

Application of hybrid machine learning models and data pre-processing to predict water level of watersheds: Recent trends and future perspective

Sarah J. Mohammed, Salah L. Zubaidi, Sandra Ortega-Martorell, Nadhir Al-Ansari, Saleem Ethaib & Khalid Hashim

To cite this article: Sarah J. Mohammed, Salah L. Zubaidi, Sandra Ortega-Martorell, Nadhir Al-Ansari, Saleem Ethaib & Khalid Hashim (2022) Application of hybrid machine learning models and data pre-processing to predict water level of watersheds: Recent trends and future perspective, Cogent Engineering, 9:1, 2143051, DOI: [10.1080/23311916.2022.2143051](https://doi.org/10.1080/23311916.2022.2143051)

To link to this article: <https://doi.org/10.1080/23311916.2022.2143051>



© 2022 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.



Published online: 11 Nov 2022.



Submit your article to this journal [↗](#)



Article views: 160



View related articles [↗](#)



View Crossmark data [↗](#)



Received: 06 May 2022
Accepted: 29 October 2022

*Corresponding author: Nadhir Al-Ansari, Department of Civil Environmental and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden
E-mail: nadhir.alansari@ltu.se

Reviewing editor:
Sanjay Kumar Shukla, School of Engineering, Edith Cowan University, Perth, Australia

Additional information is available at the end of the article

CIVIL & ENVIRONMENTAL ENGINEERING | REVIEW ARTICLE

Application of hybrid machine learning models and data pre-processing to predict water level of watersheds: Recent trends and future perspective

Sarah J. Mohammed¹, Salah L. Zubaidi^{1,2}, Sandra Ortega-Martorell³, Nadhir Al-Ansari^{4*}, Saleem Ethaib⁵ and Khalid Hashim^{6,7}

Abstract: The community's well-being and economic livelihoods are heavily influenced by the water level of watersheds. The changes in water levels directly affect the circulation processes of lakes and rivers that control water mixing and bottom sediment resuspension, further affecting water quality and aquatic ecosystems. Thus, these considerations have made the water level monitoring process essential to save the environment. Machine learning hybrid models are emerging robust tools that are successfully applied for water level monitoring. Various models have been developed, and selecting the optimal model would be a lengthy procedure. A timely, detailed, and instructive overview of the models' concepts and historical uses would be beneficial in preventing researchers from overlooking models' potential selection and saving significant time on the problem. Thus, recent research on water level prediction using hybrid machines is reviewed in this article to present the "state of the art" on the subject and provide some suggestions on research methodologies and models. This comprehensive study classifies hybrid models into four types algorithm parameter optimisation-based hybrid models (OBH), pre-processing-based hybrid models (PBH), the components combination-based hybrid models (CBH), and hybridisation of parameter optimisation-based with preprocessing-based hybrid models (HOPH); furthermore, it explains the pre-processing of data in detail. Finally, the most popular optimisation methods and future perspectives and conclusions have been discussed.

ABOUT THE AUTHOR



Nadhir Al-Ansari

Nadhir Al-Ansari: Professor at the department of Civil, Environmental and Natural Resources Engineering at Lulea Technical University Sweden. Research interests are mainly in Water Resources and Environment. Served several academic administrative post (Dean, Head of Department). Publications include more than 790 articles in international/national journals, chapters in books and 22 books. He supervised more than 70 postgraduate students. He executed more than 60 major research projects in Iraq, Jordan, Sweden and UK. Awarded several scientific and educational awards, among them is the British Council on its 70th Anniversary awarded him top 5 scientists in Cultural Relations. One patent on Physical methods for the separation of iron oxides. Supervised more than 70 postgraduate students at different universities. Member of several scientific societies e.g., International Association of Hydrological Sciences, Chartered Institution of Water and Environment Management, Network of Iraqi Scientists Abroad, etc. Member of the editorial board of 32 international journals.

Subjects: Environmental Studies; Environmental Issues; Environment & Resources;

Keywords: Water level forecasting; data pre-processing; meta-heuristic algorithms; hybrid model

1. Public interest statement

Providing enough drinkable water, especially in major cities, is a huge task. Cities have grown without proper planning, resulting in the loss of vegetative cover and soil impermeability. As a result, hydrological and meteorological problems have arisen—changes in air temperature, evapotranspiration, and flood danger. So, the prediction of water level data is essential to managing freshwater resources that lead to the achievement of sustainability. This article comprehensively analyses the literature on using various machine learning models, metaheuristic algorithms, and pre-processing data methodologies for water level forecasting. We hope that by outlining the benefits and drawbacks of various training environments for these complex models, our research study will make it easier for readers to make a choice. This review offers the reader information about how different factors impact these forecast models and what techniques perform best in various scenarios.

2. Introduction

Providing enough drinkable water, especially in major cities, is a huge task. Cities have grown without proper planning, resulting in the loss of vegetative cover and soil impermeability. As a result, hydrological and meteorological problems have arisen—changes in air temperature, evapotranspiration, and flood danger (Ethaib et al., 2022; De Souza Groppo et al., 2019). Beachfront zones are dynamic and vulnerable to natural disturbances caused by environmental change and human activities. They are particularly vulnerable since they are typically densely populated, with more than 40% of the population living in shoreline areas. The level of risk associated with this can be rather substantial (Hussaini et al., 2020). Additionally, the overuse of groundwater directly affects the sea water level (Koutsoyiannis, 2020). Water level forecasting is a vital mission for the hydrologists, relevant authorities and engineers in improving a sustainable conceptual design of water infrastructures. Forecasting water levels aid in flood and drought control and practical applications such as studying the behaviour of rivers, lakes, and reservoirs (Deng et al., 2021). The watershed's temporal and spatial variability, climate variability, seasonal patterns, and local to regional-scale heterogeneity in precipitation and temperature patterns all influence water level (Yarar et al., 2009). Flood monitoring, river level monitoring, wetland studies, tidal studies, groundwater monitoring and surface water monitoring are all typical uses for water level monitoring (Rakshitha & Maheshan, 2020).

Hurst was the first to propose a method to control the water level in the Nile River by taking a comprehensive view of the Nile River, from its headwaters in the African Great Lakes and Ethiopian plains to the great delta on the Mediterranean, where its recurrent floods and unpredictable flows were a serious development hindrance. Early attempts to regulate the flow by building a dam at Aswan were only partially successful (Graves et al., 2017). Then, this phenomenon has been identified in many key hydrological-cycle processes, such as Dimitriadis (Dimitriadis et al., 2021) used this phenomenon to identify stochastic similarities in marginal distribution and dependent structure of the hydrological cycle. The Hurst phenomenon is so important in the generation and predictability of a process, since it expresses the high variability/uncertainty observed in the time series that can influence the prediction in both the long- and short-term fluctuations.

The Hurst phenomenon was employed for long-range dependence, Dimitriadis and Koutsoyiannis (Dimitriadis & Koutsoyiannis, 2018) utilised the symmetric-moving-average (SMA) scheme for the stochastic synthesis of a stationary process for estimating any dependence structure and marginal distribution. As well as taking into consideration the Hurst effect, Rozos

(Rozos et al., 2021) employed a multilayer perceptron (MLP) network to stochastically synthesize time series of daily rainfall. Unlike other existing similar approaches, this MLP-based strategy is unique in that it replicates the statistical characteristics of the related historical time series at various scales (Hurst effect).

In the past, different water level forecasting models have been developed to increase the capabilities of water level forecasting. Machine learning (ML) techniques can learn from past experience (data) and create new (mathematical) models that can be applied to new data. ML has been applied in many areas of the hydrology fields, for example, stream flow (Kilinc, 2022; Kilinc & Haznedar, 2022). There are different types of ML used in the field of water level prediction, such as artificial neuro-fuzzy inference system (ANFIS), artificial neural network (ANN), support vector regression (SVR), and a variety of hybrid models (Çimen & Kisi, 2009; B. Li et al., 2016; Tiu et al., 2021; Yarar et al., 2009). Also, recently different techniques have been used for different areas, such as Young (Young et al., 2015), Zhang (J. Zhang et al., 2022), and Chen (X. Chen et al., 2022).

It might be challenging for researchers to select which forecasting model is best suited to their study activity due to the great number of models available. The literature on water level forecasting issues can be looked at from various angles. Focusing on the supply side of the issue, Wee (Wee et al., 2021) reviewed the paper on ML applications for studying reservoir water level prediction strategies. The main focus was on ANN, ANFIS, BA, COA, and support vector machine (SVM), as well as their main benefits. Zhu (Zhu et al., 2020) provided the literature with a complete overview of how ML models can be used to predict water-level dynamics in lakes. Seven popular ML model types are examined among the many existing ML models: ANFIS, ANN, SVM, hybrid models, evolutionary models, extreme learning machines (ELM), and deep learning (DL). In addition, model inputs, data split, model performance criteria, and model inter-comparison were all studied. Hussaini (Hussaini et al., 2020) reviewed the global water level fluctuation modelling project to provide evidence for further improvement in modelling. The scientific theories behind the modelling approaches were closely examined to better grasp their primary features and to present a comprehensive picture of the current water level fluctuation modelling effort.

Despite numerous advances in water level forecasting systems, no global strategy can consistently outperform all models throughout all research areas. Each circumstance must be studied independently, with the effectiveness of each technique or set of techniques being assessed (De Souza Groppo et al., 2019). Hybrid models have contributed tremendously to methodological advancement in water level prediction. Hybrid models have been widely used in water level forecasting during the last decade since their performance is far superior to standalone models. A hybrid model combines the advantages of each distinct model, leading to high prediction performance for time series with shorter time scales and longer lead times (Fung et al., 2019). Hybridisation has emerged as a promising technique for overcoming numerous drawbacks of standalone methods while also improving prediction accuracy (Hajirahimi & Khashei, 2022). There are several types of hybrid models, e.g., the hybrid model that combines a physical model and a machine learning model (Nualtong et al., 2021; Xu et al., 2019), and the hybrid stochastic-ML model proposed by Koutsoyiannis and Montanari (Koutsoyiannis & Montanari, 2022). Papacharalampous (Papacharalampous et al., 2019) showed that the hybrid stochastic-ML model is better than single stochastic or ML models.

To the best of the authors' knowledge, no review paper has been written specifically to study hybrid machines in water level forecasting. In this review paper, the hybrid models will be studied in detail with the pre-processing of the data for the recent year. It can be observed the increase in the use of the hybrid machine in recent years (Figure 1). According to the findings of the summarised articles, certain countries contributed more in the field of water levels than others (Figure 2). The most active societies in water level studies, for example, are China with eight articles and Malaysia with four. The results of Fig. 1 and 2 belong to the WL prediction papers

Figure 1. Studies' number of hybrid ML models for water level forecasting over the years.

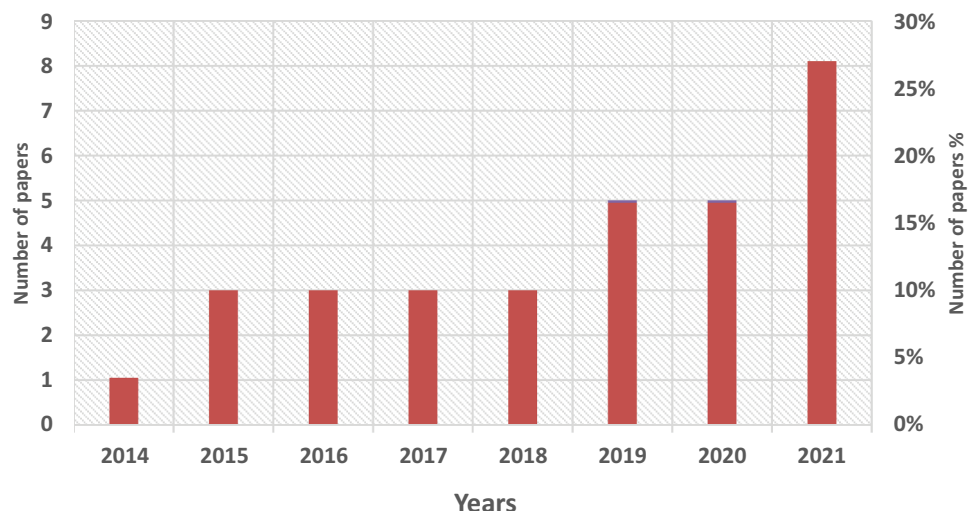
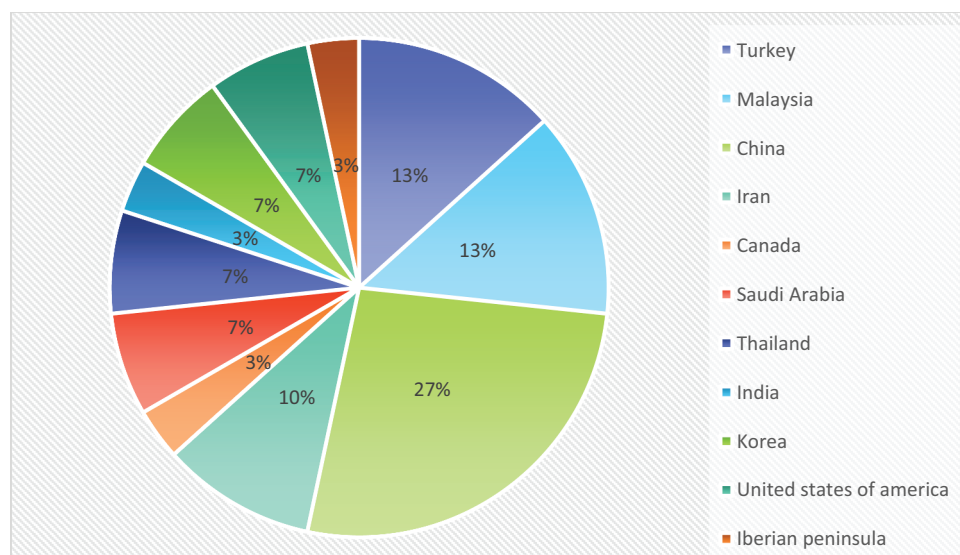


Figure 2. Countries' contribution to employing hybrid ML models from period (2014–2021).



published between 2014–2021 considered in this study. The current study's objective is demonstrating briefly of data pre-processing techniques and hybrid machine learning models of the WL prediction papers published between 2014–2021 that can aid in understanding the gaps in the literature concerning the use of hybrid techniques in WL forecasting.

