Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 119 (2015) 844 – 853

Procedia Engineering

www.elsevier.com/locate/procedia

13th Computer Control for Water Industry Conference, CCWI 2015

Short-term forecasting of hourly water consumption by using automatic metering readers data

Antonio Candelieri[a,b] *, Davide Soldi[a], Francesco Archetti[a,b]

[a]Consorzio Milano Ricerche, via Roberto Cozzi 53, Milano, 20125, Italy
[b]Department of Computer Science, Systems and Communication, University of Milano-Bicocca, viale Sarca 336, Milano, 20126, Italy

Abstract

A completely data-driven, fully adaptive self-learning algorithm for water demand forecasting in the short-term and with hourly periodicity is proposed, according to the renewed interest generated by the availability of new technological solutions such as Automatic Metering Readers (AMR), a key enabler of the “Smart Water” paradigm. The approach is based on two sequential stages: at the first stage (time-series clustering) the daily water demand patterns (i.e., time-series of hourly data) are analysed to identify a limited set of typical behaviours. At the second stage Support Vector Machine regression is used to obtain one specific forecasting model (consisting of a regression model for each hour) for each cluster identified at the first stage. The approach has been validated on real data acquired by AMRs deployed on the Italian pilot site of ICeWater, computing the widely adopted error measure MAPE (Mean Absolute Percentage Error).

Peer-review under responsibility of the Scientific Committee of CCWI 2015

Keywords: Short-term demand forecasting; Time-series mining; Time-series clustering; Support Vector Machines regression

1. Introduction

Water Distribution Networks (WDN) are large-scale systems characterized by complex decision making activities. 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 energy-related costs for caption, treatment and pumping. In the United States, examples of the adoption of a control model based on a short-

* Corresponding author. Tel.: +39-02-6448-2185.
E-mail address: candelieri@milanoricerche.it

Peer-review under responsibility of the Scientific Committee of CCWI 2015
doi:10.1016/j.proeng.2015.08.948

term water demand forecasts were validated already in 2009. [1]. In the Netherlands the penetration of short-term forecasting models is expected to rise over 90% in 2016 [2].

Supervisory Control And Data Acquisition (SCADA) systems are currently widely adopted by water utilities, usually providing time-series data about overall system or region-wide demand, typically with a resolution of 10-15 minutes. Nowadays, the availability of smart metering devices, such as Automatic Metering Readers (AMRs), makes possible the application of water demand forecasting also at individual users/meters level, even if with a lower resolution (usually 1 hour or, in the best case, 30 minutes). These technological solutions have generated a renewed interest in a new generation of individual demand forecasting as enabler of the "Smart Water" paradigm.

1.1. State of the art of water demand short-term forecasting

A recent review about urban water demand forecasting is provided in [3]: a wide variety of approaches has been reviewed, their application differs according to the management objectives as well as the variable to be forecasted, the *forecast periodicity* and the *forecast horizon*. Furthermore, the availability and choice of specific determinants can influence the selection of the forecasting approach to use. However, methods and models whose input variables can be easily collected, monitored and used by the utility (i.e. consumption data) should be preferred for practical application, reducing the risk to add noise/errors coming from data/information sources which are not under control (e.g. weather forecast provided by external sources/services).

A possible classification of the current water demand short-term forecasting approaches consists in distinguishing between *linear* and *nonlinear* methods [4]. Usually linear methods are not so effective due to the intrinsic nonlinearity into water demand data. Indeed, as reported in [5], 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 most relevant differentiation proposed in [3] 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. ANN is a machine learning approach widely applied to forecast water demand. Research papers about ANN for water demand forecasting typically involve a comparison of the performance between different ANN models and more conventional regression models [6][7][8][9] as well as (univariate) time series models [10]. Furthermore, hybrid approaches have been also proposed, in [11] linear regression is used to model the deterministic component of water demand and Artificial Neural Networks (ANNs) are used 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.

