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/

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

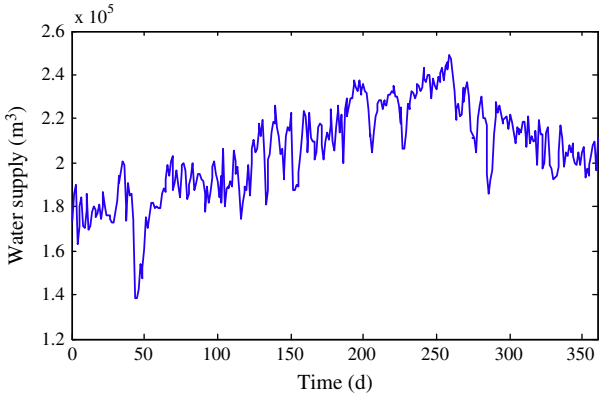

Fig. 1. Daily water supply series of #1 urban waterworks from January to December, 2011.

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.

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

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

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

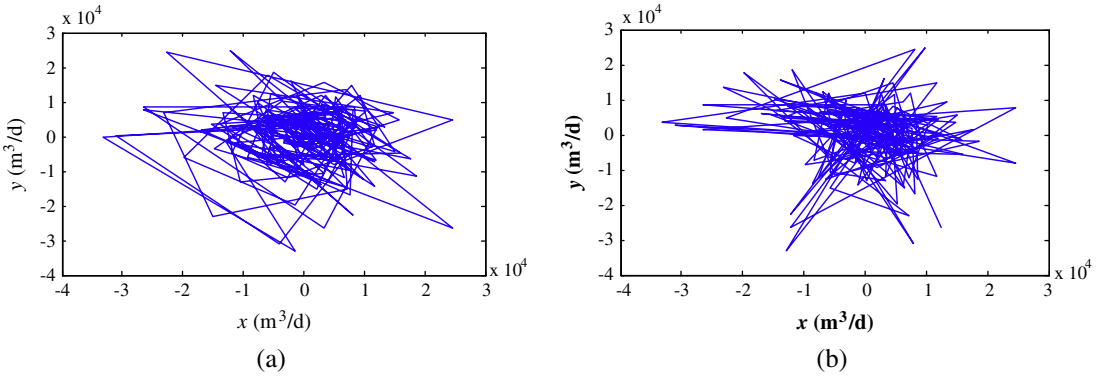

Fig. 2. The phase diagrams of daily water supply with different time delay: (a) 1d delay; and (b) 2d delay.

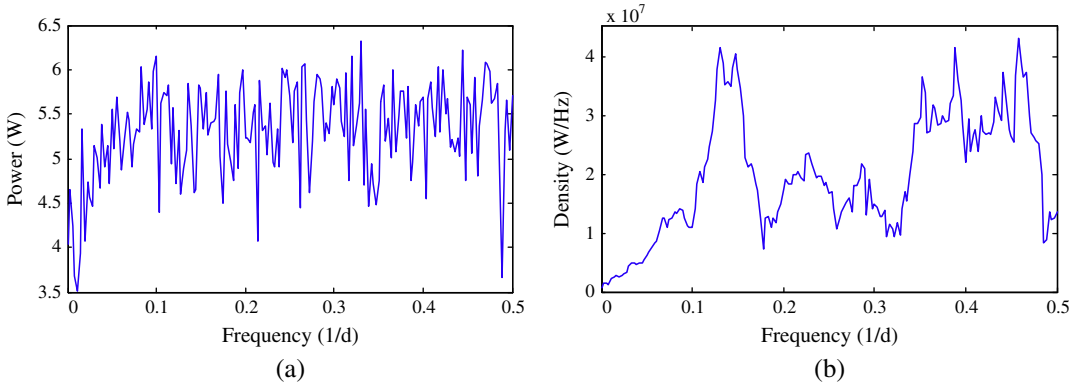

Fig. 3. Power spectrum diagrams of the water supply series: (a) power spectrum; and (b) power spectral density.

method are applied to determine the delay time t_{opt} and the embedding dimension m_{opt}, respectively (Holzfuss and Mayer-Kress, 1986).

To determine the optimal delay time using the autocorrelation method, we can calculate the autocorrelation function (ACF), which is given by

$$C(i) = \frac{1}{N}\sum_{i=1}^{N} W(i)W(i+t), \tag{2}$$

where C is the ACF and N means the length of the time series. The optimal delay time t_{opt} can be determined by decreasing the ACF to 1–1/e times the initial value. According to a numerical experiment on the time series shown in Fig. 1, the optimal delay time in our study is $t_{opt} = 1$ using the above equation.

After obtaining the optimal delay time t_{opt}, we calculate the optimal embedding dimension m_{opt} using the saturated correlation dimension method. In this algorithm, the correlation dimension $D(m)$ is defined as

$$D(m) = \lim_{r\to 0}\frac{\log(CD(r,m))}{\log(r)}, \tag{3}$$

where $CD(r, m)$ is the cumulative distribution function, which describes the probability that the distance between two points of an attractor is less than r in the phase space. Fig. 4 shows the relationship between $D(m)$ and m with $t = 1$. $D(m)$ can be seen to increase with the increase of m, and the rate of increase reduces when m is greater than 7. Hence, the optimal embedding dimension is chosen as $m_{opt} = 7$. This also partially conforms to the weekly water consumption habits.

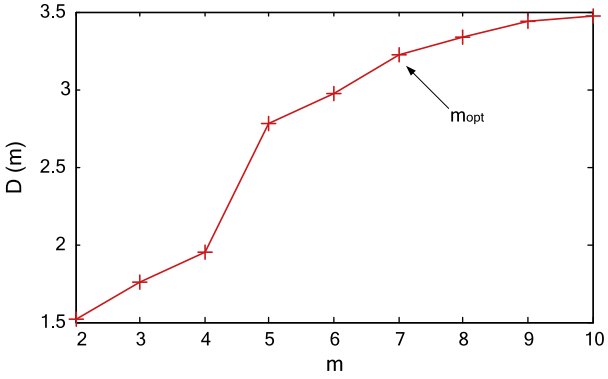

Fig. 4. Values of the correlation dimension $D(m)$ with different m candidates ($t = 1$).

