Hybrid regression model for near real-time urban water demand forecasting☆

Bruno M. Brentan[a], Edevar Luvizotto Jr.[a], Manuel Herrera[b], Joaquín Izquierdo[c,*], Rafael Pérez-García[c]

[a] *LHC - FEC, University of Campinas, Campinas, Brazil*
[b] *EDEn - Dept. of Architecture and Civil Eng., University of Bath, Bath, UK*
[c] *Fluing - IMM, Universitat Politècnica de València, Valencia, Spain*

Abstract

The most important factor in planning and operating water distribution systems is satisfying consumer demand. This means continuously providing users with quality water in adequate volumes at reasonable pressure, thus ensuring reliable water distribution. During the last years, the application of Statistical, Machine Learning and Artificial Intelligence methodologies has been fostered for water demand forecasting. However, there is still room for improvement and new challenges concerning to on-line predictive models for water demand have aroused. This work proposes applying support vector regression, as one of the currently better Machine Learning options for short-term water demand forecasting, to build a base prediction. Over this model, a Fourier time series process is built to improve the base prediction. This addition produces a tool able to get rid of part of the errors and bias inherent to a fixed regression structure in response to new incoming time series data. The final hybrid process is validated using demand data from a water utility in Franca, Brazil. Our model, being a near real-time model for water demand, may be directly exploited in water management decision-making processes.

Keywords: Demand forecasting, Water supply, Fourier series, Support vector regression, Near real-time algorithms

☆This work is partially supported by CAPES Foundation of Brazil's Ministry of Education. The data were provided by SABESP, São Paulo state water management company.

*Corresponding author: jizquier@upv.es (J. Izquierdo)

2010 MSC: 00-01, 99-00

1. Introduction

Long-term water demand forecasting is required for planning and designing new water distribution systems (WDSs) [1]. However, to satisfy safe operation and management of their systems and make right decisions about valve and pump manoeuvres, water utilities need to be acquainted with real (or near real) time end-users behaviour regarding water consumption. In addition, having a deep knowledge on water demand helps identify and control possible leakages in the network, when observed consumption and demand prediction diverge far from the expected uncertainty [2, 3].

Auto regressive integrated moving average (ARIMA) based models [4] have been traditionally considered for understanding and modelling urban water demand [5]. ARIMA approach models usually treat the problem as a linear correlation among variables and according to Voitcu and Wong (2006) [6] this technique does not always produce predictions with sufficient accuracy, which can harm other processes, such as the control of the system. To cope with this situation, a number of data analysis models have been considered more recently. For instance, several authors [7, 8, 9] have applied artificial neural networks (ANNs) architectures to both long and short term demand forecasting. The use of other Machine Learning tools has also increased during the last years. [10] has performed a comprehensible comparison of various predictive methods for hourly water demand forecasting, suggesting the use of support vector regression (SVR) as one of the models through which it is possible to reach better results. These results were in agreement with the ones previously obtained by Msiza et al. (2007) [11], just comparing SVR against ANNs architectures for daily water demand prediction. However, the straightforward use of SVR without any internal adaptation when new data arrives (off-line SVR) has two main drawbacks: the computational complexity of the training and validation phases, necessary for parameter tuning, and the selection of the best model among the validated proposals [12, 13]. These two phases are highly time-consuming and lead to slow down the whole process, since it is strongly dependent on its associate database size. In addition, for off-line predictive models is likely the existence of certain growing bias, if models are not updated with the arrival of new data. The model can also become rapidly obsolete in the case of abrupt changes occurring in the forecasting framework. These are the models known as intervened [14], and are a consequence of unexpected changes in the scenario in which the demand is computed. For example, opening and closing valves, extreme variation of weather conditions, existence of new leakages or celebration of a social events, among others, may change the end-user response regarding water demand.

The majority of hydraulic models proposed in the literature are off-line, addressing a number of purposes such as network design optimisation [15], strategy planning [16], and setting optimal pumping schedule to reduce energy [17]. However, off-line models do not represent well the current state of the water supply system for operational purposes, especially in emergency events [18, 16]. The new paradigm of on-line modelling in WDSs is a topic of growing interest with the high amount of data information with which water utilities operate nowadays, aiming at making decisions in a very short time [15]. On-line predictive models for water demand forecasting [19, 20] emerge to bridge the gap between this constant flow of available information and off-line models, which are not optimised to be updated in near real-time. Through on-line models for water demand, it is possible to improve predictions of water demand and to have better control of such system state variables as flow and pressure [19], by suitable valve operation.

