JSML
Journal of Science and Mathematics Letters
Universiti Pendidikan Sultan Idris
Sultan Idris Education University
ISSN 2462-2052 • eISSN 2600-8718

Research Article

Hybrid k-GA Model: A Robust Framework for Chaos-Based River Water Level Predictions Across Different Elevations of the Pahang River

Adib Mashuri[1,4], Nur Hamiza Adenan[1*], Azurah A Samah[2], Ahmad Basri Ruslan[3], Nor Suriya Abd Karim[1] Muhammad Faizan[5]

[1] Department of Mathematics, Faculty Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, Malaysia
[2] Intelligent Informatics Research Group, Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
[3] District Planning and Management Sector, Kedah State Education Department, 05100 Alor Setar, Kedah, Malaysia
[4] Department of General Studies, Kuala Selangor Vocational College, 35600 Bestari Jaya, Selangor, Malaysia
[5] School of Chemical, Biological, and Battery Engineering, Gachon university, Seongnam 13120, Republic of Korea
* Corresponding author: hamieza@fsmt.upsi.edu.my

ARTICLE HISTORY

Received
5 January 2026
Revised
20 April 2026
Accepted
15 May 2026
Published
12 June 2026

KEYWORDS

Hybrid k-GA model
Chaos approach
Genetic Algorithm
Time series forecasting
Pahang River

ABSTRACT

Improving the chaos-based method is important for increasing the accuracy of predictions in uncertain river water level conditions. This paper introduces the k-GA model, a hybrid method that combines chaos theory with genetic algorithms to forecast river water levels, particularly at different elevations of the Pahang River. Traditional chaos-based techniques often struggle with the uncertain and unpredictable nature of river systems, especially when water levels are high. The proposed k-GA model applies chaos theory to identify meaningful patterns in hydrological time series data. At the same time, the genetic algorithm component improves the forecasting process by selecting the most suitable parameters and features. To evaluate its effectiveness, the model was compared with existing chaos-based methods, namely the Local Mean Average Method (LMAM) and the Local Linear Approximation Method (LLAM), across three river zones which are upstream, midstream and downstream area. The results show that the k-GA model outperformed both comparison methods by effectively capturing the complex behavior of water level changes and producing more accurate forecasts, achieving over 99% accuracy compared to the 93% to 96% accuracy of existing chaos-based methods. The main contribution of this study lies in the development of a hybrid chaos-genetic algorithm model that enhances forecasting performance in complex hydrological environments, providing a more accurate and adaptive tool for flood risk management and water resource planning.

https://doi.org/10.37134/jsml.vol14.2.15.2026

1. INTRODUCTION

Precise prediction of river water level is the most important part for both flood prevention and water resource management (Lahsen & Ribot, 2022; Zakaria et al., 2021). Increasing concerns around the globe to climate uncertainties and extreme weather conditions have brought the necessity of robust prediction models to minimize the potential socio-economic consequences climax (Vagheei et al., 2023). Accurate forecasting can help not only in preventing disasters but also in developing a long-term and efficient strategy for using water resources wisely (Kamarudin et al., 2023a).

Chaos theory has increasingly been used as an effective approach to analyze complex, nonlinear and dynamic systems. For example, chaos-based methods have been applied to river water level analysis, as they are well-suited for identifying underlying structures and interdependencies in time series data. In addition, genetic algorithms (GA) have been used extensively for optimization because they can search for an optimal solution in a high-dimensional space (Tahir et al., 2022). Both methods have their own advantages, but they also encounter difficulties in the application of river water level prediction. The chaos-based methods may be sensitive to noise and parameter optimization may not be robust, whereas traditional GA as an independent model may not be able to learn the underlying dynamics of the chaotic system (Alsayed et al., 2020). Chaos theory and GA have been widely used in many forecasting problems, yet they have some weaknesses when applied independently. When using chaos-based methods for prediction, their adaptability may be insufficient to capture complex patterns in hydrological data would potentially leading to poor results (Ramadevi & Bingi, 2022; Swagatika et al., 2024). Moreover, GA are efficient in optimizing solutions, but they usually need to be supplied with distinct input values and their exploration ability in using the chaotic characteristics of chaotic systems may be limited (Mashuri et al., 2024). This creates a gap, as no single method alone can effectively address the challenges of dynamic and complex data, such as river water levels. Therefore, a more unified, integrated and synergistic approach which combines the strengths of chaos and GA is needed.

The main objective of this paper is to present a hybrid approach that integrates chaos-based method with an optimization-based model (Genetic Algorithm), which is called k-GA model. By combining the optimization capability of GA with the ability of chaos theory to detect hidden dynamics, this model aims to improve the accuracy and reliability of river water level forecasting. The hybrid methodology is developed with the aim to improve over the disadvantages of the conventional methods and to provide a more stable method for complex and nonlinear datasets. This paper makes several important contributions which are (a) We developed an innovative hybrid k-GA model using chaos theory for feature extraction and GA for optimization; (b) We introduced a methodology that overcomes the limitations of existing chaos-based and models based on GA only; (c) We demonstrated the proposed model’s accuracy through carefully designed experiments on Pahang River water level; (d) We compared our model with the traditional models, which justified the model’s utilization in hydrological forecasting applications. By addressing this research gap, the study advances the theoretical development of hybrid forecasting models and offers a practical framework to improve disaster response and resource management.

In Malaysia, unpredictable weather patterns and increasingly frequent extreme rainfall events have made river level forecasting more important than ever (Zakaria et al., 2021). The Pahang River, being one of the longest and most flood-prone rivers in the country, affects thousands of residents, especially during the monsoon season (Kamarudin et al., 2023). Timely and accurate predictions can help reduce the impact of floods on communities, agriculture, and infrastructure. This research was driven by a real need to improve early warning systems and support better decision-making in water resource management, particularly in areas where traditional forecasting methods often fall short.

2. METHODOLOGY

Chaos theory has received attention in hydrology due to the potential for it to analyze and predict non-linear, complex systems (Yousefi et al., 2020). These approaches have been useful to decipher the complex dynamics of hydrological systems in such a way that the traditional statistical models are less able to capture (Jiang et al., 2020; Mihailović et al., 2023). Nevertheless, chaos-based methods are readily susceptible to noise and quite sensitive to the choice of parameters, which restricts their use in practice for real world applications.

Genetic algorithms have been applied in numerous forecasting applications because of their ability to solve optimization problems (Pretorius & Pillay, 2024). In hydrology, GA have been used for calibration of model parameters, feature selection and improving the prediction ability of water resource models (Ghorbani et al., 2018; Jiang et al., 2020). GA has the capacity in searching a large solution space and generalizing to different problem domains makes it also a versatile forecasting tool. However, despite its capabilities, the traditional GA can lose much of its predictive ability when relying on an explicit input set and may struggle to capture the complex cascades present in chaotic processes, which can result in unstable or unreliable predictions.

