



## LJMU Research Online

Zubaidi, SL, Abdulkareem, IH, Hashim, KS, Al-Bugharbee, H, Ridha, HM, Gharghan, SK, Al-Qaim, FF, Muradov, M, Kot, P and Alkhaddar, R

**Hybridised Artificial Neural Network model with Slime Mould Algorithm: A novel methodology for prediction urban stochastic water demand**

<http://researchonline.ljmu.ac.uk/id/eprint/13719/>

### Article

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

**Zubaidi, SL, Abdulkareem, IH, Hashim, KS, Al-Bugharbee, H, Ridha, HM, Gharghan, SK, Al-Qaim, FF, Muradov, M, Kot, P and Alkhaddar, R  
Hybridised Artificial Neural Network model with Slime Mould Algorithm: A novel methodology for prediction urban stochastic water demand. Water.**

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

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

For more information please contact [researchonline@ljmu.ac.uk](mailto:researchonline@ljmu.ac.uk)

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

Article

# Hybridised Artificial Neural Network model with Slime Mould Algorithm: A novel methodology for prediction urban stochastic water demand

Salah L. Zubaidi <sup>1,\*</sup>, Iqbal H. Abdulkareem <sup>2</sup>, Khalid S. Hashim <sup>3,4</sup>, Hussein Al-Bugharbee <sup>5</sup>, Hussein Mohammed Ridha <sup>6</sup>, Sadik Kamel Gharghan <sup>7</sup>, Fuod F. Al-Qaim <sup>8</sup>, Magomed Muradov <sup>3</sup>, Patryk Kot <sup>3</sup>, Rafid Al-Khaddar <sup>3</sup>

<sup>1,\*</sup> Department of Civil Engineering, Wasit University, Wasit, Iraq, [salahlafta@uowasit.edu.iq](mailto:salahlafta@uowasit.edu.iq);

<sup>2</sup> Environmental Research and Studies Centre, University of Babylon, Hill, Babylon City, Iraq

<sup>3</sup> BEST Research Institute, Liverpool John Moores University, Liverpool, UK

<sup>4</sup> Department of Environment Engineering, University of Babylon, Babylon, Iraq

<sup>5</sup> Department of Mechanical Engineering, Wasit University, Wasit, Iraq

<sup>6</sup> Department of Electrical and Electronics Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400, Serdang, Malaysia

<sup>7</sup> Department of Medical Instrumentation Techniques Engineering, Electrical Engineering Technical College Middle Technical University, Al Doura, Baghdad, Iraq

<sup>8</sup> Faculty of Science for Women, Babylon University, Babylon, Iraq

**Abstract:** Urban water demand prediction based on climate change is always challenging for water utilities because of the uncertainty which results from a sudden rise in water demand due to stochastic patterns of climatic factors. For this purpose, a novel combined methodology including, firstly, data pre-processing techniques were employed to decompose the time series of water and climatic factors by using Empirical Mode Decomposition and identifying the best model input via tolerance to avoid multi-collinearity. Second, the Artificial Neural Network (ANN) model was optimised by an up-to-date Slime Mould Algorithm (SMA-ANN) to predict the medium term of the stochastic signal of monthly urban water demand. Ten climatic factors over 16 years were used to simulate the stochastic signal of water demand. The results reveal that SMA outperforms Multi-Verse Optimiser and Backtracking Search Algorithm based on error scale. The performance of the hybrid model SMA-ANN is better than ANN (stand-alone) based on the range of statistical criteria. Generally, this methodology yields accurate results with a coefficient of determination of 0.9 and a mean absolute relative error of 0.001. This study can assist local water managers to efficiently manage the present water system and plan extensions to accommodate the increasing water demand.

**Keywords:** Artificial neural network; backtracking search algorithm; empirical mode decomposition; multi-verse optimiser; slime mould algorithm; water demand model

## 1. Introduction

Security of municipal water is fundamental to gain a sustainable environment in modern cities especially under the impact of global warming and socio-economic variables. Additionally, most cities are located close to the freshwater sources to ensure the prosperity of both industry and agriculture. For the mentioned reasons, freshwater scarcity is a classic problem for policymakers [1,2]. Recently, the World Economic Forum confirmed that water scarcity is one of the largest international risks because of the limited amount of accessible freshwater (approximately 0.014% of the total amount of water on the planet of Earth). In addition, climate change, water pollution and poor management of freshwater

sources are other major factors that contribute to water scarcity. For example, climate change increases water demand that increases the pressure on the urban water system, especially during periods of water shortage. Accordingly, water companies must devise strategies to tackle this issue. Hence, water utilities should support and enhance the management of the municipal water system [3-5].

Hemati, *et al.* [6] reported that, during the period from 1930 to 2009, Melbourne City experienced two short-term droughts and three long-term droughts (the last one was from 1997 to 2009). As a result, annual rainfall was adversely affected, making its influence on the municipal water supply system keenly felt. In addition, Roy and Majumder [7] stated that hazardous weather events, such as heatwaves in Australia, are likely to be more regular or extreme, leading to alterations in the amount of freshwater available.

In the last few decades, there has been a surge of interest in the effects of climate change on the demand for freshwater. Although data from several sources have identified that the increase of urban water demand is associated with climatic factors, researchers have not looked at the variability of climate change in much detail. Therefore, to date, the impact of climate change on urban water demand has still not yet been extensively investigated. Consequently, the uncertainty in water sustainability has increased due to the lack in the knowledge about the impact of climate change on the demand of water[8].

Prediction of municipal water demand is crucial to manage, operate and plan urban water systems. It leads to a balance between water need and delivery, especially with the fluctuations caused by climate change that causes unexpected growth in water demand, fluctuations in stochastic patterns, and it also increases the uncertainty for water utilities. Additionally, estimating water demand is quite useful for proper management, operation and upgrading of present freshwater sources [8-10]. Ebrahim Banihabib and Mousavi-Mirkalaei [11] stated that modelling techniques for prediction of municipal water demand are required to address the potential water security. Prediction of medium-term water demand is classified as a tactical type, and it is used to improve water supply networks and installations and to manage dam reservoirs [12].

A review of different municipal water demand modelling techniques over the past few decades,[5,13-16], highlighted various issues that include: 1) Most of these methods and techniques were focused on the prediction of short-term water demand, and little research has considered medium- to long-term urban water demand. 2) Different artificial intelligence techniques were successfully applied to simulate urban water demand, such as artificial neural network [12], support vector regression [19], adaptive neuro-fuzzy inference system [20] and random forests [21]. 3) Several studies indicated that hybrid techniques yield superior outcomes, when compared with the classical single models, such as Chen, *et al.* [21] and Altunkaynak and Nigussie [22]. 4) A number of methods, techniques and models have been used to forecast water demand basing on weather or climatic factors. However, these techniques have been largely deterministic and focus generally on only a few weather variables, such as the study of Mouatadid and Adamowski [23]. Very limited research has used the stochastic signal to predict water demand based on climatic factors, such as the studies of Zubaidi, *et al.* [24] and Zubaidi, *et al.* [8]. These studies used the stochastic model to predict the stochastic signal of water demand of an area in the Melbourne City that served by the Yarra Valley Water utility considering the effects of five climatic factors and applied two hybrid models (particle swarm optimisation and artificial neural network (PSO-ANN)) and lightning search algorithm and artificial neural network (LSA-ANN).

Although the application of machine learning and data analytic methodologies in urban water estimating have grown considerably in recent years, more investigations are needed to develop new methodologies to enhance the results from present conservation and demand management programmes [5]. Altunkaynak and Nigussie [9] stated that the artificial neural network is a preferred option to simulate water demand because it can deal with non-linear time series.

Different metaheuristic optimisation algorithms could be applied to tackle a range of problems for various application domains. The main advantages of optimisation algorithms are their ability to select the optimal values of parameters of the system under different conditions and they have time-saving

qualities. Recently, a multi-verse optimiser (MVO) proposed by Mirjalili, *et al.* [25] to solve various optimisation problems, for example, it has been used for energy management in smart cities [26], multi-level image segmentation [27] and text document clustering [27]. Additionally, a backtracking search algorithm (BSA) has been utilised to tackle several optimisation issues, such as predicting urban water demand depending on previous water consumption data [28], photovoltaic models [29] and power signals [30]. Moreover, Slime Mould Algorithm (SMA) has been proposed by Li, *et al.* [31] and used in several optimisation issues, but has not been investigated in the urban water sector, such as spring design problem [32], photovoltaic models [33] and image segmentation method [34].

