Enhancement of nitrogen prediction accuracy through a new hybrid model using ant colony optimization and an Elman neural network

Pavitra Kumar\textsuperscript{a}, Sai Hin Lai\textsuperscript{a}, Nuruol Syuhadaa Mohda\textsuperscript{a}, Md Rowshon Kamal\textsuperscript{b}, Ali Najah Ahmed\textsuperscript{c,}\textsuperscript{d}, Mohsen Sheriff\textsuperscript{e,}\textsuperscript{f}, Ahmed Sefelnasr\textsuperscript{c}\textsuperscript{,}\textsuperscript{d} and Ahmed El-shafie\textsuperscript{a,}\textsuperscript{e,}\textsuperscript{f}

\textsuperscript{a}Department of Civil Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur, Malaysia; \textsuperscript{b}Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, Selangor, Malaysia; \textsuperscript{c}Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), Kajang, Malaysia; \textsuperscript{d}Institute for Energy Infrastructure (IEI), Universiti Tenaga Nasional (UNITEN), Kajang, Malaysia; \textsuperscript{e}National Water and Energy Center, United Arab Emirates University, Al Ain, United Arab Emirates; \textsuperscript{f}Civil and Environmental Engineering Department, College of Engineering, United Arab Emirates University, Al Ain, United Arab Emirates

ABSTRACT

Advanced human activities, including modern agricultural practices, are responsible for alteration of natural concentration of nitrogen compounds in rivers. Future prediction of nitrogen compound concentrations (especially nitrate-nitrogen and ammonia-nitrogen) are important for countries where household water is obtained from rivers after treatment. Increased concentrations of nitrogen compounds result in the suspension of household water supplies. Artificial Neural Networks (ANNs) have already been deployed for the prediction of nitrogen compounds in various countries. But standalone ANN have several limitations. However, the limitations of ANNs can be resolved using hybrid models. This study proposes a new ACO-ENN hybrid model developed by integrating Ant Colony Optimization (ACO) with an Elman Neural Network (ENN). The developed ACO-ENN hybrid model was used to improve the prediction results of nitrate-nitrogen and ammonia-nitrogen prediction models. The results of new hybrid models were compared with multilayer ANN models and standalone ENN models. There was a significant improvement in the mean square errors (MSE) (0.196 → 0.049 → 0.012, i.e. ANN → ENN → Hybrid), mean absolute errors (MAE) (0.271 → 0.094 → 0.069) and Nash–Sutcliffe efficiencies (NSE) (0.7255 → 0.9321 → 0.984). The hybrid model had outstanding performance compared with the ANN and ENN models. Hence, the prediction accuracy of nitrate-nitrogen and ammonia-nitrogen has been improved using new ACO-ENN hybrid model.

ARTICLE HISTORY
Received 20 April 2021
Accepted 2 October 2021

KEYWORDS
Nitrate-nitrogen; ammonia-nitrogen; ant colony optimization; Elman neural network; new ACO-ENN hybrid model

Nomenclature

Ant Colony Optimization ACO
Artificial Neural Network ANN
Department of Irrigation and Drainage DID
Discharge Q
Elman Neural Network ENN
Low Flow Criteria LFC
Mean Absolute Error MAE
Mean Square Error MSE
Nash–Sutcliffe Efficiency NSE
Peak Flow Criteria PFC
Rainfall RF
Levenberg–Marquardt back-propagation training algorithm trainml
Water Level WL

1. Introduction

Nitrogen compound concentration enhancement in rivers is the result of advanced human activity. Such activities lead to an adverse effect on water quality, nitrogen cycle and ecological functioning of the river (Jacobs et al., 2017; Kilonze et al., 2014; Maloney & Weller, 2011). Agricultural practices of using high nitrogen content fertilizers, as per the advice of specialists (Hes-song, 2019) to increase production, leaves huge amounts of unabsorbed nitrogen compounds flowing in runoff water from agricultural fields (Kumar et al., 2020a), ultimately emptying into river water or percolating downwards to groundwater (Salehi et al., 2000; Sharma et al., 2003). These unabsorbed nitrogen compounds increase their natural concentration in rivers, ultimately affecting...
human health (Fewtrell, 2004; Gallo et al., 2015; Ward et al., 2005). These compounds lead to eutrophication in rivers (Rabalais et al., 2002; Rabalais & Turner, 2006), affecting aquatic life and also leading to cancer, blue-baby syndrome (Hossain et al., 2010) and several birth defects in humans (Chen et al., 2017; Gulis et al., 2002).

Most countries, like Malaysia, supply river water for household purposes after processing it in treatment plants. Such treatment plants in Malaysia are not designed for the removal of ammonia-nitrogen (Indah Water, 2019), leading eventually to the suspension of treatment plants when there is an abrupt rise in nitrogen compounds in river water. Historically, the operation of treatment plants have been suspended several times at various locations in Malaysia for the same reason (New Straits Times, 2017; The Star, 2019). This problem of the suspension of treatment plant operation seems to persist unless there is prior knowledge of spikes in nitrogen compounds in river water, in which circumstance the government will have sufficient time for managing the situation by warning the suspected sources of these compounds. Artificial Neural Networks (ANNs) are used for the prediction of such natural variables to obtain prior knowledge of their patterns.

The ANN is a black-box model (Akrami et al., 2013) mimicking the biological neurons of the human brain (Anctil et al., 2009; Lek et al., 1999; Sharma et al., 2003). Similar to fractional calculus theory (Arqub, 2017, 2019), numerical methods such as ANNs have also gained popularity because of their applicability in a variety of problems in physics and engineering (Arqub & Abo-Hammour, 2014). ANNs have been used continuously, and without limitation to the prediction of various water quality variables. He et al. (2011) stated that ANNs have been implemented in the fields of hydrological processes (Cigizoglu & Alp, 2004; Riad et al., 2004), water resources management (He & Takase, 2006; Mazvimavi et al., 2005) and reservoir operations (Aguilera et al., 2001; Li-Chiu & Fi-John, 2001; Suen & Eheart, 2003; Tayfur et al., 2005; Zaheer & Bai, 2003). In recent years, ANNs have been used along with computational fluid dynamics for various prediction and optimization applications in reactors – as used by Mosavi et al. (2019) – and compressors – as used by Ghalandari et al. (2019). Despite good success rates, in some cases of complex natural variables, ANNs fail to learn the actual physical or chemical processes associated with the corresponding target variable. Also, an ANN model may have been found to be efficient in some cases but the same ANN model may be inefficient in other cases, leading to the need for hybridization (Voyant et al., 2012). A hybrid model is an integration of two or more ANNs and other models, formed to overcome certain limitations of standalone ANN models, and hence to increase prediction accuracy.

Cabaneros et al. (2017) compared multilayer perceptrons with three different hybrid models for the prediction of urban roadside NO2 pollution. Stepwise regression, principle component analysis and regression trees were used along with multilayer perceptrons to develop these hybrid models. Based on their results, they concluded that hybrid models are superior to multilayer perceptron models. Ebrahimpour et al. (2016) developed a hybrid model by combining a multilayer perceptron with Ant Colony Optimization (ACO) for software cost estimation. Their results on hybrid model development stated that hybrid models have better efficiency than algorithmic models. Mulia et al. (2013) developed a hybrid model using a feedforward neural network with a genetic algorithm for the prediction of turbidity and chlorophyll-a. They claimed that, owing to the simplicity and rapid estimation capability of the developed hybrid model, it can be an effective and efficient tool for decision making.