Hybrid models combining chaos theory and other computational methods such as GA have been introduced recently to overcome the limitations of individual methods (Sridhar et al., 2022). These models aim to leverage the strengths of both methods by using chaos theory to identify important patterns in time series data and GA to optimize prediction models. Although these hybrid methods have illustrated better prediction, most previous studies paid little attention to a systematic way to incorporate chaos theory and GA. Furthermore, there are also certain limitations, such us computations and lack of wider testing on different datasets preventing them to be widely applied. This work aims towards filling this gap by proposing a strong hybrid k-GA model. The developed method is based on an approach which combines the analytical capabilities of chaos theory, with the efficiency of the GA optimization to provide an accurate and reliable forecast results without the limitations presented in the literature.

In this study, the dataset was divided into training and testing sets following the approach used by Mashuri et al. (2022) in water level time series forecasting. Specifically, 4,635 data points (approximately 92.7%) were used for training, while the remaining 365 data points (approximately 7.3%) were reserved for testing. This selection follows a fixed number of final observations used for validation. Moreover, using a smaller testing portion ensures sufficient data for training the chaos-based model, which is sensitive to data length and requires adequate historical information for phase space reconstruction.

2.1. Data

The data used in this study consist of river water level measurements from the Pahang River, a major river in Malaysia known for its complex hydrological characteristics. The dynamic of Pahang River toward chaotic has been verified in previous studies by Mashuri et al. (2022). Accordingly, this paper does not rediscover the chaos but develop it more on the established result. A robust framework namely Hybrid k-GA Model is proposed here to predict the chaos-based river water level of Pahang River. Based on the previously identified chaotic properties, this paper focuses on developing and testing the performance of the k-GA model as an advanced tool for improving forecasting accuracy. Based on the previously identified chaotic properties, this paper focuses on developing and testing the performance of the k-GA model as an advanced tool for improving forecasting accuracy.

In this paper, the study area is extended to the upper, middle and lower reaches of the Pahang River. Each area possesses specific features, which affect river water level changes and show different problems for water level forecasting (Hamidon & Musa, 2022). The upper reaches near the source of the river are the fast flowing water and higher gradient section of the river (Kamarudin et al., 2023). These areas are primarily erosional zones as the high energy levels cause substantial movement of sediment in the downstream orientation. On the other hand, the middle reaches location is a transition zone where the sediment transport diminishes and some of the sediment deposits start to occur (Yusoff et al., 2021). These sub-regions are shaped by gentle valley inclinations and feature erosional and depositional dynamics. The downstream area, where the lowland areas are located, lie near to the river valleys that are prone to long term flooding (Mustaffa et al., 2023). These zones are described as sedimentation zones, in which materials eroded from the upstream and midstream areas accumulate (Lyche Solheim et al., 2019). Downstream reaches also serve as floodplains, which frequently flooded by persistent heavy rainfall, extensive precipitation and the volume of runoff (Barrera-Animas et al., 2022 & Lindenschmidt et al., 2018). These factors contribute to the high uncertainty in forecasting river stages in these areas.

The lower area of the Pahang River in Pekan, Pahang historically served as a centre for trade and social interaction. Significant activities include fishing that is provided by a highly diverse river ecosystem and established agricultural production, along the fertile banks of the river and tourism (Along & Deep, 2023). On the other hand, the upstream and midstream area have important implications on river dynamics including erosion process, sediment transport and water flow regulation. The river stations examined in this study across the upstream, midstream and downstream area are listed in Table 1. This information was obtained from the Department of Irrigation and Drainage.

Table 1. The selected station in upstream, midstream and downstream area of the Pahang River

Area	District	Station	No. Station	Dataset
Upstream	Lipis	Kuala Medang	4218416	5 000 hours
Midstream	Maran	Jam. Kg. Awah	3525405	5 000 hours
Downstream	Maran	Lubuk Paku	3527410	5 000 hours

Figure 1 illustrates the hourly time series data of river water levels at the three selected stations, upstream, midstream and downstream areas used in this study. From the plotted time series data, it can be observed that the series exhibits relatively calm patterns over the 5,000-hour period, similar to midstream areas. This is because the recorded data does not show sharp fluctuations, unlike upstream areas. This could be attributed to the gradual release of water from midstream catchments, as the downstream area is densely populated. However, all three stations also recorded irregular tidal water levels, which at times exceed the dangerous level, as shown in Figures 1 (a), (b) and (c). Despite this, there are periods where water levels drop to lower and steady values, along with sudden spikes, indicating that the time series data in this context is highly variable and challenging to predict.