In addition, the literature has emphasised the importance of using data pre-processing to improve the quality of time series and to determine the best independent variables. More attention has recently focused on data cleaning. Therefore, several signal pre-treatment techniques have been employed to clean and/or detect the trend, seasonal and stochastic components of water consumption time series, such as singular spectrum analysis (SSA) [35,36], wavelet transform (WT) [22,37], variational mode decomposition (VMD) [17] and empirical mode decomposition (EMD) [38,39]. Another significant aspect of data pre-processing is selecting the best independent variables, such as principal component analysis (PCA) [40,41], mutual information (MI) [42,43] and variance inflation factor (VIF) [18,24].

Despite that different techniques and approaches were used to forecast the future water demand, water companies still face challenges in estimating the accurate water demand, especially with the influence of climatic factors and their implications for future water demand. Therefore, additional research studies are required to accurately estimate the growing water demand [8].

In this context, the main contributions of the current research are:

1. The employment of 10 climatic factors over 16 years to assess the impact of climate change on urban water demand.
2. Development and analyse of a new hybrid algorithm SMA-ANN for water demand optimisation problem, and to choose the optimal hyperparameters of the ANN approach.
3. The application of two hybrid algorithms, MVO-ANN and BSA-ANN, for analysing and validating the proposed SMA-ANN algorithm.
4. Using the novel methodology, which contains data pre-processing techniques (EMD and tolerance) and hybrid SMA-ANN algorithm, to simulate the monthly stochastic pattern of water demand based on the best scenario of climatic factors over 16 years.
5. Minimising the uncertainty by applying three metaheuristic algorithms for more validation, and using the ANN (stand-alone) to confirm the results of SMA-ANN model. Additionally, employing 10 climatic factors that give scientific insight (i.e. to what extent climate change has driven water demand) for policymakers to achieve sustainability.

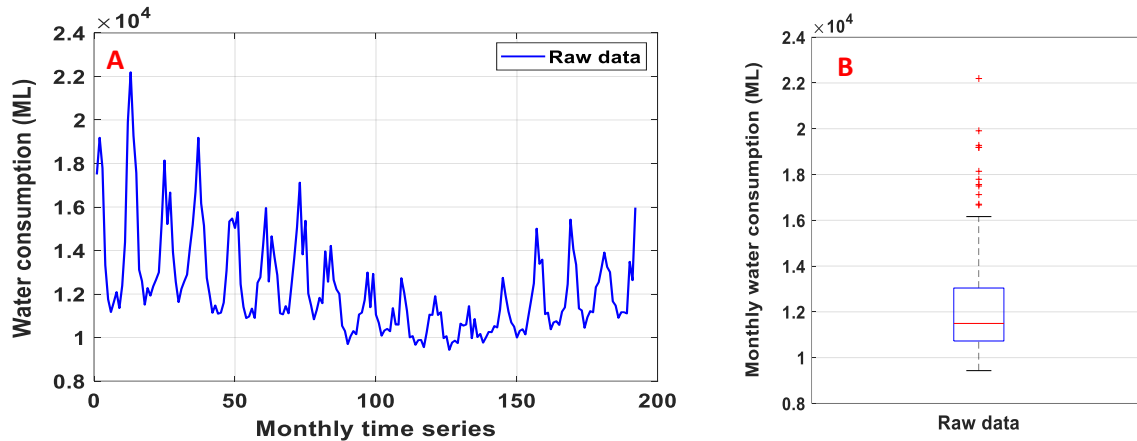
To the best of the authors' knowledge, the present study explores a novel methodology, for the first time: the effects of climate change on the monthly stochastic pattern of urban water demand. The structure of the research is organised as follows: case study and data used are presented in Section 2; the proposed methodology for predicting monthly stochastic water demand is described in Section 3; Section 4 provides the analysis and compares the obtained results. Finally, section 5 presents the final conclusions with some considerations of the study.

## 2. Case Study and Data Used

The suggested methodology was applied to the observed water consumption and climatic factor data relating to the South East Water (SEW) utility, which is one of three retail water utilities that purchase water wholesale from Melbourne Water company in Melbourne City, Australia. The sources of freshwater are 11 large storage facilities, which are refilled regularly by stormwater harvesting [44]. The urban water system network of SEW utility serves more than 1.7 million individuals in a 3640 km<sup>2</sup> area, and the company has approximately 727,00 customers, categorised into residential, industrial and commercial [45].

The collected data comprise monthly urban water consumption (megalitre, ML), maximum temperature (Tmax) (°C), minimum temperature (Tmin) (°C), mean temperature (Tmean) (°C), solar

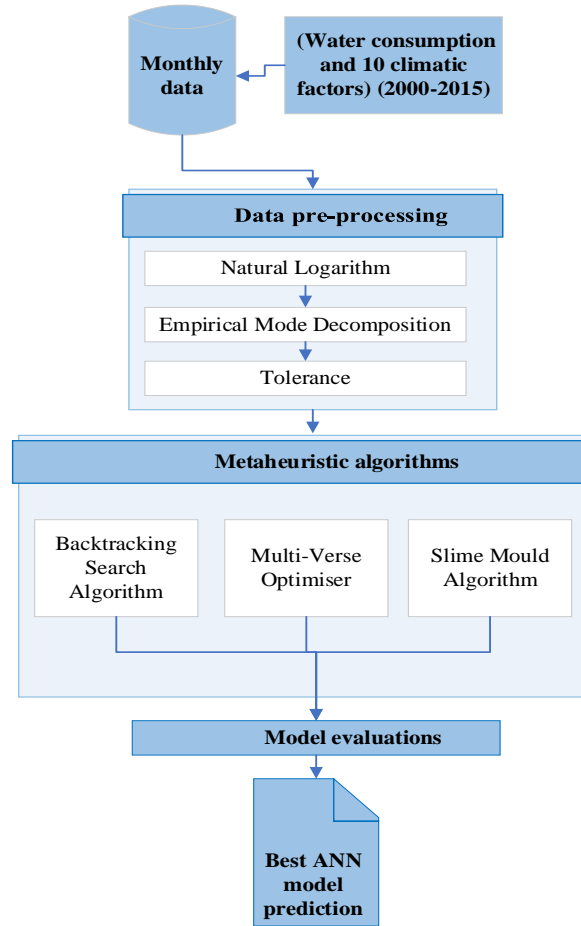
radiation (Srad) ( $\text{MJ}/\text{m}^2$ ), potential evapotranspiration (FA-O56) (mm), vapour pressure (VP) (hpa), rainfall (Rain) (mm), evaporation (Eva) (mm) maximum relative humidity (RHmax) (%), and minimum relative humidity (RHmin) (%) from 2000 to 2015. Figure 1 shows the time series and box-plot of monthly water consumption for SEW utility. The figure reveals the decrease in water consumption due to drought, and water-conserving policies and initiatives. After that, the consumption increased, possibly because restrictions were eased after the impact of the drought lessened. Also, it may be due to the strategies that Melbourne Corporation pursued by upgrading the dams and relying on other resources, such as water desalination and water recycling [44].



**Figure 1.** A) Monthly time series, B) box-plot of municipal water consumption for SEW utility.

### 3. Proposed Methodology

There is a relatively small body of literature that is concerned with the impact of climate change only on the municipal water demand. Accordingly, this paper proposes a novel combined methodology for investigating the impact of climate change on water demand. It could be divided into five main categories: data pre-processing, slime mould algorithm (SMA), artificial neural network (ANN), hybrid metaheuristic algorithm-based artificial neural network, and model evaluation. Figure 2 shows the structure of the proposed methodology to predict monthly stochastic data of water demand based on climatic factors.



**Figure 2.** The proposed methodology for predicting water demand based on climatic factors.

### 3.1. Data Pre-processing

Recent developments in urban water predictive methodologies have highlighted the need to apply different data pre-processing techniques, which could be classified into normalisation, cleaning and select best model input [5,46]. To be in accordance with Tabachnick and Fidell [47], the natural logarithm was used to normalise all raw time series of water and climatic factors to reduce both the impact of the outliers and the multicollinearity between independent factors.

Zubaidi, et al. [8] mention that the relation between water demand time series and climatic time series is stochastic. Also, stochastic models will offer a better reflection of reality and insight into the system's dynamics [48]. Hence, an empirical mode decomposition (EMD) approach was applied to decompose the original time series of dependent and independent variables into a trend, seasonal, stochastic and noise components, and detect the stochastic component after that. EMD is used in the analysis of various problems, such as machinery fault diagnosis [49] and biomedical signal analysis [50].

EMD is also used for analysing geodetic data [51]. In this study, the authors have used EMD to analyse the natural variability of sea level and its effect, among other factors, on the trend of the sea level. Recently, Chu and Huang [52] utilised EMD for synthetisation and generation of flow data and for increasing the number of flow time series for the same time period, which is used in the simulation of a water supply system. This technique decomposes a time series into a number of time-domain components called Intrinsic Mode Functions (IMFs), the latter must have two properties:

- The maximum difference between the number of local maxima and minima is one.
- The mean value of an IMF is zero.