The primary objective of this research is to propose a new hybrid model, developed by integrating ACO with an Elman Neural Network (ENN), that improves the accuracy of standalone ANN models for the prediction of nitrate-nitrogen and ammonia-nitrogen in Langat river, Malaysia. The two compounds of nitrogen (nitrate-nitrogen and ammonia-nitrogen) selected for this study are the major concerns in nitrogen pollution.

2. Effect of nitrogen

The presence of nitrogen compounds in stream water proves to be deleterious for humans as well as aquatic life. Ammonia-nitrogen may get converted to nitrate-nitrogen in river water. However, this conversion depends on the amount of oxygen dissolved in the river water (Nuruzzaman et al., 2017). The major effects of ammonia-nitrogen are eutrophication (Rabalais et al., 2002; Rabalais & Turner, 2006) and the acidification of water bodies (Kumar et al., 2020a). Eutrophication is caused by the growth of algae on the surface of water, which is mainly caused by the presence of nitrogen and phosphorus compounds. Years ago, the problem of eutrophication had been identified as a major problem. Also, nutrient control measures in rivers have commenced in the USA and other countries (Dodds & Smith, 2016; Dodds & Welch, 2000). Excess of ammonia-nitrogen may lead to stream acidification. Acidified stream water cannot be used to fulfill the daily water demand of humans. It is said that, one day, treated stream water supply will be a reality for rural areas,
which will lead to serious human health issues (Gündüz, 2015).

Nitrate-nitrogen, when consumed by humans, may leads to different types of tumors in the human body (Aslan & Turkman, 2003; Hossain et al., 2010). Nitrate-nitrogen compounds may lead to the formation of N-nitroso compounds (Della Rocca et al., 2005; Hossain et al., 2010), which are carcinogenic in nature. Nitrate-nitrogen in the human body is capable of restricting iodine update, leading to thyroid related problems (Hossain et al., 2010).

3. ANNs

An ANN is a complex network of interconnecting nodes, mostly used for the prediction of nonlinear variables (Farzad & El-Shafie, 2016). The basic structure of an ANN consists of an input layer, hidden layers and an output layer, as shown in Figure 1. The input data is transferred from the input layer to the nodes in the hidden layers, which then processes the data and transfer it to the next hidden layer, if any, through a transfer function. Finally, this data propagates to output layer as the output of the ANN model (El-Shafie & Noureldin, 2011). Nodes in the hidden layers process the data by multiplying it with weights and adding the bias values, and then forwarding it to the next hidden layer through a transfer function (El-Shafie et al., 2011), where the same process is followed. During training of the neural network, these weights and biases values are updated to enhance the performance of the model. As shown in Figure 1, the ANN structure receives three inputs ($x_1$, $x_2$ and $x_3$). These inputs are fed to the hidden layer $H_1$, which receives bias value $B_1$. The values from hidden layer $H_1$ are transferred to $H_2$, which receives bias value $B_2$. The values from hidden layer $H_2$ are then forwarded to the output layer, which produces the output $y$.

4. ACO

ACO is one of many nature-inspired algorithms for discrete optimization (Mavrovouniotis & Yang, 2013), and was proposed by Dorigo et al. (1996). The ACO algorithm is based on the foraging behavior of real ants in nature (Junior et al., 2013). Such behavior helps them to find the shortest path to the food source from their initial position (the nest) (Kucukkoc & Zhang, 2013). Ants deposit a chemical called a pheromone on their paths while moving (Shah et al., 2012; Tehrani & Khodayar, 2010), forming a pheromone trail. The following ant smells the deposited pheromone and chooses its paths with a certain probability (Kalinli et al., 2010). A path having a greater intensity of pheromone has a higher probability of being selected by the following ants. Ants choose paths having equal or no pheromone deposited on them randomly. Ants choosing the shortest path will return back sooner, thus increasing the intensity of the pheromone on the shortest path. Hence, the possibility of selecting the shortest path among the set of all paths increases. Figure 2 presents an example of ants selecting the shortest path. At time $T = 0$, all the ants are in their nest. Each ant leaving the nest will have three paths.

![Figure 1. Basic structure of an ANN.](image-url)
At time $T = 0$, there is no pheromone deposited on the paths; hence, the selection probability of each path is the same, and the ants will select the paths randomly. At time $T = 1$, ants have started moving in search of the food source, selecting every possible path randomly. At this point, those ants that have chosen the shortest path (A-B-D-F-G-H) will return faster than those ants following the other two paths. Hence, in time, the intensity of the pheromone will increase on the shortest path leading to more ants choosing the shortest path. At time $T = 2$, most of the ants are seen to be moving on the shortest path. Hence, the shortest path is chosen by the ants.

ACO has been continuously used for many discrete optimizations. The most common use of ACO is in the traveling salesman problem (Hingrajiya et al., 2012), where the minimum cost is to be calculated for the sales person to visit many cities. Many authors have modified ACO to perform certain tasks, as in the case of ElSaid et al. (2020), who developed an ant based neural topology search based on ACO for optimizing recurrent neural network topology. ACO has been used to determine the trainable connections of recurrent neural networks for predicting flight parameters (ElSaid et al., 2019). Juang and Yeh (2017) used advanced continuous ACO to optimize a fully connected recurrent neural network to generate the multi-objective gait of a biped robot. Sheikhan and Mohammadi (2012) used a hybrid model consisting of a genetic algorithm and ant colony optimization for selecting efficient features. Zheng-da et al. (2005) used a hybrid of ACO and a recurrent neural network to overcome the shortcomings of a back-propagation algorithm. In this study, ACO has been used for optimizing the weights of the ENN learning model.

5. ENN

The ENN was developed by Elman in 1990 for resolving voice processing problems (Li et al., 2019). The ENN, a typical dynamic recurrent neural network, is similar to other ANN models except for a context layer introduced into the structure that receives input from the output of the hidden layers. The stored-values of the context layer are fed back to the same hidden layers (Mahdaviani et al., 2008) in the next data point along with the model inputs. The context layer adds short-term memory to the network as it stores activations of hidden layers of the last data point, thus adding one-step delay to hidden layers (Sheela & Deepa, 2013). Context layers intervene between the hidden layer and output layer, unlike the Jordan network, in which context layers receive values from output layers (Song, 2010). The ENN has mainly feedforward connections, except for the context layers, and the model is trained using a back-propagation algorithm (Tampelini et al., 2011).
As shown in Figure 3, the ENN model has three user-defined inputs ($x_1$, $x_2$, $x_3$), one hidden layer with five nodes, one context layer and one output layer with one output ($y$). In this study, the ENN model is used along with ACO to develop a new hybrid model.

The unique feature of ENNs lies in the additional context layer. This layer adds a time-delay operator to the ENN. The advantage of the ENN over other ANN models is that it has the ability to memorize, and hence it has the characteristic of being time-varying and has stronger global stability (Jia et al., 2014). With the help of the context layers, the ENN is capable of learning different complex patterns in the given data, making it more efficient than other ANN models. Owing to its advantage of memorizing characteristics, an ENN has been chosen for this study.

