Short-Term Urban Water Demand Prediction Considering Weather Factors

Salah L. Zubaidi[a,b], Sadik K. Gharghan [c], Jayne Dooley [a], Rafid M. Alkhaddar[a], Mawada Abdellatif [a], Jareer Mohammed [b]

[a] Department of Civil Engineering, Liverpool John Moores University, Liverpool, UK

[b] Department of Civil Engineering, University of Wasit, Wasit, Iraq, salahlafta1976@yahoo.com

[c] Department of Medical Instrumentation Techniques Engineering, Electrical Engineering Technical College Middle Technical University (MTU), Al Doura, Baghdad 10022, Iraq.

Abstract

Accurate and reliable forecasting plays a key role in the planning and designing of municipal water supply infrastructures. Recent studies related to water demand prediction have shown that water demand is driven by weather variables, but the results do not clearly show to what extent. The principal aim of this research was to better understand the effects of weather variables on water demand. Additionally, it aimed to offer an appropriate and reliable technique to predict municipal water demand by using the Gravitational Search Algorithm (GSA) and Backtracking Search Algorithm (BSA) with Artificial Neural Network (ANN). Moreover, eight weather factors were adopted to evaluate their impact on the water demand. The principal findings of this research are that the hybrid GSA-ANN (Agent=40) model is superior in terms of fitness function (based on RMSE) for yearly and seasonal phases. In addition, it is evidently clear from the findings that the GSA-ANN model has the ability to simulate both seasonal and yearly patterns for daily data water consumption.

Keywords: Australia; explanatory variables; municipal water demand and neural network model.

Acknowledgements

The Iraqi Ministry of Higher Education and Scientific Research, Wasit University supported this project. I thank Peter Roberts, the Demand Forecasting Manager, Yarra Valley Water for providing all data.

1 Introduction

The environmental outlook of the Organisation for Economic Co-operation and Development (OECD) to 2050 indicates that global demand for water is anticipated to increase by 55%, depending on 2000 as a baseline. Moreover, more than 40% of the universal population may be under acute water stress (Fogden and Wood, 2009). Adamowski et al. (2012) stated that successive dry days with high temperatures and a low number of rainy days can play a crucial role in increased water demand. Accordingly, the urban water supply infrastructure faces increasing pressures related to the impact of extreme weather factors. Under these pressures, the present urban water supply infrastructure is probably insufficient to meet future water demands.

Prediction of water demand can play a significant role in optimising the design, operation and management of urban water supply infrastructures. Additionally, it can minimise the uncertainty that results from a rapid increase in water demand due to increasing the weather variables effect. Moreover, short-term forecasting is fundamentally associated with scheduling operations related to pumping and decreasing the time that water is detained in storage tanks, which can improve the water quality (Bougadis et al., 2005).

Several previous researchers have examined water consumption considering weather variables by using traditional models (Zhoua et al., 2000; Gato et al., 2005). Gato et al. (2005) developed a technique based on both a simple time series and simple linear regression analysis using total daily rainfall and daily maximum temperature. This study revealed that residential water consumption during winter months was affected by weather. Bakker et al. (2014) applied three various models: a Multiple Linear Regression, a Transfer/-noise model, and an Adaptive Heuristic with and without utilising weather input. The models' outcomes demonstrated that, when weather inputs are used, the average errors can be decreased by 7% and the largest predicting errors by 11%. Their models can be classified into traditional and new techniques.

Several previous studies have investigated and compared conventional and machine-learning models to predict water demand, and they found that machine-learning techniques have the ability to predict water demand better than the traditional model; these studies include:

Jain and Ormsbee (2002) presented an artificial neural network model (MLP) and compared it with two traditional models, regression and time series. The study employed the historical daily data for water consumption and weather variables in Kentucky state, USA, from 1982-1992. The results indicated that the simple ANN model offers a better ability for accurate prediction than the conventional techniques.

Bougadis et al. (2005) investigated three methods: time series analysis, linear and multiple linear regression, and artificial neural networks. The research adopted the historical weekly data of water consumption for summer months only for the city of Ottawa, Canada, from 1993-2002. In addition, it used the climate variables and antecedent water consumption as model inputs. The performance of the ANN models in predicting water demand consistently outperformed the traditional models.

Unlike many hydrological applications, it has been noted that the artificial neural network technique has only limited application in terms of water demand modelling (Firat et al., 2010). In addition, the majority of previous studies have adopted monthly time series data in their water demand models and sometimes used weekly data; few have adopted daily time series (Sarker et al., 2013).

Most studies of water prediction have only investigated the impact of socio-economic factors or a mix of socio-economic and weather factors (Liu et al., 2003; Firat et al., 2009; Behboudian et al., 2014). Few studies have adopted weather variables only in their water demand models as well as employing limit variables (maximum temperature and total rainfall only) (Jain et al., 2001; Jain and Ormsbee, 2002; Adamowski, 2008; Adamowski et al., 2012). Adamowski (2008) advised using extra weather variables in the water demand model to include evaporation, humidity, wind speed, and amount of cloud cover and sunshine.

Prediction of water demand is a substantial topic for policy-makers in the water industry. It is still extremely challenging to achieve the expected forecasting accuracy with respect to the prediction trends (Behboudian et al., 2014). Accordingly, much uncertainty still exists about the relationship between the capacity of water systems and a potential rapid increase in water demand resulting from acute weather factors based on seasonal and yearly base.

The aims of this research study are to:

a) Utilise two novel optimisation algorithms to enhance the capability of the ANN technique to predict water demand with high accuracy and minimum error.
b) Use statistical techniques to select the model inputs that increase forecasting accuracy compared with a trial and error approach.
c) Examine the extra weather variables employed in the model inputs to assess the weather factors' impact on water demand and reduce the uncertainty, and
d) Develop two kinds of short-term models – seasonal and yearly time series daily data – to explore the relationship between water demand and weather factors on both a seasonal / a yearly basis and explore the uncertainty.

To the best of the authors' knowledge, no previous applications for the techniques in both points (a) and (b) have been used in water prediction for short-term daily data time series analysis.

2 Studied Area and Model Data

For the purpose of this study, one catchment area in Australia, the city of Melbourne, has been used to develop the water demand model. Yarra Valley Water is one of three retail water companies which receive municipal water from the Melbourne Water Corporation. Yarra Valley Water delivers municipal water supply and sewerage services to more than 1.5 million capita who live in the catchment area of the Yarra River where it flows through Melbourne. Figure 1 shows the Licence Service Area of Yarra Valley Water (YVW, 2017).

