Journal of Hydrology 558 (2018) 392–404

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Increased performance in the short-term water demand forecasting through the use of a parallel adaptive weighting strategy

A. Sardinha-Lourenço [a], A. Andrade-Campos [a,*], A. Antunes [a], M.S. Oliveira [b]

[a] *Department of Mechanical Engineering, Centre for Mechanical Technology & Automation, GRIDS, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal*
[b] *Department of Economics, Management, Industrial Engineering and Tourism, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal*

A R T I C L E I N F O

Article history:
Received 29 November 2017
Received in revised form 19 January 2018
Accepted 21 January 2018
Available online 2 February 2018
This manuscript was handled by G. Syme, Editor-in-Chief

Keywords:
Water demand
Forecasting
Adaptive weighting model
Short-term

A B S T R A C T

Recent research on water demand short-term forecasting has shown that models using univariate time series based on historical data are useful and can be combined with other prediction methods to reduce errors. The behavior of water demands in drinking water distribution networks focuses on their repetitive nature and, under meteorological conditions and similar consumers, allows the development of a heuristic forecast model that, in turn, combined with other autoregressive models, can provide reliable forecasts. In this study, a parallel adaptive weighting strategy of water consumption forecast for the next 24–48 h, using univariate time series of potable water consumption, is proposed. Two Portuguese potable water distribution networks are used as case studies where the only input data are the consumption of water and the national calendar. For the development of the strategy, the Autoregressive Integrated Moving Average (ARIMA) method and a short-term forecast heuristic algorithm are used. Simulations with the model showed that, when using a parallel adaptive weighting strategy, the prediction error can be reduced by 15.96% and the average error by 9.20%. This reduction is important in the control and management of water supply systems. The proposed methodology can be extended to other forecast methods, especially when it comes to the availability of multiple forecast models.

1. Introduction

The main objective in the management of drinking water distribution systems is to satisfy the demand of the consumers satisficing the continuous conditions of quality, flow, and adequate pressure thus ensuring a reliable distribution system. Efficiently managing and operating a potable water supply systems requires short-term forecasts of the water demand demanded by consumers (Adamowski et al., 2012; Coelho, 2016). The estimation of the future demand of water is fundamental for the planning of a regional system of water supply, since it will allow the reduction of costs (Zhou et al., 2002). These water demands are highly variable and fluctuate according to the type and size of the consumer, the time of day, the day of the week, the season of the year, the weather and even with the days of celebrations, extraordinary events and cost of supply of this service. Specific consumption patterns for each distribution network can be handled and interpreted skillfully by operators with large accumulated experience and used for operational control of the network. However, the use of these patterns is done in many cases manually, compromising efficiency and functioning of the network costs.

In the literature, it is possible to find numerous models developed by statistical methods, typically multiple regression and time series, to predict urban water consumption. For the daily operation of treatment plants and pumping stations, a short-term forecast model is needed for the next 24 h (Bakker et al., 2013a).

Many water demand forecasts models used in the literature use a one-hour step to study time series (Jowitt and Chengchao, 1992, Shvartser et al., 1993; Homwongs et al., 1994; Alvisi et al., 2007; Santos and Filho, 2014). These works show that it is possible to generate fairly accurate forecasts using as single input the historical demand (Msiza et al., 2008). Other forecast models include meteorological information as additional input. The model proposed by Zhou et al. (2002) uses the daily rainfall, the maximum temperature of the day, the number of days since the last rainfall and the effect of evaporation. The artificial neural networks (ANNs) model of Ghiassi et al. (2008) uses hourly temperature values and the ANNs model of Herrera et al. (2010) uses daily values of temperature, wind speed, atmospheric pressure, and rain.

Babel and Shinde (2011) evaluated the effect of weather variables as ANN inputs for daily and monthly water demand forecast in the city of Bangkok (Thailand). In the daily forecasts,

* Corresponding author.
E-mail address: gilac@ua.pt (A. Andrade-Campos).

https://doi.org/10.1016/j.jhydrol.2018.01.047
0022-1694/

no significant differences were found in the forecast accuracy when including weather variables (rainfall, average temperature, and relative humidity) in their models. In the work of Tabesh and Dini (2009), the best results were obtained with the ANN models and neural-fuzzy approaches considering only past water demand variables as input.

The quality and reliability of the input data for the analysis of a time series are crucial since it is the only information used to determine the future values. Critical Infrastructure Systems (CIS), including water distribution network systems, are large in size and occupy a large geographic space. These systems require a monitoring and control system in real time to ensure the maintenance of the variables in acceptable conditions for as long as necessary (Quevedo et al., 2010). Eventually, these variables suffer appreciable deviations due to failures (e.g. sensor and/or actuator and/or malfunctioning of pipelines) (Schütze et al., 2004). In the CIS, the telecontrol system has the function of acquiring, storing and validate data collected from different types of sensors in each time sample and thus monitoring in real time the whole system. Several problems can occur during the acquisition of process data, such as for example, those related to communication failure between sensors and the telecontrol system itself. These problems cause lost or damaged data. In that case, the missing data must be replaced by a set of estimated data, which should be representative of the data that was lost (Quevedo et al., 2010).

