Environmental Engineering and Management Journal September 2015, Vol.14, No. 9, 2061-2072
http://omicron.ch.tuiasi.ro/EEMJ/

"Gheorghe Asachi" Technical University of Iasi, Romania

LAYERED MACHINE LEARNING FOR SHORT-TERM WATER DEMAND FORECASTING

Antonio Candelieri[1]*, Davide Soldi[1], Francesco Archetti[1,2]

[1]*Consorzio Milano Ricerche, via Roberto Cozzi 53, Milan, Italy*
[2]*Department of Computer Science, Systems and Communication, University of Milano-Bicocca, viale Sarca 336, 20126, Milano, Italy*

Abstract

Water Distribution Networks (WDN) are large-scale systems whose management is a complex task, with increasing environmental and socio-economic implications worldwide, and with a renewed considerable attention by all stakeholders: local authorities, regulators, environmental groups and the scientific community.
WDM managers need to reliably estimate the water demand in the short-term (typically 1 day ahead), in order to operate their reservoirs and treatment plants appropriately to meet demand while reducing costs, in particular energy-related costs for caption, treatment and pumping.
In this paper the authors propose a fully adaptive, data-driven and self-learning approach to forecast short term urban water demand in two stages: *i)* identifying and characterizing typical daily consumption patterns (based, at least, on hourly demand data) and *ii)* dynamically generating a set of forecasting models for each typical pattern identified at the previous stage. This schema permits to deal with nonlinear variability of the water demand at different levels, automatically characterizing periodicity (e.g., seasonality) and behaviour-related differences among different types of days and hour of the day.
The approach has been validated on the urban water demand data acquired through the SCADA system of the Metropolitana Milanese (MM) partner, the urban water distribution utility in Milan, Italy.
Moreover, the approach has been developed in order to work also at individual (customer) level, exploiting the new available technological solutions for smart metering (i.e., Automatic Metering Readers, AMRs) now being installed in MM.

Keywords: Short-term water demand forecasting, Support Vector Machines regression, time-series clustering

Received: December, 2014; Revised final: August, 2015; Accepted: September, 2015

1. Introduction

A recent review about urban water demand forecasting is provided by Donkor et al. (2014): a wide variety of approaches have been proposed, their application differs according to the management objectives as well as the variable to be forecasted, its periodicity and the forecast horizon.

As *forecast periodicity* and *forecast horizon*, availability and choice of specific *determinants* can influence the selection of the forecasting approach to use. Anyway methods and models whose input variables can be easily collected, monitored and used by the utility should be preferred for practical application, reducing the risk to add noise/errors coming from data/information sources which are not under control.

Reliable urban water demand forecasting is the basis for supporting decision making at operational, tactical, and strategic levels (***planning level***) (Billings and Jones, 2008). These concern decisions for capacity expansion (strategic), investment planning (tactical) and system operation, management, and optimization (operations).

For instance, utilities need to know the expected water demand in order to optimize

* Author to whom all correspondence should be addressed: e-mail: candelieri@milanoricerche.it; Phone: +39 02 6448 2184

treatment plants planning as well as pump scheduling and storage, with the aim to reduce energy costs.

In terms of forecast horizon, water demand forecasting can be categorized as either ***long-term***, ***medium-term***, or ***short-term***, with these horizons being implied by the planning levels, respectively (Alvisi et al., 2007; Ghiassi et al., 2008; Jain et al., 2001).

Fig. 1 summarizes the association between the *forecasting horizon* and the *levels of planning*. Anyway a general consensus on a time frame for the three different horizons does not exist. Indeed new technological achievements, e.g. the adoption of new (smart) metering solutions, enabling high-rate sampling, blur the distinction among the three levels with relevant implication, in particular, on the *short-term horizon*.

In the United States, examples of the adoption of a control model based on a short-term water demand forecasts were validated already in 2009 (Bunn and Reynolds, 2009). Thus, the current context will support a wider deployment of short-term water demand forecasting solutions for water utilities. In the Netherlands the penetration of short-term forecasting models is expected to rise over 90% in 2016 (Bakker et al., 2014). Measuring forecast errors is crucial for the selection of an accurate and reliable forecasting model, as well to continuously evaluate the opportunity to update existing model(s) in order to reduce deviations in future forecasts.

The basic step consists in comparing forecasts with observations. The evaluation consists of four subsequent steps:

1) Dividing the data into an estimation period and a model validation period;

2) Using the model calibration period to build the forecast model;

3) Evaluating the forecast accuracy of the model(s) by comparing the forecasts with observations. This step may be performed on both the estimation period and the hold-out period, even if only the evaluation on the second data set usually provides an estimation of the accuracy of the forecast model(s) on future prediction;

4) Selecting the best model according to the evaluated performances.

A relevant taxonomy, valid for all the approaches reviewed in this state of the art, consists in distinguishing between linear and nonlinear methods (Romano and Kapelan, 2014; Zhang 2001). As water demand data shows varying type of nonlinearity, a linear method might not have sufficient flexibility to handle such nonlinearities.

Another general categorization distinguishes between ***bottom-up*** and ***top-down*** approaches (Chen and Boccelli, 2014). The first approach is devoted to model water usage at the level of individual users/meters, usually through stochastic processes (Blokker et al., 2009), while the second approach is devoted to model water usage at the level of whole system, usually with statistical tools (Magini et al., 2008). SCADA systems, currently widely adopted by water utilities, usually provide overall system or region-wide demand data in the form of time series, typically with resolution of 10-15 minutes. This kind of data naturally fit in the (time series) statistical modelling and forecasting framework. Nowadays, the availability of smart metering devices, such as Automatic Metering Readers (AMRs), makes possible the application of *bottom-up* approaches also at individual users/meters, even if with a lower resolution (usually 1 hour or, in the best case, 30 minutes).

The most relevant differentiation proposed by Donkor et al. (2014) refers to the inclusion/exclusion of exogenous variables to build the water demand forecasting model, concluding with some more general approaches, in particular Artificial Neural Networks (ANNs), which can be – and have been – applied in both the two mentioned cases. ***Stochastic Processes Models*** are more advanced than *Moving Average* and *Exponential Moving*, as they are able to model more complex profiles (rather than the eight generic exponential smoothing models presented by Mun (2010). The advantage of using stochastic process models is the ability to estimate the level of uncertainty associated with the forecast values.