For a time series  $x(t)$ , the extraction of IMFs could be described briefly in the following steps [53]:

1. Assume  $h_{k-1}(t) = x(t)$ , and  $h_{i,k-1}(t) = x(t)$ , where  $I$  and  $k$  refer to the IMF number and the iteration number for finding the accurate  $i^{\text{th}}$  IMF, respectively.
2. Identify all the maxima and minima points of the series  $h_{i,k-1}(t)$ .
3. Connect the maxima points by cubic splines interpolation and do the same thing for the minima points. The linked maxima points are called the upper envelop,  $U_{i,k-1}(t)$ , while the linked minima points are called the lower envelope,  $L_{i,k-1}(t)$ .
4. The mean of the upper and lower envelopes is found using this formula:  $m_{i,k-1}(t) = (U_{i,k-1}(t) + L_{i,k-1}(t))/2$ .
5. Form the following formula:  $h_{i,k}(t) = h_{i,k-1}(t) - m_{i,k-1}(t)$ . The component  $h_{i,k}(t)$  is primarily described as the first IMF. To determine the first IMF accurately, the  $h_{i,k}(t)$  is considered as a new signal and the mean of upper envelope, lower envelope and the mean (i.e.  $U_{i,k}(t)$ ,  $L_{i,k-1}(t)$  and  $m_{i,k}$  of the  $h_{i,k}(t)$ ) are calculated. The new component  $h_{i,k}(t)$  is checked to see whether it has IMF properties or not. If it does, then it (i.e.  $h_{i,k}(t)$ ) is identified as an IMF. If not, the process will be repeated until IMF properties are obtained. The number of the repetitions to identify an IMF is called iterations and notated by  $k$  while the IMF number is notated by  $i$ .
6. When the  $i^{\text{th}}$  IMF is obtained, the residue is obtained;  $res_i = h_{i,k-1} - \text{IMF}_i$ .
7. The residue  $res_i$  is now treated as the signal  $h_{i+1,k-1}$  and the same steps 2-6 are repeated till no more IMF can be extracted.

The EDM process above is applied for all dependent and independent variables in this study.

Regarding the selection of the best model input, Pallant [54] recommended using a tolerance method to choose the independent variables that have a tolerance value more than 0.1 because values less than 0.1 indicate the presence of multi-collinearity.

### 3.2. Slime Mould Algorithm (SMA)

The SMA is one of the recent nature-inspired algorithms. It refers to the mathematical model of simulating the propagation wave of slime mould when forming the optimal path for connecting foods. This model adaptively simulates the process of producing negative and positive feedback during the propagation wave. This algorithm is incorporated into different optimisation problems including the engineering ones. The main two stages in the SMA algorithm are called approaching food and warp food.

#### a. Approach food:

In this stage, the slime is approaching food based on its odour in the air and this behaviour is mathematically described as follows:

$$X(t+1) = \begin{cases} \overrightarrow{X_b(t)} + \overrightarrow{vb} \cdot (\overrightarrow{W} \cdot \overrightarrow{X_A(t)} - \overrightarrow{X_B(t)}), & r < p \\ \overrightarrow{vc} \cdot \overrightarrow{X(t)}, & r \geq p \end{cases} \quad (1)$$

where

$\overrightarrow{vb}$  is a parameter which ranges from  $-a$  to  $a$ .

$\overrightarrow{vc}$  represents a parameter which decreases from one to zero in a linear form.

$X_b$  represents the current individual location corresponding to high odour concentration.

$t$  is the current iteration.

$X$  is the location of the slime mould.

$X_A$  and  $X_B$  are randomly selected individuals from the mould.

$W$  is the weight of the slime mould.

The formula of  $p$  can be represented as follows:

$$p = \tanh[S(i) - DF] \quad \text{where } i \in 1, 2, 3, \dots, n \quad (2)$$

$S(i)$  represents the fitness of  $\vec{X}$ .

$DF$  represents the best fitness over all the iterations.

As mentioned above,  $\overrightarrow{vb}$  ranges from  $-a$  to  $a$ , and  $a$  can be described as follows:

$$a = \text{arctanh} \left( -\left( \frac{t}{\max t} \right) + 1 \right) \quad (3)$$

The  $\vec{W}$  formula can be described as follows:

$$\overline{W(SmeellIndex(t))} = \begin{cases} 1 + r \cdot \log\left(\frac{bF - S(i)}{bF - wF} + 1\right), & \text{condition} \\ 1 - r \cdot \log\left(\frac{bF - S(i)}{bF - wF} + 1\right), & \text{others} \end{cases} \quad (4)$$

$$SmeellIndex = \text{sort}(S) \quad (5)$$

where

$r$  denotes the random value within the interval  $[0,1]$ .

$bF$  represents the optimal fitness obtained in the current iterative process.

$wF$  represents the worst fitness value obtained in the current iterative process.

$SmellIndex$  refers to the sequence of fitness values.

b. Warp food

In this stage, the behaviour of the slime in conducting contraction of its venous structure is mathematically described as follows.

$$\vec{X}^* = \begin{cases} rand \cdot (UB - LB) + LB, & rand < z \\ \vec{X}(t) + vb \cdot (\vec{W} \cdot \vec{X}_A(t) - \vec{X}_B(t)), & r < p \\ \vec{vc} \cdot \vec{X}(t), & r \geq p \end{cases} \quad (6)$$

where  $LB$  and  $UB$  are the lower and upper boundaries of the search range, and  $rand$  and  $r$  are random parameters ranging from 0 to 1. Further details of the SMA can be found in Li, et al. [31]. In this study, the SMA algorithm is combined with the ANN model to determine the optimum parameters of the ANN model (see section 3.4).

### 3.3. Artificial Neural Network (ANN)

In recent years, there has been an increasing interest in using the ANN model to predict urban water demand because it is capable of accurately simulating the nonlinear time series. Additionally, Rahim, et al. [5] reported that different multi-layer feedforward neural networks (MLFFNNs) have been successfully developed utilising propagation networks in a water demand estimation model for short and medium terms. The Levenberg–Marquardt (LM) backpropagation algorithm was employed for training the ANN model because it can effectively simulate any independent/dependent map [55]. The structure of the ANN model can be categorised into four layers: the input layer, two hidden layers, and an output layer. The input layer contains the independent variable (climatic factors), the output layer involves the dependent variable (water demand), and the hidden layers comprise the hidden neurons, which are responsible for stimulating the nonlinear relationship between water consumption and climatic factors. As in Zubaidi, et al. [18], the tansigmoidal activation function was considered in both the hidden layers and the linear activation function was utilised in the output layer. In this research, total data can be randomly categorised into training (70%), testing (15%) and validation datasets (15%) [56]. Following González Perea, et al. [57], the ANN model was integrated with a metaheuristic algorithm to determine the optimal hyperparameters of the ANN model that includes the learning rate coefficient (LR) and the number of neurons hidden in the first (N1) and second (N2) hidden layers.

### 3.4. Hybrid Metaheuristic Algorithm-Based Artificial Neural Network

In the ANN technique, before achieving the stages of training, testing, and validation, it is important to locate two hyperparameters, which are the learning rate coefficient (LR) and the number



of neurons hidden (N1 and N2) for the hidden layer one and two, respectively. These hyperparameters are responsible for mapping the nonlinear relationship among the stochastic signals of water consumption and climatic factors. While the determination of these hyperparameters, based on a trial and error procedure, may not present the optimum solutions. For this purpose, the ANN model is hybridised with the slime mould algorithm (SMA-ANN) (automated machine learning) to select the best LR, N1 and N2 for the ANN model. Additionally, two extra metaheuristics were hybridised with the ANN, the multi-verse optimiser (MVO-ANN) and the backtracking search algorithm (BSA-ANN), to assess and validate the results of the SMA-ANN algorithm. Five population sizes (10, 20, 30, 40 and 50 popsize) with 200 iterations were employed for each hybrid algorithm to select the popsize that could offer the lower value of fitness function (root mean square error, RMSE).