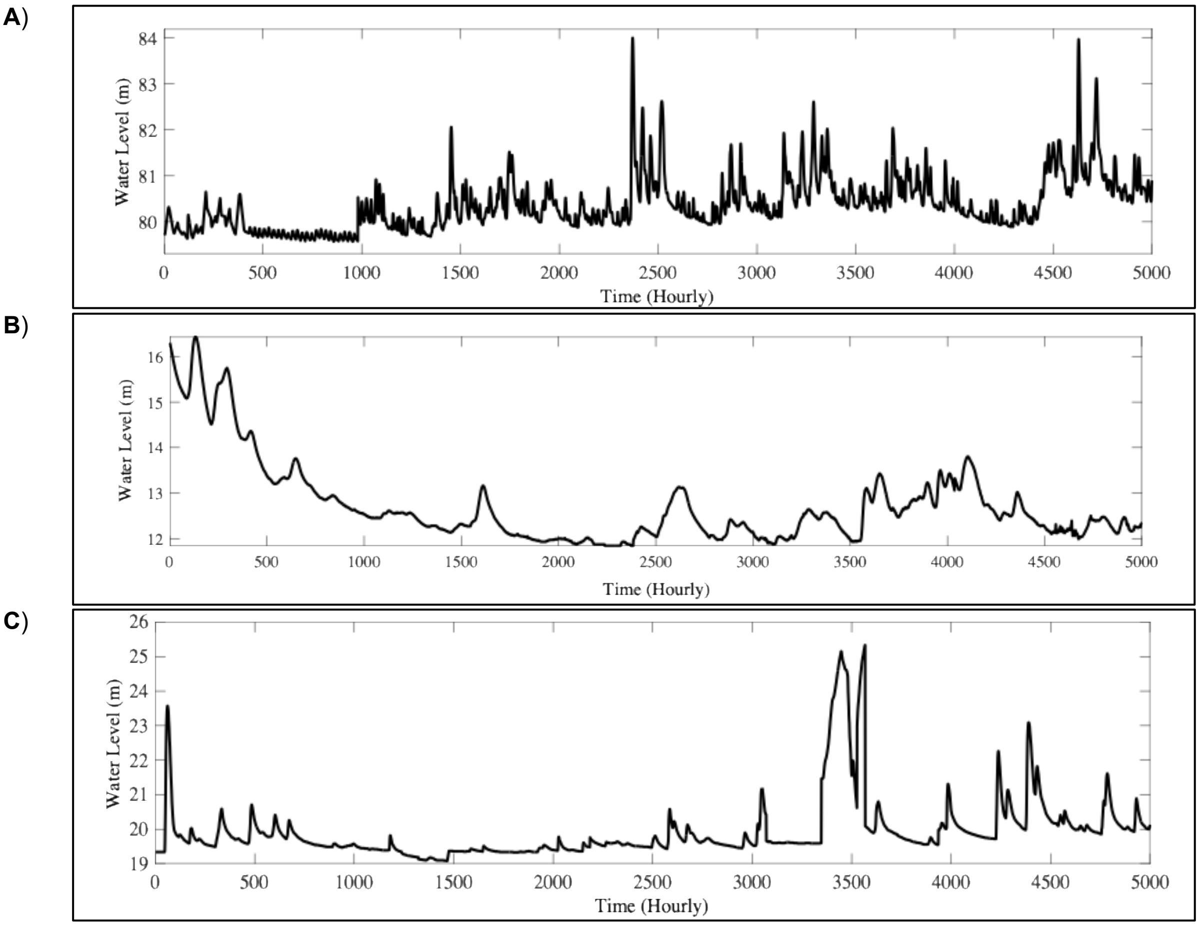

Figure 1. Water level time series data across (**A**) upstream, (**B**) midstream and (**C**) downstream areas

In this study, a total of 5,000 hourly water level data points were used. From this, the final 365 data points (equivalent to approximately 15 days) were selected as the testing set, while the preceding 4,635 data points were used for training the model. This fixed window approach allows the model to learn long-term patterns and behaviors before being evaluated on a continuous prediction horizon that reflects real-world operational use. The choice of 365 hours was inspired by practices in prior studies such as Mashuri et al. (2022), who demonstrated that short-term windows can yield meaningful predictive accuracy. However, in our case, we maintain a consistent long-range dataset while still reserving the final sequence for fair and sequential forecasting evaluation.

2.2. Chaos Approach and Hybridization

In today's context of modernization, with the recent progress in hydrology learning has also revealed the significance of time series models for the design, planning and operation of irrigation and drainage systems (Singh & Dhiman, 2018). Prior to any forecasting, one needs to carry out a chaos analysis since this becomes a precursor for application of chaos-based methods. The modeling of river water level time series with a chaos approach in this research can be divided into two stages: (1) phase space reconstruction and (2) forecasting. The phase space reconstruction requires three crucial parameters which are time delay (τ), embedding dimension (d) and number of nearest neighbors (k). It is crucial to determine these parameters to capture the underlying system dynamics and ensure reliable prediction outcomes.

2.2.1. Phase Space Reconstruction

Phase space reconstruction refers to embed a one-dimensional time series into an m-dimensional phase space vector. The original time series data, recorded as scalar values, is represented by $X=\{x_1,x_2,\ldots,x_{N-1,}x_N\}$ where *N* is the total length of the time series. The time series data was then split into two groups: training data and testing data. This separation ensures that the model-building and validation processes are feasible. The temporal series may be written as Eq. (1) and Eq. (2):

$$X_{training}=\{x_1,x_2,\ldots,x_l\} \tag{1}$$
$$X_{testing}=\{x_{l+1},x_{l+2},x_{l+3},\ldots,x_N\} \tag{2}$$

The training series is the data set used to identify chaotic pattern and estimates the required parameters. On the other hand, the testing time series is used to compare the performance of the forecasting model that the end of study developed. A multi-dimensional phase space is reconstructed using the time series $X_{training}$. According to Takens' Theory (Takens, 1981), *d*-dimensional phase space can be represented as Y_{i}, with time delay τ and embedding dimension, d defined as Eq. (3):

$$\mathrm{Y}_i=\{x_i,x_{i+\tau},x_{i+2\tau},\ldots,x_{i+(d-1)\tau}\} \tag{3}$$

To ensure a fair and coherent comparison, all the models and methods were run with the same parameters for both the time delay ($\tau=1$) and the embedding dimension ($d=2$), considering that these are the best options found by Mashuri et al. 2022 for predictions. It also enables fairer comparison over how performance differences of the methods are found, so any possible improvement or degradation is due the algorithms instead of parameter settings.

2.2.2. Local Linear Average Method (LLAM)

LLAM is an established method within the field of chaos theory. To examining the value of *k* by using this method because the Euclidean distance among the vector at the current value of Y_i and previous vector, $\mathrm{Y}_w(w=1,2,\ldots,i-1)$ by choosing the minimum distance to helps nearest vector. In order to get the value of *k*, the equation of $k=2d$ is being used as the previous research by Adenan et al. (2021). By letting $k=\mathrm{Y}\ \mathrm{p}$, estimation of one step ahead *T* = 1 will give Y_p and $\mathrm{Y}_{\mathrm{p+1}}$ where it will show the equal value of *d*. The value of Y_p and $\mathrm{Y}_{\mathrm{p+1}}$will be used to form the equation of $\mathrm{Y}_{p+1}=A\mathrm{Y}_p+B$. The constants of *A* and *B* are defined by using smallest squared value method. Hence, the prediction of Y_{i+T} can be calculate by using Eq. (4):

$$\mathrm{Y}_{i+1}=A\mathrm{Y}_i+B. \tag{4}$$

2.2.3. Local Mean Average Method (LMAM)

LLAM is an established method within the field of chaos theory. It introduces the final phase space $\mathrm{Y}_{i-(d-1)\tau}$, the nearest neighbors are $\mathrm{Y}_i=(\mathrm{Y}_{i_1},\mathrm{Y}_{i_2},\mathrm{Y}_{i_3},\ldots,\mathrm{Y}_{i_{l-(d-1)\tau-1}})$. The number of neighbors Y_i is $l-(d-1)\tau-1$ because the final phase space is $\mathrm{Y}_{i-(d-1)\tau}$. Then, the k nearest neighbors are listed as$\mathrm{Y}_p=(\mathrm{Y}_{p_1},\mathrm{Y}_{p_2},\mathrm{Y}_{p_3},\ldots,\mathrm{Y}_{p_n})$. For the one-step-ahead forecasting at *T* = 1, the d-dimensional phase space used is $\mathrm{Y}_{p+1}=(\mathrm{Y}_{p+1_1},\mathrm{Y}_{p+1_2},\mathrm{Y}_{p+1_3},\ldots,\mathrm{Y}_{p+1_n})$. Then, the first column of the time series for equation are listed as Eq. (5) and Eq. (6):

$$x_p=(x_{p_1},x_{p_2},x_{p_3},\ldots,x_{p_n}) \tag{5}$$
$$x_{p+1}=(x_{p+1_1},x_{p+1_2},x_{p+1_3},\ldots,x_{p+1_n}) \tag{6}$$