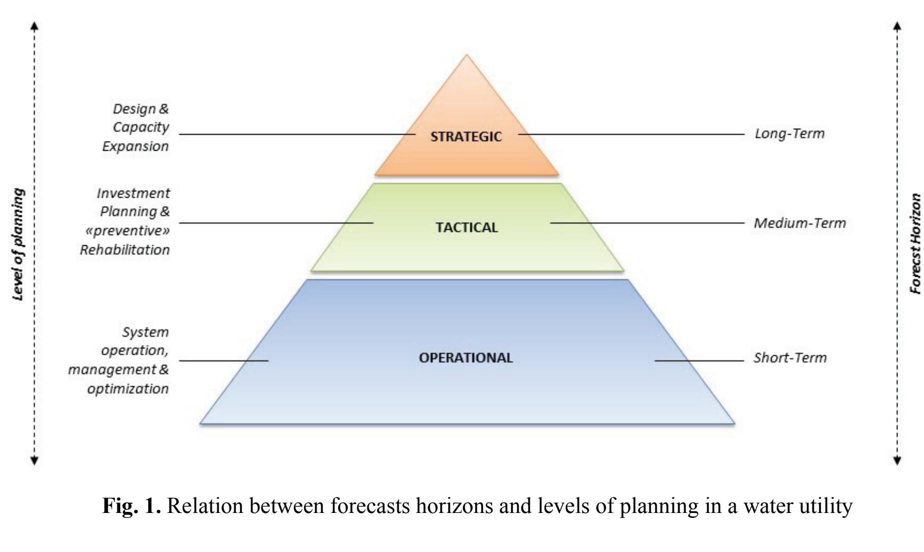

Fig. 1. Relation between forecasts horizons and levels of planning in a water utility

Time-series regression models are the most widely used to produce forecast by including exogenous variables. They have been used to forecast water demand according to its relationship with relevant determinants (Polebitski and Palmer, 2010).

However, demand at time *t* can be influenced by moving average and autoregressive terms, in addition to exogenous variables and their autoregressive terms. Other models, usually known as **ARMAX*(p,q,b)*** can be used, where a single exogenous variable is assumed.

With respect to the current water demand literature, the identification of the structure of ARMAX*(p,q,b)* models is done *ad-hoc*, usually with demand at time *t* modelled as a function of a selection of its previous values and those of its exogenous inputs variables. In a recent work by Adamowski and Karapataki (2010), determinants have been modelled on correlations coefficients, selecting variables such as peak water demand for the previous week, maximum temperature for the current and previous weeks, maximum temperature from two weeks ago.

A different approach, basically belonging to the class of regression models, is the ***scenario-based forecasting***. This approach is usually preferred in case uncertainty in demand forecasts has to be taken into account, relatively to a limited number of discrete combinations of the independent variables. Applications of this approach for urban water demand forecasting are reported by Polebitski et al. (2011) and Wei et al. (2010).

Currently, *IWR-MAIN Water Demand Management Suite*, developed by the U.S. Army Corps of Engineers' Institute for Water Resources, and *Demand-Side Management Least-Cost Planning Decision Support System*, created by Maddaus Water Management, Alamo, California, are two decision support tools providing scenario-based water demand forecasting. The demonstration of the IWR-MAIN has been reported by Mohamed and Al-Mualla (2010a, 2010b).

A strategy widely applied in water demand forecasting concerns ***composite*** (or ***hybrid***) ***models***. The general idea consists of combining forecasts provided by a set of different models which are able to capture different aspects of the relationship between water demand and its determinants. Usually, composite models report better forecasting performance for water demand (Wang et al., 2009) when compared to the application of the individual models.

The most widely used combination schemes are simple or weighted averages (Caiado, 2010; Wang et al., 2009), where the value of the weights is determined through optimization or least squares regression to minimize the mean squared error between the composite forecast and the actual data.

Wu and Zhou (2010) used linear regression to model the deterministic component of demand and Artificial Neural Networks (ANNs) to model the cyclical component. As result, the composite model offers more accurate forecasts with respect to those obtained from linear regression and ANN separately.

ANN is a machine learning approach widely applied to forecast water demand. Initially, conventional ANN models were applied (Joo et al., 2002; Jain et al., 2001) but, according to the advancements of the methodology, more complex and dynamic implementations have been proposed and applied (Ghiassi et al., 2008). As other machine learning strategies, ANN can be used to obtain both regression models and (univariate) time series models, according to the inclusion/exclusion of exogenous variables among the determinants.

Research papers about ANN for water demand forecasting typically involve a comparison of the performance between different ANN models and:

- more conventional regression models (Adamowski and Karapataki, 2010; Cutore et al., 2008; Firat et al., 2009; Herrera et al., 2010; Jentgen et al., 2007),
- (univariate) time series models (Ghiassi et al., 2008),
- or with both (Bougadis et al., 2005; Jain and Ormsbee, 2002; Jain et al., 2001).

Nonetheless, identifying the optimal architecture first requires the determination of the structure of a univariate time series or regression model.

Recently some advances have been achieved in the application of machine learning techniques for water demand forecast. In particular not only ANNs have been adopted in the last years, but also more effective and efficient strategies such as Support Vector Machine (SVM) regression (Herrera et al., 2014; Ji et al., 2014; Know et al., 2014; Sampathirao et al., 2014).

Analogously, composite (or hybrid) approaches gained a renewed interest from the application of advanced strategies devoted to optimize the parameters setting for a specific machine learning algorithm or to adequately combine information coming from different sources (e.g., different types of variables, including exogenous ones). Significant applications of strategies for the optimal parameters setting are related to the application of ***heuristic search algorithms***, such as evolution-based strategies (e.g., Genetic Programming, Genetic Algorithms, etc.). A relevant example is the approach proposed by Romano and Kapelan (2014) which uses Evolutionary Artificial Neural Networks (EANN) to implement a fully automated, data-driven and self-learning demand forecasting system. Other relevant heuristics use different paradigms inspired from other "real world intelligence" examples, such as Tabu Search, Ant/Bee Colony Optimization, and Particle Swarm Optimization. Recently a heuristic known as Teaching-Learning-based Optimization (TLBO), emulating the effect of a teachers to learners in a class (Rao et al., 2011), has been improved (Ameliorated TLBO, ATLBO) and used to optimally configure the parameters of a Least Square Support Vector Machine (LS-SVM), that is a an extension of SVM usually preferred in the case of large scale problem (Ji et al., 2014).

With respect to SVM regression, it proved to be the best computational model for forecasting hourly water demand when compared with other different approaches, such as ANNs, *Projection Pursuit Regression* (PPR), *Multivariate Adaptive Regression Splines* (MARS), *Random Forests* and *weighted pattern-based* water demand forecasting (Herrera et al., 2010). More recently, ***Multiple Kernel Learning*** has been proposed in order to improve accuracy of SVM for water demand forecasting. Herrera et al. (2014), proposed a *Multiple Kernel regression* (MKr) to extend SVM regression through a combination of different kernels from as many types as kinds of input data source are available.