### 3.5. Model Evaluation

In this research, several performance statistical criteria were employed to evaluate the performance of the suggested methodology because there are no global performance criteria that are appropriate for a particular usage. The performance criteria employed in this research are categorised into absolute, relative and dimensionless errors [17]. The absolute error contains the mean absolute error (MAE, Equation (7)) and mean square error (MSE, Equation (8)). The relative error comprises the mean absolute relative error (MARE, Equation (9)). The dimensionless error contains the coefficient of determination ( $R^2$ , Equation (10)). In addition, a Bland–Altman scatterplot is used to graphically represent the upper and lower limits of agreement area between (actual data- simulated data) on the y-axis, and ([actual data + simulated data]/2) on the x-axis. Moreover, Augmented Dickey-Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests are used to examine the stationary of the stochastic component for dependent and independent variables.

$$MAE = \frac{\sum_{i=1}^N |C_i - P_i|}{N} \quad (7)$$

$$MSE = \frac{\sum_{i=1}^N (C_i - P_i)^2}{N} \quad (8)$$

$$MARE = \frac{1}{N} \sum_{i=1}^N \frac{|C_i - P_i|}{C_i} \quad (9)$$

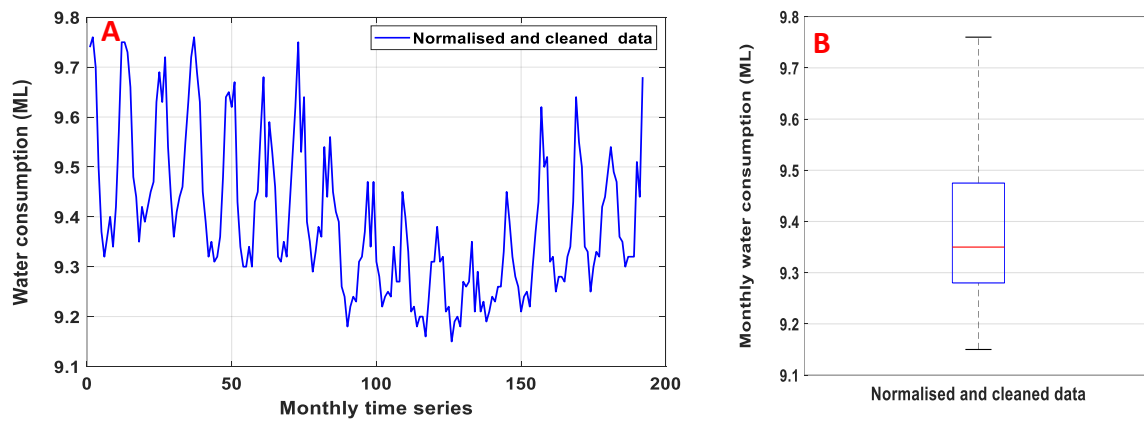
$$R^2 = \left[ \frac{\sum_{i=1}^N (C_i - \bar{C}_i)(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^N (C_i - \bar{C}_i)^2 \sum_{i=1}^N (P_i - \bar{P}_i)^2}} \right]^2 \quad (10)$$

where  $C_i$ : measured water consumption,  $P_i$ : predicted water demand,  $\bar{C}_i$ : mean of measured water consumption,  $\bar{P}_i$ : mean of predicted water demand, N: length of data.

## 4. Results and Discussion

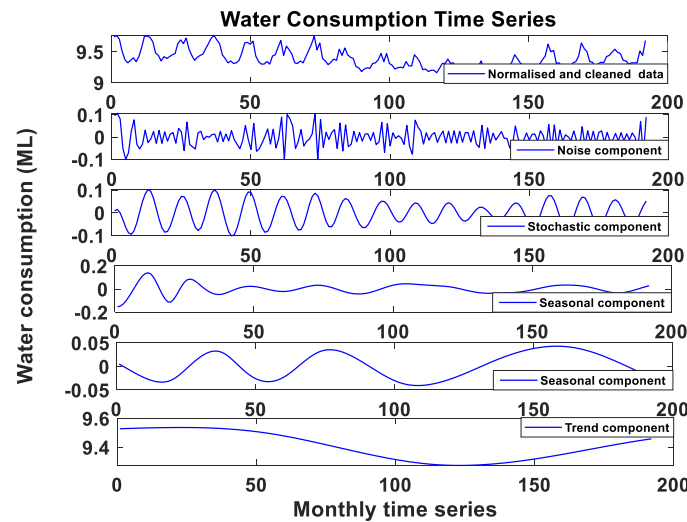
### 4.1. Preparation of Dependent and Independent Variables

First of all, data on water consumption and 10 climatic factors were normalised and cleaned according to section 3.1, Figure 3 shows the normalised and cleaned water data. Figure 3 (A) shows that the variance of the seasonal periods along the time series has been reduced in comparison with Figure 1 (A). Also, Figure 3 (B) shows that the time series, after the normalisation and treating the outliers, tends to follow a normal distribution in comparison with Figure 1 (B).



**Figure 3.** A) Monthly time series, B) box-plot of normalised and cleaned municipal water consumption data.

Then, the EMD approach was applied to analyse the normalised data of water consumption and all climatic factors to unravel the stochastic components of each time series. Figure 4 presents the normalised and cleaned data for water consumption and its decomposed components including trend, seasonal, stochastic and noise. The Augmented Dickey-Fuller and Kwiatkowski–Phillips–Schmidt–Shin tests are used to assess the stationarity of the stochastic component for each factor (i.e., these two tests are used to test and select the stochastic signal). It can be also seen that other components (trend and seasonal) represent the deterministic signal, which is driven by socio-economic factors.



**Figure 4.** Normalised and cleaned water time series and the first five components obtained by EMD.

Table 1 presents the correlation coefficients between water consumption and climatic factors time series in the raw and stochastic stage. The table shows that data pre-processing techniques produce significant improvements in the quality of the data, such as increasing the correlation coefficient between water consumption and maximum temperature time series (from 0.63 to 0.93). Additionally, the correlation coefficient between the stochastic signal of water consumption and climatic factors over sixteen years confirmed the relation between water consumption and climatic factor.

**Table 1.** Correlation coefficients between water consumption and climatic factors in the raw and stochastic stage.

<b>Data</b>	<b>Tmax</b>	<b>Tmin</b>	<b>Tmean</b>	<b>Rain</b>	<b>Eva</b>	<b>Srad</b>	<b>VP</b>	<b>RHmax</b>	<b>RHmin</b>	<b>FA-O56</b>
Raw	0.63	0.61	0.62	-0.10	0.61	0.60	0.55	-0.59	-0.54	0.63
Stochastic	0.93	0.91	0.92	-0.53	0.88	0.83	0.88	-0.89	-0.75	0.88

In the final stage of data pre-processing, a tolerance method was used to locate the best scenario of independent factors, climatic factors, that can accurately simulate the stochastic component of water demand and omit redundant factors to avoid multi-collinearity. The tolerance values for all the climatic factors in the initial stage are less than the minimum limit of acceptance (i.e. it should be more than 0.1) except the Rain, which has tolerance value equal to 0.43. So, the climatic factors that have multicollinearity were removed separately, one by one, until the tolerance values of the selected model were reached more than 0.1, as shown in Table 2. The latter shows that Tmax, RHmin and Rain were selected to be the best scenario of independent factors based on the tolerance value. As presented in Table 2, the tolerance value for each climatic factor was more than 0.1, which means the multi-collinearity assumption was not violated.

**Table 2.** Collinearity statistics to the selected model input.

<b>Climatic factors</b>	<b>Tolerance Value</b>
Tmax	0.322
RHmin	0.344
Rain	0.867

Another graphical technique can show the significance of data pre-processing. Figure 5 presents the boxplot of the stochastic components of water consumption, Tmax, RHmin and Rain. It can be seen that there are no outliers within the data, and the median is zero for water consumption and Tmax, and nearly zero for all RHmin and Rain factors. The stochastic components of all factors showed a normal distribution, which was confirmed by the Kolmogorov-Smirnov test with significance values (Sig.) more than 0.05 (i.e., the Sig. values are 2 for both minimum relative humidity and rainfall factors). Additionally, Figure 5 reveals how the EMD technique enhances the normal distribution of water consumption compared with the distribution of water consumption in Figure 3 (B).