For the LMAM model, x_{i+1} is predicted through Eq. (7):

$$x_{i+1}=\frac{\sum_{i+1}^{p}(x_{p_i+1})}{k} \tag{7}$$

2.3. k-GA Model

In this section, the fundamental concepts of the Genetic Algorithm (GA) are discussed, as it has been explored by researchers as a novel approach in forecasting method. The field of hydrology remains relatively limited uses of GA for modelling, particularly for time series data related to hydrology. Nevertheless, this method has been used by some scientists with good results. For example, Salih et al. (2020) used the GA for predicting rainfall for Selangor, Johor, and Negeri Sembilan in order to compare hydrological regimes of Malaysia and Iraq. GA is widely used and it is one of the most popular and versatility algorithms (Nagasubramanian et al., 2018). GA is an optimization technique developed by John Holland in the 1970's and it has been applied extensively in forecasting literature, particularly in combination with neural networks (NN) (Cicek & Ozturk, 2021). GA operates based on the principles of natural selection to identify optimal solutions. In conclusion, GA is a programming-based approach that simulates natural genetic processes to solve optimization problems. This approach is appropriate for this work since before prediction, one must find the best *d*-value. Table 2 shows the basic terminology in the Genetic Algorithm (Lambora et al., 2019). The k-GA model is a hybrid approach that operates continuously. This means that no parameter changes occur during calculations within the chaotic method, but GA is incorporated into the chaotic method process. This principle is supported by studies conducted by Wang and Babovic (2016) which implemented a hybrid chaotic method and Ordinary Kriging. These studies showed that the hybrid approach operates continuously without altering any parameters in the chaotic method process.

Table 2. Basic Terminologies in Genetic Algorithm (Lambora et al., 2019)

Chromosome	A set of parameters representing a solution to the problem under study. Also referred to as an individual. Known as Kr in this study.
Population	The total number of chromosomes.
Generation	A level in the optimization process.
Fitness	The value of the fitness function.
Parents	A pair of individuals selected to undergo the reproduction process.
Reproduction	A process involving selection, crossover and mutation to improve the population by creating better chromosomes (optimal solutions). Also referred to as the reproduction process.
Children	New individuals produced after reproduction. Also referred to as offspring.
Selection	A process of selecting chromosomes from the population to become parent chromosomes and undergo crossover.
Crossover	A process where genetic material is transferred from parents to produce child chromosomes.
Mutation	A process of mutating chromosomes to prevent the algorithm from converging to suboptimal solutions.

GA is chosen as it can search for optimal or best value prior the forecasting. This is because the authors verified the forecasted values in the rows of the chosen vectors after the phase space construction and verified that the values did not coincide. Hence, some rows do have the potentiality to produce the optimal forecast results, and they have faith that GA is an appropriate technique for selecting these rows. It is believed that the hybrid k-GA model, which takes advantage of both the chaos theory and genetic algorithms, that can benefit forecasting effectively. This method is aimed to cope with the difficulties of extracting trend, complex nonlinearity and noise of the time series data. The most important genes are presented in Table 3.

Table 3. The key components of k-GA model

Chaos Identification	The first stage of the hybrid method based on chaos recognition of the mathematics equation of ground system we can analyze the genetic characteristic of territorial system. Quantities such as Cao Method are used to measure how much a system is chaotic and with phase space reconstruction it is also possible to visualize a system's trajectory in a multi-dimensional space. These methods allow the identification of chaotic features that are often missed in standard models.
Feature Selection	Using insights from chaos identification, significant variables are selected to serve as inputs for the forecasting model. The chaotic characteristics guide this process, ensuring that the selected features capture the most relevant and meaningful information about the system's behavior
Genetic Algorithm (GA)	Individuals in the population evolved over successive generations used for evolving solution to forecasting. It is based on encoding solutions, a fitness function to assess quality of a solution, and crossover and mutation to create new solutions. These processes ensure the diversity, exploration and exploitation of the solution space with the aim to make precise and ideal predictions.
Hybridization	The hybrid process depends on chaos theory with the GA. According to the chaos feature, population initialization and parameter setting can be guided, enabling the algorithm to converge to an optimum solution. This complementarity enables the model to learn the deterministic as well as chaotic features of the dataset.

The hybrid model has emerged as a competitive filed for researchers and deserves more in-depth analysis. Hybrid models combining chaotic methods with other techniques have been successful in this endeavour. The success of these hybrid models has motivated researchers to further develop and refine such approaches. In this study, the adopted hybrid model is a combination of a chaotic-based method with a Genetic Algorithm (GA), referred to as the k-GA model. A detailed explanation of the k-GA model's operations is provided in the Appendix, while Figure 2 presents a schematic representation of the overall k-GA architecture.