Moreover, the paper focuses on water demand forecasting in the presence of a continuous source of information, proposing two different *on-line* learning MKr to continuously update the current forecasting model to a more accurate and reliable one, avoiding the computational efforts associated with the re-execution of the entire analytical process each time that new data are available. The two proposed approaches differs on the procedure to identify the time window to use for the analysis of data, that is *sliding* and *worm windows*; anyway the benefits of the overall approach lies in the possibility to adequately combine information coming from different data sources (e.g., weather, socio-economic factors, previous demand data, etc.).

With respect to the possibility to analyse original data and extract information at different levels of abstraction, ***Deep Learning*** represents one the most recent strategy in order to build hierarchies of data analysis and machine learning approaches.

Although it has not been strictly classified as deep learning, a recent approach proposed by Bai et al. (2014) adopts, at different level, signal analysis algorithms, machine learning and heuristic search. More in detail, wavelet transform is used to decompose historical time series of daily water supplies into different scales; at each scale the wavelet coefficients are used to train a *Relevance Vector Regression (RVR)* model. The Relevance Vector Machine (RVM) is a probabilistic machine learning method based on Bayesian theory and similar to SVM. RVM deals very well with nonlinear problem and time series with small samples; it is the core of the Multi-Scale Relevance Vector Regression (MSRVR) approach proposed by the authors. In addition, a particle swarm optimization algorithm is used to find the optimal tuning of the RVR parameters.

Fig. 2 summarizes a wider and updated representation of the state of the art approaches proposed/applied for the water demand forecasting, where the Machine Learning includes ANN. The contribution of the water demand forecasting approach proposed in this paper concerns, in particular, the ***short-term*** *forecast horizon* and the ***hourly*** *periodicity* (even if shorter time scale can be handled as well).

The water demand forecasting approach proposed in this paper has been designed and developed in order to be:

- completely data-driven
- fully adaptive
- provided with self-learning capabilities
- completely independent on the data source and, therefore, directly applicable to time series data (hourly water consumption, or higher sample rate) collected through both SCADA and AMR
- based on a two-stages learning (allowing its classification as a complex/hybrid approach or a deep/layered learning approach)

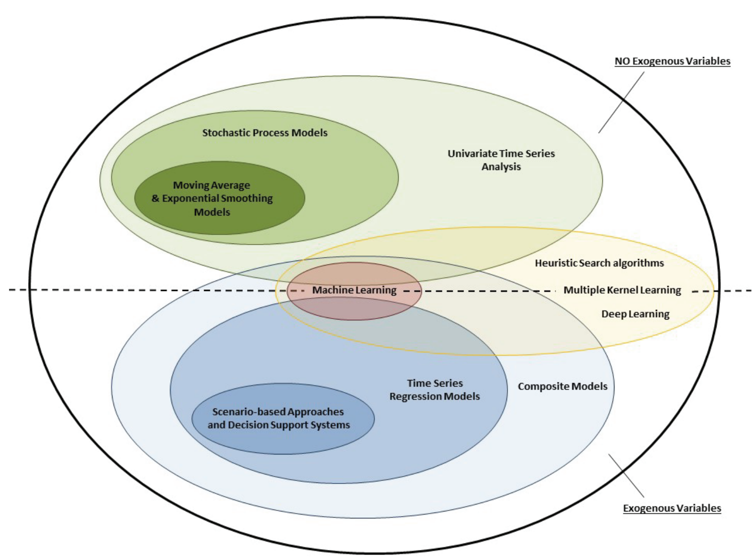

Fig. 2. Proposed extension and update of the representation of the current state of the art on water demand forecasting

As reported by Know et al. (2014), most of the existing demand forecasting models can be divided into two groups: those *modelling* the time series behaviour and those *predicting* it. One is devoted to model specific components such as periodicity (seasonality) and trends, the other one is, usually, an autoregressive model using short memory data and reproducing the underlying "generation" process of data. The main disadvantage is the limited predictability of water demand at sub-daily scale, due to the nonlinearities of the problem.

The approach presented in this paper addresses this limitation through a layered learning schema, a paradigm usually applied to tasks for which learning a direct mapping from inputs to outputs is intractable or results in not satisfactory accuracy. Given a hierarchical task decomposition into subtasks, layered learning seamlessly integrates separate learning at each subtask layer (Stone and Veloso, 2003). The application of machine learning in a subtask directly supports the learning of the next subtask layer. Usually three are the possible steps from a layer to the next one: *(i)* generating/reducing the set of training examples; *(ii)* modifying the input representation; and/or *(iii)* modifying the output representation. The proposed two-layer approach in particular addresses all the three issues by: *(i)* splitting the original training set into training subsets used in the second stage; *(ii)* modifying input representation by using different set of timely data in the two stages; and *(iii)* modifying output representation by using, in particular, clustering (unsupervised learning) at the first stage and regression (supervised learning) at the second stage.

The rest of the paper is organized as follows: the "Materials and methods" section describes the proposed approach and the error measures used to evaluate its accuracy; "Experimental" section gives an overview on the available set of data used in this study, finally "Results and discussion" section presents the relevant results obtained along with limitations and potential of the developed approach.

2. Material and methods

Rather than developing an ensemble of nonlinear models, as suggested by Know et al. (2014), a two-stage approach has been preferred in ICeWater, in order to deal with the nonlinearities at different time scale.

More in detail, at the first stage the daily water demand patterns (i.e., time-series of hourly data) are analysed in order to identify a limited set of typical behaviours, over the time period taken into account. When the time period is at least one year, possible periodicity, seasonality and trends may be automatically discovered. All the time series are grouped together according to their association to a specific typical behaviour, generating one time-series dataset for each behaviour. At the second stage, each dataset is separately analysed to obtain one specific ***daily*** *water demand forecasting model*, where, as better defined in the following, each daily water demand forecasting model consists of a set of ***hourly*** *water demand forecasting models*. The proposed approach permits to: *i)* identify and deal with (inter) variability among different water consumption behaviours and *ii)* model and exploit the (intra) variability in each typical behaviour to forecast the evolution of water demand over the day.

The approach considers as input only historical water demand data, without including any exogenous variables. It is completely data-driven and addresses one of the critical issues as perceived by the water utilities which are keen to use only variables collected, monitored and controlled by their technological systems (Donkor et al., 2014). The approach has been designed and developed to be applicable both at aggregated level (i.e., urban water demand data from SCADA) and at individual customers level (i.e., consumption data from AMRs). To be applicable on customers data, the proposed approach has been developed to be scalable on parallel/distributed architectures.