The paper is organised as follows—an overview of famous models in the second section and data pre-processing in the third section. Then the hybrid models are divided into four sections in this paper, pre-processing-based hybrid models (PBH), the components combination-based hybrid models (CBH), hybridisation of parameter optimisation-based with preprocessing-based hybrid models (HOPH), and parameter optimisation-based hybrid models (OBH) in the fourth section—finally, the conclusion and future perspective.

3. Methods of forecasting water level

There are two types of water level forecasting methods: linear and nonlinear (De Souza Groppo et al., 2019). Linear approaches, such as multiple linear regression methods (MLR), autoregressive integrated moving averages (ARIMA) (e.g., (Kisi et al., 2012; Yarar et al., 2009)). Also, nonlinear methods, such as SVM, ANN, ANFIS, and hybrid methods (e.g., (Bazartseren et al., 2003; Al Visi

et al., 2006; Yarar et al., 2009)). The hybrid models will be dealt with in particular due to their merit and superiority over traditional models and single ML models.

3.1. Artificial neural network

An artificial neural network (ANN) is a large-scale distributed parallel information analysis theory with performance characteristics that are similar to those of a biological neural network in the human brain. ANNs are technologically advanced, and they can accomplish a lot of large-scale computing in a short amount of time. They are inspired by human cognition and neurobiology by a mathematical model (Mohammadi, 2021). The ANN tool could be particularly useful in situations when mathematically establishing the relationship between the dependent and independent variables for any physical occurrence is difficult. Using historical data, ANN can create a relatively accurate prediction of the modelled parameters.

A neural network is made up of a network of processing elements (PEs) connected by various weights and topologies. PEs are typically laid out in layers. The input layer receives the input variables for forecasting, the output layer calculates the output, and the remaining levels (hidden layers) transform the inputs into outputs. An ANN's layers can be fully or partially connected, and the weights must be changed correspondingly for forecasting purposes. Training algorithms are responsible for modifying the weights to reduce the predicting error (Ghalekhondabi et al., 2017). Multilayer feed-forward neural network (FFNN) learned using the back-propagation technique, recurrent neural network (RNN), bayesian neural network (BNN), echo state network (ESN), and long short-term memory (LSTM), are the most prominent ANN models.

The main characteristics of the ANN model are that they can train and test data series continuously to increase forecast accuracy, and can handle nonlinear data series (Z. Zhang et al., 2018).

ANN can be used to replicate any physical phenomenon (M.Y.A. Khan et al., 2016), which makes it appealing for use in a variety of domains, such as water demand prediction (Ömer Faruk, 2010; Zubaidi, Abdulkareem et al., 2020; Zubaidi, Dooley et al., 2018; Zubaidi, Ortega-Martorell et al., 2020), streamflow forecasting (Apaydin et al., 2021; F.F. Li et al., 2021; Gunathilake et al., 2021; Huang et al., 2021; Kilinc, 2022; Kilinc & Haznedar, 2022; Kim et al., 2022; Niu & Feng, 2021; Rahimzad et al., 2021), water quality predictions (Abba et al., 2017; W. Li et al., 2020; Sami et al., 2021; Stamenković, 2021; Tahraoui et al., 2021; Y.-F. Zhang et al., 2020), and drought prediction (Adamowski et al., 2012; Ahmadi et al., 2021; Dikshit et al., 2020; M.M.H. Khan et al., 2020; Nabipour et al., 2020; Nourani et al., 2019). In lakes, rivers, and seas, ANNs have been routinely used to forecast water levels (Ali Ghorbani et al., 2010; Bazartseren et al., 2003; Coulibaly, 2010; Hrnjica & Bonacci, 2019; Leahy et al., 2008; Piasecki et al., 2017; Ren et al., 2020; Tiu et al., 2021; Al Visi et al., 2006).

3.2. Support vector machine

The support vector machine (SVM) technique has gained popularity as a modern kind of statistical learning over the last four decades, and it has been proven to be a quick and successful tool. Additionally, it is a classification tool that is a supervised technique in machine learning that optimises the range of differences between the two groups.

By modifying the data using a kernel trick technique, SVM can be used for classification or regression tasks. On the other hand, the SVR technique is defined as a regression or prediction method that keeps all of the fundamental characteristics of the optimum margin methodology. In a nutshell, both the SVR and the SVM, with minor differences, follow the same value (Ibrahim et al., 2022). Understanding the three main elements that make up the foundation of SVM and SVR models, which are described below, will help you comprehend the SVM/SVR models more clearly.

- Hyperplane separation into high-dimensional spaces.
- To solve a nonlinear issue, use kernel functions and their methods.

- Soft margins are used to reduce error.

One of the most important features of the SVR model is the availability of various kernel functions, which allows for multiple options (Ibrahim et al., 2022).

3.3. Artificial neuro-fuzzy inference system

Fuzzy logic models have proven to be effective for resolving complex computational issues. It can deal with nonlinearity, uncertainty, and subjective information. The adaptive neuro-fuzzy inference system (ANFIS) is the most extensively used fuzzy logic model. The ANFIS is a multi-layer feed-forward network that employs a neural network learning algorithm. It can recognise nonlinear boundaries, use fuzzy logic to differentiate nonlinear equations and map the input-output space. It is capable of achieving a very nonlinear mapping. Choosing the type of interfering system, such as mamdani, sugeno, and tsumoto; aggregation; and defuzzification are the stages of ANFIS (Karaboga & Kaya, 2020).

ANFIS is employed in a variety of domains that require dealing with significant nonlinearity, and one of these fields is hydrology, where words like drought prediction are used (Mohamadi et al., 2020), streamflow (Samanataray & Sahoo, 2021), and water quality forecasting (Al-Mukhtar & Al-Yaseen, 2019). The ANFIS models have long been used to predict water levels, e.g., (Çimen & Kisi, 2009; Mohammadi et al., 2020).

ANFIS model's characteristics include interpreting fuzzy rules using natural language and simulating nonlinear functions with arbitrary complexity and insufficient data (Fung et al., 2019).

3.4. Random forest

In an attempt to improve the accuracy of decision trees, Breiman (Breiman, 2001) proposed an improved method named a random forest (RF). In this method, classification, and regression tasks can be achieved depending on the base models whether these models are classification trees or regression trees (Tyrallis et al., 2019). The name of this algorithm is inspired by the movement from the root of one tree to its terminal's nodes. Accordingly, this method has been increasingly used due to this movement. In other words, once a tree within a particular forest is split, the RF method chooses a random subset of the independent variables. Not only the predictive accuracy but also the running time is also relatively improved (Herrera et al., 2010; Seo et al., 2018). We would refer the readers to Breiman (Breiman, 2001) to get detailed information regarding how the RF is mathematically formulated.

RF can handle large datasets with multiple features, and modelling accuracy improves as the number of trees rises (Ibrahim et al., 2022).

3.5. Deep learning model

As artificial intelligence has quickly advanced, numerous studies have successfully used deep learning models like the long short-term memory (LSTM), stack Autoencoder, and deep restricted Boltzmann machine. One class of advanced ANN network among these deep learning techniques is the LSTM, which has feedback connections built into its model architecture. In the meantime, it comprises memory blocks with self-connection (in the hidden layer), which can store the network's temporal state (Zhu et al., 2020). Applications of deep learning models for WL forecasts, such as (Kim et al., 2022; Shuofeng et al., 2021; Xie et al., 2021).

4. Data pre-processing

Data pre-processing methods are regarded as critical to the process of data mining. Data pre-processing is essential to ensure that all predictors receive adequate importance during the learning phase and speed up the procedure. These methods are critical in models because they promote high accuracy and low computational costs during the learning phase, as noisy and

unreliable information in data records will negatively impact the training stage and result in a poor model (Khudhair et al., 2022).

4.1. Normalisation

This method suggests reducing outliers' impact by smoothing the answer space. The continuous variables must be transformed for the time series to be normal or nearly normally distributed. In addition, if the time series of the dependent variable is not normally or nearly normally distributed, the model's findings are degraded (Zubaidi, Kot et al., 2018; Zubaidi, Ortega-Martorell et al., 2020).

There are various methods for normalisation such as, the natural logarithm was employed to normalise the data to decrease the effect of multicollinearity between input variables, Z-score normalisation is a traditional method of standardisation that uses the mean and standard deviation to standardise parameters, Min-Max normalisation is among the most common and widely utilised data normalisation methods, and decimal scaling is the process of moving the decimal point of the variable's values to achieve the normalised value (Alawsi et al., 2022).

4.2. Cleaning

Noise and outliers can have a negative impact on data analysis and the suggested model's performance as well. As a result, data cleaning is required to discover and eliminate data corruption (Zubaidi et al., 2022). Where an outlier is a case in which the extreme value on one variable or the strange combination of scores on a variety of variables distorts statistics, and noise is an undesirable variation in the dependent variables that are estimated by covariate scores (Tabachnick & Fidell, 2013).

There are different noise components in each time series. The most efficient techniques to denoise the original time series are by analysing them into multiple components (such as, Wavelet, Empirical Mode Decomposition, Singular Spectrum Analysis)(Alawsi et al., 2022).

4.3. Selection of best model input

The selection of explanatory factors as model input data is a crucial stage in creating any successful prediction model (Maier & Dandy, 2000).

Most of the previous research studies only employed one or two steps of data pre-processing. Also, several studies did not apply data pre-processing techniques (see Table 1), which may adversely affect the accuracy of the forecast's result (see Table 2).

5. Hybrid models classification

Researchers have developed unique models in response to the requirement for enhanced reliability, capability, and accuracy in data-driven methodologies. Hybrid models are being developed to fulfil new requirements; their main aim is to combine the benefits of two or more methods to improve the capabilities of single models. These hybrid models are typically made up of various procedures, with one serving as the principal technique and the others as pre-or post-processing methods (Zubaidi, Ortega-Martorell, Al-Bugharbee et al., 2020). According to the reviewed papers, hybrid models can be classified into four groups: the components combination-based hybrid models (CBH), pre-processing-based hybrid models (PBH), parameter optimisation-based hybrid models (OBH), and hybridisation of parameter optimisation-based with preprocessing-based hybrid models (HOPH) as in Hajirahimi and Khashei (Hajirahimi & Khashei, 2022) (see Figure 3) and (Table 2).

5.1. Pre-processing-based hybrid models (PBH)

In pre-processing-based hybrid models, the input data is first pre-processed employing different methods such as decomposition-based, denoising-based, feature selection, dimensionality reduction, and data cleaning approaches. In general, the main motivation for decomposition-based approaches is to divide a time series into several components with varying degrees of complexity.