Historical daily data of water consumption and weather variables were collected from Yarra Valley Water for the areas that were served in Melbourne city from 2010-2015. This data comprises water consumption (ML), Maximum Temperature (◦c), Mean Temperature (◦c), Minimum Temperature (˚c), Rainfall (mm), Evaporation (mm), Solar Radiation (MJ/m^2), Vapour Pressure (hpa), and Maximum Relative Humidity (%). Figure 2 depicts the historical daily water consumption data.

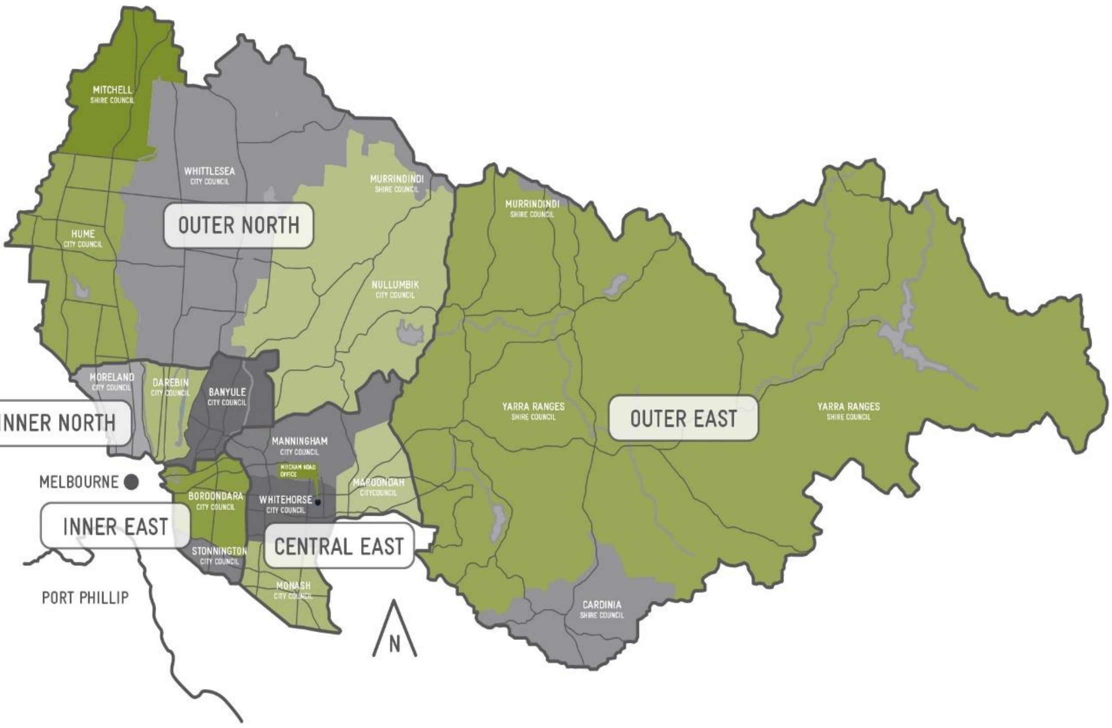

Figure 1: Yarra Valley Water's licence service area (YVW, 2017)

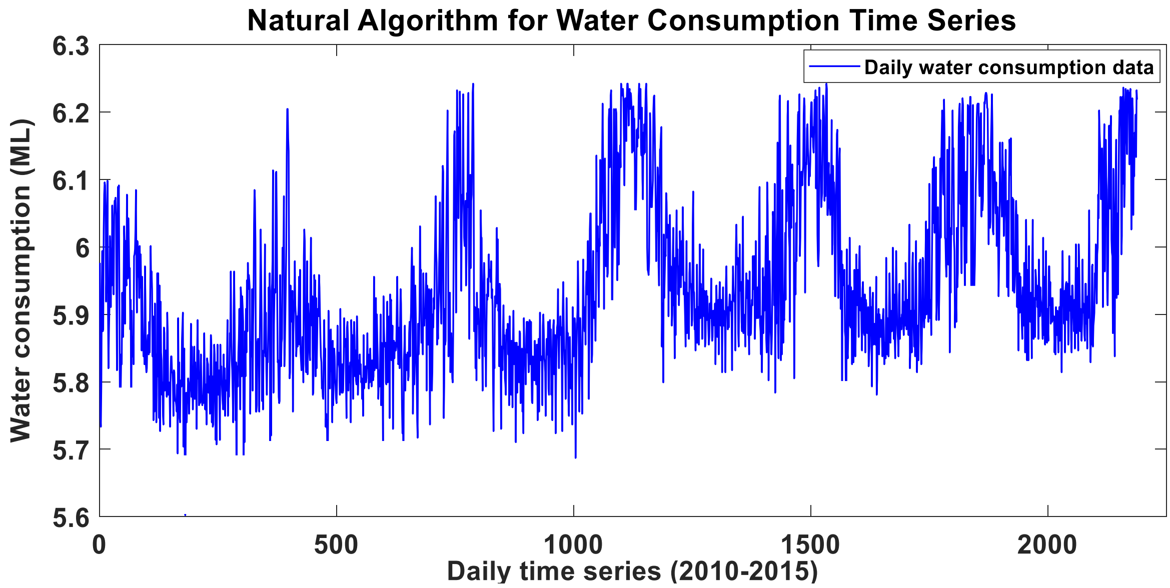

Figure 2: Daily data water consumption for Melbourne for the period 2010-2015

3 Data Pre-processing Techniques

Maier and Dandy (2000) stated that it is vital to pre-process data in an appropriate form before it is utilised in the ANN. These techniques are essential to confirm that all the data receives equal attention in the learning mode.

3.1 Data Cleaning

Data cleaning techniques comprise detection and removal of irrelevant or meaningless data, noise or outliers, to improve the outcomes of data analysis (Xiong et al., 2006). Extreme data has adverse effects on the regression solution and influences the accuracy of the model (Pallant, 2011). In this study, the box and whisker technique has been used to determine the outliers' data and then scores changed to be within the same pattern as the rest of the data.

3.2 Normalisation

This approach aims to smooth the answer space and minimise the effects of noise (ASCE, 2000; Kotsiantis et al., 2006). Tabachnick and Fidell (2013) stated that transforming the continuous variables is important in making the time series normally or near normally distributed. Additionally, the results of the model are degraded, if the time series of variables are not normally or near normally distributed. In this research study, natural algorithm, square root, and inverse function are adopted to transfer time series data depending on the type of series.