The proposed solution is based on two learning stages:

- clustering time-series of water demand data in order to identify typical daily consumption patterns
- training a Support Vector Machine (SVM) regression model for each cluster identified at the first stage and each hour of the day

Many different *error measures* have been proposed and adopted in the recent literature about water demand forecasting (Bai et al., 2014; Bakker et al., 2014; Herrera et al., 2014; Ji et al., 2014; Romano and Kapelan, 2014; Sampathirao et al., 2014), such as MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), Mean Squared Error (MSE), Root Mean Squared Error (RMSE). Other measures proposed in the forecasting literature, even if not largely used, are: ARE (Absolute Relative Error) and Average ARE (AARE), NRMSE (Normalised RMSE), CC (Correlation Coefficient) and ME (Mean Error).

For the purpose of this paper, the following widely accepted error measures have been considered: MAPE, RMSE and NS. It is important to highlight that MAPE might be the only error measure which can be used to compare forecasting performance among different utilities because it is independent of system capacity (Donkor et al., 2014).

With respect to the issue of model calibration, the entire approach has been developed in order to automatically capture changes in the behaviours and adapt the forecasting model, according to a self-learning paradigm.

The (self) re-learning of the entire system is performed at very low frequency, such as every month (and by taking into account at the least data of the latest year). To perform the updating of the system, both the stages have to be executed: the identification of typical consumption patterns (time series clustering) and the training of the pools of SVM regression models. Thus, the approach is able to automatically capture modifications of the typical water consumption behaviours, including changes to the number/shape of typical urban demand patterns.

Some relevant examples of similar approaches are related to the energy-grid field, where huge sets of high frequency data were already available some years ago, thanks to the wide adoption of smart metering devices for monitoring energy consumption (Hernández et al., 2013, 2014).

2.1. First stage: characterizing typical consumption behaviours through time-series clustering

A time-series data set consists of a set n time-series $V = \{v_1, v_2, \ldots, v_n\}$, where each time-series is represented as a vector (v_i) of l ordered values. The goal of time-series clustering is to identify structures in the unlabelled data set V by partitioning time series into disjoint groups, such that some measure of similarity is maximized within groups and minimized between groups. Specific surveys on time series clustering are provided by Liao (2005) and Kavitha and Punithavalli (2010). In time-series clustering one can work directly with ***raw data*** or preprocess them to perform ***feature extraction*** and then cluster data in the feature space.

The approach proposed in this paper allows working ***directly with the raw data***. All the time series to analyse are defined in the same time window (i.e., a day) and thus have the equal length (i.e., 24 data points in the case of hourly consumption data). The function used to measure similarity (or, analogously, distance) between 2 time-series objects is a critical component in the clustering algorithms. Classical measures from static data are used like Euclidean distance, Pearson correlation coefficient and cosine distance (Pereira and de Mello, 2013). To measure similarity, the proposed solution adopts ***clustering based on cosine similarity (type 1, similarity in time)*** – also known as triangle similarity (Zhang et al., 2011) – which handles the alignment of peaks and bursts.

More in detail, cosine similarity is given by the cosine of triangle between two vectors (Eq. 1), so the range of value of cosine similarity is [-1; 1].

$$s(v_1, v_2) = \frac{\langle v_1, v_2 \rangle}{\|v_1\|\|v_2\|} \quad (1)$$

where v_1 and v_2 are two vectors (with 24 components, in the specific case of daily time-series of hourly water consumption), < . , . > denotes the internal product between two vectors and || . || is the norm operator. A relevant consideration is that – in this case – triangle similarity may only varies in the range [0; 1] as the components of the urban water demand vectors are not negative.

To perform the time series clustering, the K-means algorithm has been experimentally shown to give good results, even in comparison with more sophisticated approaches such as Spectral or Markov clustering. In particular the implementation of the ***Spherical K-means*** provided by the R package "skmeans" (Maitra and Ramler, 2010) has been used as core of the approach developed. This specific implementation performs a simple K-means strategy based on the cosine distance (Eq. 2):

$$d(v_1, v_2) = 1 - s(v_1, v_2) = 1 - \frac{\langle v_1, v_2 \rangle}{\|v_1\|\|v_2\|} \quad (2)$$

A parameter that can be set by the manager/analyst is the overall time period of the observations to build the time-series dataset of daily water demand. Naturally, an adequate choice is to take into account at least one year in order to capture possible seasonality. In particular the clustering stage may be performed either on the entire dataset or at two levels. At the first level, the daily consumption time-series are averaged on each month and the resulting "monthly-averaged daily time-series" are preliminary clustered. At the second level, a new clustering is performed on the original daily consumption time-series but within each cluster obtained at the first level. This bi-level approach usually permits to improve the identification of seasonal behaviours within the period analysed.

The number of clusters is identified according to two cluster validity measures, which are Calinski-Harabatz and Silhouette (Arbeilatz et al., 2013); it is usually 2 or 3 for the first level and 2 for the second, when urban water demand data are analysed.

2.2. Second stage: Learning a water demand forecasting model through SVM regression

As results of the first stage (i.e., time-series clustering), a limited set of typical daily water demand behaviours (clusters) is identified, where the centroids are select as the "archetypal" daily water consumption behaviour/pattern for each cluster.

A possible relationship between each archetype and the time of its occurrence (e.g., period of the year and/or type of day) can be highlighted through the visualization of a calendar, with every day coloured according to the corresponding cluster; this supports the managers in the evaluation of possible seasonality, surprising periods, and consumption habits at different time scale.

To perform the second stage, each cluster is considered as a separate dataset. The first m columns are the input variables, and they correspond to the water demand values observed at the first m hours of the day (more generically, the first m samples in the data collection process). The output variable, to be predicted, is the j-th column of the original dataset, with $j = m+1, \ldots, 24$. For each j, a SVM regression model is trained, taking as input only the first m columns of the original dataset. Thus, $nClusters \times (24 - m)$ SVM regression models will be overall trained and used for forecasting, where *nClusters* is the number of clusters identified at the first stage; the *(24 – m)* SVM regression models related to a specific cluster are named "pool". The procedure is summarized, with respect to a specific cluster, in the following Fig. 3.

The two stages described in previous section are related to the "off-line" analysis of data, devoted to the dynamic, fully adaptive and data-driven procedure for learning a reliable and updated water demand forecasting model (consisting in a set of different pools of SVM regression models). This section concerns the usage of this model in order to "on-line" generate the forecast on water demand for the current day. When a new pattern of *m* daily demand values are collected by the system (generically SCADA or AMRs), the most suitable pool of SVMs is selected to produce forecast. This selection is performed – in the current solution – by involving directly the user who has to select the most appropriate pattern expected for the day to forecast.

