

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Intelligent Soft Computing Models in Water Demand Forecasting

Sina Shabani, Peyman Yousefi, Jan Adamowski and
Gholamreza Naser

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/63675>

Abstract

Given the increasing trend in water scarcity, which threatens a number of regions worldwide, governments and water distribution system (WDS) operators have sought accurate methods of estimating water demands. While investigators have proposed stochastic and deterministic techniques to model water demands in urban WDS, the performance of soft computing techniques [e.g., Genetic Expression Programming (GEP)] and machine learning methods [e.g., Support Vector Machines (SVM)] in this endeavour remains to be evaluated. The present study proposed a new rationale and a novel technique in forecasting water demand. Phase space reconstruction was used to feed the determinants of water demand with proper lag times, followed by development of GEP and SVM models. The relative accuracy of the three best models was evaluated on the basis of performance indices: coefficient of determination (R^2), mean absolute error (MAE), root mean square of error (RMSE), and Nash-Sutcliffe coefficient (E). Results showed GEP models were highly sensitive to data classification, genetic operators, and optimum lag time. The SVM model that implemented a Polynomial kernel function slightly outperformed the GEP models. This study showed how phase space reconstruction could potentially improve water demand forecasts using soft computing techniques.

Keywords: water demand forecasting, soft computing, genetic expression programming, support vector machines, phase space reconstruction, lag time

1. Introduction

While water scarcity has become a key concern worldwide, it is particularly so in arid and semiarid regions with limited potable water sources. In designing water distribution systems (WDS), engineers have typically used a “fixture unit” method, which considers the sum of fixture unit demands, facility types, and socioeconomic factors to determine peak demand. However, this overestimates the actual peak demand by as much as 100% [1]. Due to various uncertainties, including those associated with demand, engineers often include large safety factors when designing WDS. Given that WDS rely mainly on regional energy and resources, an oversized system can have environmental impacts that will appear in region(s) beyond the jurisdictional boundaries of the system. While short-term demand forecasts are critical to a WDS daily operations [2], long-term forecasts are required for future planning and management of the systems. In providing an accurate estimate of water demand, a robust demand-forecasting model assists managers in designing a more environmentally sustainable WDS and in managing available water resources more efficiently. When coupled with a water demand management strategy, such models can help managers overcome operational problems (e.g., low pressure during peak demands) and issues related to asset management (e.g., nonreplacement of assets or replacement by lower capacity assets reaching the end of their economic life). It has been estimated that a well-predicted monthly average demand might be up to 400% lower than peak demands that cause low pressure; however, a more realistic model can enhance resource management and operating systems. This will eventually lead to significant savings for water and energy (for running pumps, treatment plants, etc.) industries. Considering weather conditions and population, the prime objective of the present study was to develop a predictive model for monthly average water demand. While the present study proposed a generic framework that could be easily adjusted for any specific case, the City of Kelowna (British Columbia, Canada) was employed as a test case.

2. Literature review

Water demand varies greatly both regionally and seasonally. Increasing urbanization and industrialization as well as emerging issues such as shifting weather patterns and population growth have significant impacts on water demand. The main components in demand prediction are the explanatory variables and time scales used. Selecting explanatory variables for a predictive model depend on the desired time scale and the availability of data. Simple models using very few explanatory variables have shown promising accuracy for short-term prediction [3, 4]. In general, the explanatory variables affecting water demand are of two types: weather (e.g., temperature, relative humidity, and rainfall) and socioeconomic (e.g., population and income). Weather conditions affect short-term prediction while their socioeconomic counterparts can affect long-term predictions [5–7]. As has been highlighted by significant worldwide changes in climate, both in terms of weather conditions and global warming, water availability is prone to great uncertainty [8]. Therefore, the impact of evolving weather

conditions on long-term water demand predictions should receive greater attention. Furthermore, researchers who have considered weather conditions in short-term water demand prediction have established that it is not feasible to feed online automated WDS with real time weather information [9]. As a result, limited studies have considered weather conditions in their demand forecasting models [10–12]. **Table 1** summarizes the relevant literature. Temperature, precipitation, pan evaporation, and number of days since the last rainfall were used in a forecasting model [13]. Another study used temperature, relative humidity, rainfall, wind speed, and air pressure as weather parameters in their hourly water demand model for Sao Paulo, Brazil [12]. **Table 1** shows the previous researchers did not consider socioeconomic and weather conditions simultaneously since their effects are highly dependent on the forecast's time scale. Traditionally, WDS utilities have used historical patterns as explanatory variables in predicting future water demands. Scarce water reserves and the rapid increase in urbanization have raised awareness and led to implementation of statistical approaches. Multiple linear regression (MLR) and time series were the most popular techniques used in the early stages of demand forecasting [6]. While MLR has been widely used to better understand the determinants of water demand [14–18], its major drawback is the fact that it considers linear relationships among variables and water demand, such relationships are nonlinear by nature. Time series have been introduced along with regression as methods for demand forecasting [10, 19]. Due to the common belief that they can deal with complex systems [20], artificial neural networks (ANNs) have been widely applied in water demand forecasting [21–23, 2]. Comparing regression, univariate time series, and ANN models, one study found ANN models drawing on standard rainfall and maximum temperature data could better predict weekly water demand than other models [6]. Similarly, drawing on temperature and rainfall data in their forecasting models, researchers concluded that ANN models provided more reliable forecasts for peak weekly demand than time series and simple and multiple linear regressions [22]. Results of another study showed ANN models performed better for hourly forecasts, whereas regression models were more accurate in forecasting daily demand [23]. To improve the accuracy and robustness of demand forecasting models, hybrid models combining or modifying ANN, MLR, and time series techniques have been tested [24–27]. However, application of nonlinear regression in demand forecasting has remained limited to studies using support vector machines (SVMs) [28–30] and training nonlinear relationships through linear regression models [6, 31]. The present study compares gene expression programming (GEP) and SVM nonlinear approaches. Inspired by Darwin's theory of evolution, GEP was recently proposed in engineering disciplines to optimize the structure of input variables fed into predictive models [32]. Being a self-learning algorithm, GEP has several advantages over conventional predictive models. GEP defines individual block structures (input variables, response, and function sets) and selects the optimized operating functions and multipliers through the process of learning algorithms. Results of one study indicated GEP models outperformed traditional linear models in the field of hydrology [32]. Since weather information is one of the major determinants of water demand, this research employed GEP to develop a robust and accurate demand-forecasting model.

