A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China

Shuai Wang\textsuperscript{a,d}, Lean Yu\textsuperscript{b,c,*}, Ling Tang\textsuperscript{a,d}, Shouyang Wang\textsuperscript{b,c}

\textsuperscript{a}Institute of Policy and Management, Chinese Academy of Sciences, Beijing 100190, China
\textsuperscript{b}MADIS, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
\textsuperscript{c}Center for Forecasting Science, Chinese Academy of Sciences, 55 Zhongguancun East Road, Haidian District, Beijing 100190, China
\textsuperscript{d}Graduate University of Chinese Academy of Sciences, Beijing 100080, China

1. Introduction

Hydropower, as a clean and renewable energy, has been playing an increasingly important role in the world energy system due to its technical, economical and environmental benefits. Most countries have given the highest priority to its development, especially in China. It is well known that China has the richest hydraulic resource on the planet with a total hydropower theoretical potential of 694 million kilo-Watt (kW). However, the development rate of the technical exploitable amount (which has about 542 million kW) is only 27.3\%, far below the world average level \cite{1}. It is because of the low development level that the hydropower prospect in China would be very promising. Therefore, considering the inherent benefits of hydropower and the superiority of resources endowment in China, developing hydropower is of great importance to alleviate the current energy crisis and environmental pollution resulting from the rapid economic growth of China.

Nowadays, China's hydropower is taking a rapid development process and the hydropower installation capacity has been 210 million kW until 2010 \cite{2}. In last decade, the Chinese government has proposed the objectives of hydroelectric development in the “Tenth Five-Year Plan” (2001–2005) and the “Eleventh Five-Year Plan” (2006–2010). According to the “Tenth Five-Year Plan”, the hydropower installation capacity should have reached 95 million kW, 125 million kW and 150 million kW by 2005, 2010 and 2015, respectively. In terms of the “Eleventh Five-Year Plan” the hydropower installation capacity should have arrived at 190 million kW and 300 million kW by 2010 and 2020, respectively \cite{1}. In terms of the two five-year plans, it can be found that the objectives of hydropower installation capacity are often adjusted dynamically based on the practical development status. Moreover the objective of hydropower development in 2010 in the “Eleventh Five-Year Plan” has been reached. It is time to draw up the following new Five-Year plan to develop the hydropower in China. To make a proper plan of hydropower development and make sure to accomplish it successfully, the policy makers should always know the consumption demand in the hydropower market and make sure that the objective can match the actual situation. If the development of hydropower supply is faster than that of consumption demand, the overcapacity of hydropower will be appeared. Conversely, if the hydropower supply is less than hydropower...
consumption demand, it will lead to a shortage of hydropower supply. Therefore, an accurate prediction of hydropower consumption is important for policy makers to draw up hydropower development programs and modify the policy to ensure the expected objectives. Furthermore, a good hydropower consumption prediction is useful for the hydropower industry to optimize hydropower operation scheduling. For these purposes, this study tries to propose a novel ensemble learning approach to predicting the future hydropower consumption in China.

However, it is not easy to predict the hydropower consumption accurately due to the intrinsic complexity of involving driven factors. As hydropower is one kind of electric energy, the prediction of hydropower consumption is similar to other kinds of energy consumption forecasting. In the past decades, there have been already abundant studies on electricity consumption forecasting. For example, traditional linear statistical and econometric techniques, such as multiple regression [3–7], fuzzy regression [8], ARIMA [9,10], error correction models (ECM) [9], co-integration analysis [9,11] and state space model [12] have been widely applied to electricity consumption forecasting. Particularly, these models take account of economic, demographic, climatic variables. In addition, other approaches used in the electricity demand prediction include decision tree [6], probabilistic model [13] and grey prediction method [14–16]. However, most of the statistical-based models are built on linear assumptions and they cannot capture the nonlinear patterns hidden in the original data. Due to the limitations of the traditional statistical and econometric models, some nonlinear and artificial intelligence (AI) models, such as nonlinear regression model, artificial neural networks (ANN) (e.g. [17–20]), support vector machines (SVM) (e.g. [21,22]) and genetic algorithm (GA) (e.g. [23]) provided powerful nonlinear solutions to electricity consumption prediction. Nevertheless, these AI models have their own disadvantages. For example, ANN often suffers from local minima and over-fitting, while other AI models, such as SVM and GA as well as ANN, are sensitive to parameter selection. To remedy the above shortcomings, some hybrid methods have been used to predict electricity consumption. For example, Pao [24] used hybrid nonlinear models integrating generalized autoregressive conditional heteroscedasticity (GARCH) family models and ANN to forecast energy consumption in Taiwan. Nguyen and Nabney [25] combined wavelet transform (WT) with multi-layer perceptron (MLP), radial basis functions, linear regression and GARCH to forecast daily electricity demand. These papers all indicated that hybrid models outperform single methods. Particularly, Wang et al. [26] developed a hybrid AI approach integrating SVR and seasonal adjustment to predict electricity demand, but this hybrid SVR model simply assumes the same seasonal factor in different years that does not fit the reality. Particularly, it is noteworthy that multivariate modeling is an infrequent used method in energy consumption forecasting. The reasons may be that the influential external factors about energy consumption are not easy to determine, and their accurate data may be unavailable or difficult to obtain. Abdel-Aal and Al-Garni [27] proved that univariate time series models are more accurate than multivariate additive network models in the monthly energy consumption forecasting in Saudi Arabia. Moreover, using only the consumption time series can reduce the data dimensionality which may improve generalization and forecasting performance. Based on the reasons above mentioned, this study will focus on univariate time series models.

Although there are a lot of literatures about electricity consumption forecasting, very few researchers studied the topic of hydropower consumption forecasting to our knowledge. Current hydropower related forecasting studies are usually about the river flow discharges forecasting in hydropower station [28,29] and the installation capacity forecasting of small hydropower [30]. In addition, only few researches on energy demand forecasting involve hydropower. For instance, Ediger and Akar [10] employed ARIMA model with the annual datasets to predict primary energy demand including primary hydropower demand. However, ARIMA model cannot capture the nonlinear patterns hidden in the hydropower demand series. To sum up, the possible reason of few hydropower consumption forecasting researches may be caused by the difficulty not only in data collection but also in hydropower consumption modeling.

