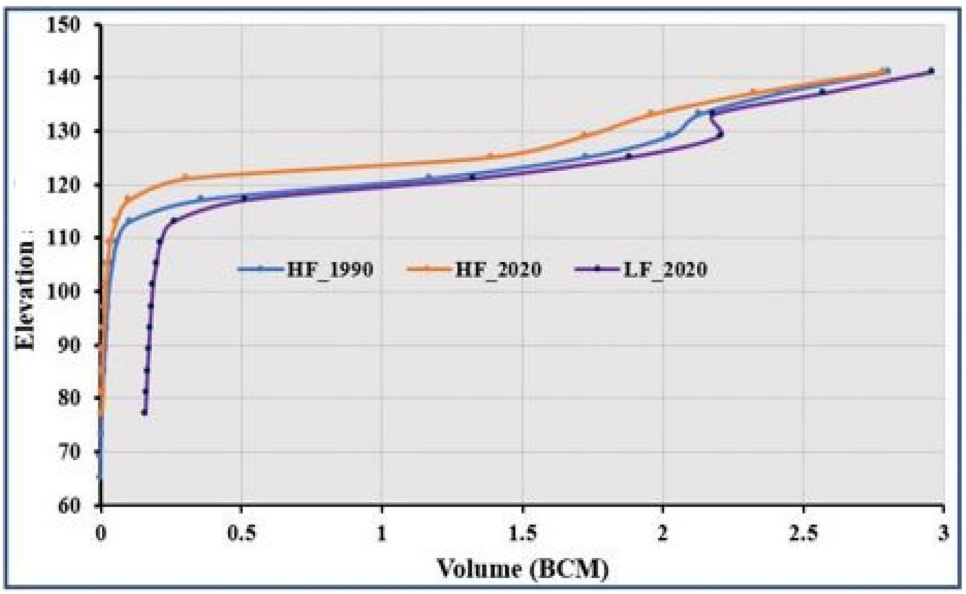

Fig. 5 Elevation capacity curve of the Kainji reservoir. The volume presented is computed from the LF and HF of 2020 acquired bathymetric data and the received HF_1990 bathymetric data. The space between LF_2020 and HF_2020 is the region consumed by lacustrine sediment from 1968 to 2020 when the measurement was carried out. Additionally, amsl refers to elevation above mean sea level while BCM is abbreviation for billion cubic metres

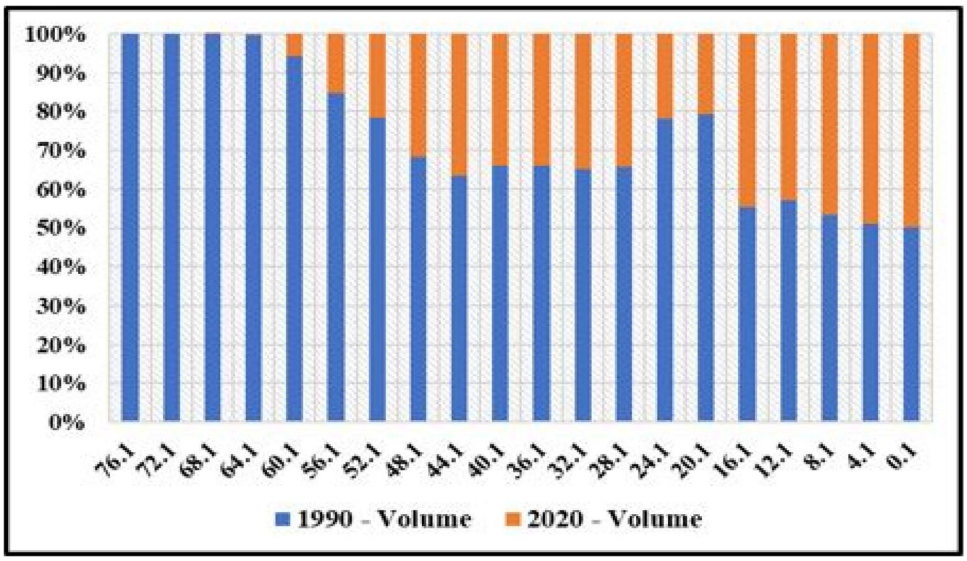

Fig. 6 100% stacked column chart of 1990 and 2020 computed volume. The graph shows the relative loss in volume to depth between the two epochs of data. Mainly, it compares the percentage that each depth range value contributes from each dataset to the computed volume of that section

in the state number; this refers, in general, to a reduction in the depth of the investigated area. The research made the same conclusions when constructing contour diagrams from historical data, as presented in Fig. 7.

