A water demand forecasting methodology for supporting day-to-day management of water distribution systems

G. Arampatzis*, N. Perdikeas, E. Kampragou, P. Scaloubakas and D. Assimacopoulos
School of Chemical Engineering, National Technical University of Athens, 15780 Zografou, Greece

* Corresponding author: E-mail: arab@chemeng.ntua.gr, Tel +302107723135, Fax: +302107723155

Abstract
Water demand forecasts are widely used for the design, operation and management of water supply systems and in order to minimize the use of resources (water, energy, etc.) and costs. A methodology for short-term water demand forecasting is presented in this paper. It is based on a "similar days" approach, which is able to explain the day-to-day demand variation and still remains simple enough to be of almost general application with very modest data requirements. The developed model uses a composite "day similarity index", based on day attributes (such as the day of week, weather conditions, special events, etc.) that affect water consumption and their state values can be accurately predicted for the forecast day. This index is used to select a small subset of similar days from a database of historical water consumption data and correlated day attributes. The expected demand for the forecast day is estimated as a weighted average of observed consumption for the days in the "similarity set". The model also provides a range and a probability distribution of forecast values, thus permitting the performance of risk analysis, necessary for resource optimization. The model is applied in two case studies: i) the city of Karlsruhe (Germany) and ii) the metropolitan area of Barcelona (Spain). Real historical daily and hourly water consumption data sets, spanning the last 3 and 6 years are used, together with the corresponding meteorological and other day characterization data. Results from both applications demonstrate that small forecasting errors can be achieved by using readily available data.

Keywords: water demand management, forecasting model, similar days, water demand drivers

1. INTRODUCTION

Water demand forecasts are used by water authorities and water service providers for the planning, designing the infrastructure, operating and managing water supply systems [1]. Depending on the specific goals, different forecasting timeframes are used and different forecasting methods are applicable. For example, planning for large-scale investments usually requires long-term forecasting, at a yearly or monthly resolution with limited accuracy. On the other hand, for optimum allocation of water supply among users, medium-term forecasting at a monthly or weekly resolution and somewhat higher accuracy might be required. A variety of water demand forecasting methodologies suitable for such purposes exist, including unit use/end-use models, regression models, time-series analysis, etc.

An analysis of conditions and management procedures along the water supply-distribution chain shows that for water-stressed regions or regions with frequent shortage problems (e.g. around the Mediterranean), resource management priorities lie mostly in the day-to-day regulation of water resources so as to minimize consumption as well as water losses and other environmental impacts. For water-rich regions (e.g. the European North), priorities lie mostly in optimizing the daily operation of the infrastructure (e.g. scheduling tank and pump work load) so as to minimize costs and energy consumption [2]. In both cases, the efficient operation and management of the water system requires the continuous update of short-term water demand forecasts, that should be both highly accurate and of high temporal resolution (daily, hourly and/or even finer time intervals) [3].

Most of the conventional water demand forecasting models that try to express causative relations between water demand and the water users' attributes (demand drivers), such as the unit use/end-use [4, 5], or regression models [6], are mostly suitable for medium-to-long-term forecasting. They are not able to provide accurate enough short-term forecasts at the required resolution or they require data, which are not readily available in most real world situations (e.g. may require special surveys, smart metering, etc.).

On the other hand, the statistical trends analysis (time-series) type of models [2, 7, 8], might provide good short-term predictions, but these depend on the day-to-day stability of water consumption and the very short-term effect of "hidden" system variables. A methodology that is able to encode such variables in a flexible way and to explain day-to-day demand variation is the "similar day method" [9, 10]. It tries to emulate the way that a human expert (e.g. working on pump scheduling for a Water Utility) does his guesswork about what a day's demand pattern might look like, by trying to find a past day that looks similar. Typically, it is used to forecast demand for the next day or the next few days. However, it may be used as far ahead as desired, provided that the states of the system variables can be forecasted. Besides using relatively easy to find data, the methodology is quite extensible, since the forecast time step (day, hour, minutes) can be parametric and the system variables themselves can be parametric (they are included in a system's mapping database, which forms part of the model). Thus, system experts using the model may continuously improve its predictive power and fit it to changing conditions and system events.