The aim of on-line prediction is to update the current model to a more accurate one, avoiding the computational burden associated with re-calculating the whole process each time new data are available. Vaerenbergh et al. [21] propose a sliding windows methodology for kernel regression together with a fast optimisation of its associate parameters. Other approaches are based on a fusion of methods for model updating [22]. In this sense, a number of hybrid methodologies for predictive models have been introduced in the literature. These are usually based on ARIMA models combined with alternative models attempting to capture forecasting non-linearities [23]. In general, this fusion of models increases model accuracy and reduces overfitting problems. For example, Aladag et al. (2009) [24] proposed a hybrid model linking ARIMA and ANN models; this hybrid model was applied to yearly demand forecast providing accurate results. A similar approach is used in [14] modelling water demand intervened (e.g. by open/close valve manoeuvres). In [25] a hybrid Particle swarm optimisation - Support vector machine (PSO-SVM) based model is built as a predictive model of chemical components concentration in water.

Our working proposal starts by choosing a suitable methodology among the standard Machine Learning options for regression analysis and run it for short-term water demand forecasting. In this case, we have selected SVR as one of the models that provides better results for water demand forecasting [11, 26]. Built on this model, an on-line process based on Fourier time series is launched to improve the predictions.

The time series error associated with the SVR model is subsequently investigated through and adaptive Fourier series (AFS) technique, enabling to model any possible variation on the original pattern that could arise with the arrival of new data. The combination of these two perspectives endows the model with enough flexibility to be efficiently adapted before any unexpected scenario comes up. Working with a near-real time predictive model for water demand forecasting speeds up decision-making processes on water supply operation and management; in particular, it is useful in water disruption scenarios.

The roadmap for the rest of the paper is as follows: Section 2 describes the hybrid forecasting method where the SVR and AFS equations, together with the hybridization process, are presented. Section 3 presents the water supply case study, the variables involved in the forecasting process, and the evaluation methodology. The results obtained at each step of the method are presented in Section 3.2, where results from SVR and SVR+AFS are analysed separately to give strong support to the hybrid model through clear improvement of results.

The efficiency evaluation is also presented in this section. Section 4 presents a discussion on the results and conclusions about this work.

2. Hybrid method: Support Vector Regression+Adaptive Fourier Series

This Section introduces our hybrid methodology proposal for near real-time water demand prediction. Firstly, both the off-line model (using SVR) and its corresponding on-line adjustments (via AFS) are briefly explained. After that, we focus on its combination through modelling the historic record of errors produced by just using the basis off-line model, and justify the use of AFS to adjust those errors. Finally, statistical parameters for evaluation the accuracy of proposed model is presented. Also an efficiency parameter is presented to define a period of retuning for SVR model.

2.1. Support Vector Regression

Common kernel-based learning methods [27, 28] use an implicit mapping, ϕ, of the input data into a high dimensional feature. Then, a kernel function, K, is used to return the inner product $\langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle$ between the images of two data points $\boldsymbol{x}$, $\boldsymbol{x}'$ in the feature space. The choice of the map ϕ aims to convert the non-linear relations into linear ones. The learning then takes place in the feature space, while the learning algorithm can be expressed so that the data points only appear inside dot products with other points, readily calculated via the kernel. This is often referred to as the "kernel trick" [27, 29].

The key characteristic of SVR is that it allows to specify a margin, ε, within which we are willing to accept errors in the sample data without they affecting the prediction quality. The SVR predictor is defined by those points or vectors which lie outside the region formed by the band of size $\pm\varepsilon$ around the regression. These vectors are the so-called *support vectors*. Giving data $\{(x_1, y_1), \ldots, (x_n, y_n)\} = \{(\boldsymbol{x}, \boldsymbol{y})\} \subset X \times \mathbb{R}$, where X denotes the space of the input patterns, the goal is to find a function

$$\hat{f}(\boldsymbol{x}) = \langle \boldsymbol{w}, \phi(\boldsymbol{x}) \rangle + b, \tag{1}$$

that at most deviates ε from the observed output, y_i, for all pairs (x_i, y_i) with $i = 1, \ldots, n$; at the same time that minimizes the so-called model complexity, which depends on the support vectors (see Eq. (2)).

$$\begin{aligned} &\min_{\boldsymbol{w},b} \frac{1}{2}||\boldsymbol{w}||^2 \\ &\text{s. to } |y_i - \langle \boldsymbol{w}, \phi(x_i) \rangle - b| \leq \varepsilon \end{aligned} \tag{2}$$

