Combined Multiple Neural Networks and Genetic Algorithm with Missing Data Treatment: Case Study of Water Level Forecasting in Dungun River - Malaysia

Antoni Wibowo, *Member, IAENG,* Siti Hajar Arbain and Norhaslinda Zainal Abidin

Abstract— **We consider water level forecasting in Dungun River where the collected data contain missing values. Therefore, we cannot utilize a prediction technique to forecast the water level directly. To overcome this difficulty, we used Ordinary Linear Regression (OLR) and mean substitution to handle the imperfect data and to make the data meaningful. ARIMA and SARIMA are well known techniques and widely used in time series forecasting. Unfortunately, they produce a linear regression model that may improper model for water level forecasting. Instead, Backpropagation Neural Network (BPNN) and Nonlinear Autoregressive Exogenous Model (NARX) are alternative techniques to address the issue of linearity in regression. Nevertheless, they also have difficulties to determine the optimal network and regression coefficients/weights due to the randomness of their initial weights. Under this circumstance, we proposed Multiple BPNN and Genetic Algorithm (GA) to overcome the limitation of ARIMA/SARIMA, standalone BPNN and NARX. Our experiment showed that our proposed technique is superior compared to ARIMA, SARIMA, BPNN and NARX.**

Index Terms — **Genetic algorithm, missing data, neural network, water level.**

I. Introduction

THE stages of water level are designed to make local authority aware of the level of danger posed by the rising water level so that a necessary emergency arrangement could be initiated for the welfare of the local community affected by the river. As the water level forecasting could reduce the damage from the impact of flooding in agriculture, public uses, avoid both life and economic loss, it is therefore important to predict its appearance. Prediction of the pattern of water level is one of the benchmark points in the flood forecasting analysis and has been one of the most important issues in hydrological research. Water level is an essential component in the process of forecasting flood resources evaluation and is considered as a central problem in hydrology [1].

We consider the forecasting of the water level at the Dungun River in Terengganu – Malaysia which is a main river in Dungun District. Dungun District is one of the seven districts in the Terengganu state and located between $4^o36'10N$ to $4^o53'02N$ and $103^o07'25E$ to $103^o25'50E$ [2]. In reports of flooding in Dungun District, Department of Irrigation and Drainage (DID) stated that there are two types of flooding which are flash floods and river flood. Flash flood usually occurs in urban areas where it is usually caused by short, intense localized thunderstorm rains, where it is usually experienced during the evening [3]-[4]. Besides flash flood, there is also river flood usually happens when the flow in a river exceeds its conveyance capacity, the water in the river rises above its bank level and overspills into adjacent low-lying areas, causing river floods.

Data pre-processing is one of the most important steps before the application of statistical model, where it usually handles the imperfect characteristics of the produced data such as missing data and inconsistent value of data. The data pre-processing such as treatment of missing data can also influence the performance of the prediction model [5]-[6]. It is noticed that the original data that are collected from DID and Malaysian Meteorological Department (MMD) involve some imperfect characteristics that need to undergo the process of treatment of missing data before proceeding to the next method procedures. The collected data from the two departments involve months, monthly rainfall, rate of evaporation, rate of temperature, relative humidity and water level. The water level is treated as a response variable and the others are regressor variables. In this paper, the weekly data comprises a total number of 75 observation data from the year 2006 until 2012

In terms of forecasting techniques, it is reported that many analyses of forecasting time series approaches had been done in hydrological problems. The choice of the forecasting model is an important factor in order to improve the forecasting accuracy [7]. The application of forecasting is becoming increasingly popular in many real-world applications such as financial market prediction, electric utility load forecasting, weather and environmental state prediction, machining, internet resource, reliability forecasting and in social science research [8] - [15]. A well-

Manuscript received December 20, 2016; revised December 17, 2017.

Antoni Wibowo is with Computer Science Department, Binus Graduate Program-Master of Computer Science, Bina Nusantara University, Indonesia (e-mail: anwibowo@ binus.edu).

Siti Hajar Arbain is with Faculty Computer Science, Universiti Tun Hussein Onn Malaysia. She is now a PhD student at Faculty of Computing, Universiti Teknologi Malaysia (e-mail: shajar@live.utm.my).

Norhaslinda Zainal Abidin is with Department of Decision Sciences, School of Quantitative Science, Universiti Utara Malaysia (e-mail: nhaslinda@uum.edu.my).

known technique such as ARIMA and SARIMA are most commonly used for time series forecasting, however, they have limitations in applications due to linearity issue.

Neural Network (NN) is one of the methods that are widely used to solve most real-world problems. As NN has the ability to recognize time series patterns and nonlinear characteristics, which gives better accuracy over other methods, it has become the most popular method in forecasting [16] - [18]. A case study predicting the Caspian Sea level compares the performance of NN and ARIMA. The results proved that NN is a more powerful tool in complementing or even substituting statistical models [19].

Nowadays, using hybrid techniques or combining several techniques has become a common practice to improve the forecasting accuracy in which combination of forecasts from more than one technique often leads to improved forecasting performance [20]. Many papers have reported that hybridization of two or more techniques offers a number of advantages in many domain problems (see for examples: [21] - [31]). A study showed that using hybridization of NN-GA increased the rainfall runoff forecasting accuracy more than any other standalone methods [32]. Besides, the study by [23] combined Neural Network and Partial Least Squares (PLS) and the finding showed the proposed method gave better result compared to PLS alone.

It is well known that Backpropagation Neural Network (BPNN), Nonlinear Autoregressive Exogenous Model (NARX) and Genetic Algorithm (GA) are standalone technique with each technique has its own advantages and disadvantages. BPNN is commonly used in forecasting studies and suitable tool for modelling the behaviour of a system since it has the following three important characteristics: generalization ability, noise tolerance and fast response once trained [33] - [34]. However, BPNN and NARX have difficulties to determine the optimal network of architecture and regression coefficients (weights) due to randomness of its initial weights [20]. This implies that the best regression coefficients may be different in each learning process and there are many possibilities of nonlinear regression models which will be used for forecasting. While GA is an effective technique for obtaining optimum values of an optimization problem and is one of the potential methods for optimization of parameters in BPNN [25] - [26]. However, GA encounters difficulties in finding a fitness function that effectively work in forecasting or classification [35].