Table 1. Summary of studies using data pre-processing in water level forecasting

Reference	Normalisation	Cleaning	Selection model input
Altunkaynak and Kartal (Altunkaynak & Kartal, 2019)	×	✓	×
Ghorbani (Ghorbani et al., 2018)	×	×	×
Ehteram (Ehteram et al., 2021)	×	×	×
Liu (Liu et al., 2021)	×	×	✓
Chen (N. Chen et al., 2019)	×	×	✓
El-Diasty and Al-Harbi (El-Diasty & Al-Harbi, 2015)	×	×	×
Loh (Loh, Ismail, Khamis et al., 2019)	✓	✓	×
Loh (Loh, Ismail, Mustafa et al., 2019)	✓	✓	×
Panyadee (Panyadee et al., 2017)	×	×	×
Imran (Imran et al., 2021)	×	×	✓
Shafaei and Kisi (Shafaei & Kisi, 2015)	×	✓	✓
Kisi (Kisi et al., 2015)	×	×	×
Seo (Seo et al., 2015)	×	✓	✓
Tao (Tao et al., 2021)	×	×	✓
Pan (Pan et al., 2020)	×	✓	×
Zhong (Zhong, Jiang et al., 2017)	×	✓	✓
Mohammadi (Mohammadi et al., 2020)	×	×	✓
Xie (Xie et al., 2021)	×	✓	✓
Ghorbani (Ghorbani et al., 2017)	×	×	✓
Lineros (Lineros et al., 2021)	×	×	×
Zhong (Zhong, Guo et al., 2017)	×	✓	✓
Soleymani (Soleymani et al., 2016)	×	×	×
Malekpour and Mohammad Rezapour Tabari (Malekpour & Mohammad Rezapour Tabari, 2020)	✓	×	×
Altunkaynak (Altunkaynak, 2019)	×	✓	×
Deng (Deng et al., 2021)	×	×	✓
Phitakwinai (Phitakwinai et al., 2016)	×	×	×
Zhang and Lou (Zhang & Lou, 2016)	✓	×	×
Wang (Wang et al., 2020)	×	✓	×

(Continued)

Reference	Normalisation	Cleaning	Selection model input
Nguyen (Nguyen et al., 2021)	×	×	×
Nie (Nie et al., 2021)	×	×	×
El-Diasty (El-Diasty et al., 2018)	×	×	×

While the primary aim of the filter and denoising-based approaches is to detect and remove existing noise in the underlying time series. In the second step, the screened time series is forecasted by the appropriate individual model (Hajirahimi & Khashei, 2022).

Wavelet transformation can be used as a pre-treatment tool to help solve different problems in a variety of disciplines of inquiry. Wavelet analysis has gained prominence in recent years, particularly for overcoming the limitations of stochastic forecasting models such as ARIMA or classic neural network techniques. Wavelet analysis can be used in statistics to denoise data, estimate nonparametrically, and compress data (Ozan Evkaya & Sevinç Kurnaz, 2020).

The daily water levels at the northern and southern boundaries of the Bosphorus Strait are predicted using a combination of discrete wavelet transform-fuzzy (DWT-Fuzzy) and continuous wavelet transform-fuzzy (CWT-Fuzzy) models. Based on evaluation criteria (i.e., RMSE and CE), the CWT-Fuzzy model outperformed both the DWT-Fuzzy and standalone Fuzzy models for forecast lead-times up to 7 days (Altunkaynak & Kartal, 2019).

Loh (Loh, Ismail, Mustafa et al.,) used discrete wavelet transformation (DWT) with different kinds of mother wavelets (i.e., haar, sym2, sym3, db2, db3, coif1, coif2, and coif3) to denoise the data. An artificial neural network (ANN) model is also applied to simulate the Kelantan River water level. The results revealed that the performance of the sym3 offers the best scenario of DWT and the hybrid model DWT-ANN better than the standalone ANN model.

Shafaei and Kisi (Shafaei & Kisi, 2015) combined the wavelet (W) method with three different prediction models, auto-regressive moving average (ARIMA), support vector regression (SVR) models, and adaptive neuro-fuzzy inference system (ANFIS), to forecast monthly lake level variations. The suggested combined models W-SVR, W-ANFIS and W-ARMA are compared against single ARMA, SVR, and ANFIS models based on their performance accuracy. The combined models provide improved precision in forecasting lake levels in the study region, according to the findings. Also, the W-SVR model slightly outperforms the other combined models.

Seo (Seo et al., 2015) applied two hybrid models, which are wavelet-based adaptive neuro-fuzzy inference system (WANFIS) and wavelet-based artificial neural network (WANN). A time series is decomposed into approximation and detail components using wavelet decomposition. The findings of this study showed that combining wavelet decomposition with artificial intelligence models can be a valuable tool for accurately forecasting daily water levels, and it can outperform standalone forecasting models. Also, the WANFIS yields the best performance.

Xie (Xie et al., 2021) proposed a deep learning approach called long-, short-term memory network combined with discrete wavelet transform (WA-LSTM) and compared it with WA-ANN and WA-ARIMA models for daily water level prediction in Yangtze River, China. A novel LSTM network is used to learn generic water level features through layer-by-layer feature granulation with a greedy layer-wise unsupervised learning algorithm. The wavelet transform is applied to decompose time series into details and approximation components to understand temporal

Table 2. Summary of studies using hybrid models for water level forecasting

year	Reference	Location	Time scale	Variable used	Period	Models used	Best model	Measures of accuracy
2019	Altunkaynak and Kartal (Altunkaynak & Kartal, 2019)	Northern to southern Bosphorus strait	Daily	Water level, wind speed, direction, atmospheric pressure, conductivity, temperature, depth	2004–2005	DWT-fuzzy CWT-fuzzy Fuzzy stand alone	CWT-fuzzy	RMSE, CE.
2018	Ghorbani (Ghorbani et al., 2018)	Winnepesaukee and Cypress lakes in United States of America	Monthly	Water level	1938–2005 1942–2011	MLP-GSA, MLP, MLP-PSO, MLP-FFA, AREMA	MLP-GSA	COE, COC, RRMSE, RMSE
2021	Ehteram (Ehteram et al., 2021)	Urmia Lake in Iran	Monthly	Rainfall, water level, temperature	1940–2004	ANFIS-SO, MLP-SO	ANFIS-SO	RMSE, RI, NSE
2021	Liu (Liu et al., 2021)	Dongjiang River in China	Daily and monthly	Water level	1989 – 2011	VC, BVC (BMA-VC), ANFIS	BVC(BMA-VC)	RMSE, R ² , NSE
2019	Chen (N. Chen et al., 2019)	Han River in China	Monthly	Evapotranspiration, precipitation, ground temperature, humidity, sunshine-hours, wind direction, atmospheric pressure, temperature water level	2010–2017	BPNN-GA, BPNN-IGA	BPNN-IGA	RMSE, NSE, R, MAE, MARE

(Continued)

Table 2. (Continued)

year	Reference	Location	Time scale	Variable used	Period	Models used	Best model	Measures of accuracy
2015	El-Diasty and Al-Harbi (El-Diasty & Al-Harbi, 2015)	Along the Saudi and Canadian water SIX gauges (1) Yarmouth (Canada) (2) Saint John (Canada) (3) Charlottetown (Canada) (4) Ras Tanura (Saudi Arabia) (5) Jizan (Saudi Arabia) (6) Al-Khafji (Saudi Arabia)	Hourly	Water level for three gauges, water level and atmospheric data for three gauges other	One month and week	Wavelet network model (WA-ANN), harmonic analysis method, ANN	WA-ANN	RMSE
2019	Loh (Loh, Ismail, Mustafa et al.,)	Kelantan River in Malaysia	Daily	Streamflow, water level, the temperature, humidity, dew point temperature, wind speed, and pressure	2005–2014	DWT (haar)-ANN, DWT (sym2)-ANN, DWT (sym3)-ANN, DWT (db2)-ANN, DWT (db3)-ANN, DWT (coif1)-ANN, DWT (coif2)-ANN, DWT (coif3)-ANN, ANN	DWT (sym3)-ANN	MAE, RMSE, MSE, MAAPE
2020	Malekpour and Mohammad Rezapour Tabari (Malekpour & Mohammad Rezapour Tabari, 2020)	Amirkabir dam in Karaj, Iran	Monthly	Water level, precipitation, evaporation, inflow, and outflow	1986–2018	ANFIS, SVR, ANN, RBFNN, SVR-SICM	SVR-SICM	R, MSE, NMSE, LA, FB, MAE, EFF, MAPE, NSE

(Continued)

Table 2. (Continued)

year	Reference	Location	Time scale	Variable used	Period	Models used	Best model	Measures of accuracy
2016	Phitakwinai (Phitakwinai et al., 2016)	Chao Phraya River in Thailand	Hourly	Water level	7 different crucial flooding events from 2005 – 2006	MLP, MLP-CS, SVM	MLP-CS	MAPE
2016	Zhang and Lou (Zhang & Lou, 2016)	Gaoliangjian gate	Monthly	Water level	1961-1972-1977-1982-1988-2010-2011-2014	The grey model, RBF, the grey-RBF	The grey model - RBF	
2019	Loh (Loh, Ismail, Khamis et al., 2019)	Kelantan River in Malaysia	Daily	Water level, rainfall, streamflow, temperature, humidity, dew point temperature, wind speed, and pressure	2005-2014	ANN, EMD-ANN	EMD-ANN	MAE, MSE, RMSE, MAAPE
2021	Nguyen (Nguyen et al., 2021)	Han River in south Korea	Hourly	Rainfall, water level	2003-2020	XGBOOST-GA, XGBOOST-DE, RF, CART	XGBOOST-GA, XGBOOST-DE	RMSE, CC, NSE, BIAS, MAE, MAPE, TI
2021	Nie (Nie et al., 2021)	Pinghe basin in China	Hourly	Water level, rainfall	2010-2020	CNN-BILSTM, SVM, TCN, BILSTM	CNN-BILSTM	RMSE, R ² , NSE
2017	Panyadee (Panyadee et al., 2017)	Mae-Bon River in Thailand	Hourly	Water level	22 days	ANN-PSO, ANN	ANN-PSO	MSE, MAPE
2020	Imran (Imran et al., 2021)	Jhelum River in India	Monthly	Water level, rainfall, temperature	1980-2010	ANFIS-GA, ANFIS	ANFIS-GA	RMSE, MAE
2015	Shafaei and Kisi (Shafaei & Kisi, 2015)	Van Lake in Turkey	Monthly	Water level	38 years	WA-AREMA, WA-ANFIS, WA-SVR, AREMA, ANFIS, SVR	WA-SVR	RMSE, R, MAE
2015	Kisi (Kisi et al., 2015)	Urmia Lake in Iran	Daily	Water level	1972-2003	SVM-FFA, ANN, GP	SVM-FFA	RMSE, R, R ²

(Continued)

Table 2. (Continued)

