© Universiti Tun Hussein Onn Malaysia Publisher's Office

IJIE

http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Metaheuristic Algorithms to Enhance the Performance of a Feedforward Neural Network in Addressing Missing Hourly Precipitation

Wai Yan Lai[1], King Kuok Kuok[1*], Shirley Gato-Trinidad[2], Md. Rezaur Rahman[3], Muhammad Khusairy Bakri[3]

[1]Faculty of Engineering, Computing and Science,
Swinburne University of Technology, 93350 Kuching, Sarawak, MALAYSIA

[2]Faculty of Engineering, Computing and Science,
Swinburne University of Technology, Hawthorn Victoria 3122, AUSTRALIA

[3]Faculty of Engineering,
Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, MALAYSIA

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2023.15.01.025
Received 23 April 2022; Accepted 20 December 2022; Available online 31 March 2023

Abstract: This research study investigates the implementation of three metaheuristic algorithms, namely, Grey Wolf Optimizer (GWO), Multi-Verse Optimizer (MVO), and Moth-Flame Optimisation (MFO), for coupling with a Feedforward Neural Network (FNN) in addressing missing hourly rainfall observations, while overcoming the limitation of conventional training algorithm of artificial neural network that often traps in local optima. The proposed GWOFNN, MVOFNN, and MFOFNN were compared against the conventional Levenberg Marquardt Feedforward Neural Network (LMFNN) in addressing the artificially introduced missing hourly rainfall records of Kuching Third Mile Station. The findings show that the proposed approaches are superior to LMFNN in predicting the 20% hourly rainfall observations in terms of mean absolute error (*MAE*) and coefficient of correlation (*r*). The best performance ANN model is GWOFNN, followed with MVOFNN, MFOFNN and lastly LMFNN.

Keywords: Grey Wolf Optimizer (GWO), Multi-Verse Optimizer (MVO), Moth-Flame Optimization (MFO), artificial neural network, missing hourly rainfall observations

1. Introduction

Rainfall is one of the most important variables that often utilized and included for further hydrological and climatological modelling, analysis and simulation studies [1]. It is required to have long term, consistent, and high resolution of climate variables that include the sub-daily rainfall observations for executing them accurately [2]. However, the occurrences of missing rainfall observations are unavoidable, mostly due to human errors in data management, natural disasters, and machinery defects on site when measuring and collecting the rainfall data. There are three conventional approaches for addressing missing data observations in various fields of research as the conventional data pre-processing approaches, and they are hot-deck imputation, listwise deletion, and zero imputation [3], [4]. Hot-deck imputation is currently adopted in Malaysia to address the missing rainfall observations by substituting the missing entries with the entries that were measured simultaneously from neighboring rainfall measurements [5].

**Corresponding author: kkuok@swinburne.edu.my*

However, the conventional approaches are not reliable, and they are not scientifically supported to be accurate and feasible in all the scenarios. Thus, data imputation approaches are encouraged by researchers to deal with the missing variables.

Artificial Neural Network (ANN) is one of the most popular Artificial Intelligence (AI) based empirical methods in predicting hydrological and climate variables [6]. The reason that ANNs are popular for hydrological predictions is due to the ability to account for the non-linear relationships of hydrological and climatological variables, and they do not rely on high data demand in order to achieve reliable and accurate predictions, which make them much more preferrable than other physical based and conceptual models [7], [8]. Back-Propagation Neural Network (BPNN) had successfully applied for estimating the monthly rainfall runoff of the Hub River in Pakistan, and reliable estimations could be obtained via BPNN [9]. Wavelet Neural Network model (WNN) was found outperformed than the conventional Multilayer Perceptron (MLP) for predicting the monthly rainfall of the Darjeeling rain gauge station [10]. On the other hand, ANNs are also proven to be feasible in addressing missing rainfall data [11]. Through the performed rainfall imputation study, the Levenberg-Marquardt (LM) algorithm was found to be superior in training the Feedforward Neural Network (FNN) when compared against the Conjugate Gradient Fletcher–Reeves update (CGF) algorithm and Broyden–Fletcher–Goldforb–Shanno (BFG) algorithm [11]. In addition to that matter, the LM algorithm is also highly recommended by MATLAB software as the supervised algorithm for training ANN due to its superior performance and time efficiency [12]. Apart from that, other AI-based empirical imputation models such as K-Nearest Neighbor (KNN) [1], [13], Sequential K-Nearest Neighbor (SKNN) [14], Regularized Expectation Maximization (RegEM) [15], Bayesian Principal Component Analysis (BPCA) [16], Expectation Maximization (EM) [17], and probabilistic principal component analysis (PPCA) [18] are also proven to be reliable in addressing missing hydrological and climatological variables.

Although ANNs are proven to be robust in various hydrological and climatological predictions, the conventional training algorithms of ANNs often traps in local optima, which it fails to achieve the global optimum hence limiting the prediction accuracy [19], [20]. In other words, the solution set in terms of Weights (W) and Biases (B) trained by the conventional training algorithms during the training stage is not feasible to be applied in the testing stage to make the predictions. In this case, metaheuristic algorithms can be applied to replace the conventional training algorithms of ANNs for local optima avoidance, which are simple and flexible for various optimization attempts [21]. Metaheuristic algorithms are designed by referring to the inspiration of natural systems, where they are mathematically formulated and developed based on the concept of evolution in nature, physics laws, or behavior of living organisms [22]. It is currently a popular approach to be coupled with ANNs to enhance the performance of the conventional ANNs. Grey Wolf Optimizer (GWO) algorithm was applied in training MLP for classifying Exclusive or (XOR), balloon, iris, breast cancer, and heart datasets [23]. It was found that the GWO algorithm outperformed the Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Ant Colony Optimization (ACO), Evolution Strategy (ES), and Population-based Incremental Learning (PBIL) in training MLP for the classification tasks. Similarly, Moth-flame Optimization (MFO) algorithm developed by Mirjalili was used to train MLP networks for solving classification problem [24], [25]. A competitive result can be observed by the authors when comparing the MFO algorithm against PSO, GA, ACO, and ES optimization approach in training MLP for classifying five different datasets obtained from the University of California at Irvine (UCI) Machine Learning Repository. On the other hand, hybrid approach of ANN with the Multi-Verse Optimizer (MVO) algorithm was adopted to predict the streamflow [26]. The findings showed that the MVO-ANN model is superior to the other ANN models trained by PSO and BP algorithms. The dragonfly algorithm (DA) medical prediction and classification tasks. It was found that the DA-ANN approach could outperform the existing classification approaches such as Bayes Network training with hill climbing algorithm, Naive Bayes, and KNN classifier.