Particularly, the difficulty in modeling hydropower consumption mainly results from the intrinsic complexity of hydropower consumption data. Generally speaking, various factors of affecting hydropower consumption can be divided into two-fold. On the one hand, similar to other kinds of energy consumption, hydropower consumption is influenced by numerous random factors such as economic development, social reform, climate change, gross energy consumption and national policy in energy field. These factors are often of irregularity and nonstationarity. On the other hand, the hydrological characteristic of water resource formulates the other important nature of hydropower consumption. It is already proved that the water resource system, i.e. the hydrological system, exhibits two distinct characteristics: seasonality and nonlinearity [31]. There is no exception in China. Most rivers have uneven runoff distribution within one year or between years and the river flows have great differences between rainy and dry seasons [1]. Therefore, considering the energy related random factors and hydrological factors, the hydropower consumption series shows obvious seasonality and nonlinearity. Thus, the hydropower consumption series is intrinsically a nonlinear time series, which is affected by obvious seasonal factors. It reveals that forecasting hydropower consumption is a rather challenging task.

Due to the seasonal nature and nonlinear characteristic of hydropower series and motivated by hybrid methodology, we try to employ the “decomposition and ensemble” principle (or “divide and conquer” strategy), introduced in [32,33], to overcome the difficulty in forecasting hydropower consumption. The main aim of decomposition is to simplify the difficult forecasting task into some relatively easy forecasting subtasks, while the goal of ensemble is to formulate a consensus forecasting result in terms of decomposed subtasks. Accordingly, some hybrid ensemble methods motivated by this principle have been proposed recently to solve some tough prediction tasks such as international crude oil price forecasting. The corresponding empirical results demonstrated that the hybrid ensemble methodologies outperform individual forecasting models in many cases [32,33]. Considering the seasonal pattern of hydropower consumption time series, the seasonal decomposition (SD) method is more suitable for hydropower consumption forecasting, relative to other decomposition methods, such as empirical mode decomposition (EMD) [34]. Therefore, based on the “decomposition and ensemble” principle, a novel SD based support vector regression (LSSVR) ensemble learning paradigm is proposed for hydropower consumption forecasting. Firstly, due to the seasonality of hydropower, hydropower consumption series is decomposed into a trend-cycle value, a seasonal factor and an irregular component, via a SD technique, simplifying the difficult task into three relatively easy subtasks. Then the LSSVR is applied to predict each component independently. Finally, prediction results of all components are aggregated, using another LSSVR model to produce an ensemble forecasting result for original hydropower series.

Particularly, LSSVR which was proposed by Suykens and Vandewalle [35] in 1999 has been used in many fields [36–39]. LSSVR is a modified version of SVR. It changes inequality constraints into equations and takes a squared loss function comparing with
traditional SVRs. Consequently, LSSVR solves a system of equations instead of a quadratic programming (QP) problem which leads to important improvement of calculating speed. In addition, it retains principle of the structural risk minimization (SRM) and has good generalization capability. Generally, LSSVR is a promising forecasting tool relative to statistical and other AI models (e.g., ANN). Due to these advantages of LSSVR, we employ it as the prediction model.

The main original contributions of this paper reflect the following three aspects. First of all, a novel SD-based least squares support vector regression (LSSVR) ensemble learning approach is proposed for Chinese hydropower consumption forecasting. To our knowledge, the SD-based LSSVR ensemble learning method has not been found in the previous studies. Second, the proposed SD-based LSSVR ensemble learning paradigm is proved to be very effective to predict complex time series with seasonality. Third, the proposed SD-based LSSVR ensemble learning approach can be used in both one-step ahead forecasting and multi-step ahead forecasting simultaneously. Furthermore, the empirical results demonstrated that the long-term prediction performance of the proposed SD-based LSSVR ensemble learning method is better than that of short-term predictions of all benchmark models.

The main motivation of this study is to propose a SD-based LSSVR ensemble learning approach for hydropower consumption prediction and compare its forecasting performance with some existing forecasting methods. The rest of the article is organized as follows. Section 2 describes the formulation process of the proposed SD-based LSSVR ensemble learning methodology. For illustration and verification purposes, some experiments on Chinese hydropower consumption are performed, and the corresponding results are reported in Section 3. Section 4 concludes the article.

2. Formulation of SD-based LSSVR ensemble learning methodology

In this section, an overall process of formulating the SD-based LSSVR ensemble learning paradigm is presented. First, the SD technique and LSSVR algorithm are briefly introduced, and then the SD-based LSSVR ensemble learning methodology is proposed.

2.1. SD approach

In existing econometric literature, a time series is usually divided into systematic pattern and random noise. In most cases, the random noise is considered as a white noise process with zero mean, while the systematic pattern is considered as a combination of trend, cycle and seasonality [40].

When a time series with clear seasonal pattern is analyzed, it is necessary to extract the seasonal factor from the original time series. In the existing SD methods, the most commonly used method is the X-12-ARIMA method, which is the Census Bureau’s latest seasonal adjustment program [41]. Thus, this study will select X-12-ARIMA as SD method. Basically, X-12-ARIMA method decomposes the time series data into three components, namely trend-cycle component $tC_t$, seasonal factor $sF_t$, and irregular component $iR_t$. In X-12-ARIMA method, multiple forms can be used to combine these components into the original data. Typically, the most widely used forms are the additive and multiplicative forms which are defined below, respectively:

\[ x_t = tC_t + sF_t + iR_t \]  \hspace{1cm} (1)

\[ x_t = tC_t \times sF_t \times iR_t \]  \hspace{1cm} (2)

Comparing the two forms of SD, the multiplicative decomposition is more suitable for most seasonal time series. The main reasons leading to the priority can be summarized into two aspects. On the one hand, the seasonal factor of the multiplicative form is a relative value of the original series. On the other hand, most seasonal time series with positive values has the characteristic that the scale of seasonal oscillations increases following the level of original series [42]. For this purpose, this study will employ the multiplicative form for SD via X-12-ARIMA program. For abbreviation purpose, X-12-ARIMA-based SD with Multiplicative Form is shortened as SD (M). For more details about X-12-ARIMA, please refer to [41].

Besides X-12-ARIMA method, another popular SD method is the ratio-to-moving-average (RMA) method, first proposed by Macaulay [43]. The basic idea of the RMA method is to separate a time series into two components, i.e., trend-cycle-irregular component and seasonal factor. Unlike SD (M) method, the RMA seasonal factor is fixed and invariable between different years. Similar to SD, the multiplicative forms of RMA is chosen for this study and it is shortened as RMA (M) for abbreviation. For comparison purpose, the RMA (M) method is selected as the benchmark method relative to SD (M) method in this study.

2.2. LSSVR methodology

In contrast to other forecasting approaches, SVM firstly proposed by Vapnik [44] in 1995 based on the principle of SRM has been proved to possess their own excellent capabilities in classification or prediction even for small sample, by minimizing an upper bound of the generalization error. Generally speaking, SVM can be applied into classification and regression, i.e., support vector classification (SVC) and support vector regression (SVR). Nevertheless, SVM training is a time consuming process when analyzing huge data. For this purpose, least squares support vector machine (LSSVM) is proposed to overcome these shortcomings [35]. Accordingly, LSSVM can be categorized into LSSVR and LSSVC for regression and classification purposes. In this study, the LSSVR is used for a prediction tool. In the following sections, there will be brief descriptions to SVR and LSSVR, respectively.