In Section 2, a brief description of the similar day methodology, including the selection of model variables, the establishment of the mapping database and the definition, measurement and use of the similarity between days is presented. In Section 3, the results from the implementation of the model to two case studies are discussed and their accuracy is assessed. Finally, Section 4 draws some conclusions about the applicability of the model.

2. METHODOLOGY

The proposed model is based on the "similar day" pattern recognition methodology, which forecasts future water demand by averaging the historical water consumption data of a set of past days that have very similar characteristics with the day of forecast. It is expected that days with similar attributes (conditions) will follow similar consumption patterns. Thus, the model comprises of two stages. At the first stage, days similar to the forecast day are selected on the basis of a number of explanatory variables while, at the second stage, demand is predicted based exclusively on the recorded historic consumption values for this set of similar days. The data required to support the forecasting procedure are composed of three main datasets (water consumption, day characterization and meteorological data). The forecasting procedure can be further analyzed into seven (7) steps:

1. Selection of factors influencing water demand to be included in the analysis.
2. Establishment of a mapping database, to obtain the range of values for each factor.
3. Calculation of the similarity index for all days in the database.
4. Selection of similar days to the forecasted day.
5. Calculation of weights of the similar days.
6. Calculation of the forecasted day demand.
7. Estimation of error statistics and confidence levels.

A number of similarity factors, influencing water consumption, are selected and incorporated to initialize the mapping database (Step 1). Such factors are the type of day (e.g. normal, public holiday, strike, major sports or other special event, etc.), day of the week, month, maximum and minimum temperatures, the number of days after rainfall (calculated from meteorological data on precipitation), etc.

The mapping database (Step 2) is filled by providing numeric ranking values for each possible state (class) of each qualitative factor, in order to capture the importance of each factor and provide a basis for the comparison of days. Quantitative factors, such as temperature, are also converted into qualitative ones by subdividing their ranges into classes correlated to distinct day states. The numeric ranking vales of all factor states, contained in the mapping database, are normalized in the [0...1] interval. The final mapping values (weights of similarity) are obtained by multiplying the normalized value of each state with a normalized weighting coefficient for each similarity factor.

The similarity index (Step 3) provides a quantitative expression of the concept of similarity. Days are considered similar to the forecast day, if they have the most similar characteristics. The calculation of the similarity index is based on the following formula [9]:

$$r_i = \frac{\sum_{k=1}^{M} v_{ik} v_{ok}}{\sqrt{\sum_{k=1}^{M} v_{ik}^2 \cdot \sum_{k=1}^{M} v_{ok}^2}} \tag{1}$$

where r_i is the similarity index of past day (i), M is the number of factors influencing demand, v_{ik} is the mapping value of factor (k) for day (i) and v_{ok} is the mapping value of factor (k) for forecast day (o). This index is analogous to a correlation coefficient and its values are assessed similarly.

The days having sufficiently high similarity index values are considered similar to the forecast day and they are selected to be taken into account in the estimation of the demand forecast (Step 4). Sufficiently high are the values which are above a pre-set similarity threshold (e.g. $r_i > 0.9$).

The weight of any similar day (i) on the forecasted demand (Step 5) is based on its degree of similarity, according to the following formula:

$$w_i = \frac{r_i}{\sum_{j=1}^{N} r_j} \tag{2}$$

where w_i is the weight of similar day (i) and N the number of similar days with $r_i >$ threshold. Accordingly, the forecasted water demand (Step 6) is estimated by using the formula:

$$d = \sum_{i=1}^{N} w_i \cdot c_i \tag{3}$$

where d is the forecasted water demand (daily, hourly or other) and c_i is the observed water consumption of similar day (i) for the same time step.

Finally, error statistics and confidence levels of the result (Step 7) are estimated using standard statistical methods [11], having the advantage that the most probable values and their probabilities (estimated by the similarity weights w_i) have already been estimated in Steps 4 and 5 respectively.

