Omega 45 (2014) 80–91

Contents lists available at ScienceDirect

Omega

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

Combining forecasts of electricity consumption in China with time-varying weights updated by a high-order Markov chain model

Weigang Zhao [a], Jianzhou Wang [a,*], Haiyan Lu [b]

[a] School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China
[b] Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Received 16 April 2013
Accepted 3 January 2014
Processed by B. Lev
Available online 18 January 2014

Keywords:
Monthly electricity consumption in China
Time-varying-weight combining method
High-order Markov chain model
Multi-step-ahead combination forecasting
Out-of-sample forecasting accuracy

ABSTRACT

Electricity consumption forecasting has been always playing a vital role in power system management and planning. Inaccurate prediction may cause wastes of scarce energy resource or electricity shortages. However, forecasting electricity consumption has proven to be a challenging task due to various unstable factors. Especially, China is undergoing a period of economic transition, which highlights this difficulty. This paper proposes a time-varying-weight combining method, i.e. High-order Markov chain based Time-varying Weighted Average (HM-TWA) method to predict the monthly electricity consumption in China. HM-TWA first calculates the in-sample time-varying combining weights by quadratic programming for the individual forecasts. Then it predicts the out-of-sample time-varying adaptive weights through extrapolating these in-sample weights using a high-order Markov chain model. Finally, the combined forecasts can be obtained. In addition, to ensure that the sample data have the same properties as the required forecasts, a reasonable multi-step-ahead forecasting scheme is designed for HM-TWA. The out-of-sample forecasting performance evaluation shows that HM-TWA outperforms the component models and traditional combining methods, and its effectiveness is further verified by comparing it with some other existing models.

1. Introduction

An accurate prediction can help decision makers to develop an optimal action program, which plays the leading effect of reducing risk and then improves the economic and social benefits of management [1]. Accordingly, forecasting applications abound in management practice at both macro and micro levels, and forecasting methods are integral components of Management Science models in many fields [2]. Specially, demand forecasting (consumption forecasting) is a crucial area of management forecasting because the most concerned issue of market participants is often the market potentiality. Enterprises can make right decisions and sound managerial planning for production, finance, personnel and organization, only if accurate future demands are obtained [3,4].

Electricity is a product that is related to the national economy and the people's livelihood, and this relationship has been strengthened continuously due to the people's increasing dependence on the electricity supply [5,6]. Like ordinary products, planning for the production of electricity is also of crucial importance. Besides, electricity is hard to store so it is generally generated and then instantly used without any kept in reserve. This increases even more the need for power utilities to plan their electricity supply in a proactive manner [5]. Moreover, a reliable anticipation of future electricity consumption level is just the primary guideline for planning [6,7].

In contrary, inaccurate electricity consumption forecasts will be counterproductive. An overestimation will waste scarce energy resources, huge amounts of capital investment and lengthy construction time [6]. An underestimation will lead to even more negative outcomes such as electricity shortage. The severe 2004 China Electricity Shortage is an example which was mainly caused by the lack of installed capacity [8]. Clearly, if early warning had been effectively made in advance, derived from good forecasts, this calamity might have been avoided by taking response measures. However, power systems along with the electricity market are affected by various unstable factors including the natural and social environment, law and policy, holidays, technical progress, population growth and more [9]. Especially, China is undergoing a period of economic transition. All these greatly increase the uncertainty of electricity consumption series, which in turn makes it very difficult to establish a valid and feasible electricity consumption forecasting model [5].

To tackle this challenge, a wide variety of methods have been proposed including: statistical and econometric models such as multiple linear regression [10], autoregressive moving average

* Corresponding author. Tel.: +86 931 8914050; fax: +86 931 8912481.
E-mail addresses: zwgstd@gmail.com (W. Zhao), wjz@lzu.edu.cn (J. Wang), haiyan.lu@uts.edu.au (H. Lu).

0305-0483/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.omega.2014.01.002

models [11,12], functional nonparametric model and semi-functional partial linear model [13]; nonlinear and artificial intelligent (AI) models such as grey forecasting approach [6], abductive network machine learning [14], fuzzy logic methodology [15], radial basis function network [16], recurrent neural network [17], support vector regression [18], genetic programming [19,20], genetic algorithm [21,22] and the most mentioned backpropagation network [23–27]. These methods can generally provide good forecasts, however, the statistical and econometric models have the limitation of linear (or near linear) assumption, and the AI models often suffer from over-fitting or the difficulty of parameter selection [28,29]. To remedy these shortcomings, some hybrid methods have been introduced and obtained even better reported performances: Grey–Markov model [30], SARIMA model with residual modification [9], backpropagation network based on moving average [31], wavelet transform combined with machine learning and time series models [32], neural networks based on trend extraction [33] and weighted hybrid model where trend and seasonal components are predicted by combined method and SARIMA respectively [34].

In fact, it is rare that a single forecasting model is always best in all cases. Each model has its own particular strengths and weaknesses [35]. When multiple forecasting models are available, consideration can be given to the combining method which is regarded as an outstanding approach to take advantage of strengths of each model. Generally, finding the optimal combination is the traditional goal of this method, which minimizes the in-sample sum of squared forecast errors to find the optimal weights of each individual model [36]. However, the properties of the individual forecast error may vary over time. For example, Moghram and Rahman [37] reviewed five forecasting methods for short-term load forecasting. They modeled the summer loads and winter loads separately, and no one method was determined to be superior. The transfer function model gave the best results over the summer months, whereas it resulted in the second worst accuracy over the winter months [38].

Electricity consumption forecasting may also suffer from the above trouble, in which case the combining method using fixed weights may perform poorly. Thus it is more appropriate to allow the combining weights to change according to the time-varying underlying process. This paper proposes a novel time-varying-weight combining method, i.e. High-order Markov chain based Time-varying Weighted Average (HM-TWA) method to predict the monthly electricity consumption in China. The in-sample combining weights are first calculated through solving one optimization problem for each time point. Then treating the combining weight vector at each time point as a state probability distribution, the out-of-sample combining weights can be predicted by the high-order Markov chain model through extrapolating the in-sample weights. Along with this step, the parameter estimation method of the high-order Markov chain model has been modified, which generalizes its application from the categorical data sequence to the common state probability distribution sequence. Finally, the combined forecasts are obtained using such out-of-sample predicted weights. In addition, considering that the properties of the sample data should be the same as those of the required forecasts, a multi-step-ahead forecasting scheme has been designed for combining methods including HM-TWA. Specifically, this scheme utilizes the h-step-ahead in-sample individual forecasts to predict the h-step-ahead out-of-sample combined forecast. It can be applied for both the constant-weight method and the time-varying-weight method.

