



LJMU Research Online

Zubaidi, SL, Al-Bugharbee, H, Alattabi, AW, Ridha, HM, Hashim, K, Al-Ansari, N and Yaseen, ZM

Forecasting urban water demand using different hybrid-based metaheuristic algorithms' inspire for extracting artificial neural network hyperparameters

<http://researchonline.ljmu.ac.uk/id/eprint/24621/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Zubaidi, SL, Al-Bugharbee, H, Alattabi, AW, Ridha, HM, Hashim, K, Al-Ansari, N and Yaseen, ZM (2024) Forecasting urban water demand using different hybrid-based metaheuristic algorithms' inspire for extracting artificial neural network hyperparameters. Scientific Reports. 14 (1).

LJMU has developed [LJMU Research Online](#) for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>



OPEN Forecasting urban water demand using different hybrid-based metaheuristic algorithms' inspire for extracting artificial neural network hyperparameters

Salah L. Zubaidi^{1,2}, Hussein Al-Bugharbee³, Ali W. Alattabi¹, Hussein Mohammed Ridha^{4,5}, Khalid Hashim^{6,7}, Nadhir Al-Ansari⁸ & Zaher Mundher Yaseen⁹

This research offers a novel methodology for quantifying water needs by assessing weather variables, applying a combination of data preprocessing approaches, and an artificial neural network (ANN) that integrates using a genetic algorithm enabled particle swarm optimisation (PSOGA) algorithm. The PSOGA performance was compared with different hybrid-based metaheuristic algorithms' behaviour, modified PSO, and PSO as benchmarking techniques. Based on the findings, it is possible to enhance the standard of initial data and select optimal predictions that drive urban water demand through effective data processing. Each model performed adequately in simulating the fundamental dynamics of monthly urban water demand as it relates to meteorological variables, proving that they were all successful. Statistical fitness measures showed that PSOGA-ANN outperformed competing algorithms.

Keywords Urban water management, ANN, PSOGA, Metaheuristic algorithms, Water demand prediction

Freshwater resources are extremely important and play a vital role in developing cities. The logical design and active management of the municipal water supply framework are of great importance to guarantee social and economic development¹. Urban water demand is rising in many nations worldwide due to the escalating severity of climate change, population expansion, and economic development^{2,3}. Water shortage has been well documented in developed and developing countries, leading to an imbalance between supply and requests for water^{4,5}.

To properly manage modern freshwater resources, sound estimates and predictions of water demands in a certain area are essential to overcome the issue of water scarcity. This should be achieved because various cities in the United States (US), particularly those in arid regions such as the state of Texas in the southwest of the US, encounter problems of water scarcity. As such, authorities and engineers are continually trying to keep up with the rising demand resulting from the growing impact of socio-economic, political, and weather factors⁶. DeMaagd and Roberts⁷ stated that shifting patterns of rainfall are anticipated to influence surface water and aquifers worldwide, some of which, such as the Southwestern US, are already increasingly strained. Reliable estimates of water requests that account for factors driving consumption are essential to understanding future municipal water demands⁶.

Additionally, short-term water demand forecasting aids in administering and operating water supply systems. As an illustration, forecasts of short-term demand help water managers balance the needs of water supply to make

¹Department of Civil Engineering, Wasit University, Wasit 52001, Iraq. ²College of Engineering, University of Warith Al-Anbiyaa, Karbala 56001, Iraq. ³Department of Mechanical Engineering, Wasit University, Wasit 52001, Iraq. ⁴Advanced Lightning, Power and Energy Research (ALPER), Department of Electrical and Electronics Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Malaysia. ⁵Department of Computer Engineering, Mustansiriyah University, Baghdad, Iraq. ⁶Department of Environmental Engineering, University of Babylon, Al-Hillah 51001, Iraq. ⁷School of Civil Engineering and Built Environment, Liverpool John Moores University, Liverpool, UK. ⁸Department of Civil Environmental and Natural Resources Engineering, Lulea University of Technology, 971 87 Lulea, Sweden. ⁹Civil and Environmental Engineering Department, King Fahd University of Petroleum & Minerals, 31261 Dhahran, Saudi Arabia. ✉email: salahlafta@uowasit.edu.iq; nadhir.alansari@ltu.se

better-informed decisions regarding water management. Reliable urban water demand forecasting models must be established in order to help ensure reliable water availability and reduce peak water use. These forecasting models help water utilities make tactical and strategic decisions, enhancing water security and the sustainability of water use⁸.

Water demand datasets typically exhibit non-stationary and non-linear behaviour at various spatial and temporal dimensions⁹. With the development of machine learning (ML) models, different versions of ML models were adopted to solve the associated non-linear nature of water demand variability. Some examples of these models are artificial neural networks (ANNs)¹⁰, support vector machines¹¹, random forests¹², gene expression programming¹³, long short-term memory neural networks¹⁴, and fuzzy logic¹⁵.

An analysis of different methods for predicting urban water demand over the last several decades^{16–20} detected that the ANN model was used effectively for different scenarios. ANN techniques have revealed optimistic progress in estimating short-, medium-, and long-term municipal water requests. Also, among several ML models, ANN showed a predominant model applied for various hydrology fields^{21,22}. However, the capacity of individual ML methods to understand intricate data patterns and relationships is typically restricted, especially when dealing with non-linear and high-dimensional data²³. This may lead to less efficient and accurate results when contrasted with more advanced ML models²⁴. To tackle these limitations, an optimising method was developed.

Metaheuristic algorithms (MHAs) show promise in solving a variety of challenging non-linear issues across hydrological disciplines when comparing hybrid models to single ML models, such as^{25–28}. These combined techniques are preferable to solving complicated real-world problems²⁹. This is why these methods have shown improvement in identifying new regions that could lead to a broader set of solutions³⁰. Not only that, but the ability to exploit is boosted so that local minima are avoided. Finally, when dealing with complex, multi-variable, and non-linear problems, these methods perform better³¹. Regardless, optimisation is still necessary, especially in forecasting hydrological factors due to their stochastic, data noise, and non-stationary nature³².

Exploration and exploitation are the two main parts of MHAs. In computer science, “exploration” refers to the steps used to identify the boundaries of a search space for an algorithm, while “exploitation” delineates the process of picking the best option out of a multitude of generated possibilities³³. Finding the optimal balance between exploring and exploiting is essential for a search algorithm’s performance³⁴. The relationship between exploration and exploitation, however, is inversely proportional³⁵. MHAs can be classified based on their inspired technique into physics (P), such as gravitational search algorithm (GSA), evolutionary algorithms (EA) like genetic algorithm (GA), and swarm intelligence (SI), for example, particle swarm optimisation (PSO) and grey wolf optimiser (GWO)³⁶. The effectiveness of improving algorithms to increase their efficiency in tackling optimising issues, either by altering existing algorithms (single-based) or combining different algorithms into hybrid ones (hybrid-based)²⁵.

The framework of the suggested research methodology depends on data preprocessing techniques (i.e., normalisation, cleaning, and selecting best predictors) and prediction models³⁷. Data preprocessing techniques are crucial steps that help to reduce the multicollinearity between predictors, improve data quality, and remove redundant variables, resulting in an improved forecasting model performance³⁸. However, previous research studies of hydrological hybrid models have suffered from notable methodological weaknesses, leading to decreased prediction range and increased uncertainty. Considering data preprocessing techniques, previous studies have not dealt with one or more steps of data preprocessing methods, such as normalisation data^{39,40}, cleaning data^{41,42}, and selecting the best predictors^{43,44} or unimplemented all these steps^{45,46}. Also, along with this growth in hydrological hybrid models, however, there are certain drawbacks associated with using hybrid models of prediction in multiple previous studies. One major drawback of these studies is that they hybridised the ML model with one MHA^{47,48}. Hybridised the ML model with MHAs, which have the same inspiration^{49,50}, is another (potential) limitation. Another problem with these studies is that they hybridised the ML model with only single-based MHAs^{51,52}. Additionally, these studies, however, suffer from the fact that they were applying each swarm of MHA with few iterations^{52,53}. Finally, applying each swarm of MHA one time^{54,55} is another potential concern.

Considering all of that, this study will compare the performance of the different hybrid MHAs based on their inspiration, and no literature has compared these MHAs before in this sector. It includes PSOGA (SI combined with EA), CPSOCSA (SI combined with P), PSOGWO (SI combined with SI), the modified PSO (MPSO), and PSO algorithm (as a benchmarking model) to forecast urban water demand. The above MHAs were applied effectively in different hydrology fields, such as PSOGA^{56,57}, CPSOCSA^{58,59}, PSOGWO⁶⁰, MPSO^{54,61}, and PSO⁶².

Water demand forecasts have lately become a very active area of research, as they lead to considerable environmental and economic benefits. A precise water demand forecast ensures a dependable water distribution system that is capable of providing users with potable water in sufficient volumes and adequate pressure⁶³. The field of water demand estimation has become more critical, resulting from water resource shortages and a growth in water usage. Consequently, real doubts persist among water utility decision-makers regarding the present urban water system’s capacity to handle the exponential increase in water demands⁶⁴. Therefore, the present study aims to address the following research objectives in light of the limitations highlighted in the literature review:

1. The capacity of the wavelet transformation (with various mother wavelets and orders) was adopted for data pre-processing as an advanced stage for the prediction process.
2. A new combined ANN-PSOGA model was developed based on the combination of the ANN model and hybrid nature-inspired algorithms for water demand prediction using ten years dataset belonging to College Station City, USA.

Variable	Maximum	Minimum	Mean
Water consumption (ML)	715,798,498	207,601,706	366,849,231
Tmax (°C)	41.2	11.4	26.8
Tmin (°C)	26.3	2.6	14.8
Tmean (°C)	33.8	7.1	20.8
Rain (mm)	326.8	2.1	75.1
Srad (MJ/m ²)	28.3	17.1	6.7
RHmax (%)	0.83	0.35	0.64
Wind (m/s)	4.7	2.1	3.1

Table 1. Statistical indicators of the collected data.

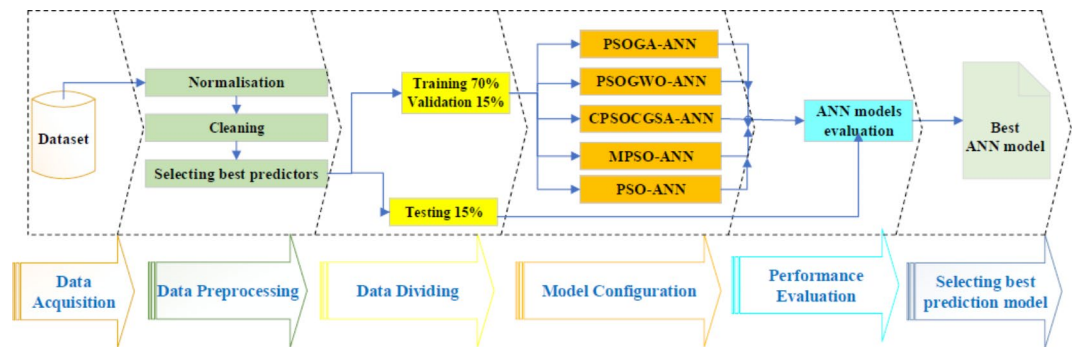


Fig. 1. Framework of the suggested research methodology.