3. APPLICATION IN TWO CASE STUDIES

The model is implemented and tested with real historical data for the city of Karlsruhe (Germany) and for the metropolitan area of Barcelona (Spain), in the course of the FP7 WatERP research project. Forecasting is incorporated as an integral part of two entirely different water resource management procedures. In Karlsruhe hourly forecasts are used to optimize the operation of numerous pumping stations, thus minimizing the use of energy and costs. In Barcelona daily forecasts are used to estimate the water that should be supplied daily from each one of five main water resources, in order to meet demand and reduce the amount of excess water rejected to the sea.

3.1 City of Karlsruhe

The model is applied in the city of Karlsruhe using historical data on hourly water consumption for a range of almost three years (May 2010 to December 2012). Daily meteorological data (minimum, maximum and average temperatures as well as precipitation height) were also collected for the same time period. Missing data and invalid values were excluded. The (aggregated) daily water consumption time-series and the correlated average temperature are presented in Figure 1.

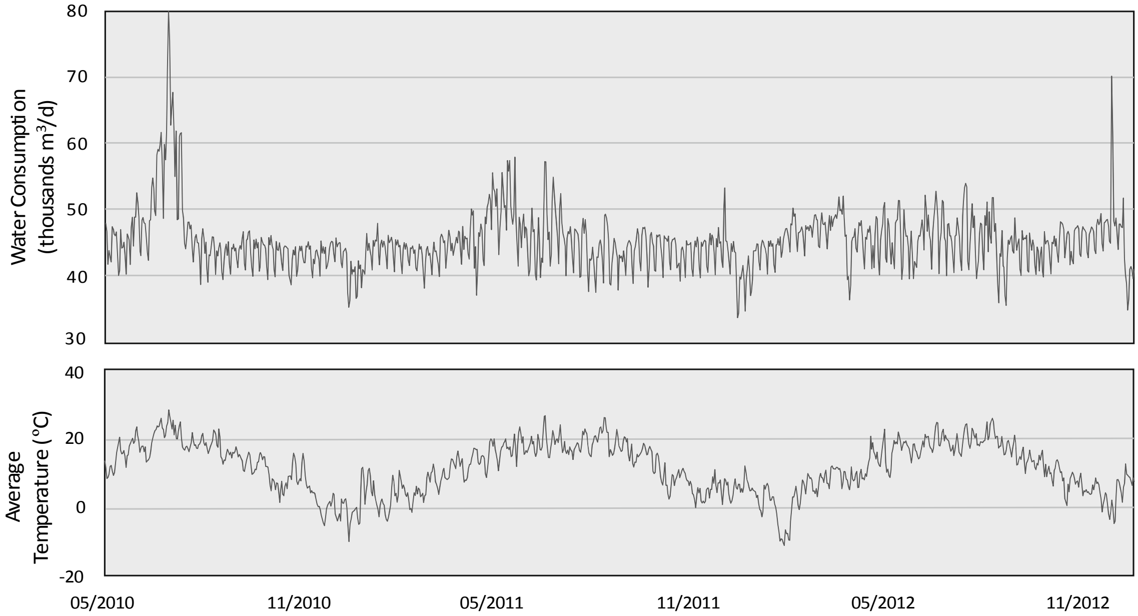
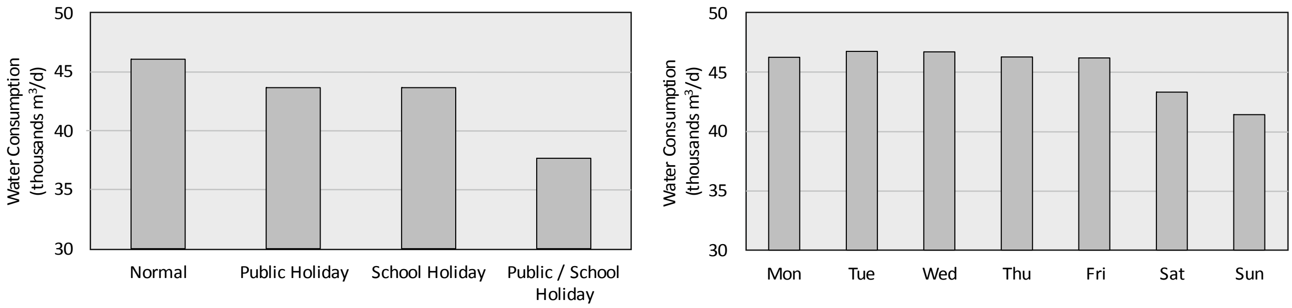


