Prediction of water consumption using Artificial Neural Networks modelling (ANN)

Elias Farah[1,2,3,*], *Amani* Abdallah[1,4], and *Isam* Shahrour[1]

[1]Laboratoire de Génie Civil et géo-Environnement, Université de Lille, 5900 Lille, France

[2]Civil Engineering Department, Lebanese University, Faculty of Engineering, Branch 1, North Campus, Ras Maska, Koura, Lebanon

[3]Civil Engineering Department, Holy Spirit University of Kaslik USEK, Faculty of Engineering, Jounieh, Lebanon

[4]Faculty of Technology, Lebanese University, Saida, Lebanon

Abstract. This paper presents an application of Artificial Neural Network models (ANN) to predict the water consumption at two scales: i) District Metered Area (DMA) located in the Scientific Campus of Lille University and ii) End user representing a restaurant inside this DMA. Data are collected from Automated Meter Readers (AMRs) that measure in near real-time the water consumption. The models are trained at both daily and hourly time intervals using historical values and the variation between the hour and the type of days. The paper shows that the ANN-based models can well predict the water consumption including peak values.

1 Introduction

Water demand prediction is widely required by water utilities for operational reasons in distribution networks and for updating the pricing policies. Nowadays, water management tools become more relevant due to the development of smart metering. The Artificial Neural Network (ANN) is one of the modelling approaches used in the hydroinformatics domain, especially for predicting water consumption [1, 2]. The ANNs are efficient in modelling complex linear and non-linear relationships without the need of implicit assumptions [3].

Many studies have been conducted to predict water demand using ANN models. [4] developed a model to forecast, at a daily basis, the water demand of a city considering meteorological indicators. It has been shown that time series data are necessary to be combined with other ANN inputs to get a reliable prediction.

Another forecasting model was presented by [5]. The study was based on statistical results and artificial neural networks to simulate the domestic water consumption for nine users. However, the developed models faced some difficulties to accurately predict the peak values. Limited research has been devoted to improve the forecasting of peak demand. [6] found that this peak is affected by the socio-economic and behavior changes of the residential sector.

This paper presents a methodology to predict water consumption at different scales using the ANN approach. Data were collected from a large scale demonstrator of the smart water system SunRise. The latter concerns the water distribution network consisting of 15 km of cast iron pipes supplying the Scientific Campus of Lille University located in the North of France which stands for a town of 25000 users. This water network is monitored by Automated Meter Readers (AMRs) that measure hourly the water consumption hourly and transmit the recorded data at daily time interval [7].

This paper presents first the methodology followed in this research including data collection and ANN modelling then discusses results obtained at daily and hourly time scales.

2 Methodology

2.1 Data collection and analysis

Water consumption data were obtained at hourly time interval from AMRs in the period from the 1st of January 2015 till the end of April 2016. This study is based on data related to (i) a DMA that comprises university faculties, administrative buildings and students' residences and (ii) a restaurant inside this DMA. Statistical analysis is conducted first in order to predict the water usage profile during the year and identify the boundaries of the study domain. Figure 1 shows the water consumption of the DMA and the restaurant at monthly time interval. It could be observed that the water consumption is

---

* Corresponding author: elias_farah@live.com

related to the university calendar, with a significant drop during summer holidays. This is expected since teaching and administrative buildings are closed during this period and the resident students leave the campus for summer break. In addition, the water consumption in this study is not related to the seasons or the weather variation. This result could be explained by the localization of the demo site in North of France, which has a marine west coast climate according to the Köppen Climate Classification [8] and experiences rainfall more than 150 days per year.
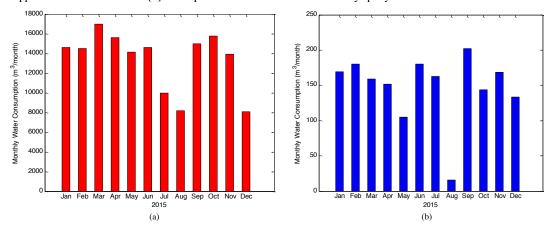

Fig. 1. Monthly water consumption in 2015 for: a) DMA and b) restaurant

2.2 Modelling approach

A feed-forward back-propagation artificial neural network is used in the numerical modelling using a Matlab script. The time series of historical water consumptions are used in the model to forecast future demand. Two main data analyses were carried out:

-   The first analysis aimed at predicting daily water consumption and forecast the demand for the next day taking into account a lag of 1 day in order to study the effect of the water consumption of the day before on the actual demand.
-   The second analysis was conducted to predict the water consumption for the next day with a step of 1 hour using the historical values of the day before, with a lag length of 24 hours. This lag interval is chosen due to the transmission of recorded data at 24 hours' time interval.