From the review above, it shows that replacing the conventional training algorithms of ANNs with metaheuristic algorithms seems to be a plausible and reliable solution to enhance the prediction performance of ANNs. However, there is a lack of hourly imputation studies, especially the hourly rainfall imputation studies, where the recent rainfall imputation studies mostly focused on addressing annually, monthly, and daily scaled rainfall observations [27]. Thus, it inspired this study to adopt the GWO, MVO, and MFO algorithm to be coupled with FNN for addressing missing hourly rainfall observations. To the best of the authors' knowledge that the selected metaheuristic algorithms are yet applied together with FNN for addressing any missing sub-daily hydrological variables, especially for missing hourly rainfall observations. Hence, a case study was formulated and executed in this study for exploring the feasibility and robustness of the GWOFNN, MVOFNN, and MFOFNN in addressing missing hourly rainfall observations. By accounting and considering the aforementioned issues, the objectives of this study are outlined as listed below:

a) To predict the hourly missing rainfall observations with the formulated hybrid approaches of the selected metaheuristic algorithms with FNN (GWOFNN, MVOFNN, and MFOFNN),
b) To compare the performance of GWOFNN, MVOFNN, and MFOFNN against an existing conventional ANN, LMFNN,
c) To evaluate the reliability and feasibility of GWOFNN, MVOFNN, and MFOFNN in addressing missing hourly rainfall observations.

2. Study Area, Datasets and Assumptions

2.1 Study Area

This research utilized the hourly rainfall observations from the rainfall stations located within the Sarawak River Basin (Fig. 1), where the locations of the selected rainfall stations are illustrated in Fig. 2. The selection of the rainfall stations was made by utilizing two to three neighboring rainfall stations as the interpolation reference for predicting the missing rainfall observations [28]. The other rainfall stations that are further from each other and with lower correlation relationships are not recommended as it will result in low imputation accuracy. The correlation relationships (r) between the selected rainfall stations were tabulated in Table 1 using Eq. (1).

$$r = \frac{\sum\left(A-\bar{A}\right)\left(B-\bar{B}\right)}{\sqrt{\sum\left(A-\bar{A}\right)^2 \sum\left(B-\bar{B}\right)^2}} \tag{1}$$

whereas A is the rainfall observations at Station A, $\bar{A}$ is the mean of rainfall observations at Station A, B is the rainfall observations at Station B and $\bar{B}$ is the mean of rainfall observations at Station B.

2.2 Datasets and Assumptions

The hourly rainfall observations from January 2009 to June 2009 were collected from DID Sarawak, where there were 4344 samples available for the formulation of train and test datasets, where the samples were divided into the ratio of 80/20 for train dataset (3475 samples) and test dataset (869 samples). It was also assumed that there were 20% missing hourly rainfall observations at Kuching Third Mile Station, where the rest of the rainfall stations were used as the interpolation reference to predict the 20% artificial missing data entries. By referring to Table 1, it is observed that there is a medium ($0.7 < r \leq 0.3$) to high ($1 < r \leq 0.7$) correlation relationship of the hourly rainfall data between all the rainfall stations. On the other hand, there is a significant pattern observed when comparing the distance between the rainfall stations and the correlation relationship, where the further the distance between the rainfall stations, the lower the correlation relationship between the rainfall station. In this case, the usage of the rainfall variables was indicated to be appropriate as high correlation relationships can be observed ($r > 0.7$) relative to the Kuching Third Mile Station [28]. On the other hand, the missingness mechanism of this imputation study was assumed to be missing completely at random (MCAR), where this makes the imputation study to be fallen within an ignorable case. This assumption should be reasonable since the missing rainfall data in Malaysia is mainly contributed by human errors in data management [17], where the occurrences of missing rainfall data do not depend on any occurrences of any random events. Hence, the missingness mechanism was not accounted for in this research study.

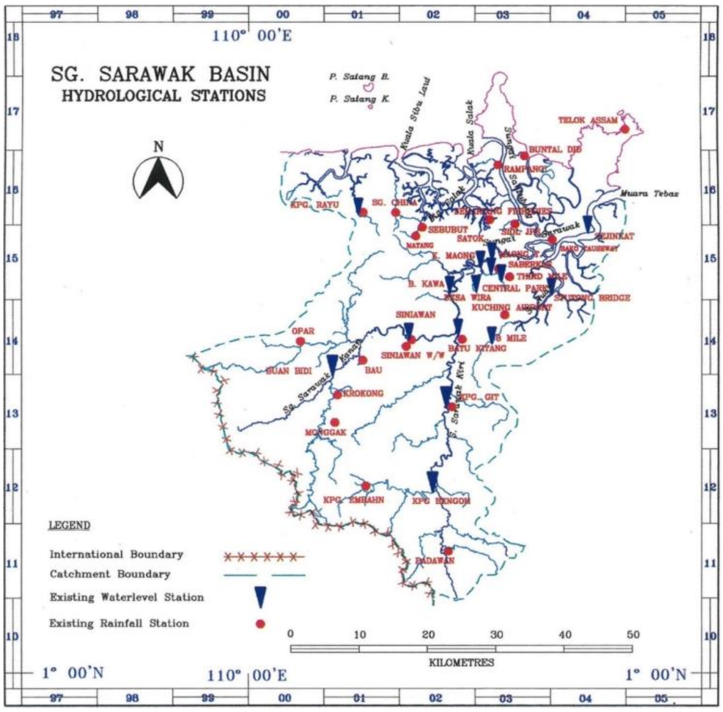

Fig. 1 - The Sarawak river basin [29]