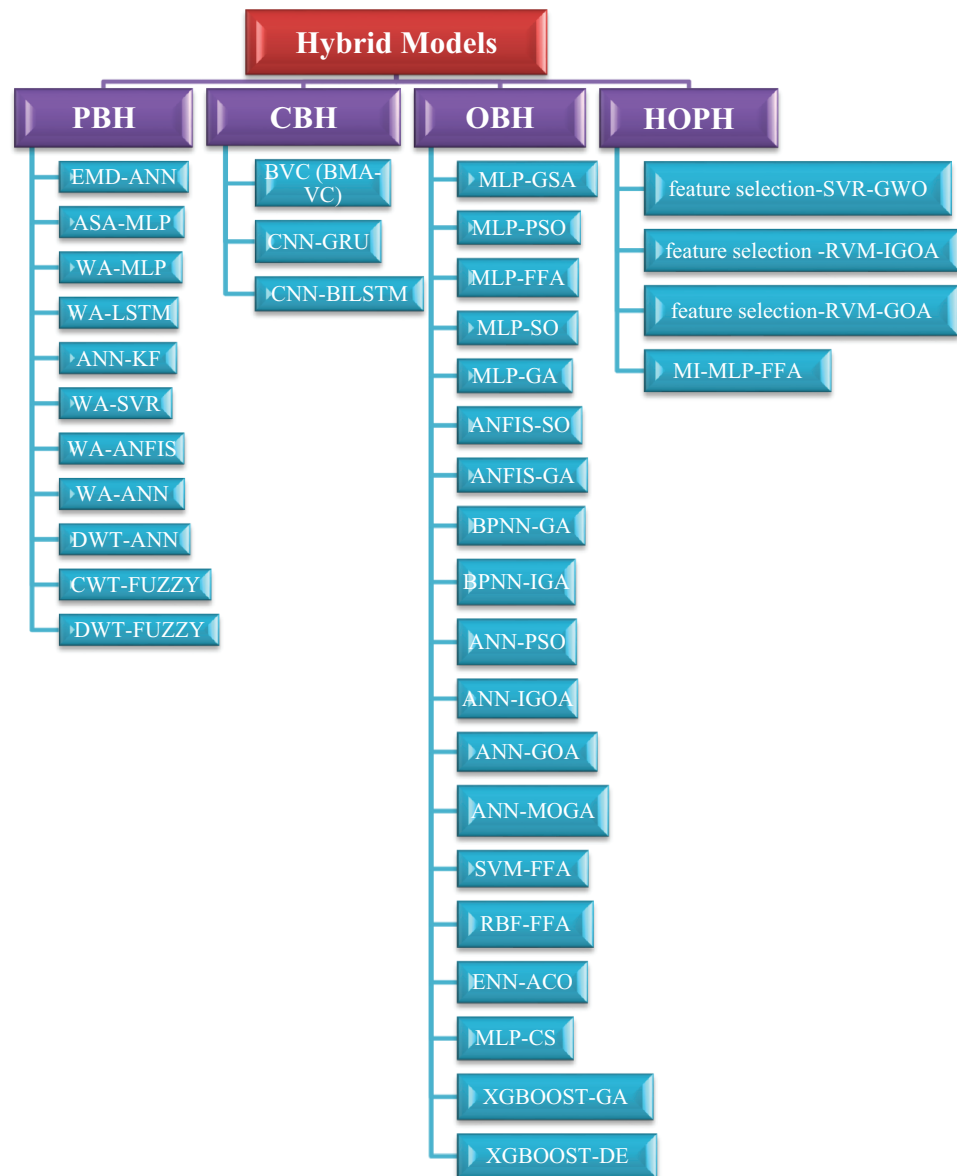
year	Reference	Location	Time scale	Variable used	Period	Models used	Best model	Measures of accuracy
2014	Seo (Seo et al., 2015)	Nadong River in Korea	Daily	Water level	2002–2013	WA-ANN, WA-ANFIS, ANN, ANFIS	WA-ANFIS, WA-ANN	CE, d, R ² , RMSE, MAE, MSRE and MS4E
2021	Tao (Tao et al., 2021)	Nerus River in Malaysia	Hourly	Water level, rainfall	2017–2019	RVM-IGOA, ANN-IGOA, RVM-GOA, ANN-GOA, RVM-PSO, ANN-PSO	RVM-IGOA	KGE, MD, NSE, NRMSE
2020	Pan (Pan et al., 2020)	Yangtze River in China	Daily	Water level	30 years	CNN-GRU, WANN, LSTM, AREMA	CNN-GRU	NSE, MRE, RMSE
2017	Zhong (Zhong, Jiang et al., 2017)	Yangtze River in China	Daily	Water level	2014–2016	ANN, ANN-KF (local), ANN-KF (global)	ANN- KF (local)	RMSE, MAPE, R ²
2020	Mohammadi (Mohammadi et al., 2020)	Titicaca Lake in South American	Monthly	Water level	1973–2017	SVR-GWO, SVR	SVR-GWO	RMSE, MAE, R ²
2021	Xie (Xie et al., 2021)	Yangtze River in China	Monthly	Water level	2010–2016	WA-LSTM, LSTM, WA-ARIMA, WA-ANN	WA-LSTM	RMSE, MAE
2018	Ghorbani (Ghorbani et al., 2017)	Lake Egirdir in Turkey	Monthly	Water level	1961–2016	MLP-FFA, MLP	MLP-FFA	MAE, RMSE, NSE
2021	Linerios (Linerios et al., 2021)	Carrión River in Northwest of the Iberian Peninsula	Every 10 minute	Water level	1999–2010	ANN-MOGA	ANN-MOGA	RMSE
2017	Zhong (Zhong, Guo et al., 2017)	Yangtze River in China	Daily	Water level	2012–2016	ANN-Kalman, ANN	ANN-Kalman	RMSE, R ²
2016	Soleymani (Soleymani et al., 2016)	Selangor River in Malaysia	Daily	Water level	—	RBF-FFA, SVM, MLP	RBF-FFA	RMSE, MAPE, R, R ²

(Continued)

Table 2. (Continued)

year	Reference	Location	Time scale	Variable used	Period	Models used	Best model	Measures of accuracy
2019	Altunkaynak (Altunkaynak, 2019)	Van Lake in Turkey	Monthly	Water level	1960–2017	ASA-MLP, WA-MLP, MP	ASA-MLP	MSE, NSE
2021	Deng (Deng et al., 2021)	Dongting Lake in China	Hourly	Water level	2004–2020	MLP, MLP-GA, ENN-ACO	ENN-ACO	RMSE, MAE, NSE, ARE, R
2018	El-Diasty (El-Diasty et al., 2018)	4 tide gauges along Saudi and USA	Hourly	Water level	2012–2014	HA, WN, HA-WN	HA-WN	RMSE
2020	Wang (Wang et al., 2020)	2 coastal tide station (1) Ship john shoal in New Jersey (2) New canal station in Louisiana	Hourly	Sea water level, harmonic tidal levels, wind speed, wind direction	2015–2019	WA-ANFIS, WA-ANN, ANFIS, ANN, HA	WA-ANFIS	MAE, CE, RMSE, MSRE

Figure 3. A list of the hybrid models that were used in forecasting water levels.



properties better. According to the results, the WA-LSTM model is stable, dependable, and extensively applicable.

Season Algorithm technique can successfully remove the seasonal component from the original time series data. The additive season algorithm (ASA) and the multiplicative season algorithm (MSA) are two decomposition algorithms. The season method decomposes observed time series data into trend-cycle, seasonal index, and irregular (error term) components based on these concepts (Altunkaynak, 2019).

Altunkaynak (Altunkaynak, 2019) suggested increasing prediction accuracy and prolonging water-level lead-time prediction, a new predictive model based on the season algorithm (SA) and multi-layer perceptron (MLP) approaches. The additive season algorithm (ASA) was employed for the first time to estimate water levels as an alternative data pre-processing technique, and its performance was compared to that of the wavelet transform (WT). Based on the mean squared error (MSE) and the Nash-Sutcliffe coefficient of efficiency (NSE) as performance evaluation criteria,

the combined additive season algorithm–multi-layer perceptron model (ASA-MLP) outperformed the WA-MLP and standalone MLP model.

Additionally, other types, such as Kalman filtering (KF) was applied:

Zhong (Zhong, Guo et al., 2017) established a hybrid ANN-Kalman filtering technique to forecast the short-term daily water level of the Yangtze River, China. The performance of the hybrid model was compared with the standalone ANN model. Historical daily water level data from 29 July 2012 to 31 July 2016 were applied to build and assess the prediction models. The findings showed that the hybrid model could accurately simulate the water level data better than the standalone ANN model.

Zhong (Zhong, Jiang et al., 2017) proposed integrating the artificial neural network with the Kalman filtering algorithm (ANN-KF) to predict the short-term water level of the Yangtze River in China. The model was built and assessed by the daily water level data over three years (2014–2016). The KF algorithm confirms the superiority of local Kalman filtering, and the ANN-KF technique predicted the water levels better than the traditional ANN model.

5.2. The components combination-based hybrid models (CBH)

The CBH models use the exceptional capacity of an individual prediction model in various combination structures to improve forecasting accuracy (Hajirahimi & Khashei, 2022).

Gated recurrent unit (GRU) and convolutional neural network (CNN) network structures are combined to create a CNN-GRU model. The GRU part learns the changing water level trend, and the CNN part learns the spatial correlation among water level data observed from adjacent water stations on the Yangtze River in China. The study employed data from multiple locations to predict the water level of the middle location. The CNN-GRU model was compared with three models based solely on GRU and other cutting-edge techniques such as the autoregressive integrated moving average model (ARIMA), wavelet-based artificial neural network (WANN), and long-short term memory model (LSTM). The outcomes showed the CNN-GRU model outperformed the ARIMA, WANN, and LSTM models depending on three assessment factors: Nash-Sutcliffe efficiency coefficient (NSE), average relative error (MRE), and root mean square error (RMSE) (Pan et al., 2020).

Nie (Nie et al., 2021) proposed the CNN-BiLSTM water level forecasting method includes an attention mechanism. The study employed hourly water level and rainfall data for Pinghe basin in China, from 2010 to 2020. CNN extracts spatial characteristics from water level data, and BiLSTM learns time period characteristics by integrating past and future sequence information. An attention mechanism has been implemented to focus on the salient features in the sequence. This method outperforms the support vector machine (SVM), bidirectional long short-term memory network (BiLSTM), and temporal convolutional neural network (TCN) models in terms of accuracy.

5.3. Parameter optimisation-based hybrid models (OBH)

The parameter optimisation-based hybrid models focusing on metaheuristics are solution approaches that coordinate an interplay between local improvement procedures and higher-level strategies to achieve the objective. Develop a procedure that can escape local optima and conduct a rigorous solution space search (Ghalekhondabi et al., 2017).

5.3.1. Particle swarm optimisation (PSO)

Particle swarm optimisation (PSO) is a heuristic approach for solving non-continuous and nonlinear problems. This method tracked the cooperative and social behaviour shown by birds and fish. It is a population-based algorithm that uses particles to optimise a parallel group of swarms. These swarms move at a set speed through the issue area in pursuit of the best-fit solution. The memory of each particle is retained, allowing it to keep its previous optimal position. Particle positions are

categorised into global and local best (Shah et al., 2021). Flow diagram and formulas of the PSO algorithm are found in Shah (Shah et al., 2021).

Panyadee (Panyadee et al., 2017) enhanced the performance of the artificial neural network (ANN) model by combining it with the particle swarm optimisation (PSO) algorithm to predict the water level of the Mea-Bong River, Thailand. The PSO algorithm proposes in this research is to fine-tune the hyperparameters of the ANN model. The evaluation results reveal that the PSO algorithm provides the best hyperparameters values that lead to shortening the time it takes to train an ANN. Also, the outcomes present that the forecast error is 1.88 percent for the training stage, whereas the error is 7.82 percent for the testing stage. The simulated water level from the suggested hybrid model is suitable for use in early warning systems for flash floods.

5.3.2. The firefly algorithm (FFA)

The firefly algorithm (FFA) is a swarm intelligence optimisation technique inspired by fire flies' movement. An optimisation problem's answer can be modelled as a firefly that lights in proportion to its quality. As a result, each brighter firefly attracts its mates, regardless of gender, making the search for space exploration more efficient. Fire flies are drawn to the light. The entire swarm makes a beeline for the most brilliant firefly. As a result, the firefly's appeal is related to their brightness. Furthermore, the brightness is determined by the agent's intensity (Ghorbani et al., 2018; Tripura et al., 2020). FFA algorithm formulas are mentioned in Yang (Yang, 2010).

Soleymani (Soleymani et al., 2016) designed a novel technique that combined radial basis function (RBF) and firefly algorithm (FFA) to forecast the water level of Selangor River, Malaysia. The FFA algorithm is employed to interpolate the RBF model to estimate the best solution. The RBF-FFA technique was compared with support vector machine (SVM) and multilayer perceptron (MLP) models to increase its validation. The results reveal that the proposed RBF-FFA model delivers more accurate predictions than other machine learning models, the SVM and MLP, based on several statistical indicators. The outcomes also specify that the established RBF-FFA model can be utilised as an efficient technique for the precise forecast of the water stage of the river.

Kisi (Kisi et al., 2015) evaluated a new methodology that coupled the support vector machine (SVM) with the firefly algorithm (FFA) to predict daily water-level data for Urmia Lake, Iran. The FFA was applied to estimate the optimum SVM hyperparameters. The SVM-FFA model was validated by comparing genetic programming (GP) and artificial neural networks (ANNs). According to the findings, the SVM-FFA technique produced better predictions with greater generalisability in 1 day ahead of lake level forecasts than GP and ANN models.

5.3.3. Ant colony optimisation (ACO)

Ant colony optimisation (ACO) is a probabilistic strategy that uses graph minimisation to locate acceptable paths to solve numerical problems. Artificial ants are multi-agent tactics modelled after real-life ant behaviour. The pheromone-based contact of biological ants is always the most popular model. An ant colony optimisation algorithm is a type of optimisation algorithm that focuses on the behaviour of ants. The artificial' ants' identify optimal solutions by iterating over a parameter space that describes all possible solutions. When true ants are out investigating their environment, they leave behind pheromones that help them find food. The simulated 'ants' keep track of their positions and the consistency of their solutions, such that in succeeding simulation iterations, more ants find better answers (Adnan, Mostafa, Elbeltagi et al., 2021). Kucukkoc and Zhang (Kucukkoc & Zhang, 2013) include a flow diagram and the ACO algorithm's formulas.

Deng (Deng et al., 2021) examined the combined Elman neural network (ENN) with ant colony algorithm (ACO) to forecast water level in Dongting Lake, China. To ensure that the ENN-ACO model captures the nonlinear pattern of independent and dependent factors, the results of the ACO-ENN model validate with a hybrid multi-layer perceptron and genetic algorithm (MLP-GA) model and standalone MLP model. The study used hourly data of water levels from 2004 to 2020.

The outcomes gained display that the ENN-ACO model can forecast the hourly water level with an error of 1.2%. It can offer an advanced water level forecast of 21 h ahead of the time step. Generally, the results indicate that the ACO-ENN model's accuracy is better than the GA-MLP and the standalone MLP models.