4 Selection of Explanatory Variables

The selection of explanatory variables influencing water demand as model input data is one of the most significant stages in evolving a satisfactory forecast model (Zhang et al., 2006) . Previous studies have trained many ANNs with various combinations of input variables to select the model that has the best performance (Jain et al., 2001; Zhang et al., 2006; Adamowski et al., 2012). A potential drawback in the above technique is that it is model-based. In other words, the best performance is not achieved by depending on model input variables only, but also on the structure and calibration of the model (Shahin et al., 2008).

After the potential weather factors were identified (screened and normalised), a three-staged process was adopted in this study to select the ANN model input data; this was to avoid this problematic issue of choice and to reduce the uncertainty in input variables. In the first stage, correlation matrix analysis will be employed to determine the strength of association between the dependent and independent variables, as well as among the independent variables. Accordingly, the explanatory independent variables that have a significant correlation at the 0.01 level (2-tailed) will be selected. In the second stage, correlation matrix analysis will be adopted to investigate the effect of lag (previous values) of the independent variables that were selected in the first stage on the dependent variable. In addition, an autocorrelation technique will be applied for water consumption time series. The final stage of the selection process, variance inflation factor (VIF), will be utilised to determine the presence of multicollinearity.

These stages of the process were carried out to ensure that as many of the potential variables as possible were properly included in the map of the input-output relationship, to avoid multicollinearity, which can lead to incorrect conclusions.

5 ANN Techniques

The ANN approach is a system of information processing that attempts to mimic the workings of the brain's neurons by utilising a network of artificial neurons which are regular in layers. In addition, it has the ability to adequately map the non-linear water demand trend (Babel and Shinde, 2011). In this study, the Backpropagation Neural Network (BP-NN) kind and the Levenberg-Marquardt (LM) learning algorithm were employed for training, testing and validation. The LM training algorithm was adopted because it offers minimum error in addition to its speed and efficiency, as proven in Payal et al. (2015). In the ANN, before achieving the phases of training, testing and validation, the parameters of the number of inputs, number of hidden layers, number of neurons in each hidden layer, learning rate and the number of outputs must be determined. In this research, to predict the short-term daily water demand, an ANN architecture with the following four layers was employed: (1) input layer; (2) first hidden layer (FHL); (3) second hidden layer (SHL); and (4) output layer (Ahmed et al., 2016; Gharghan et al., 2016b), as depicted in Figure 3. The input layer contains seven parameters consisting of weather variables and antecedent water consumption. In the first layer, the neurons work as a buffer to distribute the values of inputs to the first hidden layer. The values of inputs were weighted by connections w_{ij} and collected by each neuron of the FHL to pass the output of the FHL to the neurons of the SHL. The inputs of the SHL were weighted by connections w_{iz} and collected by each neuron of the SHL to compute the output y_k in the fourth layer. The *tansigmoidal* activation function was employed in the FHL and SHL to cover all ranges of the negative input values, whilst the output layer utilised the *linear* activation functions to cover the positive values of water demand. ANN parameters chosen were not secure and subject to the trial-and-error technique, which does not always offer the optimal answer. Hence, the learning rate and the number of neurons in the first and second hidden layers were determined depending on the optimisation algorithms (GSA and BSA) with population sizes 10, 20, 30, 40 and 50 (Gharghan et al., 2016a). The Gravitational Search Algorithm (GSA) and Backtracking Search Algorithm (BSA) are able to remedy such a problem by locating the best learning rate value and the optimum number of neurons for both hidden layers of the ANN model. Consequently, the ANN's performance can be improved. In this case, these algorithms could be combined with the ANN to form two different types of hybrid model, the GSA-ANN algorithm and the BSA-ANN algorithm, through which the ANN was capable of predicting water demand with minimum error.

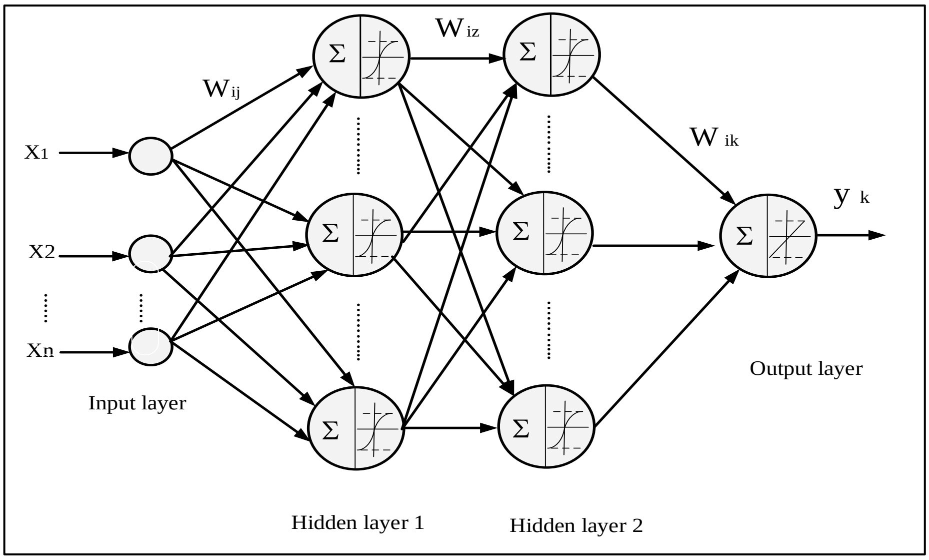

Figure 3: The ANN algorithm architecture

5.1 Heuristic Algorithms

Heuristic Algorithm is an approach that tries to catch a good solution (near optimal) at a plausible computational cost without the capability for undertaking either feasibility or optimality, or even in some situations to explain how close it is to the optimal solution (Rashedi et al., 2009). Because the conventional approaches provided a high water estimation error, ANN was employed in this research to improve the predicted water error. Due to the ANN technique's flexible modelling and learning abilities, it is likely to produce minimal errors in determining the future water demand. With a large amount of data and adequate ANN parameters, ANN has the ability to represent the relationship between dependent and independent variables. The heuristic algorithms, BSA and GSA, were hybridised with the ANN to select the optimum parameters of the ANN (i.e., the learning rate and number of neurons in both hidden layers). Choosing ANN parameters is not totally reliable and is dependent on trial and error, which in return gives a high level of error in water demand prediction.

Five population sizes, 10, 20, 30, 40 and 50, and 100 iterations were applied to let each algorithm determine the population that could attain the minimal fitness function value. In addition, the parameter settings of the heuristic algorithms were detected as F=3 for BSA, G_0

=1 and α=0.2 for GSA (Gharghan et al., 2016a). The process of ANN training was repeated many times utilising a large number of epochs (i.e., 1000 iterations) until the error between the observed and predicted municipal water reached the minimum.