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 [5][12][13][14]. Analogously, meta-heuristic based approaches gained a renewed interest from their application to optimize the parameters setting for a specific machine learning algorithm, such as evolution-based strategies (e.g., Genetic Programming, Genetic Algorithms, etc.). A relevant example is the approach proposed in [4] which uses Evolutionary Artificial Neural Networks (EANN). Recently a heuristic known as Teaching-Learning-based Optimization (TLBO), emulating the effect of a teachers to learners in a class [15], 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 [15].

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 [9]. More recently, Multiple Kernel Learning has been proposed in order to improve accuracy of SVM for water demand forecasting. In [12] a Multiple Kernel regression (MKr) has been proposed 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 computation-al 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 in [16] 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.

1.2. Forecast error measures

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, possibly by using a wide sub-set of the available data to learn/build the forecasting model and the remaining sub-set of data to validate it. The most widely adopted error measures is MAPE [2][4][12][13][14][16]. Let denote with:

Y – time-series of observed water demand (at any forecast periodicity)
Y_t – water demand observed at the time t
$\bar{Y}$ – average of the water demand observed
$\hat{Y}$ – time-series of forecasted water demand (at any forecast periodicity)
$\hat{Y}_t$ – water demand forecasted at the time t
N – time-series length

the main forecasting error measures are the following:

the Mean Absolute Percentage Error (MAPE) is computed as the average of the absolute values of the difference, in percentage, between forecasted and observed data on each time stamp; its main advantage is the independence on the unit measure.

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \tag{1}$$

It is important to highlight [3] 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.

1.3. Contribution of this paper

This paper presents a new short-term forecasting approach (24 hours horizon) with hourly periodicity (even if shorter time scale can be handled as well), designed and developed in order to be:

- completely data-driven; it considers as input only historical water demand data, without including any exogenous variables, addressing one of the critical issue as perceived by the water utilities which are keen to use only variables collected, monitored and controlled by their technological systems [3].
- fully adaptive and 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, that are: 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 previously 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 [17]. In this paper, results provided by its validation on a set of 26 AMRs (individual consumption) installed in the same pilot are reported.

2. Materials and Methods

2.1. Available data

Due to some delay in the installation and technical validation of the deployment, a relevant set of data has been collected only for the Italian pilot of the project ICeWater, at the time of analysis reported in this paper. In particular, the available set of data – after the removal of the first period needed for calibration and technical tests – is related to the water consumption data in the period September – December 2014. A selection has been performed in order to remove anomalous time-series due to reported malfunctioning of the devices: at the end of the process the available dataset consisted of 26 AMRs, with 110 time series of daily water demand, for each AMR, with an hourly resolution.

Contrary to the previous work on urban water demand [17], the limited size of the dataset did not permit to identify any seasonality, and only two different patterns have been identified, associable to working days and holidays/week-ends. The two following sub-sections will present the process of learning and usage of forecasting models, respectively.

2.2. Learning the forecasting models

The learning process in IceWater has been organized in two stages, in order to deal with the nonlinearities at different time scales. At the first stage (time series clustering) the daily water demand patterns (i.e., time-series of hourly data) are analysed in order to identify a limited set of typical daily (consumption) behaviours. As result of this first stage, 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 (i.e. SVM regression models).

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, when possible). To perform the updating of the system, both the stages have to be executed.

2.2.1. 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 a given measure of similarity is maximized within groups and minimized between groups [18]. 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 works 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, l=24). As similarity measure the *cosine similarity* – also known as *triangle similarity* – has been adopted: it is a "*similarity in time*" measure which handles the alignment of peaks and bursts [19]. More in detail, cosine similarity is given by the cosine of triangle between two vectors, so the range of value of cosine similarity is [-1; 1].

$$s(v_1, v_2) = \frac{\langle v_1, v_2 \rangle}{\|v_1\| \|v_2\|} \tag{2}$$

where v_1 and v_2 are two vectors, < . , . > 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 *Spherical K-means* provided by the R package "skmeans" [20] has been used, which performs a simple K-means strategy based on the *cosine distance*:

$$d(v_1, v_2) = 1 - s(v_1, v_2) = 1 - \frac{\langle v_1, v_2 \rangle}{\|v_1\| \|v_2\|} \tag{3}$$

The number of clusters is identified according to two cluster validity measures, Calinski-Harabatz and Silhouette [21].

