Journal of Hydrology 517 (2014) 236–245

Contents lists available at ScienceDirect

Journal of Hydrology

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

A multi-scale relevance vector regression approach for daily urban water demand forecasting

Yun Bai [a], Pu Wang [a], Chuan Li [b,*], Jingjing Xie [c], Yin Wang [a]

[a] School of Urban Construction and Environmental Engineering, Chongqing University, Chongqing 400044, China
[b] Engineering Laboratory for Detection, Control and Integrated Systems, Chongqing Technology and Business University, Chongqing 400067, China
[c] Testing Center for Science and Technology, Chongqing Academy of Science and Technology, Chongqing 401123, China

ARTICLE INFO

Article history:
Received 20 April 2013
Received in revised form 7 October 2013
Accepted 8 May 2014
Available online 22 May 2014
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

Keywords:
Water demand forecasting
Stationary wavelet transform
Relevance vector regression
Multi-scale
Adaptive chaos particle swarm optimization

SUMMARY

Water is one of the most important resources for economic and social developments. Daily water demand forecasting is an effective measure for scheduling urban water facilities. This work proposes a multi-scale relevance vector regression (MSRVR) approach to forecast daily urban water demand. The approach uses the stationary wavelet transform to decompose historical time series of daily water supplies into different scales. At each scale, the wavelet coefficients are used to train a machine-learning model using the relevance vector regression (RVR) method. The estimated coefficients of the RVR outputs for all of the scales are employed to reconstruct the forecasting result through the inverse wavelet transform. To better facilitate the MSRVR forecasting, the chaos features of the daily water supply series are analyzed to determine the input variables of the RVR model. In addition, an adaptive chaos particle swarm optimization algorithm is used to find the optimal combination of the RVR model parameters. The MSRVR approach is evaluated using real data collected from two waterworks and is compared with recently reported methods. The results show that the proposed MSRVR method can forecast daily urban water demand much more precisely in terms of the normalized root-mean-square error, correlation coefficient, and mean absolute percentage error criteria.

1. Introduction

Water is an irreplaceable resource and plays an important role in economic and social development. Water consumption increases dramatically with increased urbanization and improved living standards, and the increasing demand for water can lead to conflicts over existing water supply facilities. Therefore, the existing water infrastructure needs to be effectively used. Water demand forecasting is an effective measure for scheduling urban water facilities. To this end, researchers and engineers have developed different methods for different forecasting horizons.

The literature shows that the wide variety of existing methods and models have different applications depending on the periodicity and forecast horizon of the forecast variable(s) (Donkor et al., 2012). Billings and Jones (2008) defined water demand forecasts spanning more than 2 years as long term, those from 3 months to 2 years as medium term, and those less than 3 months as short term. For long-term forecasting, Alhumoud (2008) used a univariate time series method to assess the relationship between water consumption and its determinants. Lee et al. (2010) introduced a regression method to estimate water demand based on population density. Wei et al. (2010) proposed a scenario-based method (econometric model) to forecast water demand under different scenarios. Mohamed and Al-Mualla (2010) presented a constant rate model to forecast water demand for 20-year and 30-year horizons. Li et al. (2012a,b) pointed out that price elasticity is one of the main factors affecting urban water consumption. For medium-term forecasting, Babel and Shinde (2011) suggested an artificial neural network (ANN) method to forecast six-month water demand in Bangkok. Ziervogel et al. (2010) employed information on seasonal climatic change to manage water resources. For short-term forecasting, Herrera et al. (2010) compared different models for forecasting hourly water demand, and indicated that the support vector regression was the best for the given case. Ticlavilca et al. (2013) proposed a robust multivariate Bayesian learning model to forecast irrigation demand. Other studies (Adamowski and Karapataki, 2010; Caiado, 2010; Odan and Reis, 2012; Wong et al., 2010) recommended using hybrid approaches to improve forecasting accuracy.

Artificial intelligence (AI) algorithms have been proven effective in long-term, medium-term, and short-term water demand

* Corresponding author. Tel.: +86 (23) 6276 8469.
E-mail address: chuanli@21cn.com (C. Li).

http://dx.doi.org/10.1016/j.jhydrol.2014.05.033
0022-1694/© 2014 Elsevier B.V. All rights reserved.

forecasting. Tabesh and Dini (2009) applied fuzzy and neuro-fuzzy models to forecast short-term water demand. Beal et al. (2011) introduced a mixed smart metering approach to reconcile the differences between perceived and actual residential end use water consumption. Nasseri et al. (2011) used the extended Kalman filter and genetic programming to forecast monthly urban water demand. In addition, different wavelet transform (WT) methods have been used for scale analyses and feature extractions to improve the performance of time series forecasting (Adamowski, 2008; Campisi-Pinto et al., 2012; Maheswaran and Khosa, 2012).

We propose a multi-scale relevance vector regression (MSRVR) approach for forecasting urban daily water demand. The nonlinear mapping capability of the RVR and the multi-resolution characteristics of the WT are integrated to improve the forecasting accuracy. Time series of daily urban water demand are decomposed into different scales using the stationary wavelet transform (SWT). In each SWT scale, an RVR model is developed to generate the wavelet coefficient of the next-day water demand. The inverse SWT is subsequently employed to reconstruct the next-day water demand using the wavelet coefficients of all the scales. To facilitate the aforementioned forecasting procedure, the chaos feature of the daily water supply series is analyzed to determine the input variables of the RVR. An adaptive chaos particle swarm optimization (ACPSO) is used to simultaneously optimize the structural parameters of all the RVR model parameters. As these results in the selection of reasonable parameters, the developed approach is more suitable for real applications.