The rest of this paper is organized as follows: Section 2 specifies two traditional types of combining methods, while the high-order Markov chain model is described in Section 3. In Section 4, the proposed time-varying-weight combining method is shown and the experimental results along with a discussion of them are displayed in Section 5. Finally, Section 6 concludes this paper with the discussion on the contribution of this paper.

2. Combination forecasting methods

Let $\hat{x}_{j,t}$ denote the unbiased out-of-sample forecast for x_t, which is obtained by the jth individual model. Then the combined output at time t of the combining methods has the following weighted average form [39,40]:

$$\hat{x}_{c,t} = \sum_{j=1}^{m} w_j \hat{x}_{j,t}, \quad t = 1, 2, \ldots \tag{1}$$

where $\hat{x}_{c,t}$ is the combined output, m is the number of the component models and w_j is the weight on the jth component model. These weights are all constrained to be 0–1 and have to meet the following requirement to ensure that the combined forecast is unbiased [41–43]:

$$\sum_{j=1}^{m} w_j = 1 \tag{2}$$

Clearly, to perform the combination forecasting, the key issue is to estimate the weights $w_j, j = 1, 2, \ldots, m$. The Simple Average (SA) method and the Weighted Average (WA) method are the two most popular approaches to tackle this problem. The way of SA is to simply assign each $\hat{x}_{j,t}$ $(j = 1, 2, \ldots, m)$ an equal weight $w_j = 1/m$ [42].

The WA method is more general than the SA method. It can get an even higher accuracy through considering the individual and mutual characteristics of the individual forecasts [42]. Specifically, let $e_{j,t} = (x_t - \hat{x}_{j,t})$, $t = 1, 2, \ldots, T$, denote the residual of the jth individual model at time t, then the residual of the combined output at time t is

$$e_{c,t} = x_t - \hat{x}_{c,t} = \sum_{j=1}^{m} w_j e_{j,t} \tag{3}$$

Accordingly, to obtain the optimal weights, WA minimizes the in-sample sum of squared errors (SSE) of the combined output as follows [44,45]:

$$\min \quad J = \sum_{t=1}^{T} e_{c,t}^2 = \sum_{t=1}^{T} \sum_{j=1}^{m} \sum_{k=1}^{m} w_j w_k e_{j,t} e_{k,t}$$
$$\text{subject to} \quad \begin{cases} \sum_{j=1}^{m} w_j = 1 \\ w_j \geq 0, \quad j = 1, 2, \ldots, m \end{cases} \tag{4}$$

3. High-order Markov chain model

Given a categorical data sequence $\{y_t, t = 1, 2, \ldots, T\}$, where T is the length of this sequence and $y_t \in DOM(A)$ $(1 \leq t \leq T)$. The domain $DOM(A)$ is finite and unordered, which has m elements called categories or states. There is no harm in letting $DOM(A) = \{1, 2, \ldots, m\}$. For modeling this type of sequence, it has been shown that the high-order Markov chain model is a promising approach. However, for a long time its large number of parameters discouraged people from using it directly [46]. Raftery [47] changed this situation through reducing the number of independent parameters. However, his model involved solving a highly nonlinear optimization problem (maximum log-likelihood problem) so that it guaranteed neither convergence nor a global maximum. Ching et al. [46] extended this model to a more general form, for which they also developed an efficient parameter estimation method. Let Y_t denote the state probability distribution (column vector) at time t, i.e., each of its entry is between 0 and 1 and the sum of all entries is equal to 1. In particular, if $y_t = j$ $(t = 1, 2, \ldots, T)$, which means that the system is in state j at time t, the following form is its corresponding state probability distribution [48,49]:

$$Y_t = \left(0, \ldots, 0, \underbrace{1}_{\text{The } j\text{th entry}}, 0, \ldots, 0\right)'_{m \times 1} \tag{5}$$

Thus, there is a vector-valued (multivariate) sequence $\{Y_t, t = 1, 2, \ldots, T\}$ which is a sequence of state probability distributions and equivalent to the original categorical data sequence $\{y_t, t = 1, 2, \ldots, T\}$, and we call it the state probability distribution sequence. Define the ith-step transition probability from state k at time $t-i+1$ to state j at time $t+1$ is

$$q_{jk}^{(i)} = P(y_{t+1} = j | y_{t-i+1} = k), \quad j, k = 1, 2, \ldots, m, \quad i = 1, 2, \ldots, n \tag{6}$$

Let $q_j^{(i)} = (q_{j1}^{(i)}, q_{j2}^{(i)}, \ldots, q_{jm}^{(i)})$, then the conventional Markov chain model clearly can be written as

$$P(y_{t+1} = j | Y_t) = q_j^{(1)} Y_t, \quad j = 1, 2, \ldots, m \tag{7}$$

Specifically, if the kth entry of Y_t is one, $P(y_{t+1} = j | Y_t)$ is just equal to $q_{jk}^{(1)}$. This shows that the state probability distribution at time $t+1$ is only conditionally dependent on the state of the sequence in the one preceding epoch, i.e. Y_t.

The nth-order Markov chain model is more general. The conditional probability of observing $y_{t+1} = j$ given the past is a linear combination of contributions from each of Y_t, …, Y_{t-n+1} as follows [47–49]:

$$P(y_{t+1} = j | Y_t, Y_{t-1}, \ldots, Y_{t-n+1}) = \sum_{i=1}^{n} \lambda_i q_j^{(i)} Y_{t-i+1} \tag{8}$$

where λ_i is non-negative and

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{9}$$

Accordingly, the state probability distribution at time $t+1$ can be predicted by

$$\hat{Y}_{t+1} = \begin{pmatrix} P(y_{t+1} = 1 | Y_t, Y_{t-1}, \ldots, Y_{t-n+1}) \\ P(y_{t+1} = 2 | Y_t, Y_{t-1}, \ldots, Y_{t-n+1}) \\ \vdots \\ P(y_{t+1} = m | Y_t, Y_{t-1}, \ldots, Y_{t-n+1}) \end{pmatrix} = \sum_{i=1}^{n} \lambda_i Q_i Y_{t-i+1} \tag{10}$$

where $Q_i = [q_{jk}^{(i)}]_{m \times m}$, called the ith-step transition probability matrix, which is a non-negative matrix with column sums equal to 1 ($\sum_{j=1}^{m} q_{jk}^{(i)} = 1$). Here, Q_i and λ_i can be estimated according to the following two subsections.