There are a large number of papers detailing various methodologies for short-term forecasts using different time scales: 15-min intervals (Bakker et al., 2013b); per hour (eg Shvartser et al., 1993; Zhou et al., 2002; Alvisi et al., 2007) as well as daily/monthly time scales (Maidment and Parzen, 1984a; Franklin and Maidment, 1986; Smith, 1998; Miaou, 1990). In all these documents, recurrent patterns and the periodicities that exist in the data of water demand, at different levels of temporal aggregation, were recognized.

Most models of prediction of water demand have a limited number of demand patterns. Jowitt and Chengchao (1992) and Homwongs et al. (1994) used three different water consumption patterns for the development of their model: one for weekdays, one for Saturdays and one for Sundays. The model described by Zhou et al. (2002) uses only two different patterns: working days and another for national holidays and weekends. Alvisi et al. (2007) propose the identification of the patterns implicit in the time series of the water demand of the Castelfranco Emilia municipality, Italy. The study was conducted for daily water demand for a whole year, finding that the demands increase during the summer period and during the week from Sunday to Saturday. In the same way, the demands of water per hour show a variable behavior during the day, with different patterns depending on if it is a weekday or weekend. According to their analysis, the daily and hourly series observed present a pattern of demand during the holidays very similar to the one observed during the weekends.

Recently, Bakker et al. (2013b) proposed a model that uses a relatively greater number of demand patterns, discern not only the demand patterns for the seven days of the week, but also for a series of day types that deviate from the common days of the week, such as primary school vacation periods and private day events related to multiple activities for the Netherlands. In this study, national holidays are treated as Sundays.

The developed heuristic model predicts water demand for the next 48 h with 15-min steps. The model determines water demand in three main phases: in the first phase, the average water demand is forecasted for the next 48 h. In the second phase, normal water demands are anticipated for individual steps of 15 min. In the third phase, if necessary, additional water demands are anticipated for reasons of increasing the ambient temperature (summer) for the individual 15-min stages.

The work of Bakker et al. (2013b) focuses on the repetitive nature of the behavior of water demands in distribution networks. For the development of an adaptive heuristic forecast model, he used a large amount of historical data from the water distribution networks for conditions and type of similar consumers, allowing him to determine particular parameters of daily behavior for his forecast model. For the calculation of these parameters, the model requires many days of historical information, which represents a disadvantage if consumption undergoes shorter periodic variations in its behavior and there is a large amount of continuous historical information validated.

Winkler and Makridakis (1983) affirm that the combination of forecasts improves accuracy and using simple averages, in the combination of forecasts, also reduces the variability of the errors in the forecast and, therefore, the risk associated with the choice of forecasting method. Recent research on short-term demand forecasting has shown that models using univariate time series based on historical data are useful and can be combined with other prediction methods to reduce errors. Therefore, the forecasts can be combined by using simple and optimal weights.

Stock and Watson (1999) proposed several pooling procedures that differ by the amount of weight placed on the model as a function of the currently best performance. These procedures include the equally weighting of all the forecast, the weighting inversely proportional to their current mean squared error (MSE), using average forecast, and placing all weight on the forecasting method that currently has the lowest simulated real-time MSE. The final pooling procedure is simulated using real-time model selection.

The methodology used in Caiado (2010) was to weight the forecasts of three models with the inverse of the mean squares errors (MSE) of each of the individual methods. The average error difference when using individual prediction methods to predict a single day was 8.33% greater than combining forecasting methods. In the prediction of 2 and 3 days, combined forecasts can reduce error by 12.77% and 10.64%, respectively (Caiado, 2010).

In this study, a parallel adaptive weighting model of water consumption forecast for the next 24–48 h, using univariate time series of drinking water consumption, is proposed. For the development of the model, it is used a seasonally integrated autoregressive integrated mobile averages (ARIMA) and a short-term forecasting heuristic method (Bakker et al., 2013a,b), which takes into account the multiple daily water consumption patterns. This last method classifies the historical information of the water demands of the working days and the typical days and then makes the forecasts. A third method is developed from the classification made previously and the use of the ARIMA methodology. Adaptive weighting parameters are calculated for each one of the 24–48-h parallel forecasts methods according to the minimum mean absolute percentage error (MAPE) obtained for the day previous. The model was validated using Portuguese networks as case studies.

This forecasting strategy that uses a combination of optimal weights allows the proposed model to be adaptive, simple to implement and scalable for the incorporation of a larger number of forecast algorithms.

2. Water demand forecasting

2.1. Time series analysis

The temporal series are generally analyzed from a deterministic point of view. However, nowadays, time series are also studied from a stochastic point of view, using more complex methods and their application requires larger data. Statistical models for time series were developed by Box and Jenkins (1976) that consider the dependence between the data. The analyses are based