5.1.1 Backtracking Search Optimisation Algorithm (BSA)

BSA is a recently developed evolutionary optimisation algorithm. It has the ability to solve problems of numerical optimisation in a quick and successful way by adopting a unique technique to generate a trail individually. This approach has been employed to address different kinds of optimisation problem in engineering fields, such as home energy management (Ahmed et al., 2017), Optimisation issues (Chen et al., 2017). The BSA technique is organised into five steps: initialisation, selection-I, mutation, crossover and selection-II. Additional details can be found in Su et al. (2016).

5.1.2 Gravitational Search Algorithm (GSA)

Rashedi et al. (2009) proposed the GSA algorithm, which is based on the Newtonian law of gravity: "Every particle in the universe attracts every other particle with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them". The mathematical principle of the GSA is dependent on the Newtonian law of gravity and the motion laws, as in the following:

$$F = G = \frac{M_1 M_2}{R^2} \quad (1)$$

Where

F= gravitational force,

R = the distance between the first and second particles mass (M_1) and (M_2) respectively, and

G = the gravitational constant value.

Newton's second law states that "acceleration is inversely proportional to mass M and directly proportional to force F", as follows:

$$a = \frac{F}{M} \quad (2)$$

Due to the influence of declining gravity, the real value of the "gravitational constant (G)" relies on the universe's real age. Eq. (3) offers a reduction of the gravitational constant with age (Gharghan et al., 2016a):

$$G(t) = G(t_0) \times (\frac{t_0}{t})^{\beta} \qquad \beta < 1 \quad (3)$$

Where

G (t) = the gravitational constant at time t, and

G (t_0) = the gravitational constant at the first cosmic quantum-interval of time t_0.

The agents' positions are initialised (i.e., the masses are chosen randomly within the offered search interval). The i^{th} agent position can be known by:

$$X_i = \left(X_i^1, \dots\dots\dots, X_i^d, \dots\dots.., X_i^k\right), \quad for\ i = 1,2,3, \dots.., N \quad (4)$$

Where

N = the number of agents,

X^d_i= the i^{th} agent position in the d^{th} dimension, and

k = the space dimension.

To compute the GSA fitness function, a root mean square error (RMSE) can be adopted to select the best and the worst fit for each iteration. The purpose of the computations was to reduce the problems and locate the masses of each agent as follows (Shuaib et al., 2015):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e^2} \quad (5)$$

$$best(t) = \min_{j\in\{1,,\ldots,N\}} fit_j\ (t) \quad (6)$$

$$worst(t) = \max_{j\in\{1,,\ldots,N\}} fit_j\ (t) \quad (7)$$

$$m_i(t) = \frac{fit_i(t) - Worst(t)}{best(t) - Worst(t)} \quad (8)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m_i(t)} \quad (9)$$

Where

e = the predicted water error, and

n = the number of samples.

The actual water consumption was obtained based on observation, whereas the predicted water was gained using the GSA-ANN algorithm. The gravitational constant G at iteration t was calculated as follows:

$$G(t) = G_0 e^{(-\alpha t/T)} \quad (10)$$

Computation of the total force in different directions for the i^{th} agent, calculation of the velocity and acceleration, and the position of the agents in the next iteration are as follows:

$$F_i^d(t) = G(t)\frac{M_{pi} \times M_{aj}}{R_{ij} + \varepsilon}(X_j^d(t) - X_i^d(t)) \quad (11)$$

$$F_i^d(t) = \sum_{j\epsilon Kbest, j\neq i} rand\ _j F_{ij}^d\ (t) \quad (12)$$

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} \quad (13)$$

$$v_i^d(t+1) = rand_i\ \times v_i^d(t) + a_i^d(t) \quad (14)$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \quad (15)$$

Figure 4 presents the flowchart that shows the details of the GSA-ANN operation based on the previous equations.

