Evolutionary modelling of municipal water demand with multiple feature selection techniques

Oluwaseun Oyebode

ABSTRACT

This paper presents the development of an artificial intelligent water demand forecasting model. The model comprises a single hidden-layer feed-forward neural network trained in using a differential evolution algorithm. Multiple feature selection techniques were employed to identify the minimal subset of features for optimal learning, namely Pearson correlation, information gain, symmetrical uncertainty, Relief-F attribute and principal component analysis. The performance of the feature selection techniques was compared to a baseline scenario comprising a full set of data covering potential casual variables including weather, socioeconomic and historical water consumption data. The performance of the models was evaluated based on accuracy. Results show that the five feature selection techniques outperformed the baseline scenario. More importantly, the subset of features obtained from the Pearson correlation technique produced the most superior model in terms of model accuracy. Findings from the study suggest that the inclusion of weather and socioeconomic variables in water demand modelling could enhance the accuracy of forecasts and cater for the impacts of climate and socioeconomic variations in water demand planning and management.

Key words | artificial neural network, differential evolution, feature selection, water demand forecasting

Oluwaseun Oyebode
Centre for Research in Environmental, Coastal and Hydrological Engineering (CRECHE), Department of Civil Engineering,
University of KwaZulu-Natal,
Durban 4014,
South Africa
E-mail: *oluwaseun.oyebode@gmail.com*

INTRODUCTION

Water demand forecasting is of crucial importance in water resource planning and management as it is a prerequisite for optimal allocation of available water resources (Qu *et al.* 2010). City managers and water utilities often rely on water demand forecasts to guide their decision-making on infrastructure investments as well as the scheduling and operation of water distribution systems. For example, long-term forecasts are imperative in providing new water supplies and upgrading the capacity of existing water treatment plants while short-term forecasts guide day-to-day operation of treatment plants and reservoirs to meet daily demands. Accurate water demand forecasts are therefore required for both short- and long-term infrastructure planning, operation and coordination. Moreover, the importance of water demand forecasting in realizing the sustainable development goals (SDGs) has been stressed by the United Nations (UN). The UN in its 2015 water development report called for improvement of water demand models as competing demands may lead to increasingly difficult allocation decisions and restrict the growth of sectors critical to sustainable development (UNESCO 2015). This implies that, amidst growing demands for fresh water across the globe, sustainable development can only be achieved if competing sources of demand are well defined to enable restoration of the balance between demand and supply. Water demand forecasting therefore provides useful information for promoting a more economical use of water resources and ensuring the sustainability of water distribution systems in the short, medium and long terms.

Accurate water demand prediction is reliant on the explanatory variables adopted in model development. Research has shown that model accuracy is a function of

doi: 10.2166/aqua.2019.145

the impacts of each explanatory variable (Firat *et al.* 2009; Babel & Shinde 2011; Toth *et al.* 2018). However, many city managers, consultants and water utilities still assume that water demand will evolve simply as a function of per-capita demand and a prognosis of population, although the predictive power of such approaches is deficient under changing conditions (Toth *et al.* 2018). Babel & Shinde (2011) argue on the need to develop improved and city specific water demand models as water demand is influenced not only by population but also by various weather and socioeconomic variables as well as government policies and strategy-related factors, which are often location-based. Therefore, the need to carefully define, evaluate, understand and model the explanatory variables that directly and indirectly influence water demand is now acknowledged as crucial in obtaining accurate demand forecast. Table 1 presents a list of some relevant studies that have considered weather and socioeconomic variables in water demand forecasting.

Over the past few decades, many techniques have been used in forecasting water demand. These techniques mainly include traditional forecasting techniques such as multivariate regression and time series analysis (Babel *et al.* 2007), system dynamics modelling (Qi & Chang 2011), and more recently, advanced computational intelligence techniques like expert systems or agent-based models (Ali *et al.* 2017) and artificial neural networks (ANNs; Bennett *et al.* 2013). The application of ANN in water demand forecasting is becoming increasingly popular due to its superiority over traditional techniques and its ability to account for nonlinear patterns observed in real problems (Babel & Shinde 2011; Kofinas *et al.* 2014). ANN is capable of learning and analysing data attributes and, thereafter, implementing nonlinear approximation function without any initial assumption on the physics of the system being modelled or its data distributions (Faizollahzadeh Ardabili *et al.* 2018). As a result, ANN is now being adopted as an alternative to the traditional methods which are limited due to their linear pre-assumption of the form of the model (Kofinas *et al.* 2014). A review of the capability, implementation and application of ANN in water resources modelling including water demand forecasting is available in Ghalehkhondabi *et al.* (2017) and Oyebode & Stretch (2018a).

Water demand forecasting using ANN is characterized by some complexities. According to Kofinas *et al.* (2014), these complexities can be summarized as: (i) inability to adequately extrapolate outside the range of primary (training) data; (ii) diminishing forecast accuracy when lagged values of target variable are used as input; and (iii) disregarding the impacts of other explanatory variables affecting water demand due to the high correlation between future water demand and its historical values. In a comprehensive review of techniques used in forecasting water demand, Oyebode & Stretch (2018b) noted the non-inclusion of weather-based variables as inputs in most of the studies reviewed. It was further argued that, due to the non-inclusion of weather-based variables, most studies in the literature lack a climate variability perspective to water demand modelling. This jeopardizes the opportunity to put in place effective early warning systems and to implement adaptive interventions to deal with variations in water availability and the occurrence of extreme climate-linked events. The inclusion of climate-based parameters is likely to enhance the outcome of existing water demand forecasting models.

This study aims to suggest possible ways of addressing the limitations of ANN in water demand forecasting. First,

Table 1 | Explanatory variables considered in some relevant previous studies

Sl. no.	Author and year	Explanatory variables
1.	Kim *et al.* (1999)	HWD, T, SD
2.	Firat *et al.* (2009)	HWD, R, RH, T, P, E
3.	Pulido-Calvo & Gutierrez-Estrada (2009)	HWD, T
4.	Wu & Yan (2010)	HWD, T, RH
5.	Babel & Shinde (2011)	HWD, T, R, RH, E
6.	Perea *et al.* (2015)	HWD, T, SR, ET_R
7.	Fagiani *et al.* (2015)	HWD, T
8.	Shabani *et al.* (2016)	HWD, T, RH, R, E
9.	Papageorgiou *et al.* (2016)	HWD, T, R, P
10.	Yousefi *et al.* (2017)	HWD, T, R
11.	Toth *et al.* (2018)	HWD, T, R, P, TR, WP

E, economic-based; HWD, historical water demand; P, population; T, temperature; R, rainfall; RH, relative humidity; SD, sunshine duration; SR, solar radiation; ET_R, reference evapotranspiration; TR, tourist-based; WP, water price.

it allows for the enhancement of demand forecasting models proposed in earlier works, by integrating new explanatory variables related to climate and socioeconomic variations. The study explores the capabilities of five feature selection techniques in providing the optimal set of explanatory variables required for accurate prediction of water demand. Furthermore, it investigates the ability of an evolutionary-based technique – differential evolution (DE) – in evolving an ANN model with optimal model complexity and accuracy. Research has shown that despite the prominence and successful applications of evolutionary algorithms in water resources (Maier *et al.* 2014), DE is yet to be fully explored in water demand forecasting (Oyebode & Stretch 2018a, 2018b). In this study, the ability of DE in training a multilayer feed-forward neural network is explored and, by doing so, a water demand forecasting model with optimal complexity and accuracy is developed.

METHODOLOGY

This section presents a brief background on the modelling technique, training algorithm and feature selection techniques applied in this study.