2.3. Forecasting performance criteria

Different criteria have been used to measure forecasting performance (Cutore et al., 2008; Li et al., 2012a, 2012b; Li and Wang, 2006). We use the normalized root-mean-square error (NRMSE) to optimize our forecasting approach. The NRMSE is defined as

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(W(i)-\widehat{W}(i))^2}}{\sum_{i=1}^{N}W(i)/N}. \tag{4}$$

Because the NRMSE only represents the total precision of the forecasting, we also introduce the mean absolute percentage error (MAPE) and the correlation coefficient (CC) to assess the forecasting effectiveness. The MAPE is given by

$$\mathrm{MAPE} = \frac{\sum_{i=1}^{N}\left|\frac{W(i)-\widehat{W}(i)}{W(i)}\right|}{N} \times 100\%. \quad (5)$$

The CC is defined as (Maheswaran and Khosa, 2012)

$$\mathrm{CC} = \frac{\sum_{i=1}^{N}(W(i)-\overline{W})(\hat{W}(i)-\overline{\widehat{W}})}{\sqrt{\sum_{i=1}^{N}(W(i)-\overline{W})^2}\cdot\sqrt{\sum_{i=1}^{N}(\hat{W}(i)-\overline{\widehat{W}})^2}}, \quad (6)$$

where the hat "‾" represents the mean value.

In measuring forecasting performance, a greater CC means better agreement between the measured and forecast values, whereas a greater NRMSE or MAPE indicates worse performance of the forecasting model. Hence, the NRMSE is chosen as the main criterion for the model optimization, and the CC and the MAPE are taken as the reference criteria for the performance evaluations.

3. Illustration of the proposed MSRVR approach

In this section, we describe the systematic methodology of the multi-scale RVR (MSRVR) approach in detail. Combining the nonlinearity of the RVR and the multi-scale analysis of the wavelet, the proposed MSRVR model comprises a three-step procedure: (1) the daily water supply series is decomposed by the wavelet for multi-scale analysis; (2) the wavelet coefficients are used to build the RVR model in each scale; and (3) the final forecasting result is obtained by inverse wavelet reconstruction of the output coefficients of all the RVR models.

In the following subsections, we outline the constituent techniques involved in the MSRVR step-by-step. In the first subsection, we introduce the relevance vector machine (RVM) used to build the regression model for the water demand series. Taking into consideration the intrinsic regularity of the water series, the SWT is introduced in the second subsection for the multi-scale analysis of the regression model. In the third subsection, the ACPSO is applied to solve the parameter optimization issue of the MSRVR. The final subsection provides an overview of the MSRVR approach based on the preceding theoretical results.

3.1. RVR modeling

Developed by Tipping (2000), the RVM is a probabilistic model based on Bayesian theory. Similar to the support vector machine (SVM), the RVM transforms low-dimensional nonlinear problems into high-dimensional spaces based on kernel function mapping. However, the RVM does not require following the Mercer condition in choosing the kernel function for the SVM. As a result, the computational burden of the RVM is reduced compared with the SVM. Because the RVM is appropriate for dealing with the time series with small samples and nonlinear problems, it is used to build an RVR model for forecasting urban water demand in this paper.

As analyzed in Section 2, $m_{\mathrm{opt}} = 7$ and $t_{\mathrm{opt}} = 1$ in determining the input variables for the regression model. The input vector of the RVR model can thus be chosen as $\mathbf{X}(i) = [W(i-1), W(i-2), \ldots, W(i-7)]$, and the target vector as $\mathbf{F}(i) = [W(i)]$ for the training sample. We aim to use the "training" set to model the dependency of the target on the inputs with the objective of making accurate predictions of $\hat{\mathbf{F}}$ for previously unseen values of $\mathbf{X}^*$.

Typically, forecasts are based on the function $y(\mathbf{X}^*; \boldsymbol{\omega})$ defined over the input space and learning the process of inferring the parameter of the function

$$y(\mathbf{X}^*;\boldsymbol{\omega}) = \sum_{i=1}^{N}\omega_i K(\mathbf{X}^*,\mathbf{X}) + \boldsymbol{\omega}_0, \quad (7)$$

where $K(\mathbf{X}^*, \mathbf{X})$ is the kernel function and ω is the weight vector. With the Bayesian framework applied in RVR, the goal of the RVR model can be formulated as

$$\widehat{\mathbf{F}} = y(\mathbf{X}^*;\boldsymbol{\omega}) + \boldsymbol{\varepsilon}, \quad (8)$$

where $\boldsymbol{\varepsilon}$ is assumed to be the mean-zero Gaussian noise with variance σ^2.

RVR modeling has been detailed by Tipping (2001) and can be summarized for daily urban water demand forecasting as follows: (a) set the daily water demand $\mathbf{F}$ as the target of the model; (b) choose the input variables of the RVR model using t_{opt} and m_{opt}; (c) determine the appropriate kernel function (in this paper, the Gauss kernel function is chosen as the kernel function given by

$$K(\mathbf{X}^*,\mathbf{X}) = \exp\left[-\frac{\|\mathbf{X}^*-\mathbf{X}\|^2}{2\delta^2}\right], \quad (9)$$

where δ is the kernel bandwidth); and (d) maximize the marginal likelihood of obtaining the trained RVR model. With these four steps, the obtained RVR model can be used to forecast daily urban water demand.

As mentioned, the kernel bandwidth δ of the Gaussian kernel is the parameter used to determine the application of the RVR model for urban daily water demand forecasting. The kernel function implies a data similarity measure, while the kernel function parameter (δ in Eq. (9)) plays an amplification role. Next, we illustrate the influence of δ on the regression performance using the water supply series shown in Fig. 1.

The first 290 days of data, as shown in Fig. 1, are employed for the RVR modeling, and the remaining 70 days of data are used for the model testing. The kernel bandwidth δ is chosen as 1, 2, 3, 4, 5, and 6, respectively. The number of iteration steps is set as 1000. The RVR results in relation to the NRMSE, MAPE, and CC criteria are listed in Table 1.

It can be observed from Table 1 that with the change of δ between [1, 6], the values of the NRMSE, MAPE, and CC criteria using the modeling data set vary slightly. In contrast, the value of δ evidently affects the forecasting performance of the testing data set. When $\delta = 3$, the main performance criterion, the NRMSE, reaches the minimum, which corresponds to the best NRMSE performance. In addition, the two reference criteria, the CC and MAPE, are also acceptable. Hence, we choose $\delta = 3$ for the RVR forecasting, the results of which are shown in Fig. 5.

Fig. 5(a) and (b) display the modeling results and the scatter plots using $\delta = 3$ for the training data (the 8th day–290th day data), respectively. The forecasting results of the RVR model successfully follow the trend of the training data. However, when $\delta = 3$ is used to forecast the testing data (the 298th day–360th day data), greater errors occur in both the forecasting results and the scatter plots, as shown in Fig. 5(c) and (d), respectively. From Fig. 5, it is apparent that the forecasting performance of the testing data is unsatisfactory when solely employing the RVR model, illustrating that it is difficult to obtain good forecasting results with a single model.