Taking into account the results obtained in this study, and presented in detail in the next section, patterns are associated to two different characteristics: the ***period of the year*** (Spring-Summer, Fall-Winter, Summer-break) and the ***type of day*** (Working-day, Holiday-&-Weekend), according to the occurrence of each identified archetype over the observed period. The selected pool of SVMs is thus used to predict the water demand data at the remaining *24-m* hours.

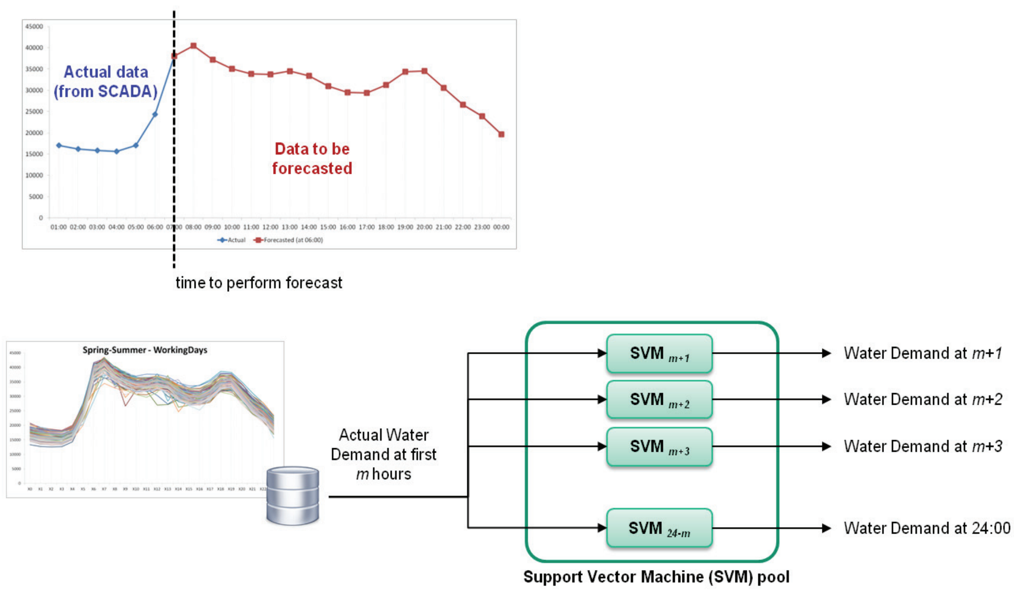

Fig. 3. Learning predictive models: one pool of SVM models for each typical pattern identified and each hour

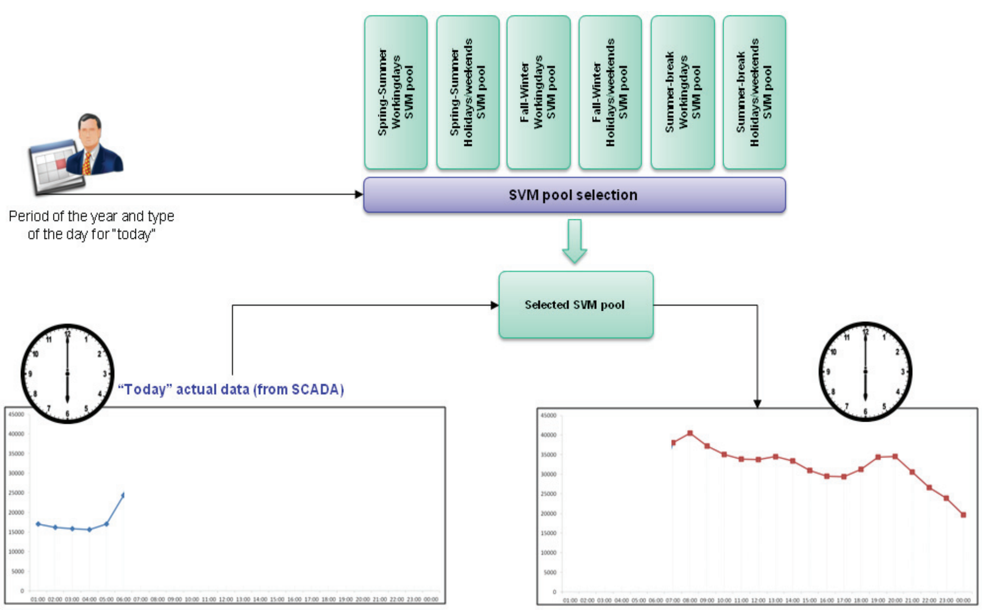

Fig. 4. Applying models learned: the most suitable pool of SVM models is selected and each model is used to forecast demand at each hour

Fig. 4 shows the on-line forecasting procedure, where demand values at the first m=6 hours of the day are considered as input of all the SVM regression models in the pools.

3. Experimental

The urban water demand data used in this study has been retrieved from the MM's SCADA, for the period 01 October 2012 to 30 September 2013, and are related to more than 5000 customers (buildings) for about 1million of inhabitants. Data has been organized into a time-series dataset, where each entry of the dataset is a vector of 24 measurements that are the hourly volume of water delivered over the day. As first step, a preliminary pre-processing of the retrieved data has been performed, aimed at identifying anomalies and replacing missing values. Anyway, this procedure affected only a very limited portion of data due to the reliability of the SCADA system.

Moreover, the urban WDN in Milan has a very low leakage level, thus distortions into the daily urban water demand time-series data are quite rare, making reliable the identification of typical daily consumption patterns.

4. Results and discussion

In this section the main results on the first learning stage are presented. The approach, at the end of the bi-level clustering procedure, identified 6 typical daily urban water demand patterns on MM's SCADA data, as reported in the following Fig. 5.

In the same Figure is possible to identify the occurrence of each cluster (i.e., archetype), over the analysed time windows. The following relative considerations have been made:

- 3 different periods of the year have been identified, namely Spring-Summer, Fall-Winter and Summer-break;
- 2 different types of day for each time period exist, namely, working-days and holidays-weekends.

Thus, every cluster is identified by the pair "*period of the year*" and "*type of the day*".

It is also really easy to note that major differences among the identified archetypes regard the peaks in consumption in the morning and in the evening. In particular, the peak in the morning of holidays and week-ends is always delayed of about 1 hour respect to that of working days, for each period of the year.

Moreover, the archetype named "Summer-break – working-days" is a really specific daily water demand pattern, more "flat" and "low" than the others, and associated to the 15 days in the middle of August, when usually citizens of Milan have their summer holidays and leave the city.

The identified clusters have been then used for training the SVM regression models by using the first m values of hourly consumption as input features. Different values of m have been considered in order to evaluate the impact on the forecasting error.