Consequently, Fig. 7 below describes the Markov chain classification of depths into state contour plots prior to applying and computing the probability matrix or confusion matrix. The vertical scale readings show the states for the 1990 and 2020 historical data as classified, and the depth gradually reduces based on historical data as presented in the forecasted MC – CA (Figs. 8 and 9).

Finally, the predicted future depth distribution was modelled based on the integration of CA and Markov chain techniques, as presented in Fig. 9. The map contains artefacts probably caused by the random choice of the following state in the Markov chains. The reason is that the Markov process is a categorization process that is best applicable to spatially categorized data, e.g., imagery depicting different land uses and covers. The artifacts are the black spots on the projected surface of 2050 as shown in Fig. 9. However, the CA-Markov chain adequately predicted the future state depths following historical patterns except for the introduced artefacts on the map (Markov 2050). This setback affected the quality of predicted map, because the map was supposed to be a clear map as that of 1990 and 2020 (Fig. 9). The future maximum depth is between 59 m and 65 m; the depth gradually decreases from upstream, and the weight is experienced at the reservoir axis region. Meanwhile, the

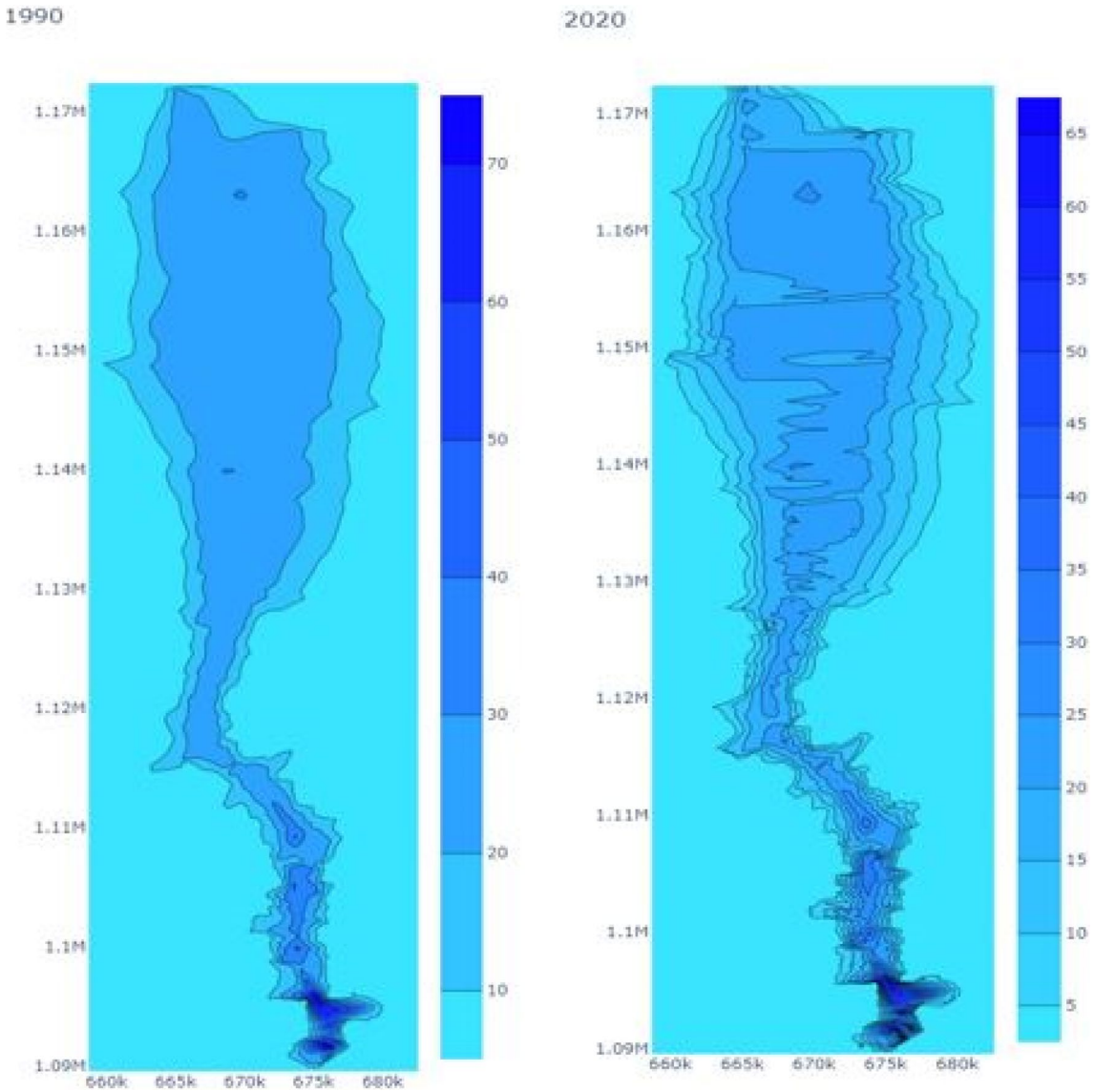