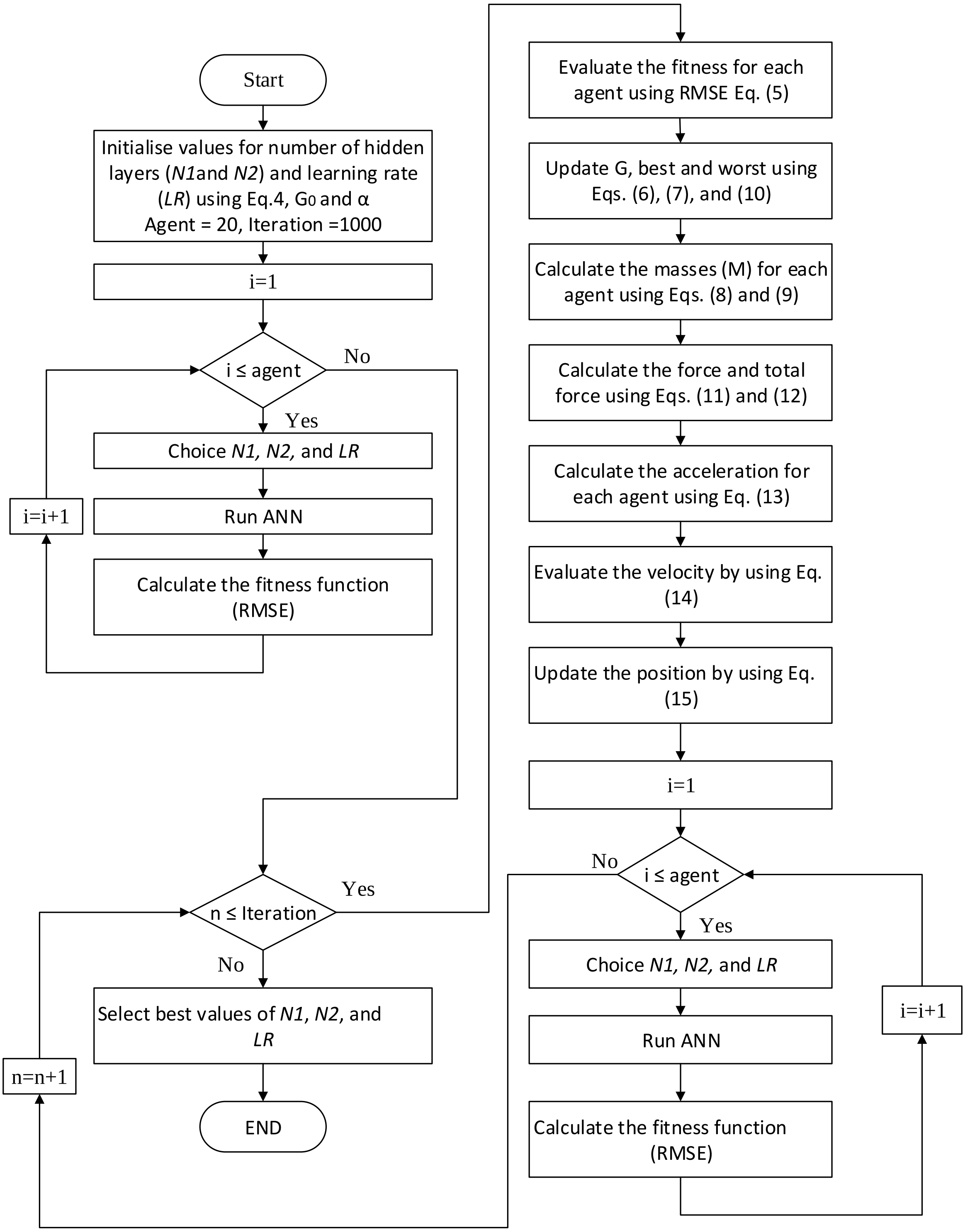

Figure 4: The GSA-ANN algorithm flowchart (Gharghan et al., 2016a)

5.2 Data Division

Data division is a vital process that needs to be addressed in the ANN. It is general practice to divide the obtainable data into three sub-sets, namely: learning, testing and validation. All these three sets must have the same pattern because the ANN does not have the capability to extrapolate outside the range of data that is employed for training (Basheer and Hajmeer, 2000). In this study, data was divided randomly between the training, testing and validation sets – 70%, 15%, 15% respectively (Babel and Shinde, 2011; Behboudian et al., 2014).

6 Performance Measurement Criteria

The statistical criteria parameters provide a means of measuring prediction accuracy, so prediction errors play a considerable role in the choice of suitable models and in providing insights in advising alterations to present models to minimise deviations in future predictions (Donkor et al., 2014). Several statistical parameters will be applied in the model's calibration such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Correlation Coefficient (R). These indicators are defined in Eqs. (16) through (20).

$$MAE = \frac{\sum_{m=1}^{N}|x_o - x_p|}{N} \quad (16)$$

$$MAPE = \frac{\sum_{m=1}^{N}\frac{|x_o - x_p|}{x_o}}{N} \quad (17)$$

$$MSE = \frac{\sum_{m=1}^{N}(x_o - x_p)^2}{N} \quad (18)$$

$$RMSE = \sqrt{\frac{\sum_{m=1}^{N}(x_o - x_p)^2}{N}} \quad (19)$$

$$R = \left[\frac{\sum_{m=1}^{N}(x_o - \overline{x_o})(x_p - \overline{x_p})}{\sqrt{\sum(x_o - \overline{x_o})^2 \sum(x_p - \overline{x_p})^2}}\right] \quad (20)$$

Where x_o= observed water consumption, x_p= predicted water demand, N= sample size, $\overline{x_p}$= mean of predicted demand, and $\overline{x_o}$= mean of observed consumption.

Bland–Altman analysis: this scatter plot test is employed to locate the area of agreement between (observed- predicted) versus ([observed+ predicted]/2), and the percentage of data that is distributed inside the limits of the agreement area

7 Results

7.1 Model Development

After data pre-processing, correlation coefficients and autocorrelations were detected between dependent and independents variables. Additionally, variance inflation factor (VIF) technique then was used to select the best model input. The values of VIF was located between (2.87-4.92), which were accepted as mention in Tabachnick and Fidell (2013). Accordingly, Eq. 21 can express the relation between dependent and the independents variables:

$$WD=f\,(WD_{t-1}, Tmax, Rad, Eva, WD_{t-2}, Tmax_{t-1}, Eva_{t-2}) \qquad (21)$$

Where: WD = water demand, Tmax = maximum temperature, Rad = solar radiation and Eva = evaporation. In addition, WD_{t-1} and WD_{t-2} represent the previous water consumption for the last first and second days and so on for the rest variables. Also, Eva and Eva_{t-2} were offered high correlation and less collinearity compare with Eva_{t-1}.

The correlation coefficients between input and output models for the best model input are 0.79, 0.66, 0.49, 0.64, 0.69, 0.57 and 0.52 for WD_{t-1}, Tmax, Rad, Eva, WD_{t-2}, $Tmax_{t-1}$ and Eva_{t-2} respectively. In addition, all these correlation is significant at the 0.01 level (2-tailed).

After selection of the best model inputs, data was divided into a seasonal (winter, spring, summer and autumn) and an annual scale to assess the effect of weather factors on water consumption in both seasonal and yearly data. Additionally, to examine the ability of each model to predict daily water demand.

To get an effective prediction, four statistical criteria were used to ensure that data set for training, testing, and validation have the same pattern. Table 1 provides a comparison of the three data sets. The results, as shown in Table 1, indicate that all the data sets have exactly the same pattern. In addition, the results of validation set will support that all sets have the same pattern.

Table 1 the statistical parameters of training, testing, and validation sets

	WD	WD_{t-1}	Tmax	Rad	Eva	WD_{t-2}	$Tmax_{t-1}$	Eva_{t-2}
Training set data								
$\bar{X}$	5.937	5.937	4.487	3.667	5.938	1.762	4.489	1.755
X_{max}	6.242	6.242	6.245	5.568	6.242	3.162	6.245	3.131
X_{min}	5.602	5.602	3.000	1.732	5.649	0.447	3.000	0.447
S_x	0.126	0.126	0.697	0.986	0.125	0.588	0.697	0.579
Testing set data								
	WD	WD_{t-1}	Tmax	Rad	Eva	WD_{t-2}	$Tmax_{t-1}$	Eva_{t-2}
$\bar{X}$	5.933	5.933	4.475	3.645	5.933	1.732	4.484	1.771
X_{max}	6.234	6.242	6.205	5.568	6.242	3.066	6.205	3.162
X_{min}	5.690	5.649	3.082	1.732	5.602	0.633	3.162	0.447
S_x	0.119	0.123	0.703	0.993	0.123	0.571	0.680	0.579
Validation set data								
	WD	WD_{t-1}	Tmax	Rad	Eva	WD_{t-2}	$Tmax_{t-1}$	Eva_{t-2}
$\bar{X}$	5.934	5.931	4.445	3.689	5.929	1.737	4.423	1.735
X_{max}	6.242	6.234	6.124	5.385	6.242	3.066	6.124	3.131
X_{min}	5.700	5.700	3.082	1.732	5.694	0.447	3.082	0.447
S_x	0.120	0.117	0.664	0.955	0.121	0.554	0.681	0.596

$\bar{X}$=mean, X_{max}=maximum value, X_{min}=minimum value, S_x=standard division

7.2 Hybrid Heuristic Algorithms-ANN Techniques

Several sizes of a population were simulated in MATLAB for the hybrids (BSA-ANN and GSA-ANN) to let each hybrid algorithm determine the optimal learning rate value and number of neurons in both hidden layers of the ANN. Thereby, the minimum value of the fitness function could be obtained as shown in Figurer 5. The figure was displayed that the GSA-ANN is capable of attaining the best fitness function at the 40 population size for all seasonal and yearly water consumption data. Thus, the output of the GSA algorithm has been selected to develop the ANN model for water demand. Accordingly, table 2 presents the ANN factors for the best population size for all data types.