6. The hybrid model

Hybridization is the integration of two or more ANNs with other models. Standalone ANN models have proved to be very efficient in some special cases and inefficient in some other cases (Voyant et al., 2012). For instance, some ANN models that are used for linear time series prediction cannot be used for nonlinear time series prediction. The concept of hybridization is to overcome such limitations of standalone ANN models. Integrating models for linear and nonlinear time series prediction will create a hybrid model that is efficient in both cases.

One of the benefits of hybridization is that it helps improve prediction accuracy. The concept of hybridization of ANNs is to attach an optimization algorithm ahead of an ANN model. In this study, ACO and an ENN have been integrated to develop a new hybrid model. An ACO algorithm has been used as a decision-making model and an ENN is used as a learning model. ACO is assigned to choose a better set of weights and biases, having lower mean square errors (MSE) values, for the ENN to train upon. By default, the ENN was initializing the set of weights and biases randomly, which might have any MSE value and would start the training. But with the ACO algorithm attached ahead of the ENN, ACO will choose a set of weights and biases that has lower MSEs than usual. Then the ENN will be allowed to start training on that set of weights and biases, which will lead the ENN to higher accuracy. Hence, the developed hybrid model significantly improves the accuracy of predicting nitrate-nitrogen and ammonia-nitrogen in Langat River, Malaysia.

7. Study area and data acquisition

For this study, the data have been collected from Kumar et al. (2020b), who ultimately obtained the data from the Department of Irrigation and Drainage (DID), Malaysia. The data consisted of RainFall (RF), Water Level (WL) and discharge ($Q$) as input and nitrate-nitrogen and ammonia-nitrogen as target variables. Monthly average, non-seasonal and non-stationary data were obtained for the period 1981–2017 from two different stations, i.e. Lui Station and Kajang Station in the Langat River basin, Malaysia. Figure 4 presents the study area for the current study, i.e. Langat River basin, Selangor, Malaysia. Lui Station and Kajang Station are indicated on the map in Figure 4.

The original data collected from these two stations had some missing data points, which were completed by interpolation. The missing data introduce some uncertainty (Arqub, 2015) into the prediction process. However, this uncertainty, and other uncertainties, are
reflected in the performance parameters. Figure 5(a) plots the nitrate-nitrogen and ammonia-nitrogen data from both stations. Figure 5(b) plots the rainfall, water level and discharge data for both stations. The y-axis in Figures 5(a) and 5(b) is the concentration of nitrogen compounds in milligrams per liter, and the x-axis of Figures 5(a) and 5(b) is the data point count. Geographically, Lui Station is located upstream of Kajang Station, in a mountainous region close to the origin of the river. Kajang Station is located in an almost flat region, in a populous city. The average nitrate-nitrogen and ammonia-nitrogen at Lui Station is 1.34 and 0.11 mg/l, respectively, which is low in comparison to the Kajang Station (7.32 and 1.96 mg/l, respectively). This is because there is less nitrogen pollution near Lui Station, possibly because there are fewer agricultural and industrial activities owing to its being a mountainous region. Along the river path to Kajang there are many agricultural and industrial areas leading to a higher level of nitrogen pollution at Kajang Station.

According to statistical analysis, the monthly average rainfall received at both stations is almost the same (6.85 mm at Lui Station and 6.89 mm at Kajang Station). However, there is a huge difference in the average water level of the river at these stations (76.17 m at Lui Station and 22.7 m at Kajang Station) with an average discharge of 2.19 m$^3$/s at Lui Station and 12.53 m$^3$/s at Kajang Station, which is possibly due to the difference in the geographical locations of the stations (the mountainous region of Lui Station and the almost flat region of Kajang Station).

8. Performance criteria

The performance of the hybrid model is measured with six different parameters, which are: MSE (Equation 1), mean absolute errors (MAE) (Equation 2), Nash–Sutcliffe efficiencies (NSE) (Equation 3), the maximum relative percentage error (Equation 4), peak flow criteria (PFC) (Equation 5) and low flow criteria (LFC) (Equation 6). MSE and MAE measure the error between the predicted and the target values. MSE and MAE are measured on a scale of zero to infinity (Malik et al., 2021). NSE quantifies the accuracy of the model on a scale of minus infinity to one (Malik et al., 2021), where one represents a perfect model that predicts the exact target values. PFC and LFC represent the ability of the model to predict the peak and low values, respectively. These two criteria measure...
the error of peak values and low flow values. Zero values of PFC and LFC represent a perfect model (Coulibaly, Anctil, et al., 2001), predicting the peak and low values exactly. The equations used for calculation of the different parameters are as follows:

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (x - y)^2 \quad |0 < \text{MSE} < \infty| \quad (1)
\]

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |x - y| \quad |0 < \text{MAE} < \infty| \quad (2)
\]

\[
\text{NSE} = 1 - \frac{\sum (y - x)^2}{\sum (x - \bar{x})^2} \quad |-\infty < \text{NSE} < 1| \quad (3)
\]

\[
\text{Maximum relative percentage error} = \max \left( \left| \frac{x_i - y_i}{x_i} \right| \times 100 \right), \quad i = 1, \ldots, n \quad (4)
\]

\[
\text{PFC} = \frac{\left( \sum_{i=1}^{T_p} (x - y)^2 \times x^2 \right)^{0.25}}{\left( \sum_{i=1}^{T_p} x^2 \right)^{0.5}} \quad |0 < \text{PFC} < \infty| \quad (5)
\]

**Figure 5.** (a) Plot of nitrate-nitrogen and ammonia-nitrogen concentrations for both stations (Kumar et al., 2020b). (b) Plot of rainfall, water level and discharge for both stations (Kumar et al., 2020b).
Figure 5. Continued.

\[
LFC = \left( \frac{\sum_{i=1}^{T_p} (x - y)^2 \times x^2}{\left( \sum_{1}^{T_L} x^2 \right)^{0.5}} \right)^{0.25} \quad |0 < LFC < \infty|
\]

where, for this study, \( n \) = number of data points, \( x \) = target data, \( \bar{x} \) = average of target data, \( y \) = predicted data, \( i \) = iterative variable, \( T_p \) = the number of flow peaks greater than one-third of the observed mean peak flow, and \( T_L \) = the number of low flows lower than one-third of observed mean low flow (Coulibaly, Bobée, et al., 2001).

9. Model structure

9.1. Modification of ACO

The general behavior of ACO is to find the shortest path between the ant nest and the ants’ food source. For this study, ACO is modified to find the optimized hidden layer weights and biases for the ENN on the basis of the MSE of the ENN’s output. As explained earlier, ants are capable of choosing their paths, from among various paths options, at a certain node based on some criteria. The criterion for ants in nature is the shortest path between the nest and the food source. In various problems, artificial ants are given different criteria. For instance, the criterion for the traveling salesman problem is the path with minimum cost. The present authors have used this feature of ACO and have set the artificial ants the criterion of minimizing the MSE.

9.2. Integrating ACO and the ENN

The new ACO-ENN hybrid model is developed on the MATLAB platform. The process of integrating ACO with the ENN involves the different processes explained below.