5.3.4. *The genetic algorithm (GA)*

The genetic algorithm (GA) is the most prevalent algorithm, which is a population-based optimisation algorithm that mimics natural evolution (natural genetics and natural selection). GA's fundamental theories are founded on Darwinian principles of mutation, operator selection, and subsequent recombination (crossover) (Ibrahim et al., 2022). In a nutshell, the method begins by producing a random population of chromosomes representing potential solutions to a given problem. Then, for each chromosome, determine the fitness function that defines the selection stage's probability. By uniting two separate chromosomes, the crossover operation is conducted on a pair of selected chromosomes to produce a new better offspring. As a result, chromosomal components (genes) at randomly selected chromosome sites are altered. Mutation is the word for this final genetic alteration. The next population to be examined is the offspring produced through genetic alteration procedures (Mulia et al., 2013). Flow diagram of GA algorithm is found in Ibrahim et al., 2022).

Chen (N. Chen et al., 2019) explored a genetic algorithm linking a back-propagation neural network (GA-BPNN) to predict water level. However, a conventional genetic algorithm is susceptible to local optimisation and local convergence when confronted with a complicated neural network. To solve this issue, a new technique termed an improved genetic algorithm (IGA) coupled with a back-propagation neural network model (IGA-BPNN) is presented using several genetic methods. Weather and hydrologic data collected between 2010 and 2017 were applied to estimate the Han River's water levels. The Pearson correlation coefficient (R), Nash-Sutcliffe efficiency (NSE) coefficient and root mean square error (RMSE) were used to assess the prediction models. The outcomes displayed that IGA-BPNN outperformed the GA-BPNN and BPNN models depending on three statistical indicators. The IGA-BPNN technique revealed suitability for water-level estimates and would offer a better effect on short-term flood anticipating.

Imran (Imran et al., 2021) inspected the performance of an adaptive neuro-fuzzy inference system (ANFIS) model that coupled with a genetic algorithm to forecast the water levels of the Jhelum River, India. Two meteorological stations' temperature and precipitation data were employed to train the model to simulate the river's water level. Several ANFIS methods with various membership functions and optimisation approaches are examined, and the best-performing technique is selected for additional modification. The parameters of chosen ANFIS technique are then optimised via the genetic algorithm to gain better outcomes. The ANFIS-GA technique exhibited better results as compared to the traditional ANFIS model.

Linerós (Linerós et al., 2021) studied the effect of using a multi-objective genetic algorithm (MOGA) framework for the design of an artificial neural network (ANN) technique that is applied for designing 1-step-ahead river water level prediction models. A design process is a semi-automatic approach that can split data into datasets and find a near-optimal model with the proper topology and inputs, performing well on unseen data (data not utilised for model design). The study used water level data every 10 minutes over eleven years of the Carrión River, Northwest of the Iberian Peninsula. The results demonstrate that the proposed framework can produce low-complex models with high performance on unseen data, obtaining an RMSE of 2.5 10⁻³, which compares favourably to outcomes produced by alternative methods.

5.3.5. *Grey wolf optimisation (GWO)*

Grey wolf optimisation (GWO) was created as a result of natural inspiration. Grey wolves' hunting habits and social leadership structure serve as inspiration for the core concept. The GWO has a number of advantages when it comes to dealing with problems. Simplistic, flexible, and avoiding local optima

are all advantages of nonlinear and multivariable functions (Adnan, Mostafa, Kisi et al., 2021). Based on the effectiveness and decision-making power of the group, each member of a grey wolf herd is classified as α , β , δ and ω . The alpha wolf is normally the strongest and most dominating wolf in the herd, and the rest of the pack should obey his or her orders. The β wolves are the second group of alphas that serve as counsellors. The beta wolves help the alpha wolf by reiterating their directives to the remainder of the pack. In terms of leadership structure, the δ wolves are below the α and β wolves and above the ω wolves. They serve as the group's guardians, sentinels, hunters, and caregivers. The ω wolves are ranked last in the decision-making hierarchy and must subordinate to all other wolves (Mohammadi et al., 2020). GWO algorithm formulae can be found in Mohammadi et al. (2020).

Other algorithms for forecasting water level have also been combined with machine learning, such as the gravitational search algorithm (GSA), sunflower optimisation (SO) algorithm, and grasshopper optimisation algorithm (GOA).

Ghorbani (Ghorbani et al., 2018) designed a new hybrid forecasting model that combines the gravitational search algorithm (GSA) with the multi-layer perceptron (MLP) method to estimate water levels in two lakes. The MLP was coupled with an additional two meta-heuristic optimisations, namely particle swarm optimisation (PSO) and firefly algorithm (FFA), to increase the validation of the suggested model. Also, the study used two stochastic models: ARMA and ARIMA. To train and assess the MLP-GSA model, monthly water level data from 1938 to 2005 and 1942 to 2011 for Lakes Winnepesaukee and Cypress were used. The results highlighted the significant efficacy of the MLP-GSA, which outperformed the other hybrid (MLP-PSO and MLP-FFA) and standalone (MLP, ARMA, and ARIMA) models.

Ehteram (Ehteram et al., 2021) enhanced the capacity of the adaptive neuro-fuzzy inference system (ANFIS) and multi-layer perceptron (MLP) models utilising a sunflower optimisation (SO) algorithm to predict the lake water level. Also, the hybrid models' performance was validated with the firefly algorithm (FFA) and practical swarm optimisation (PSO) algorithms. The optimisation algorithms were applied to determine the optimum tuning hyperparameters for ANFIS and MLP models. The rainfall, temperature, and water level lags data were used to predict the water level of Urmia Lake, Iran, from 1940 to 2004. The ANFIS-SO model was found to have a lower level of uncertainty depending on the percentage of more responses in the confidence band and the smaller bandwidth of the model.

5.4. Hybridisation of parameter optimisation-based with preprocessing-based hybrid models (HOPH)

The metaheuristic algorithms integrated with PBH models to choose optimal parameters of a forecasting model, pre-processing technique, or investigate the optimised weights to aggregate decomposition component predictions (Hajirahimi & Khashei, 2022).

Mohammadi (Mohammadi et al., 2020) inspected the accuracy of coupled support vector regression (SVR) and the grey wolf algorithm (GWO) to forecast fluctuations in lake water level. Additionally, three pre-treatment methods, i.e., random forest, relief algorithm, and principal component analysis, were applied to select the best scenario of predictors. Monthly datasets of Titicaca Lake in South America, from August 1973 to January 2017 were used to build and evaluate the model. The results show that the random forest method offers the best model input scenario with four lags. The hybrid model SVR-GWO simulates water level better than standalone SVR based on several statistical criteria.

Tao (Tao et al., 2021) used improved grasshopper optimisation (IGOA) algorithm to integrate both relevance vector machine (RVM) and artificial neural network (ANN) models to forecast the catchment water level in Malaysia. The classical GOA and particle swarm optimisation (PSO) algorithms were employed to validate the performance of the IGOA algorithm. The hourly rainfall and water level data lags from 1 January 2017 to 31 December 2019 adopted to build and validate

the prediction models. Considering different statistical criteria, the IGOA algorithm significantly improved the models' performance, and the RVM-IGOA was superior.

Ghorbani (Ghorbani et al., 2017) investigated the predictive capability of a combined model integrating the multi-layer perceptron (MLP) with the firefly algorithm (FFA) to forecast water level in Lake Egirdir, Turkey. Monthly data for 56 years (1961–2016) were used to train and test the suggested hybrid MLP-FFA model to develop and explore its veracity. Four lagged combinations of historical data were adopted as predictors using the average mutual information technique. The outcomes show that the MLP-FFA model outperforms the standalone MLP model based on different statistical score metrics, e.g., root mean square error was about 0.029 m for the MLP-FFA model and compared to 0.102 m for the standalone MLP model.

Following is a list of observations obtained from a survey of many articles:

- (1) Generally, almost the previous studies applied pre-treatment signal only and did not focus on the other steps of data pre-processing techniques.
- (2) The wavelet algorithm was shown to be effective in denoising raw data, improving the results' accuracy.
- (3) Different metaheuristic algorithms were used to integrate machine learning models. These algorithms have proved their capacity to tune all machine learning models and earn a significantly higher score on several statistical evaluations. In addition, when compared to a trial and error procedure, the chances of achieving optimal parameters are substantially better.
- (4) Few studies have employed the hybridisation of parameter optimisation based with pre-processing-based hybrid models (HOPH).

6. Future perspectives

Tao (Tao et al., 2021) suggested assessing the performance of other nonlinear input choice approaches, such as integrating REF with other ML techniques like ANN and RF. Pan (Pan et al., 2020) recommended employing cluster approaches to analyse the water levels of various river segments for clustering and extracting significant environmental parameters. Ghorbani (Ghorbani et al., 2018) advised using the hybrid MLP-GSA model for short-term forecasting of other hydrological variables (e.g., lake water levels, rainfall, evaporation, daily discharge, drought and flood indices) because of the high accuracy of the developed hybrid model over the applied standalone and hybrid methods. Seo (Seo et al., 2015) suggested developing hybrid techniques joining the wavelet decomposition method with other machine learning models and metaheuristic algorithms for predicting hydrological factors with non-stationary and nonlinear relationships. A comparative research study on various learning algorithms such as Levenberg-Marquardt and conjugate gradient algorithms can also be proposed to enhance hybrid techniques' performance. Further research studies can also be recommended to explore the model performance for various input series constructed from effective or all wavelet components. Farzad and El-Shafie (Farzad & El-Shafie, 2016) recommended that a variety of ANN techniques should be examined to identify the best fit for each cluster of data points.

Also, the literature review allows us to summarise some future perspectives:

- (1) It is suggested to apply additional approaches for pre-treatment data, such as empirical mode decomposition (EMD) and singular spectrum analysis (SSA).
- (2) The choice of input variables is crucial in determining the model's performance and accuracy. With this in mind, it is suggested that more effort be made into determining the ideal input variable combination; therefore, it is recommended to apply other methods to determine the inputs, such as dimensionality reduction methods, feature extraction, and feature selection methods.

- (3) It is recommended that the data pre-processing steps be completed to avoid outliers and noise and to determine the most dependable and accurate data to be utilised as predictors later.
- (4) Implementing the combined metaheuristic algorithms and machine learning techniques in WL forecasting has grown considerably in recent years. Nevertheless, there is still room for improvement regarding the WL forecast.

7. Conclusion

The use of machine learning to forecast water levels has expanded, as has the research and development of this technology. Optimisation models have long been used to aid decision-making. However, certain situations are complex, poorly understood and nonlinear. They have a large number of possible solutions, particularly when it comes to ecological modelling. This shows that, depending on the problems to be solved, this strategy can be used in a variety of contexts.

This paper reviews the recent water-level forecasting works for recent years, where the data pre-processing and hybrid models have been studied in detail, and reviewed a number of the machine learning models, as it turned out that the majority of the models, where it was observed that most machine learning models, such as ANN, ANFIS, and SVR deal with nonlinear data series. The papers chosen for this review demonstrated that there has been a rising trend in recent years toward using hybrid methods in the field of WL modelling. The results reveal that no single global model or technique consistently outperforms all other methods. Comparing the hybrid models with standalone models shows that hybrid models produce better results than standalone models. Also, metaheuristic has improved the single models by selecting the best hyperparameter for the model, saving time and avoiding the lengthy and imprecise trial and error method. In addition, data pre-treatment methods effectively enhance data quality before feeding it into the model by removing noise from time series and choosing the suitable model input scenario. The majority of previous studies used one or two data pre-processing stages. While several of the research did not use all steps of pre-processing data procedures. Therefore, the fact that the current study offers a thorough analysis of all the aforementioned criteria is a key strength of the study.

No new strategies such as the hybrid model or any other have emerged as the best prediction method despite recent significant breakthroughs in machine learning models. Therefore, academics still have an opportunity to create hybrid approaches for particular applications for water level prediction as a research challenge.

In the future, it is recommended to use data pre-processing approaches for their high ability to improve data accuracy, employ updated metaheuristics to simulate water levels in rivers and lakes, and utilise the hybrid model (HOPH) because it optimises both the data and the model.

Funding

This work was supported by the nadhir.alansari@ltu.se [184940].

Author details

Sarah J. Mohammed¹

Salah L. Zubaidi^{1,2}

Sandra Ortega-Martorell³

Nadhir Al-Ansari⁴

E-mail: nadhir.alansari@ltu.se

ORCID ID: <http://orcid.org/0000-0002-6790-2653>

Saleem Ethaib⁵

Khalid Hashim^{6,7}

¹ Department of Civil Engineering, Wasit University, Wasit, Iraq.