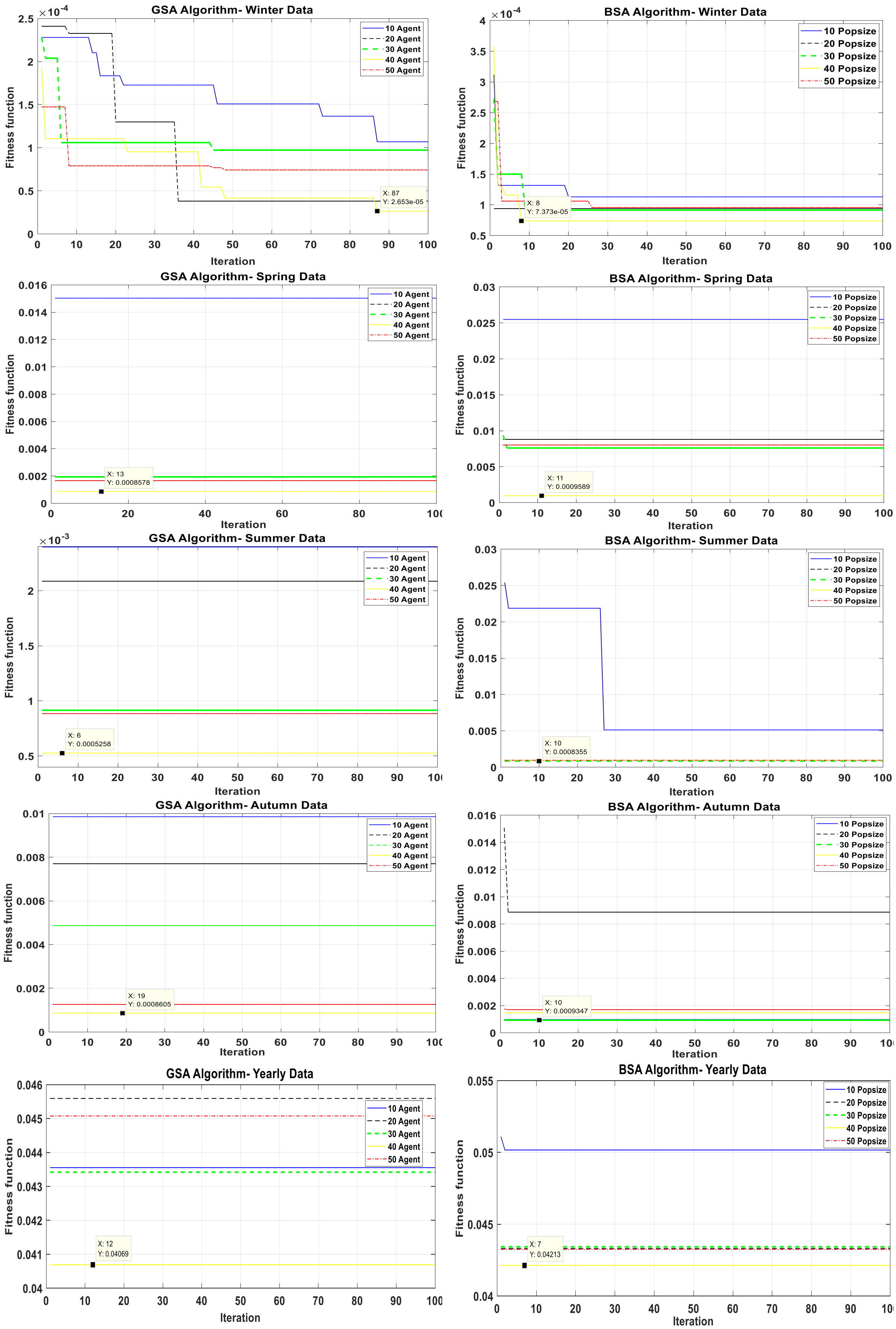

Figure 5: Fitness function versus iteration of all data kinds for GSA and BSA algorithms

Table 2: ANN parameters based on GSA-ANN algorithm for all data types

Data type	Parameters	GSA-ANN
Winter	N1	15
	N2	10
	LR	0.4434
Spring	N1	17
	N2	19
	LR	0.5198
Summer	N1	16
	N2	17
	LR	0.1477
Autumn	N1	19
	N2	19
	LR	0.9729
Yearly	N1	19
	N2	17
	LR	0.5412

N1: Number of neurons in hidden layer one, N2: Number of neurons in hidden layer two and LR: ANN's learning rate.

The regression coefficient (R) of determination between the observed and predicted municipal water is a perfect indicator for the exploration of the forecast performance of the hybrid GSA-ANN algorithm. In addition, R is equal to 0.99, 0.99, 0.97, 0.97 and 0.95 for winter, spring, summer, autumn and yearly data respectively. It can be seen that the values of (R) for seasonal data are more than the yearly data for the validation stages. These interesting findings could be because the model is more able to capture the relation between input and output factors using seasonal data compared to when using yearly data.

To examine the agreement of the model for seasonal and yearly water demand data, a Bland–Altman scatter plot was employed. It has the ability to reveal the systematic and random differences as well as the merit of exhibiting the variation in the outcomes. In this plot, mean (m) and standard deviation (SD) of the differences were obtained by applying the T test technique. In addition, m+1.96 SD and m−1.96 SD represent the upper and lower limits of agreement. From this plot, it is quite easy to evaluate the level of (systematic) variation, the scatter of the values and to display whether there is a relation between the observed and predicted error.

The most obvious finding to emerge from the analysis is that scattered data suggests an excellent distribution fit between agreement limits for seasonal and yearly data, as shown in Figure 6. Table 3 shows the percentage of data distributed between the agreement limits for seasonal and yearly. One of the issues that emerged from these findings was that the percentage range values were from 93.8 to 95.0. In addition, Table 3 presents four statistical indicators used to evaluate the model performance. These findings further indicate that no statistically significant difference was observed between measured and forecasted water demand, especially for winter season data.