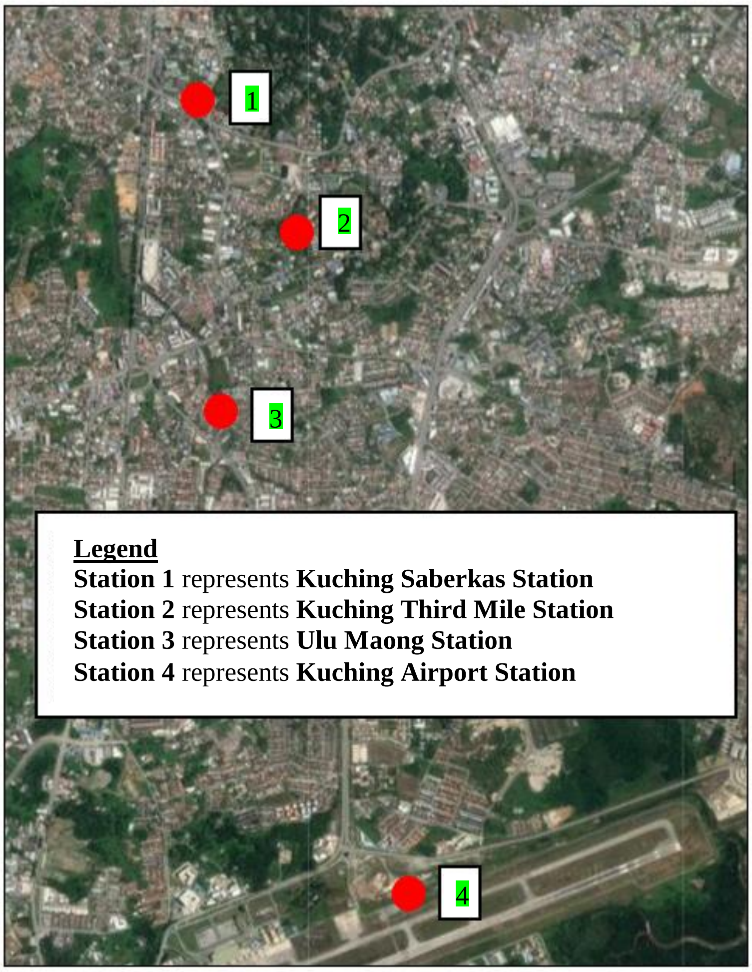

Fig. 2 - The selected rainfall stations [30]

Table 1 - Correlation relationship of the hourly rainfall observations between the selected rainfall stations

Station A	Station B	*r*
Kuching Third Mile	Kuching Saberkas	0.9307
	Ulu Maong	0.9518
	Kuching Airport	0.7478
Kuching Saberkas	Ulu Maong	0.8714
	Kuching Airport	0.6414
Ulu Maong	Kuching Airport	0.8024

3. ANN Based Imputation Approaches and Evaluation Criteria

This section is utilized to discuss some of the key features of the metaheuristic algorithms together with the formulated approaches and the conventional LMFNN. Further details of the metaheuristic algorithms can be obtained via each of the cited works. The hyper-parameters of the ANN based models are presented, and they were optimized via *k*-fold cross validation and the trial-and-error approaches.

3.1 Metaheuristic Algorithms

3.1.1 Grey Wolf Optimizer (GWO)

GWO algorithm was developed (via the inspiration of the social hierarchy and the prey hunting behavior of grey wolves' packs (as presented in Fig. 3) [21]. The wolves pack is led by the alpha (α) wolves, followed by beta (β), delta (δ), and lastly, the omega (ω) wolves as the lowest rank wolf in the society, where the wolves at the lower rank have to obey the wolves at the higher rank in their social hierarchy. As such, the GWO algorithm assumes that α has the best solution, followed by *β*, *δ*, and lastly, *ω*. The exploration of the GWO algorithm mimics the prey tracking and searching behavior of the wolves' pack, while the exploitation process of the GWO algorithm mimics the last phase of the hunting, where the wolves converge and attack the prey that stopped to move. Thus, the optimization process of the GWO algorithm starts with the initialization of random population of wolves (also known as the candidate solution or search agent), followed by the measurement of the position of targeted prey by *α*, *β*, and *δ* wolves, and the assigned search agents are required to update their respective distance from the targeted prey. The updating and optimization of the GWO algorithm will continue until it reaches the termination criterion, which is the number of maximum iterations [21].

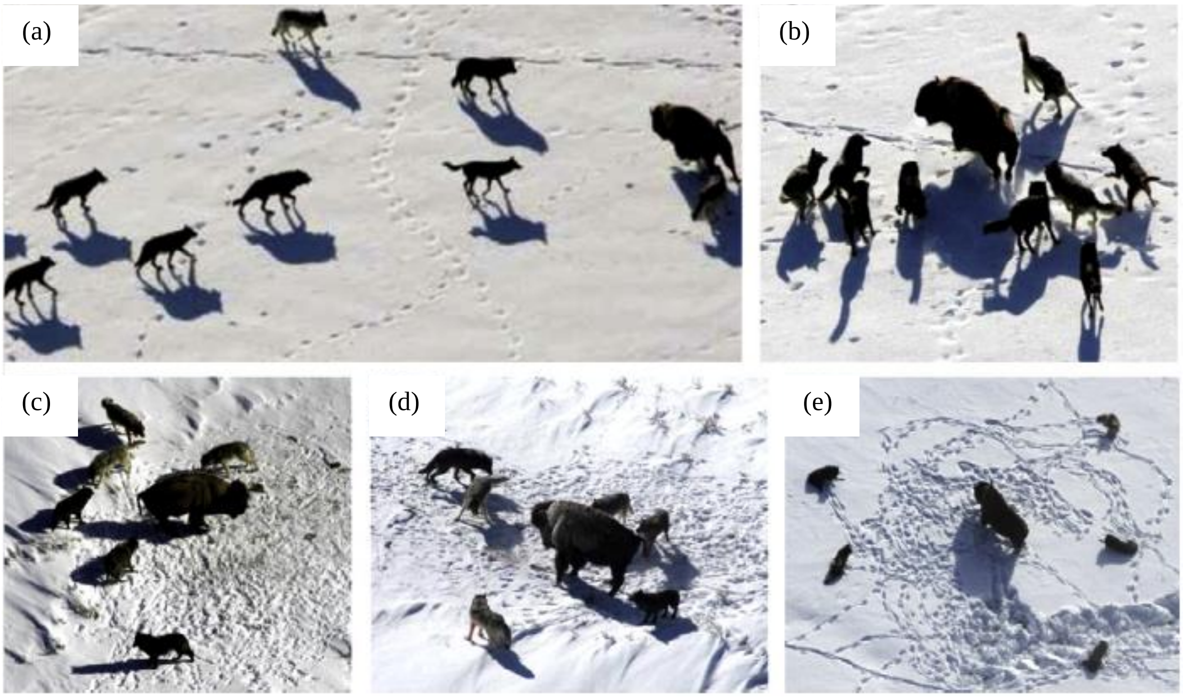

Fig. 3 - The hunting mechanism of grey wolves (a) tracking and approaching the prey; (b) to (d) following and encircling the prey, and; (e) attacking the prey when the prey stops to move [21]

3.1.2 Multi-Verse Optimizer (MVO)

MVO algorithm was created via the inspiration of the multi-verse concept, where the concept of white holes, black holes, and wormholes (as illustrated in Fig. 4) is utilized to formulate the optimization mechanism of the MVO algorithm [31]. In order to mathematically formulate the multi-verse concept, the following rules are adopted with:

- The concepts of white holes and black holes are utilized as the exploration mechanism, and the concept of wormholes is utilized as the exploration mechanism of the MVO algorithm.
- Each of the generated solutions is treated as a universe, and each variable of the solution is treated as an object from a universe.
- All the solutions are assigned with their respective inflation rate, where the inflation rate is set to be proportional to the corresponding objective function value or fitness value.
- The probability of white holes is higher when a higher inflation rate is observed and vice versa for the existence of black holes.
- The universes with higher inflation rates will send the objects out via white holes, where the objects will be received by universes with lower inflation rates via black holes. By utilizing this rule, it allows the average inflation rate of all the universes to improve across the iterations.
- All the objects from all the universes may move randomly towards the best universe through the wormholes without considering their respective inflation rates.
- The roulette wheel mechanism is utilized to formulate the mechanism and concept of white holes and black holes and the exchange and movement of objects between universes.