Under the circumstance, we present a combination of Multiple Backpropagation Neural Network (Multiple BPNN) and Genetic Algorithm (GA) to overcome the limitation performance of ARIMA/SARIMA, standalone BPNN and NARX. The basic idea of the proposed technique is done by the following steps. First, we construct Multiple BPPN, say L BPNNs with the same structure (with L is a positive integer), and collect n sets of candidate regression coefficients from the Multiple BPNN. The next step is finding the best regression coefficients by GA with the initial population of GA is the candidates founded from Multiple BPNN. When L is equal to 1, it is call Single-BPNN-GA (S-BPNN-GA), otherwise we call it the Multiple-BPNN-GA (M-BPNN-GA).

The rest of this manuscript is organized as follows: Section II provides the general research framework. In Section III, we briefly introduce BPNN, NARX and GA, followed by the hybrid techniques of Multiple BPNN and GA in Section IV. Section V presents data pre-processing including missing data treatment and data standardization. Finally, results and discussion are given in Section VI and followed by conclusions in Section VII.

II. Research Framework

Generally, this research is divided into four main stages as depicted in Figure 1. The first stage involves missing data treatment and data standardization and data splitting. To simplify, two simple treatment missing data techniques based on Ordinary Linear Regression (OLR) and mean substitutions are employed. We conducted a data standardization to omit the units of the variables of interest. In data splitting, we divided our data into three subsets of data namely training, testing and evaluation data. The training and testing data are used in the learning process for determining the best weights, while the evaluation data are used to evaluate the best Multiple BPNN-GA in the future forecasting.

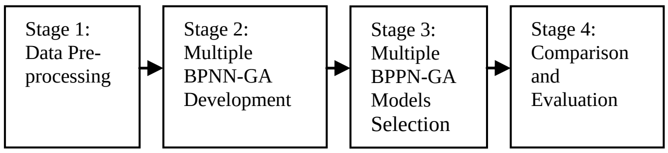


Fig. 1. General research framework.

In the second stage, we hybrid Multiple BPNN and GA in which L standalone BPNNs are used to provide sets of candidates' regression coefficients and then the candidates will be optimized by GA. In the third stage, we perform model selection of Multiple BPNN-GA, and the best model will be used in the next stage. In the last stage, we make comparisons between Multiple BPNN-GA with the other famous techniques such as ARIMA/SARIMA, standalone BPNN and NARX.

III. BPNN, NARX and Genetic Algorithm

A. BPNN

BPNN has a certain network architecture that contains input layer, hidden layer, output layer, number of nodes in each layer and the associated weights in inter-layer connection. In order to achieve a good performance, therefore, the network architecture must be determined and trained properly through a learning process [25] - [36].

In this paper, the maximum neuron input is five since we have five independent variables which are monthly index, rainfall, evaporation, temperature and relative humidity. For variable and model selection purposes, the number of input neurons and hidden nodes are changed to find the most stable structure and the most accurate prediction. The best structure and variables will be determined based on the measurement performances which will be discussed in Section IV.

B. NARX

NARX is a regression technique based on the linear autoregressive network with exogenous inputs (ARX) model, which is commonly used in time-series modelling. It uses tapped delay lines (d) to store previous values of the input, x(t) and output, y(t) sequences. The y(t) sequence is considered a feedback signal which is an input and also an output. Mathematically, NARX's model is given as follows:

$$y(t) = f(y(t-1), y(t-2), \cdots, y(t-d_1); \mathbf{x}(t-1), \mathbf{x}(t-2), \cdots, \mathbf{x}(t-d_{21})) \quad (1)$$

where f is a nonlinear function, x(t) is the input of NARX, y(t) is the output and also feedback of NARX and d is a tapped delay.

C. Genetic Algorithm

Genetic algorithms (GA) are a computerized search and optimization algorithm based on the mechanics of natural genetics and natural selection. The basic steps of genetic algorithm [10], [16], [25] can be described as follows: 1) Randomly generate an initial population, 2) Compute the fitness of each chromosome in the current population, 3) Create new chromosome by selection, crossover and applying mutation, 4) Substitute these new chromosomes for some bad chromosomes in the current population and 5) If the end condition is satisfactory, then stop; otherwise repeat step 2.

IV. The Proposed Technique

Even though BPNN can capture most nonlinear functions and gain wider applications in various fields, however, the adjustment of each regression coefficient parameter to optimize the whole network is not an easy task [37]. Technically, Multiple BPNN are employed in producing several sets of candidates of regression coefficients, whereas GA is adopted in searching optimal design based on the sets of candidates which produces best predicted fitness values. The framework of Multiple BPNN and GA is depicted in Figure 2.

In this paper, we used notation *k- j-1* to represent BPNN with k input nodes, j hidden layer nodes and 1 output node, respectively. In Multiple-BPNN-GA, we assume that there are L BPNNs with the same architecture *k-j-1* where L<=33 and the number of chromosomes is 100. It is noticed that there is a bias weight in each hidden node in our BPNN. This implies that the number of weights and biased (regression coefficients) in each BPNN are $(k+1)j$ and $(j+1)$, respectively, and the length of chromosome of GA is $(k+1)\, j+(j+1)$.

The process of finding the best coefficient regressions is conducted as follows. On the first stage, each of the L standalone BPNNs extracts the three best sets of weights and biases; and put them into the initial population P_o in GA. The second stage, GA adds *100-3L* chromosomes in P_o randomly since the initial population of GA is 100 chromosomes. The third stage, GA tries to obtain an 'optimum' solution of set of regression coefficients which repeats evaluations, selection, crossover and mutation after initialization until the stopping condition is satisfied. The final stage is the optimum regression coefficient founded by GA is used in standalone BPNN for forecasting.