2.2.1. SVR approach

SVR is a nonlinear regression model based on statistical learning theory. Due to its adoption of SRM principle it would alleviate the over-fitting and local minima issue and its solution is more stable and global optimum.

In SVR, training dataset is assumed to be \( \{x_i, y_i\} (i = 1, 2, ..., l) \) in which \( x_i \) is the input vector and \( y_i \) is its corresponding target vector. The purpose of applying SVR is to find out the underlying relationship between the input data \( x \) and output data \( y \). As the input data and the output data are assumed to be a nonlinear relationship in most cases, SVR maps the training data space into a high-dimensional feature space by a nonlinear mapping function \( \phi(x) \) and then makes linear regression in the high-dimension feature space to attain the original space nonlinear regression effect. The regression function can be formulated as follows,

\[ y(x) = w^T \phi(x) + b \]  \hspace{1cm} (3)

where \( \phi(x) \) is called the nonlinear function mapping from input space into a high-dimensional feature space, and \( y \) is the estimated value in term of input data \( x \). Usually, coefficients \( w \) and \( b \) are obtained by minimizing the upper bound of generalization error. Accordingly, Eq. (3) can be transformed into the following optimization problem:
\[
\begin{align*}
\min & \quad \frac{1}{2}w^T w + \gamma \sum_{i=1}^{l} (\xi_i + \xi_i^*) \\
\text{s.t.} & \quad w^T \varphi(x_i) + b - y_i \leq \epsilon + \xi_i, (i = 1, 2, \ldots, l) \\
& \quad y_i - (w^T \varphi(x_i) + b) \leq \epsilon + \xi_i^*, (i = 1, 2, \ldots, l) \\
& \quad \xi_i, \xi_i^* \geq 0, (i = 1, 2, \ldots, l)
\end{align*}
\]

where \(\gamma\) is the regularization parameter which is used to avoid the over-fitting, and non-negative variables \(\xi_i\) and \(\xi_i^*\) are the slack variables which represents the distance from actual values to the corresponding boundary values of \(\epsilon\)-tube. Both \(\gamma\) and \(\epsilon\) are user-determined parameters. In addition, s.t. is the abbreviation of “subject to” which marks the constraints.

It is worth noticing that the problem-solving process of the SVR is a QP problem. There inevitably exists a high computational complexity specially when computing large-scale QP problem. For this purpose, LSSVR is proposed by Suykens and Vandewalle [35] to overcome these shortcomings.

\subsection*{2.2.2. LSSVR approach}

LSSVR is the least squares version of SVR. In LSSVR, the regression problem can be transformed into the following optimization problem:

\[
\begin{align*}
\min & \quad \frac{1}{2}w^T w + 1/2\gamma \sum_{i=1}^{l} e_i^2 \\
\text{s.t.} & \quad y_i = w^T \varphi(x_i) + b + e_i, (i = 1, 2, \ldots, l)
\end{align*}
\]

where \(e_i\) are the error variables and \(\gamma\) is the penalty parameter, \(\gamma\) is applied to control the minimization of estimation error and the function smoothness.

In order to solve the optimization problem, the Lagrangian function is formulated below.

\[
L(w, b, e, \alpha) = \frac{1}{2}w^T w + \frac{1}{2}\gamma \sum_{i=1}^{l} e_i^2 - \sum_{i=1}^{l} \alpha_i [w^T \varphi(x_i) + b + e_i - y_i]
\]

where \(\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_l)\) is the Lagrange multiplier. The KKT conditions are used for optimality by differentiating \(L\) in (6) with the variable \(w, b, e, \alpha\), which is shown as follows.

\[
\begin{align*}
\frac{\partial L}{\partial w} = 0 \Rightarrow w &= \sum_{i=1}^{l} \alpha_i \varphi(x_i) \\
\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i = 0 \\
\frac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = \gamma e_i, \; i = 1, \ldots, l \\
\frac{\partial L}{\partial \alpha_i} = 0 \Rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0, \; i = 1, \ldots, l
\end{align*}
\]

By solving the upper linear system, the final solution of the primal problem can be represented in the following form.

\[
f(x) = \sum_{i=1}^{l} w_i K(x_i, x) + b
\]

In Eq. (8), \(K(\bullet, \bullet)\) is the so-called kernel function which can simplify the use of a mapping. Usually any symmetric kernel function \(K(\bullet, \bullet)\) satisfying Mercer’s condition corresponds to a dot product in some feature spaces [44]. The most used kernel functions include the Gaussian RBF \(K(x_i, x_j) = \exp(||x - x_j||/2\sigma^2)\) with a width of \(\sigma\), and the polynomial kernel \(K(x_i, x_j) = (\alpha_1 x_i + \alpha_2)^d\) with an order of \(d\) and constants \(\alpha_1\) and \(\alpha_2\). For more details about LSSVR can refer to [35,45].

Generally, LSSVR has been modified two points of traditional SVR formulation. First, it takes equality constraints instead of inequality constraints in SVR. Second, a squared loss function is taken for error variable \(e_i\) in LSSVR [45]. With these modifications, instead of solving a QP problem, LSSVR is solved by a set of linear equations which leads to important reduction in computational complexity.

In this study, the most popular kernel function, Gaussian RBF kernel, is employed, and the kernel parameter \(\sigma^2\) and \(\gamma\) should be determined beforehand. Currently, many approaches have been applied in parameter optimization of SVR, such as experience [38], grid search [22,36], cross-validation [39,46–48], particle swarm optimization (PSO) [49], genetic algorithm (GA) [49,50], simulated annealing algorithm [49,51], etc. Considering computing complexity, the most useful used method, cross-validation grid search, is selected to determine the parameters \(\sigma^2\) and \(\gamma\) in LSSVR modeling [52].

\subsection*{2.3. Overall process of the SD (M)-based LSSVR ensemble learning paradigm}

Suppose there is a time series \(x_t (t=1, 2, \ldots, T)\), in which one would like to make the \(h\)-step ahead prediction, i.e. \(x_{t+h}\). For example, \(h=1\) means one-step-ahead prediction and \(h=6\) represents six-step-ahead prediction. Specifically, the univariate model used in our study can be described in the following form:

\[
\hat{x}_{t+h} = f(x_t, x_{t-1}, \ldots, x_{t-s})
\]

where \(x_{t+h}\) is the output data, while \(x_t, x_{t-1}, \ldots, x_{t-s}\) are the input data. In addition, \(h\) is the size of horizon and \(s\) presents the lag period which is determined by analyzing the autocorrelation and partial correlation of the time series. Depending on the techniques and methods presented in the previous subsections, the SD-based LSSVR ensemble learning paradigm can be formulated, as illustrated in Fig. 1.