No.	Reference	Method	Determinant	Time scale
1	[16]	Linear regression	Seasonal dummies, derivatives of weather and price	Monthly demand
2	[17]	Linear regression	Density, building size, lot size, household size, income, price, temp, rain, drought dummies	Bimonthly demand
3	[18]	Regression using Bayesian moment entropy	Population density	Annual demand
4	[13]	Decomposed daily demand followed by composite forecasts	Daily demand and hourly demand	Daily demand
5	[19]	Univariate time series	Y_{t-1}	Annual residential demand
6	[22]	Regression and ANN	Temp, rainfall, and lags of peak demand	Peak weakly demand
7	[23]	ANN	Temperature, rainfall, and delayed demand	Daily demand
8	[2]	Time series	Univariate demand series, temperature in a multivariate model	Daily, weekly, monthly, annual
9	[6]	Time series and ANN	Delayed demands, temperature, and rainfall	Weekly demand
10	[24]	Holt-Winters multiplicative smoothing modified regression	Precipitation, temperature, humidity, lagged demand	Weekly (6 days)
11	[26]	Weighted average regression and ANN	Historical demand and time	Annual demand
12	[27]	Decomposed annual demand, regression and ANN	GDP, population, temperature, greenery coverage, delayed demand	Annual demand
13	[31]	Wavelet-deinoizing and ANN	7-year long time series of demand	Monthly demand
14	[28]	SVM with RBF function is compared with ANN	Delayed demand, population	Daily demand
15	[29]	ANN, SVM, Monte Carlo	Rain, demand, wind speed, atmospheric pressure	Hourly demand
16	[30]	SVM and adaptive Fourier series	Wind speed, temperature, demand, humidity, and rainfall	Hourly demand

Table 1. Literature on water demand forecasting.

3. Study area and data collection

This research focused on the City of Kelowna located in the Okanagan Valley (British Columbia, Canada). The City has five water districts including the City of Kelowna District (CKD), Glenmore Ellison Irrigation District (GEID), Black Mountain Irrigation District (BMID), Rutland Water District (RWD), and the South East Kelowna Irrigation District (SEKID). The CKD served as the study area of this research. Using three major pumping stations, the CKD primarily supplies water from the Okanagan Lake. The present study used monthly mean water demand data from 1996 to 2010 (<http://www.kelowna.ca/>). The population censuses of 1996, 2001, 2006, and 2011, along with the best-fit parabolic equation (with coefficient of determination of $R^2 \approx 1$) allowed estimation of the population in noncensus years. Weather indices including temperature, wind speed, relative humidity, and rainfall, were drawn from the Environment Canada weather data (<http://kelowna.weatherstats.ca/>) collected at Station A (latitude 49°57'13"N, longitude 119°22'29"W) located at the City of Kelowna's airport.

4. Methodology

4.1. Model development

To determine water demand (D) in millions of liters (ML), this research used population (P) and hotel occupancy factor (HOR) as socioeconomic parameters (the City of Kelowna is one of the hot spots for tourism in North America), and temperature (T) in °C, relative humidity (RH) in percent, and rainfall (R) in millimeters as weather parameters. As these parameters did not have the same order of magnitude, they were normalized prior to models development by

$$X = \frac{x - \mu}{\sigma} \quad (1)$$

where X is the standardized magnitude of parameter x , μ and σ are the corresponding mean and standard deviation, respectively. Phase space reconstruction of each explanatory variable was used prior to GEP modeling to define the structure of the model inputs. This was done to identify the stochastic or deterministic nature of the collected data. For a given proper lag time, the phase space was built by applying Taken's theorem [33] and transforming the time-series data into the geometry of a single moving point along a trajectory, where each point corresponds to a datum. Average mutual information (AMI) was used to determine the proper lag time of water demand for phase space reconstruction of all input factors. This was done to achieve a comprehensive understanding of input factors, variable self-interaction, and assess the use of lag times in demand forecasting models. Labeled $M_a D_b OP_c$, where a , b , and $c \in \{1, 2, 3\}$ a total of 27 models were created (**Table 2**), which combined three input types [M_1 : demand data only; M_2 : demand and climatic data; M_3 : demand, climatic, and demographic data], three lag times [D_1 : 1 month lag; D_2 : 1 and 2 month lags; D_3 : 1, 2, and 3 month lags], and three types

of genetic operators [OP₁: {+, −, x}; OP₂: {+, −, x, x², x³}; OP₃: {+, −, x, x², x³, √, e^x, log, ln}] used in developing the GEP models.

Classification	Model	Input variables combination*
Demand Data Based	M ₁ D ₁	D _{t-1}
	M ₁ D ₂	D _{t-1} , D _{t-2}
	M ₁ D ₃	D _{t-1} , D _{t-2} , D _{t-3}
Demand + Weather Data Based	M ₂ D ₁	D _{t-1} , T _{t-1} , R _{t-1} , RH _{t-1}
	M ₂ D ₂	D _{t-1} , D _{t-2} , T _{t-1} , T _{t-2} , R _{t-1} , R _{t-2} , RH _{t-1}
	M ₂ D ₃	D _{t-1} , D _{t-2} , D _{t-3} , T _{t-1} , T _{t-2} , T _{t-3} , R _{t-1} , R _{t-2} , R _{t-3} , RH _{t-1} , RH _{t-2} , RH _{t-3}
Demand + Weather + Population Data Based	M ₃ D ₁	D _{t-1} , T _{t-1} , R _{t-1} , RH _{t-1} , P, HOR
	M ₃ D ₂	D _{t-1} , D _{t-2} , T _{t-1} , T _{t-2} , R _{t-1} , R _{t-2} , RH _{t-1} , RH _{t-2} , P, HOR
	M ₃ D ₃	D _{t-1} , D _{t-2} , D _{t-3} , T _{t-1} , T _{t-2} , T _{t-3} , R _{t-1} , R _{t-2} , R _{t-3} , RH _{t-1} , RH _{t-2} , RH _{t-3} , P, HOR

*t is current month; D is demand; HOR is hotel occupancy factor; P, is population; R is rainfall; RH is relative humidity; T is temperature.

Table 2. Structure of classified models.

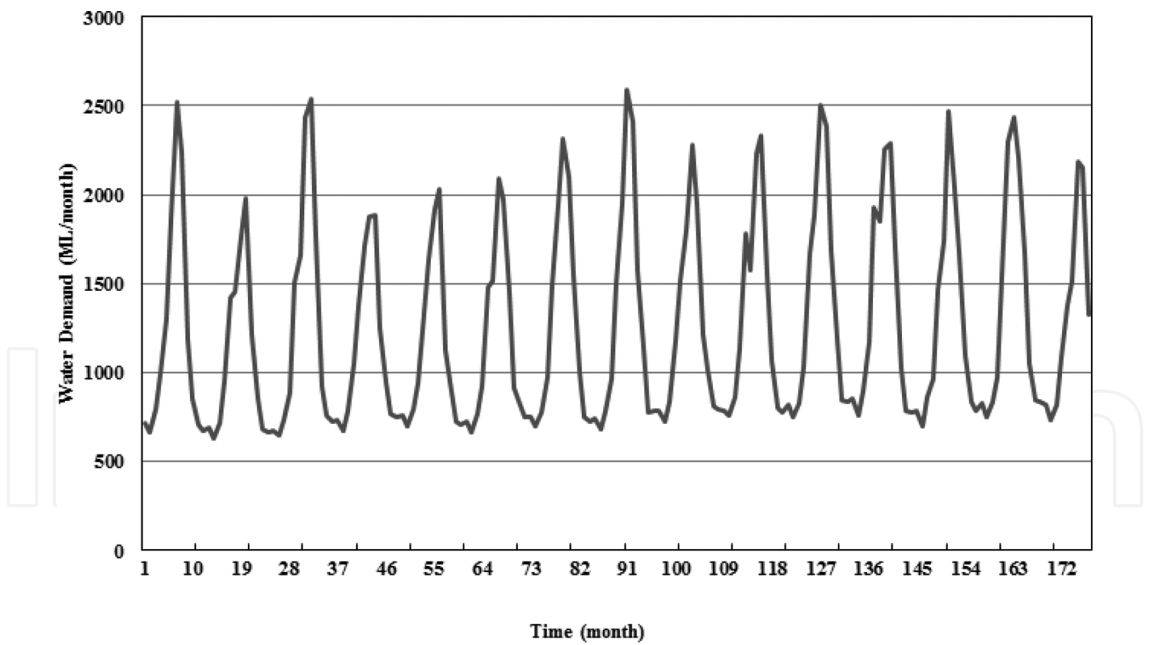


Figure 1. Time series of water demand in the City of Kelowna District (CKD) for 1966–2008.