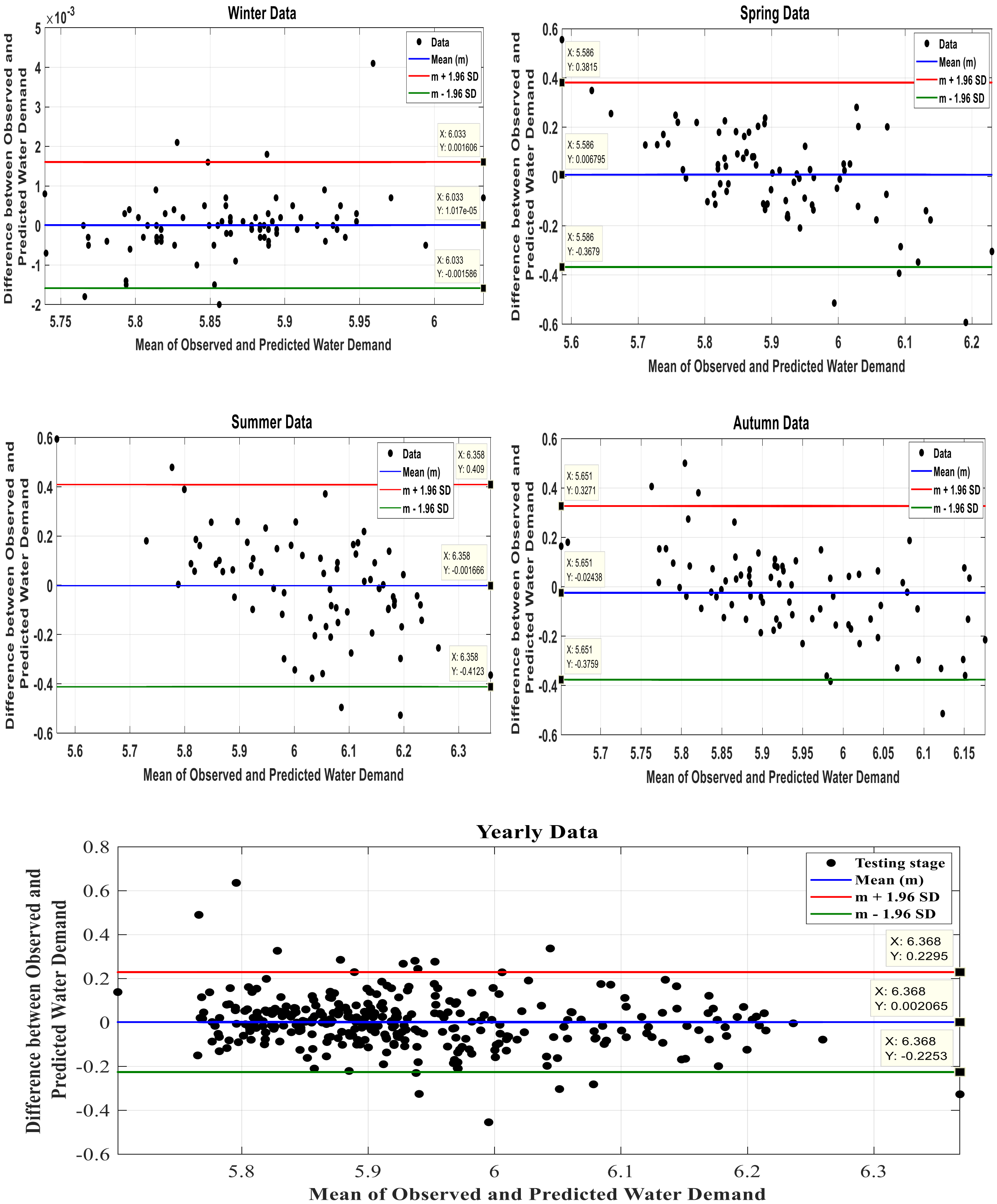

Figure 6: Bland–Altman plot of the relation between observed and predicted municipal water for seasonal and yearly data

Table 3: Several statistical parameters for seasonal and yearly data.

Data	RMSE	MSE	MAE	MAPE	BAPA %
Winter	8.094 e^{-04}	6.551 e^{-07}	5.086 e^{-04}	2.1522e^{-06}	93.8
Spring	0.19	0.0361	0.1449	0.0012	94.7
Summer	0.2081	0.0433	0.1612	1.4295e^{-04}	94.7
Autumn	0.1799	0.0324	0.1363	0.0040	93.8
Yearly	0.1159	0.0134	0.0833	4.7264e^{-04}	95.0

RMSE: root mean square error, MSE: mean square error, MAE: mean absolute error and BAPA: Bland–Altman plot accuracy

According to this data analysis and statistical criteria, it can be inferred that these results provide further support for the hypothesis that water demand is driven by weather variables. In addition, this study confirms that data pre-processing techniques, data division and selection of explanatory variables are associated with the accuracy and robustness of results. Another important finding was that the model's capability to catch the pattern of time series data depends on seasonal data rather than yearly data. Moreover, the winter season model reveals more accuracy and less error compared with the rest of the models. A possible explanation for this might be that winter weather factors have less variability than those in other seasons. Furthermore, the GSA-ANN algorithm model is a robust technique which has sufficient capability to predict water demand considering trend and seasonal pattern for seasonal and yearly time series data.

8 Conclusion

Prediction of water demand can assist in determining convenient alternatives for ensuring the balance between water supply and demand. The aim of this research was to examine the potential input variables to select the best model input by adopting several different statistical techniques. These techniques consisted of data screening, cross-correlation matrix, autocorrelation and variance inflection factor. The second aim of this study was to determine the accurate prediction of short-term future urban water demand considering weather factors. To achieve these predictions, hybrid GSA-ANN and BSA-ANN algorithms were utilised. The most obvious findings to emerge from this study are that: (a) statistical criteria are powerful techniques for selecting appropriate model inputs, and (b) the GSA-ANN (Agent=40) hybrid model is quite suitable compared with the other type of hybrid model in this study in terms of water demand estimation accuracy. A key strength of this study is that selection of best model inputs and ANN factors based on optimisation techniques is better than using a trial and error method. This research could be further advanced by assessing the effects of extra weather variables, depending on the availability of data, on water demand for different locations around the world.

References

Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B. & Sliusarieva, A. 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research,* 48**,** 1-14.

Adamowski, J. F. 2008. Peak daily water demand forecast modeling using artificial neural networks. *Journal of Water Resources Planing and Management,* 134**,** 119-128.