² College of Engineering, University of Warith Al-Anbiyaa, Karbala, Iraq.

³ Department of Applied Mathematics, Liverpool John Moores University, Liverpool, UK.

⁴ Department of Civil Environmental and Natural Resources Engineering, Lulea University of Technology, Lulea, Sweden.

⁵ Department of Civil Engineering, University of Thi-Qar, Al-Nassiriya, Iraq.

⁶ School of Civil Engineering and Built Environment, Faculty of Engineering and Technology, Liverpool John Moores University, Liverpool, UK.

⁷ Department of Environment Engineering, Babylon University, Babylon 51001, Iraq.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Abbreviations

ANFIS	Adaptive neuro -fuzzy inference system	[1]
ACO	Ant colony optimisation	[2]
ANN	Artificial neural network	[3]
ARIMA	Autoregressive moving average	[4]
ASA	Additive season algorithm	[5]
ARE	Average percentage relative error	[2]
BA	Bat algorithm	[6]
BVC	Bayesian vine copula	[7]
BPNN	Back-propagation neural network	[8]
BNN	Bayesian neural network	[9]
BILSTM	Bidirectional long short-term memory network	[10]
BMA	Bayesian model averaging	[7]
CBH	The components combination-based hybrid models	[11]
CNN	Convolutional neural network	[12]
CWT	Continuous wavelet transform	[13]
COE	Coefficient of efficiency	[14]
COC	Coefficient of correlation	[14]
CS	Cuckoo search	[15]
CE	Coefficient of efficiency	[13]
CART	Classification and registration tree	[16]
CC	Correlation coefficient	[16]
DWT	Discrete wavelet transform	[13]
d	Index of agreement	[17]
DE	Differential evolution	[16]
ENN	Elman neural network	[2]
EMD	Empirical mode decomposition	[18]
EFF	Efficiency coefficient	[19]
FFNN	Feed-forward neural network	[20]
FFA	Firefly algorithm	[21]
FB	Fractional bias	[19]
GWO	Grey wolf optimiser	[22]
GEP	Gene expression programming	[23]
GA	Genetic algorithm	[8]
GRU	Gated recurrent unit	[24]
GOA	Grasshopper optimisation	[25]
GSA	Gravitational search algorithm	[14]

(Continued)

ANFIS	Adaptive neuro -fuzzy inference system	[1]
HOPH	Hybridisation of parameter optimisation-based with preprocessing-based hybrid models	[11]
HA	Harmonic analysis	[26]
IGA	Improved genetic algorithm	[8]
IA	Index of agreement	[19]
IGOA	Improved grasshopper optimisation	[25]
KF	Kalman filtering algorithm	[27]
KGE	Kling Gupta efficient	[25]
LSTM	Long short – term memory	[28]
MAE	Mean absolute error	[8]
MLR	Multiple linear regression	[29]
MAPE	Mean absolute percentage error	[30]
MAAPE	Mean arctangent absolute percentage error	[31]
MSE	Mean squared error	[31]
MRE	Average relative error	[24]
MSRE	Mean squared relative error	[32]
MARE	Mean absolute relative error	[8]
MLP	Multilayer perceptron	[33]
md	Modified index of agreement	[25]
MOGA	Multi-objective genetic algorithm	[34]
MI	Mutual information	[35]
NSE	Nash–Sutcliffe coefficient	[36]
NMSE	Normalised mean square error	[19]
NRMSE	Normalised root means square error	[25]
OBH	Algorithm parameter optimisation-based hybrid models	[11]
PBH	Pre-processing-based hybrid models	[11]
PSO	Particle swarm optimisation	[30]
PEs	Processing elements	[37]
R ²	Coefficient of determination	[7]
R	Correlation coefficient	[8]
RI	Stability index	[36]
RMSE	Root mean squared error	[13]
RRMSE	Relative root means square error	[14]

(Continued)

(Continued)		
ANFIS	Adaptive neuro -fuzzy inference system	[1]
RNN	Recurrent neural network	[38]
RBFNN	Radial basis function neural network model	[19]
RVM	Relevance vector machine	[25]
RBF	Radial basis function neural network	[39]
REF	Recursive filter elimination	[25]
RF	Random forest	[16]
SVM	Support vector machine	[40]
SSA	Singular spectrum analysis	[41]
SICM	Supervised intelligence committee machine method	[19]
SVR	Support vector regression	[42]
SMA	Symmetric-moving-average	[43]
SO	Sunflower optimisation	[36]
TI	Time lag	[16]
TCN	Temporal convolutional neural network	[10]
VC	Vine copula	[7]
WA	Wavelet transform	[44]
WL	Water level	[9]
WN	Wavelet network	[26]
XGBOOST	Extreme gradient boosting	[16]

Citation information

Cite this article as: Application of hybrid machine learning models and data pre-processing to predict water level of watersheds: Recent trends and future perspective, Sarah J. Mohammed, Salah L. Zubaidi, Sandra Ortega-Martorell, Nadhir Al-Ansari, Saleem Ethaib & Khalid Hashim, *Cogent Engineering* (2022), 9: 2143051.

References

- Abba, S. I., Hadi, S. J., & Abdullahi, J. (2017). River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. *Procedia Computer Science*, 120, 75–82. <https://doi.org/10.1016/j.procs.2017.11.212>
- Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B., & Sliusarieva, A. (2012). Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research*, 48(1). <https://doi.org/10.1029/2010wr009945>
- Adnan, R. M., Mostafa, R. R., Elbeltagi, A., Yaseen, Z. M., Shahid, S., & Kisi, O. (2021). Development of new machine learning model for streamflow prediction: Case studies in Pakistan. *Stochastic Environmental Research and Risk Assessment*, 36(4), 999–1033. <https://doi.org/10.1007/s00477-021-02111-z>
- Adnan, R. M., Mostafa, R. R., Kisi, O., Yaseen, Z. M., Shahid, S., & Zounemat-Kermani, M. (2021). Improving streamflow prediction using a new hybrid ELM model combined with hybrid particle swarm optimization and grey wolf optimization. *Knowledge-Based Systems*, 230, 107379. <https://doi.org/10.1016/j.knosys.2021.107379>
- Ahmadi, F., Mehdizadeh, S., & Mohammadi, B. (2021). Development of bio-inspired- and wavelet-based hybrid models for reconnaissance drought index modeling. *Water Resources Management*, 35(12), 4127–4147. <https://doi.org/10.1007/s11269-021-02934-z>
- Alawsji, M. A., Zubaidi, S. L., Al-Bdairi, N. S. S., Al-Ansari, N., & Hashim, K. (2022). Drought forecasting: A review and assessment of the hybrid techniques and data pre-processing. *Hydrology*, 9(7), 115. <https://doi.org/10.3390/hydrology9070115>
- Ali Ghorbani, M., Khatibi, R., Aytak, A., Makarynskyy, O., & Shiri, J. (2010). Sea water level forecasting using genetic programming and comparing the performance with artificial neural networks. *Computers & Geosciences*, 36(5), 620–627. <https://doi.org/10.1016/j.cageo.2009.09.014>

- Al-Mukhtar, M., & Al-Yaseen, F. (2019). Modeling water quality parameters using data-driven models, a case study Abu-Ziriq marsh in South of Iraq. *Hydrology*, 6(1), 24. <https://doi.org/10.3390/hydrology6010024>
- Altunkaynak, A. (2019). Predicting water level fluctuations in Lake Van using hybrid season-neuro approach. *Journal of Hydrologic Engineering*, 24(8). [https://doi.org/10.1061/\(asce\)he.1943-5584.0001804](https://doi.org/10.1061/(asce)he.1943-5584.0001804)
- Altunkaynak, A., & Kartal, E. (2019). Performance comparison of continuous Wavelet-Fuzzy and discrete Wavelet-Fuzzy models for water level predictions at northern and southern boundary of Bosphorus. *Ocean Engineering*, 186. <https://doi.org/10.1016/j.oceaneng.2019.06.002>
- Al Visi, S., Mascellani, G., Franchini, M., & Bardossy, A. (2006). Water level forecasting through fuzzy logic and artificial neural network approaches. *Hydrology and Earth System Sciences*, 10(1), 1–17. <https://doi.org/10.5194/hess-10-1-2006>
- Apaydin, H., Sattari, M. T., Falsafian, K., & Prasad, R. (2021). Artificial intelligence modelling integrated with singular spectral analysis and seasonal-trend decomposition using loess approaches for stream-flow predictions. *Journal of Hydrology*, 600, 126506. <https://doi.org/10.1016/j.jhydrol.2021.126506>
- Bazartseren, B., Hildebrandt, G., & Holz, K. P. (2003). Short-term water level prediction using neural networks and neuro-fuzzy approach. *Neurocomputing*, 55(3–4), 439–450. [https://doi.org/10.1016/s0925-2312\(03\)00388-6](https://doi.org/10.1016/s0925-2312(03)00388-6)
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen, X., Song, W., Shi, Y., Liu, W., Lu, Y., Pang, Z., & Chen, X. (2022). Application of Cosmic-ray neutron sensor method to calculate field water use efficiency. *Water*, 14. <https://doi.org/10.3390/w14091518>
- Chen, N., Xiong, C., Du, W., Wang, C., Lin, X., & Chen, Z. (2019). An improved genetic algorithm coupling a back-propagation neural network model (IGA-BPNN) for water-level predictions. *Water*, 11(9), 1795. <https://doi.org/10.3390/w11091795>
- Çimen, M., & Kisi, O. (2009). Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey. *Journal of Hydrology*, 378(3–4), 253–262. <https://doi.org/10.1016/j.jhydrol.2009.09.029>
- Coulibaly, P. (2010). Reservoir computing approach to great lakes water level forecasting. *Journal of Hydrology*, 381(1–2), 76–88. <https://doi.org/10.1016/j.jhydrol.2009.11.027>
- Deng, B., Lai, S. H., Jaing, C., Kumar, P., El-Shafie, A., & Chin, R. J. (2021). Advanced water level prediction for a large-scale river-lake system using hybrid soft computing approach: A case study in Dongting Lake, China. *Earth Science Informatics*, 14(4), 1987–2001. <https://doi.org/10.1007/s12145-021-00665-8>
- de Souza Groppo, G., Costa, M. A., & de Souza Groppo, G. (2019). Predicting water demand: A review of the methods employed and future possibilities. *Water Supply*, 19(8), 2179–2198. <https://doi.org/10.2166/ws.2019.122>
- Dikshit, A., Pradhan, B., & Alamri, A. M. (2020). Temporal hydrological drought index forecasting for new south wales, Australia using machine learning approaches. *Atmosphere*, 11(6), 585. <https://doi.org/10.3390/atmos11060585>
- Dimitriadis, P., & Koutsoyiannis, D. (2018). Stochastic synthesis approximating any process dependence and distribution. *Stochastic Environmental Research and Risk Assessment*, 32(6), 1493–1515. <https://doi.org/10.1007/s00477-018-1540-2>
- Dimitriadis, P., Koutsoyiannis, D., Iliopoulou, T., & Papanicolaou, P. (2021). A global-scale investigation of stochastic similarities in marginal distribution and dependence structure of key hydrological-cycle processes. *hydrology*, 8(2), 59. <https://doi.org/10.3390/hydrology8020059>
- Ehteram, M., Ferdowsi, A., Faramarzpour, M., Al-Janabi, A. M. S., Al-Ansari, N., Bokde, N. D., & Yaseen, Z. M. (2021). Hybridization of artificial intelligence models with nature inspired optimization algorithms for lake water level prediction and uncertainty analysis. *Alexandria Engineering Journal*, 60(2), 2193–2208. <https://doi.org/10.1016/j.aej.2020.12.034>
- El-Diasty, M., & Al-Harbi, S. (2015). Development of wavelet network model for accurate water levels prediction with meteorological effects. *Applied Ocean Research*, 53, 228–235. <https://doi.org/10.1016/j.apor.2015.09.008>
- El-Diasty, M., Al-Harbi, S., & Pagiatakis, S. (2018). Hybrid harmonic analysis and wavelet network model for sea water level prediction. *Applied Ocean Research*, 70, 14–21. <https://doi.org/10.1016/j.apor.2017.11.007>
- Ethaib, S., Zubaidi, S. L., Al-Ansari, N., & Fegade, S. L. (2022). Evaluation water scarcity based on GIS estimation and climate-change effects: A case study of Thi-Qar governorate, Iraq. *Cogent Engineering*, 9(1). <https://doi.org/10.1080/23311916.2022.2075301>
- Farzad, F., & El-Shafie, A. H. (2016). Performance enhancement of rainfall pattern – water level prediction model utilizing self-organizing-map clustering method. *Water Resources Management*, 31(3), 945–959. <https://doi.org/10.1007/s11269-016-1556-7>
- Fung, K. F., Huang, Y. F., Koo, C. H., & Soh, Y. W. (2019). Drought forecasting: A review of modelling approaches 2007–2017. *Journal of Water and Climate Change*, 11(3), 771–799. <https://doi.org/10.2166/wcc.2019.236>
- Ghalekhondabi, I., Ardjmand, E., Young, W. A., & Weckman, G. R. (2017). Water demand forecasting: Review of soft computing methods. *Environmental Monitoring and Assessment*, 189(7). <https://doi.org/10.1007/s10661-017-6030-3>
- Ghorbani, M. A., Deo, R. C., Karimi, V., Kashani, M. H., & Ghorbani, S. (2018). Design and implementation of a hybrid MLP-GSA model with multi-layer perceptron-gravitational search algorithm for monthly lake water level forecasting. *Stochastic Environmental Research and Risk Assessment*, 33(1), 125–147. <https://doi.org/10.1007/s00477-018-1630-1>
- Ghorbani, M. A., Deo, R. C., Karimi, V., Yaseen, Z. M., & Terzi, O. (2017). Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey. *Stochastic Environmental Research and Risk Assessment*, 32(6), 1683–1697. <https://doi.org/10.1007/s00477-017-1474-0>
- Graves, T., Gramacy, R., Watkins, N., & Franzke, C. (2017). A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951–1980. *Entropy*, 19(9), 437. <https://doi.org/10.3390/e19090437>
- Gunathilake, M. B., Karunanayake, C., Gunathilake, A. S., Marasingha, N., Samarasinghe, J. T., Bandara, I. M., & Rathnayake, U. (2021). Hydrological models and artificial neural networks (ANNs) to simulate streamflow in a tropical catchment of Sri Lanka. *Applied Computational Intelligence and Soft Computing*, 9. <https://doi.org/10.1155/2021/6683389>
- Hajirahimi, Z., & Khashei, M. (2022). Hybridization of hybrid structures for time series forecasting: A review. *Artificial Intelligence Review*, 10.1007/s10462-022-10199-0. <https://doi.org/10.1007/s10462-022-10199-0>