Figure 1. Daily water consumption and average temperature in the city of Karlsruhe.

Temperature variables were directly used as similarity factors in the forecasting model while the precipitation was used to calculate a “days after rainfall” similarity factor, which has been shown to be correlated with the water consumption. Other similarity factors used are the month, the day of week, the hour of the day and the day type. The latter is used to categorize days into four types: normal day, public holiday, school holiday and public/school holiday. Figure 2 presents the variation of daily water consumption by the type of day and the day of week



(a) (b)

Figure 2. Variation of daily water consumption in the city of Karlsruhe
(a) by type of day and (b) by day of week

The hourly demand forecasting model is tested against a data subset of 168 hours (one week) were excluded from the dataset used to ‘train’ the model. The forecasted demand is compared to the actual consumption and the results are presented in Figure 3. The model performs satisfactorily with an average relative forecasting error equal to 5.4%.

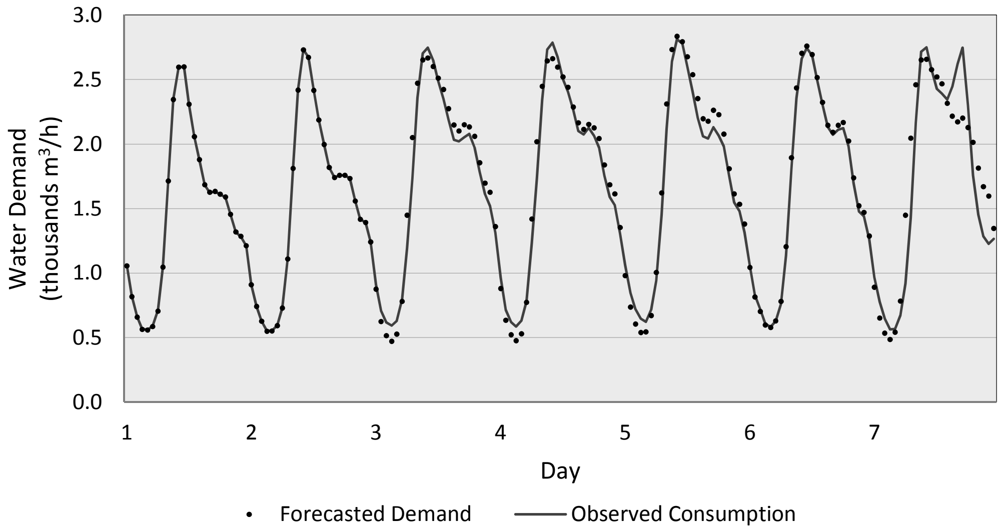


Figure 3. Hourly forecasted water demand and observed consumption in the city of Karlsruhe.

3.2. Metropolitan area of Barcelona

The application of the model to the Metropolitan Area of Barcelona uses historical data on daily water consumption for a range of six years (January 2006 to December 2012). Meteorological data (minimum, maximum and average temperature, precipitation height) and a day characterization dataset were also provided. Figure 4 presents the water consumption time-series and the correlated average temperature values.