Ahmed, M., Mohamed, A., Homod, R. & Shareef, H. 2016. Hybrid LSA-ANN Based Home Energy Management Scheduling Controller for Residential Demand Response Strategy. *Energies,* 9**,** 716.

Ahmed, M. S., Mohamed, A., Khatib, T., Shareef, H., Homod, R. Z. & Ali, J. A. 2017. Real time optimal schedule controller for home energy management system using new binary backtracking search algorithm. *Energy and Buildings,* 138**,** 215-227.

ASCE Task Committee, 2000. Artificial neural networks in hydrology. I: preliminary concepts. *Journal of Hydrologic Egineering.* ASCE Task Committee,.

Babel, M. S. & Shinde, V. R. 2011. Identifying prominent explanatory variables for water demand prediction using artificial neural networks: a case study of Bangkok. *Water Resources Management,* 25**,** 1653-1676.

Bakker, M., Van Duist, H., Van Schagen, K., Vreeburg, J. & Rietveld, L. Improving the performance of water demand forecasting models by using weather input. 12th International Conference on Computing and Control for the Water Industry, 1-1-2014 2014. 93-102.

Basheer, I. A. & Hajmeer, M. 2000. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods,* 43**,** 3-31.

Behboudian, S., Tabesh, M., Falahnezhad, M. & Ghavanini, F. A. 2014. A long-term prediction of domestic water demand using preprocessing in artificial neural network. *Journal of Water Supply: Research and Technology—AQUA,* 63**,** 31-42.

Bougadis, J., Adamowski, K. & Diduch, R. 2005. Short-term municipal water demand forecasting. *Hydrological Processes,* 19**,** 137-148.

Chen, D., Zou, F., Lu, R. & Wang, P. 2017. Learning backtracking search optimisation algorithm and its application. *Information Sciences,* 376**,** 71-94.

Donkor, E. A., Mazzuchi, T. H., Soyer, R. & Roberson, J. A. 2014. Urban water demand forecasting: review of methods and models. *Journal of Water Resources Planing and Management,* 140**,** 146-159.

Firat, M., Turan, M. E. & Yurdusev, M. A. 2010. Comparative analysis of neural network techniques for predicting water consumption time series. *Journal of Hydrology,* 384**,** 46-51.

Firat, M., Yurdusev, M. A. & Turan, M. E. 2009. Evaluation of artificial neural network techniques for municipal water consumption modeling. *Water Resources Management,* 23**,** 617-632.

Fogden, J. & Wood, G. 2009. Access to Safe Drinking Water and Its Impact on Global Economic Growth. *HaloSource Inc*.

Gato, S., Jayasuriya, N. & Hadgraft, R. 2005. A simple time series approach to modelling urban water demand. *Australian Journal of Water Resources,* 8**,** 153-164.

Gharghan, S. K., Nordin, R. & Ismail, M. 2016a. A Wireless Sensor Network with Soft Computing Localization Techniques for Track Cycling Applications. *Sensors (Basel),* 16.

Gharghan, S. K., Nordin, R., Ismail, M. & Ali, J. A. 2016b. Accurate wireless sensor localization technique based on hybrid pso-ann algorithm for indoor and outdoor track cycling. *Institute of Electrical and Electronics Engineers Sensors Journal,* 16**,** 529-541.

Jain, A. & Ormsbee, L. E. 2002. Short-term water demand forecast modeling techniques conventional methods versus AI. *American Water Works Association,* 94**,** 64-72.

Jain, A., Varshney, A. K. & Joshi, U. C. 2001. Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. *Water Resources Management,* 15**,** 299–321.

Kotsiantis, S. B., Kanellopoulos, D. & Pintelas, P. E. 2006. Data preprocessing for supervised leaning. *International Journal of Computer Science,* 1**,** 111-117.

Liu, J., Savenije, H. H. G. & Xu, J. 2003. Forecast of water demand in Weinan City in China using WDF-ANN model. *Physics and Chemistry of the Earth, Parts A/B/C,* 28**,** 219-224.

Maier, H. R. & Dandy, G. C. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling & Software,* 15**,** 101–124.

Pallant, J. 2011. *SPSS SURVIVAL MANUAL : A step by step guide to data analysis using SPSS,* Australia, Allen & Unwin.

Payal, A., Rai, C. S. & Reddy, B. V. R. 2015. Analysis of Some Feedforward Artificial Neural Network Training Algorithms for Developing Localization Framework in Wireless Sensor Networks. *Wireless Personal Communications,* 82**,** 2519-2536.

Rashedi, E., Nezamabadi-Pour, H. & Saryazdi, S. 2009. GSA: A Gravitational Search Algorithm. *Information Sciences,* 179**,** 2232-2248.

Sarker, R. C., Gato, S. & Imteaz, M. Temperature and rainfall thresholds corresponding to water consumption in Greater Melbourne, Australia. 20th International Congress on Modelling and Simulation, 1–6 December 2013 2013 Adelaide, Australia. Modelling and Simulation Society of Australia and New Zealand, 2576-2582.

Shahin, M. A., Jaksa, M. B. & Maier, H. R. 2008. State of the Art of Artificial Neural Networks in Geotechnical Engineering. *Electronic Journal of Geotechnical Engineering,* 13**,** 1-26.

Shuaib, M., Kalavathi, S. M. & Rajan, C. A., C. 2015. Optimal capacitor placement in radial distribution system using Gravitational Search Algorithm. *International Journal of Electrical Power & Energy Systems,* 64**,** 384-397.

Su, Z., Wang, H. & Yao, P. 2016. A hybrid backtracking search optimization algorithm for nonlinear optimal control problems with complex dynamic constraints. *Neurocomputing,* 186**,** 182-194.

Tabachnick, B. G. & Fidell, L. S. 2013. *Using Multivariate Statistics,* United States of America, Pearson Education, Inc.

Xiong, H., Pandey, G., Steinbach, M. & Kumar, V. 2006. Enhancing data analysis with noise removal. *Institute of Electrical and Electronics Engineers Transactions on Knowledge and Data Engineering,* 18**,** 304-319.
YVW, 2017. Yarra Valley Annual Report Water 2016-2017. Australia.
Zhang, J. J., Song, R., Bhaskar, N. R. & French, M. N. Short-term water demand forecasting: a case study. 8th Annual Water Distribution Systems Analysis Symposium, August 27-30, 2006 2006 Cincinnati, Ohio, USA. United States, 1-14.
Zhoua, S. L., Mcmahon, T. A., Walton, A. & Lewis, J. 2000. Forecasting daily urban water demand: a case study of Melbourne. *Journal of Hydrology,* 236**,** 153–164.