9.2.1. Creation of an Elman network and weights extraction

The process of the hybridization starts with the creation of an Elman network with hidden layers and their corresponding nodes defined by the user. The created Elman network is then configured with the inputs and target variables provided by the user. On configuring, the network assigns the weights and biases to the nodes of hidden layers. The hybrid model then extracts the weights and biases of the network and stores them in a single 1-D array by appending them one after the other.
9.2.2. Splitting the weights

The extracted weights and biases, stored in a single 1-D array, are then split into a user-defined number of layers to form a 2-D array with rows as the number of weight-split layers and columns as the number of weights and biases. The hybrid model then splits the weights and biases by randomly selecting values between the maximum and minimum weights and biases stored in the 1-D array. The first row of the 2-D array is the extracted weights and biases itself and the rest of the rows are the randomly selected values based on the maximum and minimum weights and biases extracted from the ENN. Splitting of the weights and biases forms a 2-D array that can be converted into the network of paths by making the split weights nodes and joining them with all the nodes of the layers ahead, as shown in Figure 6.

9.2.3. Ant movement

The hybrid model then attaches a nest to the left and a food source to the right of the network of the paths obtained by splitting the weights. Attaching the nest and the food source completes the plot for the ants to start searching for the food source. A user-defined number of ants then start the motion from the nest through each node to the food source, depositing pheromone along the way. Pheromone for each path is stored in a 3-D array with rows as the number of weight-split layers, columns as the number of weights and biases, and the third dimension as the number of the weight-split layers (representing the connection of a node to each node in the adjacent layers). The initial values of pheromone are assigned as one. The path of each ant is stored. Every ant reaching the food source has a definite path, i.e. a definite set of weights and biases. These sets of weights and

Figure 6. Plot of ant movement.

Figure 7. Movement of the queen ant (blue path).
biases for every ant are then sent to the ENN, where these weights and biases are restructured, in the same manner as they were extracted, and assigned to the ENN. This network uses the input data to predict the output and calculates the MSE using the target values. Hence, an MSE for each ant is obtained. Once all the ants reach the food source through a definite path, each ant gets their MSE. The hybrid model then selects the ant with the minimum MSE and assigns it as the queen ant (i.e. the best ant). The MSE value of the queen ant is stored for comparison in the upcoming iterations. The path of the queen ant is also stored for display on the graph, as shown in Figure 7, and for reference after the iterations of ACO end.

9.2.4. Updating and limiting pheromone

The hybrid model then allows only the queen ant to deposit pheromone while returning. The rest of the ants are forced to return to the nest without depositing pheromone. The 3-D pheromone array is then updated according to the MSE value of the queen ant – Equation (7) as used by Mavrovouniotis and Yang (2013).

\[
\text{Updated Pheromone Value} = \text{Previous Pheromone Value} + \frac{1}{\text{Mean Square Error of Best Ant}} \quad (7)
\]

![Flowchart of the hybrid model structure.](image-url)
Figure 9. Convergence of the Mean Square Error (MSE) in ACO.

The smaller the value of the MSE, the greater is the value of pheromone addition. In some cases where the MSE is less than one, the inverse of such a value will result in a higher value. Hence, a higher pheromone value will be updated, leading to one particular path having a higher intensity of pheromone in comparison with other paths. As a result, almost all ants will choose that path because the probability of choosing that particular path will be very high. As a result, all other paths that may result in an even smaller value of the MSE will be left unexplored. Hence, to avoid this limitation of the hybrid model, the pheromone values are limited to a certain value according to Equation (8) as used by Mavrovouniotis and Yang (2013):

\[
\text{Max Pheromone} = \frac{1}{\text{Evaporation rate} \times \text{Mean Square Error of Queen Ant}}
\]

The deposited pheromone keeps on evaporating after every iteration, with a user-defined evaporation rate. The hybrid model allows only the queen ant to deposit pheromone, leaving other paths (followed by ants or not) undeposited with pheromone. But the pheromone on all the paths is subject to evaporation. After a few iterations, the pheromone on a path that received no pheromone deposition but only evaporation will observe a major reduction in the intensity of the pheromone. As a result,

| Nitrogen compound | Lui Station | Kajang Station |
|-------------------|-------------|----------------|
|                   | Nitrate     | Ammonia        | Nitrate     | Ammonia        |
|                   | ANN | Hybrid | ANN | Hybrid | ANN | Hybrid | ANN | Hybrid |
| Number of hidden layers | 2 | 2 | 3 | 2 | 3 | 2 | 2 | 2 |
| Nodes in each hidden layer | 10 | 10 | 7 | 7 | 10 | 10 | 10 | 9 |
| ANN/ENN epochs | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 |
| ACO iterations | – | 100 | – | 100 | – | 100 | – | 100 |
| Number of ants | – | 50 | – | 50 | – | 50 | – | 50 |
| Number of weight-split layers | – | 50 | – | 50 | – | 50 | – | 50 |
| MSE | 0.196 | 0.012 | 0.0085 | 0.0008 | 14.09 | 1.21 | 0.505 | 0.112 |
| MAE | 0.271 | 0.069 | 0.046 | 0.017 | 2.47 | 0.40 | 0.38 | 0.16 |
| NSE | 0.726 | 0.984 | 0.603 | 0.957 | 0.586 | 0.965 | 0.87 | 0.97 |
| Maximum error | 257.5% | 52.2% | 334.3% | 91.8% | 386.3% | 89.6% | 305.7% | 75.4% |
| PFC | 0.197 | 0.070 | 0.345 | 0.118 | 0.234 | 0.103 | 0.172 | 0.117 |
| LFC | 0.764 | 0.330 | 0.727 | 0.293 | 1.212 | 0.335 | 0.888 | 0.509 |
the intensity of the pheromone on the queen ant’s path will be higher in comparison with the other paths, leading to a higher probability of the selection of that path by the ants. As a result, all other paths that may lead to an even smaller value of the MSE will be left unexplored. To overcome this situation, the lower limit of the pheromone is applied according to Equation (9), as used by Mavrovouniotis and Yang (2013):

\[
\text{Min Pheromone} = \frac{\text{Max Pheromone}}{2 \times \text{No. of Weights and Biases}} \tag{9}
\]

The minimum amount of pheromone is less than half the maximum amount of pheromone. This depends on the number of weights and biases, which in turn depends on the size of the Elman network.

9.2.5. ENN training

At the end of the ACO iterations, the hybrid model collects the final queen ant path from the ACO part of the model. Based on the queen ant path, the corresponding weights and biases are obtained from the 2-D weight split array. The hybrid model then re-structures these weights and biases and assigns them in the Elman network. The hybrid model then allows the Elman network to start training using the weights and biases received from ACO and the inputs and target values received from the user. Beginning the training of the Elman network from the lower MSE, obtained from ACO, will provide the Elman network with a better start, which will lead the Elman network to enhanced results. The trained network is then provided as the output of the hybrid model.

Figure 8 presents a flow chart of the overall process of hybrid structure development. The flow chart briefly describes each step, starting from the initialization of the Elman neural network and all the way through ACO, to finally training the ENN with the optimized weights and biases. After training the ENN, a hybrid model is developed and can be used for further analysis. The algorithm used to develop the hybrid structure is presented below. The steps presented in the algorithm are similar to the steps in the flow chart in Figure 8.