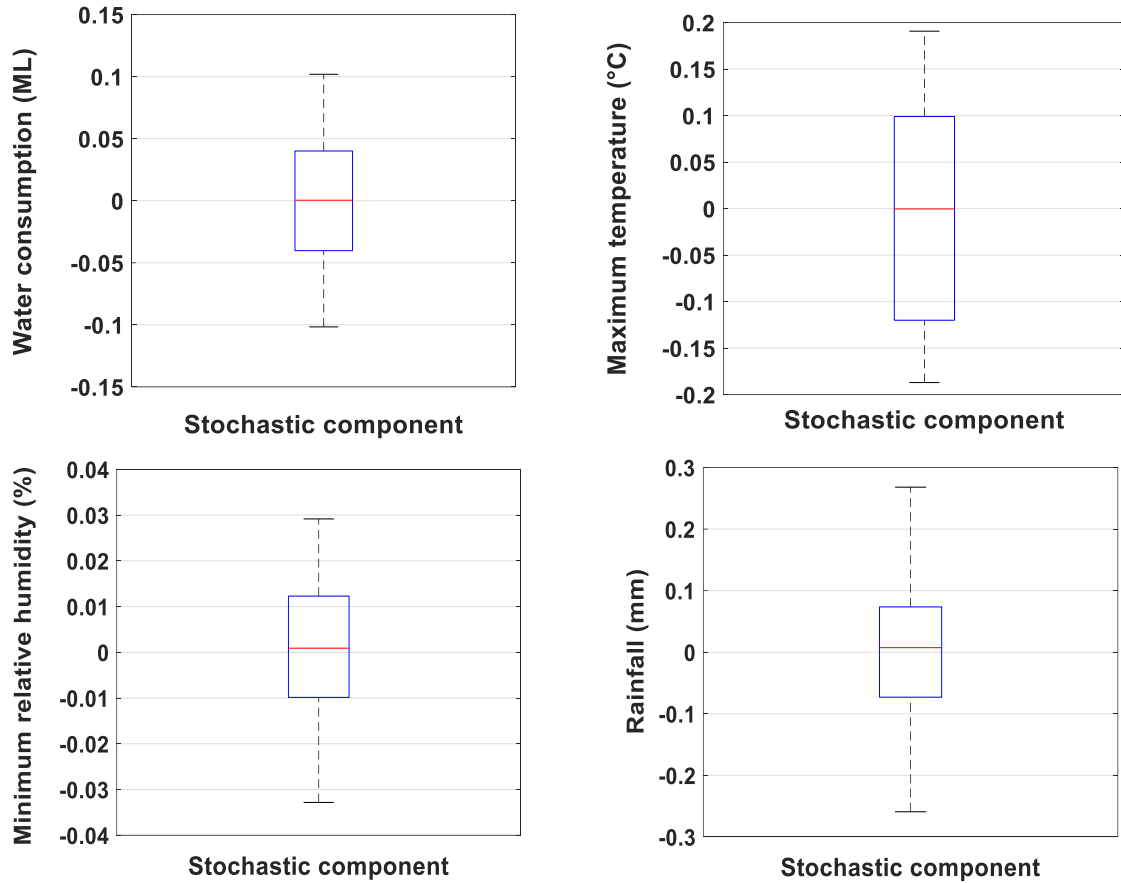
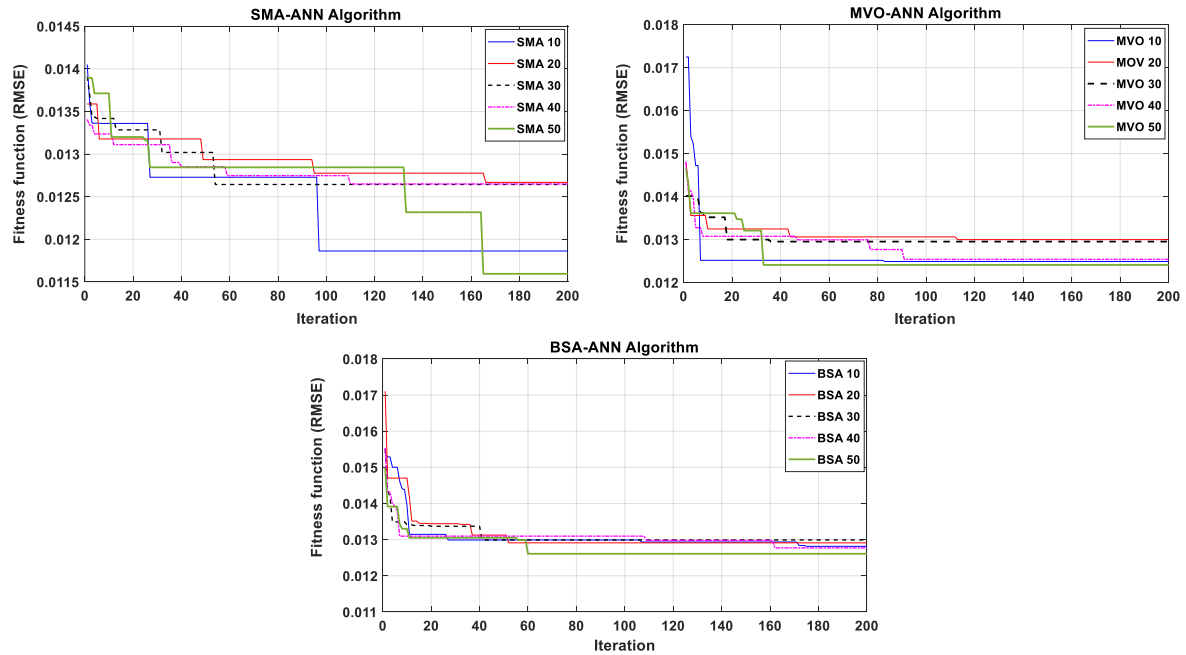


Figure 5. Boxplot of monthly stochastic components for water consumption and climatic factors.

After preparing the stochastic signals of the dependent and independent factors, data were organised into three sets (as mentioned in section 3.3): training (70%, 134 datapoints), testing (15%, 29 datapoints) and validation (15%, 29 datapoints) to build and assess the prediction model. Accordingly, this is to ensure that each dataset has data selected from along total time series (i.e., the values of the key statistical parameters, such as the maximum limit, minimum limit, and standard deviation for the mentioned datasets were very comparable).

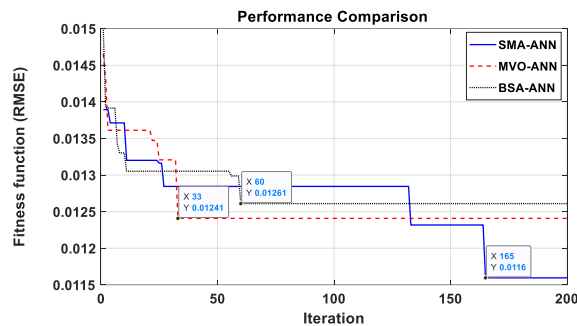
#### 4.2. Model Configuration

The ANN model needs to be integrated with the metaheuristic algorithm to determine the optimum hyperparameters of the ANN model, including LR, N1 and N2. Thus, the SMA algorithm was hybrid with the ANN model, and the results, for more validation, were compared with MVO-ANN and BSA-ANN algorithms. Each algorithm was run five times based on population sizes (10, 20, 30, 40 and 50 popsize) with 200 iterations to increase the solution range, as depicted in Figure 6. It can be noticed that popsize 50 offers the best solution for all hybrid algorithms based on the fitness function (RMSE) (i.e. offers the lowest RMSE value).



**Figure 6.** Performance of SMA-ANN, MVO-ANN and BSA-ANN algorithms.

For a closer inspection of the 50 popsize of all the three hybrid algorithms. Figure 7 shows that SMA-ANN algorithm yielded the lowest RMSE (0.0116) after 165 iterations, while the RMSE was 0.01241 (after 33 iterations) and 0.01261 (after 60 iterations) for MVO-ANN and BSA-ANN algorithm, respectively. That means that the MVO-ANN and BSA-ANN algorithms did not improve their performance, based on the value of fitness function (RMSE), along with the iteration limit. For these reasons, SMA-ANN is a better choice to determine the hyperparameters of the ANN approach comparing with MVO-ANN and BSA-ANN algorithms. So, the 50 popsize of SMA-ANN algorithm offers LR, N1 and N2 values of 0.7094, 2, and 6, respectively.



**Figure 7.** Performance comparison of the best popsize of SMA-ANN, MVO-ANN and BSA-ANN algorithms.

To scrutinise and validate the influence of integrating the ANN model with the SMA algorithm, the performance of the ANN technique was inspected. Thus, wide scenarios of the trial and error process were used to select the hyperparameters of the ANN technique. The outcomes present that the values of hyperparameters are LR= 0.6, N1= 2, and N2=3.

#### 4.3. Performance Evaluation

After determining the hyperparameters of the ANN model, the ANN model became ready to simulate the monthly stochastic signal of municipal water demand. The ANN model was implemented several times to get a better network (weights) that can accurately predict the water demand. Different

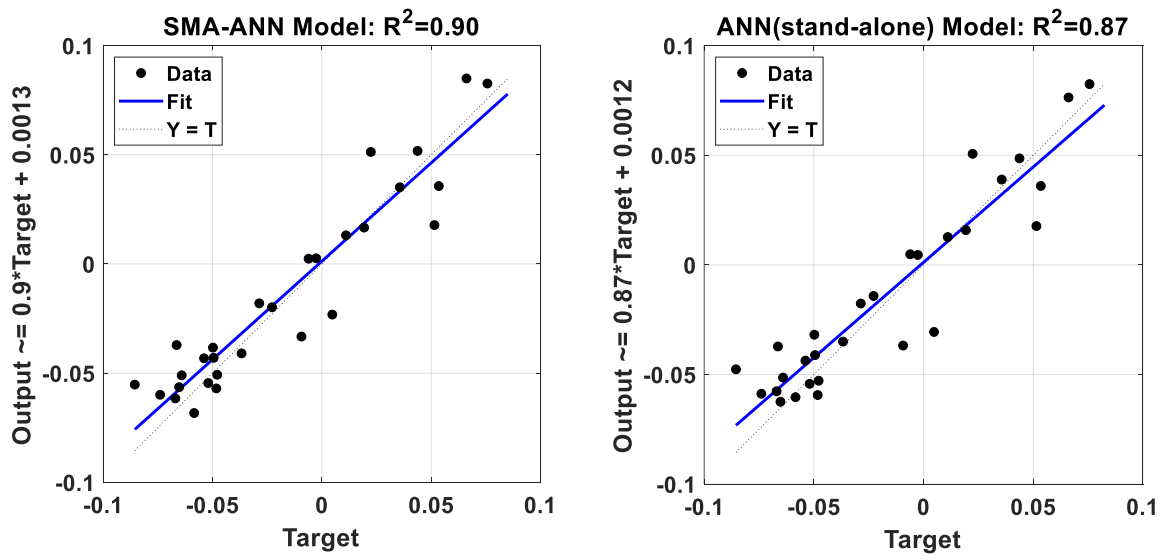
types of statistical tests were used to assess the ability of the model to forecast water demand based on climatic factors (validation stage).