Artificial neural networks

ANN is a computational intelligence technique inspired by the configuration and working principles of the human brain (Šiljić Tomić *et al.* 2018). The ANN architecture comprises a collection of processors (neurons), typically arranged in three layers which collect, interpret, and exchange information over a framework of weighted connections (Oyebode & Stretch 2018a). ANN is popularly used for mapping an input–output relationship for a given system by combining the input information and estimating their weights. The connection weights are a product of the impact of each input on the processor, and a threshold value (known as bias) must be exceeded for a processor to be triggered. Each processor returns an output based on the weighted sum of all inputs collected and according to a nonlinear activation function. ANN thus undergoes a learning process by adjusting the weights iteratively between its processors and comparing the resulting error between actual and modelled values (Shahin *et al.* 2008). Given a sigmoidal activation function, the relationship between inputs and output(s) is expressed as follows:

$$\boldsymbol{P} = \boldsymbol{1}/[\boldsymbol{1} + \boldsymbol{e}^{\boldsymbol{s}}] \tag{1}$$

$$\boldsymbol{s} = (\boldsymbol{a}_1\boldsymbol{w}_1 + \boldsymbol{a}_2\boldsymbol{w}_2 + \ \ldots) + \boldsymbol{B} \tag{2}$$

where $\boldsymbol{P}$ is the output of each node, $\boldsymbol{a_i}$ is the input value, $\boldsymbol{w_i}$ is the weight, and $\boldsymbol{B}$ is the bias. The key objective of ANN training is to reduce the overall error $\boldsymbol{E}$ between the outputs and actual observations by adjusting the weights. The overall error, $\boldsymbol{E}$, can be mathematically expressed as follows (Mafi & Amirinia 2017):

$$\boldsymbol{E} = \frac{\boldsymbol{1}}{\boldsymbol{m}}\sum \boldsymbol{E_m} \tag{3}$$

where $\boldsymbol{m}$ is the total number of training patterns and $\boldsymbol{E_m}$ can be expressed as follows:

$$\boldsymbol{E_m} = \frac{\boldsymbol{1}}{\boldsymbol{2}}\sum (\boldsymbol{O_n} - \boldsymbol{P_n})^2 \tag{4}$$

where $\boldsymbol{O_n}$ and $\boldsymbol{P_n}$ are actual and predicted values for $\boldsymbol{n}$th output processor, respectively. To be concise, details on the configuration of ANN and its implementation are not presented in this study; however, they are available in the literature (Shahin *et al.* 2008; Oyebode & Stretch 2018a).

The study investigates the ability of ANN to forecast monthly water demand considering the nonlinear and dynamic nature of input variables based on climate and socioeconomic factors.

DE training algorithm

DE is a population-based heuristic algorithm for global optimization over continuous spaces. Thus, it can find the optimal weights required for error minimization in ANNs (Ilonen *et al.* 2003). According to Piotrowski (2014), the classic DE algorithm evolves a population of $\boldsymbol{NP}$ individuals, $\boldsymbol{x_{i,g}} = \{\boldsymbol{x_{i,g}^1}, \ \ldots, \ \boldsymbol{x_{i,g}^D}\}, \ \boldsymbol{i} = 1, \ \ldots \boldsymbol{NP}$ during successive generations $\boldsymbol{g}$ to obtain the global optimum of a function $\boldsymbol{f}$ in a subset $\prod_{\boldsymbol{j}=1}^{\boldsymbol{D}} [\boldsymbol{L^j}, \ \boldsymbol{U^j}]$ within a decision domain $\boldsymbol{R^D}$.

A preliminary location of individuals is randomly initiated from a uniform distribution expressed as follows:

$$x_{i,0}^{j} = L^{j} + rand_{i}^{j}(0,\ 1) \cdot (U^{j} - L^{j});\quad j = 1,\ \ldots,\ D;\quad i = 1,\ \ldots,\ NP \tag{5}$$

where $rand_{i}^{j}(0,\ 1)$ creates an arbitrary value within the range [0,1] for every component of each individual.

In newer generations, each parent individual ($x_{i,g}$) generates an offspring ($u_{i,g}$) using a dual-staged approach. The initial stage involves creating a donor vector ($v_{i,g}$) via mutation. In the second stage, a crossover operation is executed between the donor and parent vectors, resulting in an offspring. A parent and an offspring are subjected to a competition-based selection process (greedy selection) and only the superior proceeds to the succeeding generation.

$DE/rand/1$ mutation strategy with a scaling factor F used in implementing the classic DE can be expressed as follows:

$$v_{i,g} = x_{r1,g} + F \cdot (x_{r2,g} - x_{r3,g}) \tag{6}$$

where r_1, r_2, and r_3 are randomly chosen integers within the interval [1, NP], such that $r_1 \neq r_2 \neq r_3 \neq i$, $x_{best,g}$ signifies the most superior individual in the present population at generation g.

A binomial crossover operation is executed on the parent and target vectors after mutation, producing an offspring ($u_{i,g}$) and, consequently, requiring the value of a crossover control parameter (CR) to be defined, whence:

$$u_{i,g}^{j} = \begin{cases} v_{i,g}^{j} & \text{if } rand_{i}^{j}(0,\ 1) \leq CR \quad or \quad j = j_{rand,i} \\ x_{i,g}^{j} & \text{otherwise} \end{cases} \tag{7}$$

The CR values are typically defined within the [0, 1] interval. $j_{rand,i}$ is a randomly chosen integer within the [1, D] interval, as an assurance that an offspring acquires a minimum of one element from a donor vector. Ultimately, the greedy selection between the parent and the offspring is expressed by the following equation:

$$x_{i,g+1}^{j} = \begin{cases} u_{i,g} & \text{if } f(u_{i,g}) \leq f(x_{i,g}) \\ x_{i,g} & \text{otherwise} \end{cases} \tag{8}$$

The DE algorithm typically proceeds with the exploration until a predefined number of iterations is attained. DE is employed to optimize the architecture (complexity) and network parameters of ANN models developed in this study, thus pioneering the application of DE in training multilayer feed-forward ANN models in water demand forecasting.

Feature selection

To develop a model with high degree of accuracy and minimal complexity, it is important to select only a small number of variables with significant predictive features. Research has shown that the inclusion of irrelevant or redundant or noisy variables could increase model complexity, reduce model interpretability, heighten computational demands, and consequently lead to non-convergence (Bowden *et al.* 2005). Feature selection has been widely reported as a technique that can be beneficial to learning as it seeks to identify and possibly remove all the irrelevant and redundant attributes, thereby reducing the dimensionality of the data and the size of the hypothesis space accordingly (Hall 1999; Oyebode 2014). Feature selection achieves this aim by finding a minimum set of variables, such that the resulting probability distribution of the data classes is, to a great extent, close to the original distribution obtained using all variables (Azhagusundari & Thanamani 2013). Feature selection thus enables learning algorithms to execute faster and more effectively. The techniques utilized in evaluating the worth of features (variables) used in this study are briefly described in turn.

Pearson correlation

Pearson correlation belongs to the class of 'filter' feature selection techniques which are founded on data pre-processing to isolate the features $X_1, \ldots, X_p$ that most impact the target Y. Pearson correlation provides a straightforward approach to filter features based on their correlation coefficient. The Pearson correlation coefficient between a feature X_i and the target Y is expressed as follows:

$$\rho_i = \frac{cov(X_i,\ Y)}{\sigma(X_i)\sigma_Y} \tag{9}$$

where $cov(X_i, Y)$ represents the covariance, and σ the standard deviation (Mangal & Holm 2018). The coefficient is typically bounded within the interval [−1, 1] and applicable to regression and numerical classification problems. The Pearson correlation thus serves as a quick criterion for ranking features according to the absolute correlation coefficient to the target.

Information gain

Information gain is a symmetric-based index used to rank features. The index computes the number of bits of information gained by an independent variable about a target variable (Karimi *et al.* 2013). Given the entropy is a function of impurity in a training set S, an index, IG, denoting additional information about Y as provided by X can be defined, representing the amount by which the entropy of Y decreases. This index is mathematically expressed as follows:

$$IG = H(Y) - H(Y\backslash X) = H(X) - H(X\backslash Y) \tag{10}$$

Information gain is thus founded on the premise that the information gained about the target variable Y after observing an independent variable X is equal to the information gained about X after observing Y. The limitation in using information gain is in its bias towards features with more values even when they are not more informative (Phyu & Oo 2016).

Symmetrical uncertainty

Symmetrical uncertainty is a feature selection system that operates based on the principle of mutual information. Symmetrical uncertainty measures the correlation, SU, between the features and the target class using the following expression (Karimi *et al.* 2013):

$$SU = (H(X) + H(Y) - H(X\backslash Y))/(H(X) + H(Y)) \tag{11}$$

where $H(X)$ and $H(Y)$ are the entropies according to the probability associated with each feature and class value, respectively, and $H(X, Y)$, the mutual probabilities of all combinations of values of X and Y.

Relief-F attribute

Relief-F attribute is a feature selection technique for detecting conditional dependencies between data attributes and providing an integrated assessment on the attribute estimation in regression and classification-based problems (Robnik-Šikonja & Kononenko 2003). It seeks to draw instances at random, calculate their nearest neighbours, fine-tune a feature weighting vector, and consequently, award additional weight to features that discriminate the instance from neighbours of different classes (Phyu & Oo 2016). Mathematically, Relief-F attribute attempts to assign a weight for each feature f using a probabilistic estimate expressed as follows:

$$\begin{aligned} w_f &= P(\text{different value of } f/\text{different class}) \\ &\quad - P(\text{different value of } f/\text{same class}) \end{aligned} \tag{12}$$

Principal component analysis

Principal component analysis is a feature selection which seeks to identify linear combinations of unique explanatory variables (referred to as principal components – PCs) capable of summarizing the data, with the aim of retaining maximum information during the process. Principal component analysis operates by transforming a given set of variables orthogonally such that the transformed variables are uncorrelated and independent of each other, especially if the initial variables are normally distributed (Hu *et al.* 2007).