- Initialise the Elman neural network
- Configure the Elman neural network

Figure 10. (a) Plot of target and output for each hybrid model. (b) Plot of target and output for each standalone multilayer model.
• Extract the weights and biases of the configured ENN
• Reshape the weights and biases into a 1-D array
• Split the weights and biases into a 2-D array
• Attach the nest at the beginning and a food source at the end of the 2-D array
• Start the ACO iterations
  o Start the movement of ants from the nest
  o Ants reaching the food source via any path have a set of weights and biases
  o Send this set to ENN
  o Reshape and assign the weights and biases to ENN
  o Obtain the MSE value for each ant
  o Every ant return to the nest with their MSE
    Update pheromone values (only the queen ant is allowed to deposit the pheromone)
  o Limit the pheromone within maximum and minimum values
• Weights and biases of the queen ant are sent to ENN
• Weights and biases are reshaped and assigned to ENN
• Train the ENN with the new weights and biases

9.3. Application of the ACO-ENN hybrid model on nitrogen data

The developed ACO-ENN hybrid model was tested on the nitrate-nitrogen and ammonia-nitrogen prediction data obtained for this study. Four separate models were trained for nitrate-nitrogen and ammonia-nitrogen for the corresponding stations, i.e. the Lui and Kajang Stations. Models were trained on a MATLAB platform. Input and target data were provided to the hybrid model with 90% of the data assigned for training the model and the remaining 10% of the data for the testing the trained model. The data were divided for training and testing in such a way that training and testing data were statistically same (i.e. their average values were approximately same) with the maximum and the minimum data points present in the training portion. The following configurations of the model was used for training the hybrid model:

• ACO iterations: 50–200
• Number of ants: 20–50
• Evaporation rate: 0.5
• Number of weight-split layers: 20–50
• Elman hidden layers: varied from 2 to 3
• Elman nodes in each hidden layer: varied from 7 to 10
• Elman training function: ‘trainlm’
• Elman performance function: ‘MSE’
• Elman epochs: 1000–2000

The range and values of the above-mentioned configurations provide the optimum results for the hybrid modeling. The value of the evaporation rate (0.5) used in ACO algorithm is most commonly used in optimization, providing optimum results. The range of ACO iteration selected for this hybrid training is 50–200, which is the optimal range. Iteration below 50 would be too few to obtain the most optimized result. Iterations above 200 rarely leads to the optimized results, but have the cost of extra time incurred. Hence, the range 50–200 often lead to the optimized results. The range of the number of ants used in this study is 20–50. In general, the more ants, the higher is the possibility of getting the optimized result. But a greater number of ants requires more computational resources and more time. Hence the optimal range of the number of ants selected is 20–50, which is sufficient to explore all the paths of the branched structure made by the weight-split layers of range 20–50. The number of weight-split layers controls the complexity of the optimization process. More weight-split layers lead to a more branched structure, which needs greater computational resources to move all the ants on this massive branched structure. However, less weight-split layer may lead to the inferior structure which may not yield the optimum result. Hence, the range of weight-split layers selected is 20–50, which works fine in collaboration with 20–50 ants. More weight-split layers require more ants to explore all the possible paths of the branched structure. The range of Elman network configurations used in this study such as hidden layers, nodes in hidden layers, train functions, performance functions and epochs were selected based on the study conducted by Kumar et al. (2020b).

Figure 11. (a) Plots of relative percentage error obtained from the hybrid models. (b) Plots of relative percentage error obtained from the standalone multilayer models.
The ACO-ENN hybrid model begins training with the deployment of ACO for the optimization of weights and biases extracted from the ENN. ACO performs the optimization process for the defined number of iterations. Post completion of the optimization process, ACO provides the ENN with the best set of weights and biases to train upon. The ACO-ENN hybrid model then trains the ENN model, for the defined number of epochs, to deliver the model predicted results. The default training function ‘trainlm’ is used for training the hybrid model: ‘trainlm’ is the Levenberg–Marquardt back-propagation training algorithm, which updates the weight and bias values according to the Levenberg–Marquardt procedure. The results obtained from the hybrid model for the nitrate-nitrogen and ammonia-nitrogen predictions at the corresponding stations (i.e., Lui Station and Kajang Station) are presented in the following section.

10. Results

Training the hybrid model at the above-mentioned configuration and trying all the possible configurations in the given range of hidden layers and nodes in the hidden layers, the hybrid model improved the accuracy of the results significantly. The input dataset used for this training consisted of rainfall, water level and discharge. The target dataset consisted of the nitrate-nitrogen and ammonia-nitrogen concentrations at stations Lui and Kajang. The dataset used for hybrid model training was manually divided into two sets: a training set and a testing set, 90% of the data was used for training and the remaining 10% was separated for testing. The division of the data was such that the training and testing datasets were statistically similar, i.e. both datasets had similar mean values. Also, the maximum and minimum values of the target variables were placed in the training dataset for the model to learn all the maxima and minima of the data.

One of the results of the hybrid model is the plot of MSE convergence. Figure 9 shows a plot of MSE convergence in the ACO iterations. The plot shows how the ants, by their pheromone communication, have managed to find the best path. Figure 9 is a plot of MSE convergence obtained while training the model for predicting the ammonia-nitrogen concentration at Lui Station.
figure demonstrates how the ants in the ACO had found the path with the lowest MSE from among the network of paths. Towards the end of the iteration, ants have converged to the path having the lowest MSE. As can be seen in Figure 9, ants have tried to converge several times on different paths (the portion of the plot having constant MSE) but, after a few iterations, the queen ant found a path having an even lower MSE. In similar steps, the queen ant managed to find the ultimate lowest MSE path, leading other ants to converge on that path. This MSE and its set of weights and biases was then assigned to the weights and biases of the ENN learning model. The ENN that was trained on this MSE provided enhanced results when compared with the results of the traditional standalone multilayer ANN models.

Table 1 shows a comparison of the results obtained from the hybrid models with the results obtained from the traditional standalone multilayer ANN models. The configurations of the multilayer models are presented in Table 1. The hybrid models are fed the same input and target data as that used in the standalone multilayer ANN models. Both models are compared to observe the capability of the models when both are fed the same input and target data. Both model results are compared on the basis of all six performance parameters mentioned earlier, i.e. MSE, MAE, NSE, maximum error, PFC and LFC. Results from the new hybrid model showed a significant improvement in almost all performance criteria computed. MSE and MAE represent the overall error (difference between the target and predicted values) produced by a model. With the help of the new hybrid model, MSE and MAE values have been reduced to a certain extent. The NSE criterion measures a model’s capability of predicting the values close to the target values. As stated earlier for a model’s capability measurement, NSE uses a scale of minus infinity to one, where one represents a perfect generalized model. With the help of the new hybrid model, the NSE values have been brought closer

Table 1. The hybrid models are fed the same input and target data as that used in the standalone multilayer ANN models. Both models are compared to observe the capability of the models when both are fed the same input and target data. Both model results are compared on the basis of all six performance parameters mentioned earlier, i.e. MSE, MAE, NSE, maximum error, PFC and LFC. Results from the new hybrid model showed a significant improvement in almost all performance criteria computed. MSE and MAE represent the overall error (difference between the target and predicted values) produced by a model. With the help of the new hybrid model, MSE and MAE values have been reduced to a certain extent. The NSE criterion measures a model’s capability of predicting the values close to the target values. As stated earlier for a model’s capability measurement, NSE uses a scale of minus infinity to one, where one represents a perfect generalized model. With the help of the new hybrid model, the NSE values have been brought closer

**Figure 12.** (a) Regression plot of the hybrid models. (b) Regression plot of the standalone multilayer models.
to one, making the hybrid model more nearly perfect in comparison with the multilayer model. The maximum error criteria represent the lack of proper generalization of the model. In some cases in the current study, the maximum error for the multilayer model was more than 100%, which has been reduced below 100% with the help of the hybrid model. PFC and LFC have been reduced significantly. PFC represents the ability of the model to learn the pattern in the target data provided to the model when there is peak flow. In other words, it represents the ability of the model to predict accurately the peak flow values. Similarly, LFC represents the ability of the model to predict accurately the low flow values. The hybrid model helped in reducing PFC and LFC closer to zero; which means that, after training the models with the new hybrid algorithm, the models are now capable of predicting more accurately the values of nitrate-nitrogen and ammonia-nitrogen when they are at their peak flow and low flow conditions, hence increasing the prediction capability of the models at both extreme values of the pattern of the target data.