3. The proposed combined model ANN-PSOGA was validated against several other combined models (ANN-CPSOCGSA, ANN-PSOGWO, ANN-MPSO, and ANN-PSO) for validation and benchmarking purposes.
4. For each population size, the forecasting range was increased, and the uncertainty was decreased by repeating the process five times.

The remainder of this paper has been divided as follows. In “**Study area and data set**” section deals with the study area and data set. In “**Methodology**” section explains the methodology. Results are covered in “**Results**” section. In “**Discussion**” section considers the discussion. In “**Conclusion**” section presents the conclusion.

Study area and data set

The present research adopted a catchment area in the USA located in College Station, TX, to build and evaluate the water demand model. The water services for the approximately 150,000 residents of this area are provided by the city’s municipal government. The region that is serviced is around 140 km², with a residential service population of more than 95,000 customers and around 55,000 for commercial and industrial purposes. Water is supplied by the City’s nine groundwater wells and conveyed to the Dowling Road Pump Station by the Sandy Point Pump Station and transmission lines. Table 1 provides the statistical indicators of the dependent and independents data, including the urban water usage (megalitre, ML), maximum temperature (Tmax) (°C), minimum temperature (Tmin) (°C), mean temperature (Tmean) (°C), rainfall (Rain) (mm), solar radiation (Srad) (MJ/m²), maximum relative humidity (RHmax) (%), and wind (m/s) from January 2005 to July 2014.

Methodology

This research suggests a novel methodology focused on understanding the forecast monthly consumption of urban water based on weather variables (see Fig. 1). The framework of the suggested research methodology started with data acquisition from College Station, USA, (“**Study area and data set**” section). This is followed by data preprocessing techniques that include three steps responsible for enhancing raw data quality and selecting the best set of predictors. The data was then divided into three sets (i.e., data dividing): training, validation, and testing. Afterwards, five MHAs were combined with ANN to determine the optimal hyperparameters for the ANN model (i.e., model configuration stage). Next, the performance evaluation was conducted. In this stage, all the models are conducted. Finally, all the models are compared based on different statistical and graphical tests to select the best prediction models. The methodology will be described in the subsequent sections (i.e., from 3.1 to 3.6).

Data preprocessing

It is a valuable technique for the prediction model, and it is categorised into normalisation data, cleaning data, and determining the optimum predictors²⁸. Time series normalisation is best accomplished using the natural logarithm since it mitigates the impact of outliers and eliminates collinearity across predictors^{37,65}. The cleaning of data can be implemented by applying the box and whisker method to determine and treat the outliers' data. Then, denoising the time series after that by wavelet transform (WT) technique.

WT is a popular time–frequency analysis method. The principles of WT contain shifting and scaling the so-called mother wavelet along with the original time series to obtain a time–frequency representation. This WT can also be utilised to denoise the original time series. The principles of wavelet denoising contain localising time series information at different scales. Then, the important information (i.e., large-magnitude wavelet coefficients) is preserved while the noise (i.e., small-magnitude wavelet coefficients) can be shrunk or removed using thresholding. The denoised time series can be reconstructed using inverse wavelet transform^{66,67}. Various mother wavelets are utilised and investigated in this study, and the most effective one in terms of providing a higher correlation between dependent and independent parameters is selected. The process of denoising and reconstructing can be conducted using the MATLAB toolbox.

The final step is to locate the optimum scenario of predictors, and for this reason, seven independent variables (weather variables) were determined as potential predictors of monthly urban water demand. The covariates were screened for inclusion in the prediction models utilising the tolerance technique to account for a high level of collinearity within the potential independent variables. Covariates with a tolerance coefficient of less than 0.2 were removed because they make little to no enhancement in forecast quality⁶⁵.

Genetic algorithm enabled particle swarm optimisation

The property of intensification provided by the genetic algorithm (GA) and diversification of PSO are combined to provide a better solution⁶⁸.

The first step in PSO is to propose an initial population representing the potential solution. The individuals (i.e. particles) of this population are represented by the Eq. (1)

$$X_i = (X_{i1}, X_{i2}, X_{i3}, \dots, X_{iD}) \quad (1)$$

where i is the particle number, and D is the dimension. As in birds swarm, which look for their food, these particles are proposed to keep moving with a velocity v in the search space, looking for the best solution. The initial swarm velocity is represented in Eq. (2)

$$V_i = (V_{i1}, V_{i2}, V_{i3}, \dots, V_{iD}) \quad (2)$$

Using the initial particle positions and velocities to calculate the fitness value according to a predetermined fitness function.

The genetic algorithm (GA) is inspired by the principles of Genetics and Natural Selection and is popularly used for finding optimum solutions to various problems. The first step in the GA technique is to propose a set of random populations where each individual is called a chromosome, and each chromosome consists of a fixed length of strings, where a string is called a gene^{69,70}. Then, the chromosomes of the first generation (1st set of population) follow several steps, including selection, cross-over and mutation to create a new population. In the selection step, a number of chromosomes of best fitness values are selected to create the next generation through mating and cross-over process. In the crossover operation, parts of two parent chromosomes are swapped as described in the Eqs. (4, 5) below:

$$\begin{aligned} X_i^{t+1} &= \alpha \times X_j^t + (1 - \alpha) \times X_i^t \\ X_j^{t+1} &= \alpha \times X_i^t + (1 - \alpha) \times X_j^t \end{aligned} \quad (3)$$

where X_i^{t+1}, X_j^{t+1} are two new chromosomes created at time $t+1$, α refers to the crossover variable, and $i \neq j$. After that, the new chromosomes are subjected to a mutation process where some gene values are changed, and it is given as follows:

$$X_j^{t+1} = X_i^t + m * rand(size(D)), \quad (4)$$

where m is the mutate factor and its computed as follows:

$$m = 0.1 * (H - L); \quad (5)$$

where H and L are the lower and upper boundaries. According to the PSO algorithm, the particles update their locations at every iteration based on the local and global best position according to Eq. (6).

$$X_i^{t+1} = X_i^t + V_i^t \quad (6)$$

where t is the iteration number. The velocity is also updated at every iteration according to the Eq. (7).

$$V_i^{t+1} = w * V_i^t + C_1 * r_1 * (pBest_i^t - X_i^t) + C_2 * r_2 * (gBest^t - X_i^t) \quad (7)$$

where w is inertia weight, C_1 and C_2 are non-negative constants controlling how the global and local best position affect the particle velocity, and r_1, r_2 are random constants belonging to the range $[0, 1]$. The fitness value is calculated at every iteration until the termination condition is reached and the best solution is found.

The fitness values are evaluated for the new generation. The process of creation of new generation and evaluation of fitness is repeated till the termination condition is met⁷¹.

As it was mentioned earlier that the present methodology combines their merits of PSO social thinking and GA local search ability which helps in obtaining better solution. In this methodology, the PSO performs the building solution while the GA plays as local search optimiser.

The first step in this hybrid is to propose an initial random population as in PSO and the $pBest$ and $gBest$ values are calculated in the first iteration. Next, the new position and velocity vectors of the PSO particles are updated and fitness values are also evaluated. Then, the new position sets are subjected to GA to represent the chromosomes sets and follow the same processes explained in the previous section (i.e. selection, crossover and mutation) to infer the best solution. The chromosomes associated with the optimum solution in GA are then sent back to PSO as an updated population. The process above is repeated until the meeting of the target fitness. One way to calculate the PSOGA fitness function is to use the root mean square error (RMSE) to determine the best and worst fits for each iteration.

Constriction coefficient-based particle swarm optimisation with chaotic gravitational search algorithm

The Constriction Coefficient-based Particle Swarm Optimisation with Chaotic Gravitational Search Algorithm (CPSOCGSA) can be categorised as a stochastic hybrid optimisation methodology. The suggested methodology combines Particle Swarm Optimisation (PSO), which is influenced by bird flocking behaviour, with the Gravitational Search Algorithm (GSA), a technique, which is inspired by Newton's law of universal gravity. The proposed methodology utilises the exploration and exploitation capabilities of PSO and GSA to attain the optimal result⁷².

The CPSOCGSA has been proposed to enhance the exploratory outcomes inherent in the GSA, in conjunction with the convergence possibility of the Constriction Coefficient-Based PSO. To tackle the problem of being trapped in local minima, which is usually occurs in the classic GSA, chaotic maps are proposed as a means to address the issue. The equation that merges both aforementioned techniques is presented in Eq. (8)⁷².

$$v_i^d(t+1) = (2/(\varphi - 2 + \text{sqrt}(\varphi^2 - 4\varphi)))v_i^d(t) + K\varphi_1r_{i1}(a_i^d(t) - x_i^d(t)) + K\varphi_2r_{i2}(g_{best} - x_i^d(t)) \quad (8)$$

In this context, the variable v_i^d represents the velocity of the particles in the swarm, $(\varphi, \varphi_1, \varphi_2)$ are control parameters, K is the Constriction Coefficient. a_i^d refers to the acceleration of the particles, and g_{best} refers to the particle system's social capability component.

The spatial coordinates of the particles are presented by Eq. (9)⁷².

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad (9)$$

The usefulness of optimisation algorithms in addressing continuous benchmark test functions is evident due to the inherent simplicity that allows agents to traverse the search space and identify feasible potential solutions within this specific category of functions. However, the accurate assessment of intelligent algorithms resides in their capacity to effectively address complex non-linear test operations, such as those encountered in engineering standards. In the aforementioned situations, algorithms are required to efficiently address complex restrictions and rigorous inequalities⁷². To calculate the CPSOCGSA fitness function, a RMSE can be used to choose the best and the worst fit for each iteration.

Particle swarm optimisation with grey wolf optimiser

Grey Wolf Optimiser (GWO), based on their creators, is driven by the leadership sequence and hunting technique of grey wolves in nature⁷³. The grey wolves can be considered as the top food chain's consumers since they are the tertiary consumers. Regardless of gender, grey wolves are split into four groups within the leadership sequence: alpha, beta, delta, and omega⁷⁴. After Alpha wolves, beta wolves are the best solutions, according to the GWO algorithm, followed by the delta wolves. Omega wolves are the supporters of the abovementioned wolf groups and serve as the scapegoats for the submissive wolves⁷⁴. The top wolves are thought to be the ones doing the hunting. According to Muro et al.⁷⁵, the procedure is carried out in three phases: the chasing phase, the pursuit phase, and the attacking phase. A mathematical model based on the abovementioned hunting phases is built. The following Eqs. (10, 11) can be adapted to model the GWO mathematically:

$$D = |C * X_p(t) - X(t)|, \quad (10)$$

$$X(t+1) = X_p(t) - A * D \quad (11)$$

In this context, the variable t represents the count of instantaneous iterations. D represents encircling the prey. X_p indicates the position of the prey, whereas X represents the location of grey wolves. The coefficients A and C are utilised for the vectors. The coefficients A and C are computed according to the following Eqs. (12, 13):

$$A = a(2 * r_1 - 1), \quad (12)$$

$$C = 2 * r_1 \quad (13)$$

The number of a is a linear reduction from 2 to 0 as the number of iterations decreases. The variables r_1 and r_2 indicate random numbers selected between (0, 1).