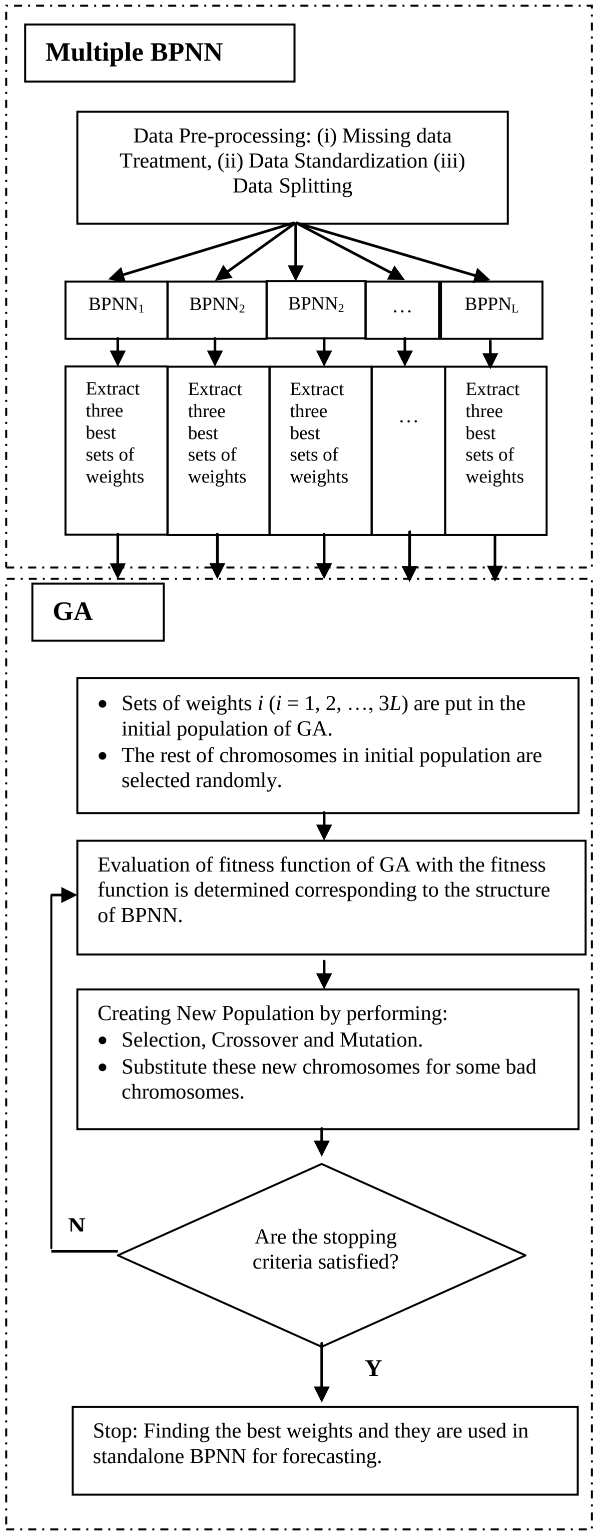


Fig. 2. Framework of Multiple BPNN-GA

V. Data Preprocessing

A. Missing Data Treatment

The missing data can be occurred due to the malfunctioned equipment, the weather was terrible, human technical problem or maybe the data were entered incorrectly. Missing data should be handled in data analysis since the missing data will influence the performance of the technique used and the quality of analysis. We may not utilize a certain technique directly when the missing data exist.

TABLE I
THE SNAPSHOT OF RAW MISSING DATA (NA: NOT AVAILABLE)

Month	t	Rf	Eva	Temp	Humid	WL
Jan	1	NA	3.8548	26.242	78.561	14.72
Feb	2	NA	3.9194	26.811	79.189	14.83
Mar	3	NA	4.8387	27.245	78.177	13.96
Apr	4	NA	5.2484	27.957	77.787	13.81
May	5	NA	4.9032	27.632	79.081	13.95
Jun	6	NA	3.8548	27.503	79.16	14.13
Jul	7	NA	3.4194	28.084	77.558	13.75
Aug	8	NA	4.0161	27.39	78.561	13.75
Sep	9	NA	3.7258	26.95	79.74	13.92
Oct	10	3.5161	4.0323	27.232	80.236	13.78
Nov	11	11.226	3.6129	26.36	83.777	14.14
Dec	12	20.548	3.1452	26.719	81.155	14.25

** **Note:** t, Rf, Eva, Temp, Humid and WL refers to index of month, rainfall, evaporation, temperature, humidity and water level, respectively.

Table 1 illustrates the snapshot of raw data from January 2006 until December 2006 which some rainfalls in January until September 2006 are missing. Deletion or elimination of the missing variable is the default method for most procedures in missing data [6]. However, in time series regression, this approach seems like not the best methods to be used since we will lose the important information of time series data. As mentioned before, we conduct two simple techniques for missing data treatment using mean and OLR substitutions which are two usual techniques in the missing data treatment [38].

Mean Substitution: This technique is very simple to be performed. First, we find the mean of a certain variable for a certain month with non-missing values. Afterward, the mean is substituted with the missing values on the associated month. Table 2 demonstrates the snapshot of replacement values of missing data using mean calculations.

OLR Substitution: In this technique, we will predict the value of missing data using regression model and non-missing values for each variable. The predictor variable in the OLR model is time (*t*) as single predictor variable. The OLR model produces the predicted value which will replace the missing data on associated variable. The regression model for rainfall, evaporation, temperature and humidity are given as follows:

Rainfall (*RF*) OLR model:

$$Rf(t) = 9.38935 + 0.0673(t) \tag{2}$$

Evaporation (*Eva*) OLR model:

$$Eva(t) = 4.09290 + 0.00228(t) \tag{3}$$

Temperature (*Temp*) OLR model:

TABLE 2
THE SNAPSHOT OF SUBSTITUTION MISSING VALUES USING MEAN APPROACH

Month	t	Rf	Eva	Temp	Humid	WL
Jan	1	9.4567	3.8548	26.242	78.561	14.72
Feb	2	9.5241	3.9194	26.811	79.189	14.83
Mar	3	9.5915	4.8387	27.245	78.177	13.96
Apr	4	9.6588	5.2484	27.957	77.787	13.81
May	5	9.7262	4.9032	27.632	79.081	13.95
Jun	6	9.7936	3.8548	27.503	79.16	14.13
Jul	7	9.8609	3.4194	28.084	77.558	13.75
Aug	8	9.9283	4.0161	27.39	78.561	13.75
Sep	9	9.9957	3.7258	26.95	79.74	13.92
Oct	10	3.5161	4.0323	27.232	80.236	13.78
Nov	11	11.226	3.6129	26.36	83.777	14.14
Dec	12	20.548	3.1452	26.719	81.155	14.25

$$Temp(t) = 27.162 + 0.00047(t) \tag{4}$$

Relative Humidity (*Humid*) OLR model:

$$Humid(t) = 78.9441 + 0.0088(t) \tag{5}$$

B. Standardization

The treatment data were transformed into standardized data with range [0, 1] by using equation (6) as follows:

$$\text{standardized data} = \frac{\text{treatment data}}{\text{maximum data}} \tag{6}$$

The predicted values of standardization scale should be transformed back to the original scale using Eq. 6. It is important to make standardized the data because standardization of data is omitting units of the variables of interest.

C. Data Splitting

As mentioned in Section I, we used months, monthly rainfall, rate of evaporation, rate of temperature, relative humidity and water level which are collected from the DID and MMD for Dungun district of Terengganu with a total number of 75 observation data from the year 2006 until 2012. In our experiment, we split the data into three subsets namely training, testing and evaluation. The learning process contains 63 observations in which 70% and 30% of 63 observations for training and testing, respectively. The twelve observation data from April 2011 till March 2012 are used as an evaluation data.

VI. RESULTS AND DISCUSSION

A comparative study is carried out to investigate the performance of Multiple BPNN-GA with missing data treatment. The performance of the Multiple BPNN-GA will then be compared with the ARIMA/ SARIMA, BPNN and NARX in the water level forecasting at Dungun River. For discussion purposes, we used the notations of X_1-X_5 representing index of month (X_1), rainfall (X_2), evaporation (X_3), temperature (X_4), and humidity (X_5) respectively.

A. Performance Evaluation

We conducted 10 runs for each technique to evaluate the performance of BPNN, NARX, S-BPNN-GA and M-BPNN. The performance of those techniques is measured based on their mean squared error (MSE) of training and testing, the absolute value of difference mean of MSE's training and MSE's testing, running time and stability predicted water level. The absolute value of difference mean of MSE's training and MSE's testing is given by the following formulae:

$$\text{DMSE} = |\,(\text{MSE's training-MSE's Testing}) \times 100\%|.$$

The DMSE is used to detect overfitting. The overfitting occurs when MSE's training provides a small value, but MSE's testing gives a relatively large value compared to MSE's training.

TABLE 3
PERFORMANCE SOME COMBINATION INPUT NODES USING BPNN WITH MISSING DATA TREATMENT

	BPNN Structure (Variables)	MEAN MSE (STDEV) Training	Testing	MEAN DMSE (%)
Mean Subst.	BPNN 2-6-1 (X_1X_2)	0.0009 (2.26E-04)	0.0011 (2.67E-04)	0.02
	BPNN 2-6-1 (X_1X_3)	0.0011 (3.09E-04)	0.0016 (4.16E-04)	0.05
	BPNN 2-6-1 (X_1X_4)	0.0019 (3.30E-04)	0.0013 (4.22E-04)	0.06
	BPNN 2-6-1 (X_1X_5)	0.0009 (3.43E-04)	0.0017 (3.16E-04)	0.08
	BPNN 2-4-1 (X_1X_2)	0.0009 (2.40E-04)	0.0011 (2.23E-04)	0.02
	BPNN 2-4-1 (X_1X_3)	0.0007 (2.98E-04)	0.0011 (3.13E-04)	0.04
	BPNN 2-4-1 (X_1X_4)	0.0006 (3.02E-04)	0.0013 (3.33E-04)	0.07
	BPNN 2-4-1 (X_1X_5)	0.0009 (3.53E-04)	0.0034 (3.46E-04)	0.025
OLR Subst.	BPNN 2-8-1 (X_1X_2)	0.001 (3.30E-04)	0.0015 (3.06E-04)	0.05
	BPNN 2-8-1 (X_1X_3)	0.002 (2.94E-04)	0.0017 (3.40E-04)	0.03
	BPNN 2-8-1 (X_1X_4)	0.0016 (4.16E-04)	0.0012 (2.83E-04)	0.04
	BPNN 2-8-1 ($X1X_5$)	0.0008 (3.43E-04)	0.0019 (3.53E-04)	0.11
	BPNN 5-6-1 ($X_1X_2X_3X_4X_5$)	0.0009 (1.63E-04)	0.0011 (2.05E-04)	0.02

The stability of the above techniques is measured using the standard deviation of 10 runs. A technique is said to be more stable if it has smaller value of standard deviation compared to the others. In terms of running time, however, it is not surprising to guess the running time of Multiple BPPN-GA will slower compared to standalone BPNN due to the effect of the multiple learning process of BPNNs and optimization process of GA. Afterwards, we select the best model using the five performance criteria and use the best obtained model to predict the evaluation data from April 2011 until March 2012.

B. Experiment

BPNN

Our first experiment is to evaluate the performance of standalone BPNN and to find the best network architecture as a basis of Multiple BPNN-GA. Standalone BPNN is used to build a non-linear model for water level at Dungun River with the logarithmic sigmoid (*logsig*) as BPNN's activation function. The sigmoid function is often used in hidden layers due to its ability of authoritative non-linear approach [24]. We used *trainlm* function as our training algorithm where the modified bias and weight values based on Lavenberg-Marquardt optimization. It is noticed that there are some combination variables X_1, X_2, X_3, X_4, and X_5 in the input layer. Therefore, the number of nodes in the input layer is either 1, 2, 3, 4 or 5. In our experiment, we set the number of nodes in the hidden layer is 4, 6, 8 and 10 for comparison purpose.