- Herrera, M., Torgo, L., Izquierdo, J., & Pérez-García, R. (2010). Predictive models for forecasting hourly urban water demand. *Journal of Hydrology*, 387(1–2), 141–150. <https://doi.org/10.1016/j.jhydrol.2010.04.005>
- Hrnjica, B., & Bonacci, O. (2019). Lake level prediction using feed forward and recurrent neural networks. *Water Resources Management*, 33(7), 2471–2484. <https://doi.org/10.1007/s11269-019-02255-2>
- Huang, X., Li, Y., Tian, Z., Ye, Q., Ke, Q., Fan, D., Mao, G., Chen, A., & Liu, J. (2021). Evaluation of short-term streamflow prediction methods in Urban river basins. *Physics and Chemistry of the Earth*, 123, 103027. <https://doi.org/10.1016/j.pce.2021.103027>
- Hussaini, A., Mahmud, M. R., Kang Wee, K. T., & A.g, A. (2020). A review of water level fluctuation models and modelling initiatives. *Journal of Computational and Theoretical Nanoscience*, 17(2), 645–653. <https://doi.org/10.1166/jctn.2020.8781>
- Ibrahim, K. S. M. H., Huang, Y. F., Ahmed, A. N., Koo, C. H., & El-Shafie, A. (2022). A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. *Alexandria Engineering Journal*, 61(1), 279–303. <https://doi.org/10.1016/j.aej.2021.04.100>
- Imran, M., Sheikh Abdul Khader, P., Rafiq, M., & Singh Rawat, K. (2021). Forecasting water level of glacial fed perennial river using a genetically optimized hybrid machine learning model. *Materials Today: Proceedings*, 46, 11113–11119. <https://doi.org/10.1016/j.matpr.2021.02.256>
- Karaboga, D., & Kaya, E. (2020). Estimation of number of foreign visitors with ANFIS by using ABC algorithm. *Soft Computing*, 24(10), 7579–7591. <https://doi.org/10.1007/s00500-019-04386-5>
- Khan, M. Y. A., Hasan, F., Panwar, S., & Chakrapani, G. J. (2016). Neural network model for discharge and water-level prediction for Ramganga River catchment of Ganga Basin, India. *Hydrological Sciences Journal*, 61(11), 2084–2095. <https://doi.org/10.1080/02626667.2015.1083650>
- Khan, M. M. H., Muhammad, N. S., & El-Shafie, A. (2020). Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. *Journal of Hydrology*, 590, 125380. <https://doi.org/10.1016/j.jhydrol.2020.125380>
- Khudhair, Z. S., Zubaidi, S. L., Ortega-Martorell, S., Al-Ansari, N., Ethaib, S., & Hashim, K. (2022). A review of hybrid soft computing and data pre-processing techniques to forecast freshwater quality's parameters: current trends and future directions. *Environments*, 9(7), 85. <https://doi.org/10.3390/environments9070085>
- Kilinc, H. C. (2022). Daily streamflow forecasting based on the hybrid particle swarm optimization and long short-term memory model in the Orontes Basin. *Water*, 14(3), 490. <https://doi.org/10.3390/w14030490>
- Kilinc, H. C., & Haznedar, B. (2022). A hybrid model for streamflow forecasting in the basin of Euphrates. *Water*, 14(1), 80. <https://doi.org/10.3390/w14010080>
- Kim, D., Lee, J., Kim, J., Lee, M., Wang, W., & Kim, H. S. (2022). Comparative analysis of long short-term memory and storage function model for flood water level forecasting of Bokha stream in NamHan River, Korea. *Journal of Hydrology*, 606, 127415. <https://doi.org/10.1016/j.jhydrol.2021.127415>
- Kisi, O., Shiri, J., Karimi, S., Shamshirband, S., Motamedi, S., Petković, D., & Hashim, R. (2015). A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm. *Applied Mathematics and Computation*, 270, 731–743. <https://doi.org/10.1016/j.amc.2015.08.085>
- Kisi, O., Shiri, J., & Nikoofar, B. (2012). Forecasting daily lake levels using artificial intelligence approaches. *Computers & Geosciences*, 41, 169–180. <https://doi.org/10.1016/j.cageo.2011.08.027>
- Koutsoyiannis, D. (2020). Revisiting the global hydrological cycle: Is it intensifying?. *Hydrology and Earth System Sciences*, 24(8), 3899–3932. <https://doi.org/10.5194/hess-24-3899-2020>
- Koutsoyiannis, D., & Montanari, A. (2022). Bluecat: A local uncertainty estimator for deterministic simulations and predictions. *Water Resources Research*, 58(1). <https://doi.org/10.1029/2021wr031215>
- Kucukkoc, I. & Zhang, D. Z. (2013). On applications of ant colony optimisation techniques in solving assembly line balancing Problem. In *Proceedings of OR55 Annual Conference - Keynote Papers and Extended abstracts*, University of Exeter; pp. 45–51.
- Leahy, P., Kiely, G., & Corcoran, G. (2008). Structural optimisation and input selection of an artificial neural network for river level prediction. *Journal of Hydrology*, 355(1–4), 192–201. <https://doi.org/10.1016/j.jhydrol.2008.03.017>
- Li, F. F., Cao, H., Hao, C. F., & Qiu, J. (2021). Daily streamflow forecasting based on flow pattern recognition. *Water Resources Management*, 35(13), 4601–4620. <https://doi.org/10.1007/s11269-021-02971-8>
- Li, W., Fang, H., Qin, G., Tan, X., Huang, Z., Zeng, F., Du, H., & Li, S. (2020). Concentration estimation of dissolved oxygen in Pearl River basin using input variable selection and machine learning techniques. *The Science of the Total Environment*, 731, 139099. <https://doi.org/10.1016/j.scitotenv.2020.139099>
- Linerós, M. L., Luna, A. M., Ferreira, P. M., & Ruano, A. E. (2021). Optimized design of neural networks for a river water level prediction system. *Sensors*, 21(19), 6504. <https://doi.org/10.3390/s21196504>
- Liu, Z., Cheng, L., Lin, K., & Cai, H. (2021). A hybrid Bayesian vine model for water level prediction. *Environmental Modelling and Software*, 142, 105075. <https://doi.org/10.1016/j.envsoft.2021.105075>
- Li, B., Yang, G., Wan, R., Yang, G., & Li, B. (2016). Comparison of random forests and other statistical methods for the prediction of lake water level: A case study of the Poyang Lake in China. *Hydrology Research*, 47(S1), 69–83. <https://doi.org/10.2166/nh.2016.264>
- Loh, E. C., Ismail, S. B., & Khamis, A. (2019). Empirical mode decomposition couple with artificial neural network for water level prediction. *Civil Engineering and Architecture*, 7(6A), 19–32. <https://doi.org/10.13189/cea.2019.071403>
- Loh, E. C., Ismail, S. B., Mustafa, A., & Shaharudin, S. M. (2019). Application of discrete wavelet transformation with artificial neural network in kelantan water level prediction. *Journal of Adv Research in Dynamical & Control Systems*, 11.
- Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environmental Modelling and Software*, 15(1), 101–124. [https://doi.org/10.1016/S1364-8152\(99\)00007-9](https://doi.org/10.1016/S1364-8152(99)00007-9)
- Malekpour, M. M., & Mohammad Rezapour Tabari, M. (2020). Implementation of supervised intelligence committee machine method for monthly water level prediction. *Arabian Journal of Geosciences*, 13(19). <https://doi.org/10.1007/s12517-020-06034-x>
- Mohamadi, S., Sammen, S. S., Panahi, F., Ehteram, M., Kisi, O., Mosavi, A., Ahmed, A. N., El-Shafie, A., & Al-Ansari, N. (2020). Zoning map for drought prediction using integrated machine learning models with a nomadic people optimization algorithm. *Natural*