One of the main difficulties in studying water demand series or any other experimental series is that all the information at our disposal consists of discrete sequences of numbers that have no reference to the potential law of the time series. Many of the important features of the series cannot be directly extracted by simple linear or nonlinear mapping, especially in nonlinear systems. Even if the equations are very complex they only produce complex behavior in terms of the particular values of the parameters in the equations. Moreover, some characteristics of water demand do not only exist in one scale, but also in other scales (Steinhauser, 2008). Hence, the forecasting performance might be poor if the features of the other scales are not considered. In addition, studies (Soltani, 2002; Mabrouk et al., 2008) have proved that multi-scale regression

Table 1
Influence of δ on the RVR regression performances.

δ	Modeling data			Testing data		
	NRMSE (m^3)	MARE (%)	CC (W/W)	NRMSE (m^3)	MARE (%)	CC (W/W)
1	0.0374	2.7788	0.9401	0.0379	2.8915	0.5348
2	0.0382	2.9264	0.9375	0.0381	4.4947	0.3393
3	0.0382	2.9257	0.9376	0.0348	2.8517	0.5269
4	0.0380	2.9195	0.9382	0.0348	4.1035	0.5192
5	0.0377	2.8873	0.9390	0.0347	18.2495	0.5259
6	0.0380	2.9125	0.9382	0.0352	2.9813	0.4987

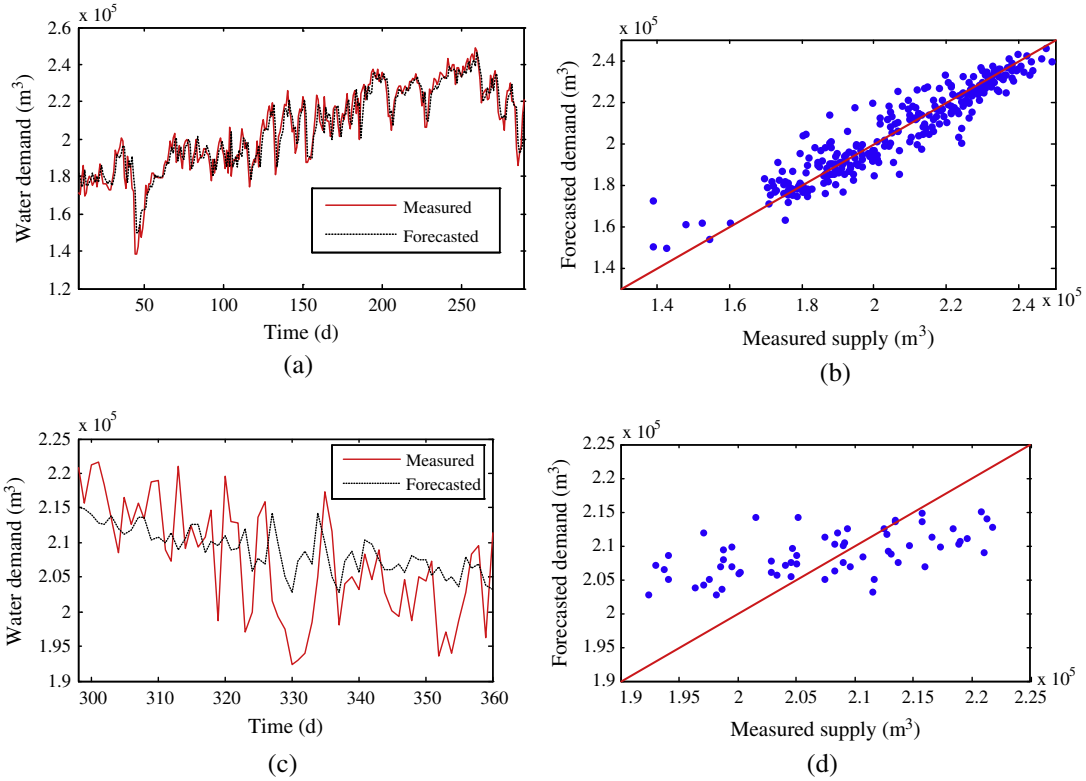

Fig. 5. Forecasting results and scatter plots based on RVR (δ = 3): (a) and (b) are the forecasting results and the scatter for the training data; and (c) and (d) are the forecasting results and the scatter for the testing data, respectively.

models have better precision than mono-scale models. In the following subsection, therefore, we introduce the SWT in collaboration with the RVR to build a multi-scale regression model for daily urban water demand forecasting.

3.2. Multi-scale regression performance

Complicated changes in nonlinear and non-stationary series are hard to be forecasted by mono-scale models (Mitani et al., 2003). Multi-scale analysis is an effective alternative for establishing a local prediction models at different resolutions. As mentioned, multi-scale regression provides the possibility of revealing all of the characteristics of a time series. In this subsection, we use the SWT to decompose the daily water series into different resolutions (scales) and use the RVR model to forecast the water demand for each subsequence. The multi-scale RVR outputs are used to generate the final forecasting result through the wavelet reconstruction.

Compared to the original wavelet transform, SWT exhibits translation invariance. In other words, the length of the wavelet coefficient at each SWT scale is equal to that of the original data. For daily water demand series **W** with a length N, the first step of the SWT is to divide **W** into two coefficient sets: the approximation coefficients $\mathbf{WA}_1$, obtained by low-pass filter *Lo_D* convolution, and the detail coefficients $\mathbf{W}_1$, obtained by high-pass filter *Hi_D* convolution. Note that the lengths of $\mathbf{WA}_1$ and $\mathbf{W}_1$ are all N, rather than $N/2$ of the original discrete wavelet transform. The next step is to divide $\mathbf{WA}_1$ into two parts, $\mathbf{WA}_2$ and $\mathbf{W}_2$, in the same way. The rest can be divided using the same methods. Unlike the original discrete wavelet transform, the approximation coefficients and the detail coefficients transformed by the SWT do not require down-sampling. Hence, the length is still the same as that of the original time series without loss of information (Li and Liang, 2011).

A key issue is to determine the decomposition depth n for the multi-scale regression of the time series using the SWT. After the SWT decomposition, the central frequency f_i at the ith scale is given by (Fan et al., 2007)

$$f_i = \frac{f_s f^* f_a^*}{2^{i+1}}, \tag{10}$$

where f_s is the sampling frequency, and f^* and f_a^* represent the center frequency of the mother wavelet and the scaling function, respectively. In this study, the db4 wavelet is chosen as the mother wavelet, for which f^* = 0.714 and $f_a^* = 0.4$ (Fan et al., 2007). As the

input variable of the RVR model is $\mathbf{X} = [W(i-1), W(i-2), \ldots, W(i-7)]$, we choose 7d as the target decomposition period, which has a frequency of 7/N (1/d) in an N-day time series. Hence, the decomposition depth can be calculated as

$$n = \log_2\left(\frac{N}{7}\cdot f_s f_a^*\right) - 1, \tag{11}$$

For the time series $\mathbf{W}$ shown in Fig. 1, with a length $N = 360$, the decomposition depth is calculated as $n = 3$ using the foregoing equation.