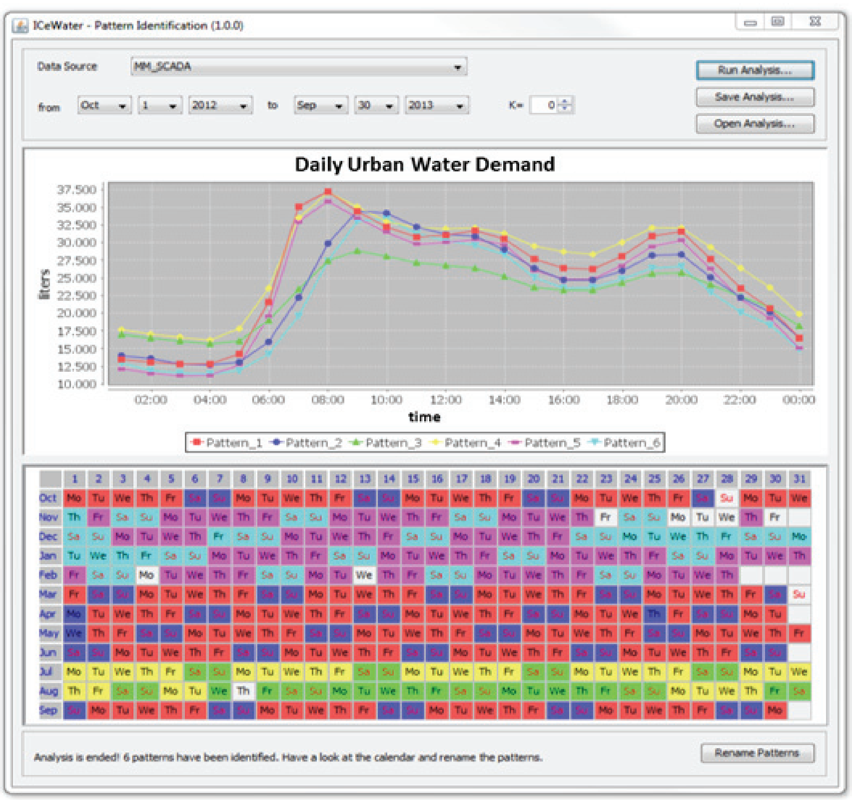

Fig. 5. Distribution of the identified clusters/patterns over the analyzed time window

For each *m*, a SVM regression has been trained for each cluster and for each hour of the day (from the *(m+1)*th to the *24*th hour of the day), that is the target variable.

Several possible configurations for each SVM regression model have been taken into account, using both Polynomial and Radial Basis Function (RBF) kernels. Optimal tuning of the parameter has been addressed in a different way for each kernel: degree of the Polynomial kernel varied from 1 to 5, while an iterative process has been used on gamma of the RBF kernel.

In particular, the possible interval for gamma is initially fixed in [0.001; 1000] and 10 values are selected by simple binning, meaning that the range of values is partitioned in 9 interval of the same size. Let g_i be and h_i be, respectively, the best value of gamma and the size of interval at iteration *i*, the next iteration is performed by considering the new interval $[g_i-h_i , g_i+h_i]$ and performing again the same simple binning procedure to identify the new values of gamma to be evaluated. The process stops when no more improvements in the forecasting performance are obtained (i.e. error decreases less than a given threshold).

It is important to highlight that this parameter optimization procedure is performed for each SVM regression model (on for each hour), independently. As result, a different kernel – in terms of type and/or value of the internal parameter – has been selected as optimal for each SVM regression model, modelling the different relationship existing between the consumptions at first hour of the day and the consumption at each one of the remaining hour of the day. Forecasting performances, measured as MAPE and NS, have been evaluated through ***leave-one-out validation***, in order to estimate the reliability of the predictions on new coming (urban) water demand time-series data.

In the following 2 tables (Table 1 and Table 2) the values of MAPE and NS are reported. Results obtained through SVM are compared to those provided by ANN, more conventionally used for generating water demand forecasting models.

In particular, the implementation of SVM regression and ANN provided by the open-source and Java-based suite WEKA (http://www.cs.waikato.ac.nz/ml/weka/citing.html) – Waikato Environment for Knowledge Analysis (Hall et al., 2009) – has been used. While the issue of optimal parameter tuning has been addressed for SVM through the iterative procedure aforementioned, the optimal setting of the ANN has been managed by WEKA, by enabling the automatic generation feature in the Multi-Layer Perceptron (MLP) algorithm.

Results are also differentiated with respect to the value of *m*. As many time-series are available, maximum, minimum and average measures are computed and compared.

Forecasting accuracy is quite higher for SVM than ANN. Furthermore, results may be compared with some others reported in the literature. In particular, the average MAPE for SVM ranges in 3.3% to 4.9%, that is lower – even in the worst case – than:

- 5.6% reported in (Romano and Kapelan, 2014);
- 5.3% with a time horizon of 1 hour (lead-1) and 10.2% with a time horizon of 2 hours (lead-2) reported in (Chen and Boccelli, 2014);
- 3.35% to 10.44%, computed on 6 different utilities (average value 6.19%), reported in (Bakker et al., 2013).

With respect to NS, SVM regression proved to be more accurate in forecasting water demand than ANNs. However, values of NS obtained in this study (0.8836 to 0.9286) are lower than 0.97 reported in (Romano and Kapelan, 2014), even if higher than 0.905 to 0.907 (average value on the 6 utilities, 0.906) reported in (Baker et al., 2013).

Finally, Figs. 6-8 show the wort, the best and the average case, respectively. It is easy to note that the worst case is associated to some anoumalous behavior potentially due to some unexpected modification in the consumption pattern during that specific day or a possible error in the data acquisition/transmission.