Data were used in partitions of 144 samples for training (1996–2007) and 35 samples for validation (2008–2010). The time series of water demand over the time period of 1996–2010 (Figure 1) shows a relatively regular periodic cycle of water demand in CKD that is mainly due to seasonal changes.

4.2. Genetic expression programming (GEP)

Introduced by Ferreira, GEP is an emerging soft computing technique [34]. The strategy used for the learning algorithms was the optimal evolution using the genetic operators. Following Ferreira, this research defined the overall structure of the GEP model by: 30 chromosomes, eight head sizes, and three genes [35]. The selected head size determined how complex each model parameter was. Each of the gene heads underwent a set of different arrangements to model the feeding data. Selecting new random populations was followed by reproduction in order to reach the most suitable model under optimized stopping conditions. Models were developed based on three genes linked together by an addition function. The number of genes per chromosome specified the layers or blocks involved in building the whole model. Although a large gene was useful, dividing the chromosomes into simpler units resulted in a more efficient and manageable learning process. RMSE was used as a fitness function to fit a curve to target values. The stopping condition was a maximum fitness and coefficient of determination (R^2). Ten numerical constants were used as floating data point in each gene.

4.3. Lag time

The literature lists three methods for estimating lag time, AMI, autocorrelation function (ACF), and correlation integral (CI) [36–38]. AMI is considered the best since ACF reflects only linear properties and CI requires a large set of data [39]. Consequently, the present study employed AMI defined as:

$$I_{\tau} = \sum_{i=1}^{i=n} P(X_i, X_{i+\tau}) \cdot \log_2 \frac{P(X_i, X_{i+\tau})}{P(X_i) \cdot P(X_{i+\tau})} \quad (2)$$

where the joint probability of two successive time series, $P(X_i, X_{i+\tau})$ and the product of their individual marginal probability, $P(X_i) \cdot P(X_{i+\tau})$, were used to find the optimum lag time. This lag can contribute to the maximum information added on X_i by the successive time series $X_{i+\tau}$. The prime objective of using this approach was to make sure these time series were independent and thereby better represented the dynamics of the system in the phase space. In other words, a balanced independency was desirable in identifying an optimum delay time.

4.4. Support vector machines (SVM)

For SVM models, in which genetic operators are not used, the input types remained M_1 , M_2 , or M_3 , while the lag times remained D_1 , D_2 , or D_3 . This study compared the performance of radial basis function (RBF), polynomial (Poly), and Linear (Lin) kernels. These were appended to the input type and lag, e.g., M_1D_1 RBF, M_1D_1 Poly, or M_1D_1 Lin. **Figure 2** shows the structure of the SVM model. Kernel functions (RBF, Poly, or Lin) were used to map the input vectors into higher dimensions in space.

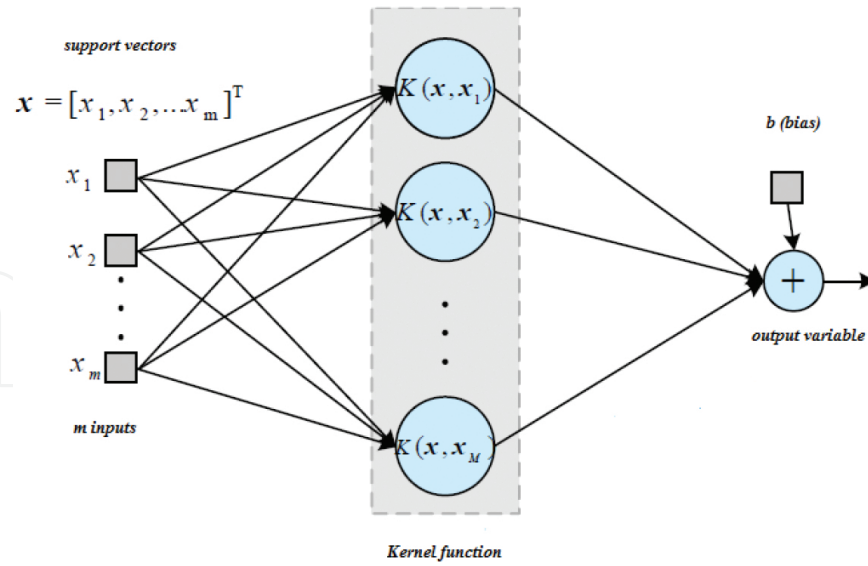


Figure 2. Support vector machine structure.

In this method, the input vectors are considered as supports forming the backbone of the whole model structure through a training process. If N samples of the population given by $X \in R^m$, $\{X_K, Y_K\}_{K=1}^N$, $Y \in R$, a function or SVM estimator on a regression can be considered as:

$$f(x) = W\phi(X) + b \quad (3)$$

where X is an input parameter with m components and Y is its response output variable, W is a weight vector, b represents a bias, and ϕ is a transfer function which exhibits nonlinear behavior, mapping the input vectors into a higher dimensional space. As these mapped vectors can compromise the complex nonlinear regression of the input space, Cortes and Vapnik introduced the convex optimization problem with an insensitivity loss function [40]:

$$\text{minimize } w, b, \xi, \xi^* \quad \frac{1}{2} \|W\|^2 + C \sum_{k=1}^N (\xi_k - \xi_k^*) \quad (4)$$

$$\text{subject to } \begin{cases} Y_k - W^T \phi(X_k) - b \leq \varepsilon + \xi_k \\ W^T \phi(X_k) + b - Y_k \leq \varepsilon + \xi_k^* \\ \xi_k, \xi_k^* \geq 0 \end{cases} \quad k = 1, 2, \dots, N \quad (5)$$

where ξ_k and ξ_k^* are slack variables that penalize training errors by the loss function over the error tolerance ε , and C is a positive trade-off parameter that determines the degree of the empirical error in the optimization problem. Following previous researchers [41, 42], the

optimization was simultaneously undertaken through Lagrangian multipliers under Karush Kuhn-Tucker (KTT) conditions.