The remainder of the paper is organized as follows. Section 2 analyzes the characteristics of daily water supply series and presents the forecasting performance criteria. The proposed MSRVR approach is outlined in Section 3. In Section 4, water supply data collected from a waterworks are employed to evaluate the effectiveness of the proposed method, which is also compared with other recently reported methods. Conclusions are given in Section 5.

2. Time series characteristics and forecasting performance criteria

In this section, we first analyze the characteristics of daily water supply series data from a real waterworks. The chaos characteristics of the time series are then utilized to determine the input variables of the regression models. Three criteria for evaluating forecasting performance, namely, the normalized root-mean-square error (NRMSE), the correlation coefficient (CC), and mean absolute percentage error (MAPE), are also introduced in this section.

2.1. Chaos characteristics of daily water demand series

To analyze the characteristics of daily water demand series, water supply records for 360 days from January to December 2011 were collected from an urban waterworks with a daily supply capacity of up to 25 million cubic meters. The waterworks is located at 106°34′10″E/29°32′26″N (Chongqing municipality, China) and is referred to as #1 urban waterworks hereafter.

The time series of the collected data is shown in Fig. 1. Although an upward trend is observable in Fig. 1, the evolution law and variation characteristics of the data cannot be deduced directly from the time series. For this reason, we use a phase diagram and a power spectrum (Kugiumtzis, 1996) to qualitatively analyze the characteristics of the time series.

Fig. 2(a) and (b) plot the phase diagrams of the first order derivative of the water supply series using 1d delay and 2d delay, respectively. All possibilities of the steady-state system can be intuitively observed from the phase space geometry of the phase diagram. The orbital characteristics and the direction of each steady state can also be discerned. Furthermore, an attraction domain that attracts the movement of each phase can be identified in both the 1d (Fig. 2(a)) and 2d delays (Fig. 2(b)), illustrating that the time series has only one possible future. This indicates that the time series of the water demand exhibits a predictable pattern rather than unpredictable random motion.

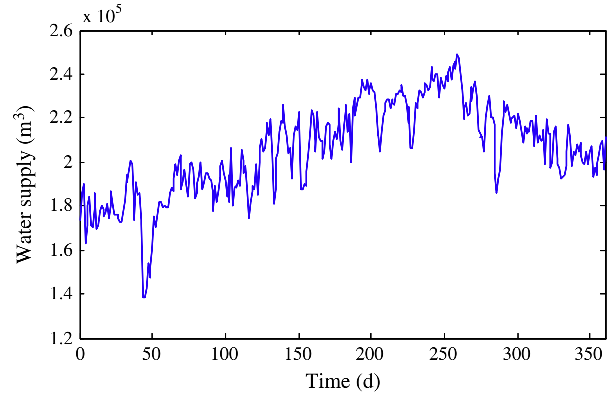

Fig. 1. Daily water supply series of #1 urban waterworks from January to December, 2011.

The power spectrum and the power spectral density (PSD) of the water supply series are displayed in Fig. 3(a) and (b), respectively. The power spectrum generated by the Fourier analysis is capable of distinguishing states of regularity and irregularity in the time series. From Fig. 3(a), no evident peak can be observed in the continuous spectrum. Moreover, as revealed in Fig. 3(b), the PSD displays continuity and broad-peak characteristics. In all, Fig. 3 suggests that the time series can be directly linked to the chaotic motion (Oshima and Kosuda, 1998).

The analysis of the water supply series using the phase diagram and the power spectrum reveals the chaos feature that is a common phenomenon in nonlinear systems. Our result is identical to that obtained by Zhao and Zhang (2008) who also found a chaos daily water demand series. According to chaos theory (Li and Chen, 2010), it is meaningful to forecast for a chaos time series, as this reflects the ergodicity, randomness, and regularity of the series itself. Therefore, in the following subsection, we can determine the input variables for the daily water demand forecasting.

2.2. Determining input variables for regression models

As described in the previous subsection, the daily water demand is predictable due to the predictability of the chaos time series. The first task for water demand forecasting (regression modeling) is to choose the input variables for the estimation model. With the 360-day series shown in Fig. 1, the first 290 daily data are used as the training data set, while the rest comprise the testing set.

We use a data-driven method to determine the general structure of the regression model for the daily urban water demand forecast. The forecast can be expressed as

$$\widehat{W}(i) = f(W(i-t_{\mathrm{opt}}), W(i-2t_{\mathrm{opt}}), \ldots\ldots, W(i-m_{\mathrm{opt}}t_{\mathrm{opt}})), \tag{1}$$

where $\widehat{W}(i)$ represents the forecasted value of the ith daily water demand, W denotes the historical data of the real water supply, t_{opt} is the optimal delay time, and m_{opt} stands for the optimal embedding dimension of the input variables.

According to Eq. (1), the input variables of the regression model are in fact determined by the selection of t_{opt} and m_{opt}. The autocorrelation method and the saturated correlation dimension