Three statistical metrics were applied to examine the ability of the model to generalise data, in the validation stage, and compare the results of the SMA-ANN with the ANN model. Table 3 provides three metrics, MAE, MSE (absolute error) and MARE (relative error), to assess the non-linear dependency between the actual and simulated water demand for both models. According to Dawson, *et al.* [58], both models showed good accuracy, but the SMA-ANN can predict water demand rather well based on the MARE value.

**Table 3.** Statistical indicators of SMA-ANN and ANN models in the validation stage.

Models	MAE	RMSE	MARE
SMA-ANN	0.012	0.015	0.001
ANN	0.013	0.017	0.015

In addition, Figure 8 presents the coefficient of determination ( $R^2$ ) for SMA-ANN and the ANN models. The values of  $R^2$  delivers information for the linear relationship between the actual water consumption (Target, ML) and predicted water demand (Output, ML) for both models. Similar to the error tests (absolute and relative), both models offered good results according to Dawson, *et al.* [58]. However, the value of  $R^2$  for the SMA-ANN model was 0.9, which is more accurate than that of the ANN model (0.87). Additionally, the scatter data for the SMA-ANN model were falling closer to the ideal line than the scatter data for the ANN model.



**Figure 8.** The performance of SMA-ANN and ANN techniques in the validation stage.

Figure 9 shows the Bland-Altman for the SMA-ANN and ANN models. The SMA-ANN has a mean = -0.002998 ML with limits of the agreement being -0.03358, 0.02759 ML, while these values for the ANN model were -0.003439, -0.03536 and 0.02947 ML for the mean, lower and upper limits of agreement area, respectively. Additionally, a good agreement was noticed for the SMA-ANN model because up to 97% of the data are scattered between the limits of agreement, whereas the proportion was 90% for the ANN model. Generally, the obtained results revealed that the SMA-ANN model has limits of agreement much closer to the mean, and it has a higher agreement percentage compared with the ANN model. Additionally, there was no trend in the scattered data for both Figures 8 and 9, and the randomness of the residual data was assessed and confirmed using the ADF test. That means the tolerance method was successfully used to select the best model input.

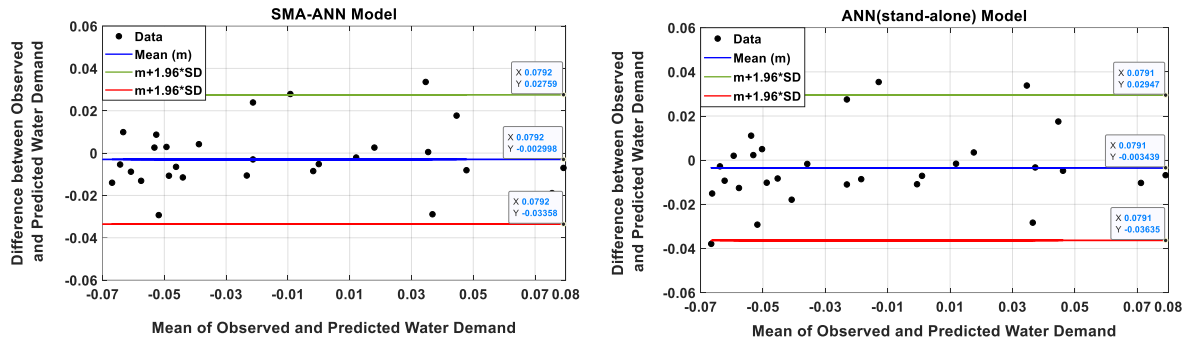


Figure 9. Bland Altman plot of SMA-ANN and ANN techniques in the validation stage.

The obtained results emphasised that the determination of the ANNs' parameters using hybridisation with SMA algorithm is better than that conducted by trial and error procedure. In the former process (i.e., SMA-ANN) the parameters are determined automatically, whereas the latter is conducted manually.

All the statistical tests examined and validated the SMA-ANN model and, for more examination, ADF and KPSST tests were used to check the stationarity of the simulated stochastic signal of water demand and residual analysis. The results show that the simulated time series of water demand is stationary and the residual data are normally distributed, which was assessed by Kolmogorov–Smirnov test. Furthermore, a graphical test was used to confirm the SMA-ANN model by comparing the observed and simulated water time series in the validation stage as shown in Figure 10, which shows that the model can closely follow the trend and cycles of the observed stochastic time series based on the scale of error. There were several slight deviations in the simulated time series that may come from the influence of fluctuation of the climatic factors. However, based on both scale of error and the result of Bland-Altman for the SMA-ANN, the error can be considered statistically insignificant.

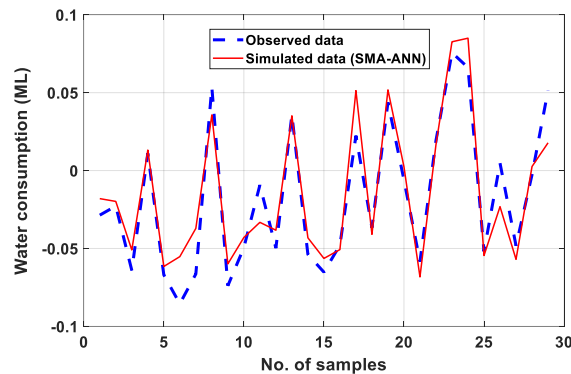


Figure 10. A comparison between observed and simulated stochastic data in the validation stage.

The most interesting conclusions that can be drawn from the above results are: (1) EMD technique has a significant rule to decompose the raw data to select the stochastic signals of dependent and independent variables. Additionally, the tolerance method was effective in determining the best scenario of independent factors. (2) The optimal hyperparameters of the ANN model were determined based on the novel hybrid model, SMA-ANN, which outperforms both MVO-ANN and BSA-ANN algorithms depending on the RMSE value. (3) The novel combined methodology that comprises data analytic and machine learning can effectively simulate the stochastic signal of water demand concerning climatic factors. (4) An automated machine learning outperforms trial and error procedure based on several statistical tests. (5) Using three metaheuristic algorithms to build the prediction model and validate the results by the ANN (stand-alone) and employing 10 climatic factors decreased the uncertainty and increased the forecasting range. (6) The hybridisation of the model, as well as the way of categorisation of the training, testing and validation samples, present a promising application of the

developed model for covering unknown extreme events, particularly when it was applied to predict data that were not used before in the model configuration. (7) The research provides important scientific insights for managers and policymakers in SEW utility to manage the water supply system under sudden increases in water demand due to the variability of the stochastic pattern of climatic factors. They can, for example, feed the model with predicting climatic factors to forecast water demand for medium-term (i.e., future). After that, they compare the water demand in the future with the capacity of the water system and decide whether the water system is capable to successfully work under the extreme events or not. (8) The obtained results confirmed the association between climate change and water demand for the medium term.

## 5. Conclusions

In this study, the potential of novel coupled data pre-processing and automated machine learning for monthly stochastic urban water demand prediction based on several climatic factors was investigated. Data for water consumption and 10 climatic factors for South East Water Company in Melbourne City were utilised for building and assessing the proposed methodology. Data pre-processing techniques were considered to analyse and select the stochastic signals of water consumption and climatic factors time series by EMD approach, and to detect the best independent variables by the tolerance method. The automated machine learning includes the ANN model, which integrated with the SMA algorithm to find optimal hyperparameters of the ANN model. The results highlighted the importance of data pre-processing to prepare the stochastic pattern of dependent and independent variables and select the best scenario of independent variables. Additionally, the SMA-ANN was found to be superior to both BSA-ANN and MVO-ANN algorithms based on RMSE as an objective function. Moreover, the performance of the hybrid model, SMA-ANN, was more accurate than the ANN (stand-alone) approach depending on different statistical tests. Furthermore, the outcomes indicated that the suggested methodology can be successfully applied in regions that suffer from climate change (i.e. drought), such as Melbourne City. Consequently, South East Water Company can take advantage of this study's findings to establish effective strategies for optimised system operation and to maintain a balance between water requested and delivered. It also helps to establish appropriate pricing plans, schedule new changes in the network and optimise the operating procedures, such as pumps, to enhance the water quality, and to reduce the uncertainty. Based on recent literature, severe weather will probably become more common in the future. Thus, there is an urgent need for more studies that use the same or different data analytic and artificial intelligence techniques to simulate the stochastic component of urban water demand based on climate factors for regions that suffer from climate change. However, the availability of reliable data for water consumption and climatic factors for the medium or long term is considered a principal limitation of this methodology.