According to the literature, PSO provides promising results in several engineering challenges. One of the most recent projects utilising PSO featured microgrid energy scheduling⁷⁶, and the outcomes were impressive. The exploration ability of GWO, as indicated by Şenel et al.⁷⁷, was introduced to reduce the likelihood of swarms being drawn to a local minima. It may be deduced that the hybrid PSO-GWO is efficient in terms of its capacity for exploration as well as congregation⁷⁸. For each iteration, to calculate the PSOGWO fitness function, a RMSE can be used to determine the best and the worst fit.

ANN model

Most current ML applications in hydrology involve ANNs, particularly feedforward back-propagation (FFBP) learning. Accurate simulations of municipal water needs across several spatial and temporal dimensions were generated by mapping the non-linear behaviour of water data using the FFBP^{79,80}.

It has been proven that with two hidden layers, ANNs can accurately simulate the non-linear relationship between predictors and targets, and this method has been effectively used by a wide range of researchers across a broad variety of applications, such as Sadeghifar et al.⁸¹, Shah et al.⁸², and Nunes Carvalho et al.⁸³. Based on this, an ANN was constructed with four layers (Fig. 2): one for the predictors (i.e., climate factors), two hidden layers for data processing, and one for the target (i.e., urban water usage). Each of both hidden layers employs a tangential activation function, while using a linear activation function in the output layer. The dataset was divided into a training set (containing 70% of the data), a validation set (15%), and a testing set (15%), as per prior studies Zubaidi et al.⁵ and Mohammed et al.⁸⁴. A perfect fit for ANN training is the procedure for calculating the ANN coefficients, which are the interlayer biases and weights. Consequently, optimisation approach algorithms (bayesian regularisation) are executed to improve precision⁸⁵.

Four MHAs (i.e., PSOGA, PSOGWO, CPSOCGSA, MPSO, and PSO) were utilised separately to develop the ANN technique by finding optimum values for the ANN's hyper-parameters (i.e., the learning rate coefficient (Lr) and the number of neurons in the first and second hidden layers (N1 and N2)).

Validation of the model

This research used different performance criteria to assess the model performance. Due to the lack of a universally applicable performance metric, it is essential to choose the appropriate criteria for a certain application. Additionally, it is usual to utilise multiple performance criteria due to the pros and cons that exist in each criterion⁸⁶. Also, conduct different statistical tests help to ensure the superiority of the proposed approach⁸⁷. Several types of performance criteria used in this research include root mean squared error (RMSE, Eq. 14), mean absolute error (MAE, Eq. 15), mean absolute relative error (MARE, Eq. 16), nash sutcliffe coefficient (NSC, Eq. 17), normalised mean square error (NMSE, Eq. 18), and coefficient of determination (R^2 , Eq. 19). For a perfect model, RMSE and MAE would be zero, and NMSE, NSC, R^2 , and correlation coefficient (R) would be one^{21,88,89}. Moreover, different graphical tests were used to assess the methodology, such as the Taylor diagram and Violin

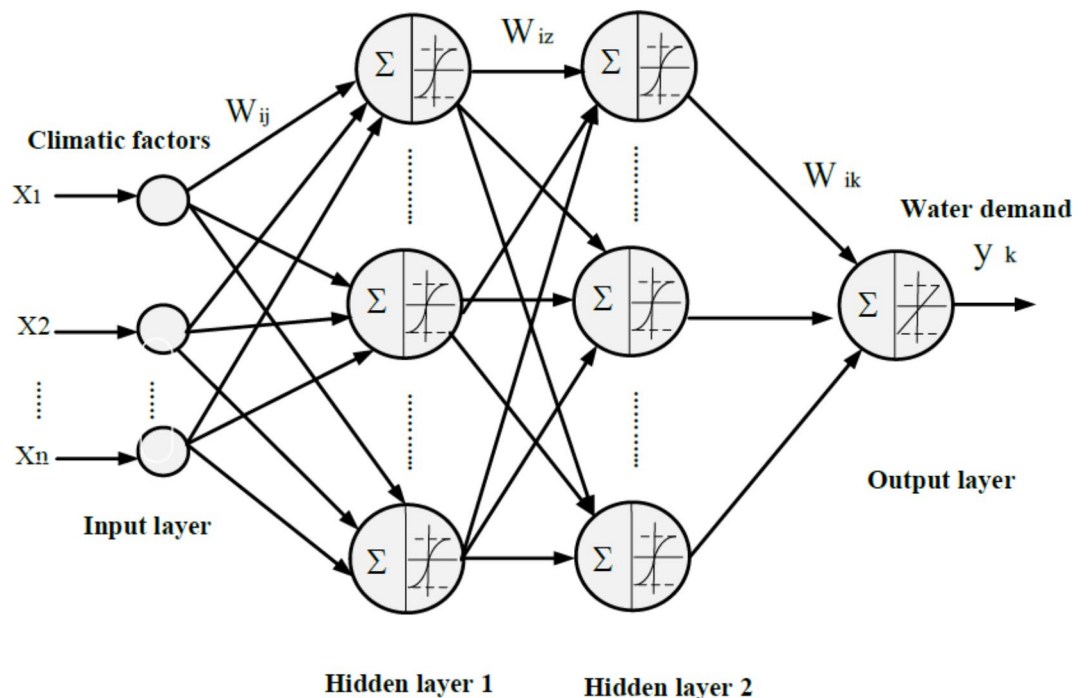


Fig. 2. ANN model architecture.

plot. Furthermore, four tests, including the Kolmogorov–Smirnov, Shapiro–Wilk, Augmented Dickey–Fuller (ADF), and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests, were used to evaluate the residual data.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - F_i)^2}{N}} \quad (14)$$

$$MAE = \frac{\sum_{i=1}^N |O_i - F_i|}{N} \quad (15)$$

$$MARE = \frac{1}{N} \sum_{i=1}^N \frac{|O_i - F_i|}{O_i} \quad (16)$$

$$NSC = 1 - \frac{\sum_{i=1}^N (O_i - F_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2} \quad (17)$$

$$NMSE = 1 - \frac{\sum_{i=1}^N |O_i - F_i|}{\sum_{i=1}^N |O_i - \bar{O}_i|} \quad (18)$$

$$R^2 = \left[\frac{\sum_{i=1}^N (O_i - \bar{O}_i)(F_i - \bar{F}_i)}{\sqrt{\sum_{i=1}^N (O_i - \bar{O}_i)^2 \sum_{i=1}^N (F_i - \bar{F}_i)^2}} \right]^2 \quad (19)$$

where O_i : observed water consumption, F_i : forecast water demand, \bar{O}_i : mean of observed water consumption, \bar{F}_i : mean of forecast water demand, N : number of data points.

Results

Data preprocessing techniques

Data should be preprocessed before being used to build the prediction model (as mentioned in “Data preprocessing” section). Accordingly, firstly, dependent and independent time series were normalised. Then, outliers were detected and treated.

Afterwards, the WT technique was employed for denoising all the time series. Initially, sym, db, and coif wavelets were examined independently in various orders to determine the optimum order for each type. The outcomes reveal that the optimum order is 5 for sym, db, and coif wavelets. The five kinds of mother wavelets (sym5, Haar, dmey, db5, and coif5) were applied separately for denoising all the time series, as shown in Fig. 3A, and for more details, Fig. 3B provides a clear view. The most interesting aspect of this figure is that all the types increase the model’s accuracy, and coif5 is the best.

The tolerance technique was used with different scenarios to locate the optimum predictors (weather variables) to simulate water consumption. Table 2 presents the optimum scenario of the predictors (i.e., T_{\max} , Rain, wind, and Rh_{\max}). It can be seen that the tolerance constants for all the predictors are more than 0.2, which means that there is no violation of the multicollinearity assumption.

Models configuration

It is important to methodically design the prediction model in order to accurately estimate water demand after data has been divided into training, validation, and testing stages. For the ANN model, it is combined with a metaheuristic algorithm (PSOGA-ANN) to find the optimum hyperparameters (i.e., N_1 , N_2 , and L_r). Other metaheuristic algorithms (i.e., PSOGWO-ANN, CPSOCGSA-ANN, and MPSO-ANN) are used to confirm the validation of the PSOGA-ANN’s results. Further, the performance of all the above hybrid models was compared with a benchmarking model (i.e., PSO-ANN) to examine to what extent the performance of the PSO algorithm was enhanced by modification or hybridised with another MHA to forecast water demand based on several weather factors.

For each algorithm, five population sizes (10, 20, 30, 40, and 50) were employed to implement the combined technique. For each population size, the forecasting range was increased, and the uncertainty was decreased by repeating the process five times (e.g., see Fig. 4 for the PSOGA-ANN algorithm). The best implementation that resulted in the lowest error (best fitness function, RMSE) was chosen for each population size (e.g., implementation two for the 10 population size of PSOGA-ANN is the best) and combined with the best implementation for the remaining population sizes (see Fig. 5 for all MHAs). Figure 5 demonstrates that the optimal population size across all methods is 50. So, the 50 population size of each hybrid model offered L_r , N_1 , and N_2 values for their ANN model, with the exception of the PSO-ANN model, which provides the ANN model’s hyperparameters via a population size of 40.

Therefore, the ANN method for water demand simulation has been fine-tuned using the output of each MHA. Consequently, the optimal ANN hyperparameters for the best population size for each case are tabulated in Table 3.

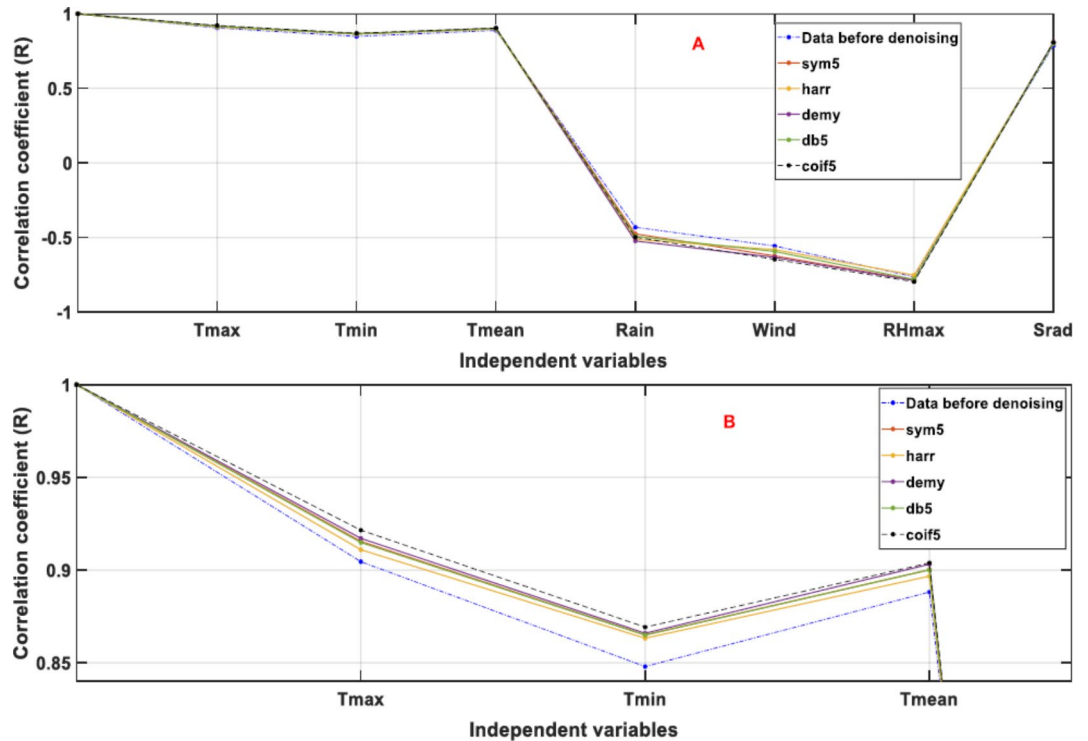


Fig. 3. Correlation coefficients between dependent and independent factors (both raw and denoised using five distinct wavelets).