As illustrated, with the SWT decomposition, the time series $\mathbf{W}$ is decomposed into the n-level detail coefficients $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_n$, and the 1-level approximation coefficients $\mathbf{WA}_n$. For convenience of expression, $\mathbf{WA}_n$ is named $\mathbf{W}_{n+1}$ in this context. For the j-level coefficients, Eq. (8) is used to build the jth RVR model illustrated as

$$\widehat{\mathbf{F}}_j = y_j(\mathbf{X}_j; \boldsymbol{\omega}_j) + \boldsymbol{\varepsilon}_j, \tag{12}$$

where $\widehat{\mathbf{F}}_j = \widehat{\mathbf{W}}_j$. It should be noted that the output $\widehat{\mathbf{F}}_j$ of the jth RVR model is $\widehat{\mathbf{W}}_j$, not $\widehat{\mathbf{W}}$. As $\widehat{\mathbf{W}}_j$ are the j-level wavelet coefficients of $\widehat{\mathbf{W}}$, the inverse stationary wavelet transform (ISWT) can be used to reconstruct $\widehat{\mathbf{W}}$ using all levels of $\widehat{\mathbf{W}}_j$, that is,

$$\widehat{\mathbf{W}} = \mathrm{ISWT}(\widehat{\mathbf{W}}_1, \widehat{\mathbf{W}}_2, \ldots, \widehat{\mathbf{W}}_j, \ldots, \widehat{\mathbf{W}}_{n+1}). \tag{13}$$

We have built the multi-scale RVR model for daily urban water demand forecasting. However, multi-scale analysis introduces a new problem for real applications. For the mono-scale analysis illustrated in Table 1, there is only one parameter δ to be optimized for the RVR model. Accordingly, the enumeration method can be used to find the optimal δ. Unfortunately, there are $n+1$ parameters, $\delta_1, \delta_2, \ldots, \delta_{n+1}$, to be optimized for the multi-scale RVR approach. It is worth noting that the $n+1$ parameters are coupled to each other. In other words, the $n+1$ parameters cannot be optimized separately. Hence, we anticipate conducting a global optimization of the $n+1$ parameters. Assuming that the calculation complexity of the δ optimization is a, using the enumeration method, the optimization complexity of the coupled $n+1$ parameters $(\delta_1, \delta_2, \ldots, \delta_{n+1})$ is a^{n+1}, which is obviously unaffordable compared to the mono-scale RVR model. To solve this problem, in the following subsection we use the ACPSO to optimize the coupled $n+1$ parameters of the multi-scale RVR model.

3.3. Parameter optimization using the ACPSO

Particle swarm optimization (PSO), which has the advantages of easy implementation, a fast convergence speed, fewer parameter settings, and being a concise concept, is one of the most successful optimization techniques. Proposed by Kennedy and Eberhart (1995), PSO is based on the natural flocking and swarming behavior of birds and insects. With PSO, potential solutions, known as particles, move through the solution space by following the current optimum particles (Kuo et al., 2010). The PSO algorithm works by attracting particles to search for space positions with the best fitness. Each particle retains a memory of its previous best position, and the best value of the swarm is called the global best. The PSO algorithm operates by iteration. The solution produced in each iteration is compared with the self-local best and the global best of the swarm.

Although the PSO algorithm has proven to be effective in solving the optimization problem, it is similar to other intelligent algorithms in being vulnerable to premature convergence and local optimization. The ACPSO approach, which utilizes chaos theory and adaptive adjustment, was introduced to avoid these shortages. We apply the ACPSO to simultaneously optimize the coupled $n+1$ parameters $(\delta_1, \delta_2, \ldots, \delta_{n+1})$.

Let K denote the number of the particles. The kth $(k \in [1, K])$ particle at the $(u+1)$th $(u \in [1, U])$ iteration can be updated using the following equations (Kiran et al., 2012)

$$V_k^{u+1} = wV_k^u + c_1 r_1 (P_k^u - X_k^u) + c_2 r_2 (M^u - X_k^u), \tag{14}$$

$$X_k^{u+1} = X_k^u + V_k^{u+1}, \tag{15}$$

where c_1 and c_2 are learning factors that determine the relative influence of the social and cognitive components, r_1 and r_2 are random numbers uniformly distributed in the interval [0, 1], w is the inertia weight for weighing the local optimum and the global optimum (Shi and Eberhart, 1998), and X is the position, V the velocity, P the best local position, and M the best global position of all the particles.

To balance the conflict between the global and local searches, we employ the ACPSO algorithm with the adaptive learning and weight factor (ALWF) for the global exploration together with the chaotic local search (CLS) (Cai et al., 2012). The ALWF is determined as

$$c_1 = c_2 = c_{\max} - (c_{\max} - c_{\min}) \times u/U, \tag{16}$$

$$w = w_{\text{mean\_min}} + (w_{\text{mean\_max}} - w_{\text{mean\_min}}) \times \text{rand} + w_{\text{sigma}} \times \text{randn}, \tag{17}$$

where $c_{\min}$ and $c_{\max}$ are the minimum and the maximum of the learning factors, respectively, U is the total iterations, $w_{\text{mean\_max}}$ and $w_{\text{mean\_min}}$ are the minimum and the maximum of the random weights, w_{sigma} is the random weights variance, rand is a random number in [0, 1] following the uniform distribution, and randn is a random number following the normal distribution with a mean zero variance.

The logistic equation, which is a well-known chaos system, is introduced into the CLS process as

$$X_{kd}^{u+1} = \mu \times X_{kd}^u \times (1 - X_{kd}^u), \text{and } V_{kd}^{u+1} = \mu \times V_{kd}^u \times (1 - V_{kd}^u), \tag{18}$$

where d denotes the dimension of the search space and μ is the control parameter. According to the research results presented by Steeb (2011), for $\mu = 4$, $X_{kd}^u \notin \{0, 0.25, 0.5, 0.75, 1\}$, and $V_{kd}^u \notin \{0, 0.25, 0.5, 0.75, 1\}$, the sequence produced by the one-dimension nonlinear mapping is a chaotic series.