As shown in Fig. 1, the proposed SD-based LSSVR ensemble learning methodology is generally composed of the following three main steps:

1. The original time series \(x_t (t=1, 2, \ldots, T)\) is decomposed into seasonal factor (SF) \(s_i\), trend cycle (TC) \(t_c\), and irregular component (IR) \(i_r\) via SD (M).
2. For the decomposed components, \(s_i, t_c\) and \(i_r\), the LSSVR is used as a forecasting tool to fit the decomposed components, and to make the corresponding prediction for each one.
3. The prediction results of SF, TC and IR are combined as an aggregated output using another LSSVR model, which can be used as the final prediction result for the original time series.

To summarize, the proposed SD (M)-based LSSVR ensemble learning paradigm is actually an “SD (M)–LSSVR (single prediction)–LSSVR (ensemble prediction)” ensemble learning approach, based on the principle of “decomposition and ensemble”.

In order to verify the effectiveness of the proposed SD (M)-based LSSVR ensemble learning paradigm, the hydropower consumption series in China is used as a testing target, which is illustrated in the next section.

\subsection*{3. Experimental analysis}

In this section, the data and evaluation criteria used in this study are first described. Then, experimental results and corresponding analysis and explanations are reported.
3.1. Data description and experiment design

In this study, the hydropower consumption data in China are obtained from Wind Database (http://www.wind.com.cn/) and originally provided by National Bureau of Statistics of China. The sample data are monthly data of hydropower consumption in China, covering the period from January 1986 to February 2010 with a total of 290 observations, as shown in Fig. 2.

In these hydropower consumption data, the data from January 1986 to December 2006 is taken as training dataset (252 observations), which are used to evaluate the forecasting performance. In order to evaluate the prediction capability of the proposed ensemble learning methodology, we carry out 1-step-ahead and 6-step-ahead prediction to test the short-term and long-term forecasting capability, respectively, and 2-step-ahead and 3-step-ahead forecasts are performed for medium-term prediction capability evaluation.

For comparison of the forecasting performance of multiple different models, three criteria are used for evaluation of level prediction and directional forecasting. The mean absolute percent error (MAPE) and the root mean squared error (RMSE) are used as the evaluation criteria of the level prediction:

\[
\text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{x_t - \hat{x}_t}{x_t} \right|
\]

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left( x_t - \hat{x}_t \right)^2}
\]

where \( \hat{x}_t \) is the predicted value for \( x_t \), and \( N \) is the number of predictions.

Apart from the level prediction accuracy, directional prediction accuracy is another important criteria for forecasting models. The ability to predict movement direction can be measured by a directional statistic (\( D_{\text{stat}} \)) [33], which can be represented as

\[
D_{\text{stat}} = \frac{1}{N} \sum_{t=1}^{N} a_t \times 100\%
\]

where \( a_t = 1 \) if \( (x_{t+1} - x_{t})(\hat{x}_{t+1} - \hat{x}_{t}) \geq 0 \), and \( a_t = 0 \) otherwise.

In addition to the evaluation criteria, Diebold–Mariano statistic (DM statistic) [53] is used to test the statistical significance of accuracy of two competing prediction models. Consider two forecasts \( \{\hat{x}_{it}\}_{t=1}^{N} \) and \( \{\hat{x}_{jt}\}_{t=1}^{N} \) of the time series \( x_t \) using model \( i \) and model \( j \), the associated forecasts absolute errors are defined as \( \{e_{it}\}_{t=1}^{N} \) and \( \{e_{jt}\}_{t=1}^{N} \). Diebold–Mariano test aims to test the null hypothesis of equal accuracy of two competing models in terms of an arbitrary loss function \( g(e_t) \). The loss difference between model \( i \) and model \( j \) is defined as \( d_t = g(e_{it}) - g(e_{jt}) \). If the loss difference series is covariance stationary and short memory, the asymptotic distribution of the sample mean loss difference \( \bar{d} = \frac{1}{N} \sum_{t=1}^{N} [g(e_{it}) - g(e_{jt})] \) is given by

\[
\sqrt{N} (\bar{d} - \mu) \rightarrow N(0, \sigma(\bar{d}))
\]

where \( \gamma_0 + 2 \sum_{k=1}^{h-1} \gamma_k \) and \( \gamma_k = \text{cov}(d_t - d_{t-k}) \). \( h \) is the prediction horizon. Therefore, the asymptotically standard normal DM statistic can be obtained as
where $\hat{V}(\overline{d})$ is the consistent estimate of $V(\overline{d})$, based on the sample auto-covariance $\gamma_k$ is given by $\gamma_k = 1/N \sum_{t=|k|+1}^N (d_t - \overline{d})(d_{t-k} - \overline{d})$. In this study, the loss function is set to absolute error (AE) and the null hypothesis is that the AE of a specific model is greater than that of another model.

To evaluate the forecasting ability of the proposed SD (M)-LSSVR-LSSVR ensemble learning model, this study compares its performance with some popular single forecasting models, such as ANN, ARIMA, SARIMA, SVR, LSSVR, and some variants of ensemble learning with other decomposition methods, e.g. EMD method and RMA (M) method, which are employed as benchmark models.

Amongst these single methods, LSSVR and SVR approaches have been already introduced above. Thus, we only give a brief review for ARIMA, SARIMA and ANN algorithms.

In an ARIMA model [40], the future value of a variable is assumed to be a linear function of several past observations and random errors. That is, the underlying process that generates a time series takes the following form

$$q_p(B)(1 - B)^d x_t = \theta_q(B) e_t$$  \hfill (15)

Considering the seasonal pattern of some time series, Box and Jenkins developed an ARIMA model to cope with seasonal time series, usually called SARIMA method [40]. A time series $x_t$ is generated by SARIMA $(p,d,q)(P,D,Q)$ process if

$$q_p(B)\phi_p(B^s)(1 - B)^d(1 - B^s)^D x_t = \theta_q(B)\Theta_Q(B^s) e_t$$  \hfill (16)

where $p, d, q, P, D,$ and $Q$ are integers, $s$ is the seasonal length; $x_t$ and $e_t$ are the observed value and random error at time $t$, respectively; random errors, $e_t$, are assumed to be independently and identically distributed with a mean of zero and a constant variance of $\sigma^2$; i.e., $e_t \sim IID(0, \sigma^2)$. The four abbreviate functions are presented by

$$\phi_p(B) = 1 - \sum_{k=1}^p \phi_k B^k \quad \theta_q(B) = 1 - \sum_{k=1}^q \theta_k B^k$$
$$\phi_p(B^s) = 1 - \sum_{k=1}^p \phi_k B^{ks} \quad \theta_Q(B^s) = 1 - \sum_{k=1}^Q \Theta_k B^{ks}$$

where $\phi_p(B)$ is the autoregressive operator of order $p$, $\theta_q(B)$ is a moving average operator of order $q$, $\phi_p(B^s)$ and $\Theta_Q(B^s)$ are polynomials in $B^s$ of order $P$ and $Q$. $B$ denotes the backward shift operator, i.e., $Bx_t = x_{t-1}$, $B^2x_t = x_{t-2}$ and so on; $d$ is the number of regular difference, $D$ is the order of seasonal differences.