5. Results and discussion

The prime objective of using phase space reconstruction was to find a proper lag time for developing the models in this study. In order to have a comprehensive understanding of model performance, GEP models were defined by all lag times up to the optimum value determined for water demand in the CKD. The AMI calculations of the water demand in the CKD resulted in a lag time of 3 months. **Figure 3** shows that the first local minimum point occurs at 3 months,

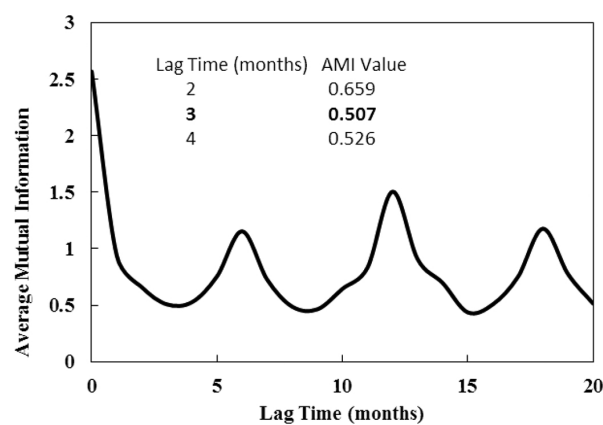


Figure 3. Average mutual information (AMI) for water demand.

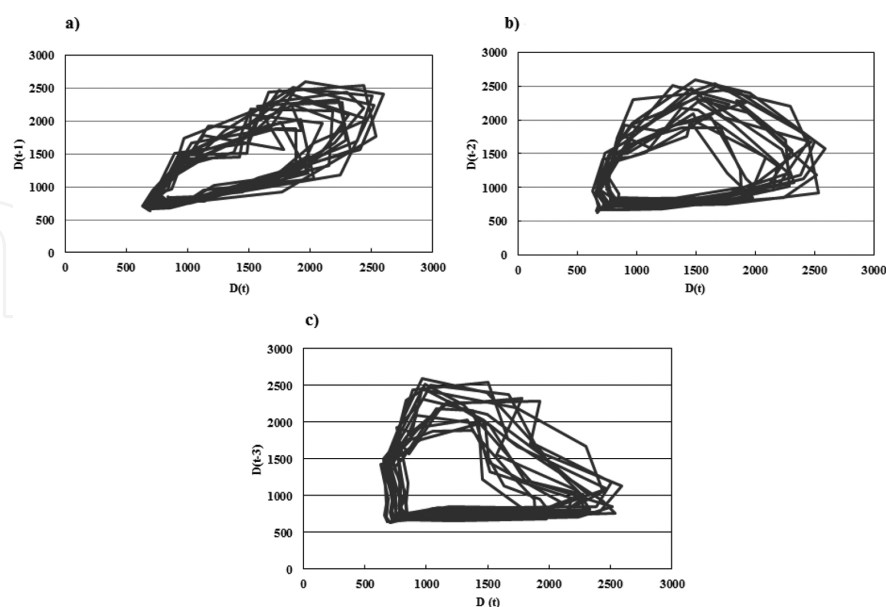


Figure 4. Phase space diagram lag times (1–3 months).

allowing the AMI an optimum lag time for phase space reconstruction ($\tau = 0.6591$ for 2 months, $\tau = 0.5073$ for 3 months).

Figure 4a–c shows the phase space diagrams of water demand for $\tau = 1, 2$, and 3 months, respectively. Each figure represents the state of WDS demand at the given time. The evolution of phase space in this time series was given by reconstructing a pseudo phase space in which the demand of CKD, a nonlinear system, was considered by its self-interaction using AMI [43]. **Figure 4c** ($\tau = 3$) has a more regular pattern in comparison with the other two previous states of phase space ($\tau = 1, 2$; **Figure 4a** and **b**, respectively), showing a lag time of 3 months to be optimum.

Prior to analysis with GEP models, a correlation table between the explanatory variables and water demand provided a better understanding of how to define the input factors (**Table 3**). The correlations were 0.92, 0.84, -0.83 , 0.11, and -0.01 for D vs. T , D vs. HOR, D vs. RH, D vs. P , and D vs. R , respectively. Interestingly, water demand was highly correlated to temperature and hotel occupancy rate in CKD, showing the periodic cycle of demand due to seasonal changes. This research, however, employed all input factors in evolving the GEP models.

	<i>D</i>	<i>T</i>	<i>R</i>	<i>RH</i>	<i>P</i>	<i>HOR</i>
<i>D</i>	1.00	0.92	-0.01	-0.83	0.11	0.84
<i>T</i>	0.92	1.00	0.10	-0.89	0.00	0.92
<i>R</i>	-0.01	0.10	1.00	-0.05	-0.26	0.11
<i>RH</i>	-0.83	-0.89	-0.05	1.00	0.02	-0.84
<i>P</i>	0.11	0.00	-0.26	0.02	1.00	-0.09
<i>HOR</i>	0.84	0.92	0.11	-0.85	-0.09	1.00

D, demand; *P*, population; HOR, hotel occupancy factor; *T*, temperature, RH, relative humidity, and *R*, rainfall.

Table 3. Correlation between water demand and factors impacting demand.

Table 4 shows all 27 GEP models developed in the present study. Three superior models were highlighted in each category or classification of determinants. Interestingly, a lag time of 3 months outperformed other combinations in all different classifications which show the importance of using phase space construction in studying complex systems. This shows that an appropriate lag time determined by AMI can significantly improve the performance of the forecasting model. Different genetic operators were also used to understand which mathematical operations better define the nature of these determinants. The first operator $\{+, -, x\}$ showed better performance in the first two classifications, i.e., for demand based and demand plus climatic info based categories. The second operator (OP_2) $\{+, -, x, x^2, x^3\}$ outperformed other operators in (OP_3) (demand + socioeconomic + climatic information) of input parameters in which socioeconomic factors were included. It is interesting that using more complex mathematical operations, as in OP_3 $\{+, -, x, x^2, x^3, \sqrt{}, e^x, \log, \ln\}$ consistently reduced the quality of the models' performance. This showed that water demand forecasting could be reasonably

explained by models using basic mathematical operations despite its complexity. Used to investigate the sensitivity of the models to determinant classification, the genetic operator, and lag time, the performance indices of MAE and RMSE did little to distinguish among the best performing models ($M_1D_3OP_1$, $M_2D_3OP_1$, and $M_3D_3OP_2$) in each category, i.e., MAE = 0.304, 0.3035, and 0.291, respectively, and RMSE = 0.3984, 0.3664, and 0.3660. While R^2 values showed M2 and M3 models to slightly outperform M1 models, plotting observed and predicted demand over time, as well as scatter plots of observed vs. predicted demand served to further delineate differences in performance (**Figure 5**). Comparing cumulative water demand calculated by each of the three top models to observed values showed the $M_1D_3OP_1$ and $M_3D_3OP_2$ models to be more accurate than $M_2D_3OP_1$ (**Figure 6**). In order to distinguish between $M_1D_3OP_1$ and $M_3D_3OP_2$ a plot of cumulative (observed – predicted) was plotted (**Figure 7**). This