The constraints of Eq. (2) assume that $\hat{f}(\boldsymbol{x})$ exists for all y_i with precision $\pm\varepsilon$. Nevertheless, the solution may actually not exist or, interestingly, it would be possible to achieve better predictions if outliers were allowed. The inclusion of *slack variables* on the regression penalizes deviations larger than $\pm\varepsilon$ but leaves room for both supporting approaches with possible infeasibilities and exploring a wider set of models. Thus we consider ξ^+ and ξ^- such that:

$$\xi^+ = \hat{f}(x_i) - y(x_i) > \varepsilon \tag{3}$$

$$\xi^- = y(x_i) - \hat{f}(x_i) > \varepsilon, \tag{4}$$

so that the objective function and constraints for SVR is finally stated as

$$\begin{aligned} &\min_{\boldsymbol{w},b} \ \frac{1}{2}||\boldsymbol{w}||^2 + C\frac{1}{n}\sum_{i=1}^{n}(\xi_i^+ + \xi_i^-) \\ &\text{s. to } \ y_i - \langle \boldsymbol{w}, \phi(x_i) \rangle - b \leq \varepsilon + \xi_i^+, \\ &\qquad \langle \boldsymbol{w}, \phi(x_i) \rangle + b - y_i \leq \varepsilon + \xi_i^-, \\ &\qquad \xi_i^+, \xi_i^- \geq 0 \ \ i = 1, \ldots, n, \end{aligned} \tag{5}$$

where n is the number of training patterns and C is a trade-off parameter between model complexity and training error. As said, ξ^+ and ξ^- are slack variables for exceeding the target value by more than ε and for being below the target value by less than ε, respectively. This method of tolerating errors is said to be ε-*insensitive* [27].

2.2. Adaptive Fourier series

A Fourier series is an expansion of a periodic signal or function into the sum of a set of simple oscillating functions. It is a fact that water demand follows a quasi-periodic behaviour, with different periodic components (corresponding mainly to daily, weekly and seasonal demands). As a consequence, Fourier series are ideal candidates to model such variations. The Fourier series set of equations presented here is based on [30], where trigonometric adjustment is applied to data coming from both equally and not-equally spaced measurements.

Taking equally spaced values of t in the period of interest, Eq. (6) normalizes the times to the interval $[0\ ,\ 2\pi]$.

$$t_i = \frac{2\pi i}{N}, \tag{6}$$

where t_i is the normalized time, $0 \leq t_i \leq 2\pi$, i is the position of the point in the interval of evaluation and N is the total number of time intervals in the period of interest.

The Fourier series approximation for a general function f^* at a point t_i is written according to Eq. (7),

$$f^*(t_i) = a_0 + \sum_{j=1}^{M}(a_j cos(jt_i) + b_j sin(jt_i)), \tag{7}$$

where M is the length of the Fourier polynomial, and a_0, a_j and b_j are the adjustable Fourier coefficients.

The differences between the real (measured, in our case) values of the function f and the estimated values , calculated by Eq. (7) enables us to obtain the square deviation s, given by Eq. (8).

$$s = \sum_{i=0}^{N-1}\{f(t_i) - [a_0 + \sum_{j=1}^{M} a_j \cdot cos(jt_i) + \sum_{j=1}^{M} b_j \cdot sin(jt_i)]\}^2. \tag{8}$$

The least square method to minimize s usually takes into account orthogonality conditions and the parameters: a_0, a_j and b_j [30]. It may be written as:

$$
\begin{aligned}
a_0 =& \frac{\sum_{i=0}^{N-1} f(t_i)}{N}, \\
a_j =& 2\frac{\sum_{i=0}^{N-1} f(t_i)cos(jt_i)}{N}, \\
b_j =& 2\frac{\sum_{i=0}^{N-1} f(t_i)sin(jt_i)}{N}.
\end{aligned} \tag{9}
$$

Once these parameters have been determined, Eq. (7) is used to estimate f to functionally adjust the measured data.

2.3. Hybrid Model SVR+AFS

Hourly water demand is not a simple phenomenon to describe. The non-linear correlation between exogenous variables and the near periodical behaviour of consumption is difficult to predict while maintaining a certain error threshold. In addition, off-line models can not respond to incidences or anomalies that happen in water distribution along the time. The proposed hybrid model provides a fast response to new weather or physical factors in the water demand environment by adding up an on-line predictive layer to a prior off-line regular predictor. The proposal is to use Fourier theory, as described in the last section, to model and predict the deviation for the prediction obtained by SVR. That is to say, a Fourier series layer over the SVR output will adjust better the minimum and maximum demand peaks and capture parts of the time series periodicity that SVR can not reproduce. As a Fourier time series model has not special requirements to work out, it is also efficiently adapted to new scenarios regarding the exogenous inputs affecting water demand. The hybrid model aims to improve as well the usual SVR results by enhancing the capacity to capture, through Fourier series, other periodical features of the already considered variables based on weather and calendar data.

The initial SVR model has to be well calibrated and trained before validation. The corresponding parameters are the following:

- Parameter C, which gives a trade-off between model complexity and the amount up to which deviations larger than ε are tolerated. Furthermore, the robustness of the regression model depends on the choice of the C value; this means that the choice of the C value influences the significance of the individual data points in the training set.