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

As results of the first stage, 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. At the end of this step, 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 "labelled" 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 (i.e. the hourly water demand values observed at the first m hours of the day) are the input variables while 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; all the SVM regression models have the same input that is the first m columns of the original dataset. Thus, K x $(24 - m)$ SVM regression models will be overall trained and used for forecasting, where K is the number of clusters identified at the first stage. The procedure is summarized in the following figure.

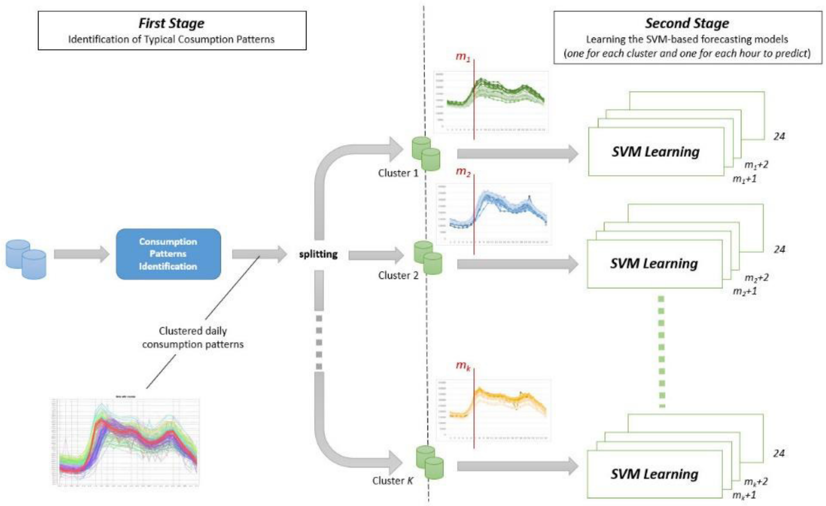

Fig. 1. Learning predictive models: one pool of SVM models for each typical pattern identified and each hour.

2.3. Performing forecasts

The two stages described in the previous section are only related to the "off-line" analysis of time-series data, devoted to the dynamic, fully adaptive and data-driven procedure for learning reliable and updated water demand forecasting models. The "on-line" usage of these models concerns how to generate the forecast on water demand for the current day. When a new pattern of *m* daily demand values are collected by the system, the most suitable pool of SVMs is selected to produce forecast. Currently, the retrieval of the most suitable pool is performed 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.

3. Results

As already anticipated, only two clusters for each AMR has been identified at the first stage of the learning process: limited size of the dataset did not permit to identify any seasonality; the two "archetypes" identified are associable to working days and holidays/week-ends, respectively The following table summarizes the average MAPE computed on the entire set of AMRs. With respect to the aggregated data [17] the MAPE is significantly higher, mainly due the limited size of the available dataset (about 100 valid time-series for each AMR) which produced significantly different errors across the AMRs. Indeed, it is important to highlight that a forecasting model is learnt for each AMR, according to the different behaviours that the correspondent customer may have. Thus, AMRs showing high variability in terms of different behaviours may produce, in some cases, a huge forecasting errors. To possibly overcome this limitation, more data are needed to capture and model this wide variability.

Table 1. Average MAPE, with respect to *m*, on AMRs data overall.

m	MAPE
3	65,32%
4	65,79%
5	65,12%
6	66,52%
7	67,08%
8	67,33%

Since *m=6* has been identified, along with the water utility, as a suitable value in order to have a reliable forecast while giving sufficient time for making decisions for the current day, in the following figure the value of MAPE on each AMR is reported according to *m=6*. It is easy to note that for most of the AMRs, the MAPE is lower than 66.52% (i.e. the overall average value for *m=6*, as reported in the previous table 1).