In the ANN model, a standard three-layer feed-forward neural network (FNN) is used [54,55]. Usually, an FNN-based forecasting model can be trained by the in-sample dataset and applied to out-of-sample dataset for prediction. Thus, the standard FNN $(I, H, O)$-based model comprises of an input layer with $I$ input neurons, a hidden layer with $H$ nodes and an output layer with $O$ output neurons. The model parameters, connection weights and node biases, are adjusted iteratively by a process of minimizing the forecasting error function. Basically, the final output of the FNN-based forecasting model can be expressed as

$$f(x) = \phi_0 \left( a_0 + \sum_{j=1}^H w_j \phi_H \left( a_j + \sum_{i=1}^I w_{ij} x_i \right) \right)$$  \hfill (21)

where $x_t(i = 1, 2, ..., I)$ represents the input patterns, $f(x)$ is the output, $a_j / (j = 0, 1, ..., H)$ is a bias on the fth unit, and $w_{ij}/(i = 1, 2, ..., I; j = 1, 2, ..., H)$ and $w_j/(j = 0, 1, ..., H)$ are the connection weights between layers of
the model; $\varphi_h(\cdot)$ and $\varphi_o(\cdot)$ present the transfer function of the hidden layer and the output layer.

In the ensemble learning methodologies, EMD and RMA (M) technologies are employed as decomposition benchmark methods. Thus, two variants of the ensemble learning paradigm, EMD-LSSVR-LSSVR, RMA (M)-LSSVR-MUL approaches are used for comparison. In these two ensemble learning models, the first part mentioned in these ensemble learning methodologies is referred as decomposition algorithms, and the second part is a single prediction approach for each decomposed components, while the last part is the ensemble approach. Taking EMD-LSSVR-LSSVR methodology for example, this methodology applied EMD technique to the original data for decomposition purpose and then the derived decomposed modes are predicted via LSSVR model, and finally another LSSVR model is employed to combine all the prediction results obtained from the previous single prediction models into final results. In the methodology of RMA (M)-LSSVR-MUL, the ensemble approach (MUL) is applied to multiply the predicted result of trend-cycle-irregular component by the constant seasonal factor.

As RMA (M) and SD (M) have been already introduced above, we only simply introduce EMD algorithm [34]. The empirical mode decomposition (EMD) technique is a form of adaptive time series decomposition technique using the Hilbert–Huang transform (HHT) especially for nonlinear and non-stationary data. It decomposes the original time series into a series of oscillatory functions, namely, intrinsic mode functions (IMFs) and a residual. That is, after performing EMD method, the original data series $x_t$ can finally be expressed as the sum of IMFs and a residual:

$$x_t = \sum_{j=1}^{n} c_j + r^M_t$$  \hspace{1cm} (22)

where $n$ is the number of IMFs, $r^M_t$ is the final residual, and $c_j (j = 1, 2, \ldots, n)$ is the jth IMF, which are nearly orthogonal to each other, and all have nearly zero means. Thus, the data series can be decomposed into $n$ empirical mode functions and one residual. The IMF components contained in each frequency band are different and they change with variation of the original time series $x_t$ while $r^M_t$ represents the central tendency of data series $x_t$.

Therefore, in this study, five single models, including single LSSVR, SVR, ANN, SARIMA, and ARIMA models, and two variants of the ensemble learning paradigm, including EMD-LSSVR-LSSVR, RMA (M)-LSSVR-MUL methodologies are employed to predict hydropower consumption for comparison purpose.

3.2. Experimental results

In this study, two decomposition methods, SD (M) and RMA (M) are utilized in the EVIEWS software package, EMD method are applied in the MATLAB software package, which is produced by MathWorks Laboratory Corporation. Three single prediction models, SVR, LSSVR and ANN, are implemented, respectively, via SVMlab2.91-1 toolbox, LSSVMlab1.5 toolbox and neural network toolbox 5.0 of MATLAB software package, while the single SARIMA and ARIMA models are built using the EVIEWS software package, which is produced by Quantitative Micro Software Corporation.

In terms of the experiment design and the overall steps described in Section 2.3, the proposed SD (M)-based LSSVR ensemble learning paradigm is first used to perform numerical experiments, and the prediction performance comparisons with the other benchmark models listed in Section 3.1 are then reported. In the proposed SD (M)-based LSSVR ensemble learning paradigm, the first step is to apply X-12-ARIMA algorithm to decompose the original hydropower consumption data series into seasonal factor, trend cycle and irregular component, as shown in Fig. 3. The second step of the SD (M)-based LSSVR ensemble learning paradigm is the single forecasting of SF, TC and IR via LSSVR. Finally, another LSSVR model is used to combine the prediction results of each decomposed component into an ensemble forecasting result. Thus there are four LSSVR should be determined in the proposed SD (M)-based LSSVR ensemble model. Besides, 1, 2, 3, 6-step-ahead predictions are performed, the total number of LSSVR is 16. In the benchmark methods, EMD-LSSVR-LSSVR, RMA (M)-LSSVR-MUL and single LSSVR should determine LSSVR parameters. For EMD,
the original data is decomposed into seven IMFs and one residual, there are 36 LSSVR models in EMD-LSSVR-LSSVR for four different step-ahead forecasting. Similarly, RMA (M)-LSSVR-MUL has four LSSVR models corresponding to four different step-ahead prediction. In addition, four LSSVR models should be selected in single LSSVR method considering four different step-ahead predictions. To sum up, 60 LSSVR models should be chosen. In this study, we employ Gaussian RBF as the kernel function of LSSVR, and the values of $\gamma$ and $\sigma^2$ parameters are first determined via 5-fold cross-validation grid search method in the range of $[0.01, 10000]$, and then they are adjusted by the trial-and-error approach to produce the smallest error in the training set [52]. Considering there are 60 LSSVR models, the details about the parameters of 60 LSSVM models are not listed for space consideration and they can be obtained by the above cross-validation grid search method and trial-and-error method [52].