Predictors	Tolerance constants
T _{max}	0.294
Rain	0.392
wind	0.454
RH _{max}	0.272

Table 2. Collinearity statistics.

Evaluation of the prediction models performance

After configuring the ANN models of prediction by locating the optimum hyperparameters for the ANN as presented in Table 3, all the models run multiple times to select the best ANN model architecture (i.e., interlayer biases and weights) to simulate the monthly water consumption. During the testing stage, various statistical criteria were applied to evaluate the methods' capability to extrapolate urban water requests from meteorological factors.

Firstly, absolute error criteria (RMSE and MAE), maximum error criterion (Max.(error)), relative error (MARE), and dimensionless error criterion (NSC, NMSE, and R²) were applied for assessing and comparing technique performances. The performance assessment outcomes for the testing stage are tabulated in Table 4. The table reveals that the performance of all the proposed models is good in generalisation water demand data in accordance with Dawson et al.⁸⁸. However, the PSOGA-ANN technique yielded lower RMSE, MAE, MARE, and Max.(error) and higher NSC, NMSE, and R² than the rest of the models. These denoted that the PSOGA-ANN model represented the greatest overall performance in comparison to the other techniques.

Additional testing was conducted on the proposed models to confirm their ability to forecast College Station City's water use. As illustrated in Fig. 6, the correlation coefficient (R) was calculated to compare the simulated and actual water consumption. The target (i.e., measured water consumption) on the x-axis is plotted versus output (i.e., forecast water request) on the y-axis. All the models offer R of more than 0.9 at the testing stage. It is equal to 0.97301, 0.961, 0.96078, 0.92379, and 0.92562 for PSOGA-ANN, PSOGWO-ANN, CPSOCGSA-ANN, MPSON-ANN, and PSO-ANN, respectively. This test confirms the capability of the PSOGA-ANN model to forecast water needs according to the criteria limitation mentioned in the section "Validation of the model." Also, it is clear that the measured and forecasted data for the PSOGA-ANN model reveal a high degree of consistency compared with other hybrid models.

Additionally, Fig. 7 presents the Taylor diagrams that are used for preparing a visual comprehension among observed patterns (Reference) and five simulated patterns performed by PSOGA-ANN, PSOGWO-

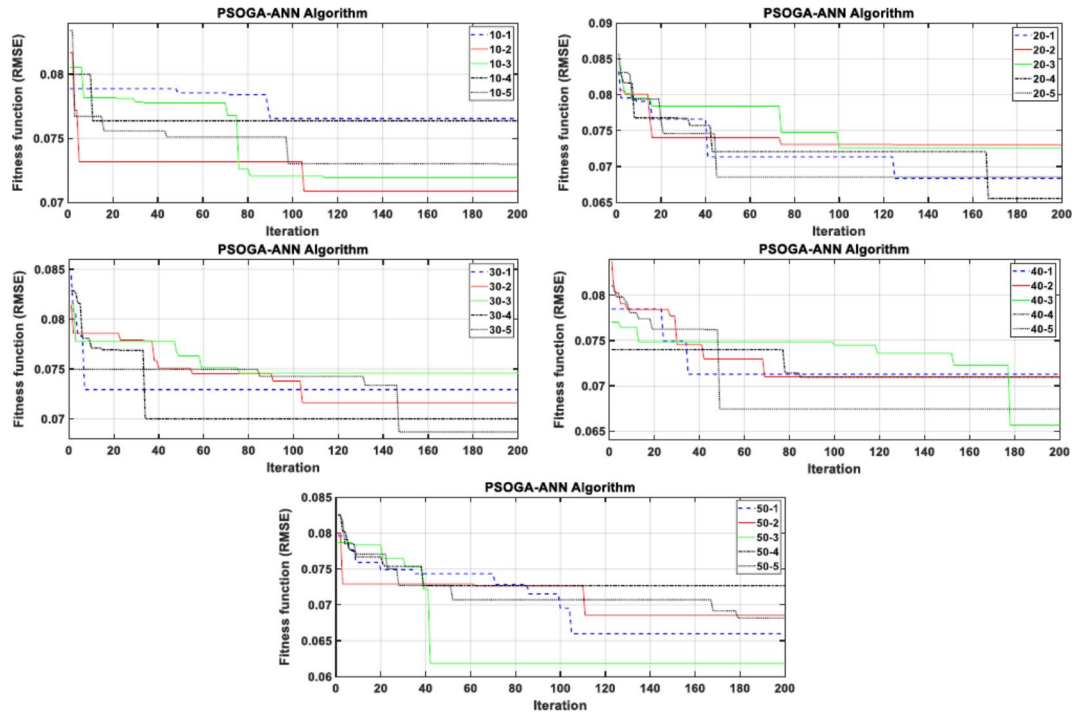


Fig. 4. Evaluation of the PSOGA-ANN algorithm with five independent trials for different population sizes.

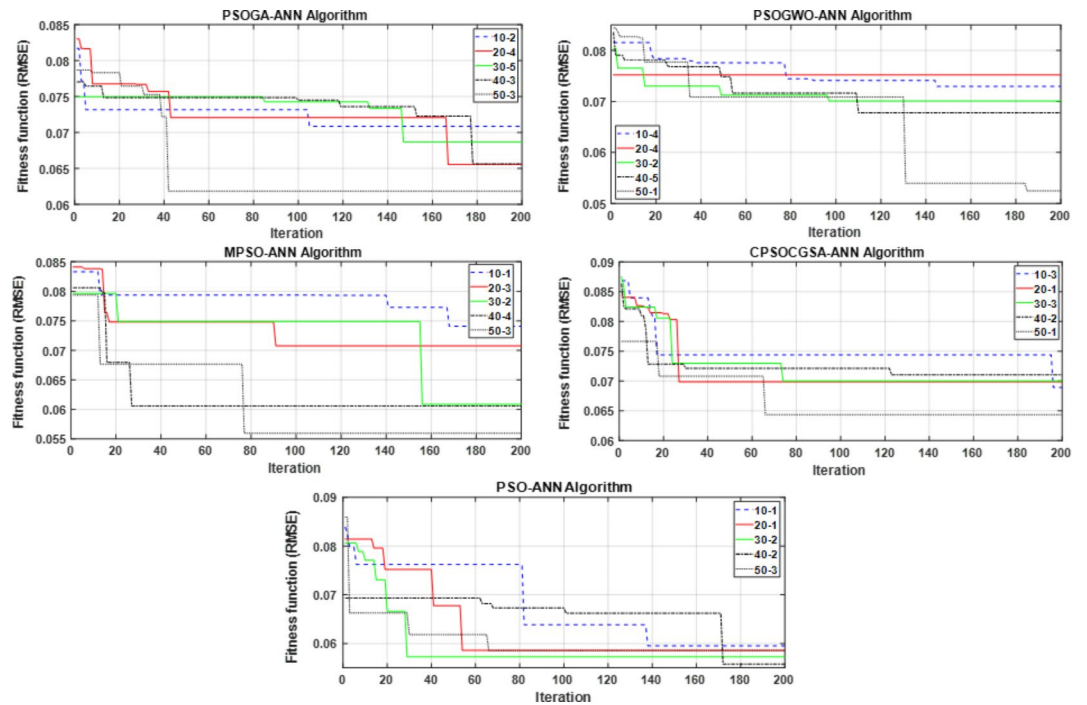


Fig. 5. The best trial for each population size of PSOGA-ANN, PSOGWO-ANN, MPSON-ANN, CPSOCGSA-ANN, and PSO-ANN algorithms is presented.

ANN, CPSOCGSA-ANN, MPSON-ANN, and PSO-ANN models. The diagram considers the root-mean-square difference (RMSD, green contour line), the standard deviation (SD, grey arc), and the correlation coefficient (R, blue azimuthal line). The diagram shows that compared to the other models (i.e. PSOGWO-ANN, CPSOCGSA-ANN, MPSON-ANN, and MPSON-ANN), the PSOGA-ANN model provided the highest R, the lowest SD and RMSD, and the nearest one from the observed pattern (Reference point).

Parameter	Hybrid ANN-metaheuristic models				
	PSOGA	PSOGWO	CPSOCGSA	MPSO	PSO
N1	6	1	8	1	4
N2	3	20	19	20	4
Lr	0.8194	0.9531	0.1511	1	0.2680

Table 3. ANN-designed parameters.

Model	Statistical criteria						
	RMSE (ML)	MAE (ML)	MARE	Max. (error) (ML)	NSC	NMSE	R ²
PSOGA-ANN	0.06745	0.04771	0.00243	0.15245	0.929	0.939	0.947
PSOGWO-ANN	0.07492	0.05597	0.00286	0.17003	0.913	0.899	0.924
CPSOCGSA-ANN	0.07479	0.05285	0.00269	0.17936	0.913	0.904	0.923
MPSO-ANN	0.10206	0.10206	0.00438	0.21568	0.838	0.808	0.853
PSO-ANN	0.09701	0.07217	0.00367	0.19026	0.854	0.828	0.857

Table 4. Performance assessment of four machine learning models for testing stage.