- Parameter ε regulates the radius of the ε tube around the regression function and thus the number of support vectors that finally will be selected to construct the regression function (leading to a sparse solution). A too large ε value results in less support vectors (more data points will be fit in the ε tube) and, consequently, in a more smooth (less complex) regression function.

The training begins by creating a mesh with pairs of C and ϵ. The values used in this work were based on [31], and have the following bounds: $0.05 \leq \epsilon \leq 0.9$ and $1 \leq C \leq 1500$. This develops a Grid Search Method (GSM) as a classical option to tune the parameters in which equispaced sampled points are used to represent all their possible combinations [32, 33]. After the model selection process, consecutive runs of N demand observations are used to compute the deviation, f, between predicted and observed values of demand. We typically have that the SVR model error is larger at demand peaks and, as expected, has periodical behaviour.

Figure 1 shows the overall on-line process. As a novelty, we highlight the embedding of an optimal cycle for model (SVR) regeneration. This cycle is related with the model structure obsolescence. It does not need continuous updating but only after a certain number of predictions. This cycle is determined by controlling the model accuracy as Section 3.2 explains in the case study.

Deviations between the observed demand and the demand predicted by the SVR model are computed at every time interval. The Fourier Series coefficients are updated using this new time series, modelling the on-line error of the model.

The final demand predicted by the on-line hybrid model combines the water demand predicted by the SVR model, y_{SVR}, and the predicted deviation

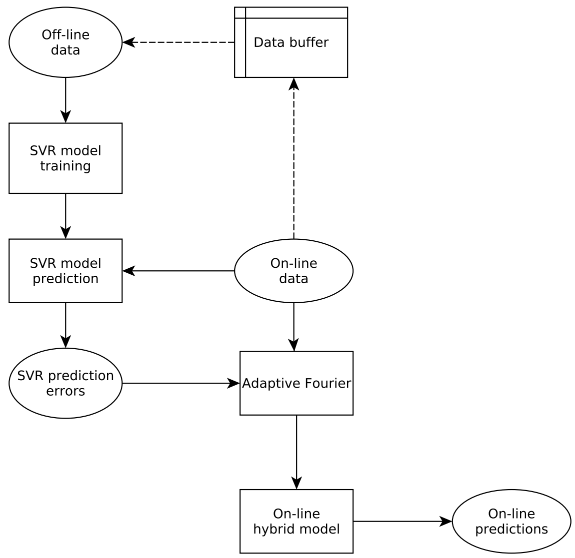


Figure 1: Flowchart of the on-line hybrid process SVR-AFS

estimated by the AFS model, f_{AFS}. Eventually, the near-real time predicted demand may be written following Eq. (10).

$$y = y_{SVR} + f_{AFS}. \tag{10}$$

Figure 2 shows the on-line updating scheme using AFS. The predicted correction not only incorporates the on-line attribute to the entire model, but also adds the last (more recent) measured water consumption as a new variable to the forecasting process sliding the window one time step further, this updating the database. The last water consumption has also been adopted in previous research as a way of incorporating the well-known inertial features of water demand [34, 5].

2.4. Computational Efficiency

The off-line water demand forecasting model must be regularly updated as new demand events arrive. Despite the on-line model is able to respond fast to new demands thus satisfying its main objective, namely contributing to the best

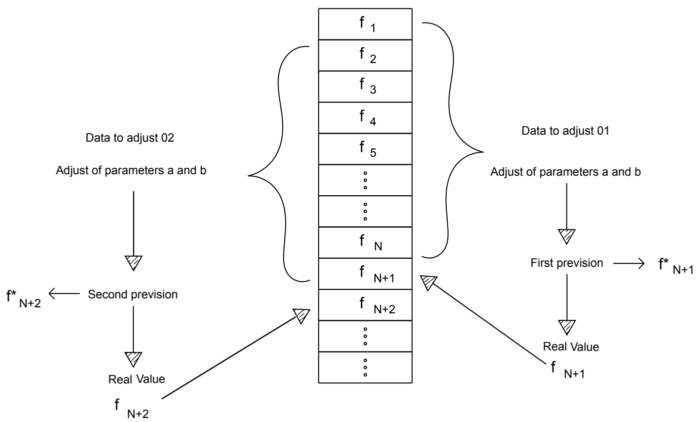


Figure 2: Diagram of AFS update

operation and management of the system, as new data arrives, new correlations between water demand and input variables may appear turning obsolete the current base model (even in the case one works with on-line corrections).