For the iterative optimization of the ACPSO, the NRMSE as shown in Eq. (4) is chosen as the fitness. With the aforementioned ACPSO method, the coupled $n+1$ parameters $(\delta_1, \delta_2, \ldots, \delta_{n+1})$ of the MSRVR can be optimized simultaneously. The optimized MSRVR is then employed to forecast the daily urban water demand shown in Fig. 1. Again, the first 290d data are used for the ACRVR modeling and the rest are used for the testing. The parameter settings that we use for the ACPSO are shown in Table 2. The performance of the optimized MSRVR model with different particle numbers is shown in Table 3.

As shown in Table 3, the performance values with particle numbers 10, 20, 30, and 40 are similar, regardless of which criterion (NRMSE, MAPE, or CC) is used. In other words, when optimized using the ACPSO, the MSRVR model is immune from the variability of the particle numbers.

Taking particle number 30 as an example, Fig. 6 plots the forecasting results using the proposed MSRVR approach. Compared to the RVR forecasting method shown in Table 1, the MSRVR

Table 2
Parameter settings for the ACPSO.

Parameters	K	U	$c_{\min}$	$c_{\max}$	w
Settings	10, 20, 30, 40	50	1	2.5	rand

approach exhibits better performance in terms of the NRMSE, CC, and MAPE criteria. Comparisons of Figs. 5 and 6 show that the MSRVR model achieves more accurate estimation than the RVR models.

3.4. Overview of the MSRVR approach

Having separately examined the constituent steps, the MSRVR modeling procedure, which is illustrated in Fig. 7, can now be summarized as follows:

- *Step 1.* Collect the modeling (training) subset of the daily urban water supply series **W**.
- *Step 2.* Initialize parameters K, U, $c_{\min}$, $c_{\max}$, and w for the ACPSO.
- *Step 3.* Calculate t_{opt}, m_{opt}, and n using Eqs. (2), (3), and (11), respectively.
- *Step 4.* Perform the SWT to decompose the time series **W** into $n+1$ levels.
- *Step 5.* Perform the ACPSO to obtain the optimized RVR parameters $(\delta_1, \delta_2, \ldots, \delta_{n+1})$.
- *Step 6.* In the jth level ($j \in [1, n+1]$), train the RVR model using the ACPSO-optimized parameter δ_j. The trained RVR model is used to estimate the model output $\widehat{\mathbf{W}}_j$.
- *Step 7.* Perform the inverse SWT to generate the estimated daily water demand $\widehat{\mathbf{W}}$ using Eq. (13).
- *Step 8.* Output the forecasting result. End.

The trained MSRVR model can be used to forecast the daily urban water demand using the historical water supply series. Similar to the modeling procedure, the daily water supply series is first decomposed into $n+1$ levels using the SWT. In the jth level, the wavelet coefficients are used by the trained RVR model (RVR_j) to generate the estimated $\widehat{\mathbf{W}}_j$. The ISWT is finally used to output the forecast next-day water demand $\widehat{\mathbf{W}}$. In this way, the proposed MSSVR approach can be used to forecast daily urban water demand by combining the nonlinear mapping characteristics of the RVR model and the multi-scale analysis capability of the SWT.

The wavelet transform, which has good time–frequency characteristics and multi-resolution capability, can extract the characteristics of the water demand series by decomposing the time series into different scales. In this way, the wavelet transform can reduce the complexity of the subsequent forecasting with the RVR model, and is capable of achieving better forecasting performance. In the next section, to evaluate the forecasting capability of the MSRVR, we forecast water demand use from data collected from another real waterworks.

Table 3
Forecasting performances of the MSRVR optimized using the ACPSO with different particle numbers.

Particle number	Training data			Testing data		
	NRMSE (m^3)	MARE (%)	CC (W/W)	NRMSE (m^3)	MARE (%)	CC (W/W)
10	0.0060	0.1854	0.9985	0.0207	1.5293	0.8542
20	0.0061	0.4845	0.9985	0.0166	1.3204	0.9086
30	0.0064	0.4954	0.9983	0.0010	0.7511	0.9687
40	0.0063	0.3671	0.9983	0.0192	1.4584	0.8764

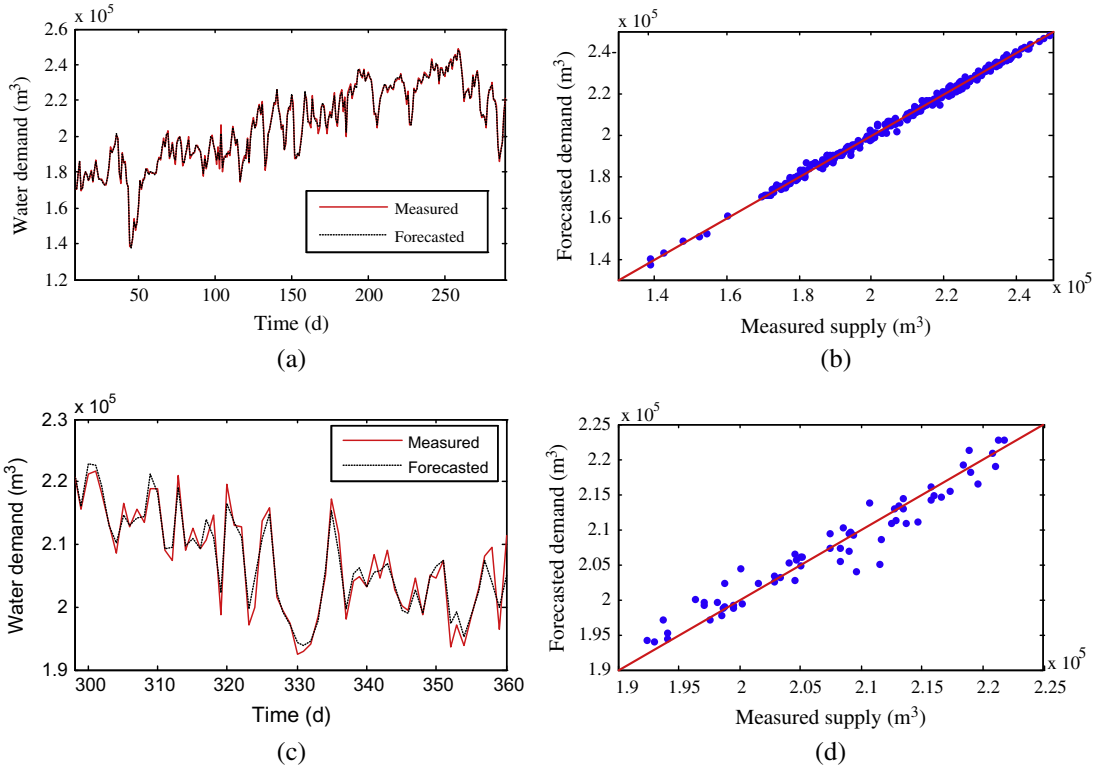

Fig. 6. Forecasting results and scatter plots using the present MSRVR approach optimized by the ACPSO (particle number 30): (a) and (b) are the forecasting results and the scatter of the training data; and (c) and (d) are the forecasting results and the scatter of testing data, respectively.