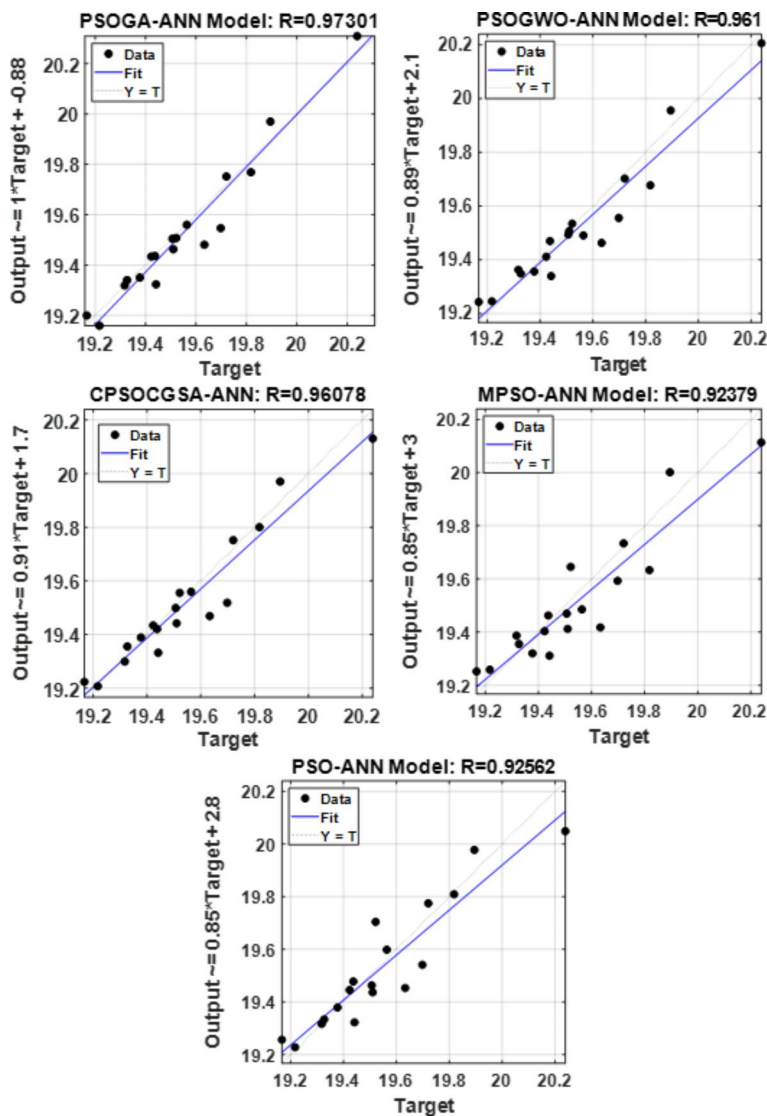


Fig. 6. Correlation coefficient test for all proposed hybrid models.

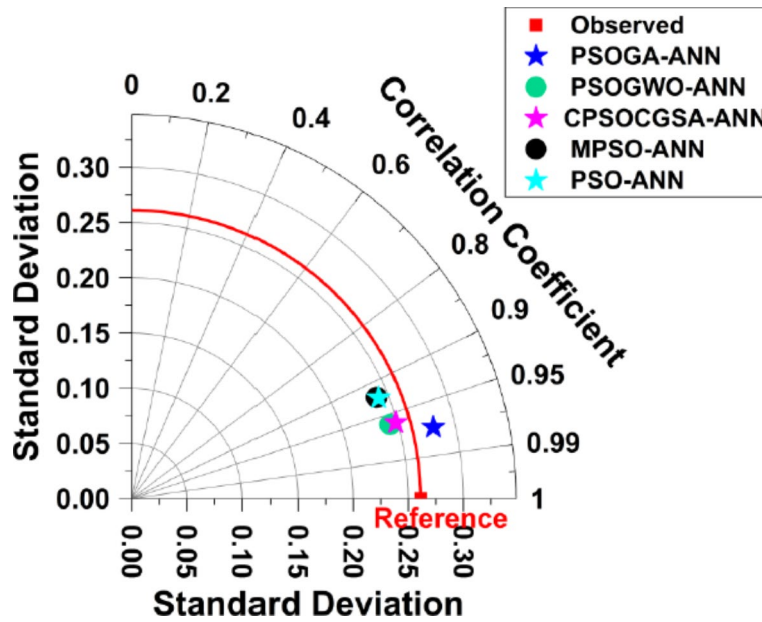


Fig. 7. Taylor diagram for PSOGA-ANN, PSOGWO-ANN, CPSOCGSA-ANN, and MPSO-ANN techniques.

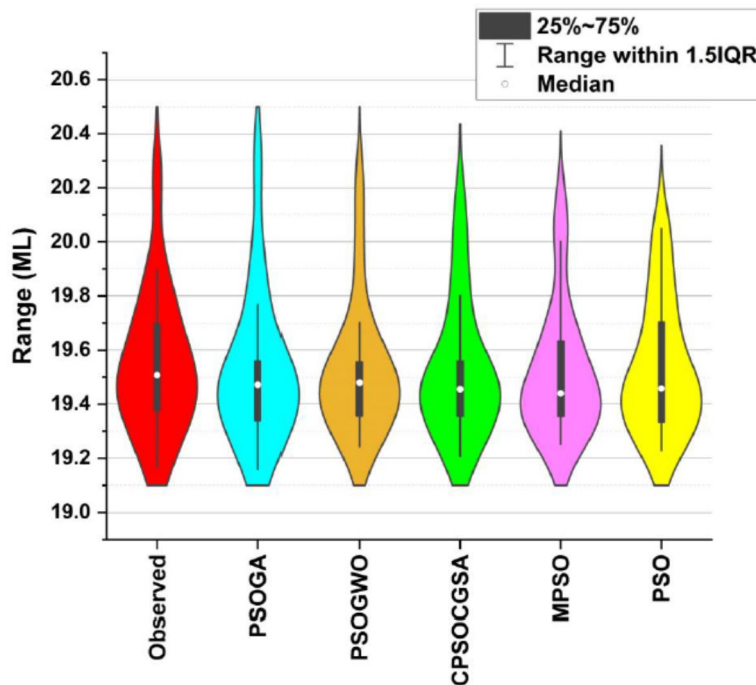


Fig. 8. Violin plot of observed and forecasted urban water demand related to the testing stage of all suggested hybrid models.

While both violin plots and box plots are similar, the former provides more useful information. A single visualisation that integrates the box plot and density trace (or smoothed histogram) to show datasets in a complementary manner is the violin plot. Violin plots, which highlight data clusters, give a clearer picture of the distribution's form. At each point, the violin plot indicates the amount of data that has been acquired; the top tip represents the highest data value, and the bottom tip represents the lowest data value⁹⁰. This section utilises the violin plot diagram to compare the distribution of the measured and forecast water demand datasets during the testing period (Fig. 8). Based on the box plot limitations and violin plot distribution, the ANN-PSOGA model is more in line with the observed urban water demand than that of the other ANN-based techniques.

Figure 9 also shows the testing stage results for the predictions of the PSOGA-ANN, PSOGWO-ANN, CPSOCGSA-ANN, MPSO-ANN and PSO-ANN models. Figure 9A presents the observed and predicted urban water time series data. In terms of pattern (trend+periodicity), the simulated time series from the PSOGA-ANN and PSOGWO-ANN models closely match the observed data. However, PSOGA-ANN performs the best, while MPSO-ANN and PSO-ANN fare the worst regarding the error scale. An error analysis was accomplished in the testing phase to examine the goodness of fit of the five hybrid models. The error scatter plots versus the sample counts for the testing phase are offered in Fig. 9B. It can be seen that relative to the other models, the PSOGA-ANN model's error was much closer to zero (ranging from -0.0746 to 0.1525) ML, while the rest of the error ranges are $(-0.0766$ to $0.1700)$, $(-0.0751$ to $0.1794)$, $(-0.1239$ to $0.2157)$, and $(-0.1829$ to $0.1903)$ ML for PSOGWO-ANN, CPSOCGSA-ANN, MPSO-ANN, and PSO-ANN models, respectively. The distribution also exhibited no discernible trend. According to the results shown above, PSOGA-ANN outperformed other hybrid models in terms of accuracy.

Accordingly, the PSOGA-ANN model outperformed PSOGWO-ANN, CPSOCGSA-ANN, MPSO-ANN, and PSO-ANN models over the entire range of testing data, and the MPSO-ANN model yielded the lowest performance. Finally, to increase the testing of the PSOGA-ANN model, the results of the ADF and KPSS tests indicate that the residual data points are stationary, and the normal distribution of the residuals was confirmed by the Kolmogorov–Smirnov and Shapiro–Wilk tests. Consequently, the distribution of the pattern and values of the residual data endorse the PSOGA-ANN technique's capacity.

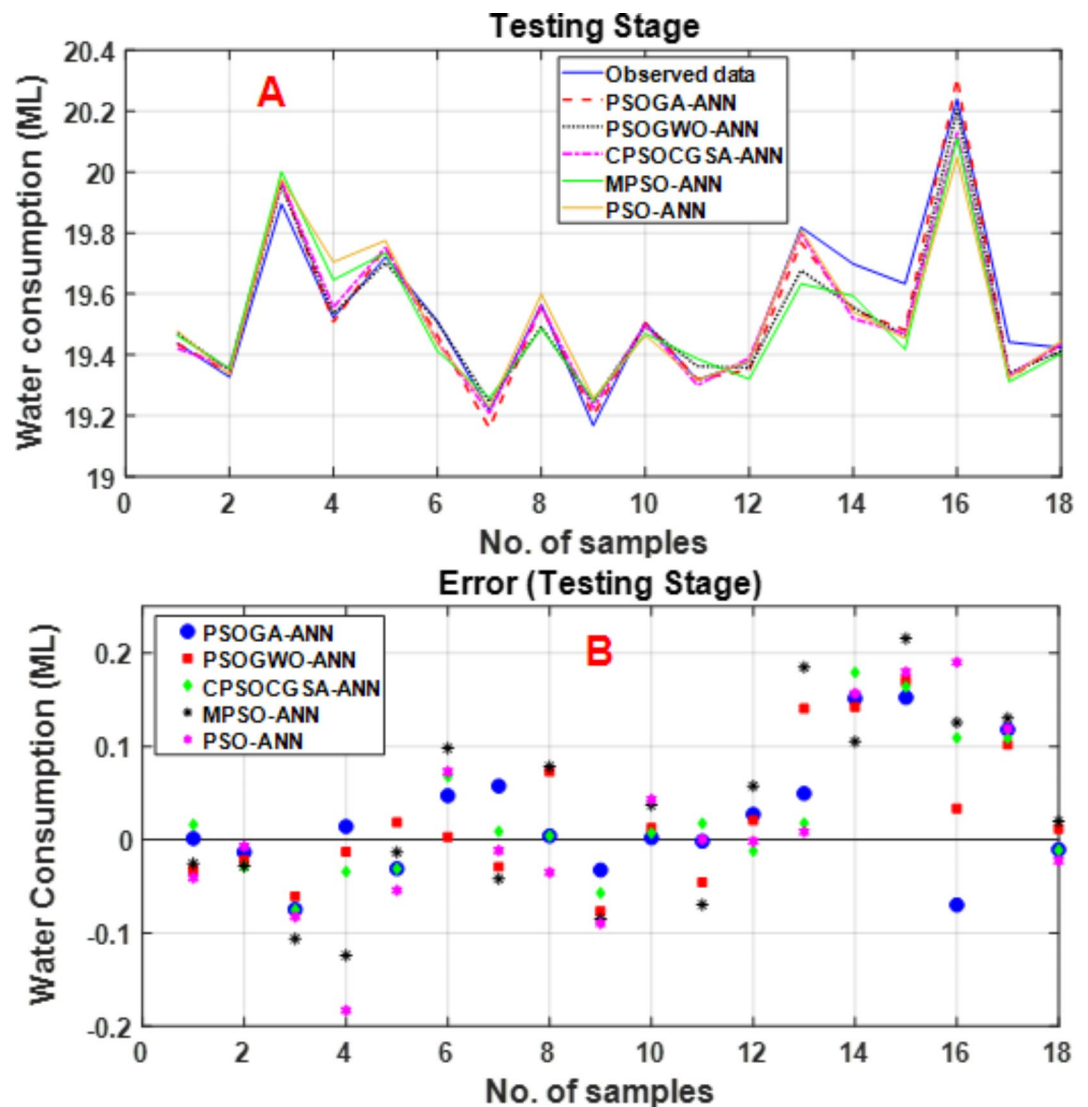


Fig. 9. (A) Observed and forecasted urban water time series comparison and (B) Residual scatterplots for the PSOGA-ANN, PSOGWO-ANN, CPSOCGSA-ANN, MPSO-ANN, and PSO-ANN approaches in the testing phase.