Accordingly, the off-line model structure needs to be periodically updated through new data. Thus, determining an optimal updating cycle is of paramount importance, while it is strongly dependent on the on-line model efficiency. Let T the total CPU time spent to run the calibrated hybrid model corresponding to SVR prediction and AFS deviation adjustment. We define the efficiency φ by Eq. 11:

$$\varphi = \frac{N_{tr}}{e \cdot T}, \tag{11}$$

where e is the root mean square error (RMSE) and N_{tr} is the training data size. The product of e and T in the denominator accounts for the accuracy and the computational cost of the method. The training data endows the model of real data information; then, the higher the value of N_{tr} the lower the error, i.e. the model augments its reliability. However, the time necessary to obtain the new model structure will be higher for large values of N_{tr}. The value φ, thus describes a trade-off between accuracy and computational cost, that is to say, the behaviour of the model when it is trained using different sizes of training data.

3. Experimental Study

This study uses water demand data collected from a real district metered area (DMA) in Franca, Brazil. This DMA is mostly based on residential customers. Water consumption data corresponds to metered data at the DMA's inlet every 20 minutes from May 2012 until December 2013. These measures are aggregated to handle an hourly demand rate data structure during the 570 days of this period. The analyses are conducted through a training window of 400 days. The next 140 days are taken to validate the model. New fresh 30 days are used to test the performance of the proposal.

3.1. Exploratory Data Analysis

Previous studies found in the literature use variables taking into account weather and calendar information for generating models to predict water demand [35, 36]. Figure 3 visualizes the behaviour of water demand and different weather inputs; where each time series data for each variable is plotted along with the box plot showing its individual distribution. Cross-correlation analyses have been carried out to check the impact of weather variables in water demand time series. As a result, the SVR model proposed for this paper uses rain, temperature, humidity, and wind velocity as the most important physical variables to consider for further analysis.

While temperature and air humidity time series follow a regular distribution along all the time frame, wind velocity and rain present extreme values as it is observed in the box plots of Figure 3. This figure also evidences direct correlations between air humidity and rain as well as temperature and water consumption. Cross-correlations computed for these time series shows evidence enough from which one cannot reject the existence of correlation between these variables.

Calendar information is important to analyse water demand given that human behaviour is usually depending on the day of the week. Thus, week days

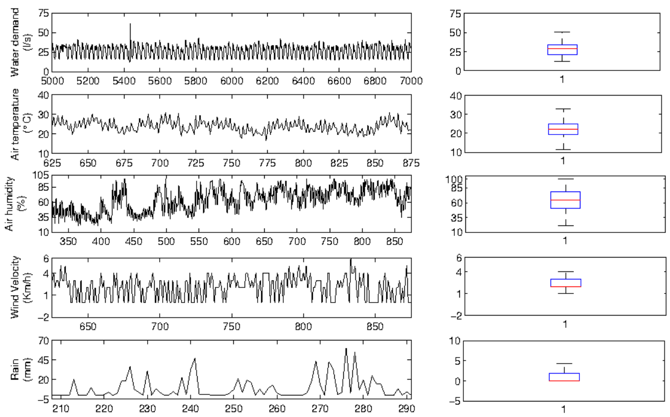
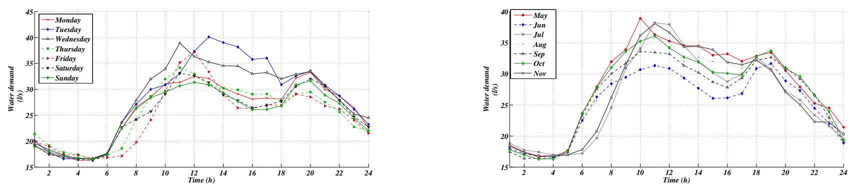


Figure 3: Water demand and weather variables series

have a different water consumption when compared with weekend days. In addition, holidays follow different consumption patterns than regular days. Taking this into account, we use markers for weekday/weekend and occurrence of holiday, while the model also handles information of the time of the day and the month of the year. Week day, hour and month attributes are enumerated with discrete variables while the occurrence of a holiday was classified with binary variables. Figure 4a shows different patterns for water demand when comparing days belonging to a typical week (no holiday).

Cyclical behaviours and possible differences between months are also considered. For instance, there is a difference between June and July, as Figure 4b shows. July is a typical Winter holiday month in Brazil with mild temperatures in which water consumption is higher than in June, a month with similar temperatures but not a vacation month.

Figure 5 shows a comparison between holidays and week days regarding water consumption using two representative day for each category, common week day and holiday, respectively. We can observe that holiday water demand pattern usually has a lower consumption peak and a lower oscillation level during



(a) Water demand by weekday

(b) Water demand by months

Figure 4: Comparison: (a) water demand by weeks and (b) water demand at same weekday along 9 months

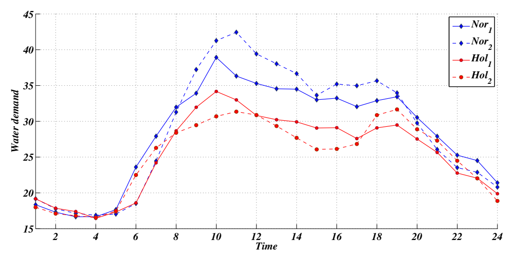


Figure 5: Water demand at various week days showing the holiday effect

the day.