3.1. Estimating Q_i

For two values in $\{y_t\}$ which are i epochs apart, if the former and the latter ones are equal to k and j respectively, we say that there is one ith-step transition from state k to state j. Let $f_{jk}^{(i)}$ denote the ith-step transition frequency from state k to state j, then it should be naturally calculated by counting the number of those ith-step transitions. Furthermore, the ith-step transition probabilities can be estimated by [46]

$$\hat{q}_{jk}^{(i)} = \begin{cases} \dfrac{f_{jk}^{(i)}}{\sum_{j=1}^{m} f_{jk}^{(i)}}, & \sum_{j=1}^{m} f_{jk}^{(i)} \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Accordingly, the estimation of Q_i is

$$\hat{Q}_i = \begin{bmatrix} \hat{q}_{11}^{(i)} & \cdots & \cdots & \hat{q}_{1m}^{(i)} \\ \hat{q}_{21}^{(i)} & \cdots & \cdots & \hat{q}_{2m}^{(i)} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{q}_{m1}^{(i)} & \cdots & \cdots & \hat{q}_{mm}^{(i)} \end{bmatrix} \tag{12}$$

3.2. Estimating λ_i

Let $\overline{Y}$ denote the stationary distribution to which Y_t will converge when $t \to \infty$. It can be estimated by computing the proportion of the occurrence of each state in the sequence as follows [46]:

$$[\hat{Y}]_k = \frac{f_k}{\sum_{j=1}^{m} f_j}, \quad k = 1, 2, \ldots, m \tag{13}$$

where $\hat{Y}$ is the estimation of $\overline{Y}$, $[\cdot]_k$ denotes the kth entry of the vector and f_j is the occurrence frequency of state j in $\{y_t\}$. Then one would expect

$$\sum_{i=1}^{n} \lambda_i \hat{Q}_i \hat{Y} \approx \hat{Y} \tag{14}$$

This supplies one way to estimate λ_i ($i = 1, 2, \ldots, n$). Specifically, solving the following min–max optimization problem can obtain their estimations:

$$\begin{aligned} &\min_{\lambda} \max_{k} \left| \left[\sum_{i=1}^{n} \lambda_i \hat{Q}_i \hat{Y} - \hat{Y} \right]_k \right| \\ &\text{subject to} \begin{cases} \sum_{i=1}^{n} \lambda_i = 1 \\ \lambda_i \geq 0, \quad i = 1, 2, \ldots, n \end{cases} \end{aligned} \tag{15}$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)'$. Clearly, the above optimization problem can be transformed to the following linear programming problem:

$$\begin{aligned} &\min_{\lambda} \quad \mu \\ &\text{subject to} \begin{cases} \begin{bmatrix} C & I \\ -C & I \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} \geq \begin{bmatrix} \hat{Y} \\ -\hat{Y} \end{bmatrix} \\ \mu \geq 0 \\ \sum_{i=1}^{n} \lambda_i = 1 \quad \text{and} \quad \lambda_i \geq 0, \ i = 1, 2, \ldots, n \end{cases} \end{aligned} \tag{16}$$

where $C = [\hat{Q}_1 \hat{Y} | \hat{Q}_2 \hat{Y} | \cdots | \hat{Q}_n \hat{Y}]_{m \times n}$, $I = (1, 1, \ldots, 1)_{m \times 1}$ and μ is a scalar quantity.

4. The proposed time-varying-weight combining method

As discussed earlier, the WA method can often attain more accurate forecasts than the SA method through estimating constant combining weights to reflect the different proportional contribution of each individual model. However, properties of the forecast error of each model may vary over time, and even the accuracy of some excellent models is sometimes undesirable. In this case, constant weights could not reflect the validity of each forecasting model well, and WA may perform poorly [42,50]. Figlewski and Urich [51] and Kang [52] even pointed out that SA is often preferable when the estimated weights are unstable [42]. For overcoming this instability problem, the combining weights should be allowed to change over time. Accordingly, the combining model (1) could be transformed to the following formula:

$$\hat{x}_{c,t} = \sum_{j=1}^{m} w_{j,t} \hat{x}_{j,t}, \quad t = 1, 2, \ldots \tag{17}$$

where $w_{j,t}$ is the weight on the jth individual model at time t. According to this time-varying formula, motivated by the aforementioned high-order Markov chain model, we proposed a time-varying-weight combining method in this section. The name of this model is shortened as HM-TWA (High-order Markov chain based Time-varying Weighted Average).

As we know, the individual forecasts are often one-step-ahead. On this basis, combining models can also perform one-step-ahead prediction. Suppose we have actual observations $\{x_t, t=1,2,\ldots,T\}$ and their one-step-ahead individual forecasts $\{\hat{x}_{j,t}, t=1,2,\ldots,T,T+1\}$ $(j=1,2,\ldots,m)$ computed at time $t-1$. HM-TWA can acquire the combined one-step-ahead forecast at time $T+1$ through the following main algorithmic steps:

1. Calculate the in-sample time-varying combining weight vector sequence $\{W_t, t=1,2,\ldots,T\}$ (see Section 4.1), where W_t is a column vector $(w_{1,t}, w_{2,t},\ldots,w_{m,t})'$ called combining weight vector.
2. Predict the out-of-sample combining weight vector $\hat{W}_{T+1}$ using the high-order Markov chain model (see Section 4.2). Here, two key points need to be noted, which are modified from the conventional high-order Markov chain model introduced in Section 3:
   (a) The ith-step transition probability matrix Q_i is estimated according to Section 4.2.1.
   (b) The coefficient λ_i $(i=1,2,\ldots,n)$ is estimated according to Section 4.2.2.
3. Obtain the combined forecast at time $T+1$ as $\hat{x}_{c,T+1}=\sum_{j=1}^{m}\hat{w}_{j,T+1}\hat{x}_{j,T+1}$ according to (17), where $\hat{w}_{j,T+1}$ is the jth entry of $\hat{W}_{T+1}$.

Besides, to perform multi-step-ahead forecasting, we have designed a reasonable scheme for both WA and HM-TWA in Section 4.3.

4.1. Calculating the in-sample combining weights

In-sample combining weights at time t $(t=1,2,\ldots,T)$ can be calculated by addressing the following quadratic programming problem:

$$\min \quad J_t = e_{c,t}^2 = \sum_{j=1}^{m}\sum_{k=1}^{m} w_{j,t}w_{k,t}e_{j,t}e_{k,t}$$
$$\text{subject to} \quad \begin{cases} \sum_{j=1}^{m} w_{j,t}=1 \\ w_{j,t}\ge 0, \quad j=1,2,\ldots,m \end{cases} \tag{18}$$

Let

$$W_t=(w_{1,t},w_{2,t},\ldots,w_{m,t})'$$
$$E_t=(e_{1,t},e_{2,t},\ldots,e_{m,t})'$$
$$R=(1,1,\ldots,1)_{m\times 1}$$

Then (18) can be rewritten in the following matrix form:

$$\min \quad J_t = W_t'E_tE_t'W_t$$
$$\text{subject to} \quad \begin{cases} R'W_t=1 \\ W_t\ge 0 \end{cases} \tag{19}$$

In this case, when a new observation along with its m individual forecasts is available, only a simple computation is needed instead of recalculating all of the in-sample weights. This is the reason why the optimization problem "$\min J=\sum_{t=1}^{T}W_t'E_tE_t'W_t$" is not used to replace the above one.