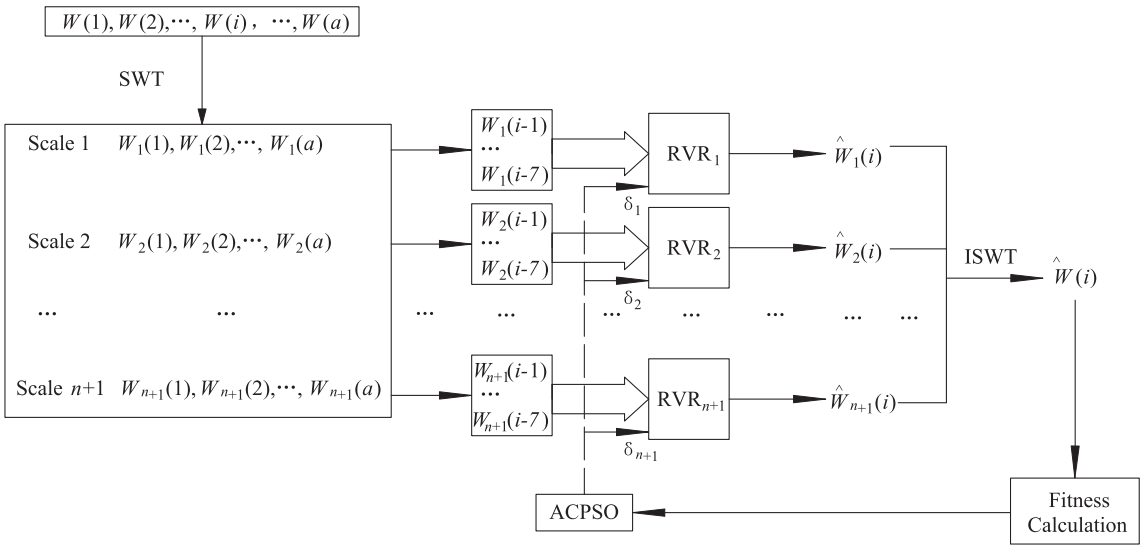

Fig. 7. MSRVR model training procedure.

4. Evaluation of the proposed MSRVR approach

In this section, we evaluate the proposed MSRVR model using real water demand series from another waterworks. The performance of the model is then compared with that of two existing methods to show its superiority.

4.1. Real application evaluation

In this section, the MSRVR approach is evaluated using real data on the daily water supply from another urban waterworks (#2 as shown in Fig. 8), which is around 100 km west of the #1 waterworks. Fig. 8 shows the location of #2 waterworks, which has a daily water supply capacity of up to 10 million cubic meters. Fig. 9 plots the 180-day water supply of #2 waterworks from January to June 2012. The data from the second waterworks are used to validate the generalization of the proposed MSRVR approach.

The first 140-day series shown in Fig. 9 is employed as the training dataset, and the remaining 40-day series is used for the testing. We obtained $n = 2$ using Eq. (11). The remaining parameters are the same as illustrated in Subsection 3.4 of Section 3. The forecasting results for the training dataset (the 8th day–140th day data) and the testing dataset (the 148th day–180th day data) are shown in Fig. 10.

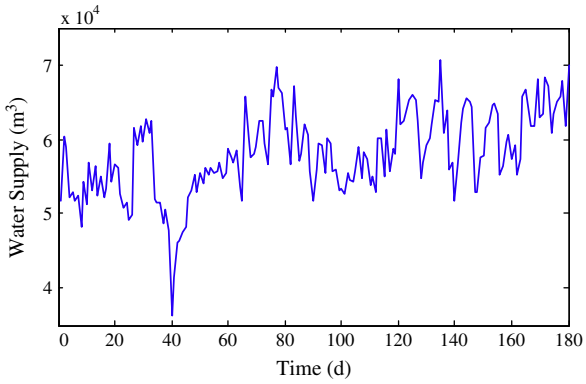

Fig. 9. Daily water supply series of the #2 waterworks from January to June, 2012.

To quantitatively evaluate the forecasting performance of the MSRVR approach, we calculate the NRMSE, CC, and MAPE values

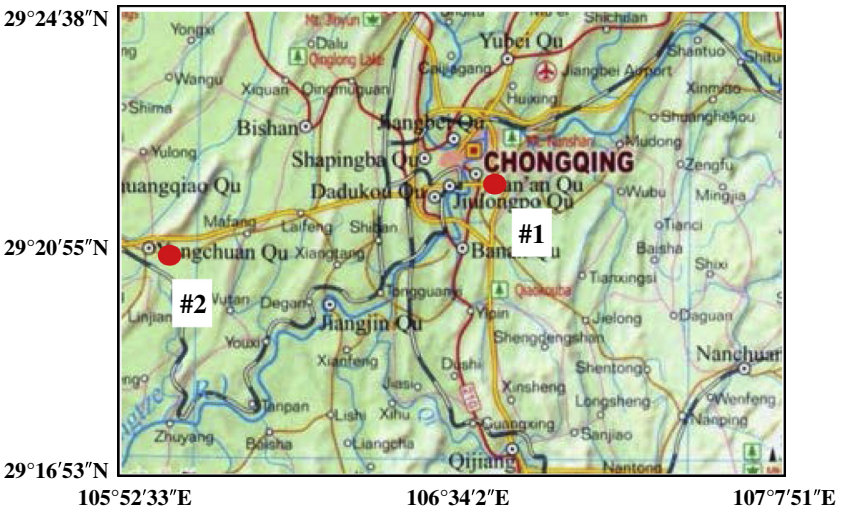

Fig. 8. Locations of the two urban waterworks.