3.2. SVR-AFS Results

To evaluate the final predicted demand by the hybrid use of SVR demand and AFS deviation correction, this work uses RMSE Eq. (12) , the mean absolute percentage error (MAE%), Eq. (13) , and correlation coefficient (R^2) [14, 10].

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{T} e_i^2}, \tag{12}$$

$$MAE\% = \frac{1}{n} \cdot \sum_{i=1}^{T} \frac{e_i}{\mu_{obs}} \cdot 100. \tag{13}$$

where e_i is the deviation between observed and final predicted demand and μ_{obs} is the mean value for observed series. These statistical parameters allow model accuracy assessment. While the RMSE shows the squared difference between real and predicted demand, the MAE % outcome is the deviation compared with the mean value of the observed series.

Regarding the training process, Table 1 shows the top 5 pairs of parameters. The best values of C and ϵ are 50 and 0.05, respectively. The process to find this best pair is based on computations with data for 20 weeks, and we work with the average values of the error, adding statistical robustness to the tuning process.

Table 1: Best values C and ϵ in the training process

C	ϵ	MSE
50	0.05	20.83
50	0.125	20.86
50	0.200	20.91
50	0.275	20.95
50	0.375	21.00

The predicted demand using the best validated model is presented in Figure 6. We can observe that the largest deviations occur at the maxima and minima.

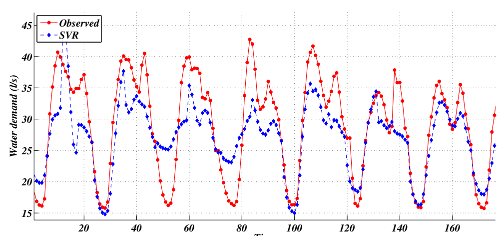
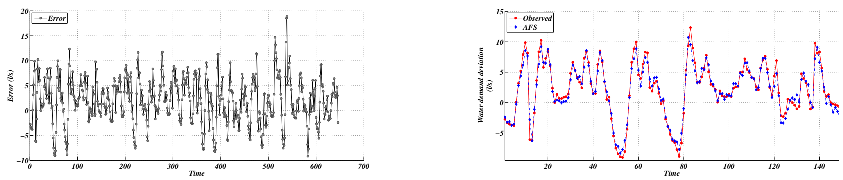


Figure 6: Water demand prediction using SVR compared with observed demand

Figure 7a shows the performance of the SVR model in terms of the deviation respect to the observed data. The near-periodical deviation behaviour of this error time series justifies the use of Fourier methods. Figure 7b shows the correction of this deviation error adjusted by the AFS.

Finally, the SVR model prediction model for water demand with corrected deviation via the AFS model is presented in Figure 8a compared with the real



(a) Error time series

(b) Predicted and Real Error

Figure 7: (a) deviation between real and SVR predicted demand and (b) AFS approximation of this deviation

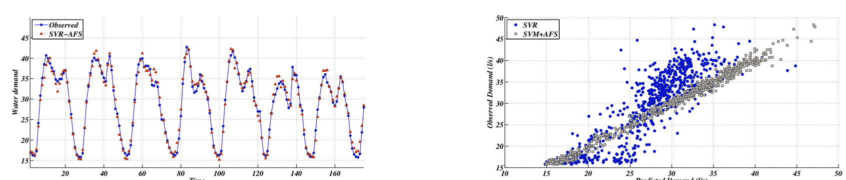
water demand.

Table 2 shows the statistical parameters for both the single SVR and the hybrid model.

Table 2: Model comparison

Model	RMSE	MAE%	R^2
SVR	4.767	12.91	0.745
SVR-AFS	1.318	3.45	0.974

Figure 8b presents the correlation between the observed demand and the SVR prediction and also between the observed demand and the SVR-AFS prediction.



(a) Hybrid SVR-AFS model predictions vs. observed demand

(b) Comparison between SVR and SVR-AFS prediction vs. observed demand

Figure 8: Hybrid SVR-AFS results: (a) time series comparison and (b) correlations with observed water demand

The *a posteriori* application of AFS allows the use of a single off-line SVR as demand base model, since AFS is able to quickly adapt the predictions to the new data available. The use of AFS is based on the quasi-periodic shape of the deviations between observed data and predicted by SVR (see Figure 6). AFS is accurate in predicting these deviations and the results of the hybrid model have been significantly improved regarding the off-line proposal. Figure 8a proves that the maximum and minimum values are better described by the hybrid model. The CPU time for the AFS model is constant, in this case equal to 16.5s. This fact together with the agility of the algorithm allow near real-time water demand prediction.