Given a data set of G variables X on every n individuals, $X = (x_1, x_2, \ldots, x_G)$ such as water consumption-explanatory variables, the aim is to find a new set of variables $\xi = (\xi_1, \xi_2, \ldots, \xi_G)$, that are linearly related to the X's but are themselves uncorrelated with a declining variance from most significant to least significant:

$$\xi_i = \alpha_{i1}x_1 + \alpha_{i2}x_2 + \ldots + \alpha_{ij}x_j + \ldots + \alpha_{iG}x_G \tag{13}$$

To apply a condition that the modification is self-orthogonal, the requisite constraints are expressed as follows:

$$\sum_{i=1}^{G} \alpha_{ij}\alpha_{ik} = 0 \qquad j \neq k \tag{14}$$

$$\sum_{i=1}^{G} \alpha_{ij}\alpha_{ik} = 1 \qquad j = k \tag{15}$$

The set-up of PCs depends on the magnitude of importance. In particular, the first PC should provide the most important relation between the original variables based on the largest variance, while the second PC should give the second most important relation and is orthogonal to the first PC, etc. The variances of the succeeding PCs would be smaller if high correlation between the original variables occurs. Consequently, principal component analysis can provide guidance in reducing the number of potential explanatory variables and offer the best representation in a lower number of transformed PCs (Hu *et al.* 2007).

Study area and data description

This research focused on the City of Ekurhuleni, a metropolitan municipality, located in the Gauteng province of South Africa – the most populous province in South Africa with a population of approximately 14.7 million people (Stats-SA 2018). The City of Ekurhuleni was established in the year 2000 from the amalgamation of two existing regional entities, namely Kyalami Metropolitan and the Eastern Gauteng Services Council, thereby agglomerating a set of nine relatively small and fragmented towns (Figure 1). The City of Ekurhuleni currently accounts for about 26% of Gauteng's population and plays a dominant role in the national economy, contributing 8.8% to South Africa's Gross Added Value as of 2016 (IDP 2018). The city also has a Human Development Index (HDI) of 0.704, greater than the National value of about 0.653. However, it is at the epicentre of migration, resulting in increased pressure on limited water resources (IDP 2018). The Gauteng province has already used up its available water resources and is now importing more expensive water, via bulk purchase, from the Lesotho Highlands transfer scheme and feeding it into the Vaal dam (IDP 2018). Table 2 provides current figures relevant to Ekurhuleni Water Infrastructure.

The city management seeks to ensure that Ekurhuleni transitions from being a fragmented city to being a 'Delivering City' from 2012 to 2020, a 'Capable City' from 2020 to 2030, and lastly a 'Sustainable City' from 2030 to 2055 (IDP 2018). To achieve these milestones, a long-term development strategy referred to as the Ekurhuleni Growth and Development Strategy 2055 (GDS 2055) has been developed to systematically analyse Ekurhuleni's history and its

Figure 1 | Overview of Ekurhuleni Metropolitan Municipality Service Area.

development challenges, wherein it therefore outlines the desired growth and development trajectory. One of the strategic objectives and the key focus areas and/interventions is the promotion of urban integration and continued investment in water infrastructure to ensure security of supply. This is critical to attaining the state of a 'Sustainable City' and realizing the African Union Agenda 2063 and 2030 UN SDGs which include access to clean water and

Table 2 | Ekurhuleni water infrastructure figures

Ekurhuleni water infrastructure	Data
Average water demand (ml/annum)	365,000
Water resources/supply	Vaal dam
Number of reservoirs	73
Number of towers	32
Number of bulk connections	186
Pipes (km)	11,448
Number of distribution zones	124
Population (million, 2016)	3.5
Annual population growth	2.51%

Source: Gubuza (2017).

sanitation, innovation and infrastructure as well as reduced inequality. A reliable water demand forecasting model which considers not only population but also other factors related to the weather and socioeconomic profile of the city is therefore of vital importance. This would assist in the planning and management of the city's water resources, thereby fostering the achievement of its objectives.

One of the initial steps in water demand forecasting is identifying explanatory variables that directly and indirectly influence water demand. The identification of explanatory variables forms the basis upon which final input parameters are selected for model development. Details on key explanatory variables to be considered in water demand forecasting are available in Oyebode & Stretch (2018b). Based on availability, the explanatory variables considered in this study include monthly total rainfall (***R***), monthly average minimum and maximum temperatures ($\boldsymbol{T_{min}}$ and $\boldsymbol{T_{max}}$), monthly average relative humidity (***RH***), monthly average wind speed (***WS***), number of household connections (***HH***), population (***P***), human development index (***HDI***) and water consumption (***WC***). Although the weather-based variables usually employed in water demand forecasting studies are temperature and rainfall, considering the semi-arid characteristics of the City of Ekurhuleni (and generally, South Africa), RH and wind speed were included as potential explanatory variables as they could influence outdoor water consumption (Huntra & Keener 2017). Moreover, previous studies conducted in areas with similar characteristics have reported the inclusion of these parameters in forecasting water demand (Firat *et al.* 2009; Babel & Shinde 2011; Huntra & Keener 2017).

Monthly data records for each variable were obtained from relevant government departments (South African Weather Service, Statistics South Africa (Stats SA), and the City of Ekurhuleni) for the period August 2010 to March 2018. Water consumption data were based on total monthly billed (revenue) water consumption of water users. Weather information was supplied from a

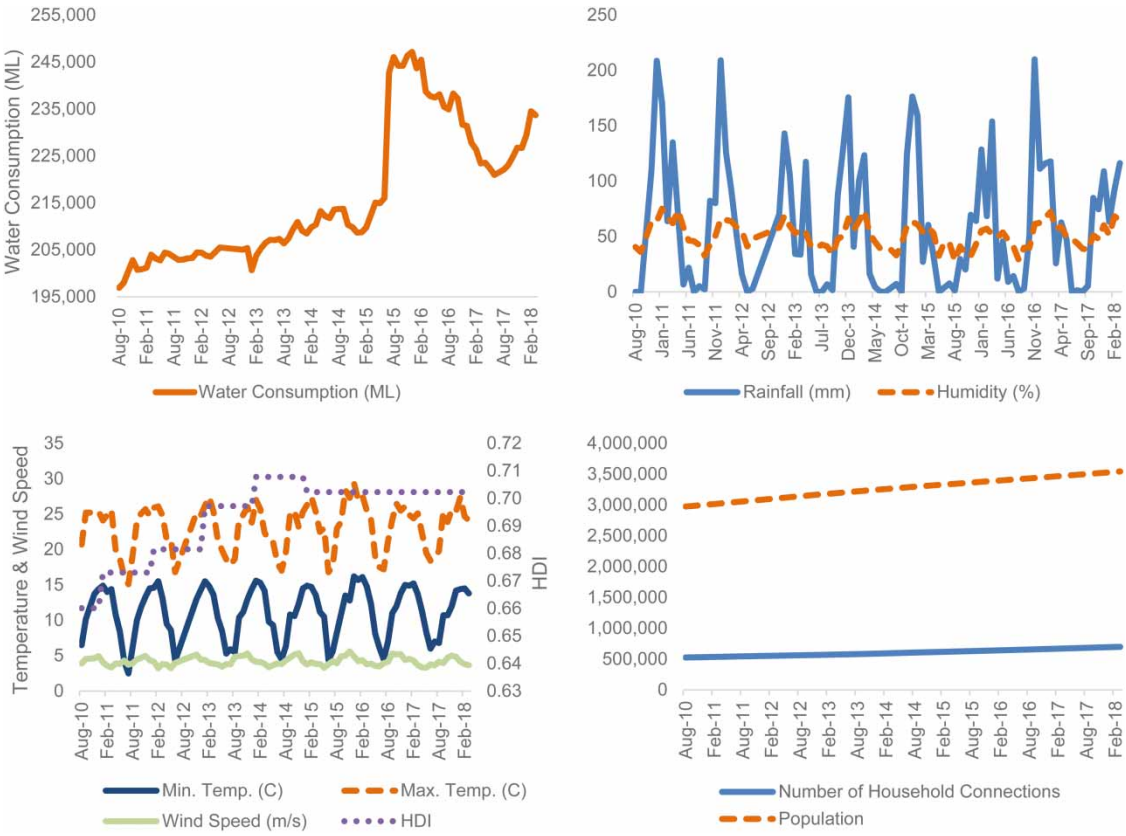

Figure 2 | Historical trend of variables considered for model development.