**Author Contributions:** Formal analysis, S.Z. and K.H.; Data curation, S.G. and H.M.R.; Investigation, N.A.-B.; Resources, P.K.; Project administration, R.A.-K, Methodology, I.H.A.; Review & Editing, F. A.-Q.; M.M., Conceptualization.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Farhan, S.L.; Abdelmonem, M.G.; Nasar, Z.A. The Urban Transformation of Traditional City Centres: Holy Karbala as a Case Study. *International Journal of Architectural Research: ArchNet-IJAR* **2018**, *12*, 53-67, doi:10.26687/archnet-ijar.v12i3.1625.
- Farhan, S.L.; Jasim, I.A.; Al-Mamoori, S.K.J.J.o.U.R.; Renewal. The transformation of the city of Najaf, Iraq: Analysis, reality and future prospects. **2019**, *13*, 160-171.
- Zubaidi, S.L.; Kot, P.; Hashim, K.; Alkhaddar, R.; Abdellatif, M.; Muhsin, Y.R. Using LARS –WG model for prediction of temperature in Columbia City, USA. *IOP Conference Series: Materials Science and Engineering* **2019**, *584*, 1-9, doi:10.1088/1757-899x/584/1/012026.



4. Hashim, K.S.; Kot, P.; Zubaidi, S.L.; Alwash, R.; Al Khaddar, R.; Shaw, A.; Al-Jumeily, D.; Aljefery, M.H. Energy Efficient Electrocoagulation Using Baffle-Plates Electrodes for Efficient Escherichia Coli Removal from Wastewater. *Journal of Water Process Engineering* **2020**, *33*, 1-7, doi:10.1016/j.jwpe.2019.101079.
5. Rahim, M.S.; Nguyen, K.A.; Stewart, R.A.; Giurco, D.; Blumenstein, M. Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review. *Water* **2020**, *12*, 1-27, doi:10.3390/w12010294.
6. Hemati, A.; Rippy, M.A.; Grant, S.B.; Davis, K.; Feldman, D. Deconstructing Demand: The Anthropogenic and Climatic Drivers of Urban Water Consumption. *Environ Sci Technol* **2016**, *50*, 12557-12566, doi:10.1021/acs.est.6b02938.
7. Roy, U.; Majumder, M. *Vulnerability of Watersheds to Climate Change Assessed by Neural Network and Analytical Hierarchy Process*; 3: 2016; 10.1007/978-981-287-344-6pp. 97.
8. Zubaidi, S.L.; Ortega-Martorell, S.; Kot, P.; Alkhaddar, R.M.; Abdellatif, M.; Gharghan, S.K.; Ahmed, M.S.; Hashim, K. A Method for Predicting Long-Term Municipal Water Demands Under Climate Change. *Water Resources Management* **2020**, *34*, 1265-1279, doi:10.1007/s11269-020-02500-z.
9. Altunkaynak, A.; Nigussie, T.A. Monthly water demand prediction using wavelet transform, first-order differencing and linear detrending techniques based on multilayer perceptron models. *Urban Water Journal* **2018**, *15*, 177-181, doi:10.1080/1573062x.2018.1424219.
10. Rasifaghihi, N.; Li, S.S.; Haghighat, F. Forecast of urban water consumption under the impact of climate change. *Sustainable Cities and Society* **2020**, *52*, doi:10.1016/j.scs.2019.101848.
11. Ebrahim Banihabib, M.; Mousavi-Mirkalaei, P. Extended linear and non-linear auto-regressive models for forecasting the urban water consumption of a fast-growing city in an arid region. *Sustainable Cities and Society* **2019**, *48*, doi:10.1016/j.scs.2019.101585.
12. Pacchin, E.; Alvisi, S.; Franchini, M. A Short-Term Water Demand Forecasting Model Using a Moving Window on Previously Observed Data. *Water* **2017**, *9*, doi:10.3390/w9030172.
13. House-Peters, L.A.; Chang, H. Urban water demand modeling: Review of concepts, methods, and organizing principles. *Water Resources Research* **2011**, *47*, doi:10.1029/2010wr009624.
14. Donkor, E.A.; Mazzuchi, T.H.; Soyer, R.; Roberson, J.A. Urban water demand forecasting: review of methods and models. *Journal of Water Resources Planning and Management* **2014**, *140*, 146-159, doi:10.1061/(ASCE)WR.1943-5452.
15. Ghalekhondabi, I.; Ardjmand, E.; Young, W.A., 2nd; Weckman, G.R. Water demand forecasting: review of soft computing methods. *Environ Monit Assess* **2017**, *189*, 313, doi:10.1007/s10661-017-6030-3.
16. De Souza Groppo, G.; Costa, M.A.; Libânio, M. Predicting water demand: a review of the methods employed and future possibilities. *Water Supply* **2019**, *10.2166/ws.2019.122*, doi:10.2166/ws.2019.122.
17. Seo, Y.; Kwon, S.; Choi, Y. Short-Term Water Demand Forecasting Model Combining Variational Mode Decomposition and Extreme Learning Machine. *Hydrology* **2018**, *5*, doi:10.3390/hydrology5040054.
18. Zubaidi, S.L.; Gharghan, S.K.; Dooley, J.; Alkhaddar, R.M.; Abdellatif, M. Short-Term Urban Water Demand Prediction Considering Weather Factors. *Water Resources Management* **2018**, *32*, 4527-4542, doi:10.1007/s11269-018-2061-y.
19. Candelieri, A. Clustering and Support Vector Regression for Water Demand Forecasting and Anomaly Detection. *Water* **2017**, *9*, doi:10.3390/w9030224.
20. Firat, M.; Turan, M.E.; Yurdusev, M.A. Comparative analysis of fuzzy inference systems for water consumption time series prediction. *Journal of Hydrology* **2009**, *374*, 235-241, doi:10.1016/j.jhydrol.2009.06.013.
21. Chen, G.; Long, T.; Xiong, J.; Bai, Y. Multiple Random Forests Modelling for Urban Water Consumption Forecasting. *Water Resources Management* **2017**, *31*, 4715-4729, doi:10.1007/s11269-017-1774-7.