The evaluation of the model efficiency with respect to the size of the training data enable us to find the best training cycle for completely updating the off-line model. Figure 9 presents the evolution of the training cycle efficiency for a range of data sizes.

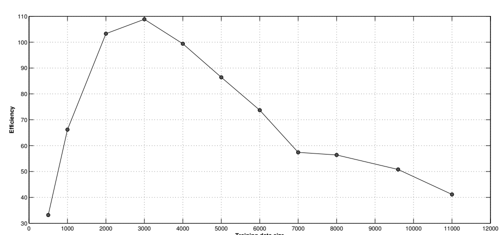


Figure 9: Efficiency values for different training data sizes

In this case, the optimal data size to refresh the SVR model is, accordingly, 3000 registers, which corresponds to 125 days. With less than that number of data the model does not reach its maximum efficiency, mainly because the error is big (although the on-line model supplements the gap of new data). After this point the computational cost increases without a significant RMSE reduction.

All the computations for SVR and AFS models were performed within MatLab, on a 64 bit Linux Debian-based operating system, Ubuntu 15.04, installed on a 2.4 GHz Intel® Xeon(R) CPU E5-2665 0 with 16 Cores. It has been used and adapted `LibSVM` Matlab toolbox [37] for relevant routines dealing with support vector machines regression.

4. Conclusions

This article has presented a hybrid model for hourly water demand forecasting. The model builds over an off-line Support Vector Regression model, constituting a base forecasting, an an on-line Adaptive Fourier Series responsible for forecasting deviation.

SVR accounts for physical and calendar information necessary for water demand forecast, as these parameters are the SVR input. The prediction using a SVR model is able to describe the shape or general pattern of the daily demand. However, it is not able to capture well the extremes and, as a result, the accuracy diminishes at the demand peaks. In addition, the obsolescence of the model is a common issue for any predictive model based on the straightforward use of Machine Learning regression models. The use of Adaptive Fourier Series aggregates to the SVR model a way to update the prediction in near-real time by correcting the demand predicted by the off-line base model. A further advantage of AFS is the addition of another social variable involved in the problem: the last water demand.

The paper also introduces a simple way to determine the optimum training data size for the off-line model, which considers computational effort and error. The ideal cycle of training can help water companies to organize interruptions of the model activity for update with new data. Updating the model is important since new data supplement the model with new correlations among the demand and input variables.

The proposed model can be an important tool for water utilities, since the on-line feature supports operation and management of WDSs, allowing operators to programme efficient and right manoeuvres to save energy and water.

References

[1] S. Behboudian, M. Tabesh, M. Falahnezhad, F. A. Ghavanini, A long-term prediction of domestic water demand using preprocessing in artificial neural

network, Journal of Water Supply: Research and Technology-AQUA 63 (1) (2014) 31–42.

[2] R. Klempous, J. Kotowski, J. Nikodem, J. Ułasiewicz, Optimization algorithms of operative control in water distribution systems, Journal of Computational and Applied Mathematics 84 (1) (1997) 81–99.

[3] F. Odan, L. Reis, Hybrid water demand forecasting model associating artificial neural network with fourier series, Journal of Water Resources Planning and Management 138 (3) (2012) 245–256.

[4] G. P. Box, G. M. Jenkins, Times series analysis: Forecasting and control, Holden-Day, 1976.

[5] S. Alvisi, M. Franchini, A. Marinelli, A short-term, pattern-based model for water-demand forecasting, Journal of Hydroinformatics 9 (1) (2007) 39–50.

[6] O. Voitcu, Y. S. Wong, On the construction of a nonlinear recursive predictor, Journal of Computational and Applied Mathematics 190 (1) (2006) 393–407.

[7] M. Tiwari, J. Adamowski, Medium-term urban water demand forecasting with limited data using an ensemble waveletbootstrap machine-learning approach, Journal of Water Resources Planning and Management 141 (2) (2015) 04014053.

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

[9] S. Zhou, T. McMahon, A. Walton, J. Lewis, Forecasting operational demand for an urban water supply zone, Journal of Hydrology 259 (1) (2002) 189–202.

[10] M. Herrera, L. Torgo, J. Izquierdo, R. Pérez-García, Predictive models for forecasting hourly urban water demand, Journal of Hydrology 387 (1-2) (2010) 141 – 150.

[11] I. S. Msiza, F. V. Nelwamondo, T. Marwala, Artificial neural networks and support vector machines for water demand time series forecasting, in: Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on, IEEE, 2007, pp. 638–643.