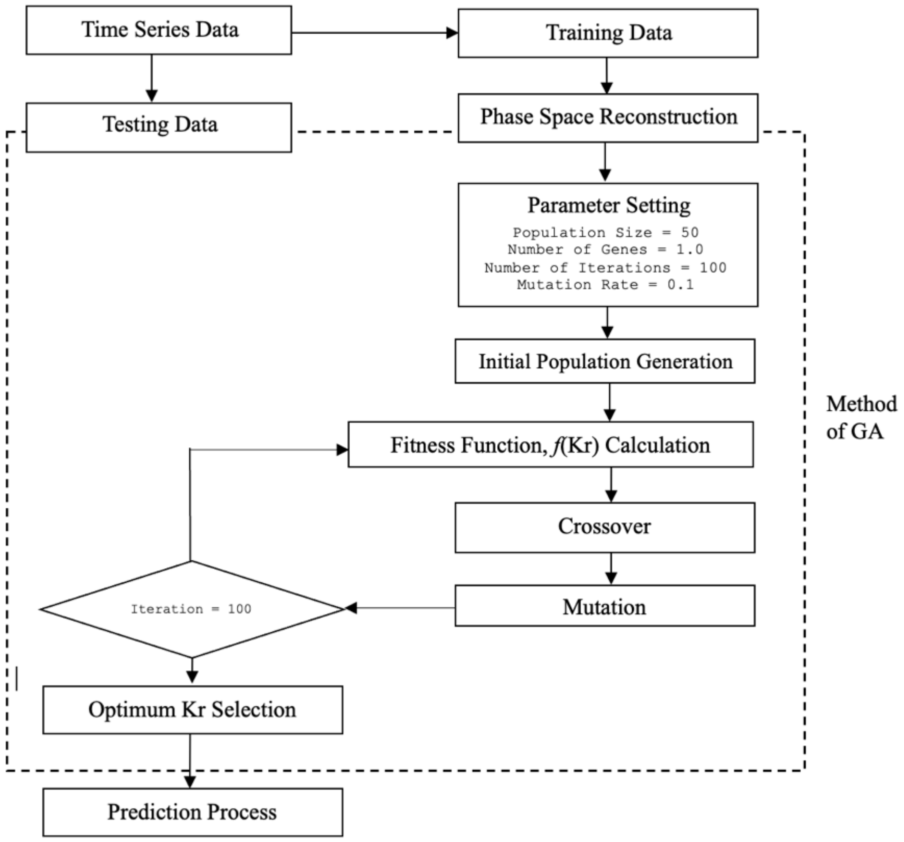

Figure 2. Basic structure of k-GA model

The k-GA model in the current paper intends to enhance the forecasting performance of the suitable forecasting. The following are steps of the k-GA process: (1) Phase space reconstruction (2) Generating an initial population (3) calculation of the fitness function (4) crossover and update of the fitness function (5) mutation and update of the fitness function and (6) forecasting. These calculations are simplified by the MATLAB R2023a to improve its speed. Such evolutionary biological evolution terms including selection, crossover and mutation are used to improve vector selection in terms of the fitness function. By iterative optimization, this method can optimize better vectors generated by solutions. Prior to execution of the GA, the following parameters should be specified. Table 4 lists the size of the parameters and their descriptions that were chosen using the study conducted by Afan et al. (2025) when they employed a hybrid model combining Genetic Algorithm (GA) and Neural Network (NN) to predict the suspended sediment load (SSL) in the Johor River.

Table 4. Explanations of parameters and components of equation for k-GA model.

Parameters and components	**Explanations**
X_{t+1}	One-step ahead forecast value
X_t	Training time series data
d	Embedding dimension
Reconstruct (X_t, d)	Phase space reconstruction for the time series
Kr	Chromosome
X_f	Forecasting target vector
g	Random gene selection
Krg	New GA chromosome
$\left\|Krg - X_f\right\|^2$	Fitness function for the new chromosome and forecasting target vector

$\underset{g}{arg\ min} \left\|Krg - X_f\right\|^2$	Optimization calculation phase for the chromosome. The gene values that minimize the forecasting error in the phase space
$\left\|Kr,_{col_i} - X_f\right\|_2$	Euclidean distance between the iii-th column in the chromosome and the forecasting target vector
$\underset{i}{arg\ max}\left(\left\|Kr,_{col_i} - X_f\right\|_2\right)$	Components for selecting the best column. The space in the chromosome closest to the first forecasting column based on Euclidean distance (fitness function)

2.3.1. Step 1: Phase Space Reconstruction

After setting the parameters, several vectors obtained during the process of phase space reconstruction are stacked to a matrix and employed as starting population for the GA process. Manual calculations of the GA method are illustrated through the following treatment examples. These examples serve to demonstrate the operational process of GA and do not represent actual computations, as real calculations involve large parameters and are performed using MATLAB R2023a software. In this hybrid approach, a row from the constructed phase space vectors is selected and referred to as a chromosome (*Kr*). For instance, Eq. (8) shows the vectors obtained after phase space reconstruction are as follows:

$$\begin{aligned} \mathbf{Y}_a &= \{\boxed{x_1}, \boxed{x_2}, \boxed{x_3}\} \\ \mathbf{Y}_b &= \{\boxed{x_3}, \boxed{x_4}, \boxed{x_5}\} \Rightarrow \boxed{} \Rightarrow \mathrm{Kr} \\ \mathbf{Y}_c &= \{\boxed{x_5}, \boxed{x_6}, \boxed{x_7}\} \end{aligned} \qquad (8)$$

Variables a*,* b and c represent the components of the chromosome. The number of *Kr* corresponds to the number of rows in the vector or the value of d. Thus, in the example above, the number of chromosomes is three.

2.3.2. Step 2: Initial Population Generation

In the GA method, each row of the vector, from the first to the last, is defined and referred to as a chromosome *Kr*. Each row in the matrix represents a chromosome, which can be expressed mathematically as Eq. (9):

$$Population = \{Kr_1, Kr_2, \ldots, Kr_d\} \qquad (9)$$

For example, if the number of rows or the embedding dimension of the vector is 6, then *d* = 6. Each *Kr* is arranged horizontally for the subsequent process. This group of chromosomes is known as the initial population.

2.3.3. Step 3: Fitness Function, f(Kr) Calculation

The fitness function, *f*(*Kr*) represents the distance between each *Kr*. This function is used to select chromosomes, with *f*(*Kr*) values sorted to identify the best ones. The best *f*(*Kr*) value is the smallest distance (McCall, 2005). This step evaluates *f*(*Kr*) for every listed chromosome. There are several methods for calculating *f*(*Kr*), such as the Euclidean distance and Manhattan distance (Truong et al., 2022). In this study, the researcher chose the Euclidean distance method, as it was already applied during the phase space reconstruction. The smallest Euclidean distance (D) is calculated with reference to the last chromosome *Kr* and other chromosomes $\left\|\mathrm{Kr}_i - \mathrm{Kr}_{j-1}\right\|$ where Kr_i represents chromosomes excluding the last one, and Kr_{j-1} is the last chromosome. The calculation for fitness function values for each chromosome using the Euclidean distance is as Eq. (10) follows:

$$f(\mathrm{Kr}) = \sqrt{(\mathrm{Y}_i - \mathrm{Y}_j)^2} \qquad (10)$$

2.3.4. Step 4: Crossover

The one-point crossover method selects the top two chromosomes to serve as parent chromosomes (*P*). For the crossover method, the offspring chromosome (*O*) can be mathematically represented as follows:

$$\begin{aligned} O_1 &= P_1[1{:}j] \cup P_2[j+1{:}l] \\ O_2 &= P_2[1{:}j] \cup P_1[j+1{:}l] \end{aligned} \qquad (11)$$

The mutation rate for the crossover method is expressed as:

$$P(i) = \frac{f(x_i)}{\sum_{i=1}^{n} f(x_i)} \tag{12}$$

where $P(i)$ is the mutation rate for selecting a random gene (i), $f(x_i)$ is the fitness function for the gene, and n is the population size.

2.3.5. Step 5: Mutation

This step will result in a single offspring chromosome generated from the mutation method applied to a selected chromosome from Step 3, which has the smallest $f(Kr)$ value. This chromosome is called as the single parent chromosome. The chromosome can be further modified by randomly selecting one gene and changing its value (El Alaoui El Fels et al., 2018). Typically, chromosomes that have undergone mutation will yield the best fitness values. Therefore, this mutated chromosome can be used for the forecasting process. Thus, the mutation method for the gene can be written as:

$$C'(p) = b \tag{13}$$

where b is a random value, and the random value obtained from the new gene population will replace the value at position p. After several new chromosomes are generated, one chromosome with the best fitness function value will be selected for forecasting. Mathematically, it can be represented by the following equation:

$$\mathrm{Kr}_{best} = arg\ max_{Kr_i} f(\mathrm{Kr}_i) \tag{14}$$

where $\mathrm{Kr}_{best} = arg\ max_{Kr_i} f(\mathrm{Kr}_i)$ is the chromosome with the highest fitness function and $\mathrm{Kr}_{best} = arg\ max_{Kr_i} f(\mathrm{Kr}_i)$ is the chromosome with the highest fitness function value in the entire population.