Figures 10(a) and 10(b) present plots of the target and output of all the datasets (training and testing datasets) for each hybrid and each standalone multilayer model, respectively. For the hybrid models, the output plot line appears to cover the target plot line entirely. This demonstrates that the model had generalized the relation between the input and target vectors, hence predicting almost the same values. For the multilayer models, the output line seems to go off the track of the target line at various points. This demonstrates that the model was not fully capable of generalizing the relation between the input and target vectors.

Figure 11(a) presents a plot of the relative percentage error for all the four models (nitrate-nitrogen and ammonia-nitrogen concentrations for the corresponding stations, i.e. Lui Station and Kajang Station) obtained from the hybrid models. In comparison with these plots,
Figure 11(b) presents a plot of the improved relative percentage error of all the four standalone multilayer models. As can be seen in Figure 11(b), the relative percentage errors surpass 100% in the case of multilayer models, which are limited to within 100% by the hybrid models, as shown in Figure 11(a).

Figures 12(a) and 12(b) present regression plots, along with their $R^2$ values and regression equations, of the hybrid and multilayer models for each station (i.e. Lui Station and Kajang Station) corresponding to nitrate-nitrogen and ammonia-nitrogen. For the hybrid models, the $R^2$ square values of all the four models are more than 0.95, which is acceptable and fit for future prediction of nitrate and ammonia compounds at Lui and Kajang Stations. For the multilayer models, the $R^2$ values are between 0.61 and 0.87, which is considered to be low. As stated above, this is because the multilayer models have failed properly to generalize the relation between the input and target vectors.

Observing the plots of the relative percentage error and the regression plots for both the ANN and hybrid models, it can be stated that the new hybrid training algorithm enhances the generalization capability of the model, thus improving the accuracy of the prediction results as the errors of the predicted values have been reduced to a certain extent.

### 11. Discussion

The hybrid model developed by integrating ACO with an Elman network shows improvement on the ANN results when trained on the same data. Table 2 presents the percentage improvement of different performance criteria when compared with the ANN model results. As shown in Table 2, the new hybrid training algorithm has improved the model capability in almost all respects. The error parameters, such as MSE, MAE and maximum relative percentage error (maximum error), have been improved significantly, leading to a reduction in the prediction error to a certain extent. PFC and LFC have been enhanced closer to zero, leading to increased accuracy of the model for predicting the peak flow values and low flow values. Hence, the new ACO-ENN hybrid model finds itself to be more suitable in comparison with other ANN models for developing a model with enhanced capabilities.

The results of the ACO-ENN hybrid model were also compared with the results of a standalone Elman neural network model and a multilayer model at the same time. This comparison proved the effectiveness of the ACO-ENN hybrid model over the ANN models. The comparison is presented in Table 3. The Elman network was trained on the same dataset with the same Elman network configurations as used by the hybrid model and multilayer ANN model. The standalone ANN model provided inferior results compared to the ACO-ENN hybrid model, yet better results than ANN models. As presented in Table 3, the error parameters (i.e. MSE, MAE) of the ENN were higher than the corresponding hybrid models and lower than the corresponding ANN model. The regression values and the NSE values were found to be closer to those of the hybrid models yet inferior to the hybrid models. The uncertainty values, calculated as the average of the absolute relative percentage error, of the ENN is observed to be superior to that of the ANN model but inferior to that of the ACO-ENN hybrid model. The uncertainty value represents the possible percentage error in the predicted results. The ACO-ENN hybrid model has the least uncertainty percentage among the ENN and ANN models, thus confirming the ability of the ACO-ENN hybrid model to produce enhanced prediction results.

Figures 13–15 present the target and output plots, relative percentage error plot and regression plot for all four ENN models for both the stations (Lui and Kajang) corresponding to both the nitrogen compounds. In Figure 13, the output plot seems to cover almost all the target line.

### Table 2. Percentage improvement of different performance criteria of hybrid model.

| Nitrogen compound | Lui Station | Kajang Station |
|-------------------|-------------|----------------|
| Nitrate | Ammonia | Nitrate | Ammonia |
| ANN MSE | 0.196 | 0.0085 | 14.09 | 0.505 |
| Hybrid MSE | 0.012 | 0.0008 | 1.21 | 0.112 |
| MSE improvement (%) | 93.88 | 90.59 | 91.41 | 77.82 |
| ANN MAE | 0.271 | 0.046 | 2.47 | 0.38 |
| Hybrid MAE | 0.069 | 0.017 | 0.40 | 0.16 |
| MAE improvement (%) | 74.54 | 63.26 | 83.97 | 57.78 |
| ANN NSE | 0.726 | 0.603 | 0.586 | 0.957 |
| Hybrid NSE | 0.984 | 0.957 | 0.965 | 0.97 |
| NSE improvement (%) | 35.63 | 58.71 | 64.68 | 12.01 |
| ANN max error (%) | 257.50 | 334.3 | 396.3 | 305.7 |
| Hybrid max error (%) | 52.21 | 91.8 | 89.6 | 75.4 |
| Max error improvement (%) | 79.72 | 72.54 | 76.80 | 75.34 |
| ANN PFC | 0.197 | 0.345 | 0.234 | 0.172 |
| Hybrid PFC | 0.070 | 0.118 | 0.103 | 0.117 |
| PFC improvement (%) | 64.60 | 65.84 | 55.98 | 32.39 |
| ANN LFC | 0.574 | 0.727 | 1.212 | 0.888 |
| Hybrid LFC | 0.330 | 0.293 | 0.335 | 0.509 |
| LFC improvement (%) | 56.75 | 59.75 | 72.33 | 42.65 |

### Table 3. Comparison of results of ANN, ENN and hybrid models at Lui Station.

| Model | Nitrate | Ammonia |
|-------|---------|---------|
| Regression | 0.859 | 0.968 | 0.992 | 0.783 | 0.907 | 0.979 |
| MSE | 0.196 | 0.049 | 0.012 | 0.0085 | 0.0038 | 0.0008 |
| MAE | 0.271 | 0.094 | 0.069 | 0.046 | 0.024 | 0.017 |
| NSE | 0.726 | 0.932 | 0.984 | 0.603 | 0.819 | 0.957 |
| Uncertainty (%) | 26.938 | 12.648 | 7.255 | 64.245 | 37.355 | 23.867 |
However, at some point the output line went off the track of target line. Figure 14 presents a plot of the relative percentage error within the plot limits of −400% to 400%. The errors at most of the points cross the 100% mark and at very few points cross the 200% mark. In Figure 15, the $R^2$ values are in the range 0.82 to 0.93, which is slightly better than the multilayer models yet inferior to the hybrid models.