Fig. 7 Generated contours from linear interpolation. According to these plots, it is clear that the limiting depth decreases with time due to sedimentation and siltation from 1990 to 2020. Additionally, the maximum depth at construction (1968) of 85 m was reduced to 77.3 m (1990) and 71.3 m based on 2020 HF data

actual cause of the artefacts is not clearly ascertained; therefore, it is subject to future investigation. Additionally, further investigation should be conducted using bathymetric datasets that are densely captured about a particular waterbody, however, this is a postulation because the data used is sparsely distributed.

Quality and accuracy evaluation of forecasted surfaces

The practical approach to investigating the quality and reliability of projected surfaces is implementing interactive performance checks or matching spatial coordinates (Ibrahim and Sternberg 2021). Thus, spatial coordinates matching was conducted to evaluate the accuracy of the projected bathymetric surface of the Kainji reservoir from the CA – Markov models of 2050. Table 6 shows the quality and accuracy evaluation of forecasted bathymetric surfaces of Kainji reservoir. The difference between 2020 HF and the CA – Markov surfaces is the change on spatial coordinates, neglecting the negative signs.

Similarly, the change between 1990 and 2020 measured data shows a remarkable reduction in depth in the year 2020 due to the menace of sedimentation, and this also translates to a decrease in depth in the projected surfaces of 2050 depending on the characteristics of the existing sediment