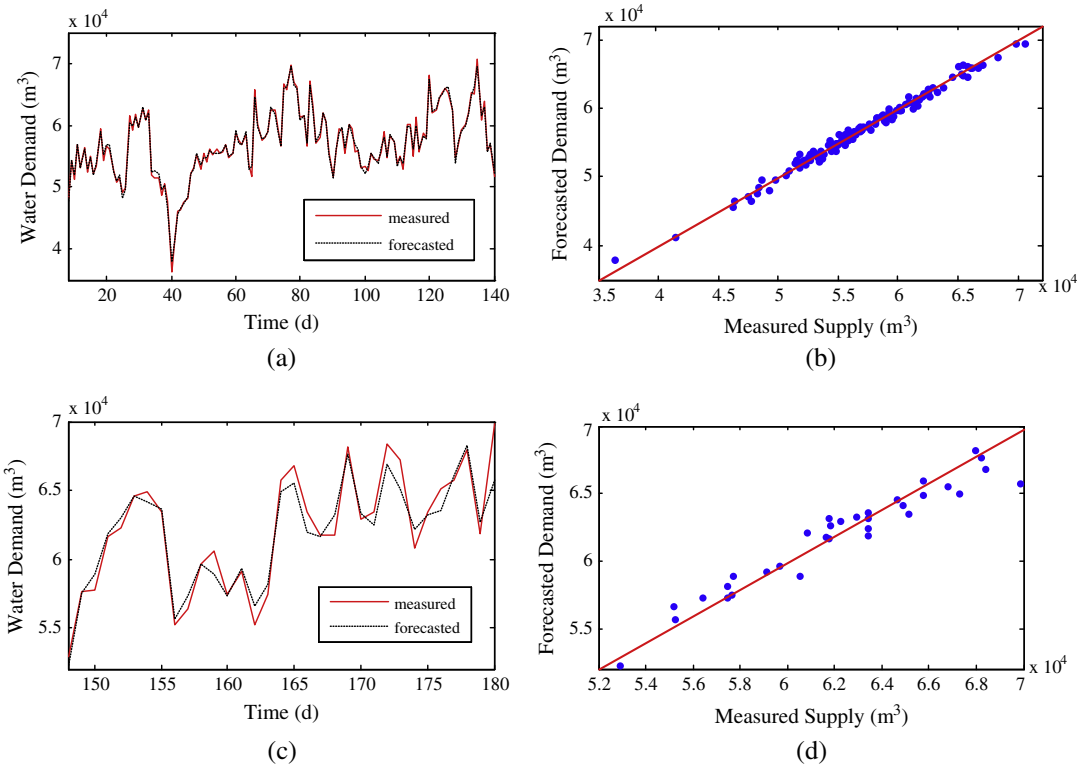

Fig. 10. Forecasting results and scatter plots for #2 urban waterworks using the present MSRVR approach: (a) and (b) are the forecasting results and the scatter of the training dataset; and (c) and (d) are the forecasting results and the scatter of the testing dataset, respectively.

Table 4
Forecasting performances for #2 waterworks using the present MSRVR approach.

Method	Training data			Testing data		
	NRMSE (m^3)	MARE (%)	CC (W/W)	NRMSE (m^3)	MARE (%)	CC (W/W)
MSRVR	0.0105	0.8455	0.9944	0.0192	1.3854	0.9653

for the #2 waterworks as shown in Table 4. Returning to Table 3, it is apparent that the forecasting capability of the proposed MSRVR approach is equally effective for the two waterworks. In other words, the proposed model can be used to forecast the daily water demand for different scales of urban waterworks. This confirms that the MSRVR approach is generalizable to the forecasting of daily urban water demand.

4.2. Comparison with existing methods

In addition to evaluating the generalizability of the proposed approach, we compare the proposed MSRVR approach with two recently reported methods (Firat et al., 2010), namely, generalized regression neural networks (GRNN) and feed forward neural networks (FFNN), using the testing dataset shown in Fig. 9. Table 5 displays the forecasting performance of the two existing methods evaluated using the three criteria. Compared to the two existing methods, the proposed MSRVR has the best forecasting performance in all three criteria.

Table 5
Comparison of the forecasting performances.

Method	Testing data		
	NRMSE (m^3)	MARE (%)	CC (W/W)
GRNN	0.0632	6.0645	0.4086
FFNN	0.0801	6.3127	0.3622
MSRVR (best)	0.0192	1.3854	0.9653

The forecasting results (Figs. 6 and 10) for real water demand data from the #1 and #2 waterworks demonstrate the superiority and generalizability of the MSRVR model. In addition, the comparison results (Table 5) indicate the forecasting superiority of the MSRVR model. The results sufficiently illustrate that the MSRVR, which combines multi-scale analysis with the RVR, demonstrates better forecasting performance than the existing models.

Based on our characterization of the water demand series, this paper can be regarded as contributing to the modeling of the potential law of a nonlinear chaos system. Our results reveal the existence of a hierarchy of multiple time scales and the superposition of different details (Weinan and Engquist, 2003). The proposed MSRVR approach forecasts the urban daily water demand by transforming the time–frequency into different scales. Hence, the MSRVR model is capable of generating better forecasting performance for daily water demand.

5. Conclusions

We propose a multi-scale relevance vector regression approach for forecasting daily urban water demand. The daily water supply series is decomposed by the SWT for multi-scale analysis of the water demand. In each scale, wavelet coefficients are employed

to build the RVR model, the input variables of which are determined by the chaos time series analysis. The ACPSO algorithm is suggested to globally optimize the parameters of all the RVR models. The final forecasting result is generated by inverse SWT reconstruction of the output coefficients of the RVR models. Three criteria, NRMSE, MAPE, and CC, are introduced to measure the forecasting performance. The proposed method is applied to the data from two real waterworks in Chongqing, China. The forecasting performance of the model is also compared with two existing approaches. The results show that the proposed approach outperforms the two existing models in terms of all of the NRMSE, CC, and MAPE criteria. As it employs both multi-scale characteristics and nonlinear mapping capability, the proposed MSRVR approach is capable of following the chaos pattern of daily urban water demand reasonably well.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China (51375517), the Natural Science Foundation of CQ CSTC (2012JJJQ70001), the Project of Chongqing Innovation Team in University (KJTD201313), and the Ministry of Housing and Urban-Rural Development of China (2001-45).

References

Adamowski, J.F., 2008. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. J. Hydrol. 353 (3–4), 247–266.

Adamowski, J., Karapataki, C., 2010. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. J. Hydrol. Eng. 15 (10), 729–743.

Alhumoud, J.M., 2008. Freshwater consumption in Kuwait: analysis and forecasting. J. Water Supply Res. Technol.-AQUA 57 (4), 279–288.

Babel, M.S., Shinde, V.R., 2011. Identifying prominent explanatory variables for water demand prediction using artificial neural networks: a case study of Bangkok. Water Resour. Manage. 25, 1653–1676.

Beal, C.D., Stewart, R.A., Fielding, K., 2011. A novel mixed method smart metering approach to reconciling differences between perceived and actual residential end use water consumption. J. Cleaner Prod.. http://dx.doi.org/10.1016/j.jclepro.09.007.

Billings, B., Jones, C., 2008. Forecasting Urban Water Demand, second ed. American Waterworks Association.

Cai, J., Li, Q., Li, L., Peng, H., Yang, Y., 2012. A hybrid CPSO-SQP method for economic dispatch considering the valve-point effects. Energy Convers. Manage. 53 (1), 175–181.