Further comparison of the hybrid models with those of the standalone ANN and ENN models is made on Taylor diagrams. Figure 16 presents plots of the Taylor diagrams for both the stations corresponding to both compounds, i.e. nitrate-nitrogen and ammonia-nitrogen. The Taylor diagrams compare the models based on their standard deviations between observed and predicted values and correlation values. The standard deviation of observed values is marked as an actual point on the plots. The model closest to the actual point is considered to be the model predicting most accurately the values closest to the observed values. In all the four plots, the hybrid model is closest to the actual point. Hence, the hybrid model is considered to be the most accurate model and can be used for further analysis. Furthermore, based on the comparison between ANN, ENN and hybrid in Tables 3 and 4 and a Taylor diagram, it is concluded that the ENN provides better results than the multilayer models. The ENN has better learning capability than ANN and has the ability to provide superior results to multilayer models (Abulalqader & Ali, 2018; Chiu et al., 2019; Oyewola et al., 2021).

The accuracy of the ACO-ENN hybrid model depends on its parameters, which are the number of ACO iterations, the number of ants, the number of weight-split layers, the number of Elman network hidden layers, the number of nodes in hidden layers, and the number of Elman network epochs. Increasing the number of ACO iterations will increase the probability of exploring all the paths. However, after sufficient iterations, all the ants start converging to a particular path. Increasing the number of ants will increase the probability of each path being explored at least once. Fewer ants may lead to poor accuracy. Increasing the number of weight-split layers will increase the combination of weights to be explored, which may lead to higher accuracy. The number of

**Figure 13.** Plots of the target and output for every ENN model.
hidden layers in the Elman network, the number of nodes in the hidden layers and the number of epochs depend on the data and the pattern in it to be learnt. Networks with fewer hidden layers, nodes and epochs may not be capable of learning all the patterns of the target data. However, higher numbers of hidden layers, nodes and epochs may lead the model to overfit, hence reducing the model accuracy. To reduce the overfitting of the neural network used in this hybrid model (i.e. the Elman neural network), the user has to track the testing regression. If the testing regression is decreasing with increasing complexity of the network, then the network is overfitting. Hence, the complexity of the network needs to be reduced by reducing the number of hidden layers, nodes in the hidden layers or epochs. The optimum configuration of the network leads to a higher accuracy model.

The rules and properties of the new ACO-ENN hybrid model are listed as follows.

1. The ACO-ENN hybrid model searches for the optimal values of the internal parameters of the Elman neural network

2. It accelerates the convergence process during the training stage to achieve the performance goal – the faster the convergence rate, the better for attaining a better model output and hence proving suitable for real-time application of the proposed model as well.

The developed hybrid model has the advantage of increasing the learning capability of the ANN model and hence increasing the prediction accuracy of the model. However, the disadvantage of the developed hybrid model is that it needs higher computation time. Hybrid model has to run two process one after other, first is the optimization process and later the network learning process. Whereas, the standalone ANN model has to go for the network leaning process only. Hence, the total computation time is more than the standalone ANN learning time. Albeit, the computation time depends on the complexity of the network and the number of data samples. However, the computation time can be reduced with modern computation resources.

All the four models trained by the new ACO-ENN hybrid model for both the stations (i.e. Lui Station and Kajang Station) corresponding to both the compounds

---

**Figure 14.** Plots of the relative percentage error obtained from the ENN models.
Table 4. Comparison of results of ANN, ENN and hybrid models at Kajang Station.

| Nitrogen compound | Nitrate | Ammonia |
|-------------------|---------|---------|
| Model             | ANN     | ENN     | Hybrid  |
| Regression        | 0.78    | 0.92    | 0.98    |
| MSE               | 14.09   | 5.31    | 1.21    |
| MAE               | 2.47    | 1.26    | 0.40    |
| NSE               | 0.59    | 0.84    | 0.97    |
| Uncertainty (%)   | 50.22   | 21.20   | 8.06    |

(i.e. Nitrate-nitrogen and ammonia-nitrogen) are proved to be of higher accuracy than other model developed on the same dataset. These models are ready for their application of prediction of nitrate-nitrogen and ammonia-nitrogen at both the stations. The information of surge in these two nitrogen compounds, if any, will be provided to the concerned authority, in advance, by these models. The concerned authority will have ample time to take proper action to prevent it.

12. Conclusion

In this study, a new ACO-ENN hybrid model is successfully developed by integrating ACO with an ENN. Here, ACO is used as a decision-making model that decides the initial weights and biases of the ENN by optimizing the weights and biases of the ENN model. In this hybrid model, the ENN is used as a learning model that learns various patterns of the target variable and provides a network for data prediction. Selecting the initial weights and biases for the ENN provides it with a better start, which leads the ENN to achieve high accuracy results. The new ACO-ENN hybrid model is tested by developing four different models for the nitrate-nitrogen and ammonia-nitrogen concentration prediction data for the two different stations, i.e. Lui Station and Kajang Station in Langat River basin, Malaysia. A multilayer ANN model and an ENN model were also trained on the same prediction data for the same two stations, i.e. Lui Station and Kajang Station. The results of the new

![Figure 15. Regression plot of the ENN models.](image-url)
Figure 16. Taylor diagrams for the applied models.

ACO-ENN hybrid model were compared with that of the multilayer ANN models and the ENN models. The hybrid model yielded improved results in comparison with the results from the multilayer ANN models and the ENN models. There was a significant improvement in the MSE, MAE, the maximum relative percentage error, NSE, PFC and LFC in the results of the ACO-ENN hybrid model. Training the model on the new ACO-ENN hybrid model algorithm resulted in enhanced capabilities of the prediction model, especially in predicting the conditions when the nitrogen compounds were in low flow and peak flow. The results of the ENN models were superior to those of the multilayer ANN model, yet inferior to those of the hybrid model. The new ACO-ENN hybrid model has the advantage of enhancing the model’s prediction capabilities. However, the new hybrid model has some disadvantage such as: it requires more training time and it requires modern computational resources. To enhance the results further, this study proposes a future recommendation of integrating other optimization algorithms with the ENN model, and to explore other possibilities of combinations that may lead to even lower errors and more accurate results.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work was supported by the Institut Pengurusan dan Pemantauan Penyelidikan, Universiti Malaya [grant numbers GPF082A-2018, GPF031A-2019].