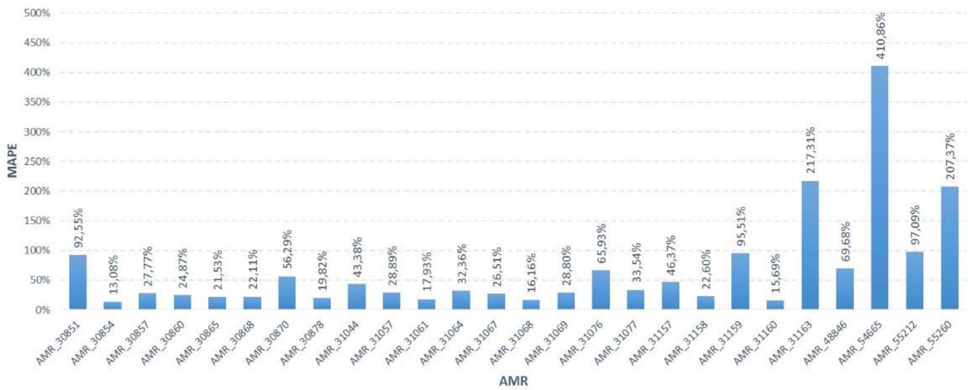

Fig. 2. MAPE for each AMR

The following table better summarizes these results: the MAPE is lower than 30% for the 50% of the AMRs.

Table 2. MAPE distribution on AMRs.

MAPE	Percentage of AMRs
<=30%	50%
>30% and <=50%	15%
>50% and <=100%	23%
>100%	12%

Finally, in the following four figures the best and worst forecasts are reported for the two AMRs associated to the minimum (AMR_30854) and maximum (AMR_54665) value of MAPE, respectively.

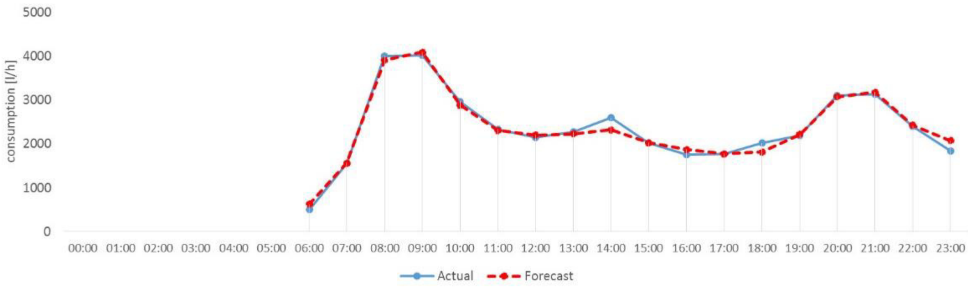

Fig. 3 AMR_30854 (minimum MAPE in Fig. 2) best case: actual vs forecasted demand

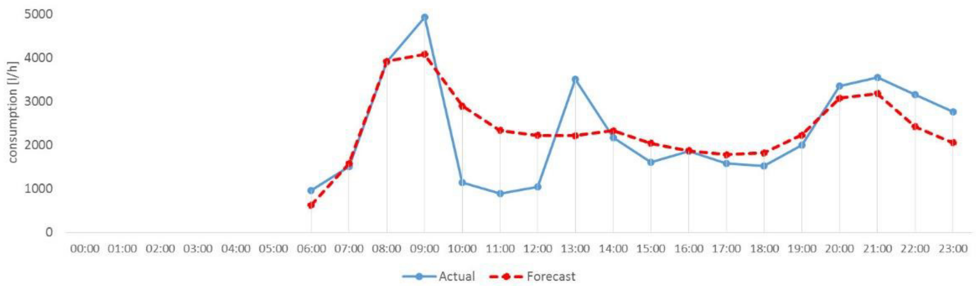

Fig. 4 AMR_30854 (minimum MAPE in Fig. 2) worst case: actual vs forecasted demand

According to the two previous figures, the error seems to be associated to a relevant variation in terms of consumption behaviour by the customer: hourly consumptions forecasted for the time window from 09:00 to 13:00 seem to be moved before (from 08:00 to 09:00) and after (from 13:00 to 14:00) and, partially, in the evening.