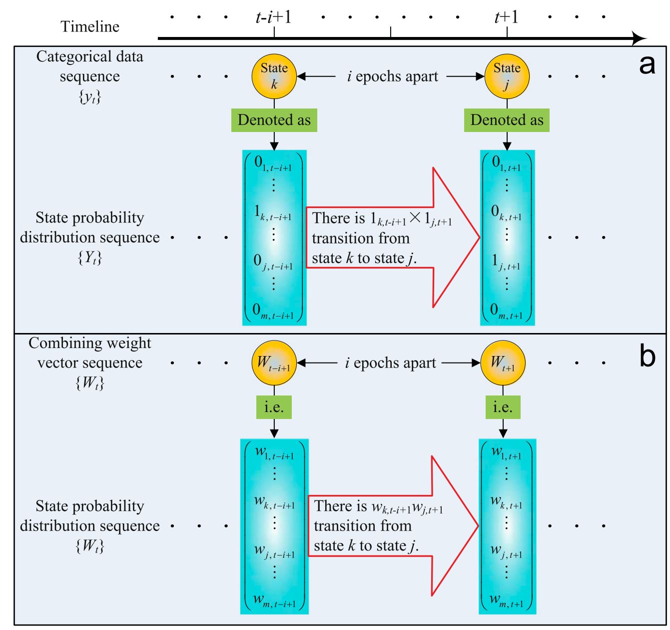

Fig. 1. Cases of ith-step transition from state k to state j. (a) For categorical data sequence, there is 1 transition, so $f_{jk}^{(i)}$ should be increased by 1. (b) For time-varying weight vector sequence, there is $w_{k,t-i+1}w_{j,t+1}$ (≤ 1) transition[1], so $f_{jk}^{(i)}$ should be increased by $w_{k,t-i+1}w_{j,t+1}$.

4.2. Extrapolating the time-varying weights

Clearly, for each epoch t $(t=1,2,\ldots,T)$, the above obtained combining weight vector $W_t=(w_{1,t},w_{2,t},\ldots,w_{m,t})'$ describes the possession ratios of each individual forecast in the optimal combined output $\hat{x}_{c,t}$. If define Markov state j is that $\hat{x}_{c,t}$ is equal to $\hat{x}_{j,t}$, $w_{j,t}$ can be naturally regarded as the probability with which the system is in state j, and $\hat{x}_{c,t}$ is the mathematical expectation of m individual forecasts $\hat{x}_{j,t}$, $j=1,2,\ldots,m$.

Accordingly, the combining weight vector W_t is a state probability distribution the sum of whose entries is one, and the sequence of weight vectors $\{W_t, t=1,2,\ldots,T\}$, called the time-varying combining weight vector sequence, can be regarded as a state probability distribution sequence. Since, in Section 3, what is dealt with by the high-order Markov chain model is the state probability distribution sequence $\{Y_t\}$ gained from the categorical data sequence, $\{W_t\}$ can also be extrapolated to predict the out-of-sample weights by this model. Like the formula (10), the weights at time $t+1$ is predicted by

$$\hat{W}_{t+1}=\sum_{i=1}^{n}\lambda_i Q_i W_{t-i+1} \tag{20}$$

where the estimating methods of Q_i and λ_i differ from those for the categorical data sequence.

To estimate Q_i and λ_i for $\{W_t\}$, the number of occurrence of states at each epoch t should be defined in advance. Let us review the case of the categorical data sequence at first. As shown in (5), the state probability distribution Y_t that is equivalent to the categorical data $y_t=j$ is degenerate, i.e., the probability of state j is 1 and the others are 0, which means that the system is in state j with probability 1.

[1] Note that it has one and only one transition between two data which are i epochs apart. For the categorical data sequence, besides the transition from state k to state j, there are 0 other transitions, so the sum of the numbers of various transitions is 1. However, for time-varying weight vector sequence, there are additional transitions such as $w_{1,t-i+1}w_{j,t+1}$ transition from state 1 to state j and $w_{m,t-i+1}w_{1,t+1}$ transition from state m to state 1, but the sum of the numbers of these transitions is also 1 since $\sum_{k=1}^{m}\sum_{j=1}^{m}w_{k,t-i+1}w_{j,t+1}=1$.

Thus there is one occurrence of state j and no occurrence of other states at time t, which accords the fact of the original categorical data.

However, W_t is not necessarily degenerate, and each of its entries represents the probability with which the system is in one state. Since the system is in state j ($j=1,2,\ldots,m$) with probability $w_{j,t}$, we consider that there is $w_{j,t}$ (≤ 1) occurrence of state j at time t. Clearly, if W_t obeys degenerate distribution and its jth entry is 1, there is one occurrence of state j at time t, which is the situation for the above categorical data.

Following this explanation with regard to occurrences of the states, Q_i and λ_i for the weight vector sequence $\{W_t\}$ are estimated below through modifying and generalizing the methods of Sections 3.1 and 3.2 respectively.

4.2.1. Modified estimation of Q_i

In fact, the calculating method of the ith-step transition frequency $f_{jk}^{(i)}$ for categorical data sequence $\{y_t\}$ introduced in Sections 3.1 can be described by the following process. Let $f_{jk}^{(i)}=0$ at first and then, for $t=i,i+1,\ldots,T-1$, repeat this procedure: if $y_{t-i+1}=k$ and $y_{t+1}=j$ which are i epochs apart, i.e., both of the kth entry of Y_{t-i+1} and the jth entry of Y_{t+1} are 1 (represents there is one occurrence of state k at time $t-i+1$ and one occurrence of state j at time $t+1$) and the other entries are 0 (represents there is no occurrence of other states at both time $t-i+1$ and time $t+1$), then we say that there is $1_{k,t-i+1}\times 1_{j,t+1}$ ($=1$) ith-step transition from state k to state j, so $f_{jk}^{(i)}$ should be increased by 1. This is the case illustrated in Fig. 1(a) in which $b_{k,t}$ represents the kth entry of Y_t where $b=\{0,1\}$.

Likewise, for the time-varying combining weight vector sequence $\{W_t\}$, see Fig. 1(b), there is $w_{k,t-i+1}$ occurrence of state k at time $t-i+1$ and $w_{j,t+1}$ occurrence of state j at time $t+1$, thus there is $w_{k,t-i+1}w_{j,t+1}$ (≤ 1) ith-step transition from state k to state j for each given t ($t=i,i+1,\ldots,T-1$). Thus the ith-step transition frequency from state k to state j can be calculated by summing up the number of these ith-step transitions:

$$f_{jk}^{(i)}=\sum_{t=i}^{T-1} w_{k,t-i+1}w_{j,t+1} \tag{21}$$

Following this up, we can estimate $\hat{Q}_i$ according to (11) and (12).