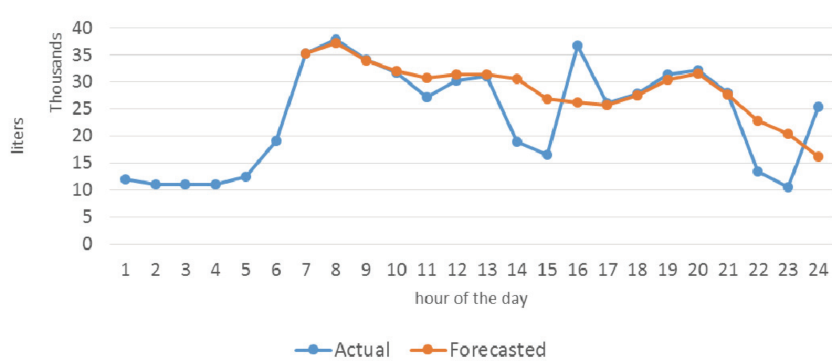

Fig. 6. Actual versus forecasted water demand (worst case)

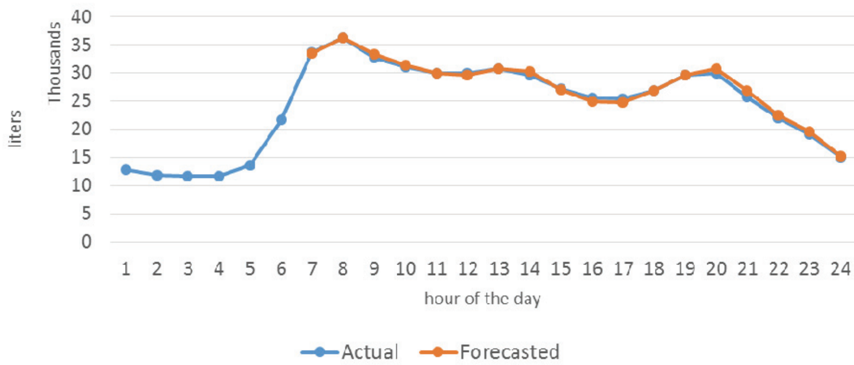

Fig. 7. Actual versus forecasted water demand (best case)

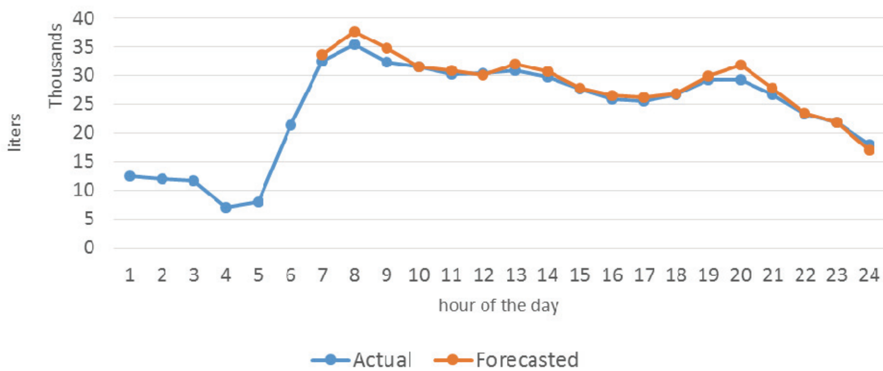

Fig. 8. Actual versus forecasted water demand (average case)

Table 1. MAPE obtained through SVM and ANN learning, for different values of m (%)

	m=3	m=4	m=5	m=6	m=7	m=8
ANN_MAPE_Max	28.8	30.3	30.9	34.2	33.8	30.6
ANN_MAPE_Min	1.4	1.3	1.2	1.1	1.1	1.1
ANN_MAPE_Avg	7.6	7.6	7.9	7.6	8.0	8.0
SVM_MAPE_Max	21.9	25.7	19.4	21.2	23.1	24.5
SVM_MAPE_Min	0.8	0.7	0.8	0.7	0.6	0.7
SVM_MAPE_Avg	4.5	4.9	4.5	3.8	3.5	3.3

Table 2. NS obtained through SVM and ANN learning, for different values of m

	m=3	m=4	m=5	m=6	m=7	m=8
ANN_NS_Max	0.9971	0.9967	0.9951	0.9950	0.9961	0.9959
ANN_NS_Min	-2.9764	-3.7410	-6.6258	-6.1000	-10.3513	-8.5048
ANN_NS_Avg	0.7673	0.7270	0.6162	0.5667	0.5317	0.5241
SVM_NS_Max	0.9982	0.9980	0.9974	0.9977	0.9978	0.9979
SVM_NS_Min	-0.6000	-1.0625	-0.9837	-0.1161	-0.0107	0.1582
SVM_NS_Avg	0.9178	0.8921	0.8836	0.9136	0.9258	0.9286

5. Conclusions

The approach for short-term water demand forecasting designed and developed within the ICeWater adopts a two-stage learning schema based on time-series data clustering (first stage) and Support Vector Machine for regression (second stage).

This completely data-driven, fully adaptive and self-learning approach has been designed and developed to be applicable both at aggregated level (i.e., urban water demand data from SCADA) and at individual customers level (i.e., consumption data from AMRs). The approach has been currently tested on real data retrieved from the SCADA system of Metropolitana Milanese, the WDN in Milan and one of the two use cases of ICeWater.

It is important to remark that accurate short-term water demand forecasting – even if at urban level – can effectively drive processes at the operational planning level, in particular the optimization of the operations planning aimed at reducing energy-related costs for caption, treatment, storage and distribution.

As future work, authors plan to use the same approach on individual water consumption data, acquired through Automatic Metering Readers (AMR), as well as in other application domain where behaviours may be inferred and predicted starting from available time series data, such as energy grid, public transportation and peri-urban traffic.

Aknowledgments

This work has been performed within ICeWater, which has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstrations under Grant Agreement No. 317624.

References

Adamowski J., Karapataki C., (2010), Comparison of multivariate regression and artificial neural networks for peak urban water demand forecasting: evaluation of different ANN learning algorithms, *Journal of Hydrological Engineering*, **15**, 729-743.

Alvisi S., Franchini M., Marinelli A., (2007), A short-term, pattern-based model for water demand forecasting, *Journal of Hydroinformatics*, **9**, 39-50.

Bai Y., Wang P., Li C., Xie J., Wang Y., (2014), A multi-scale relevance vector regression approach for daily urban water demand forecasting, *Journal of Hydrology*, **517**, 236-245.

Bakker M., Vreeburg J.H.G., van Schagen K.M., Rietveld L.C., (2013), A fully adaptive forecasting model for short-term drinking water demand, *Journal of Environmental Modelling & Software*, **48**, 141-151.

Bakker M., van Duist H., van Schagen K.M., Vreeburg J.H.G., Rietveld L.C., (2014), Improving the performance of water demand forecasting models using weather input, *Procedia Engineering*, **70**, 93-102.

Billing B., Jones C., (2008), *Forecasting Urban Water Demand*, 2nd Ed., American Waterworks Association, Denver, CO.

Blokker E., Vreeburg J., van Dijk J., (2009), Simulating residential water demand with a stochastic end-se model, *Journal of Water Resources Planning and Management*, **136**, 19-26.

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

Bunn S.M., Reynolds L., (2009), The energy-efficiency benefits of pump scheduling optimization for potable water supplies, *IBM Journal of Research and Development*, **53**, 5:1-5:13.