ORCID
Ali Najah Ahmed http://orcid.org/0000-0002-5618-6663
Mohsen Sherif http://orcid.org/0000-0002-6368-8143
Ahmed Sefelnasr http://orcid.org/0000-0002-0281-6455
Ahmed El-shafie http://orcid.org/0000-0001-5018-8505
Ghalandari, M., Ziamolki, A., Mosavi, A., Shamshirband, S., Chau, K.-W., & Bornassi, S. (2019). Aeromechanical optimization of first row compressor test stand blades using a hybrid machine learning model of genetic algorithm, artificial neural networks and design of experiments. *Engineering Applications of Computational Fluid Mechanics, 13*(1), 892–904. https://doi.org/10.1080/19942060.2019.1649196

Gulis, G., Czompolyova, M., & Cerhan, J. R. (2002). An ecologic study of nitrate in municipal drinking water and cancer incidence in Trnava District, Slovakia. *Environmental Research, 88*(3), 182–187. https://doi.org/10.1006/ersn.2002.4331

Gündüz, O. (2015). Water quality perspectives in a changing world. *Water Quality, Exposure and Health, 7*(1), 1–3. https://doi.org/10.1007/s12403-015-0161-y

He, B., Oki, T., Sun, F., Komori, D., Kanae, S., Wang, Y., Kim, H., & Yamazaki, D. (2011). Estimating monthly total nitrogen concentration in streams by using artificial neural network. *Journal of Environmental Management, 92*(1), 172–177. https://doi.org/10.1016/j.jenvman.2010.09.014

He, B., & Takase, K. (2006). Application of the artificial neural network method to estimate the missing hydrologic data. *Journal of Japan Society of Hydrology & Water Resources, 19*(4), 249–257. https://doi.org/10.3178/jjswhr.19.249

Hessong, A. (2019). *The composition of fertilizers*. Retrieved June 26, 2019, from http://homeguides.sfgate.com/composition-fertilizers-48898.html

Hingrajiya, K. H., Gupta, R. K., & Chandel, G. S. (2012). An ant colony optimization algorithm for solving travelling salesman problem. *International Journal of Scientific and Research Publications, 2*(8).

Hossain, F., Wanielista, M., Xuan, Z., & Daran-Jacobs, S. R., Breuer, L., Butterbach-Bahl, K., Pelster, D. E., Hingrajiya, K. H., Gupta, R. K., & Chandel, G. S. (2012). *Anant Hessong, A. (2019). (2012). Application of the artificial neural network method to estimate the missing hydrologic data. Journal of Japan Society of Hydrology & Water Resources, 19(4), 249–257. https://doi.org/10.3178/jjswhr.19.249.*

Indah Water, M. (2019). *Ammonia*. Indah Water Konsortium Sdn Bhd. Retrieved July 6, 2019, from https://www.iwk.com.my/do-you-know/ammonia

Jacobs, S. R., Breuer, L., Butterbach-Bahl, K., Pelster, D. E., & Rufino, M. C. (2017). Land use affects total dissolved nitrogen and nitrate concentrations in tropical Montane streams in Kenya. *Science of the Total Environment, 603-604*, 519–532. https://doi.org/10.1016/j.scitotenv.2017.06.100

Jia, W., Zhao, D., Shen, T., Tang, Y., & Zhao, Y. (2014). Study on optimized Elman neural network classification algorithm based on PLS and CA. *Computational Intelligence and Neuroscience, 2014*, 1–13. https://doi.org/10.1155/2014/724317

Juang, C.-F., & Yeh, Y.-T. (2017). Multiobjective evolution of biped robot gaits using advanced continuous ant-colony optimized recurrent neural networks. *IEEE Transactions on Cybernetics, 48*, 1910–1922. https://doi.org/10.1109/TCYB.2017.2718037

Juniora, L. S., Nedjah, N., & Mourelle, L. D. M. (2013). Routing for applications in NoC using ACO-based algorithms. *Applied Soft Computing, 13*(5), 2224–2231. https://doi.org/10.1016/j.asoc.2013.01.009

Kalini, A., Acar, M. C., & Gündüz, Z. (2010). New approaches to determine the ultimate bearing capacity of shallow foundations based on artificial neural networks and ant colony optimization. *Engineering Geology, 117*(1-2), 29–38. https://doi.org/10.1016/j.enggeo.2010.0.002

Kilonzo, F., Masese, F. O., Van Griensven, A., Bauwens, W., Obando, J., & Lens, P. N. L. (2014). Spatial–temporal variability in water quality and macro-invertebrate assemblages in the Upper Mara River basin, Kenya. *Physics and Chemistry of the Earth, Parts A/B/C, 67-69*, 93–104. https://doi.org/10.1016/j.pce.2013.10.006

Kucukkoc, I., & Zhang, D. Z. (2013, September). *On applications of ant colony optimisation techniques in solving assembly line balancing problems*. OR55 annual conference, University of Exeter.

Kumar, P., Lai, S. H., Mohd, N. S., Kamal, M. R., Afan, H. A., Ahmed, A. N., Sherif, M., Sefelnsar, A., & El-Shafie, A. (2020b) Optimised neural network model for river-nitrogen prediction utilizing a new training approach. *PLoS One, 15*(9), e0239509. https://doi.org/10.1371/journal.pone.0239509

Kumar, P., Lai, S. H., Wong, J. K., Mohd, N. S., Kamal, M. R., Afan, H. A., Ahmed, A. N., Sherif, M., Sefelnsar, A., & El-Shafie, A. (2020a) Review of nitrogen compounds prediction in water bodies using artificial neural networks and other models. *Sustainability, 12*(11), 4359. https://doi.org/10.3390/su12114359

Lek, S., Guissete, M., & Giraudel, J.-L. (1999). Predicting stream nitrogen concentration from watershed features using neural networks. *Water Research, 33*(16), 3469–3478. https://doi.org/10.1016/S0043-1354(99)00061-5

Li, C., Zhu, L., He, Z., Gao, H., Yang, Y., Yao, D., & Qu, X. (2019). Runoff prediction method based on adaptive Elman neural network. *Water, 11*(6), 1113. https://doi.org/10.3390/w11061113

Li-Chiu, C., & Fi-John, C. (2001). Intelligent control for modelling of real-time reservoir operation. *Hydrological Processes, 15*(9), 1621–1634. https://doi.org/10.1002/hyp.226

Mahdaviani, K., Mazyar, H., Majidi, S., & Saraee, M. H. (2008, June 1–8). *A method to resolve the overfitting problem in recurrent neural networks for prediction of complex systems' behavior*. International joint conference on neural networks (IJCNN 2008).

Malik, A., Tikhamarine, Y., Souag-Gamane, D., Rai, P., Sammen, S. S., & Kisi, O. (2021). Support vector regression integrated with novel meta-heuristic algorithms for meteorological drought prediction. *Meteorology and Atmospheric Physics, 133*, 891–909. https://doi.org/10.1007/s00703-021-00787-0

Maloney, K. O., & Weller, D. E. (2011). Anthropogenic disturbance and streams: Land use and land-use change affect stream ecosystems via multiple pathways. *Freshwater Biology, 56*(3), 611–626. https://doi.org/10.1111/j.1365-2427.2010.02522.x

Mavrovouniotis, M., & Yang, S. (2013). Evolving neural networks using ant colony optimization with pheromone trail limits. *IEEE, 13*, 16–23. https://doi.org/10.1109/UKCI.2013.6651282

Mazvimavi, D., Meijerink, A. M. J., Savenije, H. H. G., & Stein, A. (2005). Prediction of flow characteristics using multiple regression and neural networks: A case study in Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C, 30*(11-16), 639–647. https://doi.org/10.1016/j.pce.2005.08.003

Mosavi, A., Shamshirband, S., Salwana, E., Chau, K.-W., & Tah, J. H. M. (2019). Prediction of multi-inputs bubble column reactor using a novel hybrid model of computational