2.3.6. Step 6: Prediction

After completing 100 iterations, this section will select the best (optimal) chromosome for forecasting. The selection of the optimal chromosome is done randomly, as the iterative process will generate up to one hundred new chromosomes. Once completed, the forecasting process using LMAM will be carried out as follows:

$$x_{n+1} = Average(\mathrm{Kr}_{best}) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Kr}_{best,i} \tag{15}$$

Therefore, the following is the complete k-GA model equation, representing the principal novelty of this research as follows:

$$X_{t+1} = \left(Reconstruct\left(X_t, d\right) \cdot arg\ \underset{g}{min} \left\| \mathrm{Kr}g - X_f \right\|^2 \right) \cdot arg\ \underset{i}{max} \left(\left\| \mathrm{Kr},_{col_i} - X_f \right\|_2 \right) \tag{16}$$

3. RESULTS AND DISCUSSION

This study evaluates how well the new hybrid model, k-GA, performs by comparing it with two established chaos-based methods, LLAM and LMAM. These two methods were chosen because they are widely used and trusted for forecasting chaotic time series data. By measuring how k-GA stacks up against them, we can better understand the improvements the hybrid model offers. This comparison not only shows that k-GA is reliable but also highlights its strengths in making accurate and robust predictions, especially when dealing with complex water level data.

Figure 3 presents the water level predictions from the three models which are LMAM, LLAM, and k-GA across three parts of the Pahang River: upstream, midstream, and downstream. The graphs display the real water levels alongside the predictions, with time shown in hours on the x-axis and water level in meters on the y-axis. These visuals make it easy to see how close each model's predictions are to the actual measurements. Table 5 complements this by showing the correlation values for each model in the different river sections. Looking at the correlation coefficients (*CC*), k-GA consistently outperforms the other two models in all three areas. For example, in the upstream section, k-GA achieves a high *CC* of 0.9908, which is better than LMAM's 0.9487 and LLAM's 0.9587. The same pattern holds for the midstream area where k-GA scores 0.9971, compared to 0.9372 for LMAM and 0.9436 for LLAM.

Downstream, k-GA again leads with 0.9959, while LMAM and LLAM lag behind at 0.9393 and 0.9627 respectively. These results suggest that k-GA is especially good at capturing the complex flow behaviors in the river.

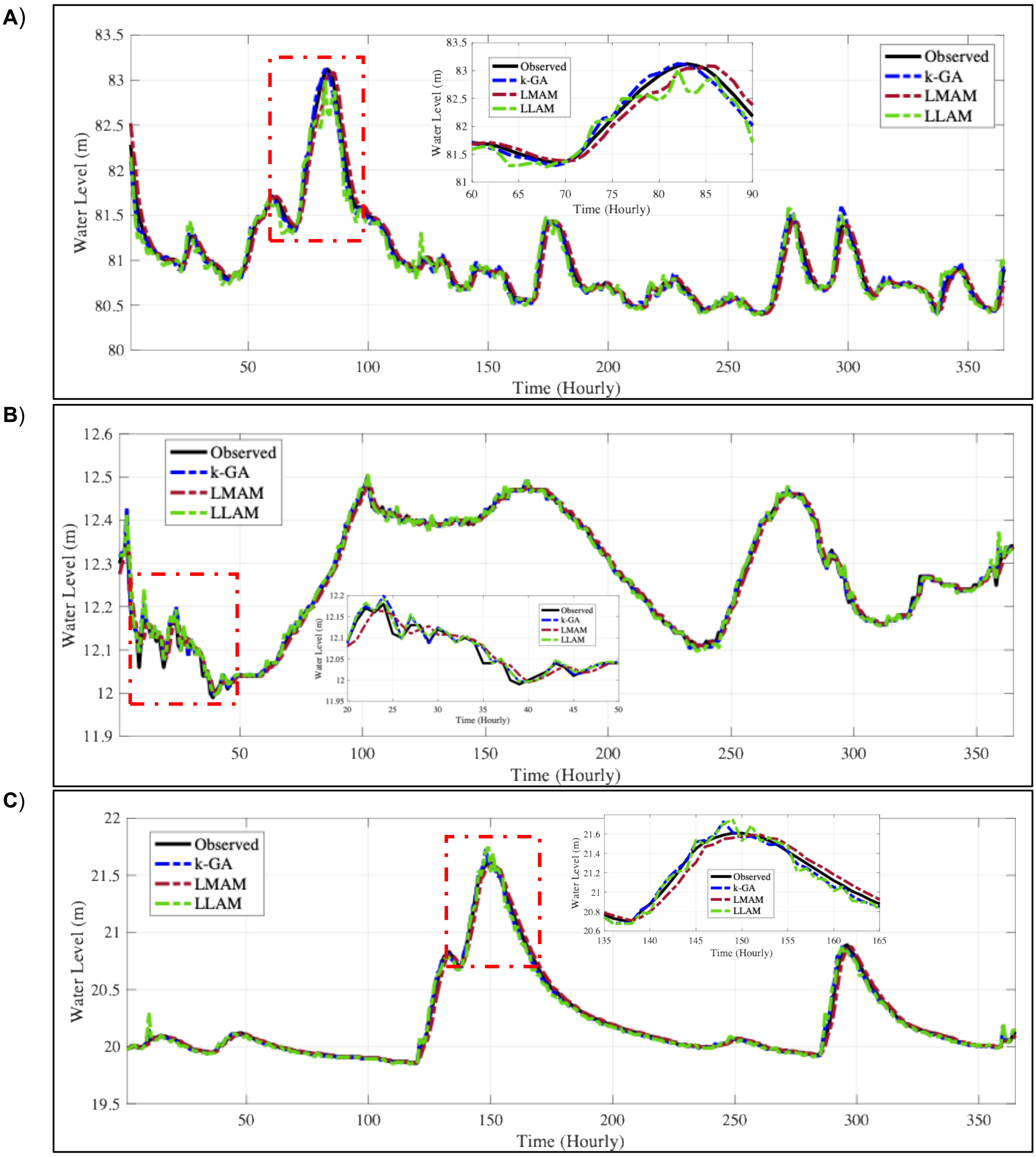

Figure 3. The analysis of the model performance of LMAM (**A**), LLAM (**B**), and k-GA (**C**) for water level prediction across three research areas at Pahang River

High correlation numbers show that the model's predictions match the real water levels closely, which means the model is reliable. The fact that k-GA performs well in all three river zones highlights how effective genetic algorithms are at handling the complicated and changing nature of river water levels. On the other hand, LMAM and LLAM do fairly well but seem to struggle more in areas where the water flow is more chaotic. Overall, the strong performance of k-GA shows that this hybrid approach could be very useful in real-life situations where accurate forecasting is vital, like managing water resources, preventing floods, and reducing disaster risks.