The optimization process of the MVO algorithm is started with the random initialization of the universes and followed by the exchange of objects between the universes via white holes and blackholes, where the universes with higher inflation rates will tend to send the object out with random travelling patterns to be received by the universes with lower inflation rates via blackholes. The process of the exchange of objects within the universes will continue until the pre-determined end criterion, the number of maximum iterations, is achieved, where the inflation rates of the universes are set to be improved continuously across the iterations.

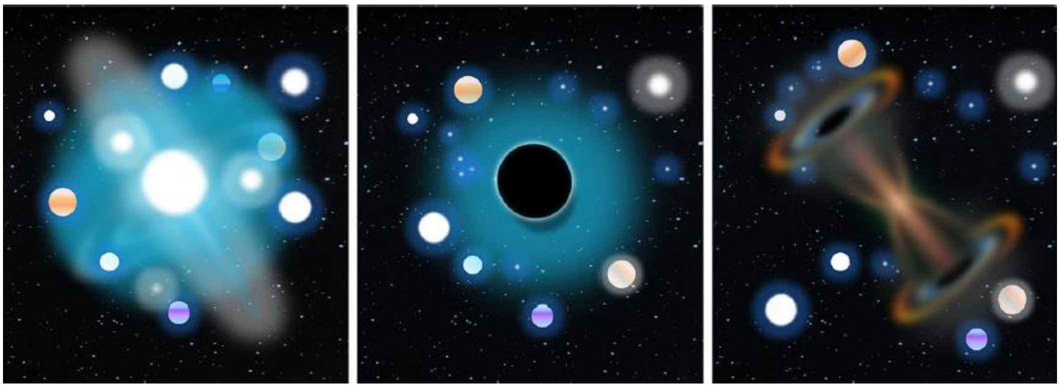

Fig. 4 - The white hole, black hole, and wormhole (from left to right) [31]

3.1.3 Moth-Flame Optimization (MFO)

MFO algorithm formulated was formulated using the concept of the navigation system of moth at night, which is also known as the transverse orientation, where the moths usually reference to the moonlight in order to fly in a fixed angle [25]. However, the MFO algorithm mimics the movement of moths when the moths are confused by artificial lights, where the moths will fly around the artificial lights as illustrated in Fig. 5. In order to mimic the movement of moths, the following rules are applied within the MVO algorithm, where:

- The moths are treated to be the search agents and candidate solutions.
- The logarithmic spiral is chosen to be the movement of the moths in the MFO algorithm, where the initial and final point of the spiral is set to be the position of moth and flame, respectively, and the spiral's fluctuation range is limited within the search space.
- Each of the moths is assigned to each flame, where moths are treated as the actual search agents which move and explore around the search space. On the contrary, flames are treated as the best position of moths obtained so far throughout the iterations. In addition to that matter, each of the moths is required to update its position by referring to only one flame.
- The number of flames is set to be reduced across the iterations to aid with the exploitation of the promising solution.
- The first moth is assumed to have the best solution and vice versa for the last moth within the population of moths.
- The moths will search and update its position continuously for finding a better solution. This feature makes the MFO algorithm never loses its best solution across the iterations.

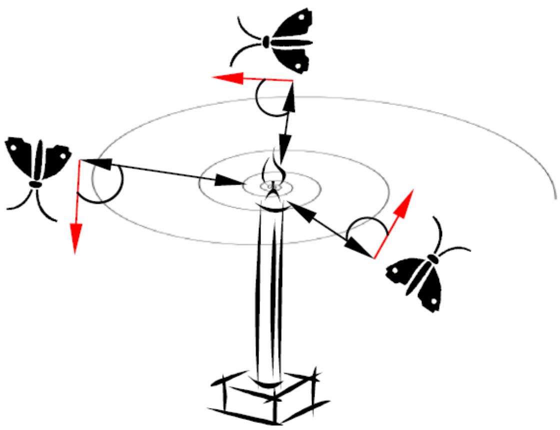

Fig. 5 - Spiral movement of moths around an artificial light [25]

The optimization process of the MFO algorithm starts with random generation of positions of moths within the solution space, where the corresponding objective function value will also be tabulated at the same time. The position of moths is updated in a logarithmic spiral pattern to achieve better positions with respect to each flame, while the number of flames will also be reduced simultaneously across the iterations to aid with the exploitation of the candidate solutions. The updating process will continue until it reaches the maximum number of iterations.

3.1.4 Copulation of Metaheuristic Algorithm with Feedforward Neural Networks (FNN)

The selected metaheuristic algorithms were utilized for training the single layered FNN (as illustrated in Fig. 6, where the optimal weights (*W*) and biases (*B*) of the FNN will be determined through the optimization process of the metaheuristic algorithms. The overall workflow of the prediction process of GWOFNN, MVOFNN, and MFOFNN can be observed in Fig. 7, where *W* and *B* are set to be optimized across the iterations with respect to an objective function. The mathematical expression of the FNN utilized in this study can be observed in Eq. (2), while a normalization function, as presented in Eq. (3), is also included to normalize all the rainfall variables within the range of 0 to 1. The log-sigmoid transfer function is utilized in the hidden layer to handle the non-linear rainfall variables. On the contrary, the positive linear transfer function is utilized in the output layer to transfer the information directly while eliminating the negatively predicted entries. These transfer functions are utilized in this study as they do not allow any negative predicted output to be generated in the output matrix, as the minimum measurement of rainfall is zero.

$$Y = poslin\left(W_o\left[\log sig\left(W_H X + B_H\right)\right] + B_o\right) \tag{2}$$

whereas *Y* is the output matrix of FNN, *X* is the input data matrix, W_H and W_o are weight matrices for the hidden layer and output layer respectively, B_H and B_O are the bias matrices for the hidden layer and output layer respectively, *poslin* is positive linear transfer function and log*sig* is log-sigmoid transfer function.

$$X = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{3}$$

whereas *X* is normalized value of rainfall, X_i is the original rainfall value at i^{th} entry, X_{max} is the maximum rainfall value among the rainfall stations and X_{min} is the minimum rainfall value among the rainfall stations where it is 0 in this case.