- Hazards, 104(1), 537–579. <https://doi.org/10.1007/s11069-020-04180-9>
- Mohammadi, B. (2021). A review on the applications of machine learning for runoff modeling. *Sustainable Water Resources Management*, 7(6). <https://doi.org/10.1007/s40899-021-00584-y>
- Mohammadi, B., Guan, Y., Aghelpour, P., Emamgholizadeh, S., Pillco Zolá, R., & Zhang, D. (2020). Simulation of Titicaca lake water level fluctuations using hybrid machine learning technique integrated with grey wolf optimizer algorithm. *Water*, 12(11), 3015. <https://doi.org/10.3390/w12113015>
- Mulia, I. E., Tay, H., Roopsekhar, K., & Tkalich, P. (2013). Hybrid ANNeGA model for predicting turbidity and chlorophyll-a concentrations. *Journal of Hydro-environment Research*, 7(4), 279–299. <https://doi.org/10.1016/j.jher.2013.04.003>
- Nabipour, N., Dehghani, M., Mosavi, A., & Shamshirband, S. (2020). Short-term hydrological drought forecasting based on different nature-inspired optimization algorithms hybridized with artificial neural networks. *IEEE Access*, 8, 15210–15222. <https://doi.org/10.1109/access.2020.2964584>
- Nguyen, D. H., Hien Le, X., Heo, J.-Y., & Bae, D.-H. (2021). Development of an extreme gradient boosting model integrated with evolutionary algorithms for hourly water level prediction. *IEEE Access*, 9, 125853–125867. <https://doi.org/10.1109/access.2021.3111287>
- Nie, Q., Wan, D., & Wang, R. (2021). CNN-BiLSTM water level prediction method with attention mechanism. *Journal of Physics: Conference Series*, 2078, 012032. <https://doi.org/10.1088/17426596/2078/1/012032>
- Niu, W. J., & Feng, K. (2021). Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management. *Sustainable Cities and Society*, 64, 102562. <https://doi.org/10.1016/j.scs.2020.102562>
- Nourani, V., Molajou, A., Uzelaltinbulat, S., & Sadikoglu, F. (2019). Emotional artificial neural networks (EANNs) for multi-step ahead prediction of monthly precipitation; case study: Northern Cyprus. *Theoretical and Applied Climatology*, 138(3–4), 1419–1434. <https://doi.org/10.1007/s00704-019-02904-x>
- Nualtong, K., Panityakul, T., Khwanmuang, P., Chinram, R., & Kirtsang, S. (2021). A hybrid seasonal Box Jenkins-ANN approach for water level forecasting in Thailand. *Environment and Ecology Research*, 9(3), 93–106. <https://doi.org/10.13189/eer.2021.090301>
- Ömer Faruk, D. (2010). A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering Applications of Artificial Intelligence*, 23(4), 586–594. <https://doi.org/10.1016/j.engappai.2009.09.015>
- Ozan Evkaya, O., & Sevinç Kurnaz, F. (2020). Forecasting drought using neural network approaches with transformed time series data. *Journal of Applied Statistics*, 48(13–15), 2591–2606. <https://doi.org/10.1080/02664763.2020.1867829>
- Panyadee, P., Champrasert, P., & Aryupong, C. Water level prediction using artificial neural network with particle swarm optimization model. In *Proceedings of Fifth International Conference on Information and Communication Technology*, Melaka, Malaysia, 17–19 May 2017; pp. 1–6. IEEE.
- Pan, M., Zhou, H., Cao, J., Liu, Y., Hao, J., Li, S., & Chen, C.-H. (2020). Water level prediction model based on GRU and CNN. *IEEE Access*, 8, 60090–60100. <https://doi.org/10.1109/access.2020.2982433>
- Papacharalampous, G., Tyralis, H., & Koutsoyiannis, D. (2019). Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Stochastic Environmental Research and Risk Assessment*, 33(2), 481–514. <https://doi.org/10.1007/s00477-018-1638-6>
- Phitakwinai, S., Auephanwiriyakul, S., & Theera-Umpon, N. Multilayer perceptron with cuckoo search in water level prediction for flood forecasting. In *Proceedings of International Joint Conference on Neural Networks (IJCNN)*, Vancouver, BC, Canada, 24–29 July 2016. IEEE.
- Piasecki, A., Jurasz, J., & Skowron, R. (2017). Forecasting surface water level fluctuations of lake Serwy (Northeastern Poland) by artificial neural networks and multiple linear regression. *Journal of Environmental Engineering and Landscape Management*, 25(4), 379–388. <https://doi.org/10.3846/16486897.2017.1303498>
- Rahimzad, M., Nia, A. M., Zolfanoon, H., Soltani, J., Mehr, A. D., & Kwon, H. H. (2021). Performance Comparison of an LSTM-based deep learning model versus conventional machine learning algorithms for streamflow forecasting. *Water Resources Management*, 35(12), 4167–4187. <https://doi.org/10.1007/s11269-021-02937-w>
- Rakshitha, M. R., & Maheshan, C. M. (2020). A review on water level & quality monitoring system. *International Journal of Engineering Research & Technology (IJERT)*, 8. <https://doi.org/10.17577/IJERTCONV8IS14019>
- Ren, T., Liu, X., Niu, J., Lei, X., & Zhang, Z. (2020). Real-time water level prediction of cascaded channels based on multilayer perceptron and recurrent neural network. *Journal of Hydrology*, 585, 124783. <https://doi.org/10.1016/j.jhydrol.2020.124783>
- Rozos, E., Dimitriadis, P., Mazi, K., & Koussis, A. D. (2021). A multilayer perceptron model for stochastic synthesis. *Hydrology*, 8(2), 67. <https://doi.org/10.3390/hydrology8020067>
- Samanatary, S., & Sahoo, A. (2021). A comparative study on prediction of monthly streamflow using hybrid ANFIS-PSO approaches. *KSCE Journal of Civil Engineering*, 25(10), 4032–4043. <https://doi.org/10.1007/s12205-021-2223-y>
- Sami, B. H. Z., Khai, W. J., Sami, B. F. Z., Fai, C. M., Essam, Y., Ahmed, A. N., & El-Shafie, A. (2021). Investigating the reliability of machine learning algorithms as a sustainable tool for total suspended solid prediction. *Ain Shams Engineering Journal*, 12, 1607–1622. <https://doi.org/10.1016/j.asej.2021.01.007>
- Seo, Y., Kim, S., Kisi, O., & Singh, V. P. (2015). Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *Journal of Hydrology*, 520, 224–243. <https://doi.org/10.1016/j.jhydrol.2014.11.050>
- Seo, Y., Kim, S., & Singh, V. P. (2018). Comparison of different heuristic and decomposition techniques for river stage modeling. *Environmental Monitoring and Assessment*, 190(7), 392. <https://doi.org/10.1007/s10661-018-6768-2>
- Shafaei, M., & Kisi, O. (2015). Lake level forecasting using WAVELET-SVR, Wavelet-ANFIS and Wavelet-ARMA conjunction models. *Water Resources Management*, 30(1), 79–97. <https://doi.org/10.1007/s11269-015-1147-z>
- Shah, M. I., Javed, M. F., Alqahtani, A., & Aldrees, A. (2021). Environmental assessment based surface water quality prediction using hyper-parameter optimized machine learning models based on consistent big data. *Process Safety and Environmental Protection*, 151, 324–340. <https://doi.org/10.1016/j.psep.2021.05.026>

- Shuofeng, L., Puwen, L., & Koyamada, K. (2021). LSTM based hybrid method for basin water level prediction by using precipitation data. *Journal of Advanced Simulation in Science and Engineering*, 8(1), 40–52. <https://doi.org/10.15748/Jasse.8.40>
- Soleymani, S. A., Goudarzi, S., Anisi, M. H., Hassan, W. H., Idris, M. Y. I., Shamshirband, S., Noor, N. M., & Ahmedy, I. (2016). A novel method to water level prediction using RBF and FFA. *Water Resources Management*, 30(9), 3265–3283. <https://doi.org/10.1007/s11269-016-1347-1>
- Stamenković, L. J. (2021). Application of ANN and SVM for prediction nutrients in rivers. *Journal of Environmental Science and Health, Part A*, 56(8), 867–873. <https://doi.org/10.1080/10934529.2021.1933325>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Pearson Education, Inc.
- Tahraoui, H., Belhadj, A.-E., Hamitouche, A.-E., Bouhedda, M., & Amrane, A. (2021). Predicting the concentration of sulfate (SO₄²⁻) in drinking water using artificial neural networks: A case study: Médéa-Algeria. *Desalination and Water Treatment*, 217, 181–194. <https://doi.org/10.5004/dwt.2021.26813>
- Tao, H., Al-Bedry, N. K., Khedher, K. M., Shahid, S., & Yaseen, Z. M. (2021). River water level prediction in coastal catchment using hybridized relevance vector machine model with improved grasshopper optimization. *Journal of Hydrology*, 598, 126477. <https://doi.org/10.1016/j.jhydrol.2021.126477>
- Tiu, E. S. K., Huang, Y. F., Ng, J. L., AlDahoul, N., Ahmed, A. N., & Elshafie, A. (2021). An evaluation of various data pre-processing techniques with machine learning models for water level prediction. *Natural Hazards*, 110(1), 121–153. <https://doi.org/10.1007/s11069-021-04939-8>
- Tripura, J., Roy, P., & Barbhuiya, A. K. (2020). Simultaneous streamflow forecasting based on hybridized neuro-fuzzy method for a river system. *Neural Computing & Applications*, 33(8), 3221–3233. <https://doi.org/10.1007/s00521-020-05194-x>
- Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, 11(5), 910. <https://doi.org/10.3390/w11050910>
- Wang, B., Wang, B., Wu, W., Xi, C., & Wang, J. (2020). Sea-water-level prediction via combined wavelet decomposition, neuro-fuzzy and neural networks using SLA and wind information. *Acta Oceanologica Sinica*, 39(5), 157–167. <https://doi.org/10.1007/s13131-020-1569-1>
- Wee, W. J., Zaini, N. B., Ahmed, A. N., & El-Shafe, A. (2021). A review of models for water level forecasting based on machine learning. *Earth Science Informatics*, 14(4), 1707–1728. <https://doi.org/10.1007/s12145-021-00664-9>
- Xie, Z., Liu, Q., & Cao, Y. (2021). Hybrid deep learning modeling for water level prediction in yangtze river. *Intelligent Automation & Soft Computing*, 28(1), 153–166. <https://doi.org/10.32604/iasc.2021.016246>
- Xu, G., Cheng, Y., Liu, F., Ping, P., & Sun, J. A water level prediction model based on ARIMA-RNN. In *Proceedings of 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, Newark, CA, USA, 04–09 April 2019; pp. 221–226. IEEE.
- Yang, X. S. F. A. (2010). stochastic test functions and design optimisation. *International Journal of Bio-Inspired Computation*, 2(2), 78. <https://doi.org/10.1504/ijbic.2010.032124>
- Yarar, A., Onucyildiz, M., & Coptay, N. K. (2009). Modelling level change in lakes using neuro-fuzzy and artificial neural networks. *Journal of Hydrology*, 365(3–4), 329–334. <https://doi.org/10.1016/j.jhydrol.2008.12.006>
- Young, -C.-C., Liu, W.-C., & Hsieh, W.-L. (2015). Predicting the water level fluctuation in an alpine lake using physically based. *Artificial Neural Network, and Time Series Forecasting Models. Mathematical Problems in Engineering*, 2015, 1–11. <https://doi.org/10.1155/2015/708204>
- Zhang, Y.-F., Fitch, P., & Thorburn, P. J. (2020). Predicting the trend of dissolved oxygen based on the kPCA-RNN model. *Water*, 12. <https://doi.org/10.3390/w12020585>
- Zhang, J., Huo, A., Zhao, Z., Yang, L., Peng, J., Cheng, Y., & Wang, Z. (2022). Impact of Mountain Reservoir construction on groundwater level in downstream loess areas in Guanzhong basin, China. *Water*, 14. <https://doi.org/10.3390/w14091470>
- Zhang, J., & Lou, Y. Water level prediction based on improved grey RBF neural network model. In *Proceedings of IEEE, Xi'an, China, 03-05 October 2016*. IEEE.
- Zhang, Z., Zhang, Q., & Singh, V. P. (2018). Univariate streamflow forecasting using commonly used data-driven models: Literature review and case study. *Hydrological Sciences Journal*, 63(7), 1091–1111. <https://doi.org/10.1080/02626667.2018.1469756>
- Zhong, C., Guo, T., Jiang, Z., Liu, X., & Chu, X. A hybrid model for water level forecasting: A case study of Wuhan station. In *Proceedings of 2017 4th international conference on transportation information and safety (ICTIS)*, August 8–10, 2017, banff, Canada, 8–10 August 2017. IEEE.
- Zhong, C., Jiang, Z., Chu, X., Guo, T., & Wen, Q. (2017). Water level forecasting using a hybrid algorithm of artificial neural networks and local Kalman filtering. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 233, 174–185. <https://doi.org/10.1177/1475090217727135>
- Zhu, S., Lu, H., Ptak, M., Dai, J., & Ji, Q. (2020). Lake water-level fluctuation forecasting using machine learning models: A systematic review. *Environmental Science and Pollution Research International*, 27(36), 44807–44819. <https://doi.org/10.1007/s11356-020-10917-7>
- Zubaidi, S. L., Abdulkareem, I. H., Hashim, K. S., Al-Bugharbee, H., Ridha, H. M., Gharghan, S. K., Al-Qaim, F. F., Muradov, M., Kot, P., & Al-Khaddar, R. (2020). Hybridised artificial neural network model with slime mould algorithm: A novel methodology for prediction of urban stochastic water demand. *Water*, 12(10), 2692. <https://doi.org/10.3390/w12102692>
- Zubaidi, S. L., Dooley, J., Alkhaddar, R. M., Abdellatif, M., Al-Bugharbee, H., & Ortega-Martorell, S. (2018). A novel approach for predicting monthly water demand by combining singular spectrum analysis with neural networks. *Journal of Hydrology*, 561, 136–145. <https://doi.org/10.1016/j.jhydrol.2018.03.047>
- Zubaidi, S. L., Hashim, K., Ethaib, S., Al-Bdairi, N. S. S., Al-Bugharbee, H., & Gharghan, S. K. (2022). A novel methodology to predict monthly municipal water demand based on weather variables scenario. *Journal of King Saud University - Engineering Sciences*, 34(3), 163–169. <https://doi.org/10.1016/j.jksues.2020.09.011>
- Zubaidi, S. L., Kot, P., Alkhaddar, R. M., Abdellatif, M., & Al-Bugharbee, H. Short-term water demand prediction in residential complexes: Case study in Columbia city, USA.

In *Proceedings of 2018 11th International Conference on Developments in eSystems Engineering (DeSE)*, Cambridge, UK, 02-05 September 2018; pp. 31–35. IEEE.
Zubaidi, S. L., Ortega-Martorell, S., Al-Bugharbee, H., Olier, I., Hashim, K. S., Gharghan, S. K., Kot, P., & Al-Khaddar, R. (2020). Urban water demand prediction for a city that suffers from climate change and population growth: Gauteng province case study.

Water, 12(7), 1885. <https://doi.org/10.3390/w12071885>

Zubaidi, S. L., Ortega-Martorell, S., Kot, P., Alkhaddar, R. M., Abdellatif, M., Gharghan, S. K., Ahmed, M. S., & Hashim, K. (2020). A method for predicting long-term municipal water demands under climate change. *Water Resources Management*, 34(3), 1265–1279. <https://doi.org/10.1007/s11269-020-02500-z>



© 2022 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



Cogent Engineering (ISSN: 2331-1916) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com