4.2.2. Modified estimation of λ_i

Since there is $w_{j,t}$ occurrence of state j ($j=1,2,\ldots,m$) at time t, the occurrence frequency of state j is

$$f_j=\sum_{t=1}^{T} w_{j,t} \tag{22}$$

Thus, according to (13), the estimation of the stationary distribution $\overline{Y}$ is

$$\hat{Y}=\frac{1}{T}\sum_{t=1}^{T} W_t \tag{23}$$

Then the estimation of λ_i can be gotten according to (15) and (16).

4.3. A multi-step-ahead forecasting scheme for combining methods

For the combining methods, it is a significant problem that how to employ actual observations $\{x_t, t=1,2,\ldots,T\}$ to acquire the forecasts during the out-of-sample period $[T+1,T+H]$. This is related to a multi-step-ahead forecasting problem. As we all know, to perform prediction with time series, the properties of the sample data used to estimate predictive models should be the same as those of the required forecasts. According to this requirement, a reasonable multi-step-ahead forecasting scheme for combining methods is proposed below.

Let $\hat{x}_{j,t}^{h}$ denote an h-step-ahead forecast of x_t, that is, the forecast computed by the jth individual model using the observations up to epoch $t-h$. Then Fig. 2 shows an instance of such scheme with three individual models. In this case, to obtain the h-step-ahead ($h=1,2,\ldots,H$) combined forecast $\hat{x}_{c,T+h}^{h}$, the h-step-ahead individual forecasts are gotten hold of in advance, i.e. $\{\hat{x}_{j,t}^{h}, t=1,2,\ldots,T,\ldots,T+h\}$ ($j=1,2,3$). Clearly, there are three h-step-ahead forecast series whose lengths are $T+h$, and the forecasts at epochs $t=T+1,\ldots,T+h$ are all computed before/at time T. Depending on the individual forecasts during the in-sample period $[1,T]$ and the original series, the in-sample weights are calculated according to (4) for WA and (19) for HM-TWA. Then based on these weights, the weights at time $T+h$ are just the above in-sample ones for WA or can be extrapolated by the high-order Markov chain model for HM-TWA. Finally, the h-step-ahead combined forecast is obtained.

In particular, the way that HM-TWA obtains combined forecasts during the out-of-sample period $[T+1,T+H]$ is shown in the following steps, where variables with the superscript h are related to the h-step-ahead forecast series:

1. Let $h=1$.
2. Calculate the in-sample time-varying combining weight vector sequence $\{W_t^h, t=1,2,\ldots,T\}$ according to (19).
3. Estimate the ith-step transition probability matrix Q_i^h and coefficient λ_i^h ($i=1,2,\ldots,n$) according to Sections 4.2.1 and 4.2.2 respectively.
4. Predict the combining weights W_t^h at time $t=T+h$ (i.e., $\hat{W}_{T+h}^{h}$) according to (20) along the following iterated multi-step-ahead forecasting process successively until $\hat{W}_{T+h}^{h}$ has been attained:

$$\hat{W}_{T+1}^{h}=\hat{\lambda}_1^h\hat{Q}_1^hW_T^h+\hat{\lambda}_2^h\hat{Q}_2^hW_{T-1}^h+\cdots+\hat{\lambda}_n^h\hat{Q}_n^hW_{T-n+1}^h$$
$$\hat{W}_{t+1}^{h}=\hat{\lambda}_1^h\hat{Q}_1^h\hat{W}_t^h+\cdots+\hat{\lambda}_{t-T}^h\hat{Q}_{t-T}^h\hat{W}_{T+1}^h+\hat{\lambda}_{t-T+1}^h\hat{Q}_{t-T+1}^h W_T^h+\cdots+\hat{\lambda}_n^h\hat{Q}_n^hW_{t-n+1}^h,\quad T<t<T+n$$
$$\hat{W}_{t+1}^{h}=\hat{\lambda}_1^h\hat{Q}_1^h\hat{W}_t^h+\hat{\lambda}_2^h\hat{Q}_2^h\hat{W}_{t-1}^h+\cdots+\hat{\lambda}_n^h\hat{Q}_n^h\hat{W}_{t-n+1}^h,\quad t\ge T+n$$

5. Obtain the combined forecast at time $T+h$ as $\hat{x}_{c,T+h}^{h}=\sum_{j=1}^{m}\hat{w}_{j,T+h}^{h}\hat{x}_{j,T+h}^{h}$ according to (17), where $\hat{w}_{j,T+h}^{h}$ is the jth entry of $\hat{W}_{T+h}^{h}$.
6. If $h=H$, stop; otherwise, let $h=h+1$ and return to Step 2.

Through above steps, forecasts of HM-TWA at epochs $T+1$, $T+2$, …, $T+H$ can be procured. They are $\{\hat{x}_{c,T+1}^{1},\hat{x}_{c,T+2}^{2},\ldots,\hat{x}_{c,T+H}^{H}\}$. Here, $\hat{x}_{c,T+h}^{h}$ is the h-step-ahead combined forecast gotten on the basis of the h-step-ahead forecasts computed by the m individual models. For viewing convenience, the method for determining the h-step-ahead combined forecast of HM-TWA is shown in Fig. 3 in detail where $F_j(\cdot)$ represents the jth individual model. Note that this multi-step-ahead forecasting scheme is also used by the WA method in Section 5 for impartial forecasting performance comparison.

5. Forecasting the monthly electricity consumption of China

5.1. Individual models

Since the monthly electricity consumption data series usually exhibits strong month of the year seasonality, these three seasonal data forecasting methods are used as the individual models for our combination forecasting: predetermined seasonal term method

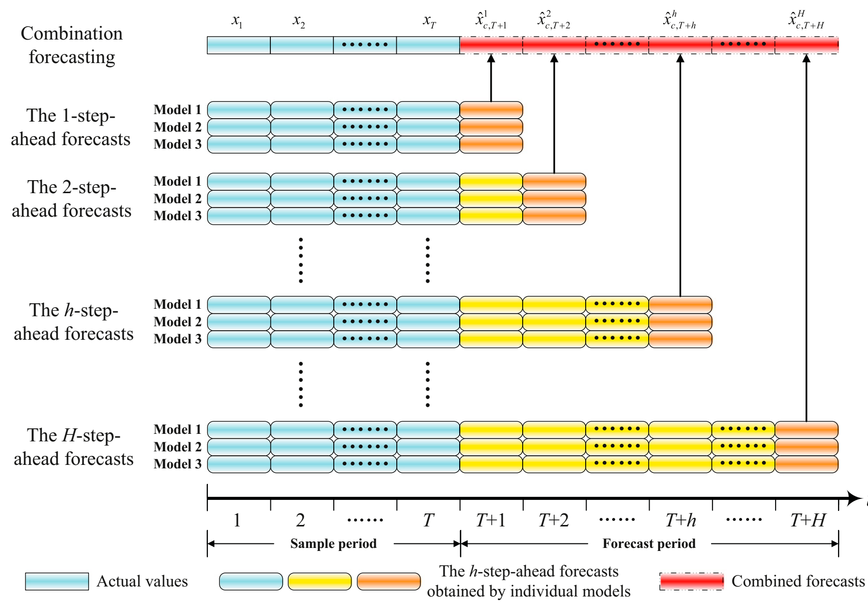

Fig. 2. The illustration for the multi-step-ahead forecasting scheme.