The availability of more data, for example one year of consumption data, could help to identify, automatically, if this is a cyclic behaviour – a typical pattern – or a completely casual variation in consumption pattern just for that day (anomaly).

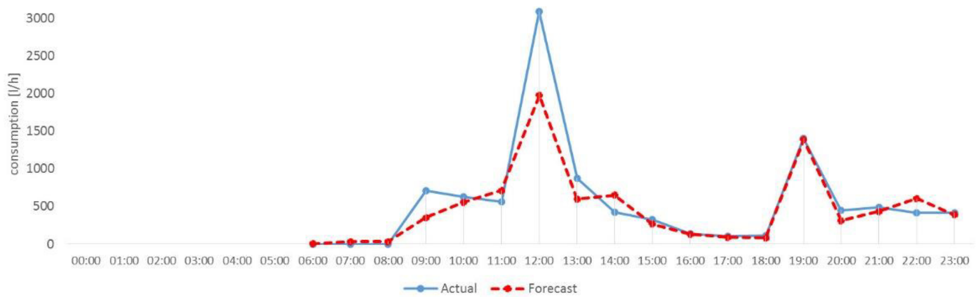

Fig. 5. AMR_54665 (maximum MAPE in Fig. 2) best case: actual vs forecasted demand

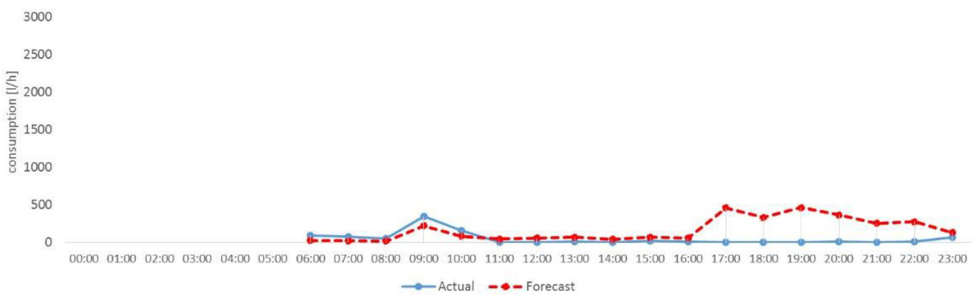

Fig. 6. AMR_54665 (maximum MAPE in Fig. 2) worst case: actual vs forecasted demand

According to the two previous figures, it is possible to observe that the error in the worst case is related to the prediction of consumption in the evening that did not actually occur. It is important to highlight that this specific AMR seems to refer to a customer that is not a household: the volume of hourly consumption is completely different in the best and worst cases, which are a working day and a holiday/week-end, respectively. Although the MAPE in the worst case is really high, the error in terms of overall volume of water is not so wide. Also in this case, the availability of a larger set of data could help to identify, automatically, if the worst case can be associated to a cyclic behaviour – and therefore a typical pattern – or a completely casual variation in consumption pattern just for that day (anomaly).

4. Conclusions

The proposed approach for short-term water demand forecasting 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 already 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, and currently validated at individual customer level. 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. In the case of individual consumption, a reliable forecast is useful in order to improve effectiveness of hydraulic simulation and decision making based on what-if scenarios. On the other hand, it is a useful tool in order to improve demand side management, both through information to individual customer, about its forecasted consumption, and for the possible design of alternative tariff schemes, in particular Time-Of-Use (TOU) tariffs.