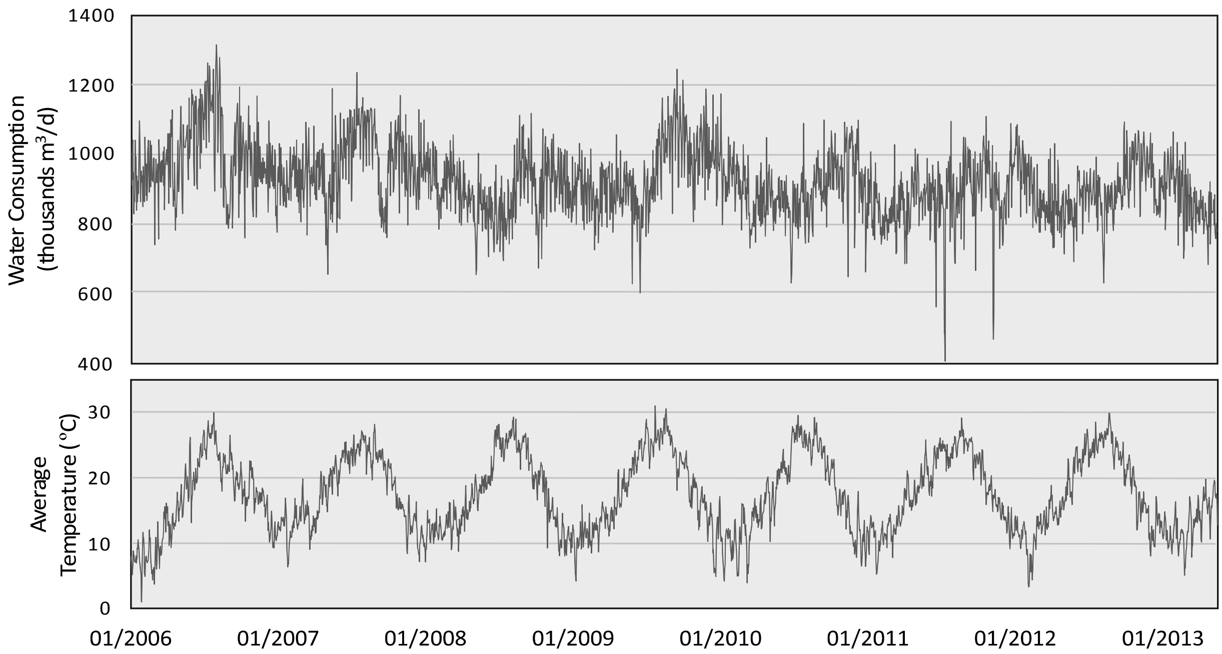


Figure 4. Daily water consumption and average temperatures in the area of Barcelona.

The day characterization dataset uses five discreet day types: normal day, strike, public holiday, school holiday and world congress. The variation of water consumption by the type of day and day of week is presented in Figure 5.

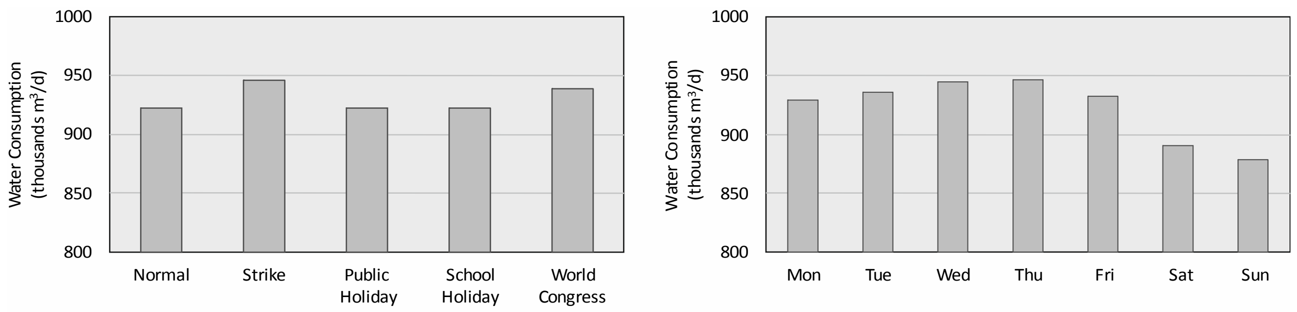


Figure 5. Variation of daily water consumption in the area of Barcelona (a) by type of day and (b) by day of week.