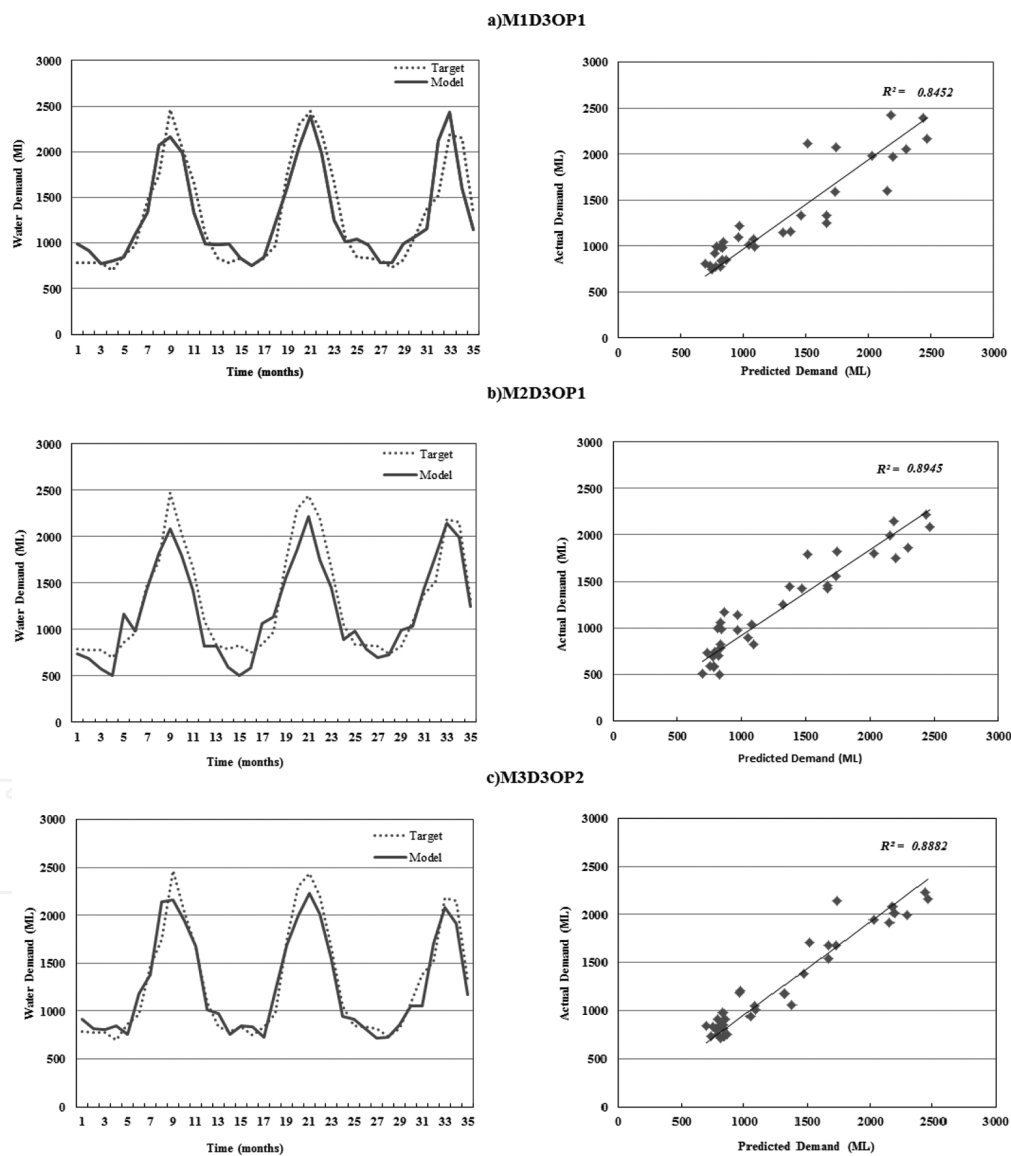


Figure 5. Observed and predicted demand over time (left), and scatter plots of observed vs. predicted demand (right) using superior GEP models: (a) $M_1D_3OP_1$; (b) $M_2D_3OP_1$; c) $M_3D_3OP_2$.

showed model $M_3D_3OP_2$ to be the best given the lesser fluctuations in errors and a consistent pattern throughout the plot's time period. This better performance may be attributable to the combination of socioeconomic factors with demand and climatic data; this might having resulted in a more consistently accurate model, which lowered the error associated compared to the other two models.

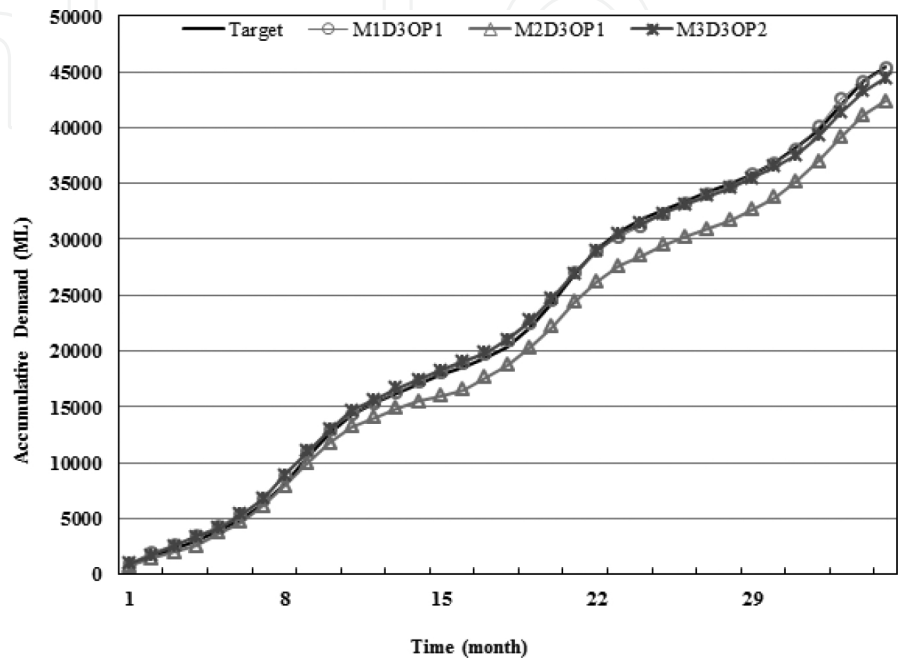


Figure 6. Cumulative demand with time.