Table 3 | Descriptive statistics of data used in the study

Statistical parameter	R (mm)	T_{min} (°C)	T_{max} (°C)	RH (%)	WS (m/s)	HH	P	HDI	WC (ml)
Mean	59.37	11.11	23.15	51.10	4.19	607,096	3,280,134	0.69	216,917
Maximum	210.00	16.20	29.40	75.07	5.60	698,407	3,543,077	0.71	247,135
Minimum	–	2.50	15.10	28.07	3.23	526,700	2,975,216	0.66	196,908
Standard deviation	59.42	3.65	3.38	11.56	0.58	50,953	165,046	0.01	14,524
Kurtosis coefficient	−0.30	−0.91	−0.72	−0.87	−0.64	−1.21	−1.12	0.01	−0.86
Skewness coefficient	0.81	−0.55	−0.60	0.07	0.48	0.12	−0.21	−1.15	0.70

representative weather station located in the OR Tambo International Airport. The number of household connections provides an indication of the number of dwelling units served by the authority while population represents the total number of people domiciled in the city. The HDI is a measure of the city's overall achievement in its socioeconomic dimensions including life expectancy, education and income levels. Yearly population and HDI data were transformed into monthly values for use in this study by linear interpolation.

The statistical properties and historical trends of the data collected in this study are presented in Figure 2 and Table 3, respectively. A surge in water consumption and huge variations thereafter can be observed between mid-2015 and March 2018. As clarified by the city's water services planning manager, in mid-2015, the city management implemented one of its strategic objectives which aimed at reducing water loss within the city's water distribution system. The implementation of this strategy entailed the installation of new water meters and repair of faulty water pipes to address high water losses which were mainly due to leaks, theft and metering inaccuracies. As a result, water initially categorized as non-revenue water (i.e. real and apparent water losses) was transformed into revenue water.

MODEL DEVELOPMENT

To identify feature subsets that can describe the water consumption data of the City of Ekurhuleni as good or better than the primary data set, the five feature selection techniques described above were investigated. The feature selection algorithms were implemented by means of a Ranker search method (Witten *et al.* 2016). The features selected by each of the techniques are presented in Table 4.

Table 4 | Functional relationship between water consumption and features selected

Feature selection technique	Functional relationship of features selected	
Pearson correlation	$WC = f(HH, P, HDI, WS)$	(16)
Information gain	$WC = f(HH, P, T_{max}, T_{min}, RH)$	(17)
Symmetrical uncertainty	$WC = f(HH, P, T_{max}, T_{min})$	(18)
Relief-F attribute	$WC = f(HH, P, R, T_{min}, HDI)$	(19)
Principal component analysis	$WC = f(HH, P, HDI, RH, T_{max})$	(20)

The functional relationship between water consumption and the original data set (i.e. all the potential explanatory variables) is expressed below. This is henceforth referred to as 'baseline scenario'.

$$WC = f(R, T_{min}, T_{max}, RH, WS, HH, P, HDI) \quad (21)$$

The data sets were split into two subsets of similar statistical properties with 70% of the data (61 instances) used for model training and the outstanding 30% (26 instances) for validation.

To investigate the performance of the feature selection techniques, a multilayer feed-forward ANN comprising three layers such as one input, one hidden and one output layer was developed. The feature subsets produced by each of the feature selection techniques were used as model inputs in turn. The baseline scenario was also implemented on the ANN. The optimal architecture of the models was established

by incrementally changing the number of hidden layer processors from 1 to 10 using a single stepping function. The output layer consists of only one neuron, representing the target variable, water consumption, while a logistic sigmoidal-type activation function within [0, 1] interval was utilized in the hidden layer to rescale the inputs in the range [0.1, 0.9]. A linear activation function was used in the output layer to transform nonlinearities in the inputs into a linear space.

The ANN was trained using a classic DE algorithm (Storn & Price 1997). The crossover probability, ***CR***, and mutation scale factor, ***F***, were used to govern the genetic operations during the algorithm run. Following the suggestion of Montgomery & Chen (2010), ***NP*** was set at '***D*** multiplied by 10', where ***D*** is the number of weights and biases in the selected architecture. Adopting a stepping value of 0.1, the DE algorithm was subjected to sensitivity analysis by varying ***CR*** and ***F*** incrementally within [0.5, 0.9] and [0.1, 0.5] intervals, respectively. This was aimed at determining the optimal parameter settings to govern the evolution process. The algorithm was thereafter run for 1,000 generations for each of the models. Early stopping (Raskutti *et al.* 2014) was integrated in the ANN models to address overfitting problems. Early stopping aims to identify the point where minimum error on the validation data set begins to rise and then halts training to prevent overfitting. Early stopping thus ensures that model performance balances model complexity with the errors observed during training and validation.

The methodological framework developed and implemented for this study is depicted in Figure 3.

MODEL EVALUATION

To evaluate the predictive capabilities of the models developed using the baseline scenario and feature selection techniques, three statistical measures were applied, namely root-mean-square error (RMSE), Nash–Sutcliffe efficiency index (NSE) and coefficient of determination (R^2). The RMSE is a measurement of the error variance in the model prediction, while the NSE scores the error variance within the interval $[-\infty; 1]$ (Amaranto *et al.* 2018). R^2 measures the degree of collinearity between observed values and predicted values, thereby defining the proportion of variance in observed values as explained by the models. Both NSE and R^2 indicate a better model as their value approaches 1. The mathematical expression for the three statistical measures is expressed as follows:

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\boldsymbol{O_i} - \boldsymbol{P_i})^2}{\boldsymbol{N}}} \tag{22}$$

$$\boldsymbol{NSE} = 1 - \frac{\sum_{i=1}^{N} (\boldsymbol{O_i} - \boldsymbol{P_i})^2}{\sum_{i=1}^{N} (\overline{\boldsymbol{O_i}} - \boldsymbol{O_i})^2} \tag{23}$$

$$\boldsymbol{R^2} = \left[\frac{\sum_{i=1}^{N} (\boldsymbol{O_i} - \overline{\boldsymbol{O}})(\boldsymbol{P_i} - \tilde{\boldsymbol{P}})}{\sqrt{\sum_{i=1}^{N} (\boldsymbol{O_i} - \overline{\boldsymbol{O}})^2 \cdot \sum_{i=1}^{N} (\boldsymbol{P_i} - \tilde{\boldsymbol{P}})^2}} \right]^2 \tag{24}$$

where $\boldsymbol{N}$ is the number of instances in the set, and $\boldsymbol{P_i}$, $\boldsymbol{O_i}$, $\tilde{\boldsymbol{P}}$ and $\overline{\boldsymbol{O}}$ are the predicted and observed values, and their respective average values.

RESULTS AND DISCUSSION

The performance of the ANN models developed in the study was evaluated based on learning accuracy and model complexity, and the performance evaluation results are presented in Tables 5 and 6. Table 5 compares the performance of the ANN models in reproducing the actual water consumption at the City of Ekurhuleni, while Table 6 presents the optimal model architectures, optimal DE control parameters and ranks for each of the ANN models. The results show a highly competitive performance amongst the techniques employed in this study, with minimal errors (RMSEs) observed in all the ANN models. All the models were ranked based on their average performance across the three statistical measures and over the validation data sets. Overall, the ANN model developed using the Pearson correlation subset performed better than other techniques, producing the lowest error (RMSE) estimate of 4,172 mL. Similarly, the Pearson correlation-based ANN model produced the highest R^2 and NSE values at 0.9233 and 0.9001, respectively.

The ANN model developed using the Relief-F attribute technique produced the second-best performance, while those developed using the information gain, principal component analysis and symmetrical uncertainty came third, fourth and fifth, respectively. It is interesting to note that