The input vector for the daily basis includes the day of the week DW_i (i varies from 1 to 7, i.e. Monday, Tuesday, Wednesday, etc.), the holidays HO (i.e. automn, winter, spring and summer holidays) and the special days SD (i.e. New Year's day, Easter Monday, Labor day, etc.). The input values are either 0 or 1. In other terms, for the DW_i input, if the day is a Monday, the value will be set to 1 and the others days will be assigned as 0. For the daily prediction, a new vector is added to the input consisting of the water consumption of the day before Q_{d-1}. The target data set correspond to the daily water consumption.

Analysis at hourly time interval required the consideration of the additional following parameters: a vector representing the hour of the day HD_i (i varies from 0 to 23) and the time series of historical water consumption with a lag of 24 hours Q_{h-24}.

A pre-processing step is necessary to filter the values from any abnormal event due to a reported leak or to any faults in the signal. The anomalies in time series are identified using Chebyshev's inequality [9]. This theorem guarantees that in any probability distribution, the majority of the values are close to the mean value. Using this inequality, the values are filtered to eliminate the outliers that exceed the average plus three standard deviations (i.e. 89% of the values lies within two standard deviations).

The input dataset is then divided into three sets: 70% for training, 15% for validation and the remaining 15% corresponds to the testing set. The input and target variables are normalized in the range of 0 and 1 according the following equation:

$$\overline{X} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where X^- is the normalized input or target variables X, X_{min} and X_{max} represent the minimum and the maximum of observation values respectively. The aim of normalization is to avoid any prioritization of the variables and to remove any arbitrary effect of similarity between the objects [10].

The performances of the ANN prediction models are evaluated by the root mean square error (RMSE) and the coefficient of determination (R^2):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{2}$$

$$R^2 = 1 - \frac{\sum\limits_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum\limits_{i=1}^{N}(y_i - \bar{y})^2} \qquad (3)$$

3 Results and Discussions

3.1 Daily prediction analysis

Collected data for the DMA and the restaurant are trained using the water consumption of the previous day (i.e. lag equal to 1 day). Table 1, figures 2 and 3 show the quality of the ANN predictions.

For the DMA, a good correlation is observed between recorded and predicted values (R^2 is equal to 0.827 and RMSE is equivalent to 0.110); the ANN model predicts well the water consumption, including the peak and minimum values. The ANN model performances for the restaurant are yet better with a factor R^2 equal to 0.902 and RSME equivalent to 0.1.

Table 1. Statistical accuracy measures of ANN models at training and testing sets (daily analysis)

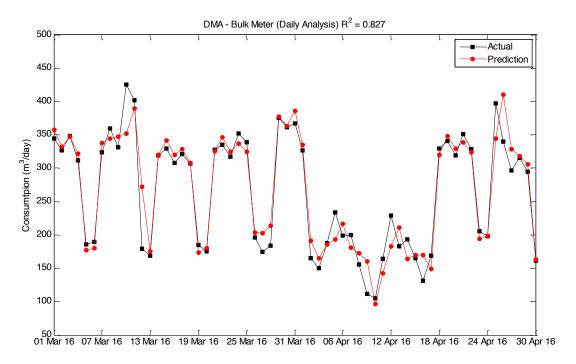| AMR	Set	Lag = 1 day	
		RMSE	R^2
DMA	Training	0.112	0.814
(Bulk meter)	Testing	0.110	0.827
Restaurant	Training	0.089	0.918
(End user meter)	Testing	0.100	0.902

Fig. 2. Comparison between recorded and ANN prediction for the DMA (daily-time interval)
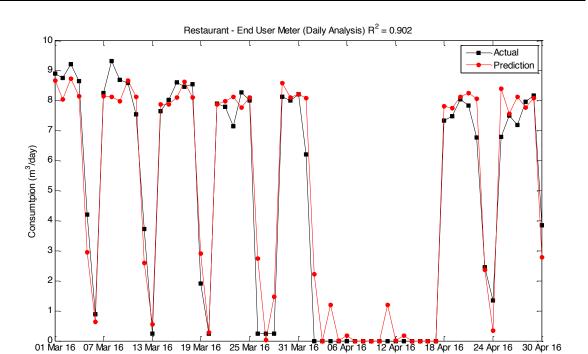
Fig. 3. Comparison between recorded and ANN prediction for the restaurant (daily-time interval)

3.2 Hourly prediction analysis

In order to model the water demand in quasi real time even, different ANN models were trained taking into account the water consumption at the hourly scale. A set of vector representing the hour of the day HD_i (i varies from 0 to 23) is added to the input matrix in addition to the day of the week, holidays and special days variables and the time series of historical values with a lag of 24 hours.