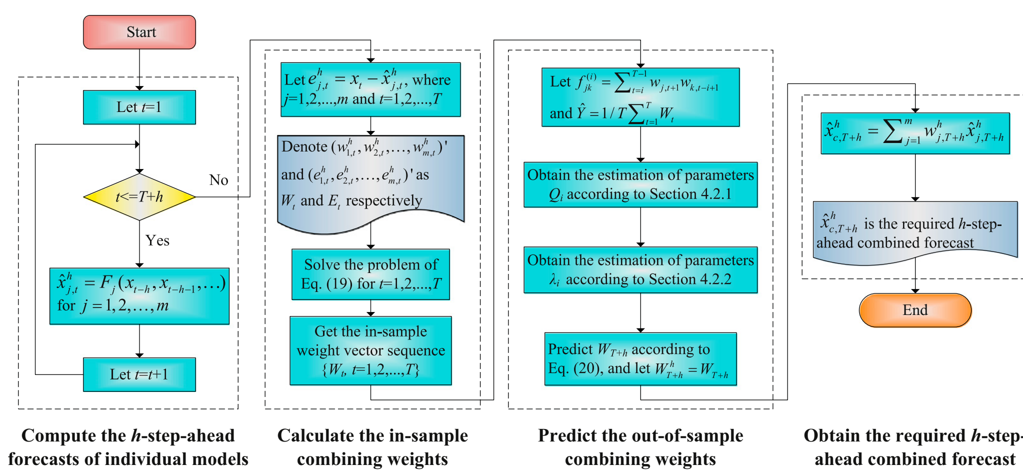

Fig. 3. The flowchart for obtaining the h-step-ahead combined forecast for HM-TWA.

(PSTM), predetermined trend term method (PTTM) and Holt–Winters method (HWM). Given a time series $\{x_t, t=1,2,\ldots,N\}$ which shows trend and seasonality (periodicity) and has m cycles whose length are s, how these three forecasting models obtain the h-step-ahead forecast $\hat{x}_{N+h}$ is shown in Appendices A–C.

5.2. Data collection and problem description

CEInet Statistics Database has released the monthly generated electricity of China up to May 2012. Considering the production and instant use property of electricity, these data can be regarded as the electricity consumption. However, these data cannot be used directly due to the Chinese New Year which always lasts a few days in either January or February. It is so important for the Chinese people that almost all of the companies and factories stop working. Accordingly, electricity consumption is abnormal sometimes in January and sometimes in February.

To avoid this problem, the average value of January and February is regarded as the observation of a new month "1&2" in each year, i.e., there are 11 monthly values every year and the cycle length is 11. This study collects electricity consumption data from the beginning of 1&2 2003 to the end of May 2012 in order to remain relevant to the current situation of electricity development. These original data are shown in Fig. 4.

In this section, these data will be employed to evaluate the effectiveness of the proposed HM-TWA method both in one-step-ahead and multi-step-ahead situations for the last 15 months (from 1&2 2011 to May 2012). Furthermore, HM-TWA will be compared with some existing models in the last year (from June 2011 to May 2012) to ensure its applicability in annual electricity planning and policy-making.

5.3. Statistics measures of forecasting performance

In this paper, these three common criteria were used to evaluate the forecasting performance: root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE)

$$\mathrm{RMSE}=\sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t-\hat{x}_t)^2} \tag{24}$$

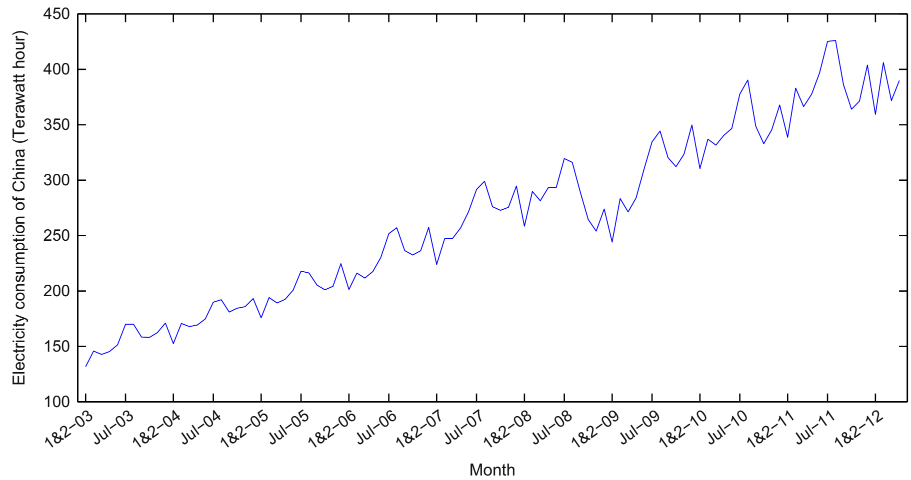

Fig. 4. The monthly electricity consumption of China from 1&2 2003 to May 2012.

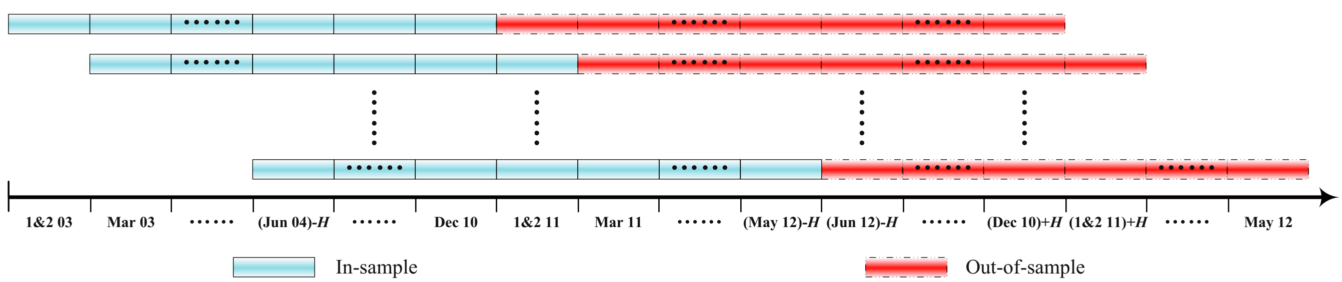

Fig. 5. Evaluation for the forecasting performance at lead times up to H-step-ahead.