The model is applied to provide daily water demand forecasts for 30 test days. The results presented in Figure 6 show that the model performs satisfactorily, with an average relative forecasting error equal to 6%.

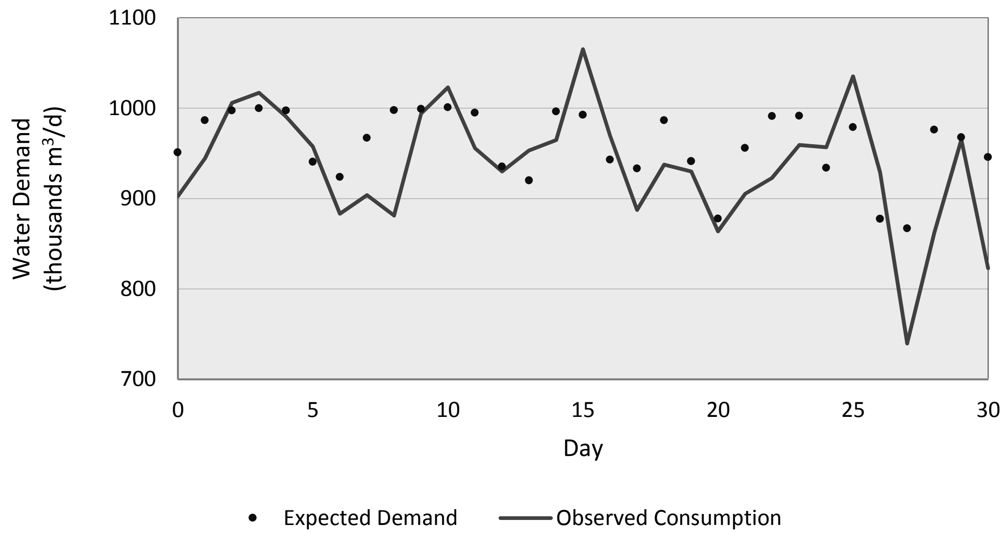


Figure 6. Daily forecasted water demand and observed consumption in the area of Barcelona.

A unique feature of the model is that, besides the forecast value, it also produces a number of similar days with their consumptions. These data can be analyzed, using standard statistical methods to provide ranges of demand or even confidence levels. As an example, Figure 7 presents the confidence interval for the forecasted water demand at a 95% level.

4. CONCLUSIONS

A methodology for short-term water demand forecasting, based on a "similar days" approach, has been presented and tested in two case studies. The model is able to provide short-term forecasts, to support the day-to-day management procedures along the water supply-distribution chain. It has the advantage that it provides accurate daily, hourly or *n*-minute demand forecasts using relatively easy to find data. It is also transparent, which means that it is easy to improve and calibrate and that it is able to generate new knowledge about the water system. It can provide a range and distribution of forecast values, in addition to a best estimate, permitting the performance of probabilistic (risk) analysis, necessary for resource optimization. Applying the model to new demand nodes or even other water systems is straightforward.