The approach proved to be reliable also on individual data, even if a wider set of data – currently collected at the ICeWater pilots – is needed to make results more concrete and statistically sound.

Acknowledgements

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

References

[1] Bunn S.M., and Reynolds L. (2009), The energy-efficiency benefits of pump scheduling optimization for potable water supplies, IBM Journal of Research and Development, 53(3).
[2] Bakker M., van Duist H., van Schagen K.M., Vreeburg J.H.G., and Rietveld L.C. (2014), Improving the performance of water demand forecasting models using weather input, 12th International Conference on Computing and Control for the Water Industry (CCWI 2013), Procedia Engineering 70, 93-102.
[3] Donkor E.A., Mazzucchi T.A., Soyer R., and Roberson J.A. (2014), Urban Water Demand Forecasting: Review of Methods and Models, Journal of Water Resources Planning and Management, 140, 146-159.
[4] Romano M., and Kapelan Z. (2014), Adaptive water demand forecasting for near real-time management of smart water distribution system, Environmental Modelling & Software, 60, 265-276.
[5] Know H., So B., Kim S., and Kim B. (2014), Development of ensemble model based water demand forecasting model, Geographical Research Abstracts, 16.
[6] Adamowski J., and 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, 729-743.
[7] Cutore P., Campisano A., Kapelan Z., Modica C., and Savic D. (2008), Probabilistic prediction of urban water consumption using the SCEMUA algorithm, Urban Water Journal, 5(2), 125-132.
[8] Firat M., Yurdusev M., and Turan M. (2009), Evaluation of artificial neural network techniques for municipal water consumption modeling, Water Resource Management, 23(4), 617-632.
[9] Herrera M., Torgo L., Izquierdo J., and Pérez-Garcìa R. (2010), Predictive models for forecasting hourly urban water demand, Journal of Hydrology, 387(1-2), 141-150.
[10] Ghiassi M., Zimbra D., and Saidane H. (2008), Urban water demand forecasting with a dynamic artificial neural network model, Journal of Water Resources Planning & Management, 138, 138-146.
[11] Wu L., and Zhou H. (2010), Urban water demand forecasting based on HP filter and fuzzy neural network, Journal of Hydroinformatics, 12(2), 172-184.

[12] Herrera M., Izquierdo J., Pérez-Garcìa R., and Ayala-Cabrera D. (2014), On-line learning predictive kernel models for urban water demand in a smart city, 12th International Conference on Computing and Control for the Water Industry (CCWI 2013), Procedia Engineering, 70, 791-799.

[13] Ji G., Wang J., Ge Y., and 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).

[14] Sampathirao A.K., Grosso J.M., Sopasakis P., Ocampo-Martinez C., Bemporad A., and Puig V. (2014), water demand forecasting for the optimal operation of large-scale drinking water networks: The Barcelona Case Study, In Proceedings of the 19th International Federation of Automatic Control (IFAC) World Congress.

[15] Rao R.V., Savsani, V.J., and Vakharia D.P. (2011), Teaching-learning-based optimization: a novel method for constrained mechanical design optimization problems, Computer-Aided Design, 43(3), 303-315.

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

[17] Candelieri A., Conti D., and Archetti F. (2014), Identifying typical urban water demand patterns for a reliable short-term forecasting – The ICeWater project approach, 16th Conference on Water Distribution System Analysis, WDSA, 2014 (Procedia Engineering).

[18] Kavitha V., and Punithavalli M. (2010), Clustering Time Series Data Stream – A Literature Sur-vey, (IJCSIS) International Journal of Computer Science and Information Security, 8, 1.

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

[20] Maitra R., and Ramler I. P. (2010). A k-mean-directions algorithm for fast clustering of data on the sphere. Journal of Computational and Graphical Statistics, 19(2), 377–396.

[21] Arbelaitz O., Gurrutxaga I., Muguerza J., Pérez J., and Perona I. (2013), An extensive comparative study of cluster validity indices, Pattern Recognition, 46(1), 243-256.