Apart from single LSSVR, other four single prediction models ANN, SVR, ARIMA and SARIMA are used for comparison purpose. The ANN ($I\rightarrow H\rightarrow O$) in this study used $l$ input neurons, $10$ hidden nodes (i.e., $H = 10$) and one output neuron (i.e., $O = 1$). Here $l$ is referred to lag period of the predicted time series, which are determined by the testing of autocorrelation and partial correlation of the time series. That is, the prediction value of $x_t$ is estimated based on the previous data of $x_{t-1}, \ldots, x_{t-l}$ and thus there are $l$ input neurons in the ANN model. The hidden nodes use logistic transfer function and the output layer uses the linear transfer function because the prediction performance is best when these transfer functions are used. The ANN models are iteratively run 10,000 times to train the model using the training subset.

In the single SVR models, the Gaussian RBF kernel function is also selected, and the values of parameters $\gamma$, $\sigma^2$, $\epsilon$ are determined by cross-validation method and the trial-and-error method, the parameter value of $\epsilon$ in SVR is set to 0.05. In the ARIMA ($p$-$d$-$q$) model and SARIMA ($p$-$d$-$q$)($P$-$D$-$Q$) model, the best model for each training sample is determined through the minimization of Schwarz Criterion (SC).

Using the experiment design and methodologies mentioned above, the forecasting experiments for Chinese hydropower consumption are performed and accordingly the prediction performances are evaluated by the three main measure criteria and the DM test. Tables 1–4 show the performance comparisons of different models through DM test in 1-step-ahead, 2-step-ahead, 3-step-ahead, and 6-step-ahead prediction, respectively. The three main measurement criteria RMSE, MAPE and $D_{stat}$ of different methods are illustrated in Figs. 4–6.

From Figs. 4–6, it is easy to find that the proposed SD (M)-LSSVR-LSSVR method is the best one for hydropower consumption forecasting in all prediction horizons (i.e.1-step-ahead, 2-step-ahead, 3-step-ahead, 6-step-ahead), relative to other models listed in this study. In all models listed in this study, the proposed SD (M)-based LSSVR ensemble learning paradigm do not only achieve the highest accuracy from the level measurement, which is measured by the RMSE and MAPE criteria, but also gets the highest hit rate in direction prediction, which is measured by $D_{stat}$ criterion.

On the other hand, among all of the models used in this study, the single ARIMA model performs the poorest in all step-ahead predictions. ARIMA model not only has the lowest level accuracy which is measure by RMSE and MAPE, but also gets the worst hit rate in direction prediction, which is measured by $D_{stat}$ criterion. The main reason may be that ARIMA is a class of typical linear model and it cannot capture the nonlinear patterns and seasonal characteristic existing in the data series.

Apart from the proposed SD (M)-LSSVR-LSSVR method and ARIMA model which perform the best and the poorest, respectively, in all the models listed in this study, other models produce some interestingly mix results. These results are analyzed from three accuracy measurement criteria (i.e. RMSE, MAPE and $D_{stat}$), respectively.

Firstly, in the case of level prediction accuracy, the results of RMSE criteria show that the SD (M)-LSSVR-LSSVR performs the best, followed by EMD-LSSVR-LSSVR, RMA (M)-LSSVR model, the single AI model and SARIMA model, and the poorest model is ARIMA model. The RMSE values of SD (M)-LSSVR-LSSVR are 26.54, 28.60, 28.85 and 32.86 of 1, 2, 3, 6-step-ahead predictions which is clearly less than other methods. In addition, the small prediction step, the better performance. Comparing ensemble methods with single models, the results indicate that the ensemble methods outperform single methods in most cases. The main reason could

<table>
<thead>
<tr>
<th>Tested model</th>
<th>Reference model</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMD-LSSVR-LSSVR</td>
<td>3.1088 (0.0009)</td>
</tr>
<tr>
<td>EMD-LSSVR-LSSVR</td>
<td>-1.7857 (0.0371)</td>
</tr>
<tr>
<td>RMA (M)-LSSVR-MUL</td>
<td>5.8349 (0.0000)</td>
</tr>
<tr>
<td>LSSVR</td>
<td>-3.6444 (0.0001)</td>
</tr>
<tr>
<td>ANN</td>
<td>-4.1879 (0.0000)</td>
</tr>
<tr>
<td>SVR</td>
<td>-2.6998 (0.0035)</td>
</tr>
<tr>
<td>SARIMA</td>
<td>-3.3753 (0.0004)</td>
</tr>
<tr>
<td>ARIMA</td>
<td>-3.3753 (0.0004)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tested model</th>
<th>Reference model</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMD-LSSVR-LSSVR</td>
<td>-3.1088 (0.0009)</td>
</tr>
<tr>
<td>EMD-LSSVR-LSSVR</td>
<td>-0.2737 (0.4060)</td>
</tr>
<tr>
<td>RMA (M)-LSSVR-MUL</td>
<td>0.6817 (0.7523)</td>
</tr>
<tr>
<td>LSSVR</td>
<td>-2.5214 (0.0058)</td>
</tr>
<tr>
<td>ANN</td>
<td>1.8829 (0.9701)</td>
</tr>
<tr>
<td>SVR</td>
<td>2.7295 (0.9968)</td>
</tr>
<tr>
<td>SARIMA</td>
<td>-3.1522 (0.0008)</td>
</tr>
<tr>
<td>ARIMA</td>
<td>-3.5641 (0.0002)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tested model</th>
<th>Reference model</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMD-LSSVR-LSSVR</td>
<td>-2.5932 (0.0048)</td>
</tr>
<tr>
<td>EMD-LSSVR-LSSVR</td>
<td>-0.1442 (0.4427)</td>
</tr>
<tr>
<td>RMA (M)-LSSVR-MUL</td>
<td>0.2063 (0.5817)</td>
</tr>
<tr>
<td>LSSVR</td>
<td>-1.5843 (0.0566)</td>
</tr>
<tr>
<td>ANN</td>
<td>1.6096 (0.9463)</td>
</tr>
<tr>
<td>SVR</td>
<td>-0.10785 (0.1404)</td>
</tr>
<tr>
<td>SARIMA</td>
<td>-3.5955 (0.0002)</td>
</tr>
</tbody>
</table>
be that the decomposition strategy does effectively improve the prediction performance. Interestingly, the 1-step-ahead prediction of EMD-LSSVR-LSSVR and 3-step-ahead forecasts of RMA (M) are inferior to single LSSVR and ANN. The possible reasons are unknown, which is worth exploring further in the future. Focusing on the single methods, all AI models and SARIMA model outperform ARIMA model whose RMSE values are close to 143. The reason may be that ARIMA is a typical linear model, which is not suitable for capturing the nonlinear and seasonal characteristic of hydropower consumption series. In AI models (i.e., LSSVR, ANN, SVR), it can be seen that ANN performs slightly better than LSSVR except for 3-step-ahead prediction and SVR performs poorest in all steps prediction. The main reason leading to this may be from the parameter selection. The RMSE values of ARIMA are from 69.74 to 84.96 which are slightly inferior to SVR. The possible reason is that the prediction results of SARIMA which is under the linear hypothesis are more volatile than those of the AI models.