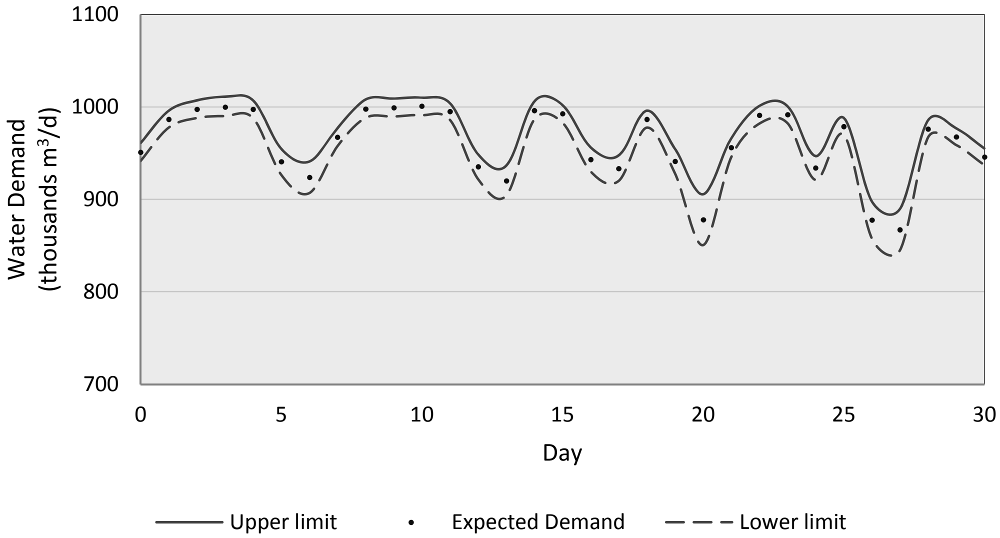


Figure 7. 95% confidence interval of forecasted demand in the metropolitan area of Barcelona.

The model has been tested using a preliminary set of data for two case studies with different water demand characteristics. Results from both applications demonstrate that small forecasting errors can be achieved by using readily available data. The average forecasting error is relatively small (4.5% for the city of Karlsruhe and 6% for the metropolitan area of Barcelona). It is concluded that this methodology can potentially provide accurate results and it could be of high value for practical use, in cases where very short term forecasting with high accuracy and temporal resolution is a requirement.

Acknowledgements

This work has been financed by the EU-funded FP7 WatERP project (Grant agreement no: 318603). The authors would like to thank partners from Water Catalan Agency (ACA), Deutsche Vereinigung des Gas und Wasserfache (DVGW) and Stadtwerke Karlsruhe GmbH (SWKA) for the provision of the data used in the analysis.

References

1. Bougadis, J., Adamowsk, K. and Diduch, R. (2005). Short-term municipal water demand forecasting. *Hydrological Processes,* **19**, 137–148.
2. Alvisi, S., Franchini, M. and Marinelli, A. (2007). A short-term, pattern-based model for water demand forecasting. *Journal of Hydroinformatics*, **9**, 39-50.
3. Jain, A., Joshi, U. C. and Varshney, K. A. (2001). Short-Term Water Demand Forecast Modelling at IIT Kanpur Using Artificial Neural Networks. *Water Resources Management*, **15**, 299–321.
4. Christian-Smith, J., Heberger, E. and Allen, L. (2012). *Urban Water Demand in California to 2100: Incorporating Climate Change*.Pacific institute, California, USA.
5. Davis, W. Y. (2003). *Water Demand Forecast Methodology for California WaterPlanning Areas - Work Plan and Model Review*.Planning and Management Consultants, California, USA.
6. Stanton, E. A. and Fitzgerald, E. (2011). *California Water Supply and Demand: Technical Report.* Stockholm Environment Institute, Somerville, USA.
7. Gato, S., Jayasuriya, N. and Roberts, P. (2004). Temperature and rainfall thresholds for base use urban water demand modelling. *Journal of Hydrology*, **337**, 364-376.

8. Zhou, S., McMahon, T., Walton, A. and Lewis, J. (2000). Forecasting daily urban water demand: a case study of Melbourne. *Journal of Hydrology*, **236**, 153-164.
9. Mu, Q., Wu, Y., Pan, X., Huang, L. and Li, X. (2010). Short-term Load Forecasting Using Improved Similar Day Method. *Power and Energy Engineering Conference (APPEEC).* Chengdu, China.
10. Srinivas, E. and Jain, A. (2009). A Methodology for Short Term Load Forecasting Using Fuzzy Logic and Similarity. *The National Conference on Advances in Computational Intelligence Applications in Power, Control, Signal, Processing and Telecommunications (NCACI- 2009).* Bhibaneswar, India.
11. Gatz, D. F. and Smith, L. (1995). The Standard Error of a Waighted Mean Concentration-I. Bootstraping vs Other Methods. *Atmospheric Environment,* **29**, 1185-1193.