Discussion

The prediction power of ML models in several branches of hydrology—including streamflow²⁸, water quality²⁷, and water level²⁶—has been shown to be enhanced by data preprocessing and hyperparameter optimisation. Considering the limitations of data preprocessing techniques that are reported in “Introduction” section, this paper has examined all the data preprocessing steps. The predictors and target time series underwent three preprocessing methods—normalisation (natural logarithm), a cleaning approach (WT), and choice of the optimum independents’ scenario (tolerance approach)—to maximise forecast accuracy. The WT method aids in removing random noise from the data series. Improved data quality led to higher correlation coefficients between model inputs and output, as seen in Fig. 3. A possible explanation for this result may be that particular attention is paid to normalisation and cleaning data techniques, especially the WT method, which was applied with different mother wavelets and multiple orders, resulting in better improved raw data quality.

We have also considered the consequences of selecting predictors. Accordingly, carefully choosing the predictors will also make use of all the choices in the design space that can improve the performance of the predictions and shed light on which metrological factors have the biggest impact on the output response. Thus, in order to reduce the likelihood of multicollinearity among predictors, only four of the seven climatic factors were chosen using the tolerance technique. These factors (i.e., Tmax, Rain, wind, and RHmax) have tolerance coefficients between 0.272 and 0.454, as tabulated in Table 2. These findings are in keeping with those of earlier research^{82,91}, which demonstrated that reducing computation time and increasing ML model accuracy by choosing predictors using a systematic technique rather than a trial-and-error procedure.

The author’s attention was focused not only on data preprocessing but also on hybrid models of prediction. One big issue with ML techniques is their slow convergence. Another is how difficult it is to avoid local minima. Enhanced ML, thanks to recent advancements in hybrid modelling, has paved the ground for more accuracy in standalone models to be developed in the future²⁸. This paper has presented several solutions to the drawbacks of hybrid models in previous studies mentioned in “Introduction” section. This study compares the single ML model with five MHAs instead of one. Also, these MHAs are related to three inspirations instead of one. Additionally, these MHAs are three hybrid-based and two single-based algorithms. Moreover, each MHA was applied with an iteration equal to 200. Furthermore, each swarm for each MHA was implemented five times.

Since the MHAs utilised follow various strategies depending on their behaviour during the optimisation phase, the hybridisation process results in a wide range of hyperparameter values, creating multiple model scenarios. In this research, three different hybrid-based MHAs (i.e., different combinations based on their behaviour) PSOGA (SI combined with EA), CPSOCGSA (SI combined with P), and PSOGWO (SI combined with SI) were applied to locate the optimum ANN model’s hyperparameters. The performance of hybrid-based MHAs is compared with that of single-based MHA (MPSO and PSO).

According to the findings of the various statistical and graphical tests utilised to evaluate the models during the testing stage (detailed in “Results” section), it is generally accepted that both MHAs (i.e., the single-based and hybrid-based) simulated urban water demand data with good accuracy according to Dawson et al.⁸⁸. It is probable that the improved quality of data made possible by the data preparation technique is responsible for this superiority. Additionally, the optimum solution was likely found due to running the swarm of each algorithm five times, resulting in a more robust prediction range and less uncertainty. Moreover, hybrid-based MHAs (PSOGA, PSOGWO, and CPSOCGSA) outperformed the single-based MHA (MPSO and PSO). These results confirm the findings of prior studies^{92,93} that MHAs that use a hybrid-based approach outperform those that rely on a single-based approach. However, the PSOGA-ANN algorithm is superior to generalise water demand data compared with other MHAs in the testing stage. Evidence for this can be seen in Table 4, which displays the outcomes of RMSE, MAE, MARE, NSC, NSME, and R² after only a few iterations throughout the optimisation procedure (Fig. 5). Another important finding, as shown in Fig. 7, was that the hybrid-based MHAs (PSOGWO and CPSOCGSA) function similarly. Similarly, the single-based MHAs (MPSO and PSO) performance is rather close.

These results lend credence to the claim made in the previous studies^{25,94} that avoidance of local minima is possible with hybrid-based algorithms, leading to improved accuracy, stability, and reliability in solving real-world issues. In terms of future research, it would be useful to extend the current findings by examining different kinds of data preprocessing techniques. Much work remains to be done before fully understanding the extent of hybrid-based MHAs’ performance with other ML techniques is established. Also, there is a need for research studies that explore the hybrid-based MHAs’ performance with long-term datasets.

Moreover, an urban water company might theoretically utilise short-term weather forecast data to optimise the schedule of water production if they calibrate this model with their data. A water company, for instance, could lower its energy use by shifting more production to shoulder and off-peak hours if short-term weather projections called for cooler temperatures.

Conclusion

In the last few decades, as water supplies have become scarcer and human consumption of water has rapidly grown, water utilities have focused heavily on developing more accurate methods of predicting future water needs. Motivated by data collected for College Station City, USA, over a decade, this study aimed to find a novel methodology to improve the prediction of urban water demand based on meteorological variables. This paper compared the performance of five MHAs, three of which were hybrid-based and two of which were single-based MHAs for urban water demand forecasting. The hybrid-based MHAs are PSOGA (SI combined with EA), CPSOCGSA (SI combined with P), and PSOGWO (SI combined with SI), and the single-based MHA are MPSO and PSO. These MHAs’ performance have not been compared before in terms of urban water demand. The methodology contains three combined techniques, including data preprocessing (WT and tolerance) and a

prediction model (ANN that is integrated by PSOGA). The results of PSOGA-ANN were compared with those of the CPSOCGSA-ANN, PSOGWO, MPSO, and PSO algorithms.

In light of the findings, it was concluded that data pre-processing is an appropriate strategy for enhancing data quality (denoising) via the use of WT and for identifying the optimal set of predictors using tolerance. The optimum scenario of the predictors in which the multicollinearity condition is not violated was provided by T_{max}, Rain, wind, and Rh_{max}. The performance of the PSOGA-ANN outperformed all other suggested models over the entire range of testing data stages, and the single-based models yielded the lowest performance. The PSOGA-ANN algorithm yielded R², NSC, NMSE, RMSE, and MAE, of 0.947, 0.929, 0.939, 0.06745 and 0.04771 megalitres, respectively. These results indicate that the proposed methodology offers a guide to choosing appropriate predictors controlling water demand.

These conclusions have significant implications for policymakers and managers as they plan for, evaluate, and compare the accessibility of potable water supplies and growing water needs. Lastly, this research filled a gap in the literature by examining the quality and uncertainty of data analytic machine learning techniques for predicting monthly urban water demand considering weather variables. More research is needed to determine how various meteorological variables impact water demand prediction at different scales.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Received: 24 May 2024; Accepted: 12 September 2024

Published online: 14 October 2024

References

- Guo, G. *et al.* Short-term water demand forecast based on deep learning method. *J. Water Resour. Plan. Manag.* [https://doi.org/10.1061/\(asce\)wr.1943-5452.0000992](https://doi.org/10.1061/(asce)wr.1943-5452.0000992) (2018).
- Sharafati, A., Asadollah, S. B. H. S. & Shahbazi, A. Assessing the impact of climate change on urban water demand and related uncertainties: A case study of Neyshabur, Iran. *Theor. Appl. Climatol.* **145**, 473–487. <https://doi.org/10.1007/s00704-021-03638-5> (2021).
- Danilenko, A., Dickson, E. & Jacobsen, M. *Climate Change and Urban Water Utilities: Challenges and Opportunities* (World Bank, 2010).
- Fan, L., Liu, G., Wang, F., Ritsema, C. J. & Geissen, V. Domestic water consumption under intermittent and continuous modes of water supply. *Water Resour. Manag.* **28**, 853–865. <https://doi.org/10.1007/s11269-014-0520-7> (2014).
- Zubaidi, S. L. *et al.* Assessing the benefits of nature-inspired algorithms for the parameterization of ANN in the prediction of Water demand. *J. Water Resour. Plan. Manag.* **149**, 1–10. [https://doi.org/10.1061/\(asce\)wr.1943-5452.0001602](https://doi.org/10.1061/(asce)wr.1943-5452.0001602) (2023).
- Capt, T., Mirchi, A., Kumar, S. & Walker, W. S. Urban water demand: Statistical optimization approach to modeling daily demand. *J. Water Resour. Plan. Manag.* [https://doi.org/10.1061/\(asce\)wr.1943-5452.0001315](https://doi.org/10.1061/(asce)wr.1943-5452.0001315) (2021).
- DeMaagd, N. & Roberts, M. J. How will climate change affect residential water demand? Evidence from Hawai'i microclimates. *Water Econ. Policy* **7**, 1–51. <https://doi.org/10.1142/S2382624X21500053> (2021).
- Tiwari, M. K. & Adamowski, J. F. An ensemble wavelet bootstrap machine learning approach to water demand forecasting: A case study in the city of Calgary, Canada. *Urban Water J.* **14**, 185–201. <https://doi.org/10.1080/1573062X.2015.1084011> (2017).
- Miaou, S. P. A class of time series urban water demand models with nonlinear climatic effects. *Water Resour. Res.* **26**, 169–178. <https://doi.org/10.1029/WR026i002p00169> (1990).
- Huang, H., Zhang, Z. & Song, F. An ensemble-learning-based method for short-term water demand forecasting. *Water Resour. Manag.* **35**, 1757–1773. <https://doi.org/10.1007/s11269-021-02808-4> (2021).
- Lu, H., Matthews, J. & Han, S. A hybrid model for monthly water demand prediction: a case study of Austin, Texas. *AWWA Water Sci.* <https://doi.org/10.1002/aws2.1175> (2020).
- Chen, G., Long, T., Xiong, J. & Bai, Y. Multiple random forests modelling for urban water consumption forecasting. *Water Resour. Manag.* **31**, 4715–4729. <https://doi.org/10.1007/s11269-017-1774-7> (2017).
- Shabani, S., Candelieri, A., Archetti, F. & Naser, G. Gene expression programming coupled with unsupervised learning: A two-stage learning process in multi-scale, short-term water demand forecasts. *Water* <https://doi.org/10.3390/w10020142> (2018).
- Zanfei, A., Brentan, B. M., Menapace, A. & Righetti, M. A short-term water demand forecasting model using multivariate long short-term memory with meteorological data. *J. Hydroinform.* **24**, 1053–1065. <https://doi.org/10.2166/hydro.2022.055> (2022).
- Ghandehari, A., Davary, K., Khorasani, H. O., Vatanparast, M. & Pourmohamad, Y. Assessment of urban water supply options by using fuzzy possibilistic theory. *Environ. Process.* **7**, 949–972. <https://doi.org/10.1007/s40710-020-00441-8> (2020).
- House-Peters, L. A. & Chang, H. Urban water demand modeling: Review of concepts, methods, and organising principles. *Water Resour. Res.* **47**, 1–15. <https://doi.org/10.1029/2010wr009624> (2011).
- Donkor, E. A., Mazzuchi, T. H., Soyer, R. & Roberson, J. A. Urban water demand forecasting: Review of methods and models. *J. Water Resour. Plan. Manag.* **140**, 146–159. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452](https://doi.org/10.1061/(ASCE)WR.1943-5452) (2014).
- Ghalekhondabi, I., Ardjmand, E., Young, I. I. & Weckman, W. A. Water demand forecasting: Review of soft computing methods. *Environ. Monit. Assess.* **189**, 1–13. <https://doi.org/10.1007/s10661-017-6030-3> (2017).
- De Souza Groppo, G., Costa, M. A. & Libânio, M. Predicting water demand: A review of the methods employed and future possibilities. *Water Supply* **19**, 2179–2198. <https://doi.org/10.2166/ws.2019.122> (2019).
- Rahim, M. S., Nguyen, K. A., Stewart, R. A., Giurco, D. & Blumenstein, M. Machine learning and data analytic techniques in digital water metering: A review. *Water* **12**, 1–27. <https://doi.org/10.3390/w12010294> (2020).
- Ren, T., Liu, X., Niu, J., Lei, X. & Zhang, Z. Real-time water level prediction of cascaded channels based on multilayer perception and recurrent neural network. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2020.124783> (2020).
- Xenochristou, M. & Kapelan, Z. An ensemble stacked model with bias correction for improved water demand forecasting. *Urban Water J.* **17**, 212–223. <https://doi.org/10.1080/1573062x.2020.1758164> (2020).
- Yaghoubzadeh-Bavandpour, A., Bozorg-Haddad, O., Rajabi, M., Zolghadr-Asli, B. & Chu, X. Application of swarm intelligence and evolutionary computation algorithms for optimal reservoir operation. *Water Resour. Manag.* **36**, 2275–2292. <https://doi.org/10.1007/s11269-022-03141-0> (2022).
- Ehteram, M. *et al.* Design of a hybrid ANN multi-objective whale algorithm for suspended sediment load prediction. *Environ. Sci. Pollut. Res. Int.* **28**, 1596–1611. <https://doi.org/10.1007/s11356-020-10421-y> (2021).
- Almubaidin, M. A. A., Ahmed, A. N., Sidek, L. B. M. & Elshafee, A. Using metaheuristics algorithms (MHAs) to optimise water supply operation in reservoirs: A review. *Arch. Comput. Methods Eng.* **29**, 3677–3711. <https://doi.org/10.1007/s11831-022-09716-9> (2022).