Table 2 and figure 4 show the quality of the ANN predictions. For the DMA, the ANN model predicts well the water consumption including the peak and minimum values with $R^2 = 0.825$ for the ANN testing phase. Results obtained for the restaurant are also very good with $R^2 = 0.884$ which is slightly higher than the correlation from the data of the DMA. This can be explained by the fact that data from the restaurant's water meter represent a unique profile whereas the DMA includes different behaviors of water consumption.

Table 2. Statistical accuracy measures of ANN models at training and testing sets (hourly analysis)

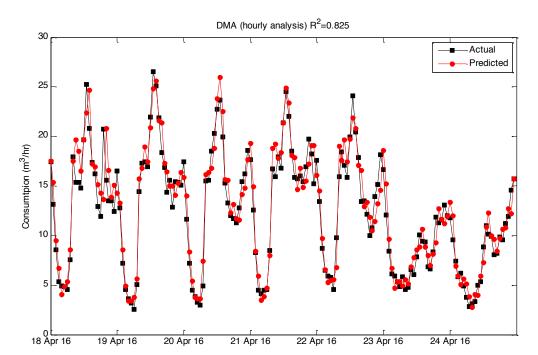| AMR	Set	Lag = 24 hours	
		RMSE	R^2
DMA	Training	0.062	0.857
(Bulk meter)	Testing	0.070	0.825
Restaurant	Training	0.054	0.900
(End user meter)	Testing	0.059	0.884

Fig. 4. Comparison between actual and ANN predicted hourly consumption values for the DMA

4 Conclusion

This paper presented the use of feed-forward back propagation ANN models for the prediction of the water consumption in the campus of Lille University at both daily and hourly time intervals. Analyses showed good performances of the ANN model for both time-scales, including peak values. For the hourly prediction, considering the historical values with a lag of 24 hours, the time and day of the week, holidays and special days, the root mean square error equals 0.07 at the DMA scale and 0.059 at the end-user one.

The ANN-based model could be easily implemented and used for short-time consumption prediction. Coupled to real-time consumption recording, they could indicate anomalies in water consumption, in particular those related to water leak. This method could also be used to estimate "recover" missing data.

References

1. Guo, G., and Liu, S., Short-term water demand forecast based on deep neural network. *1st International WDSA / CCWI 2018 Joint Conference, Kingston, Ontario, Canada* (2018)
2. Romano, M., and Kapelan, Z., Adaptive water demand forecasting for near real-ti-me management of smart water distribution systems, *ENVIRON MODELL SOFTW*, **60**, 265-276 (2014)
3. Somers, M. J., and Casal, J. C., Using artificial neural networks to model nonlinearity: The case of the job satisfaction-job performance relationship, *Organ. Res. Methods* , **12 (3)**, 403-417 (2009)
4. Al-Zahrani, M. A., and Abo-Monasar, A., Urban residential water demand prediction based on artificial neural networks and time series models, *WATER RESOUR MANAG*, **29 (10)**, 3651-3662 (2015)
5. Walker, D., Creaco, E., Vamvakeridou-Lyroudia, L., Farmani, R., Kapelan, Z., and Savić, D., Forecasting domestic water consumption from smart meter readings using statistical methods and artificial neural networks, *Procedia Eng.,* **119**, 1419-1428 (2015)
6. Day, D., and Howe, C., Forecasting peak demand-what do we need to know? *WATER SCI TECH-W SUP*, **3(3)**, 177-184 (2003)
7. Farah, E., and Shahrour, I., Smart water technology for leakage detection: feedback of large-scale experimentation, *ANALOG INTERGR CIRC S*, **96 (2)**, 235-242 (2018)
8. Peel, M. C., Finlayson, B. L., and McMahon, T. A., Updated world map of the Köppen-Geiger climate classification, *HYDROLOGY AND EARTH SYSTEM SCIENCES*, **4 (2)**, 439-473 (2007)
9. Kvanli, A., Pavur, R., and Keeling, K., Concise managerial statistics. Mason, OH: South-Western Thomson Learning (2006)
10. Riad, S., Mania, J., Bouchaou, L., and Najjar, Y., Rainfall-runoff model usingan artificial neural network approach, *MATH COMPUT MODEL*, **40 (7)**, 839-846 (2004)