Table 5. Correlation coefficient value by prediction models across three research areas at Pahang River

Research Area	Correlation Coefficient		
	k-GA	LMAM	LLAM
Upstream	0.990842	0.948738	0.958723
Midstream	0.997112	0.937188	0.943647
Downstream	0.995886	0.939305	0.962738

The highest values for correlation coefficients in all the research areas were achieved by the k-GA model, indicating that this method has the best predictive ability compared to the LMAM and LLAM models This performance difference underscores the k-GA model's ability to outperform other chaos-based approaches that have been used in similar forecasting tasks. The superior performance obtained by the k-GA model is due to its novel combination of genetic algorithms, a strong optimization method, to optimize feature selection and learning. In this way, it can fully capture the complex structures in the chaotic time series data that is highly unpredictable and nonlinear in general. The capability of the k-GA model to discover and exploit informative attributes from a data source allows it to generate more accurate predictions, which would be highly useful in a task where forecast accuracy is crucial. Such results have also important theoretical implications, indicating that a coupling with genetic algorithms in chaos-based forecasting methods may lead to a significant improvement in the accuracy of prediction. This observation opens new path to enhance the reliability of predictive models in domains where high precision is indispensable. From a practical perspective, the k-GA model has great potential for use in hydrological problems, such as water level prediction. With growing importance of water resource management especially under the conditions of climate change, accurate forecasting is very much needed to manage the risk posed by both flood and drought. The k-GA model provides one potential avenue forward for addressing these key challenges and, accordingly, exemplifies how advanced optimization strategies (specifically, genetic algorithms) may open new possibilities for improving the performance of predictive modeling in complex and evolving systems. Application of this model could inform decisions and increase natural disaster preparedness, a signal of the potential that such models must impact environmentally and socially resilient systems.

4. CONCLUSION

The main findings of this study demonstrate that the proposed hybrid forecasting model, k-GA, which integrates chaos theory with genetic algorithms, significantly enhances the prediction accuracy of river water levels. The model effectively captures the nonlinear and chaotic characteristics inherent in hydrological time series, offering improved reliability and precision compared to existing chaos-based approaches. The incorporation of genetic algorithms plays a critical role in optimizing key parameters and feature selection, thereby strengthening overall forecasting performance. These results underscore the model's practical relevance, particularly in applications such as flood risk management and sustainable water resource planning. Moreover, the model shows promising adaptability for other domains involving nonlinear time series data, including financial analysis, climate studies, and medical applications. Nevertheless, certain limitations remain, notably the relatively high computational cost associated with the hybrid approach and the reliance on data from a single case study, the Pahang River. Future research should therefore prioritize enhancing computational efficiency, validating the model across more diverse datasets, and integrating real-time data to support more robust and practical implementations.

ACKNOWLEDGEMENT
The researchers sincerely thank the Malaysian Department of Irrigation and Drainage for their invaluable cooperation and for providing the time series data from three stations along the Pahang River, which were crucial for this study.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

AUTHOR CONTRIBUTION

Adib Mashuri: Conceptualization, methodology, software, data analysis, writing original draft, visualization. Nur Hamiza Adenan: Supervision, validation, methodology, writing, review, editing. Azurah A Samah: Supervision, resources, project administration, writing, review, editing. Ahmad Basri Ruslan: Data curation, software testing, formal analysis, visualization, writing. Nor Suriya Abd Karim: Investigation, interpretation. Noraini Ramlan: Literature review, data acquisition. Muhammad Faizan: Technical support, writing, review, editing.

DATA AVAILABILITY

The data that support the findings of this study are publicly available and the corresponding author can provide upon request.

DECLARATION OF GENERATIVE AI

Not applicable.

ETHICS
Not applicable.

REFERENCES

Adenan NH, Karim NSA, Mashuri A, Hamid NZA, Adenan MS, Armansyah, Siregar I. (2021). Traffic flow prediction in urban area using inverse approach of Chaos Theory. *Civil Engineering and Architecture*, 9(4), 1277-1282. doi:10.13189/CEA.2021.090429

Afan HA, Wan Mohtar WHM, Aksoy M, Ahmed AN, Khaleel F, Khan MMH, Kamel AH, Sherif M, El-Shafie A. (2025). A multi-functional genetic algorithm-neural network model for predicting suspended sediment loads. *Water Resources Management*, 39(5), 2033-2048. doi:10.1007/S11269-024-04054-W/TABLES/3

Along NZB, Deep S. (2023). Indigenous local knowledge (ILK) and authorities' knowledge integration: Case study of flood risk management in Pekan district, Pahang, Malaysia. *XVII Mexican Symposium on Medical Physics*, 2947(1), 080003. doi:10.1063/5.0167885/2915014

Alsayed A, Sadir H, Kamil R, Sari H. (2020). Prediction of epidemic peak and infected cases for COVID-19 disease in Malaysia, 2020. *International Journal of Environmental Research and Public Health*, 17(11), 1-15. doi:10.3390/ijerph17114076

Barrera-Animas AY, Oyedele LO, Bilal M, Akinosho TD, Delgado JMD, Akanbi LA. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, 7, 100204. doi:10.1016/J.MLWA.2021.100204

Cicek ZIE, Ozturk ZK. (2021). Optimizing the artificial neural network parameters using a biased random key genetic algorithm for time series forecasting. *Applied Soft Computing*, 102, 107091. doi:10.1016/J.ASOC.2021.107091

El Alaoui El Fels A, Alaa N, Bachnou A. (2018). Use of genetic algorithm in new approach to modeling of flood routing. *Journal of Oceanology and Limnology*, 37(1), 72-78. doi:10.1007/S00343-018-7293-4

Ghorbani MA, Khatibi R, Danandeh Mehr A, Asadi H. (2018). Chaos-based multigene genetic programming: A new hybrid strategy for river flow forecasting. *Journal of Hydrology*, 562, 455-467. doi:10.1016/j.jhydrol.2018.04.054

Hamidon AM, Musa S. (2022). Short-term time-series forecasting by using exponential smoothing techniques at Sungai Pahang, Malaysia. *Springer Proceedings in Physics*, 273, 443-452. doi:10.1007/978-981-16-8903-1_38