26. Mohammed, S. J. *et al.* Application of hybrid machine learning models and data pre-processing to predict water level of watersheds: Recent trends and future perspective. *Cogent Eng.* <https://doi.org/10.1080/23311916.2022.2143051> (2022).
27. Khudhair, Z. S. *et al.* A review of hybrid soft computing and data pre-processing techniques to forecast freshwater quality's parameters: Current trends and future directions. *Environments* <https://doi.org/10.3390/environments9070085> (2022).
28. Abdul Kareem, B., Zubaidi, L., Al-Ansari, S., Raad, N. & Muhsen, Y. Review of recent trends in the hybridisation of preprocessing-based and parameter optimisation-based hybrid models to forecast univariate streamflow. *Comput. Model. Eng. Sci.* **138**, 1–41. <https://doi.org/10.32604/cmescs.2023.027954> (2024).
29. Merchaoui, M., Sakly, A. & Mimouni, M. F. Particle swarm optimisation with adaptive mutation strategy for photovoltaic solar cell/module parameter extraction. *Energy Conv. Manag.* **175**, 151–163. <https://doi.org/10.1016/j.enconman.2018.08.081> (2018).
30. Chen, H., Jiao, S., Wang, M., Heidari, A. A. & Zhao, X. Parameters identification of photovoltaic cells and modules using diversification-enriched Harris hawks optimisation with chaotic drifts. *J. Clean. Prod.* <https://doi.org/10.1016/j.jclepro.2019.118778> (2020).
31. Ridha, H. M. Parameters extraction of single and double diodes photovoltaic models using marine predators algorithm and Lambert W function. *Sol. Energy* **209**, 674–693. <https://doi.org/10.1016/j.solener.2020.09.047> (2020).
32. Adnan, R. M. *et al.* Estimating reference evapotranspiration using hybrid adaptive fuzzy inferencing coupled with heuristic algorithms. *Comput. Electron. Agric.* <https://doi.org/10.1016/j.compag.2021.106541> (2021).
33. Rather, S. A. & Bala, P. S. Hybridization of constriction coefficient-based particle swarm optimization and chaotic gravitational search algorithm for solving engineering design problems. In *Appl. Soft Comput. Commun. Netw.* Vol. 125 95–115 (Springer, Singapore, 2020).
34. Črepinšek, M., Liu, S. H. & Mernik, M. Exploration and exploitation in evolutionary algorithms. *ACM Comput. Surv.* **45**, 1–33. <https://doi.org/10.1145/2480741.2480752> (2013).
35. Eiben, A. E. & Schippers, C. A. On evolutionary exploration and exploitation. *Fundam. Inform.* **35**, 35–50 (1998).
36. Adetunji, K. E., Hofsajer, I. W., Abu-Mahfouz, A. M. & Cheng, L. A. Review of metaheuristic techniques for optimal integration of electrical units in distribution networks. *IEEE Access* **9**, 5046–5068. <https://doi.org/10.1109/access.2020.3048438> (2021).
37. Tabachnick, B. G. & Fidell, L. S. *Using Multivariate Statistics* 6th edn. (Pearson Education, Inc, 2013).
38. Zamili, H., Bakan, G., Zubaidi, S. L. & Alawsi, M. A. Water quality index forecast using artificial neural network techniques optimised with different metaheuristic algorithms. *Model. Earth Syst. Environ.* **9**, 4323–4333. <https://doi.org/10.1007/s40808-023-01750-1> (2023).
39. Zhang, Y., Yang, H., Cui, H. & Chen, Q. Comparison of the ability of ARIMA, WNN and SVM models for drought forecasting in the Sanjiang Plain, China. *Nat. Resour. Res.* **29**, 1447–1464. <https://doi.org/10.1007/s11053-019-09512-6> (2019).
40. Khan, M. M. H., Muhammad, N. S. & El-Shafie, A. Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2020.125380> (2020).
41. Kisi, O., Docheshmeh Gorgij, A., Zounemat-Kermani, M., Mahdavi-Meymand, A. & Kim, S. Drought forecasting using novel heuristic methods in a semi-arid environment. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2019.124053> (2019).
42. Banadkooki, F. B., Singh, V. P. & Ehteram, M. Multi-timescale drought prediction using new hybrid artificial neural network models. *Nat. Hazards* **106**, 2461–2478. <https://doi.org/10.1007/s11069-021-04550-x> (2021).
43. Niu, W. J., Feng, Z. K., Yang, W. F. & Zhang, J. Short-term streamflow time series prediction model by machine learning tool based on data preprocessing technique and swarm intelligence algorithm. *Hydrol. Sci. J.* **65**, 2590–2603. <https://doi.org/10.1080/02626667.2020.1828889> (2020).
44. Zhao, X. *et al.* Enhancing robustness of monthly streamflow forecasting model using gated recurrent unit based on improved grey wolf optimiser. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2021.126607> (2021).
45. Ehteram, M. *et al.* Hybridization of artificial intelligence models with nature inspired optimisation algorithms for lake water level prediction and uncertainty analysis. *Alex. Eng. J.* **60**, 2193–2208. <https://doi.org/10.1016/j.aej.2020.12.034> (2021).
46. Riahi-Madvar, H., Dehghani, M., Memarzadeh, R. & Gharabaghi, B. Short to long-term forecasting of river flows by heuristic optimization algorithms hybridised with ANFIS. *Water Resour. Manag.* **35**, 1149–1166. <https://doi.org/10.1007/s11269-020-02756-5> (2021).
47. Zhu, B. *et al.* Hybrid particle swarm optimisation with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. *Comput. Electron. Agric.* <https://doi.org/10.1016/j.compag.2020.105430> (2020).
48. Yan, S. *et al.* A novel hybrid WOA-XGB model for estimating daily reference evapotranspiration using local and external meteorological data: Applications in arid and humid regions of China. *Agric. Water Manag.* <https://doi.org/10.1016/j.agwat.2020.106594> (2021).
49. Dong, J. *et al.* Comparison of four bio-inspired algorithms to optimise KNEA for predicting monthly reference evapotranspiration in different climate zones of China. *Comput. Electron. Agric.* <https://doi.org/10.1016/j.compag.2021.106211> (2021).
50. Malik, A. *et al.* Support vector regression integrated with novel meta-heuristic algorithms for meteorological drought prediction. *Meteorol. Atmos. Phys.* **133**, 891–909. <https://doi.org/10.1007/s00703-021-00787-0> (2021).
51. Tripura, J., Roy, P. & Barbhuiya, A. K. Simultaneous streamflow forecasting based on hybridised neuro-fuzzy method for a river system. *Neural Comput. Appl.* **33**, 3221–3233. <https://doi.org/10.1007/s00521-020-05194-x> (2020).
52. Deng, B. *et al.* Advanced water level prediction for a large-scale river-lake system using hybrid soft computing approach: A case study in Dongting Lake, China. *Earth Sci. Inf.* **14**, 1987–2001. <https://doi.org/10.1007/s12145-021-00665-8> (2021).
53. Karami, H. *et al.* Multi-reservoir system optimization based on hybrid gravitational algorithm to minimise water-supply deficiencies. *Water Resour. Manag.* **33**, 2741–2760. <https://doi.org/10.1007/s11269-019-02238-3> (2019).
54. Xu, X., Wang, X., Li, Y. & Cao, N. Prediction of the height of water flowing fractured zone based on the MPSO-BP neural network model. *Math. Probl. Eng.* <https://doi.org/10.1155/2022/2133695> (2022).
55. Danandeh Mehr, A., Ghadimi, S., Marttila, H. & Torabi Haghighi, A. A new evolutionary time series model for streamflow forecasting in boreal lake-river systems. *Theor. Appl. Climatol.* **148**, 255–268. <https://doi.org/10.1007/s00704-022-03939-3> (2022).
56. Chang, J., Bai, T., Huang, Q. & Yang, D. Optimisation of water resources utilisation by PSO-GA. *Water Resour. Manag.* **27**, 3525–3540. <https://doi.org/10.1007/s11269-013-0362-8> (2013).
57. Akbari, R., Hessami-Kermani, M. R. & Shojaei, S. Flood routing: Improving outflow using a new non-linear Muskingum model with four variable parameters coupled with PSO-GA algorithm. *Water Resour. Manag.* **34**, 3291–3316. <https://doi.org/10.1007/s11269-020-02613-5> (2020).
58. Khudhair, Z. S., Zubaidi, S. L., Al-Bugharbee, H., Al-Ansari, N. & Ridha, H. M. A CPSOCGSA-tuned neural processor for forecasting river water salinity: Euphrates river, Iraq. *Cogent Eng.* <https://doi.org/10.1080/23311916.2022.2150121> (2022).
59. Meshram, S. G., Ghorbani, M. A., Shamshirband, S., Karimi, V. & Meshram, C. River flow prediction using hybrid PSOGSA algorithm based on feed-forward neural network. *Soft. Comput.* **23**, 10429–10438. <https://doi.org/10.1007/s00500-018-3598-7> (2018).
60. Khairan, H. E. *et al.* Assessing the potential of hybrid-based metaheuristic algorithms integrated with ANNs for accurate reference evapotranspiration forecasting. *Sustainability* <https://doi.org/10.3390/su151914320> (2023).
61. Samany, N. N., Sheybani, M. & Zlatanova, S. Detection of safe areas in flood as emergency evacuation stations using modified particle swarm optimisation with local search. *Appl. Soft Comput.* <https://doi.org/10.1016/j.asoc.2021.107681> (2021).
62. Yang, X., Maihemuti, B., Simayi, Z., Saydi, M. & Na, L. Prediction of glacially derived runoff in the Muzati River watershed based on the PSO-LSTM model. *Water* <https://doi.org/10.3390/w14132018> (2022).