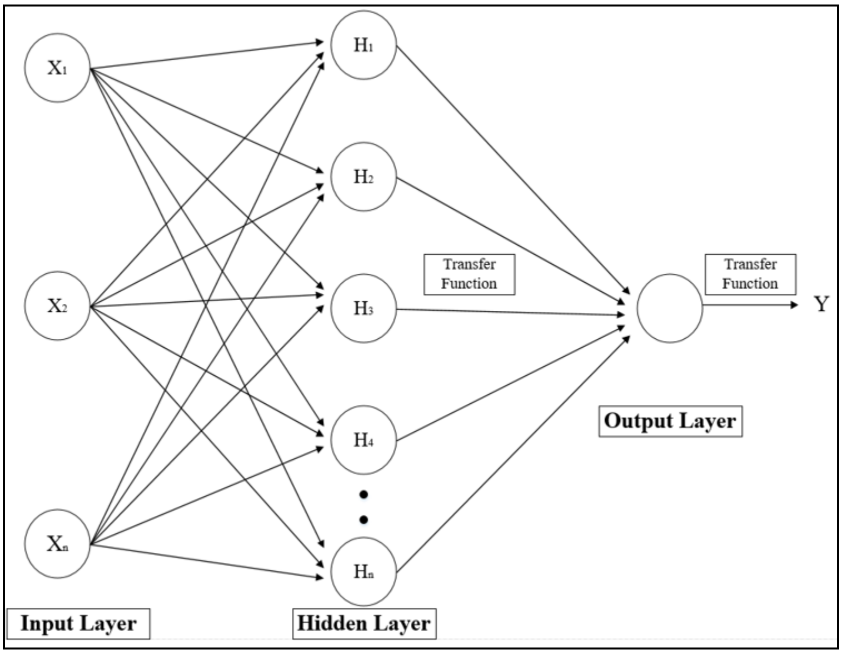

Fig. 6 - The single layer FNN

The Mean Absolute Error (*MAE*), as presented in Eq. (4), is utilized as the objective function in this study, where the optimal *W* and *B* are usually determined at the end of the iterations with the lowest objective function value shown at the end of the iterations. The optimal *W* and *B* will be adopted during the testing stage for predicting the missing hourly rainfall observations of Kuching Third Mile Station.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|O_i - F_i\right| \tag{4}$$

whereas *N* is the total number of predicted data, F_i is the i^{th} predicted value and O_i is the i^{th} actual value.

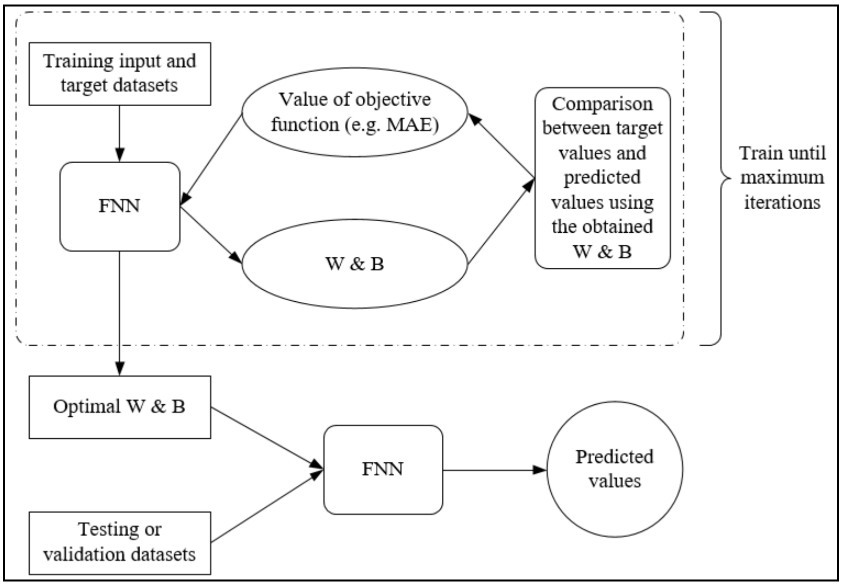

Fig. 7 - The optimization and prediction process of GWOFNN, MVOFNN, and MFOFNN

3.2 Levenberg Marquardt Feedforward Neural Network (LMFNN)

The conventional built-in LMFNN model in the MATLAB software was utilized in this study as a comparison against the proposed metaheuristic-based ANN models for predicting the missing hourly rainfall records of Kuching Third Mile Station. Similar to the proposed metaheuristic-based ANN models, the similar architecture structure of FNN and transfer functions were also utilized in LMFNN. The only difference between the LMFNN and the metaheuristic-based ANN models is the way of optimizing the W and B during the training stage, where LMFNN utilized the conventional LM algorithm to optimize the solution set.

3.3 Hyper-Parameters of the ANN Based Imputation Approaches

In order to avoid overfitting and underfitting issues of the ANN based models, the hyper-parameters of the ANN based imputation approaches, as presented in Table 2, were optimized using the trial-and-error approach and 5-fold cross validation, using the training data, where each of the hyper-parameter consists of 5 different settings. The common hyper-parameters of the ANN based models are the number of hidden nodes and the number of iterations, while for the metaheuristic-based ANN models, there is an additional hyper-parameter, which is the number of search agents. These hyper-parameters settings are optimized one by one using 5-fold cross validation to determine their optimal hyper-parameters recommended k = 5 the number of folds to perform the k-fold cross validation due to the fact that this value was empirically proven to be capable of achieving a balance in terms of bias-variance trade-off under a limited amount of data availability [32]. The optimal hyper-parameters setting is set to be the hyper-parameters setting that resulted in the lowest average *MAE* for both training and validation errors (also known as the cross-validation score) when predicting the output of 5 different validation subsets. The 5 different validation subsets were divided equally from the training data, which then resulted in 695 samples for each validation subset. In other words, different combinations of 2780 hourly rainfall samples were utilized to validate each k^{th} validation subset that consisted of 695 samples.