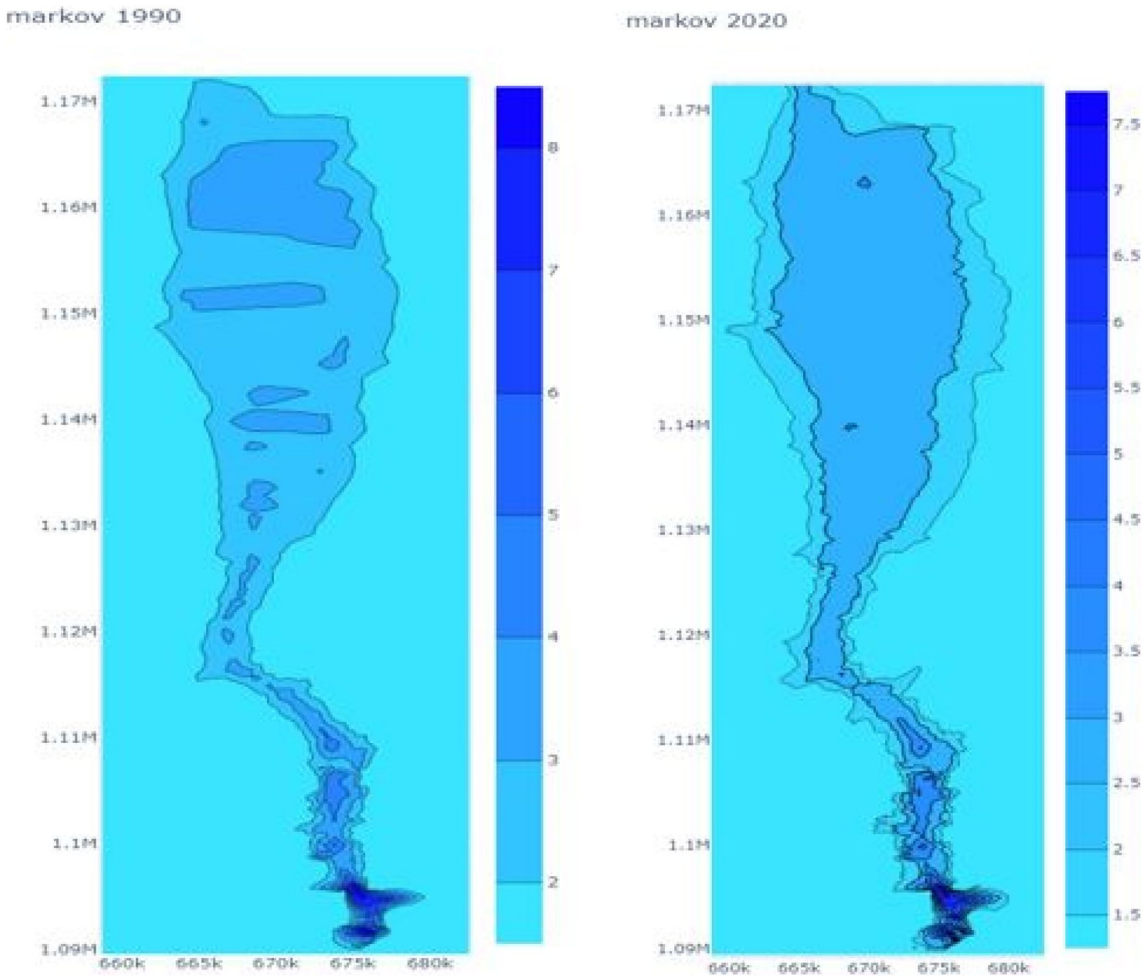

Fig. 8 Contour plots for depth states based on historical data. Horizontal and vertical scale readings are in meters (m). The plots present the Markov chain depth states based on the transformation conducted on the historical datasets of 1990 and the measured datasets of 2020

inflow pattern. In addition, Table 7 is the results obtained from the quantitative precision testing approach used to evaluate the extent of dependency of the predicted bathymetric surface of Kainji reservoir.

The evaluation results revealed that the ME (0.046) is small, which agrees that the bias between the existing and forecasted bathymetric surface is negligible. This is also reflected in the MSE; the MSE has a significant influence on the RMSE as seen in Table 7. Hence, the result of the RMSE reflects the weight of MSE, because it is sensitive to outliers and influences large error; for this reason, the MSE and EMSE suffer the same setback (Merwade et al. 2003; Eriksson and Siska 2000). It is obvious that the model experiment of R – squared or coefficient of determination ($\boldsymbol{R}^2$) of 0.012 indicates that there is no unique correlation between the two datasets considering depth value. R – squared value that tends to one, indicates strong relationship, while when the value tends to zero shows weak or no correlation (Palmate et al. 2022). This is collaborated by the outcome of coefficients of variation in depth assessment, which gave a value of 0.629. The weak correlation between the 2020 and 2050 MC – CA bathymetric surfaces shows the menace of sedimentation and siltation taking its toll on the reservoir due to environmental impact characteristics.

Conclusion

The research carried out sediment deposit analysis of the study area and forecasting by the use of a Cellular Automata – Markov Chain. The outcome revealed that Cellular Automata – Markov chain models effectively predicted the future reservoir depth pattern for 2050. The transition matrix shows similar results and how sediments are distributed based on the data trend. The lower diagonal has a majority of the higher value (Table 5), and the upper diagonal has a lower value (Table 5), suggesting loss in depth due to sedimentation. However, the projected surface map has