- 543 22. Altunkaynak, A.; Nigussie, T.A. Monthly Water Consumption Prediction Using Season  
544 Algorithm and Wavelet Transform–Based Models. *Journal of Water Resources Planning and*  
545 *Management* **2017**, *143*, doi:10.1061/(asce)wr.1943-5452.0000761.
- 546 23. Mouatadid, S.; Adamowski, J. Using extreme learning machines for short-term urban water  
547 demand forecasting. *Urban Water Journal* **2016**, *14*, 630-638,  
548 doi:10.1080/1573062x.2016.1236133.
- 549 24. Zubaidi, S.L.; Dooley, J.; Alkhaddar, R.M.; Abdellatif, M.; Al-Bugharbee, H.; Ortega-Martorell,  
550 S. A Novel approach for predicting monthly water demand by combining singular spectrum  
551 analysis with neural networks. *Journal of Hydrology* **2018**, *561*, 136-145,  
552 doi:10.1016/j.jhydrol.2018.03.047.
- 553 25. Mirjalili, S.; Mirjalili, S.M.; Hatamlou, A. Multi-Verse Optimizer: a nature-inspired algorithm for  
554 global optimization. *Neural Computing and Applications* **2015**, *27*, 495-513,  
555 doi:10.1007/s00521-015-1870-7.
- 556 26. Ullah, I.; Hussain, I.; Uthansakul, P.; Riaz, M.; Khan, M.N.; Lloret, J. Exploiting Multi-Verse  
557 Optimization and Sine-Cosine Algorithms for Energy Management in Smart Cities. *Applied*  
558 *Sciences* **2020**, *10*, doi:10.3390/app10062095.
- 559 27. Wang, X.; Pan, J.-S.; Chu, S.-C. A Parallel Multi-Verse Optimizer for Application in Multilevel  
560 Image Segmentation. *IEEE Access* **2020**, *8*, 32018-32030, doi:10.1109/access.2020.2973411.
- 561 28. Zubaidi, S.L.; Ortega-Martorell, S.; Al-Bugharbee, H.; Olier, I.; Hashim, K.S.; Gharghan, S.K.; Kot,  
562 P.; Alkhaddar, R.M. Urban Water Demand Prediction for a City That Suffers from Climate  
563 Change and Population Growth: Gauteng Province Case Study. *Water* **2020**, *12*, 1-17,  
564 doi:10.3390/w12071885.
- 565 29. Zhang, Y.; Jin, Z.; Zhao, X.; Yang, Q.J.E.C.; Management. Backtracking search algorithm with  
566 Lévy flight for estimating parameters of photovoltaic models. **2020**, *208*, 112615.
- 567 30. Mehmood, A.; Shi, P.; Raja, M.A.Z.; Zameer, A.; Chaudhary, N.I.J.N.C.; APPLICATIONS. Design  
568 of backtracking search heuristics for parameter estimation of power signals. **2020**.
- 569 31. Li, S.; Chen, H.; Wang, M.; Heidari, A.A.; Mirjalili, S. Slime mould algorithm: A new method for  
570 stochastic optimization. *Future Generation Computer Systems* **2020**, *111*, 300-323,  
571 doi:10.1016/j.future.2020.03.055.
- 572 32. Ghafil, H.N.; Jármai, K.J.A.S.C. Dynamic differential annealed optimization: New metaheuristic  
573 optimization algorithm for engineering applications. **2020**, 106392.
- 574 33. Jiao, S.; Chong, G.; Huang, C.; Hu, H.; Wang, M.; Heidari, A.A.; Chen, H.; Zhao, X.J.E.  
575 Orthogonally adapted Harris Hawk Optimization for parameter estimation of photovoltaic  
576 models. **2020**, 117804.
- 577 34. Rodríguez-Esparza, E.; Zanella-Calzada, L.A.; Oliva, D.; Heidari, A.A.; Zaldivar, D.; Pérez-  
578 Cisneros, M.; Foong, L.K.J.E.S.w.A. An Efficient Harris Hawks-inspired Image Segmentation  
579 Method. **2020**, 113428.
- 580 35. Zubaidi, S.L.; Kot, P.; Alkhaddar, R.M.; Abdellatif, M.; Al-Bugharbee, H. Short-Term Water  
581 Demand Prediction in Residential Complexes: Case Study in Columbia City, USA. In  
582 Proceedings of 2018 11th International Conference on Developments in eSystems Engineering  
583 (DeSE), Cambridge, United Kingdom, 2-5 Sept. 2018; pp. 31-35.
- 584 36. Zubaidi, S.L.; Al-Bugharbee, H.; Muhsen, Y.R.; Hashim, K.; Alkhaddar, R.M.; Hmeesh, W.H. The  
585 Prediction of Municipal Water Demand in Iraq: A Case Study of Baghdad Governorate. In  
586 Proceedings of 2019 12th International Conference on Developments in eSystems Engineering  
587 (DeSE), Kazan, Russia, 7–10 October; pp. 274-277.
- 588 37. Campisi-Pinto, S.; Adamowski, J.; Oron, G. Forecasting Urban Water Demand Via Wavelet-  
589 Denoising and Neural Network Models. Case Study: City of Syracuse, Italy. *Water Resources*  
590 *Management* **2012**, *26*, 3539-3558, doi:10.1007/s11269-012-0089-y.
- 591 38. Vecchio; Anzidei; Serpelloni; Florindo. Natural Variability and Vertical Land Motion  
592 Contributions in the Mediterranean Sea-Level Records over the Last Two Centuries and  
593 Projections for 2100. *Water* **2019**, *11*, doi:10.3390/w11071480.

39. Chu, T.-Y.; Huang, W.-C. Application of Empirical Mode Decomposition Method to Synthesize Flow Data: A Case Study of Hushan Reservoir in Taiwan. *Water* **2020**, *12*, doi:10.3390/w12040927.
40. Gedefaw, M.; Hao, W.; Denghua, Y.; Girma, A.; Khamis, M.I. Variable selection methods for water demand forecasting in Ethiopia: Case study Gondar town. *Cogent Environmental Science* **2018**, *4*, doi:10.1080/23311843.2018.1537067.
41. Haque, M.M.; Rahman, A.; Hagare, D.; Chowdhury, R.K. A Comparative Assessment of Variable Selection Methods in Urban Water Demand Forecasting. *Water* **2018**, *10*, 1-15, doi:10.3390/w10040419.
42. Zhang, X.; Qiu, J.; Leng, G.; Yang, Y.; Gao, Q.; Fan, Y.; Luo, J. The Potential Utility of Satellite Soil Moisture Retrievals for Detecting Irrigation Patterns in China. *Water* **2018**, *10*, doi:10.3390/w10111505.
43. Kim, K.; Joo, H.; Han, D.; Kim, S.; Lee, T.; Kim, H.S. On Complex Network Construction of Rain Gauge Stations Considering Nonlinearity of Observed Daily Rainfall Data. *Water* **2019**, *11*, doi:10.3390/w11081578.
44. MW. Corporate Plan 2016/17 to 2020/21. MW: Australia, 2017; p 39.
45. SEW. South East Water Annual report 2015-16. Including the Financial Report. SEW: Australia, 2016.
46. Araghinejad, S. *Data-Driven Modeling: Using MATLAB® in Water Resources and Environmental Engineering*; Springer: USA, 2014; 10.1007/978-94-007-7506-0.
47. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*, Sixth ed.; Pearson Education, Inc: United States of America, 2013.
48. Bailey, O.; Zlatanovic, L.; van der Hoek, J.P.; Kapelan, Z.; Blokker, M.; Arnot, T.; Hofman, J. A Stochastic Model to Predict Flow, Nutrient and Temperature Changes in a Sewer under Water Conservation Scenarios. *Water* **2020**, *12*, doi:10.3390/w12041187.
49. Tabrizi, A.A.; Al-Bugharbee, H.; Trendafilova, I.; Garibaldi, L.J.M. A cointegration-based monitoring method for rolling bearings working in time-varying operational conditions. **2017**, *52*, 1201-1217.
50. Abdou, A.D.; Ngom, N.F.; Sidibé, S.; Niang, O.; Thioune, A.; Ndiaye, C.H. Neural networks for biomedical signals classification based on empirical mode decomposition and principal component analysis. In *Innovation and Interdisciplinary Solutions for Underserved Areas*, Springer: 2017; pp. 267-278.
51. Vecchio, A.; Anzidei, M.; Serpelloni, E.; Florindo, F.J.W. Natural Variability and Vertical Land Motion Contributions in the Mediterranean Sea-Level Records over the Last Two Centuries and Projections for 2100. **2019**, *11*, 1480.
52. Chu, T.-Y.; Huang, W.-C.J.W. Application of Empirical Mode Decomposition Method to Synthesize Flow Data: A Case Study of Hushan Reservoir in Taiwan. **2020**, *12*, 927.
53. Zeiler, A.; Faltermeier, R.; Keck, I.R.; Tomé, A.M.; Puntonet, C.G.; Lang, E.W. Empirical mode decomposition-an introduction. In *Proceedings of The 2010 International Joint Conference on Neural Networks (IJCNN)*; pp. 1-8.
54. Pallant, J. *SPSS Survival Manual: A step by step guide to data analysis using IBM SPSS*; Open University Press/McGraw-Hill: 2016.
55. Bayatvarkeshi, M.; Mohammadi, K.; Kisi, O.; Fasihi, R. A new wavelet conjunction approach for estimation of relative humidity: wavelet principal component analysis combined with ANN. *Neural Computing and Applications* **2018**, 10.1007/s00521-018-3916-0, doi:10.1007/s00521-018-3916-0.
56. Zubaidi, S.L.; Al-Bugharbee, H.; Ortega-Martorell, S.; Gharghan, S.K.; Olier, I.; Hashim, K.S.; Al-Bdairi, N.S.S.; Kot, P. A Novel Methodology for Prediction Urban Water Demand by Wavelet Denoising and Adaptive Neuro-Fuzzy Inference System Approach. *Water* **2020**, *12*, 1-17, doi:10.3390/w12061628.

- 644 57. González Perea, R.; Camacho Poyato, E.; Montesinos, P.; Rodríguez Díaz, J.A. Optimisation of  
645 water demand forecasting by artificial intelligence with short data sets. *Biosystems*  
646 *Engineering* **2019**, *177*, 59-66, doi:10.1016/j.biosystemseng.2018.03.011.
- 647 58. Dawson, C.W.; Abrahart, R.J.; See, L.M. HydroTest: A web-based toolbox of evaluation metrics  
648 for the standardised assessment of hydrological forecasts. *Environmental Modelling &*  
649 *Software* **2007**, *22*, 1034-1052, doi:10.1016/j.envsoft.2006.06.008.

650