Model ID*	Training			Testing		
	MAE	RMSE	R ²	MAE	RMSE	R ²
$M_1D_1OP_1$	0.4687	0.6974	0.6284	0.4833	0.6067	0.6343
$M_1D_1OP_2$	0.4718	0.6100	0.6252	0.4849	0.6120	0.6300
$M_1D_1OP_3$	0.4672	0.6118	0.6235	0.4800	0.6112	0.6281
$M_1D_2OP_1$	0.3552	0.4721	0.7754	0.378	0.4607	0.7892
$M_1D_2OP_2$	0.3574	0.4721	0.7756	0.3794	0.4608	0.7892
$M_1D_2OP_3$	0.3008	0.4049	0.8481	0.4188	0.5188	0.8346
$M_1D_3OP_1$	0.3229	0.4317	0.8156	0.3040	0.3984	0.8452
$M_1D_3OP_2$	0.2858	0.3641	0.8691	0.3488	0.3106	0.8452
$M_1D_3OP_3$	0.3545	0.4647	0.7849	0.3637	0.4548	0.8029
$M_2D_1OP_1$	0.3777	0.4790	0.7735	0.4529	0.5296	0.7552
$M_2D_1OP_2$	0.3955	0.4933	0.7560	0.4423	0.5169	0.7546
$M_2D_1OP_3$	0.3914	0.4893	0.7903	0.4596	0.5488	0.7643

Model ID*	Training			Testing		
	MAE	RMSE	R^2	MAE	RMSE	R^2
$M_2D_2OP_1$	0.2463	0.3359	0.8867	0.3015	0.3981	0.8426
$M_2D_2OP_2$	0.3236	0.4022	0.8438	0.3455	0.4176	0.8473
$M_2D_2OP_3$	0.3580	0.4450	0.8048	0.3987	0.4798	0.8077
$M_2D_3OP_1$	0.2957	0.3758	0.8623	0.3035	0.3664	0.8945
$M_2D_3OP_2$	0.3619	0.4445	0.8085	0.3893	0.4649	0.8139
$M_2D_3OP_3$	0.3033	0.4184	0.8502	0.3339	0.4562	0.8260
$M_3D_1OP_1$	0.2776	0.3810	0.8542	0.4201	0.5869	0.7087
$M_3D_1OP_2$	0.3474	0.4194	0.8237	0.4154	0.5348	0.7919
$M_3D_1OP_3$	0.2780	0.3601	0.8861	0.3933	0.5410	0.7714
$M_3D_2OP_1$	0.2875	0.3694	0.8778	0.4987	0.6332	0.6999
$M_3D_2OP_2$	0.3514	0.4543	0.8147	0.5694	0.6959	0.7027
$M_3D_2OP_3$	0.3944	0.2205	0.7827	0.5219	0.6408	0.7401
$M_3D_3OP_1$	0.3213	0.3961	0.8609	0.5624	0.6556	0.6839
$M_3D_3OP_2$	0.2483	0.3230	0.9005	0.2910	0.3660	0.8882
$M_3D_3OP_3$	0.3907	0.4801	0.7800	0.3655	0.4582	0.8236

* M_1 , Demand; M_2 , Demand + Climactic; M_3 , Demand + Climactic + Socioeconomic; D_1 , τ (lag) = 1 month; D_2 , τ = 2 months; D_3 , τ = 3 months; OP_1 , {+, -, x}; OP_2 , {+, -, x, x2, x3}; OP_3 , {+, -, x, x2, x3, $\sqrt{}$, ex, log, ln}; R^2 , coefficient of determination; MAE, mean absolute error; RMSE, root mean square error.

Table 4. Performance of GEP models.

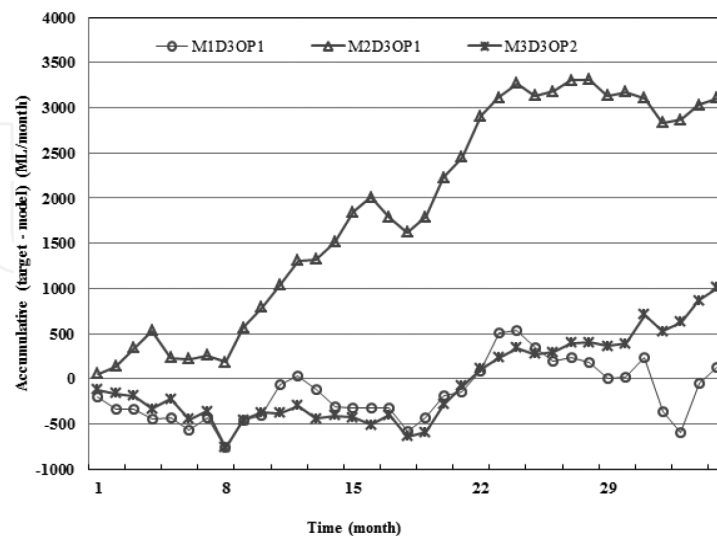


Figure 7. Cumulative (target-model) demand with time.

The superior GEP models from each classification were compared to SVM models implementing three different kernel functions (RBF, Poly, and Lin). Training and testing performance indices for the SVM models developed with each of the three kernel functions showed *Poly* kernel functions to outperform RBF and Lin functions (**Table 5**). The fact that Lin kernels performed poorly indicates that the nature of input parameters could not be considered using such functions. The M_2D_3 Poly model was selected as the superior SVM model to be compared with the GEP models (**Figure 8**).