Second, in the case of another level accuracy of MAPE criteria, RMA (M)-LSSVR is the second to SD (M)-LSSVR-LSSVR, followed by ANN, LSSVR, EMD-LSSVR-LSSVR, SARIMA, SVR, and ARIMA model. The MAPE values of SD (M)-LSSVR-LSSVR are 0.0562, 0.601, 0.0587 and 0.0602 of 1, 2, 3, 6-step-ahead predictions which are the minimum. On the other hand, all the MAPE values of ARIMA are close to 0.235 which is the maximum. Unlike the results of RMSE criterion, EMD-LSSVR-LSSVR is inferior to some single AI models (i.e., ANN, LSSVR) with one exception of 6-step-ahead prediction which ranks the second. The main reason leading to the above-mentioned results may be that the EMD method does not efficiently catch the seasonal pattern of hydropower consumption. Except for the proposed SD (M)-LSSVR-LSSVR, RMA (M)-LSSVR outperforms other models other than 6-step-ahead prediction. In AI models (i.e., LSSVR, ANN, SVR), SVR performs poorest in all steps prediction and ANN performs slightly better than LSSVR except for 6-step-ahead prediction. In addition, SARIMA outperforms slightly SVR other than 1-step-ahead forecasts. The possible reason is that the SARIMA model can capture the seasonal pattern while the SVR can model the nonlinear factors hidden in the dataset. For this reason, one of them cannot be superior to the other one in all cases.

Third, the high level accuracy does not necessarily mean that there is a high hit rate in forecasting direction of hydropower consumption. The correct forecasting direction is important for policy manager to make investment plan in hydropower. Therefore the $D_{\text{stat}}$ comparison is necessary. In Figs. 4–6, some similar conclusions can be drawn in terms of $D_{\text{stat}}$ criterion. (i) The proposed SD (M)-based LSSVR ensemble learning paradigm perform significantly better than all other models in all cases, followed by the other two ensemble models, then two of the single AI models (i.e., LSSVR, ANN), SARIMA, SVR, and lastly the ARIMA model. Specifically, the $D_{\text{stat}}$ values of SD (M)-LSSVR-LSSVR are 91.89% for 1, 2 step-ahead predictions and 89.19% for 3, 6 step-ahead predictions. (ii) The three ensemble methods mostly outperform the single prediction models. Furthermore, among the ensemble methods, SD (M)-LSSVR-LSSVR performs the best, and EMD-LSSVR-LSSVR outperforms RMA (M)-LSSVR-LSSVR model except for 3-step-ahead prediction. (iii) LSSVR outperforms other single models, SARIMA has the similar performance as ANN except for the one-step-ahead prediction and it outperforms SVR. The possible reason leading to this phenomenon may be the parameters choice. We also find that ARIMA model has the lowest directional accuracy about 56.76%.

In addition, comparing different prediction horizons, the shorter the prediction horizon, the better the models perform. Take 1-step-ahead forecasting and 6-step-ahead prediction for example, in SD (M)-LSSVR-LSSVR, RMA (M)-LSSVR-LSSVR, ANN, SVR and SARIMA models, the 1-step-ahead forecasting outperform the 6-step-ahead prediction, no matter level accuracy or directional accuracy. Apart from the models mentioned above, EMD-LSSVR-LSSVR and LSSVR performs better in 1-step-ahead prediction in view of directional accuracy. However, from the point of level accuracy, both these approaches only have slight superiority in 6-step-ahead prediction. The ARIMA perform almost the same between 1-step-ahead and 6-step-ahead prediction.

In order to further compare the predictive accuracy of different forecasting models, the DM statistic is used to test the statistical significance of the prediction performance of different models. Note that the results listed in Tables 1–4 are the DM test values, and $p$ values are in brackets.

It can be seen from Tables 1–4 that several interesting conclusions can be found. Firstly, the proposed SD (M)-LSSVR-LSSVR model outperforms other models at 5% statistical significance level. The correct forecasting direction is important for policy manager to make investment plan in hydropower. Therefore the $D_{\text{stat}}$ comparison is necessary. In Figs. 4–6, some similar conclusions can be drawn in terms of $D_{\text{stat}}$ criterion. (i) The proposed SD (M)-based LSSVR ensemble learning paradigm perform significantly better than all other models in all cases, followed by the other two ensemble models, then two of the single AI models (i.e., LSSVR, ANN), SARIMA, SVR, and lastly the ARIMA model. Specifically, the $D_{\text{stat}}$ values of SD (M)-LSSVR-LSSVR are 91.89% for 1, 2 step-ahead predictions and 89.19% for 3, 6 step-ahead predictions. (ii) The three ensemble methods mostly outperform the single prediction models. Furthermore, among the ensemble methods, SD (M)-LSSVR-LSSVR performs the best, and EMD-LSSVR-LSSVR outperforms RMA (M)-LSSVR-LSSVR model except for 3-step-ahead prediction. (iii) LSSVR outperforms other single models, SARIMA has the similar performance as ANN except for the one-step-ahead prediction and it outperforms SVR. The possible reason leading to this phenomenon may be the parameters choice. We also find that ARIMA model has the lowest directional accuracy about 56.76%.

In addition, comparing different prediction horizons, the shorter the prediction horizon, the better the models perform. Take 1-step-ahead forecasting and 6-step-ahead prediction for example, in SD (M)-LSSVR-LSSVR, RMA (M)-LSSVR-LSSVR, ANN, SVR and SARIMA models, the 1-step-ahead forecasting outperform the 6-step-ahead prediction, no matter level accuracy or directional accuracy. Apart from the models mentioned above, EMD-LSSVR-LSSVR and LSSVR performs better in 1-step-ahead prediction in view of directional accuracy. However, from the point of level accuracy, both these approaches only have slight superiority in 6-step-ahead prediction. The ARIMA perform almost the same between 1-step-ahead and 6-step-ahead prediction.

In order to further compare the predictive accuracy of different forecasting models, the DM statistic is used to test the statistical significance of the prediction performance of different models. Note that the results listed in Tables 1–4 are the DM test values, and $p$ values are in brackets.