Table 2 - Hyper-parameters settings of all the ANN based imputation approaches

ANN based models	Hyper-parameters	Settings of each hyper-parameter
LMFNN	• Number of hidden nodes • Number of iterations	• 5, 10, 15, 20, and 25 • 500, 750, 1000, 1250, and 1500
GWOFNN, MVOFNN, and MFOFNN	• Number of hidden nodes • Number of iterations • Number of search agents (SA)	• 5, 10, 15, 20, and 25 • 500, 750, 1000, 1250, and 1500 • 50, 100, 150, 200, and 250

3.4 Evaluation Criteria

Other than using r and *MAE* to evaluate the imputation performance of the ANN based imputation models, the criterion of Bias (B_S), as presented in Eq. (5), was also utilized to measure the degree to which the imputation model overestimated ($B_S > 1$) or underestimated ($B_S < 1$) the total missing hourly rainfall amount. A perfect estimation of the missing hourly rainfall values will result in MAE = 0 mm, r = 1, and B_S = 1. The criterion of MAE was utilized to measure the absolute average difference between the actual rainfall values and the predicted rainfall values. On the other hand, the criterion of r was utilized to measure the positive linear relationship between the actual rainfall values and the predicted rainfall values, while this also presents the similarities between the actual rainfall pattern and the predicted rainfall pattern of all the utilized ANN models. The graphical illustrations of actual rainfall versus the predicted rainfall values were also attempted to aid with the analysis and visualization of the predicted rainfall pattern against the actual rainfall pattern.

$$B_s = \frac{\sum_{i=1}^{N} F_i}{\sum_{i=1}^{N} O_i} \tag{5}$$

where *F* represents the imputed values or predicted values and *O* represents the actual values or observed values.

4. Results and Discussion

4.1 Optimal Configurations of Each ANN Models

The optimal configurations of each ANN based models are presented in Table 3, together with their respective cross validation scores. From the tabulated results, it shows that the adopted hyper-parameters settings of all the ANN based models fit well with the problem as the average validation errors are slightly higher than their respective average training errors, which shows that the adopted hyper-parameters settings are neither overfits nor underfits the problem. Hence, they are considered appropriate and adequate to be adopted during the testing stage for predicting the missing hourly rainfall entries of Kuching Third Mile Station. However, from the tabulated values, it shows that all the proposed metaheuristic-based ANN approaches outperformed the conventional LMFNN model, which shows that replacing the LM algorithm with metaheuristic algorithms could enhance the performance and accuracy of ANN. The performance of the proposed metaheuristic-based ANN approaches is further validated and compared with LMFNN in the testing stage when predicting the 20% artificial missing hourly rainfall observations with respect to *MAE*, *r*, and B_S.

Table 3 - Optimal configurations of each ANN based models and their respective cross validation scores with respect to MAE

ANN Model	Optimized Hyper-parameters	Cross Validation Score (mm)	
		Training	Validation
LMFNN	• 5 hidden nodes and 500 iterations *(*log*sig – poslin)*	0.1631	0.2050
GWOFNN	• 150 SA, 5 hidden nodes, and 1500 iterations *(*log*sig – poslin)*	0.1456	0.1555
MVOFNN	• 250 SA, 5 hidden nodes, and 1500 iterations *(*log*sig – poslin)*	0.1502	0.1600
MFOFNN	• 250 SA, 5 hidden nodes, and 1500 iterations *(*log*sig – poslin)*	0.1535	0.1770

#The optimal hyper-parameters are determined via the trial-and-error approach and k-fold cross validation attempts.

4.2 Prediction of Missing Hourly Rainfall Observations

Table 4 tabulates the imputation performance of all the ANN based models when estimating the artificial missing hourly rainfall entries of Kuching Third Mile Station. From the tabulated result, it shows that all the ANN models achieved lower test errors than their respective train error. Normally, one would expect that the test errors should be higher than the training errors. This might be since the ratio and number of extremes in the training dataset is higher than the test dataset, where the training dataset consisted of the larger sample size and with a longer period of time when compared to the test dataset. This is also reflected from the illustrations of the actual rainfall versus the predicted rainfall values of all the ANN based models in Fig. 8 to Fig. 11, where the illustrations show that the number of non-zero value entries and extremes is lower than the zero value entries, hence causing the test dataset seems to be easier to be predicted with lower test errors for all the ANN based models. Overall, a competitive performance can be observed between all the ANN based models when predicting the 20% missing hourly rainfall observations. However, all the ANN models have a slight tendency to overestimate and underestimate the total hourly rainfall amount. On the other hand, all the ANN based models are performing adequately as test high correlation values can be observed ($0.9402 \geq r \geq 0.9685$), together with a very similar rainfall pattern predicted by all the ANN based models when compared against the actual rainfall pattern. However, in terms of *r* and *MAE*, all the proposed metaheuristic-based models are superior to the conventional LMFNN model. However, in terms of the criterion of B_S, only MFOFNN is capable of outperforming LMFNN as it achieved a B_S that is closer to 1 when compared against LMFNN. In a nutshell, the evidence shows that replacing the conventional training algorithm of ANN with metaheuristic algorithms is feasible to enhance the imputation performance of ANN, especially in terms of *MAE* and *r*.

Table 4 - Imputation performance of all the ANN based models in estimating the 20% artificial missing hourly rainfall entries

Models	Criteria	LMFNN	GWOFNN	MVOFNN	MFOFNN
Train	***MAE*** **(mm)**	0.1700	0.1459	0.1446	0.1634
	Bs	0.9796	0.9607	0.9263	0.9020
	r	0.8846	0.9732	0.9714	0.9666
Test	***MAE*** **(mm)**	0.0705	0.0575	0.0572	0.0597
	Bs	0.9789	1.0441	0.9770	1.0101
	r	0.9402	0.9685	0.9627	0.9568