Caiado, J., 2010. Performance of combined double seasonal univariate time series models for forecasting water demand. J. Hydrol. Eng. 15 (3), 215–222.

Campisi-Pinto, S., Adamowski, J., Oron, G., 2012. Forecasting urban water demand via wavelet-denoising and neural network models, case study: city of Syracuse, Italy. Water Resour. Manage. 26 (12), 3539–3558.

Cutore, P., Campisano, A., Kapelan, Z., Modica, C., Savic, D., 2008. Probabilistic prediction of urban water consumption using the SCEM-UA algorithm. Urban Water J. 5 (2), 125–132.

Donkor, E.A., Mazzuchi, T.A., Soyer, R., Roberson, J.A., 2012. Urban water demand forecasting : a review of methods and models. J. Water Resour. Plan. Manage.. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000314.

Fan, J., Liu, M., Hai, Y., Wang, L., 2007. GUI for computing center-frequency of the scale and wavelet functions and its applications. Sci. Technol. Rev. 25 (24), 36–39.

Firat, M., Turan, M.E., Yurdusev, M.A., 2010. Comparative analysis of neural network techniques for predicting water consumption time series. J. Hydrol. 384 (1–2), 46–51.

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

Holzfuss, J., Mayer-Kress, G., 1986. An approach to error-estimation in the application of dimension algorithms. Dimensions Entropies Chaot. Syst. 32, 114–122.

Kennedy, J., Eberhart, R.C., 1995. A new optimizer using particles swarm theory. In: Proceedings of Sixth International Symposium on Micro Machine and Human Science, pp. 39–43.

Kiran, M.S., Özceylan, E., Gündüz, M., Paksoy, T., 2012. A novel hybrid approach based on particle swarm optimization and ant colony algorithm to forecast energy demand of Turkey. Energy Convers. Manage. 53 (1), 75–83.

Kugiumtzis, D., 1996. State space reconstruction parameters in the analysis of chaotic time series the role of the time window length. Physics D 95 (1), 13–28.

Kuo, R.J., Hong, S.Y., Huang, Y.C., 2010. Integration of particle swarm optimization-based fuzzy neural network and artificial neural network for supplier selection. Appl. Math. Model. 34 (12), 3976–3990.

Lee, S., Wentz, E., Gober, P., 2010. Space-time forecasting using soft geostatistics: a case study in forecasting municipal water demand for Phoenix, Arizona. Stoch. Environ. Res. Risk Assess. 24 (2), 283–295.

Li, Z., Chen, G., 2010. Integration of fuzzy logic and chaos theory, first ed. Springer Publishing Company.

Li, C., Liang, M., 2011. Separation of vibration-induced signal of oil debris sensor for vibration monitoring. Smart Mater. Struct. 20 (4), 045016.

Li, C., Wang, S., 2006. Next-day power market clearing prince forecasting using artificial fish-swarm based neural network. Lect. Notes Comput. Sci. 3972, 1290–1295.

Li, C., Liang, M., Wang, Y., Dong, Y., 2012a. Vibration suppression using two-terminal flywheel. Part I: Modeling and characterization. J. Vib. Control 18 (8), 1096–1105.

Li, C., Wang, J., Wang, H., Zhang, S., 2012b. Urban water consumption long-term prediction model based on the water price elasticity in China. World Environmental and Water Resources Congress 2012: Crossing Boundaries, pp. 2491–2500.

Mabrouk, A.B., Abdallah, N.B., Dhifaoui, Z., 2008. Wavelet decomposition and autoregressive model for time series prediction. Appl. Math. Comput. 199 (1), 334–340.

Maheswaran, R., Khosa, R., 2012. Multiscale nonlinear model for monthly stream flow forecasting: a wavelet based approach. J. Hydroinform. 14 (2), 424–442.

Mitani, Y., Tsutsumoto, K., Kagawa, N., 2003. Time series prediction of acoustic signals using neural network model and wavelet shrinkage. In: Proceedings of the tenth International Congress on Sound and Vibration, pp. 4189–4196.

Mohamed, M., Al-Mualla, A., 2010. Water demand forecasting in Umm Al-Quwain (UAE) using the IWR-MAIN specify forecasting model. Water Resour. Manage. 24 (14), 4093–4120.

Nasseri, M., Moeini, A., Tabesh, M., 2011. Forecasting monthly urban water demand using extended Kalman filter and genetic programming. Expert Syst. Appl. 38, 7387–7395.

Odan, F.K., Reis, L.F.R., 2012. Hybrid water demand forecasting model associating artificial neural network with Fourier series. J. Water Resour. Plan. Manage. 138, 245–256.

Oshima, N., Kosuda, T., 1998. Distribution reservoir control with demand prediction using deterministic-chaos method. Water Sci. Technol. 37 (12), 389–395.

Shi, Y., Eberhart, R., 1998. A modified particle swarm optimizer. Evolut. Comput. Proc., 69–73.

Soltani, S., 2002. On the use of the wavelet decomposition for time series prediction. Neurocomputing 48, 267–277.

Steeb, W.H., 2011. The Nonlinear Workbook, fifth ed. World Scientific.

Steinhauser, M.O., 2008. Computational Multiscale Modeling of Fluids and Solids: Theory and Applications, fitst ed. Springer Publishing Company.

Tabesh, M., Dini, M., 2009. Fuzzy and neuro-fuzzy models for short-term water demand forecasting in Tehran. Iran. J. Sci. Technol., Trans. B 33, 61–77.

Ticlavilca, A.M., Mckee, M., Walker, W.R., 2013. Real-time forecasting of short-term irrigation canal demands using a robust multivariate Bayesian learning model. Irrig. Sci. 31 (2), 151–167.

Tipping, M.E., 2000. The relevance vector machine. Adv. Neur. Inform. Process. Syst. 12, 652–658.

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. 1, 211–244.

Wei, S., Lei, A., Islam, S., 2010. Modeling and simulation of industrial water demand of Beijing municipality in China. Front. Environ. Sci. Eng. China 4 (1), 91–101.

Weinan, E., Engquist, B., 2003. Multiscale modeling and computation. Notices Am. Math. Soc. 50 (9), 1062–1070.

Wong, J.S., Zhang, Q., Chen, Y.D., 2010. Statistical modeling of daily urban consumption in Hong Kong: trend, changing patterns, and forecast. Water Resour. Res. 46 (3).

Zhao, P., Zhang, H., 2008. Chaotic characters and forecasting of urban water consumption. China Water Wastewater 24 (5), pp. 90–93, 97.

Ziervogel, G., Johnston, P., Matthew, M., Mukheibir, P., 2010. Using climate information for supporting climate change adaptation in water resource management in South Africa. Climat. Change 103, 537–554.