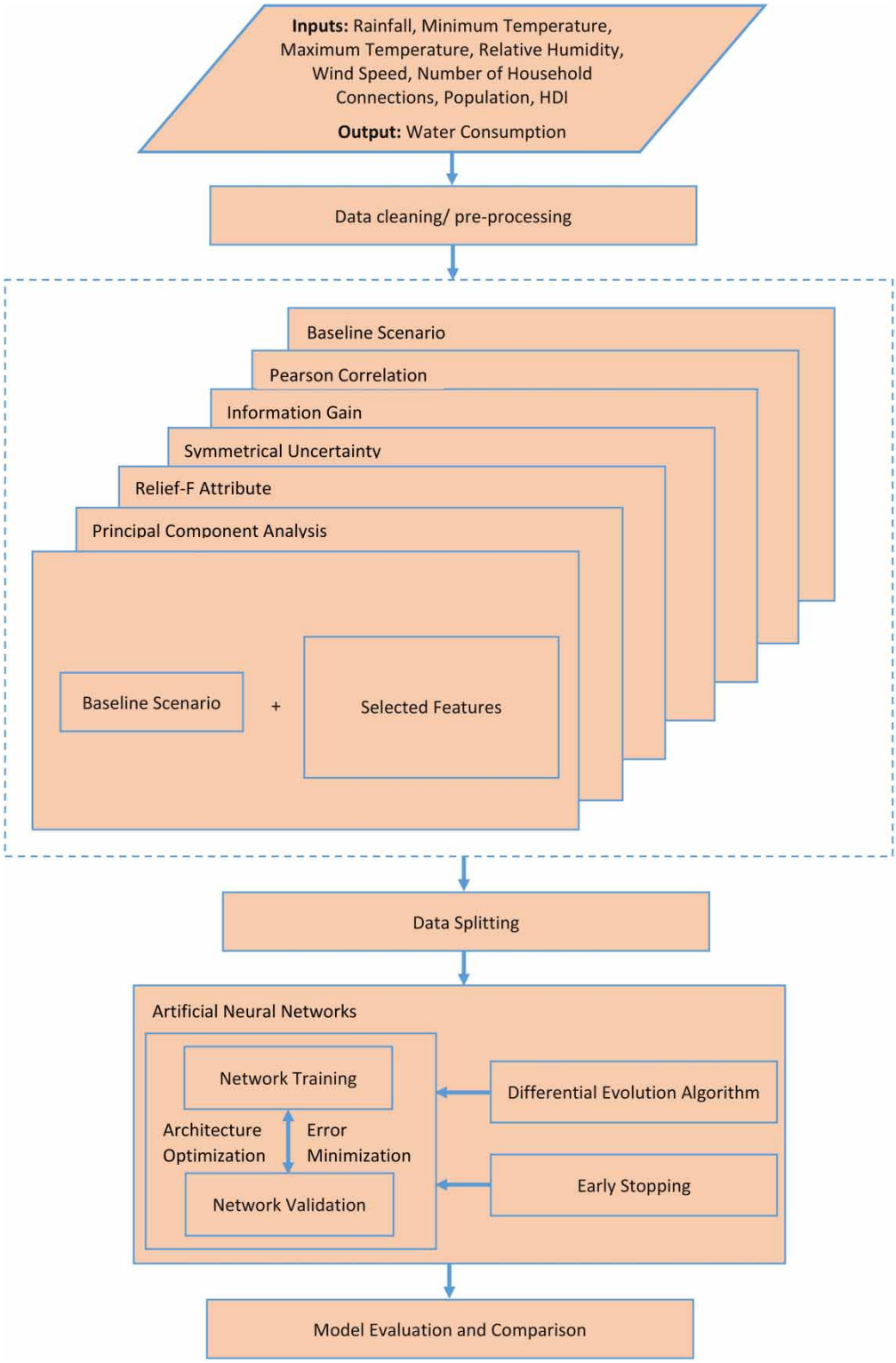

Figure 3 | Methodological framework.

Table 5 | Performance of models developed from each scenario

	Statistical parameters					
Techniques	**R^2 Training**	**R^2 Validation**	**RMSE Training**	**RMSE Validation**	**NSE Training**	**NSE Validation**
Baseline scenario	0.9038	0.8576	4,766	5,160	0.8808	0.8472
Pearson correlation	0.8812	0.9233	5,092	4,172	0.8639	0.9001
Information gain	0.8647	0.8961	5,105	4,505	0.8632	0.8835
Symmetrical uncertainty	0.8375	0.8611	5,659	5,227	0.8319	0.8454
Relief-F attribute	0.8576	0.9075	5,372	4,178	0.8485	0.8998
Principal component analysis	0.8397	0.8943	5,720	4,528	0.8283	0.8823

Table 6 | Model comparison and ranks based on model architecture and forecast accuracy during validation

		Optimal DE algorithm control parameters		Rank			
Techniques	**Model architecture**	**Cr**	**F**	**R^2**	**RMSE**	**NSE**	**Average: overall model accuracy**
Baseline scenario	8-2-1	0.7	0.5	6	6	6	6
Pearson correlation	4-4-1	0.8	0.3	1	1	1	1
Information gain	5-2-1	0.7	0.3	3	3	3	3
Symmetrical uncertainty	4-3-1	0.7	0.3	5	5	5	5
Relief-F attribute	5-3-1	0.7	0.3	2	2	2	2
Principal component analysis	5-2-1	0.8	0.4	4	4	4	4

all the ANN models developed using the five feature selection techniques converged better during validation than training, implying that the models do not suffer from the 'curse of dimensionality' and overfitting which typically plagues ANN models (Adeyemo *et al.* 2018). This also suggests that the early stopping criterion was effective in preventing overfitting.

Contrastingly, the ANN developed using the baseline scenario had the worst model performance amongst the six ANN models. A slight overfit can be noticed in its training and validation results. This slight overfit could be due to overparameterization, suggesting parameter irrelevancy or redundancy in the full set of potential explanatory variables considered. The performance of the baseline scenario ANN models agrees with the argument of Phyu & Oo (2016) that, if feature selection techniques are employed, the consistency of the full set of attributes can never be higher than that of any subset of attributes. Notwithstanding the rank of the ANN model developed from the baseline scenario, its performance could be referred to as reasonable considering its model architecture (8-2-1), which seems to have the minimal complexity.

Sensitivity analysis performed on the DE control parameters shows that the optimal crossover and mutation probabilities were in the [0.7, 0.8] and [0.3, 0.5] intervals, respectively, across the six ANN models. This suggests a high exploratory search by the DE algorithm, which is often a product of continuous productive search (Montgomery & Chen 2010).

Figures 4 and 5 present plots of observed and forecasted water demands for the training and validation phases, respectively. The plots clearly show that all the models produced a good representation of the water demand pattern in the City of Ekurhuleni. Both the peaks and troughs including sharp spikes in the water demand pattern were reproduced by the feature selection-based models. Some constant values are, however, noticeable in the baseline scenario model during training, possibly due to the inclusion of irrelevant variables

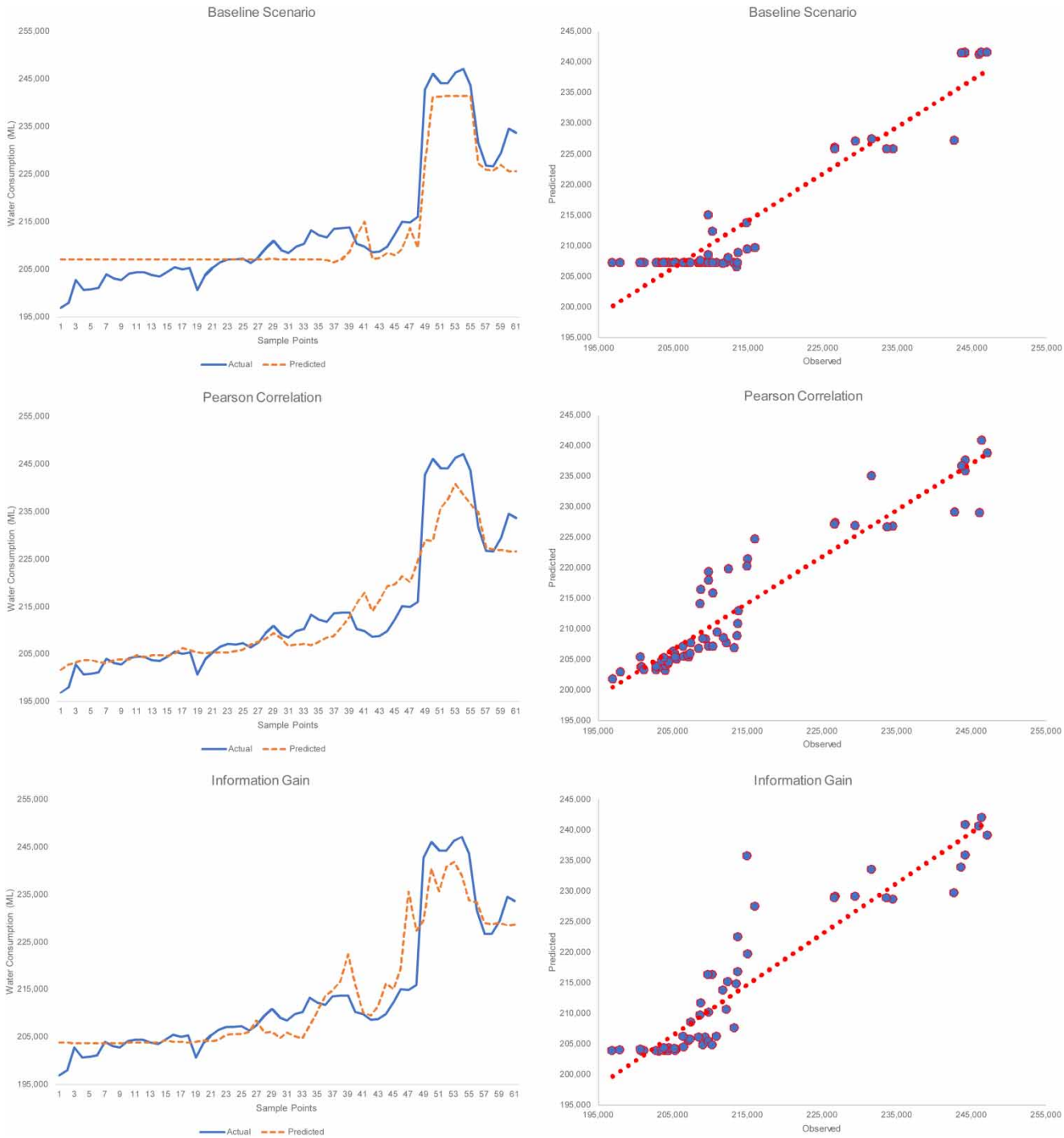

Figure 4 | Comparison of performances of developed models during the training phase. (*Continued*.)

that have little or no significant influence on the learning process. The corresponding scatter plots also depict high accuracy and correlation as the observed and forecasted values are close to the line of equality in all the models. The best model representation and the best line of fit are produced by the Pearson correlation-based model.