TABLE 4
PERFORMANCE NARX 4-6-1 AND NARX 3-10-1 WITH TWO MISSING DATA TREATMENTS AND SEVERAL TAPPED DELAY

	d	MEAN MSE (STDEV) Train	Test	MEAN DMSE (%)
Mean Subst. (NARX 4-6-1 with variables: X_1 X_2 X_3 X_5)	2	0.0012 (2.11E-04)	0.0014 (2.62E-04)	0.02
	3	0.0008 (1.49E-04)	0.0009 (1.89E-04)	0.01
	4	0.0010 (1.63E-04)	0.0013 (2.31E-04)	0.03
OLR Subst. (NARX 3-10-1 with variables: X_1 X_2 X_3)	2	0.0011 (2.21E-04)	0.0014 (2.62E-04)	0.03
	3	0.0012 (2.62E-04)	0.0017 (1.83E-04)	0.05
	4	0.0009 (1.56E-04)	0.0007 (1.63E-04)	0.02

The performance's result of standalone BPNN with missing data treatments for 2 and 5 input nodes is given in Table 3. Table 3 summarise the best performance of standalone BPNN and shows that both BPNN 2-6-1 and BPNN 2-4-1 with mean substitution and input nodes of X_1 and X_2 gave the best result in terms of MSE's training, MSE's testing, standard deviation and percentage error. While the standalone BPNN 5-6-1 with five input predictors also gave the best result when we conducted the treatment missing data using OLR substitution.

NARX

The performance of NARX 4-6-1 and NARX 3-10-1 with *d* is equal 2, 3 and 4 is shown in Table 4. It is noticed that NARX 4-6-1 and NARX 3-10-1 are the best network architectures among the other architectures of NARX. Referring to Table 4, we can obtain that NARX 4-6-1 (with *d*=3 and mean substitution) and NARX 3-10-1 (with *d*=4 and OLR substitution) provided better results compared to others.

Multiple BPNN-GA

In Multiple BPNN-GA, we set L=1 and L=10, and choose the best founded standalone BPNN structures from the previous experiment, namely BPNN 2-6-1, BPNN 2-4-1 and BPNN 5-6-1. Since each standalone BPNN extracts the three best sets of weights and biases, therefore, they produced 30 sets of acceptable weights or regression coefficients. Afterward, the 30 sets were inserted into the initial population of GA. It is noticed that we used standard GA in the Multiple BPNN-GA and the maximum iteration of GA was 1000.

The performance of Multiple BPNN-GA is presented in Table 5. From the results, it shows that M-BPGA 5-6-1 with OLR substitution provides the smallest MSE's training, MSE's testing, DMSE and standard deviations. This result also give information that the best model for forecasting in Dungun River involves the predictor variables of months, rainfall, evaporation, temperature and relative humidity.

TABLE 5
PERFORMANCE OF MULTIPLE BPNN-GA WITH MISSING DATA TREATMENT

	Technique	MEAN MSE (STDEV)		MEAN DMSE
		Train	Test	(%)
Mean Subst. (Variable: X_1 X_2)	S-BPNN-GA 2-6-1	0.00018 (2.94E-05)	0.00012 (2.87E-05)	0.006
	S-BPNN-GA 2-4-1	0.00028 (2.64E-05)	0.00019 (2.67E-05)	0.009
	M-BPNN-GA 2-6-1	0.00015 (1.56E-05)	0.00032 (1.94E-05)	0.017
	M-BPNN-GA 2-4-1	0.00025 (1.15E-05)	0.00012 (1.76E-05)	0.013
OLR Subst. (Variable: $X_1X_2X_3X_4X_5$)	S-BPNN-GA 5-6-1	0.00016 (2.67E-05)	0.00019 (2.21E-05)	0.003
	M-BPNN-GA 5-6-1	0.00013 (2.36E-06)	0.00012 (6.67E-06)	0.001

C. Discussion

In this section, the performances of ARIMA/SARIMA, BPNN, NARX, S-BPNN-GA and Multiple BPNN-GA for water level forecasting were compared. We used the performance evaluation criteria as stated before to select the best model for water level forecasting of Dungun River. The explanations for each performance are as follows:

MSE Training and MSE Testing

Table 6 provides the comparison of average MSE Training and MSE testing of the five techniques. The comparisons of MSE training and MSE testing are also depicted in Figure 3 a) and Figure 3 b), respectively. From Table 6 and the two figures, the evidence shows that Multiple BPNN-GA with mean substitution gives smallest MSEs and significantly improves the MSE of NARX by about 84% and 87% in training and testing, respectively.

DMSE

The information about the mean of DMSE of the five techniques is presented in Table 6. From this table, it can be seen that DMSE of all techniques is relatively small and there is no large difference between MSE training and MSE testing. The results explain that overfitting had not happened in all techniques.

TABLE 6
COMPARISONS OF SARIMA, BPNN, NARX, S-BPNN-GA AND MULTIPLE BPNN-GA.

Technique (Variables)	MSE		MEAN DMSE
	Training	Testing	(%)
SARIMA (0,1,0)(0,1,1)$_{10}$ (t and WL)	0.0024	0.00186	0.05
BPNN 5-6-1 with OLR Subst. ($X_1X_2X_3X_4X_5$)	0.0009	0.0011	0.02
NARX 4-6-1 with Mean Subst. (X_1 X_2 X_3 X_5)	0.0008	0.0009	0.01
S-BPNN-GA 5-6-1 with Mean Substitution ($X_1X_2X_3X_4X_5$)	0.00016	0.00019	0.003
M-BPNN-GA 5-6-1 with Mean Substitution ($X_1X_2X_3X_4X_5$)	0.00013	0.00012	0.001

Running Time

The running time of Multiple BPPN-GA is slower compared to standalone BPNN due to the effect of multiple learning processes of several BPNNs. If we set L=10 in Multiple BPNN-GA, therefore, it needs about 30 times learning process of standalone BPNN (since each BPNN performs three repetitions) and processing time of GA to optimize the best regression coefficients. However, Multiple BPPN-GA improves the quality of the predicted water level of standalone BPNN in reasonable time as shown in Table 6 since our data set is not large.