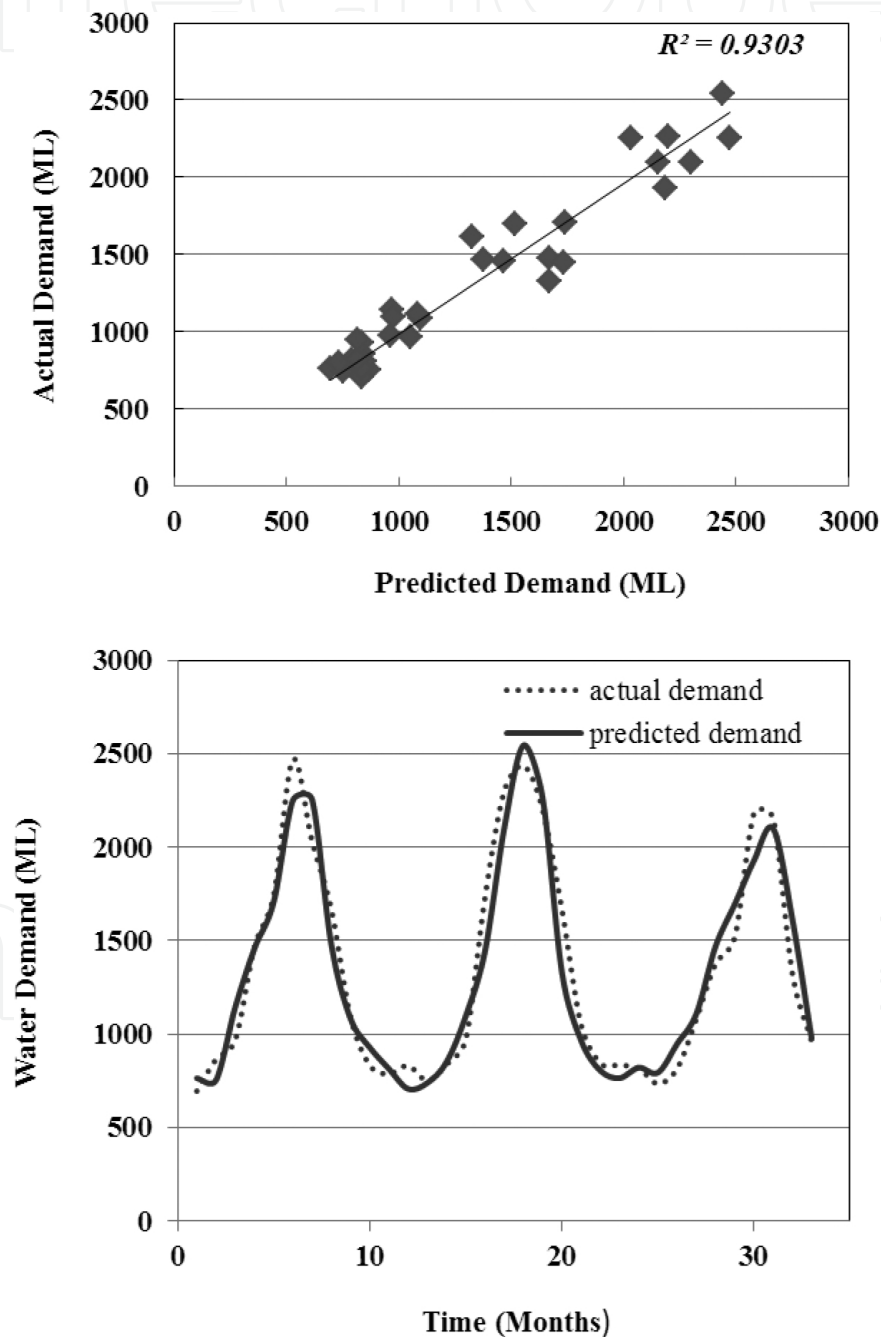


Figure 8. The best SVM model.

Model ID*	Training			Testing		
	R^2	RMSE	E	R^2	RMSE	E
M_1D_3RBF	0.9545	0.2123	0.9546	0.8397	0.4051	0.8387
M_2D_3RBF	0.9856	0.1201	0.9855	0.8701	0.3678	0.867
M_3D_3RBF	0.9416	0.2407	0.9415	0.9258	0.3014	0.9107
M_1D_3Poly	0.9308	0.2618	0.9309	0.8206	0.4278	0.8201
M_2D_3Poly	0.9372	0.2497	0.9371	0.9343	0.2593	0.9339
M_3D_3Poly	0.9428	0.239	0.9424	0.9279	0.3002	0.9114
M_1D_3Lin	0.7864	0.4602	0.7864	0.7945	0.4592	0.7927
M_2D_3Lin	0.8894	0.3311	0.8894	0.8977	0.323	0.8974
M_3D_3Lin	0.9093	0.2998	0.9004	0.9084	0.3344	0.8901

* M_1 , Demand; M_2 , Demand + Climactic; M_3 , Demand + Climactic + Socioeconomic; D_1 , τ (lag) = 1 month; D_2 , τ = 2 months; D_3 , τ = 3 months; RBF, Poly, Lin R^2 , coefficient of determination; RMSE, root mean square error; E , Nash-Sutcliffe coefficient.

Table 5. Performance of SVM models.

6. Conclusion

In an attempt to improve model prediction accuracy, a wide range of modeling techniques has been proposed by researchers over recent years in the water demand forecasting field. The present research explored a new approach to modeling water demand, namely genetic expression programming along with phase space reconstruction. In this method, input factors are not randomly chosen as in previous studies. Instead, appropriate lag time determinations made by the AMI method defined the structure of the explanatory variables employed in the models. The outcome of this research demonstrated GEP models to be highly sensitive to classification of input factors, proper lag time, and selection of genetic operators. In general, soft computing techniques like GEP should receive more attention in forecasting behaviors of complex systems such as WDS. These models can offer valuable information to WDS operators and designers to deploy optimum determinants in their forecast models. The three best GEP models proposed in this research were compared using different performance indices, however, differentiating between them was difficult due to the similarity in statistical index values. One of three GEP models was selected due to lower cumulative error in predicting demand and less fluctuation in comparison with the other two GEP models. However, these models were slightly outperformed by a SVM model, which showed even better performance indices. This shows that both GEP and SVM can be useful techniques in water demand forecasting and can account for nonlinearity of the input parameters

Acknowledgements

The authors received financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada. The Okanagan Basin Water Board and the City of Kelowna are thanked for providing water consumption data.

Author details

Sina Shabani¹, Peyman Yousefi², Jan Adamowski³ and Gholamreza Naser^{4*}

*Address all correspondence to: bahman.naser@ubc.ca

1 Ph.D. Candidate, School of Engineering, University of British Columbia, Kelowna, BC, Canada

2 Ph.D. Student, School of Engineering, University of British Columbia, Kelowna, BC, Canada

3 Associate Professor, School of Engineering, McGill University, Montréal, QC, Canada

4 Assistant Professor, School of Engineering, University of British Columbia, Kelowna, BC, Canada

References

- [1] Blokker M, Vloerbergh I, Buchberger S. Estimating peak water demands in hydraulic systems II – future trends. 14th Water Distribution Systems Analysis Conference, 24–27 September 2012; Adelaide. Australia: WDSA, 2012. p. 1138–1147.
- [2] Ghiassi M, Zimbra D, Saidane H. Urban water demand forecasting with a dynamic artificial neural network model. *Journal of Water Resources Planning and Management*. 2008; 134(2):138–146. DOI: 10.1061/(ASCE)0733-9496.
- [3] Troy P, Holloway D. The use of residential water consumption as an urban planning tool: a pilot study in Adelaide. *Journal of Environmental Planning and Management*. 2004; 47(1):97–114. DOI: 10.1080/0964056042000189826.
- [4] Koo JY, Yu MJ, Kim SG, Shim MH, Koizumi A. Estimating regional water demand in Seoul, South Korea, using principal component and cluster analysis. *Water Science & Technology: Water Supply*. 2005;5(1):1–7.