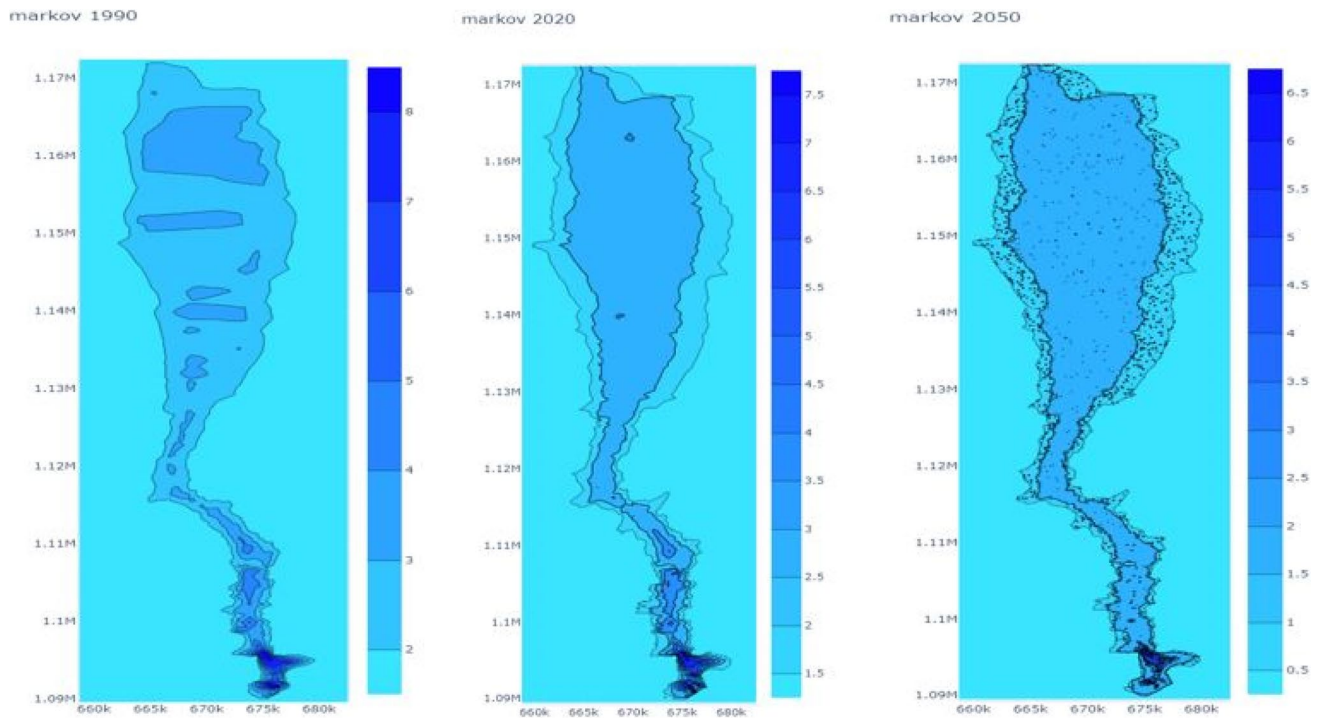

Fig. 9 CA-Markov historical and forecasted plots. The vertical scale is the Markov states in all plots. The predicted plot (Markov 2050) contains artefacts caused by the uncertainty of the modelled Markov states. However, this assertion is subject to future investigation

some artefacts on it. The actual cause of artifacts on the projected surface cannot be ascertained for now. However, the researchers suggest that it is either CA-MC is not a good model for forecasting bathymetric point cloud data or there is an unidentified data distribution error, meanwhile, these assertions are subject to future investigations. In addition, the research projected surface is in agreement with the International Sediment Initiative (2011), which says that by 2050, most reservoirs will lose more than 50% of their designed capacity to sedimentation if left uncheck. This implies that most major dams would not adequately serve their purpose of construction within the projected period and Kainji dam is not an exception.

Acknowledgements The authors sincerely acknowledge the German Academic Exchange (DAAD) and the Nigeria's Petroleum Trust Development Fund (PTDF) for awarding the leading Author scholarship for his PhD program. Additionally, thanks go to all other survey team members who supported during data acquisition.

Author contributions Each author played a critical role in this research. The co-authors were the leading author supervisors during his PhD study and made significant input in this article.

Funding Open Access funding enabled and organized by Projekt DEAL. No funding was received to assist with the preparation of this manuscript.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

This article adhered to all relevant ethical standards.

Ethical approval This study does not involve any studies with human participants or animal experiments.

Informed consent Informed consent was obtained from all the individual participants, or third parties included in the study.

Conflict of interest The authors have no conflict of interest to declare that are relevant to this article.

References

Ahmed S, De Marsily G (1987) Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. Water Resour Res 23:1717–1737

Aqil T, Shu H (2020) CA-Markov chain analysis of seasonal land surface temperature and land use land cover change using optical multi-temporal satellite data of Faisalabad Pakistan. Remote Sens 12(20):3402. https://doi.org/10.3390/rs12203402

Batty M, Yichun X, Zhanli S (1999) Modelling urban dynamics through GIS-based cellular automata. Comput Environ and Urban Systems 23:205–33. http://www.complexcity.info/files/2011/07/batty-ceus-1999.pdf

Barman N, Chatterjee S, Khan A (2014) Trends of shoreline position: an approach to future prediction for balasore shoreline, Odisha, India. Open J Mar Sci 5:13–25. https://doi.org/10.4236/ojms.2015.51002