Jiang Y, Bao X, Hao S, Zhao H, Li X, Wu X. (2020). Monthly streamflow forecasting using ELM-IPSO based on phase space reconstruction. *Water Resources Management*, 34(11), 3515-3531. doi:10.1007/S11269-020-02631-3/METRICS

Kamarudin MKA, Toriman ME, Abd Wahab N, Abu Samah MA, Abdul Maulud KN, Mohamad Hamzah F, Mohd Saudi AS, Sunardi S. (2023). Hydrological and climate impacts on river characteristics of pahang river basin, Malaysia. *Heliyon*, 9(11), e21573. doi:10.1016/J.HELIYON.2023.E21573

Lahsen M, Ribot J. (2022). Politics of attributing extreme events and disasters to climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 13(1), e750. doi:10.1002/WCC.750

Lambora A, Gupta K, Chopra K. (2019). Genetic algorithm - A Literature review. *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects,* 380-384. doi:10.1109/COMITCON.2019.8862255

Lindenschmidt KE, Huokuna M, Burrell BC, Beltaos S. (2018). Lessons learned from past ice-jam floods concerning the challenges of flood mapping. *International Journal of River Basin Management,* 16(4), 457-468. doi:10.1080/15715124.2018.1439496

Lyche Solheim A, Globevnik L, Austnes K, Kristensen P, Moe SJ, Persson J, Phillips G, Poikane S, van de Bund W, Birk S. (2019). A new broad typology for rivers and lakes in Europe: Development and application for large-scale environmental assessments. *Science of the Total Environment*, 697, 134043. doi:10.1016/J.SCITOTENV.2019.134043

Mashuri A, Adenan NH, Karim NSA, Tho SW, Zeng Z. (2024). Application of Chaos Theory in different fields - A literature review. *Journal of Science and Mathematics Letters*, 12(1), 92-101. doi:10.37134/jsml.vol12.1.11.2024

Mashuri A, Adenan NH, Karim SA, Shahriman A, Rani CNA. (2022). Performance of water level forecasting based on Chaos approach using data splitting. *Environment and Ecology Research*, 10(2), 218-224. doi:10.13189/eer.2022.100211

McCall J. (2005). Genetic algorithms for modelling and optimisation. *Journal of Computational and Applied Mathematics*, 184(1), 205-222. doi:10.1016/J.CAM.2004.07.034

Mihailović DT, Malinović-Milićević S, Han J, Singh VP. (2023). Complexity and chaotic behavior of the U.S. Rivers and estimation of their prediction horizon. *Journal of Hydrology*, 622, 129730. doi:10.1016/J.JHYDROL.2023.129730

Mustaffa H, Kamarudin MKA, Toriman ME, Rosli MH, Sunardi S. (2023). Impact of suspended sediment on Pahang River development using geographic information system. *Planning Malaysia*, 21(1), 116-133. doi:10.21837/PM.V21I25.1228

Nagasubramanian K, Jones S, Sarkar S, Singh AK, Singh A, Ganapathysubramanian B. (2018). Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods*, 14(1), 1-13. doi:10.1186/S13007-018-0349-9/FIGURES/6

Pretorius K, Pillay N. (2024). Neural network crossover in genetic algorithms using genetic programming. *Genetic Programming and Evolvable Machines*, 25(1), 1-30. doi:10.1007/S10710-024-09481-7/FIGURES/9

Ramadevi B, Bingi K. (2022). Chaotic time series forecasting approaches using machine learning techniques: A review. *Symmetry*, 14(5), 955. doi:10.3390/SYM14050955

Salih SQ, Sharafati A, Ebtehaj I, Sanikhani H, Siddique R, Deo RC, Bonakdari H, Shahid S, Yaseen ZM. (2020). Integrative stochastic model standardization with genetic algorithm for rainfall pattern forecasting in tropical and semi-arid environments. *Hydrological Sciences Journal*, 65(7), 1145-1157. doi:10.1080/02626667.2020.1734813

Singh P, Dhiman G. (2018). A hybrid fuzzy time series forecasting model based on granular computing and bio-inspired optimization approaches. *Journal of Computational Science*, 27, 370-385. doi:10.1016/J.JOCS.2018.05.008

Sridhar V, Karim F, Tshimanga RM, Kilinc HC, Yurtsever A. (2022). Short-term streamflow forecasting using Hybrid Deep Learning Model based on Grey Wolf Algorithm for Hydrological Time Series. *Sustainability*, 14(6), 3352. doi:10.3390/SU14063352

Swagatika S, Paul JC, Sahoo BB, Gupta SK, Singh PK. (2024). Improving the forecasting accuracy of monthly runoff time series of the Brahmani River in India using a hybrid deep learning model. *Journal of Water and Climate Change*, 15(1), 139-156. doi:10.2166/WCC.2023.487

Tahir M, Tubaishat A, Al-Obeidat F, Shah B, Halim Z, Waqas M. (2022). A novel binary chaotic genetic algorithm for feature selection and its utility in affective computing and healthcare. *Neural Computing and Applications*, 34(14), 11453-11474. doi:10.1007/S00521-020-05347-Y/METRICS

Takens F. (1981). *Detecting strange attractors in turbulence*. Dynamical Systems and Turbulence, Warwick. doi:10.1007/BFb0091924

Truong DS, Hien LT, Tung NT. (2022). An effective algorithm for computing reducts in decision tables. *Journal of Computer*

Science and Cybernetics, 38(3), 277-292. doi:10.15625/1813-9663/38/3/17450

Vagheei H, Laini A, Vezza P, Palau-Salvador G, Boano F. (2023). Climate change impact on the ecological status of rivers: The case of Albaida Valley (SE Spain). *Science of The Total Environment*, 893, 164645. doi:10.1016/J.SCITOTENV.2023.164645

Wang X, Babovic V. (2016). Application of hybrid Kalman filter for improving water level forecast. *Journal of Hydroinformatics*, 18(5), 773-790. doi:10.2166/hydro.2016.085

Yousefi P, Courtice G, Naser G, Mohammadi H. (2020). Nonlinear Dynamic Modeling of Urban Water Consumption Using Chaotic Approach (Case Study: City of Kelowna). *Water*, 12(3), 753. doi:10.3390/W12030753

Yusoff SHM, Hamzah FM, Jaafar O, Tajudin H. (2021). Long Term trend analysis of upstream and middle-stream river in Langat Basin, Selangor, Malaysia. *Sains Malaysiana*, 50(3), 629-644. doi:10.17576/jsm-2021-5003-06

Zakaria NH, Adenan NH, Karim NSA, Mashuri A. (2021). Prediction of water level time series data for Dam at Selangor using Chaotic Approach and Local Linear Approximation method. *Journal of Science and Mathematics Letters*, 9, 10-17. doi:10.37134/jsml.vol9.sp.2.2021