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|x_t - \hat{x}_t| \tag{25}$$

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \times 100\% \tag{26}$$

where $\hat{x}_t$ is the forecast of x_t. Clearly, as three kinds of deviation measure between the forecasts and the actual values, their smaller values represent the higher accuracy.

5.4. *Evaluation of the out-of-sample forecasting performance*

To evaluate the out-of-sample forecasting performance of HM-TWA for lead times up to H-step-ahead ($H = 1, 2, \ldots, 11$), 88 months (eight years) are elected to be used as the in-sample period and H months for the out-of-sample forecast period. As shown in Fig. 5 which illustrates this performance evaluation process, firstly, monthly observations from 1&2 2003 through December 2010 are used as in-sample data, then the combination forecasting for the horizons up to H-step-ahead can be made for the forecast period from 1&2 2011 through (December 2010)+H that is the Hth month after December 2010. Secondly, 1&2 2003 is dropped and 1&2 2011 is added in the in-sample period. Then the in-sample period is from March 2003 through 1&2 2011 and the out-of-sample forecast period is from March 2011 through (1&2 2011)+H. The last in-sample period is from (June 2004)-H through (May 2012)-H, and the corresponding out-of-sample forecast period is from (June 2012)-H through May 2012.

Two details should be noted: (1) as shown in Fig. 2, for each H=1, 2, …, 11, obtaining the combined forecasts at epochs from $T+1$ to $T+H$ depends on the 1-step-ahead through the H-step-ahead individual forecasts. Here, we use eight-year CEInet Statistics Database electricity consumption data to predict these individual forecasts. Specifically, the h-step-ahead individual forecasts ($h = 1, 2, \ldots, H$) at time t (i.e., $\hat{x}_{j,t}^{h}$) are acquired using the data at epochs from $(t-h-87)$ through $(t-h)$. (2) Considering the month of the year seasonality of the consumption data, the order of the high-order Markov chain model is set to $n = 11$ for the moment.

For prediction up to each step ahead, since there are several experiments in the performance evaluation process, we comprehensively measure the performance of each step ahead forecasting by "Avg RMSE (MAE, MAPE)" which represents the average of RMSEs (MAEs, MAPEs) brought from these experiments. Note that Avg RMSE is the same as Avg MAE for one-step-ahead forecasting. These values of each models are all shown in Table 1.

From Table 1, three cases can be found:

- For all of the 11 forecast horizons: WA and HM-TWA both outperform each individual model and SA.
- For horizons up to (1–7,11)-step-ahead[2]: HM-TWA performs better than WA.
- For horizons up to (8–10)-step-ahead: the three errors of HM-TWA are generally higher than WA. However, their differences are small. Take Avg MAPE for example, the difference between WA and HM-TWA is only reflected in the second or third decimal place, and HM-TWA is even slightly better than WA for horizons up to 10-step-ahead. Thus the performance of HM-TWA is comparable to that of WA.

[2] We denote "(1–7,11)-step-ahead" as "1-step-ahead" through "7-step-ahead" along with "11-step-ahead" in this paper. Similarly, "(8–10)-step-ahead" represents "8-step-ahead", "9-step-ahead" and "10-step-ahead".

Table 1
The out-of-sample forecasting performance comparisons for horizons up to H-step-ahead ($H = 1, 2, \ldots, 11$).

Horizons up to	PSTM	PTTM	HWM	SA	WA	HM-TWA
One-step-ahead						
Avg RMSE	29.322	12.179	10.620	14.017	10.511	9.180
Avg MAE	29.322	12.179	10.620	14.017	10.511	9.180
Avg MAPE (%)	7.597	3.158	2.818	3.633	2.789	2.420
Two-step-ahead						
Avg RMSE	28.757	13.117	11.751	14.701	11.590	10.798
Avg MAE	28.142	12.734	11.053	14.309	10.956	10.313
Avg MAPE (%)	7.250	3.282	2.921	3.686	2.893	2.703
Three-step-ahead						
Avg RMSE	28.070	13.294	11.963	14.633	11.755	10.958
Avg MAE	26.932	12.718	10.919	14.082	10.819	10.183
Avg MAPE (%)	6.916	3.263	2.878	3.614	2.848	2.660
Four-step-ahead						
Avg RMSE	26.881	13.405	12.403	14.425	12.126	11.327
Avg MAE	25.456	12.594	11.188	13.690	10.990	10.435
Avg MAPE (%)	6.521	3.216	2.941	3.502	2.884	2.718
Five-step-ahead						
Avg RMSE	25.840	13.591	12.751	14.410	12.367	11.639
Avg MAE	24.044	12.562	11.517	13.330	11.169	10.639
Avg MAPE (%)	6.135	3.191	3.015	3.395	2.918	2.763
Six-step-ahead						
Avg RMSE	24.804	13.853	12.980	14.480	12.478	11.933
Avg MAE	22.510	12.619	11.778	12.981	11.305	10.768
Avg MAPE (%)	5.729	3.194	3.076	3.297	2.946	2.794
Seven-step-ahead						
Avg RMSE	24.090	14.208	13.033	14.719	12.420	12.350
Avg MAE	21.209	12.762	11.836	12.841	11.258	11.148
Avg MAPE (%)	5.390	3.227	3.088	3.260	2.930	2.895
Eight-step-ahead						
Avg RMSE	23.865	14.507	12.751	15.017	12.081	12.225
Avg MAE	20.236	12.860	11.477	12.799	10.824	10.934
Avg MAPE (%)	5.138	3.250	2.992	3.251	2.815	2.837
Nine-step-ahead						
Avg RMSE	23.918	14.993	12.405	15.401	11.811	11.961
Avg MAE	19.609	13.094	11.003	12.887	10.411	10.462
Avg MAPE (%)	4.975	3.306	2.866	3.273	2.704	2.709
Ten-step-ahead						
Avg RMSE	23.950	15.614	12.050	15.781	11.669	11.772
Avg MAE	19.224	13.532	10.547	13.033	10.179	10.206
Avg MAPE (%)	4.887	3.418	2.742	3.313	2.639	2.635
Eleven-step-ahead						
Avg RMSE	23.928	16.043	11.810	16.090	11.749	11.711
Avg MAE	19.009	13.942	10.193	13.174	10.131	10.124
Avg MAPE (%)	4.857	3.531	2.646	3.360	2.622	2.607

In summary, among these six investigated models (PSTM, PTTM, HWM, SA, WA and HM-TWA), HM-TWA has the best performance in eight evaluations, and is close to having the best performance for the rest three evaluations. Thus the proposed HM-TWA method is generally a more effective method compared with the others.