Burrough PA, McDonnell RA (1998) Principles of geographical information systems. Oxford University Press

Chapra SC (1997) Surface water-quality modelling. McGraw-Hill

Chih-Hung C (2014) Some properties of topological pressure on cellular automata. J Algebra Combinatorics Discrete Struct Appl 1(1):41–51. https://dergipark.org.tr/tr/download/article-file/148543

Dargahi B (2012) Reservoir sedimentation. In: Bengtsson L, Herschy RW, Fairbridge RW (eds) Encyclopedia of lakes and reservoirs. Springer, pp 628–649. https://doi.org/10.1007/978-1-4020-4410-6_215

Dutta S (2016) Soil erosion, sediment yield and sedimentation of reservoir: a review. Model Earth Syst Environ 2:123. https://doi.org/10.1007/s40808-016-0182-y

EL-Hattab AI (2014) Single beam bathymetric data modelling techniques for accurate maintenance dredging. Egypt J Remote Sens Space Sci 17(2):189–195. https://doi.org/10.1016/j.ejrs.2014.05.003

Eriksson M, Siska PP (2000) Understanding Anisotropy computations. Math Geol 32:683–700

Estigoni M, Matos A, Mauad F (2014) Assessment of the accuracy of different standard methods for determining reservoir capacity and sedimentation. J Soils Sediments 14:1224–1234. https://link.springer.com/article/10.1007/s11368-013-0816-x. Accessed 11 Dec 2020

Gagniuc PA (2017) Markov Chains: from theory to implementation and experimentation. Wiley

Girish G, Ashitha MK, Jayakumar KV (2014) Sedimentation assessment in a multipurpose reservoir in Central Kerala, India. Environ Earth Sci 72:4441–4449. https://doi.org/10.1007/s12665-014-3344-0

Hansen JT, Boss SK (2000) Bathymetry and empirical modelling of sedimentation in the prairie creek sub-basin of beaver lake, Northwest Arkansas. Geol Soc Am Abstracts Programs 32(3):A13

Hernandez-Stefanoni JL, Ponce-Hernandez R (2006) Mapping the spatial variability of plant diversity in a tropical forest: comparison of spatial interpolation methods. Environ Monit Assess 117:307–334

Horton RE (1945) Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology. Bull Geol Soc Am 56:275–370. https://pdodds.w3.uvm.edu/research/papers/others/1945/horton1945a.pdf

Hou XY, Chang B, Yu XF (2004) Land use change in Hexi corridor based on CA–Markov methods. Transactions of the CSAE 20(5):286–29

Hu K, Li B, Lu Y, Zhang F (2004) Comparison of various spatial interpolation methods for non-stationary regional soil mercury content. Environ Sci 25(3):132–137

Ibrahim PO, Sternberg H (2021) Bathymetric survey for enhancing the volumetric capacity of Tagwai Dam in Nigeria via leapfrogging Approach. J Geomatics 1(2):246–257. https://doi.org/10.3390/geomatics1020014

International Sediment Initiative (ISI-2011) (2011). https://unesdoc.unesco.org/ark:/48223/pf0000212885. Accessed 14 Sept 2021

Isaaks EH, Srivastava RM (1989) Applied Geostatistics. Oxford University Press, New York, p 561

Jamal J, Arsanjani WK, Ali JM (2011) Tracking dynamic land-use change using spatially explicit Markov Chain based on cellular automata: the case of Tehran. Int J Image Data Fusion 2(4):329–345. https://doi.org/10.1080/19479832.2011.605397

Jiang G, Zhang F, Kong X (2009) Determining conversion direction of rural residential land consolidation in Beijing mountainous areas. Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering 25:214–21

Kumar KS, Kumari KP, Bhaskar PU (2016) Application of Markov chain & cellular automata based model for prediction of Urban transitions. International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp 4007–4012. https://doi.org/10.1109/ICEEOT.2016.7755466

Lingling S, Chao Z, Jianyu Y, Dehai Z, Wenju Y (2011) Simulation of land use spatial pattern of towns and villages based on CA–Markov model. Math Comput Model 54(3–4):938–943. https://doi.org/10.1016/j.mcm.2010.11.019

Martin JT (2015) Modeling sediment movement in reservoirs. Prepared by the USSD committee on hydraulics of Dams, subcommittee on reservoir sedimentation ISBN 978-1-884575-70-9

Merwade VM, Maidment DR, Goff JA (2003) Anisotropic considerations while interpolating river channel bathymetry. J Hydrol 331:731–741