- [5] Miao, SP. A class of time series urban water demand models with non-linear climatic effects. *Water Resources Research*. 1990;26(2):169–178. DOI: 10.1029/WR026i002p00169.
- [6] Jain A, Varshney AK, Joshi UC. Short term water demand forecast modeling at IIT Kanpur using artificial neural networks. *Journal of Water Resources Management*. 2001;15(5):299–321. DOI: 10.1023/A:1014415503476.
- [7] Gato S, Jayasuriya N, Roberts P. Understanding urban residential end uses of water. *Water Science and Technology*. 2011;64(1):36–42. DOI: 10.2166/wst.2011.436.
- [8] Beck L, Bernauer T. How will combined changes in water demand and climate affect water availability in the Zambezi River Basin? *Global Environmental Change*. 2011;21(3):1061–1072. DOI: 10.1016/j.gloenvcha.2011.04.001.
- [9] Bakker M, Vreeburg JHG, van Schagen KM, Rietveld LC. A fully adaptive forecasting model for short-term drinking water demand. *Environmental Modelling and Software*. 2013;48:141–151. DOI: 10.1016/j.envsoft.2013.06.012.
- [10] Zhou SL, McMahon TA, Walton A, Lewis J. Forecasting daily urban water demand: a case study of Melbourne. *Journal of Hydrology*. 2000;236(3):153–164. DOI: 10.1016/S0022-1694(00)00287-0.
- [11] Mukhopadhyay A, Akber A, Al-Awadi E. Analysis of freshwater consumption patterns in the private residences of Kuwait. *Urban Water*. 2001;3(1–2):53–62.
- [12] Dos Santos CC, Pereira Filho AJ. Water demand forecasting model for the metropolitan area of São Paulo, Brazil. *Water Resources Management*. 2014;28(13):4401–4414. DOI: 10.1007/s11269-014-0743-7.
- [13] Zhou S, McMahon T, Walton A, Lewis J. Forecasting operational demand for an urban water supply zone. *Journal of Hydrology*. 2002;259(1–4):189–202. DOI: 10.1016/S0022-1694(01)00582-0.
- [14] Anderson R, Miller T, Washburn M. Water savings from lawn watering restrictions during a drought year in Fort Collins, Colorado. *Journal of the American Water Resources Association*. 1980;16(4):642–645. DOI: 10.1111/j.1752-1688.1980.tb02443.x.
- [15] Maidment D, Parzen E. Monthly water use and its relationship to climatic variables in Texas. *Journal of the American Water Resources Association*. 1984; 19(8):409–418.
- [16] Brekke L, Larsen M, Ausburn M, Takaichi L. Suburban water demand modeling using stepwise regression. *Journal of American Water Works Association*. 2002;94(10):65–75.
- [17] Polebitski A, Palmer R, Waddell P. Evaluating water demands under climate change and transitions in the urban environment. *Journal of Water Resources Planning and Management*. 2010;137(3):249–257. DOI: 10.1061/(ASCE)WR.1943-5452.0000112.
- [18] Lee SJ, Wentz EA, Gober P. Space-time forecasting using soft geostatistics: a case study in forecasting municipal water demand for Phoenix, Arizona. *Stochastic Environmen-*

- tal Research and Risk Assessment. 2010;24(2):283–295. DOI: 10.1007/s00477-009-0317-z.
- [19] Alhumoud J. Freshwater consumption in Kuwait: analysis and forecasting. *Journal of Water Supply Research and Technology. AQUA*. 2008;57(4):279–288. DOI: 10.2166/aqua.
 - [20] Vemuri VR. *Artificial Neural Networks: Forecasting Time Series*. Los Alamitos, CA: IEEE Computer Society Press; 1994.
 - [21] Crommelynck V, Duquesne C, Mercier M, Miniussi C. Daily and hourly water consumption forecasting tools using neural networks. *Proceedings of the AWWA's Annual Computer Specialty Conference*. Nashville, Tennessee; 1992. p. 665–676.
 - [22] Bougadis J, Adamowski K, Diduch R. Short-term municipal water demand forecasting. *Hydrological Processes*. 2005;19(1):137–148. DOI: 10.1002/hyp.5763.
 - [23] Jentgen L, Kiddler H, Hill R, Conrad S. Energy management strategies use short-term water consumption forecasting to minimize cost of pumping operations. *Journal of American Water Works Association*. 2007;99(6):86–94.
 - [24] Aly A, Wanakule N. Short-term forecasting for urban water consumption. *Journal of Water Resources Planning and Management*. 2004;130(5):405–410. DOI: 10.1061/(ASCE)0733-9496(2004)130:5(405).
 - [25] Alvisi S, Franchini M, Marinelli A. A short-term, pattern based model for water-demand forecasting. *Journal of Hydroinformatics*. 2007;9(1):39–50. DOI: 10.2166/hydro.2006.016.
 - [26] Wang X, Sun Y, Song L, Mei C. An eco-environmental water demand based model for optimising water resources using hybrid genetic simulated annealing algorithms. II: Model application and results. *Journal of Environmental Management*. 2009;90(8): 2612–2619. DOI:10.1016/j.jenvman.2009.02.009.
 - [27] Wu L, Zhou H. Urban water demand forecasting based on HP filter and fuzzy neural network. *Journal of Hydroinformatics*. 2010;12(2):172–184. DOI: 10.2166/hydro.2009.082.
 - [28] Msiza IS, Nelwamondo FV, Marwala T. Water demand prediction using artificial neural networks and support vector regression. *Journal of Computers*. 2008;3(11):1–8.
 - [29] Herrera M, Torgo L, Izquierdo J, Pérez-García R. Predictive models for forecasting hourly urban water demand. *Journal of Hydrology*. 2010;387(1):141–150.
 - [30] Brentan BM, Luvizotto E, Herrera M, Izquierdo J, Pérez-García R. Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics*. 2016; Article in Press.
 - [31] Adamowski J, Chan HF, Prasher SO, Ozga-Zielinski B, Sliusarieva A. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average,

- artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research*. 2012;48(1):W01528. DOI: 10.1029/2010WR009945.
- [32] Shiri J, Kim S, Kisi O. Estimation of daily dew point temperature using genetic programming and neural networks approaches. *Hydrology Research*. 2014;45(2):165–181.
- [33] Takens F. Detecting strange attractors in turbulence. In: Rand, D.A., Young, L.S. (Eds.), *Lectures Notes in Mathematics*, vol. 898. New York: Springer Verlag; 1981. p. 366–381.
- [34] Ferreira C. Mutation, transposition, and recombination: An Analysis of the Evolutionary Dynamics. *Proceedings of the 6th Joint Conference on Information Sciences (JCIS)*, North Carolina, USA. 2002. 614–617.
- [35] Ferreira C. *Gene expression programming: mathematical modeling by an artificial intelligence 2006*. (Vol. 21). Springer Berlin Heidelberg.
- [36] Fraser AM, Swinney HL. Independent coordinates for strange attractors from mutual information. *Physical Review A*. 1986;33(2):1134–1140.
- [37] Holzfuss J, Mayer G. An approach to error-estimation in the application of dimension algorithms. In: Mayer-Kress, G. (Ed.), *Dimensions and Entropies in Chaotic Systems*. New York: Springer; 1986. p. 114–122.
- [38] Hegger R, Kantza B, Schreiber T. Practical implementation of nonlinear time series methods: the TISEAN package. *Chaos*. 1999;9(2):413–435.
- [39] Khatibi R, Sivakumar B, Ghorbani MA, Kisi O, Kocak K, Zadeh DF. Investigation chaos in river stage and discharge time series. *Journal of Hydrology*. 2012;414–415:108–117.
- [40] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995; 20(3):273–297.
- [41] Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *Journal of Hydrology*. 2011;396:128–138.
- [42] Jafarzadeh AA, Pal M, Servati M, FazeliFard MH, Ghorbani MA. Comparative analysis of support vector machine and artificial neural network models for soil cation exchange capacity prediction. *International Journal of Environmental Science and Technology*. 2016;13(1):87–96.
- [43] Ghorbani MA, Khatibi R, Asadi H, Yousefi P. Inter-comparison of an evolutionary programming model of suspended sediment time-series with other local models. In: Ventura, S. (Ed.), *Genetic Programming – New Approaches and Successful Applications*; 2012. Rijeka, Croatia: Intech. p. 255–283. DOI: 10.5772/47801.