[12] C. L. Z. Tu-Qiao, Hourly water demand forecast model based on bayesian least squares support vector machine, Journal of Tianjin University 9 (2006) 005.

[13] I. Banicescu, R. L. Cariño, J. L. Harvill, J. P. Lestrade, Investigating asymptotic properties of vector nonlinear time series models, Journal of Computational and Applied Mathematics 236 (3) (2011) 411–421.

[14] M. Herrera, J. García-Díaz, J. Izquierdo, R. Pérez-García, Municipal water demand forecasting: tools for intervention time series, Stochastic Analysis and Applications 29 (6) (2011) 998–1007.

[15] I. Montalvo, J. Deuerlein, Driving online simulations in water distribution systems, Procedia Engineering 70 (2014) 1183 – 1191.

[16] J. Machell, S. Mounce, J. Boxall, Online modelling of water distribution systems: a UK case study, Drinking Water Engineering and Science 3 (2010) 21–27.

[17] M. López-Ibáñez, T. D. Prasad, B. Paechter, Ant colony optimization for optimal control of pumps in water distribution networks, Journal of Water Resources Planning and Management 134 (4) (2008) 337–346.

[18] A. Preis, A. Whittle, A. Ostfeld, Online Hydraulic State Prediction for Water Distribution Systems, 2009, Ch. 31, pp. 1–23.

[19] I. Okeya, Z. Kapelan, C. Hutton, D. Naga, Online modelling of water distribution system using data assimilation, Procedia Engineering 70 (2014) 1261 – 1270.

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

[21] S. Van Vaerenbergh, J. Vía, I. Santamaría, A sliding-window kernel RLS algorithm and its application to nonlinear channel identification, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006), Toulouse, France, 2006.

[22] S. Hoi, R. Jin, P. Zhao, T. Yang, Online multiple kernel classification, Machine Learning 90 (2) (2013) 289–316.

[23] G. P. Zhang, Time series forecasting using a hybrid arima and neural network model, Neurocomputing 50 (2003) 159–175.

[24] C. H. Aladag, E. Egrioglu, C. Kadilar, Forecasting nonlinear time series with a hybrid methodology, Applied Mathematics Letters 22 (9) (2009) 1467–1470.

[25] P. G. Nieto, E. García-Gonzalo, J. A. Fernández, C. D. Muñiz, A hybrid PSO optimized SVM-based model for predicting a successful growth cycle of the Spirulina platensis from raceway experiments data, Journal of Computational and Applied Mathematics (2016) 293–303.

[26] M. Herrera, S. Canu, A. Karatzoglou, R. Pérez-García, J. Izquierdo, An approach to water supply clusters by semi-supervised learning, in: D. A. Swayne, W. Yang, A. Voinov, A. Rizzoli, T. Filatova (Eds.), International Congress on Environmental Modelling and Software, International Environmental Modelling and Software Society (iEMSs), 2010, pp. 1925–1932, student's award.

[27] B. Schölkopf, A. J. Smola, Learning with kernels: support vector machines, regularization, optimization, and beyond (Adaptive computation and Machine Learning), The MIT press, Cambridge, MA, USA, 2002.

[28] J. Shawe-Taylor, N. Cristianini, Kernel methods for pattern analysis, Cambridge University Press, 2006.

[29] B. Schölkopf, The kernel trick for distances, Tech. rep., Microsoft Research (2000).

[30] E. Luvizotto Jr, Representation of characteristic curves of hydraulic machines for computational simulations, Ph.D. thesis, University of São Paulo (1992).

[31] H.-T. Lin, C.-J. Lin, A study on sigmoid kernels for svm and the training of non-psd kernels by smo-type methods, submitted to Neural Computation (2003) 1–32.

[32] J. Wang, X. Wu, C. Zhang, Support vector machines based on k-means clustering for real-time business intelligence systems, International Journal of Business Intelligence and Data Mining 1 (1) (2005) 54–64.

[33] C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al., A practical guide to support vector classification (2003).

[34] J. Zhang, R. Song, N. R. Bhaskar, M. N. French, Short term water demand forecasting: a case study, in: 8th annual water distribution systems analysis symposium. ASCE Library, Cincinnati, 2006.

[35] M. A. Al-Zahrani, A. Abo-Monasar, Urban residential water demand prediction based on artificial neural networks and time series models, Water Resources Management 29 (2015) 3651–3662.

[36] M. Bakker, H. Van Duist, K. Van Schagen, J. Vreeburg, L. Rietveld, Improving the performance of water demand forecasting models by using weather input, Procedia Engineering 70 (2014) 93–102.

[37] C.-C. Chang, C.-J. Lin, Libsvm: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology (TIST) 2 (3) (2011) 27.