Stability

Figure 4 a) and Figure 4 b) depicts the standard deviation of training and testing of the best BPNN, NARX, S-BPNN-GA and M-BPPN-GA, respectively. The evidence shows that M-BPPN-GA gives better stability in prediction of water level compared to the other techniques. Referring to Table 5 and Table 6, it is found that Multiple BPNN-GA with mean substitution gives the smallest standard deviation for both training and testing. It also reduces the standard deviation of NARX by about 98.4% and 96.5% in training and testing, respectively.

Comprehensive Comparison

Referring to Table 3 to Table 6 and the above performance evaluation criteria, we have the following important conclusions as follows:

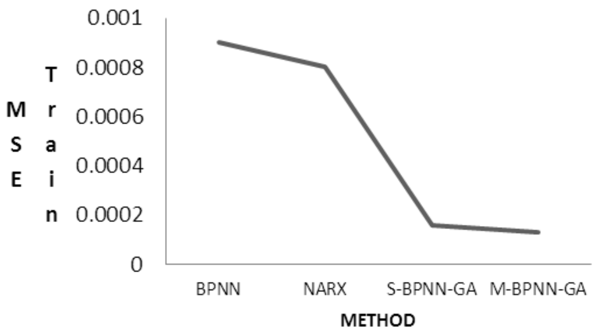
(i) BPNN is better than ARIMA/SARIMA,
(ii) NARX is superior compared to BPNN,
(iii) S-BPNN-GA gives better result compared to NARX,
(iv) Multiple BPNN-GA with mean substitution outperforms the technique of BPNN, NARX and S-BPNN-GA.

Furthermore, from our analysis, it shows that Multiple BPNN-GA is better than the other techniques by showing Multiple BPNN-GA's prediction for the rest of twelve months (evaluation data) is closest to the actual water level. The comparison performance between NARX 4-6-1 and M-BPNN-GA 5-6-1 using our evaluation data from April 2011 to March 2012 is presented in Figure 5. Using these evaluation data, we also calculated the MSE of NARX 4-6-1, S-BPNN-GA 5-6-1 and M-BPNN-GA 5-6-1 are 0.000094, 0.000085 and 0.000024, respectively. It means that the predicted values with M-BPGA 5-6-1 are closest to the actual value of water level in Dungun River.

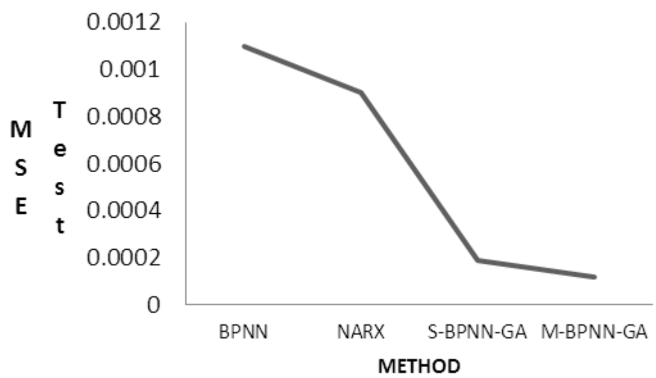
VII. Conclusions

We presented a hybrid Multiple BPNN and Genetic Algorithm (GA) to overcome the limitation of ARIMA/SARIMA, standalone BPNN and NARX. Our proposed techniques have been applied to forecast the water level at the Dungun River as our case study. The mean and OLR substitution were used to overcome the presence of the missing data in our collected data. Our experiments showed that M-BPNN-GA with mean substitution outperformed ARIMA/SARIMA, BPNN and NARX, and M-BPNN-GA improved significantly the performance of those techniques. It was noticed that the performance standalone NARX is better than standalone BPNN.

For future work, we are planning to hybrid NARX and GA, and compare its performance with M-BPNN-GA and the other existing nonlinear regressions such as kernel principal component regression and support vector machine based models.



(a)

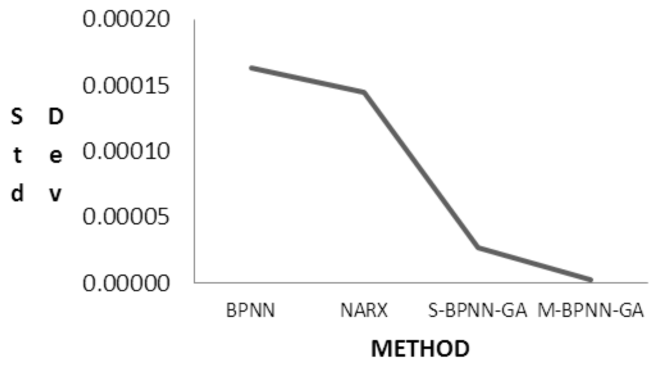


(b)

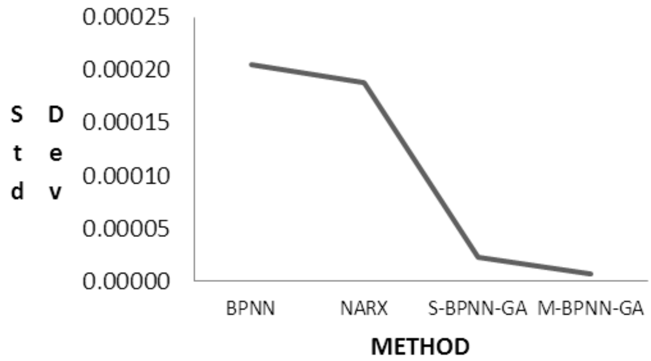
Fig. 3. Comparison of BPNN 5-6-1 NARX 4-6-1, S-BPNN-GA 5-6-1 and M-BPNN-GA 5-6-1. a) MSE's testing, and b) MSE's training.

Acknowledgment

The authors would like to express a sincere gratitude to the anonymous reviewers for their valuable comments and suggestions to improve the quality of this manuscript. In addition, the authors would also like to thank Bina Nusantara University, Universiti Teknologi Malaysia and Universiti Utara Malaysia for supporting this research project.



(a)



(b)

Fig. 4. Comparison of BPNN 5-6-1 NARX 4-6-1, S-BPNN-GA 5-6-1 and M-BPNN-GA 5-6-1. a) Standard Deviation's testing and b) Standard Deviation's training.