63. Smolak, K. *et al.* Applying human mobility and water consumption data for short-term water demand forecasting using classical and machine learning models. *Urban Water J.* **17**, 32–42. <https://doi.org/10.1080/1573062x.2020.1734947> (2020).
64. Shah, S., Ben Miled, Z., Schaefer, R. & Berube, S. Differential learning for outliers: A case study of water demand prediction. *Appl. Sci.* <https://doi.org/10.3390/app8112018> (2018).
65. Cleophas, T. J. & Zwinderman, A. H. *SPSS for Starters and 2nd Levelers* 2nd edn. (Springer, 2016). <https://doi.org/10.1007/978-3-319-20600-4>.
66. Dohan, K. & Whitfield, P. Identification and characterization of water quality transients using wavelet analysis. I. Wavelet analysis methodology. *Water Sci. Technol.* **36**, 325–335. [https://doi.org/10.1016/S0273-1223\(97\)00490-3](https://doi.org/10.1016/S0273-1223(97)00490-3) (1997).
67. Okkan, U. & Ali Serbes, Z. The combined use of wavelet transform and black box models in reservoir inflow modeling. *J. Hydrol. Hydromech.* **61**, 112–119. <https://doi.org/10.2478/johh-2013-0015> (2013).
68. Agarwal, M. & Srivastava, G. M. S. Genetic algorithm-enabled particle swarm optimisation (PSOGA)-based task scheduling in cloud computing environment. *Int. J. Inform. Technol. Decis. Mak.* **17**, 1237–1267. <https://doi.org/10.1142/S0219622018500244> (2018).
69. Kim, K. & Han, I. Genetic algorithms approach to feature discretisation in artificial neural networks for the prediction of stock price index. *Expert Syst. Appl.* **19**, 125–132. [https://doi.org/10.1016/S0957-4174\(00\)00027-0](https://doi.org/10.1016/S0957-4174(00)00027-0) (2000).
70. Koza, J. R. Genetic programming as a means for programming computers by natural selection. *Stat. Comput.* **4**, 87–112. <https://doi.org/10.1007/BF00175355> (1994).
71. Mirjalili, S. Genetic algorithm. *Evol. Algorithms Neural Netw. Theory Appl.* **780**, 43–55. https://doi.org/10.1007/978-3-319-93025-1_4 (2019).
72. Rather, S. A. & Bala, P. S. Hybridization of constriction coefficient-based particle swarm optimization and chaotic gravitational search algorithm for solving engineering design problems. In *Applied Soft Computing and Communication Networks: Proceedings of ACN 2019* (eds. Thampi, S. M., Sherly, E., Dasgupta, S., Lloret Mauri, J., Abawajy, J. H., Khorov, E., Mathew, J.) 95–115 (Springer, 2020). https://doi.org/10.1007/978-981-15-3852-0_7
73. Mirjalili, S., Mirjalili, S. M. & Lewis, A. Grey Wolf optimizer. *Adv. Eng. Softw.* **69**, 46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007> (2014).
74. Hatta, N. M., Zain, A. M., Sallehuddin, R., Shayfull, Z. & Yusoff, Y. Recent studies on optimisation method of Grey Wolf Optimiser (GWO): A review (2014–2017). *Artif. Intell. Rev.* **52**, 2651–2683. <https://doi.org/10.1007/s10462-018-9634-2> (2019).
75. Muro, C., Escobedo, R., Spector, L. & Coppinger, R. P. Wolf-pack (*Canis lupus*) hunting strategies emerge from simple rules in computational simulations. *Behav. Process.* **88**, 192–197. <https://doi.org/10.1016/j.beproc.2011.09.006> (2011).
76. Hossain, M. A., Pota, H. R., Squartini, S., Zaman, F. & Guerrero, J. M. Energy scheduling of community microgrid with battery cost using particle swarm optimisation. *Appl. Energy* **254**, 113723. <https://doi.org/10.1016/j.apenergy.2019.113723> (2019).
77. Şenel, F. A., Gökçe, F., Yüksel, A. S. & Yiğit, T. A novel hybrid PSO–GWO algorithm for optimisation problems. *Eng. Comput.* **35**, 1359–1373. <https://doi.org/10.1007/s00366-018-0668-5> (2019).
78. Suman, G. K., Guerrero, J. M. & Roy, O. P. Optimisation of solar/wind/bio-generator/diesel/battery based microgrids for rural areas: A PSO–GWO approach. *Sustain. Cities Soc.* **67**, 102723. <https://doi.org/10.1016/j.scs.2021.102723> (2021).
79. Zounemat-Kermani, M. *et al.* Neurocomputing in surface water hydrology and hydraulics: A review of two decades retrospective, current status and future prospects. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2020.125085> (2020).
80. Shirkoohi, M. G., Doghri, M. & Duchesne, S. Short-term water demand predictions coupling an Artificial neural network model and a genetic algorithm. *Water Supply* <https://doi.org/10.2166/ws.2021.049> (2021).
81. Sadeghifar, T., Lama, G. F. C., Sihag, P., Bayram, A. & Kisi, O. Wave height predictions in complex sea flows through soft-computing models: Case study of Persian Gulf. *Ocean Eng.* <https://doi.org/10.1016/j.oceaneng.2021.110467> (2022).
82. Shah, M. I., Javed, M. F., Alqahtani, A. & Aldrees, A. Environmental assessment based surface water quality prediction using hyper-parameter optimised machine learning models based on consistent big data. *Process Saf. Environ. Prot.* **151**, 324–340. <https://doi.org/10.1016/j.psep.2021.05.026> (2021).
83. Nunes Carvalho, T. M., de Souza Filho, F. A. & Porto, V. C. Urban water demand modeling using machine learning techniques: Case study of Fortaleza, Brazil. *J. Water Resour. Plan. Manag.* [https://doi.org/10.1061/\(asce\)wr.1943-5452.0001310](https://doi.org/10.1061/(asce)wr.1943-5452.0001310) (2021).
84. Mohammed, S. J. *et al.* Hybrid technique to improve the river water level forecasting using artificial neural network-based marine predators algorithm. *Adv. Civ. Eng.* **2022**, 1–14. <https://doi.org/10.1155/2022/6955271> (2022).
85. Monteiro, R. V. A., Guimarães, G. C., Moura, F. A. M., Albertini, M. R. M. C. & Albertini, M. K. Estimating photovoltaic power generation: Performance analysis of artificial neural networks, support vector machine and Kalman filter. *Electr. Power Syst. Res.* **143**, 643–656. <https://doi.org/10.1016/j.epsr.2016.10.050> (2017).
86. Seo, Y., Kwon, S. & Choi, Y. Short-term water demand forecasting model combining variational mode decomposition and extreme learning machine. *Hydrology* **5**, 1–19. <https://doi.org/10.3390/hydrology5040054> (2018).
87. Li, M. W. *et al.* Optimisation approach of berth-quay crane-truck allocation by the tide, environment and uncertainty factors based on chaos quantum adaptive seagull optimisation algorithm. *Appl. Soft Comput.* <https://doi.org/10.1016/j.asoc.2023.111197> (2024).
88. Dawson, C. W., Abraham, R. J. & See, L. M. HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environ. Model. Softw.* **22**, 1034–1052. <https://doi.org/10.1016/j.envsoft.2006.06.008> (2007).
89. Pan, M. *et al.* Water level prediction model based on GRU and CNN. *IEEE Access* **8**, 60090–60100. <https://doi.org/10.1109/access.2020.2982433> (2020).
90. Kikon, A., Dodamani, B. M., Barma, S. D. & Naganna, S. R. ANFIS-based soft computing models for forecasting effective drought index over an arid region of India. *AQUA Water Infrastruct. Ecosyst. Soc.* **72**, 930–946. <https://doi.org/10.2166/aqua.2023.204> (2023).
91. Karbasi, M., Jamei, M., Ali, M., Malik, A. & Yaseen, Z. M. Forecasting weekly reference evapotranspiration using Auto Encoder Decoder Bidirectional LSTM model hybridised with a Boruta-CatBoost input optimiser. *Comput. Electron. Agric.* <https://doi.org/10.1016/j.compag.2022.107121> (2022).
92. Zeinolabedini Rezaabad, M., Ghazanfari, S., Salajegheh, M. A. N. F. I. S. & Modeling with ICA., BBO, TLBO, and IWO optimization algorithms and sensitivity analysis for predicting daily reference evapotranspiration. *J. Hydrol. Eng.* [https://doi.org/10.1061/\(asce\)he.1943-5584.0001963](https://doi.org/10.1061/(asce)he.1943-5584.0001963) (2020).
93. El-Kenawy, E. M. *et al.* Improved weighted ensemble learning for predicting the daily reference evapotranspiration under the semi-arid climate conditions. *Environ. Sci. Pollut. Res. Int.* <https://doi.org/10.1007/s11356-022-21410-8> (2022).
94. Ridha, H. M. *et al.* A novel theoretical and practical methodology for extracting the parameters of the single and double diode photovoltaic models. *IEEE Access* **10**, 11110–11137. <https://doi.org/10.1109/access.2022.3142779> (2022).

Author contributions

Salah L. Zubaidi: Methodology, software, formal analysis, investigation, writing—original draft preparation. Hussein Al-Bugharbee: Methodology, software, formal analysis, writing—original draft preparation. Ali W. Alatabi: validation, writing—original draft preparation. Hussein Mohammed Ridha: validation, writing—original draft preparation. Khalid Hashim: formal analysis, writing—original draft preparation, J.A.; writing—review and editing. Nadhir Al-Ansari: validation, funding acquisition, writing—original draft preparation. Zaher Mundher Yaseen: validation, writing—original draft preparation.

Funding

Open access funding provided by Lulea University of Technology.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.L.Z. or N.A.-A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024