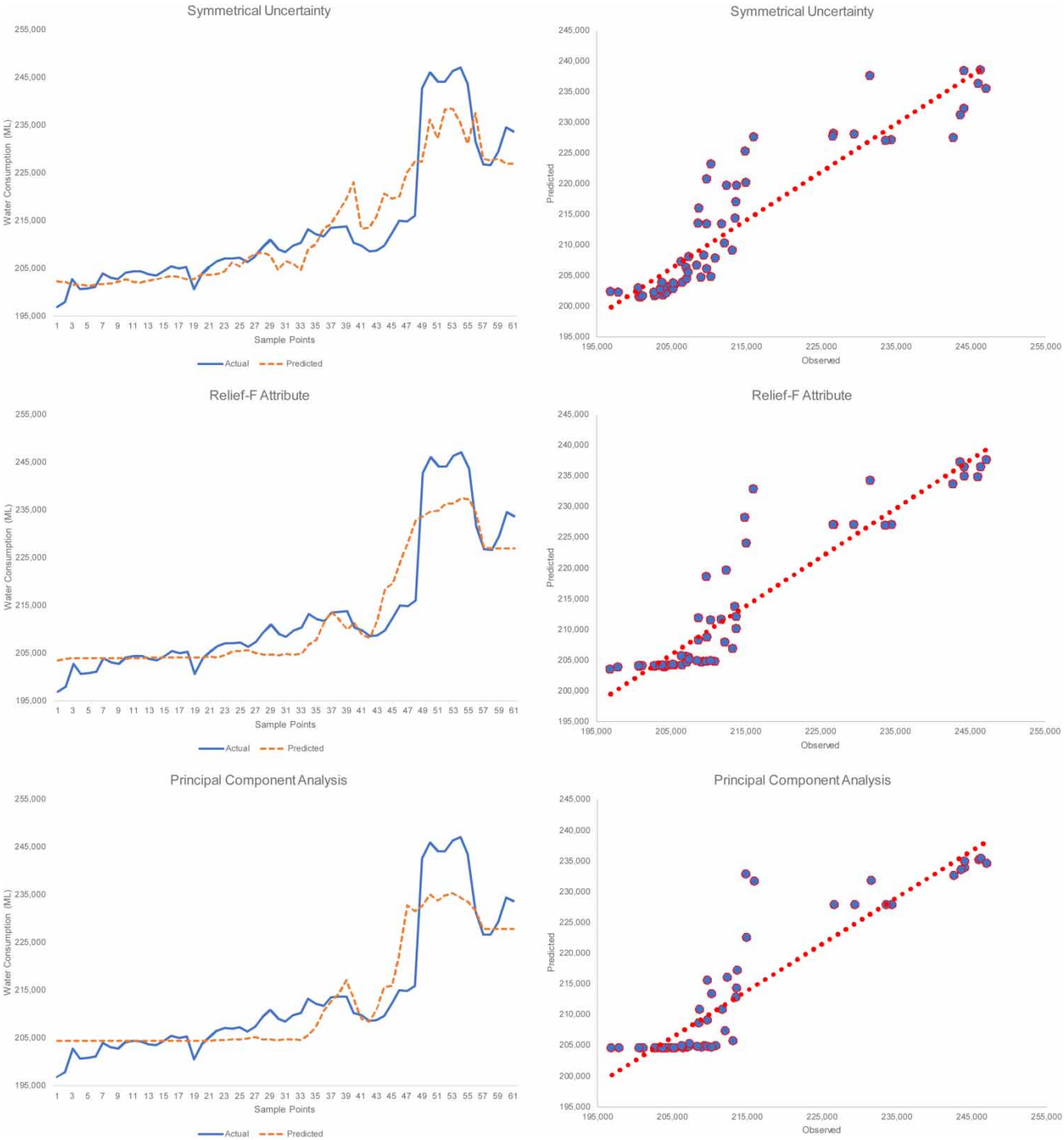

Figure 4 | Continued.

Table 7 shows the contribution of each potential explanatory variable. The contribution of each variable was determined by a total count across the subsets derived from the five feature selection techniques. The results show that the number of household connections and population contributed the most to model performance,

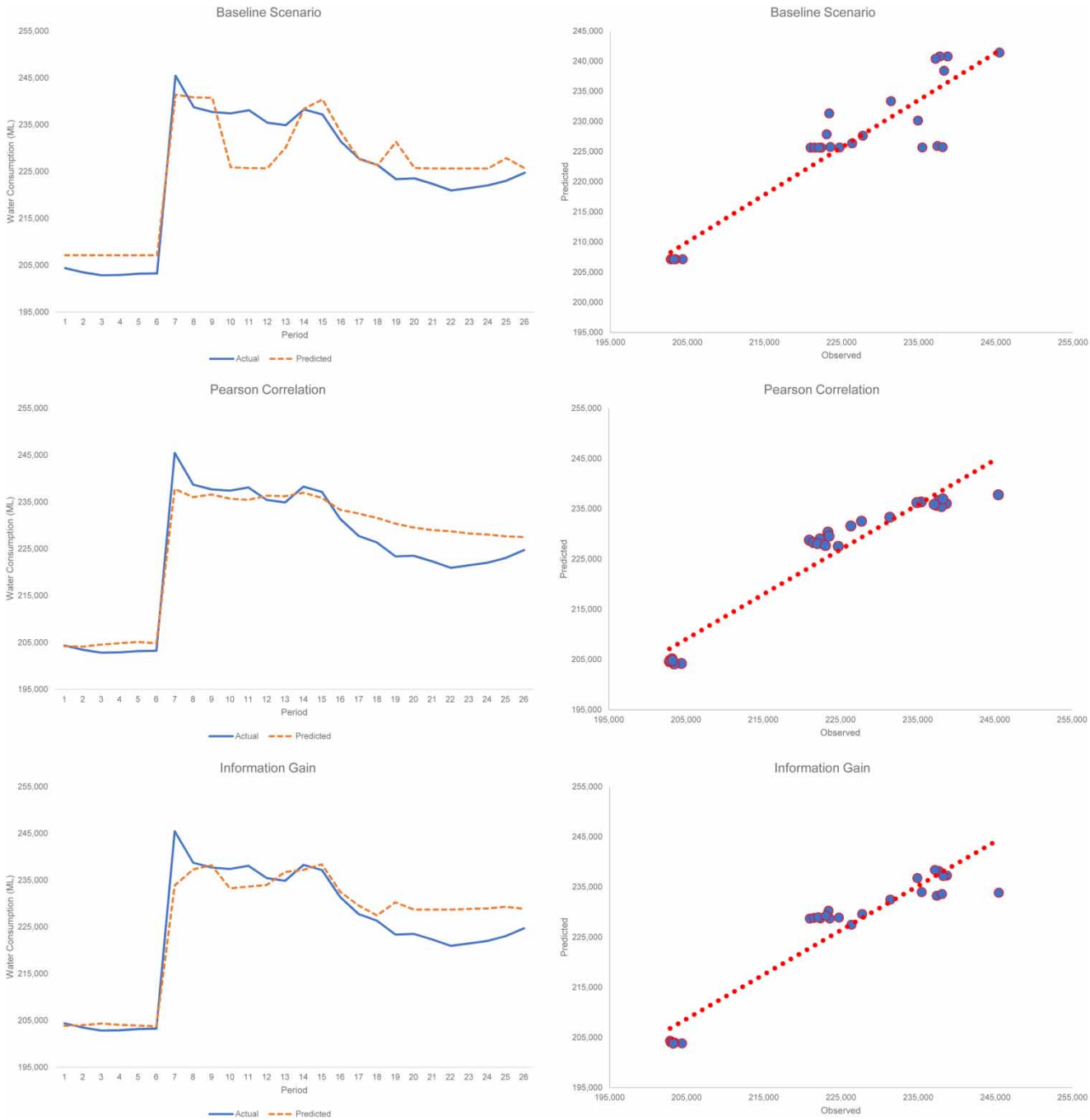

Figure 5 | Comparison of performances of developed models during the validation phase. (*Continued*.)