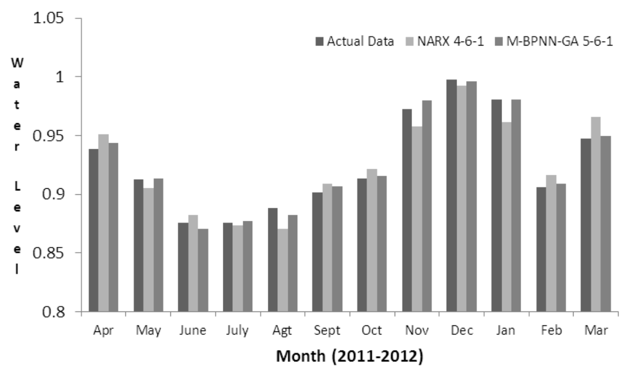


Fig. 5. Comparison performance of NARX 4-6-1and M-BPNN-GA 5-6-1 for evaluation data.

References

[1] M.T. Ekhwah, H. Juahir, M. Mokhtar, M.B. Gazim, S.M.S. Abdullah, and O. Jaafar, "Predicting for Discharge characteristics in Langat River, Malaysia using Neural Network Application Model," Research *Journal of Earth Sciences*, vol. 191, pp. 15-21, 2009.

[2] M.B. Gasim, J.H. Adam, M.E. Toriman, S.A. Rahim and H. Juahir, "Coastal Flood Phenomenon in Terengganu, Malaysia: Special Reference to Dungun,",*Research Journal of Environmental Sciences*, vol. 1, pp. 102-109, 2007.

[3] Department of Irrigation and Drainage (DID), *Laporan Banjir 2000/2001*, Unit Hidrologi Jabatan Pengairan dan Saliran Negeri Terengganu, 2002.

[4] Department of Irrigation and Drainage (DID), Flood *Forecasting and Warning System Report*, Unit Hidrologi Jabatan Pengairan dan Saliran Malaysia, 2009.

[5] G.P. Zhang, "Time Series Forecasting Using Hybrid ARIMA and ANN Model," Neurocomputing, vol. 50, pp. 159-175, 2002.

[6] N. Suguna, and K.G. Thanuskodi, "Predicting Missing Attribute Values using K-means Clustering," *J. Comp. Sci*., vol. 7, pp. 216-224, 2011, DOI: 10.3844/jcssp.2011.216.224

[7] P. Areekul, T. Senjyu, H. Toyama, and A. Yona, *A Hybrid ARIMA and Neural Network Model for Short-Term Price Forecasting in Deregulated Market. Japan*, Department of Electric & Electron, Engineer University of the Ryukyus, Nishihara, 2010

[8] N. I. Sapankevych and R. Sankar," Time series prediction: Using Support Vector Machine," *IEEE Computational Intelligence Magazine*, pp. 24-38, 2009.

[9] A. Wibowo, *Nonlinear predictions in regression models based on kernel method*, PhD Dissertation, Graduate School of Systems and Information Engineering, Univ. of Tsukuba, Japan, 2009.

[10] A. Wibowo and M.I. Desa, "Kernel Based Regression and Genetic algorithms for Estimating Cutting Conditions of Surface Roughness in End Milling Machining Process*," Expert System with Applications*, Elsevier, 2012.

[11] A. Wibowo and Y. Yamamoto, "A Note on Kernel Principal Component Regression," *Computational Mathematics and Modeling*, vol 23, Springer, 2012.

[12] N. Ibrahim and A. Wibowo, "Support Vector Regression Based Variables Selection for Water Level Predictions of Galas River in Kelantan Malaysia," *WSEAS Transaction on Mathematics*, 2014a.

[13] N. Ibrahim and A. Wibowo, "Time Series Support Vector Regression With Missing Data Treatment Based Variables Selection For Water Level Prediction Of Galas River In Kelantan Malaysia", *International Journal of Applied Research in Engineering and Science*, 2014b.

[14] A. Wibowo, "A Note of Hybrid GR-SVM for Prediction of Surface Roughness in Abrasive Water Jet Machining,", Meccanica, Springer, 2017.

[15] S. P. Meenakshi, and S. V. Raghavan, "Forecasting and Event Detection in Internet Resource Dynamics Using Time Series Models," *Engineering Letters*, vol. 23 issue 4, pp.245-257, 2015.

[16] S. Rajasekaran and G.A. Vijayalakshmi, *Neural Networks, Fuzzy Logic, and Genetic Algorithms: Synthesis and Applications*, Prentice Hall of India Private Limited, New Delhi, 2007.

[17] R. Sharda and K. Patil, K. (1994), "Neural Networks for the MS/OR Analysis," *International Journal of Economics*, vol 24, pp. 116-130, 1994.

[18] S.H. Arbain and A. Wibowo, "Neural Networks Based Nonlinear Time Series Regression for Water Level Forecasting of Dungun River," *American Journal of Computer Science*, Science Publications, 2012.

[19] M. Vaziri, 1997, "Predicting Caspian Sea Surface Water Level by ANN and ARIMA models," *Journal of Waterway*, Port, Coastal, and Ocean Engineering, vol. 123 No. 4, 1997.

[20] H. Ganji and L. Wang, "Hybrid Neural Network Models for Hydrologic Time Series Forecasting Based on Genetic Algorithm," *Fourth International Joint Conference on Computational Sciences and Optimizatio*n, 2011.

[21] B.B. Nair, S.G. Sai, A.N. Naveen, A. Lakshmi, G.S. Venkatesh and V.P. Mohandas, *A GA-Artificial Neural Network Hybrid System for Financial Time Series Forecasting*, Springer-Verlag Berlin Heidelberg, pp. 499-506, 2011.

[22] L. Wang, "A hybrid Genetic Algorithm-Neural Network Strategy for Simulation Optimization," *Appl. Math. Comput*. vol. 170, pp. 1329-1343, 2005.

[23] L. Shu, G. Dong, L. Liu, Y. Tao and M. Wang, "Water Level Variation and Prediction of the Pingshan Sinkhole in Guizhou, Southwestern China," *Sinkholes and the Engineering and Environmental Impacts of Karst*, American Society of Civil Engineers, pp. 423-432, 2008, doi: 10.1061/41003(327)40

[24] Y. Zhang and L. Wu, "Stock market prediction of S&P 500 via combination of improved BCO approach and BP Neural Network," *Expert Systems with Application*, vol. 36, pp. 8849-8854, 2009.