Mohammad H, Assefa MM, Hector RF (2018) Erosion and sediment transport modelling in shallow waters: a review on approaches, models and applications. Int J Environ Res Public Health 15:518. https://doi.org/10.3390/ijerph15030518

Morris GL, Fan J (1998) Reservoir sedimentation handbook. McGraw-Hill, Tata

Nalder IA, Wein RW (1998) Spatial interpolation of climatic normals: test of a new method in the Canadian boreal forest. Agric For Meteorol 92:211–225

Nurmiaty A, Sumbangan B, Samsu A (2014) GIS-Based modelling of land use dynamics using cellular automata and markov chain. J Environ Earth Sci 4(4). https://core.ac.uk/download/pdf/234663312.pdf. Accessed 15 Aug 2021

Odhiambo BK, Boss SK (2004) Integrated echo sounder, GPS and GIS for reservoir sedimentation studies examples from two Arkansas Lakes. J Am Water Resour Assoc Res 40(4):981–999

Palmate SS, Pandey A, Mishra SK (2017) Modelling spatiotemporal land dynamics for a trans-boundary river basin using integrated Cellular Automata and Markov Chain approach. Appl Geogr 82:11–23. https://doi.org/10.1016/j.apgeog.2017.03.001

Palmate SS, Wagner P, Fohrer N, Pandey A (2022) Assessment of uncertainties in modelling land use change with an integrated cellular automata–Markov Chain Model, vol 27. Environmental Modeling & Assessment. https://doi.org/10.1007/s10666-021-09804-3

Pankin MD (2017) Markov chain models: Theoretical background. Archived from the original. https://web.archive.org/web/20071209122054/, https://www.pankin.com/markov/theory.htm. Accessed 26 Mar 2021

Peterson LK, Bergen KM, Brown DG, Vashchuk L, Blam Y (2009) Forested land-cover patterns and trends over changing forest management eras in the Siberian Baikal region. For Ecol Manag 257(3):911–22. https://doi.org/10.1016/j.foreco.2008.10.037

Psilovikos A, Margoni S (2010) An empirical model of sediment deposition processes in Lake Kerkini, Central Macedonia Greece.

Environ Monit Assess 164:573–592. https://doi.org/10.1007/s10661-009-0914-9

Samat N (2009) Integrating GIS and CA-Markov model in evaluating urban spatial growth. Malaysian J Environ Manage 10(1):83–97. https://core.ac.uk/download/pdf/11491356.pdf

Serfozo R (2009) Basics of applied stochastic processes: probability and its applications. https://doi.org/10.1007/978-3-540-89332-5

Tadese S, Soromessa T, Bekele T (2021) Analysis of the current and future prediction of land use/land cover change using remote sensing and the CA-Markov Model in Majang Forest Biosphere Reserves of Gambella, Southwestern Ethiopia. Sci World J 2021. https://doi.org/10.1155/2021/6685045

USACE (2002) https://www.publications.usace.army.mil/USACE-Publications/EngineerManuals/u43544q/487964726F67726170686963/. Accessed 14 Jan 2021

Vahid R, Jude HK, Frank deN, Mark EJ, Edward AM, Donald HH, Christian G, Paul ML, Scott WC, Ryan AC, Adam JB (2018) Examining storage capacity loss and sedimentation rate of large reservoirs in the central U.S. Great Plains. Water 10:190. https://doi.org/10.3390/w10020190

Veli I, Ibrahim MO, Reha MA, Serdar E, Murat U, Yunus K, Serdar DZS (2019) Determination of reservoir sedimentation with Bathymetric Survey: a case study of Obruk Dam Lake. Fresenius Environ Bull 28(3):2305–2313

Vicente-Serrano SM, Saz-Sánchez MA, Cuadrat JM (2003) Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. Clim Res 24:161–180

Xavier L (2010) An introduction to underwater acoustics: principles and applications second edition. Springer

Yang G, Liu Y, Wu Z (2007) Analysis and simulation of land-use temporal and spatial pattern based on CA-Markov model. Geomatics and Information Science of Wuhan University 32:414–18

Zouiten M, Jamal C, Ibtissam N (2021) Predicting Land Use changes within the Tazekka Park and its Borders via a Cellular Automata-Markov modelling of satellite images. J Southwest Jiaotong Univ 56:2. https://doi.org/10.35741/issn.0258-2724.56.2.43