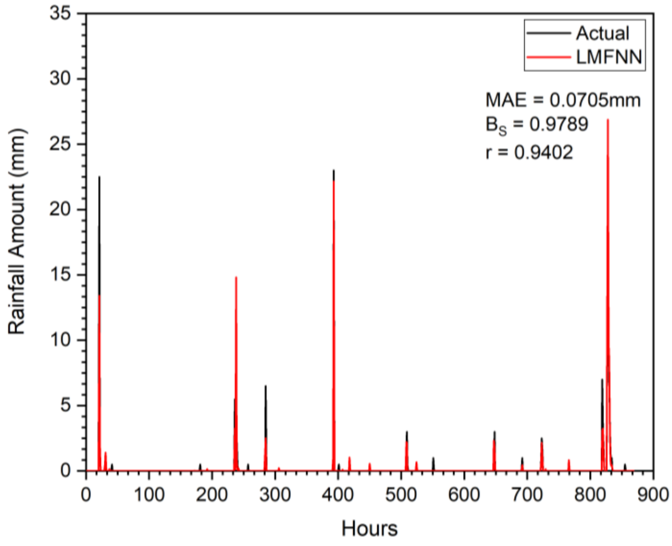

Fig. 8 - Actual versus the predicted hourly rainfall of LMFNN

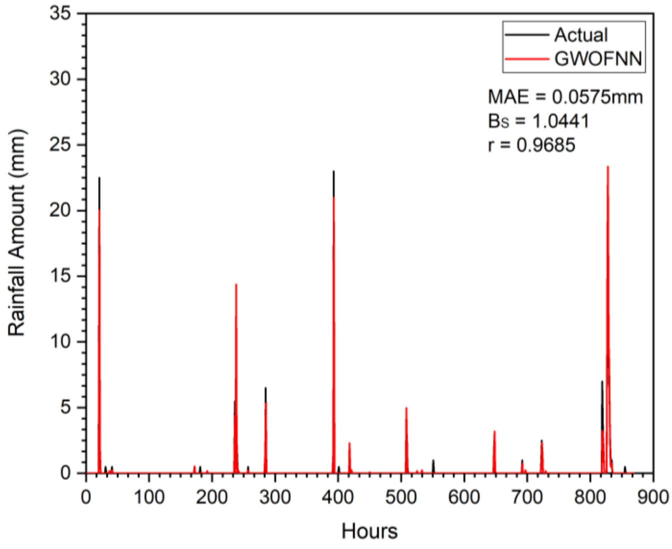

Fig. 9 - Actual versus the predicted hourly rainfall of GWOFNN

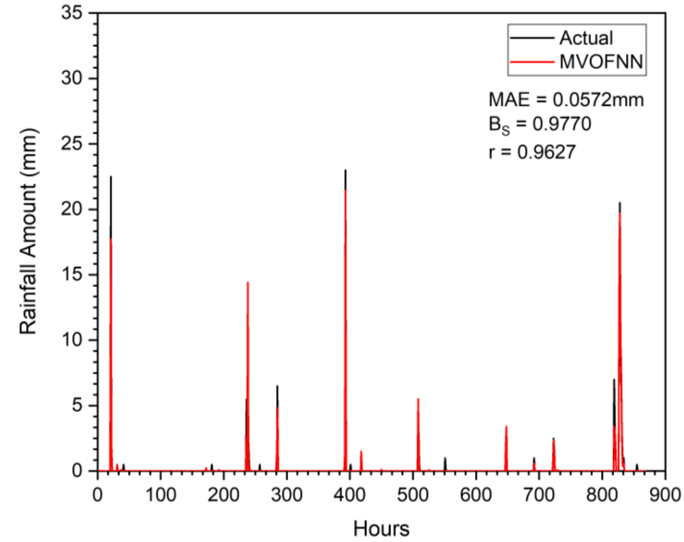

Fig. 10 - Actual versus the predicted hourly rainfall of MVOFNN

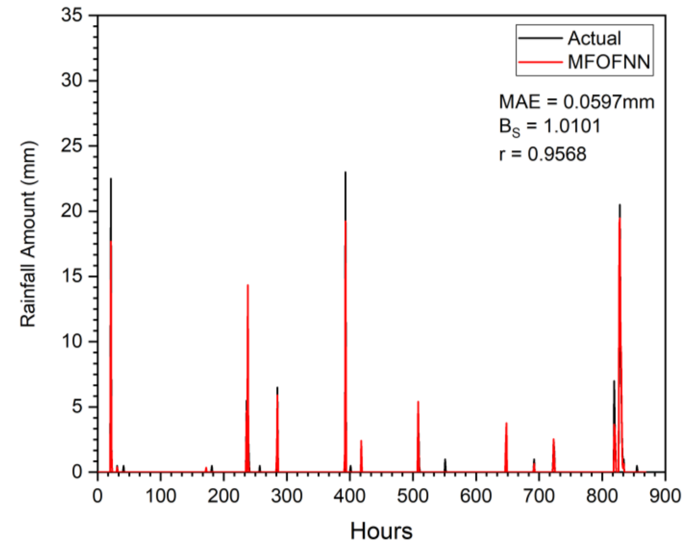

Fig. 11 - Actual versus the predicted hourly rainfall of MFOFNN

5. Conclusion

This research study proposed three novel ANNs based hourly imputation models that coupled with the GWO, MVO, and MFO algorithm to overcome the limitation of conventional training algorithm of ANN that often traps in local optima hence limiting the imputation performance. The proposed metaheuristic-based ANN models (GWOFNN, MVOFNN, and MFOFNN) were formulated in this research study to predict the 20% artificial missing hourly rainfall observations of Kuching Third Station by referring to the hourly rainfall observations of neighboring rainfall stations, where 80% of the collect data samples were utilized for training. The performance of the proposed metaheuristic-based ANN models was compared against a conventional ANN model, LMFNN. In order to avoid overfitting and underfitting issues of the utilized ANN models, their hyper-parameters were optimized with respect to MAE via the trial-and-error approach and k-fold cross validation. Their optimal configurations with the lowest achievable errors in terms of MAE during the validation stage were adopted in the testing stage for estimating the missing hourly rainfall observations of Kuching Third Mile Station. A competitive imputation performance can be observed between all the utilized ANN based models. In addition to that matter, all the proposed metaheuristic-based ANN models are superior to LMFNN, especially in terms of *MAE* and *r*. The best performance ANN model is GWOFNN, followed with MVOFNN, MFOFNN and lastly LMFNN. For future research direction, it is recommended to apply the proposed metaheuristic-based ANN models to predict other hourly hydrological or climatic variables such as runoff, evapotranspiration, or wind speed to test their robustness and feasible in estimating other hydrological and climatic variables. On the other hand, they can be also applied in other geographical locations with different climate types for further exploring their imputation performance in other study areas.

References

[1] Lee, H. & Kang, K. (2015). Interpolation of Missing Precipitation Data Using Kernel Estimations for Hydrologic Modeling. Advances in Meteorology, 2015, 1-12. https:doi.org/10.1155/2015/935868

[2] Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L.V., Hegerl, G., Seneviratne, S.I., Vautard, R., Zhang, X. & Zwiers, F.W. (2017). Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. Weather and Climate Extremes, 18, 65-74. https://doi.org/10.1016/j.wace.2017.10.003

[3] Pagano, T.C., Wood, A.W., Ramos, M-H., Cloke, H.L., Pappenberger, F., Clark, M.P., Cranston, M., Kavetski, D., Mathevet, T., Sorooshian, S. & Verkade, J.S. (2014). Challenges of Operational River Forecasting. Journal of Hydrometeorology, 15(4), 1692-1707. https://doi.org/10.1175/JHM-D-13-0188.1

[4] Chiu, P.C., Selamat, A., Krejcar, O. & Kuok, K.K. (2021). Imputation of Rainfall Data Using Improved Neural Network Algorithm. In International Conference on Pattern Recognition. Springer International Publishing, pp. 390-406.