[25] L.A. Wulandhari, A. Wibowo and Desa M.I., "Condition Diagnosis of Multiple Bearings Using Adaptive Probabilistic Based Genetic Algorithms and Back Propagation Neural Networks," *Neural Computing and Applications*, Springer, 2014a

[26] L.A. Wulandhari, A. Wibowo and Desa M.I., "Condition Diagnosis of Bearing System Using Multiple Classifiers of ANNs and Adaptive Probabilities in Genetic Algorithms," *WSEAS Transaction on Systems and Controls*, 2014b.

[27] Y. Ghanou, and G. Bencheikh, "Architecture Optimization and Training for the Multilayer Perceptron using Ant System," *IAENG International Journal of Computer Science*, vol. 43 issue 1, pp 20-26, 2016.

[28] Z. Zhong and D. Pi, "Forecasting Satellite Attitude Volatility Using Support Vector Regression with Particle Swarm Optimization," *IAENG International Journal of Computer Science*, vol. 41 issue 3, pp. 153-162, 2014

[29] X. Zeng, L. Shu and J. Jiang, "Fuzzy Time Series Forecasting based on Grey Model and Markov Chain," *IAENG International Journal of Applied Mathematics*, vol. 46 issue 4, pp. 464-472, 2016.

[30] A. Wibowo, "Hybrid Kernel Principal Component Regression And Penalty Strategy Of Multiple Adaptive Genetic Algorithms For Estimating Optimum Parameters In Abrasive Waterjet Machining", *Applied Soft computing*, vol. 62, pp. 1102-1112, 2018.

[31] L. W. Loon, A. Wibowo, M.I. Desa and H. Haron, "A Biogeography-based Optimization Algorithm Hybridized with Tabu Search for Quadratic Assignment Problems", *Computational Intelligence and Neuroscience*, vol. 2016, pp. 1-12, 2016.

[32] G. Huang and L. Wang, "Hybrid Neural Network Models for Hydrologic Time Series Forecasting Based on Genetic Algorithm," *Fourth International Joint Conference on Computational Sciences and Optimization*, 2011.

[33] G. Puscasu, V. Palade, A. Stancu, S. Buduleanu and G. Nastase, *Sisteme de Conducere Clasicesi Intelegente a Proceselor*, MATRIX ROM, Bucharest, Romania, 2000.

[34] C.D. Bocaniala and V. Palade V., "Computational Intelligence Methodology in Fault Diagnosis: Review and State of the Arts. Computational Intelligence in Fault Diagnosis," *Advanced Information and Knowledge Processing*, 2006, pp. 1-36.

[35] Z. Yangping, Z. Bingquan, X. W. Dong. "Application of Genetic Algorithms to Faults Diagnosis in Nuclear Power Plants," *Reliability Engineering and System Safety*, vol. 67, pp. 153-160, 2000.

[36] M.T. Hagan, H.B. Demuth and M.H. Beale, *Neural Network Design*, PWS Publishing Company, 1996.

[37] T.L. Lee, "Neural network prediction of a storm surge," *Journal of Ocean Engineering*, vol. 33, pp. 483-494, 2006.

[38] D.C. Howell, *The Analysis of Missing Data*, In Outhwaite, W. & Turner, S. Handbook of Social Science Methodology, London, 2008.

Antoni Wibowo (M'12) is a Member (M) of IAENG since 2012. He has received my first degree of Applied Mathematics in 1995 and master degree of Computer Science in 2000. In 2003, He awarded a Japanese Government Scholarship (Monbukagakusho) to attend Master and PhD programs at Systems and Information Engineering in University of Tsukuba-Japan. He completed the second master degree in 2006 and PhD degree in 2009, respectively. His PhD research focused on machine learning, operations research, multivariate statistical analysis and mathematical programming, especially in developing nonlinear robust regressions using statistical learning theory. He has worked from 1997 to 2010 as a researcher in the Agency for the Assessment and Application of Technology – Indonesia. From April 2010 – September 2014, he worked as a senior lecturer in the Department of Computer Science - Faculty of Computing, and a researcher in the Operation Business Intelligence (OBI) Research Group, Universiti Teknologi Malaysia (UTM) – Malaysia. From October 2014 – October 2016, he was an Associate Professor at Department of Decision Sciences, School of Quantitative Sciences in Universiti Utara Malaysia (UUM). Dr. Eng. Wibowo is currently working at Binus Graduate Program (Master in Computer Science) in Bina Nusantara University-Indonesia as a Specialist Lecturer and continues his research activities in machine learning, optimization, operations research, multivariate data analysis, data mining, computational intelligence and artificial intelligence.

Siti Hajar Arbain was born in Tapah, Perak Malaysia on 27th July 1990. She has received her first degree of Industrial Mathematics in 2011 and master degree of Computer Science in 2014. In September 2016, she awarded a Malaysian Government Scholarship (MOHE) attached with University of Tun Hussein Onn Malaysia (UTHM) and is currently attending PhD programs at Software Engineering in University of Technology Malaysia (UTM). Her PhD research focuses on application of soft computing in software engineering. She has worked from 2014 to 2015 as a former lecturer at Department of Mathematics- Faculty of Computing and Mathematics, University of Technology MARA, Malaysia.

Norhaslinda Zainal Abidin is a senior lecturer in Decision Science at Universiti Utara Malaysia, Malaysia. She holds an MSc in Decision Science from Universiti Utara Malaysia and obtained her Ph.D. in Operations Research from University of Salford, United Kingdom. She has managed to secure several national and university research grants. She has worked in several projects and her latest project including determining a competitive optimal export duty structure in Malaysian palm oil Industry using system dynamics and genetic algorithm approaches. Her areas of research interests include system dynamics, simulation, optimization, and MCDM. She practices her quantitative discipline in various areas including healthcare, transportation, agriculture, as well as supply chain incorporating with researchers from different field of studies.