appearing in the subsets produced by the five feature selection techniques. This is followed by the minimum and maximum temperatures as well as HDI, which appeared in three of the five subsets. Although wind speed and rainfall are individually evident in only one of the subsets, their contribution is noteworthy as they, respectively, belong to the subsets that produced the most superior (Pearson correlation) and second-best (Relief-F attribute) performances. Similarly, RH appeared in the subsets of the third- and fourth-best models. These results suggest that, besides

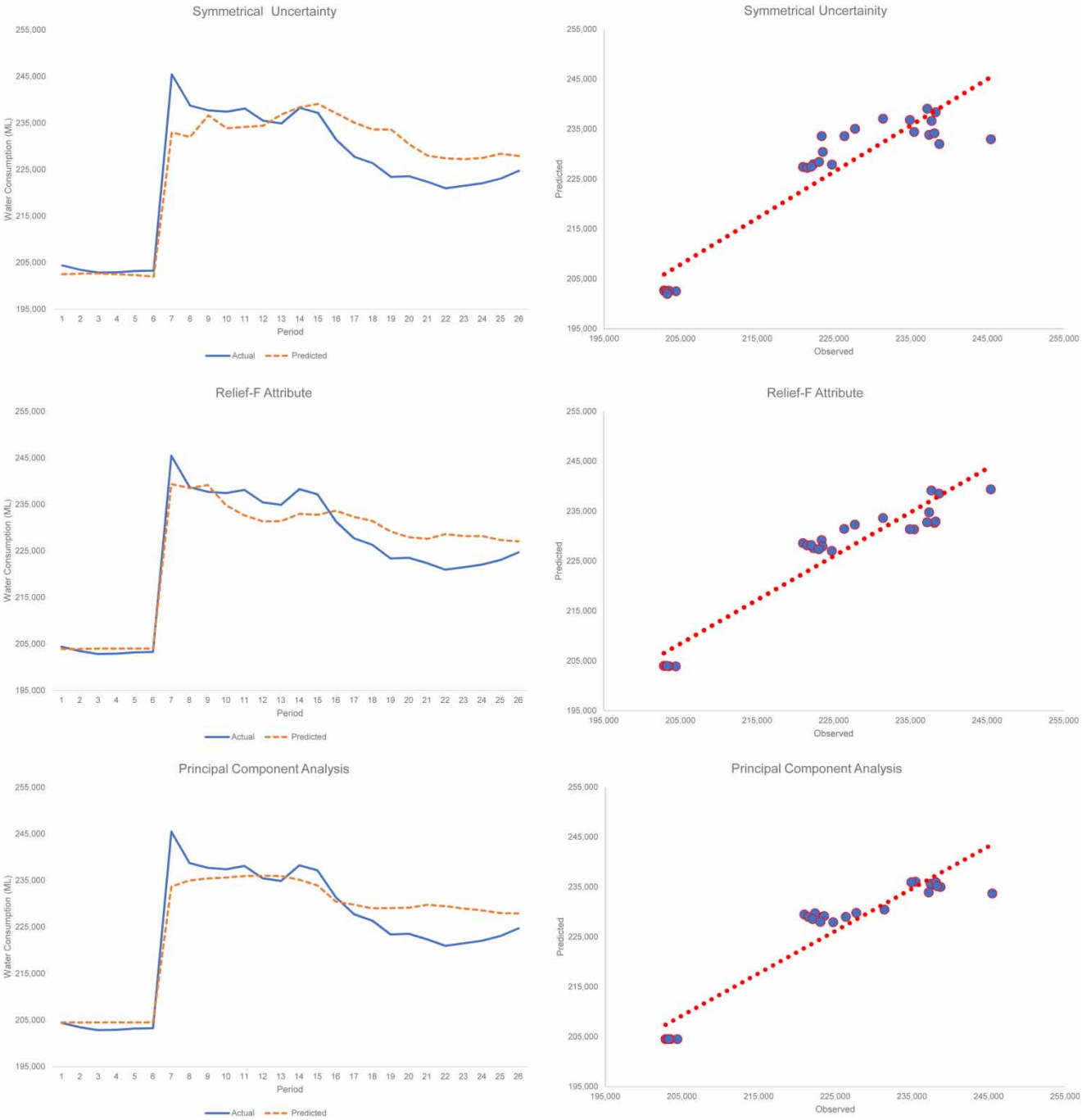

Figure 5 | Continued.

temperature, wind speed, rainfall and RH have some influence on water consumption in the City of Ekurhuleni. This may be explained by the climate of the study area which is characterized by wet, windy and humid summers, resulting in higher water use. Results from this study thus agree with the findings of Babel & Shinde (2011) and Huntra & Keener (2017), which found that RH, wind speed and rainfall could have some degree of influence on water consumption, especially in semi-arid regions. Generally, these results suggest that weather and socioeconomic status could have

Table 7 | Contribution of explanatory variables used for model development

Feature selection techniques	*R*	T_{min}	T_{max}	*RH*	*WS*	*HH*	*P*	*HDI*
Pearson correlation					✓	✓	✓	✓
Information gain		✓	✓	✓		✓	✓	
Symmetrical uncertainty		✓	✓			✓	✓	
Relief-F attribute	✓	✓				✓	✓	✓
Principal component analysis			✓	✓		✓	✓	✓
Total count	1	3	3	2	1	5	5	3

significant impacts on water demand, and thus incorporating them in water demand forecasting studies could result in a more reliable model that considers the impacts of weather and socioeconomic variations.

CONCLUSIONS AND FUTURE WORK

The capability of five feature selection techniques in finding the optimal subset of features for a water demand forecasting model has been investigated in this study. The performance of the subsets generated by the five feature selection techniques was compared to that of a baseline scenario comprising eight potential explanatory variables, totalling six scenarios. The aim was to develop an improved and reliable municipal water demand model that accounts for the impacts of weather and socioeconomic variations. HDI was introduced for the first time in water demand forecasting as a socioeconomic variable and used alongside weather-, population- and water demand-based variables. Using a combination evolutionary computation and artificial intelligence approach, DE-inspired ANN models were developed; one for each scenario. Results show that minimum and maximum temperatures as well as HDI were selected alongside population and number of household connections which are popularly used in water demand forecasting. Results further show that these three variables contributed significantly to the performance of three of the five models. Pearson correlation proved to be the most superior feature selection technique. DE showcased robustness in fine-tuning algorithm parameter values, thereby producing good performance in terms of the solution efficiency and quality. Generally, this study demonstrates that ANN water demand models can now account for the impacts of weather and socioeconomic variations by incorporating explanatory variables based on weather and socioeconomic factors. This study also suggests that the synergetic use of feature selection techniques, DE algorithm and early stopping criterion could be used to address the limitations of ANN, thereby improving model generalization and forecast accuracy as well as providing a climate variability perspective to water demand forecasting. The methodologies, principles and techniques behind this study foster sustainable development and thus could be adopted in planning and management of water resources. This study is limited to the use of historical water demand, weather and socioeconomic variables in predicting water demand. However, to enhance the applicability of the current ANN predictions, future research will focus on the impacts of other explanatory factors like land use, recharge and run-off on the city's water demand when the information becomes available.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to the City of Ekurhuleni, Statistics South Africa (Stats SA) and South African Weather Service (SAWS) for providing the data used for this study.

REFERENCES

Adeyemo, J., Oyebode, O. & Stretch, D. 2018 River flow forecasting using an improved artificial neural network. In: *EVOLVE-A Bridge Between Probability, Set Oriented Numerics, and Evolutionary Computation VI* (A. A. Tantar, E. Tantar, M. Emmerich, P. Legrand, L. Alboaie & H. Luchian, eds). Springer, Cham, Switzerland, pp. 179–193.

Ali, A. M., Shafiee, M. E. & Berglund, E. Z. 2017 Agent-based modeling to simulate the dynamics of urban water supply: climate, population growth, and water shortages. *Sustainable Cities and Society* **28**, 420–434.

Amaranto, A., Munoz-Arriola, F., Corzo, G., Solomatine, D. P. & Meyer, G. 2018 Semi-seasonal groundwater forecast using multiple data-driven models in an irrigated cropland. *Journal of Hydroinformatics* **20** (6), 1227–1246.