[5] Chiu, P.C., Selamat, A., Krejcar, O. & Kuok, K.K. (2019). Missing Rainfall Data Estimation Using Artificial Neural Network and Nearest Neighbor Imputation. In SoMeT. IOS Press, pp. 132-143.

[6] Devia, G.K., Ganasri, B.P. & Dwarakish, G.S. (2015). A Review on Hydrological Models. Aquatic Procedia, 4, 1001-1007. https://doi.org/10.1016/j.aqpro.2015.02.126

[7] Dawson, C.W. & Wilby, R.L. (2001). Hydrological modelling using artificial neural networks. Progress in Physical Geography, 25(1), 80-108. https://doi.org/10.1177/030913330102500104

[8] Kuok, K.K., Chiu, P.C. & Selamat, A. (2019). Comparison of Bat Neural Networks and Bat Optimisation Neural Networks for Rainfall Forecasting: Case Study for Kuching City. In SoMeT. IOS Press, pp. 79-92.

[9] Ghumman, A.R., Ghazaw, Y.M., Sohail, A.R. & Watanabe, K. (2011). Runoff forecasting by artificial neural network and conventional model. Alexandria Engineering Journal, 50(4), 345-350. https://doi.org/10.1016/j.aej.2012.01.005

[10] Venkata Ramana, R., Krishna, B., Kumar, S.R. & Pandey, N.G. (2013). Monthly Rainfall Prediction Using Wavelet Neural Network Analysis. Water Resources Management, 27(10), 3697-3711. https://doi.org/10.1007/s11269-013-0374-4

[11] Roman, U.C., Patel, P.L. & Porey, P.D. (2012). Prediction of missing rainfall data using conventional and artificial neural network techniques. ISH Journal of Hydraulic Engineering, 18(3), 224-231. https://doi.org/10.1080/09715010.2012.721660

[12] The MathWorks, Inc. (2020). Trainlm. https://www.mathworks.com/help/deeplearning/ref/trainlm.html>

[13] Schneider, T. (2001). Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values. Journal of Climate, 14(5), 853-871. https://doi.org/10.1175/1520-0442(2001)014<0853:AOICDE>2.0.CO;2

[14] Lai, W.Y., Kuok, K.K., Gato-Trinidad, S. & Ling, K.X. (2019). A study on Sequential K-Nearest Neighbor (SKNN) imputation for treating missing rainfall data. International Journal of Advanced Trends in Computer Science and Engineering, 8(3), 363-368. https://doi.org/10.30534/ijatcse/2019/05832019

[15] Miró, J.J., Caselles, V. & Estrela, M.J. (2017). Multiple imputation of rainfall missing data in the Iberian Mediterranean context. Atmospheric Research, 197, 313-330. https://doi.org/10.1016/j.atmosres.2017.07.016

[16] Lai, W.Y. & Kuok, K.K. (2019). A Study on Bayesian Principal Component Analysis for Addressing Missing Rainfall Data. Water Resources Management, 33(8), 2615-2628. https://doi.org/10.1007/s11269-019-02209-8

[17] Malek, M.A. (2008). Rainfall Data in-Filling Model with Expectation Maximization and Artificial Neural Network. PhD Thesis, Universiti Teknologi Malaysia.

[18] Teegavarapu, R.S.V. (2014). Missing precipitation data estimation using optimal proximity metric-based imputation, nearest-neighbour classification and cluster-based interpolation methods. Hydrological Sciences Journal, 59(11), 2009-2026. https://doi.org/10.1080/02626667.2013.862334

[19] Gill, M.K., Asefa, T., Kaheil, Y. & McKee, M. (2007). Effect of missing data on performance of learning algorithms for hydrologic predictions: Implications to an imputation technique. Water resources research, 43(7), 1-12. https://doi.org/10.1029/2006WR005298

[20] Louis, Y.H.T., Kuok, K.K., Imteaz, M., Lai, W.Y. & Derrick, K.X.L. (2019). Development of whale optimization neural network for daily water level forecasting. International Journal of Advanced Trends in Computer Science and Engineering, 8(3), 354-362. https://doi. org/10.30534/ijatcse/2019/04832019.

[21] Mirjalili, S., Mirjalili, S.M. & Lewis, A. (2014). Grey Wolf Optimizer. Advances in Engineering Software, 69, 46-61. https://doi.org/10.1016/j.advengsoft.2013.12.007

[22] Kuok, K.K. & Chiu, P.C. (2013). Particle Swarm Optimization and Shuffle Complex Evolution for Calibrating Xinanjiang Model Parameters. International Journal of Computers & Technology, 10(10), 2036-2048.

[23] Mirjalili, S. (2015). How effective is the Grey Wolf optimizer in training multi-layer perceptrons. Applied Intelligence, 43(1), 150-161. https://doi.org/10.1007/s10489-014-0645-7

[24] Yamany, W., Fawzy, M., Tharwat, A. & Hassanien, A.E. (2015). Moth-flame optimization for training Multi-Layer Perceptrons. In 2015 11th International computer engineering Conference (ICENCO), 267-272. https://doi.org/10.1109/ICENCO.2015.7416360
[25] Mirjalili, S. (2015). Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. Knowledge-Based Systems, 89, 228-249. https://doi.org/10.1016/j.knosys.2015.07.006
[26] Peng, T., Zhou, J., Zhang, C. & Fu, W. (2017). Streamflow Forecasting Using Empirical Wavelet Transform and Artificial Neural Networks. Water, 9(6), 406. https://doi.org/10.3390/w9060406
[27] Hema, N. & Kant, K. (2017). Reconstructing missing hourly real-time precipitation data using a novel intermittent sliding window period technique for automatic weather station data. Journal of Meteorological Research, 31(4), 774-790. https://doi.org/10.1007/s13351-017-6084-8
[28] Kim, J-W. & Pachepsky, Y.A. (2010). Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. Journal of Hydrology, 394(3), 305-314. https://doi.org/10.1016/j.jhydrol.2010.09.005
[29] Department of Irrigation & Drainage (DID) Sarawak (2002). Hydrological Year Book 2002, Department of Irrigation & Drainage (DID) Sarawak, Kuching.
[30] Department of Irrigation & Drainage (DID) Sarawak (2019). Rainfall Stations. https://ihydro.sarawak.gov.my/iHydro/en/map/rainfall-map.jsp>
[31] Mirjalili, S., Mirjalili, S.M. & Hatamlou, A. (2016). Multi-Verse Optimizer: a nature-inspired algorithm for global optimization. Neural Computing and Applications, 27(2), 495-513. https://doi.org/10.1007/s00521-015-1870-7
[32] James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). Resampling methods. In an Introduction to Statistical Learning, Springer Texts in Statistics. Springer Publishing, pp. 175-201.