It can be seen from Tables 1–4 that several interesting conclusions can be found. Firstly, the proposed SD (M)-LSSVR-LSSVR model outperforms other models at 5% statistical significance level.
with one exception that it is not significantly outperforms EMD-LSSVR-LSSVR model in 6-step-ahead prediction with p value 0.2928. ARIMA model is inferior to all of other models at 1% significance level and the maximum p value in Table 4 is 0.0094 compared with LSSVR.

Secondly, in the two other ensemble learning model, EMD-LSSVR-LSSVR is not significantly superior to RMA (M)-LSSVR-LSSVR model except for six-step-ahead prediction. However, the two ensemble learning methods outperform ANN, SARIMA and ARIMA at certain statistical significance level in most cases, but not for LSSVR and SVR.

Thirdly, focusing on the single models, both LSSVR and SVR outperform ANN at 10% statistical significance. LSSVR is not superior to SVR significantly, and similarly LSSVR and ANN do not significantly outperform SARIMA. Particularly, SVR outperforms SARIMA at 10% statistical significance level in 1-step-ahead prediction, but SVR do not significantly outperform SARIMA in 2, 3, 6-step-ahead predictions.

Generally speaking, from the analysis of the experiments presented in this study, we can draw the following several important conclusions.

(1) The proposed SD (M)-based LSSVR ensemble learning methodology is significantly superior to the other benchmark methods in this study in terms of both level accuracy and direction accuracy. Through DM test, the SD (M)-LSSVR-LSSVR outperforms other models at 5% statistical significance level.

(2) The prediction performance of the SD (M)-based LSSVR ensemble model, EMD-LSSVR-LSSVR and RMA (M)-LSSVR-LSSVR are generally better than the single methods. This indicates that the “divide and conquer” principle or “decomposition and ensemble” strategy can effectively improve the prediction performance in the case of hydropower consumption.

(3) Due to the seasonal, nonlinear and non-stationary appearance of hydropower consumption, nonlinear models with seasonal...
adjustment are more suitable to predict the time series with seasonal volatility than linear methods.
(4) The proposed SD (M)-based LSSVR ensemble learning paradigm, which is an effective SD and nonlinear prediction methodology, can be used as a promising solution to complex time series forecasting with nonlinearity and seasonality.

In addition, to show the predictability of the proposed SD (M)-based LSSVR ensemble learning model, we extrapolate the hydropower consumption for future 24 months from August 2011 to July 2013. It can be found that the structure of the hydropower consumption begins to change in April 2003 at 1% statistic significance level through Chow test [56]. The main reason is that the first machine sets of the Three Gorges hydroelectric station is used to generate electricity energy in June 2003 [1]. For this reason, we take the data from June 2003 to July 2011 as the sample data for testing. Based on the previous experiment designs and model descriptions, the extrapolated results are illustrated in Fig. 7 where the sample data from June 2003 to July 2011 are also shown for comparison purpose. Particularly, the curve with stars (*) in Fig. 7 represents the actual data covering from June 2003 to July 2011, and the curve with circles (.) is the extrapolate data for next 24 months (from August 2011 to July 2013). The main reason about division of actual data and extrapolated data is that there are no new data for hydropower consumption after July 2011 due to the delay of statistical data.

![Fig. 6. Performance comparisons for different methods in terms of $D_{stat}$ criteria.](image)

![Fig. 7. The extrapolated prediction for Chinese hydropower consumption. (The actual data period: 06/2003–07/2011, the extrapolate data period: 08/2011–07/2013).](image)
As can be seen from Fig. 7, we can find that obvious fluctuation remains as before. The fluctuation reflects the seasonality of hydrological system and the peak value of hydropower consumption in 2011 and 2012 are obtained at August. In addition, the first extrapolate data (August 2011) reduces over the same period of 2010, it can mainly attributed to the less rainfall in most parts of China in first half year of 2011 [57]. However, the overall tendency of hydropower consumption will grow steady. The main reasons leading to the increase tendency of hydropower consumption can be summarized into two points. On the one hand, the rapid economic growth of China will give rise to the increase of hydropower energy demand. On the other hand, the role of hydropower has become more and more important for its technical, economic and environmental benefits confronting with the energy shortage and environment degradation. Particularly the Eleventh Five-Year Plan (2006–2010) of Chinese Government pointed out that the hydropower installation capacity will have reached 190 million kW by 2010 and moreover 300 million kW by 2020 [1]. This plan has strikingly accelerated the development of hydropower, and significantly enhanced its consumption. However, the rate of growth of hydropower will be smooth in terms of the extrapolated results. The main reason leading to the steady growth reflects that the Three Gorges project is completed in 2009 and no such large hydropower project will be constructed in the next few years.

4. Conclusions

In this paper, a novel ensemble learning paradigm integrating SD and LSSVR based on the principle of “decomposition and ensemble” is proposed for the hydropower consumption forecasting in China and the good prediction performance are obtained through empirical experiments. Particularly, the proposed SD-based LSSVR ensemble learning model performs better in long-term prediction than other benchmark methods in short-term predictions. This shows that the proposed SD-based LSSVR ensemble learning paradigm can be used as a very promising methodology for hydropower consumption forecasting with seasonality.

In terms of the empirical results, several important conclusions can be drawn. First of all, the ensemble learning paradigms can outperform the single models, indicating that “decomposition and ensemble” strategy can effectively improve the prediction performance. Second, the nonlinear models with seasonal adjustment perform better than linear methods. This shows that the effects of seasonality should be taken into account in complex time series forecasting with seasonality. Third, the shorter the prediction horizon is, the better the prediction performance of the used forecasting models. This indicates the difficulty of long-term forecasting for any prediction model. Finally, the proposed SD-based LSSVR ensemble learning approach can be used in both one-step ahead forecasting and multi-step ahead forecasting, indicating the extensive applicability of the proposed SD-based LSSVR ensemble learning approach. This implies that the proposed SD-based LSSVR ensemble learning paradigm can be applied to other complex time series forecasting problems with seasonality. We will continue investigating these important issues in the future research.

Acknowledgments

The author would like to express their sincere appreciation to the Editor-in-Chief and the two independent referees for making valuable comments and suggestions. Their comments have helped improve the quality of the paper immensely. This work is partially supported by grants from the National Science Fund for Distinguished Young Scholars (NSFC No. 71025005), the National Natural Science Foundation of China (NSFC Nos. 90924024 and 70601029), and the Knowledge Innovation Program of the Chinese Academy of Sciences. The author gratefully acknowledges the support of K.C. Wong Education Foundation, Hong Kong.

References


Box GEP, Jenkins GM. Time series analysis: forecasting and control. 2nd ed. San Francisco: Holden Day; 1976.

U.S. Census Bureau. X-12 ARIMA reference manual version 0.3. See also: http://www.census.gov/ts/x12a/v03/x12adocv03.pdf; 2011.