Caiado J., (2010), Performance of combined double seasonal univariate time series models for forecasting water demand, *Journal of Hydrological Engineering*, **13**, 215-222.

Chen J., Boccelli D.L. (2014), Demand forecasting for water distribution systems, *Procedia Engineering*, **70**, 339-342.

Cutore P., Campisano A., Kapelan Z., Modica C., Savic D., (2008), Probabilistic prediction of urban water consumption using the SCEMUA algorithm, *Urban Water Journal*, **5**, 125-132.

Donkor E.A., Mazzucchi T.A., Soyer R., Roberson J.A., (2014), Urban water demand forecasting: review of methods and models, *Journal of Water Resources Planning and Management*, **140**, 146-159.

Firat M., Yurdusev M., Turan M., (2009), Evaluation of artificial neural network techniques for municipal water consumption modeling, *Water Resource Management*, **23**, 617-632.

Ghiassi M., Zimbra D., Saidane H., (2008), Urban water demand forecasting with a dynamic artificial neural network model, *Journal of Water Resources Planning & Management*, **138**, 138-146.

Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I.H., (2009), The WEKA Data Mining Software: An Update, *SIGKDD Explorations*, **11**, 1.

Hernández L., Baladrón C., Aguiar J.M., Carro B., Sanchez-Esguevillas A.J., Lloret J., (2013), Short-term load forecasting for microgrids based on artificial neural networks, *Energies*, **6**, 1385-1408.

Hernández L., Baladrón C., Aguiar J.M., Carro B., Sánchez-Esguevillas A., Lloret J., (2014), Artificial neural networks for short-term load forecasting in microgrids environment, *Energy*, **75**, 252-264.

Herrera M., Torgo L., Izquierdo J., and Pérez-Garcìa R., (2010), Predictive models for forecasting hourly urban water demand, *Journal of Hydrology*, **387**, 141-150.

Herrera M., Izquierdo J., Pérez-Garcìa R., Ayala-Cabrera D., (2014), On-line learning predictive kernel models for urban water demand in a smart city, *Procedia Engineering*, **70**, 791-799.

Jain A., Varshney A., Joshi U., (2001), Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks, *Water Resource Management*, **15**, 299-321.

Jain A., Ormsbee L., (2002), "Short-term water demand forecast modelling techniques – conventional methods versus AI", *Journal of American Water Works Association*, **94**, 64-72.

Jentgen L., Kidder H., Hill R., Conrad S., (2007), Energy management strategies use short-term water consumption forecasting to minimize cost of pumping operations, *Journal of American Water Works Association*, **99**, 86-94.

Ji G., Wang J., Ge Y., Liu H., (2014), *Urban Water Demand Forecasting by LS-SVM with Tuning Based on Elitist Teaching-Learning-Based Optimization*, 26th Chinese Control and Decision Conference (CCDC 2014).

Joo C.N., Koo J.Y., Yu M.J., (2002), Application of short-term water demand prediction model to Seoul, *Water Science and Technology*, **46**, 255-261.

Kavitha V., Punithavalli M., (2010), Clustering time series data stream - a literature survey, *International Journal of Computer Science and Information Security*, **8**, 289-294.

Know H., So B., Kim S., Kim B., (2014), Development of ensemble model based water demand forecasting model, *Geographical Research Abstracts*, **16**. EGU2014-3711, On line: http://meetingorganizer.copernicus.org/EGU2014/EGU2014-3711.pdf.

Liao T.W., (2005), Clustering of time series data – a survey, *Pattern Recognition*, **38**, 1857-1874.

Magini R., Pallavicini I., Guercio R., (2008), Spatial and temporal scaling properties of water demand, *Journal of Water Resources Planning and Management*, **164**, 276-284.

Maitra R., Ramler I.P., (2010), A k-mean-directions algorithm for fast clustering of data on the sphere, *Journal of Computational and Graphical Statis*tics, **19**, 377–396.

Mohamed M., Al-Mualla A., (2010a), Water demand forecasting in Umm Al-Quwain (UAE) using the IRW-MAIN specify forecasting model, *Water Resource Management*, **24**, 4093-4120.

Mohamed M., Al-Mualla A., (2010b), Water demand forecasting in Umm Al-Quwain using the constant rate model, *Desalination*, **259**, 161-168.

Mun J., (2010), *Modeling Risk: Applying Monte Carlo Risk Simulation, Strategic Real Options, Stochastic Forecasting, and Portfolio Optimization*, 2nd Ed., Wiley, Hoboken, NJ.

Pereira C.M.M., de Mello R.F., (2013), TS-stream: clustering time series on data streams, *Journal of Intelligent Information Systems*, **42**, 531-566.

Polebitski A., Palmer R., (2010), Seasonal residential water demand forecasting for census tracts, *Journal of Water Resources Planning & Management*, **136**, 27-36.

Polebitski A., Palmer R., Waddell P., (2011), Evaluating water demands under climate change and transition in the urban environment, *Journal of Water Resources Planning & Management*, **137**, 249-257.

Rao R.V., Savsani, V.J., Vakharia D.P., (2011), Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems, *Computer-Aided Design*, **43**, 303-315.

Romano M., Kapelan Z., (2014), Adaptive water demand forecasting for near real-time management of smart water distribution system, *Environmental Modelling & Software*, **60**, 265-276.

Sampathirao A.K., Grosso J.M., Sopasakis P., Ocampo-Martinez C., Bemporad A., Puig V., (2014), *Water Demand Forecasting for the Optimal Operation of Large-Scale Drinking Water Networks: The Barcelona Case Study*, Proc. of the 19th International Federation of Automatic Control (IFAC) World Congress.

Stone P., Veloso M., (2003), *Layered Learning*, In: *Machine Learning: ECML 2000*, *Lecture Notes in Computer Science*, **1810**, Springer Berlin Heidelberg, 369-381.

Wang X., Sun Y., Song L., Mei C., (2009), An eco-environmental water demand based model for optimizing water resources using hybrid genetic simulated annealing algorithms II: model application and results, *Journal of Environmental Management*, **90**, 2612-2619.

Wei S., Lei A., Islam S., (2010), Modelling and simulation of industrial water demand of Beijing municipality in China, *Frontiers of Environmental Science and Engineering in China*, **4**, 91-101.

Wu L., Zhou H., (2010), Urban water demand forecasting based on HP filter and fuzzy neural network, *Journal of Hydroinformatics*, **12**, 172-184.

Zhang G., (2001), An investigation of neural networks for linear time-series forecasting, *Computational Operations Research*, **28**, 1183-1202.

Zhang X., Liu J., Du Y., Lv T., (2011), A novel clustering method on time series data, *Expert Systems with Applications*, **38**, 11981-11900.