Azhagusundari, B. & Thanamani, A. S. 2013 Feature selection based on information gain. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* **2** (2), 18–21.

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** (6), 1653–1676.

Babel, M., Gupta, A. D. & Pradhan, P. 2007 A multivariate econometric approach for domestic water demand modeling: an application to Kathmandu, Nepal. *Water Resources Management* **21** (3), 573–589.

Bennett, C., Stewart, R. A. & Beal, C. D. 2013 ANN-based residential water end-use demand forecasting model. *Expert Systems with Applications* **40** (4), 1014–1023.

Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005 Input determination for neural network models in water resources applications. Part 1 – background and methodology. *Journal of Hydrology* **301** (1–4), 75–92.

Fagiani, M., Squartini, S., Gabrielli, L., Spinsante, S. & Piazza, F. 2015 Domestic water and natural gas demand forecasting by using heterogeneous data: a preliminary study. In: *Advances in Neural Networks: Computational and Theoretical Issues* (S. Bassis, A. Esposito & F. Morabito, eds). Springer, Cham, Switzerland, pp. 185–194.

Faizollahzadeh Ardabili, S., Najafi, B., Shamshirband, S., Minaei Bidgoli, B., Deo, R. C. & Chau, K.-w. 2018 Computational intelligence approach for modeling hydrogen production: a review. *Engineering Applications of Computational Fluid Mechanics* **12** (1), 438–458.

Firat, M., Yurdusev, M. A. & Turan, M. E. 2009 Evaluation of artificial neural network techniques for municipal water consumption modeling. *Water Resources Management* **23** (4), 617–632.

Ghalehkhondabi, I., Ardjmand, E., Young, W. A. & Weckman, G. R. 2017 Water demand forecasting: review of soft computing methods. *Environmental Monitoring and Assessment* **189** (7), 313.

Gubuza, D. 2017 On-site leak repair. In: *Water conservation and water demand management in the City of Ekurhuleni* (City of Ekurhuleni, ed.). Presentation at Rand Water Services Forum, Johannesburg.

Hall, M. A. 1999 *Feature selection for discrete and numeric class machine learning*. In: Working Paper Series, Working Paper 99/4. The University of Waikato, Hamilton, New Zealand, pp. 1–16.

Hu, T., Wu, F. & Zhang, X. 2007 Rainfall–runoff modeling using principal component analysis and neural network. *Hydrology Research* **38** (3), 235–248.

Huntra, P. & Keener, T. 2017 Evaluating the impact of meteorological factors on water demand in the Las Vegas Valley using time-series analysis: 1990–2014. *ISPRS International Journal of Geo-Information* **6** (8), 249.

IDP 2018 *Integrated Development Plan of City of Ekurhuleni 2017/2018 to 2020/2021*. City of Ekurhuleni.

Ilonen, J., Kamarainen, J.-K. & Lampinen, J. 2003 Differential evolution training algorithm for feed-forward neural networks. *Neural Processing Letters* **17** (1), 93–105.

Karimi, Z., Kashani, M. M. R. & Harounabadi, A. 2013 Feature ranking in intrusion detection dataset using combination of filtering methods. *International Journal of Computer Applications* **78** (4), 21–27.

Kim, J. H., Hwang, S. H. & Shin, H. S. 1999 An optimal neural network model for daily water demand forecasting. In: *29th Annual Water Resources Planning and Management Conference (WRPMD '99)*. ASCE, Arizona, pp. 1–10.

Kofinas, D., Mellios, N., Papageorgiou, E. & Laspidou, C. 2014 Urban water demand forecasting for the island of Skiathos. *Procedia Engineering* **89**, 1023–1030.

Mafi, S. & Amirinia, G. 2017 Forecasting hurricane wave height in Gulf of Mexico using soft computing methods. *Ocean Engineering* **146**, 352–362.

Maier, H. R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L. S., Cunha, M., Dandy, G. C., Gibbs, M. S., Keedwell, E. & Marchi, A. 2014 Evolutionary algorithms and other metaheuristics in water resources: current status, research challenges and future directions. *Environmental Modelling & Software* **62**, 271–299.

Mangal, A. & Holm, E. A. 2018 A comparative study of feature selection methods for stress hotspot classification in materials. *Integrating Materials and Manufacturing Innovation* **7** (3), 87–95.

Montgomery, J. & Chen, S. 2010 An analysis of the operation of differential evolution at high and low crossover rates. In: *Evolutionary Computation (CEC), 2010 IEEE Congress on*. IEEE, pp. 1–8.

Oyebode, O. K. 2014 *Modelling Streamflow Response to Hydro-Climatic Variables in the Upper Mkomazi River, South Africa*. Master's Thesis, Department of Civil Engineering and Surveying, Durban University of Technology, Durban.

Oyebode, O. & Stretch, D. 2018a Neural network modeling of hydrological systems: a review of implementation techniques. *Natural Resource Modeling* **32** (1), 1–14.

Oyebode, O. & Stretch, D. 2018b Water demand modelling using evolutionary computation techniques: integrating water equity and justice for realization of the sustainable development goals. *Environment, Development and Sustainability* (Submitted – in review).

Papageorgiou, E. I., Poczęta, K. & Laspidou, C. 2016 Hybrid model for water demand prediction based on fuzzy cognitive maps and artificial neural networks. In: *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, Vancouver, pp. 1523–1530.

Perea, R. G., Poyato, E. C., Montesinos, P. & Díaz, J. R. 2015 Irrigation demand forecasting using artificial neuro-genetic networks. *Water Resources Management* **29** (15), 5551–5567.

Phyu, T. Z. & Oo, N. N. 2016 Performance comparison of feature selection methods. In: *MATEC Web of Conferences*. EDP Sciences, Barcelona, p. 06002.

Piotrowski, A. P. 2014 Differential evolution algorithms applied to neural network training suffer from stagnation. *Applied Soft Computing* **21**, 382–406.

Pulido-Calvo, I. & Gutierrez-Estrada, J. C. 2009 Improved irrigation water demand forecasting using a soft-computing hybrid model. *Biosystems Engineering* **102** (2), 202–218.

Qi, C. & Chang, N.-B. 2011 System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts. *Journal of Environmental Management* **92** (6), 1628–1641.

Qu, J., Cao, L. & Zhou, J. 2010 Differential evolution-optimized general regression neural network and application to forecasting water demand in Yellow River Basin. In: *2010 2nd International Conference on Information Science and Engineering (ICISE)*. IEEE, pp. 1129–1132.

Raskutti, G., Wainwright, M. J. & Yu, B. 2014 Early stopping and non-parametric regression: an optimal data-dependent stopping rule. *The Journal of Machine Learning Research* **15** (1), 335–366.

Robnik-Šikonja, M. & Kononenko, I. 2003 Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning* **53** (1–2), 23–69.

Shabani, S., Yousefi, P., Adamowski, J. & Naser, G. 2016 Intelligent soft computing models in water demand forecasting. In: *Water Stress in Plants* (I. M. M. Rahman, ed.) InTech, London, pp. 99–117.

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* **8**, 1–26.

Stats-SA 2018 Mid-Year Population Estimates, 2018. In: *Statistical Release P0302* (S. S. Africa, ed.). Pretoria.

Storn, R. & Price, K. 1997 Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* **11** (4), 341–359.

Šiljić Tomić, A., Antanasijević, D., Ristić, M., Perić-Grujić, A. & Pocajt, V. 2018 Application of experimental design for the optimization of artificial neural network-based water quality model: a case study of dissolved oxygen prediction. *Environmental Science and Pollution Research* **25** (10), 9360–9370.

Toth, E., Bragalli, C. & Neri, M. 2018 Assessing the significance of tourism and climate on residential water demand: panel-data analysis and non-linear modelling of monthly water consumptions. *Environmental Modelling & Software* **103**, 52–61.

UNESCO 2015 *The United Nations World Water Development Report 2015: Water for a Sustainable World*. United Nations World Water Assessment Programme – WWAP Report 9231000713, UNESCO Publishing, Paris.

Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J. 2016 *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, MA.

Wu, Z. Y. & Yan, X. 2010 Applying genetic programming approaches to short-term water demand forecast for district water system. In: *Water Distribution Systems Analysis 2010* (K. E. Lansey, C. Y. Choi, A. Ostfeld & I. L. Pepper, eds). ASCE, Virginia, USA, pp. 1498–1506.

Yousefi, P., Shabani, S., Mohammadi, H. & Naser, G. 2017 Gene expression programing in long term water demand forecasts using wavelet decomposition. *Procedia Engineering* **186**, 544–550.

First received 11 November 2018; accepted in revised form 15 April 2019. Available online 30 April 2019