Here, since there are a large number of estimated λ (i.e. $(\lambda_1, \lambda_2, \ldots, \lambda_{11})'$) vectors, they are not shown in detail but described in brief as follows: the first entries of all λ vectors are equal to 1 approximately; as for the second through eleventh entries of these λ vectors, except one of which reaches 10^{-4} order of magnitude, the order of magnitude is less than 10^{-6} for others and even less than 10^{-8} for most of them. This means the weights generally only depend on the weights in the one preceding epoch, and the forecasting results will be the same when the order of the high-order Markov chain model is set to $n=1$. Accordingly, we can use first-order Markov chain model hereinafter. Note that this is just a model identification process for Chinese electricity consumption and does not mean that the high-order model is defeated by the first-order model because: (1) the first-order model is a special case of the high-order one; (2) we actually even cannot know whether directly using first-order model (i.e. setting $n=1$) from the outset is reasonable without the simulation

Table 2
The one-step-ahead forecasting performance comparisons during the period from June 2011 to May 2012.

Month	Actual value (TW h)	Predicted value (TW h) of										Weight of WA on			Weight of TSSE on			Weight of HM-TWA on		
		BPN	LSSVR	SARIMA	PSTM	PTTM	HWM	SA	WA	TSSE	HM-TWA	PSTM	PTTM	HWM	PSTM	PTTM	HWM	PSTM	PTTM	HWM
June-11	396.82	368.57	356.42	384.08	352.92	371.89	384.11	369.64	384.10	378.10	383.05	0.000000	0.000804	0.999200	0.062549	0.331870	0.605580	0.006166	0.070731	0.923100
July-11	425.15	390.25	374.16	429.41	384.07	405.03	418.21	402.44	418.21	411.95	417.07	0.000000	0.000000	1.000000	0.065609	0.304820	0.629580	0.006166	0.070731	0.923100
August-11	426.04	410.90	389.28	438.57	388.49	412.12	425.92	408.85	425.92	419.60	424.71	0.000000	0.000000	1.000000	0.064196	0.284230	0.651580	0.006166	0.070731	0.923100
September-11	386.06	390.99	355.18	383.64	358.82	382.89	392.42	378.04	392.38	388.04	391.53	0.000000	0.004196	0.995800	0.056758	0.259020	0.684230	0.006166	0.070731	0.923100
October-11	364.04	407.97	364.48	370.91	349.37	372.90	380.97	367.75	380.87	377.03	375.19	0.000000	0.013197	0.986800	0.057059	0.265050	0.677900	0.066870	0.454220	0.478910
November-11	371.30	408.22	361.72	375.89	353.32	376.03	385.25	371.54	385.00	380.48	370.79	0.000000	0.026877	0.973120	0.063805	0.296250	0.639950	0.379950	0.253010	0.367030
December-11	403.81	442.74	370.65	393.60	377.94	400.93	411.24	396.70	410.81	405.56	397.93	0.000000	0.041077	0.958920	0.070956	0.321300	0.607750	0.245740	0.497250	0.257010
1&2-12	359.38	417.30	381.91	375.77	336.87	357.71	365.70	353.43	365.33	361.02	359.50	0.000000	0.046162	0.953840	0.071787	0.327470	0.600750	0.081941	0.480600	0.437460
March-12	405.99	432.99	397.10	402.93	373.79	394.15	403.71	390.55	403.22	398.45	396.13	0.000000	0.050652	0.949350	0.070853	0.328440	0.600710	0.085498	0.525370	0.389140
April-12	371.83	397.95	385.46	389.08	364.28	384.02	393.43	380.58	392.92	388.83	392.25	0.000000	0.054469	0.945530	0.058545	0.307960	0.633490	0.010065	0.094271	0.895660
May-12	389.86	423.30	408.69	381.43	374.00	392.88	402.06	389.64	401.36	396.42	385.81	0.000000	0.075817	0.924180	0.076313	0.380560	0.543120	0.467260	0.342080	0.190660
RMSE		34.329	28.375	10.281	28.327	12.210	11.471	13.583	11.241	10.521	9.422									
MAE		31.590	24.191	8.9783	26.038	9.7574	9.7135	10.592	9.5407	8.8189	7.3321									
MAPE (%)		8.198	6.061	2.323	6.561	2.451	2.547	2.638	2.501	2.263	1.887									

Table 3
The (1–5)-step-ahead individual forecasts during the forecast period.

Month	The 1-step-ahead forecast			The 2-step ahead forecast			The 3-step-ahead forecast			The 4-step-ahead forecast			The 5-step-ahead forecast		
	PSTM	PTTM	HWM	PSTM	PTTM	HWM	PSTM	PTTM	HWM	PSTM	PTTM	HWM	PSTM	PTTM	HWM
June-11	352.92	371.89	384.11	355.82	371.45	382.78	358.68	371.37	381.41	361.33	370.21	378.68	364.23	371.37	376.86
July-11	–	–	–	387.18	403.61	416.55	390.35	403.14	415.00	393.48	403.06	413.41	396.37	401.82	410.26
August-11	–	–	–	–	–	–	394.69	407.15	423.23	397.94	406.69	421.55	401.14	406.61	419.83
September-11	–	–	–	–	–	–	–	–	–	367.67	375.25	389.77	370.70	374.83	388.12
October- 11	–	–	–	–	–	–	–	–	–	–	–	–	361.21	364.98	378.86

test of the 11th-order model. Thus the proposed model in this paper still should be discussed in the high-order framework which gives the proposed model more flexibility and conduces to its generalization in other fields.

5.5. Forecasting the monthly electricity consumption up to one year ahead

Since the effectiveness of HM-TWA has been evaluated, this subsection will compare it with some existing models to ensure its applicability for one-year-ahead electricity consumption forecasting to support decision-making. Here, its forecasts for horizons up to one-step-ahead and eleven-step-ahead from June 2011 through May 2012 are specifically evaluated. Besides the three individual models and the two traditional combining methods, three popular models are also used as benchmark models, including BPN (back-propagation network), [29] LSSVR (least squares support vector regression) [53] and SARIMA (seasonal autoregressive integrated moving average) [9,54]. They are established by the neural network toolbox of Matlab, LS-SVMlab1.5 toolbox and Minitab 15 respectively. Their in-sample periods are also the eighty-eight months before the forecast period. Inputs of both BPN and LSSVR are the past 11 lags (one year) considering the periodicity of the data. The parameter setting method of BPN is the same as that of [29], specifically, BPN uses a $11 \times 23 \times 1$ structure. In addition, through model identification, parameters of SARIMA are $p=d=q=1$, $P=Q=0$ and $D=1$, i.e., the model is $\mathrm{ARIMA}(1,1,1)(0,1,0)_{11}$. Furthermore, the TSSE (time-varying sum of squared error) method [42], an existing time-varying-weight combining method, is also involved to compare with HM-TWA, but it is only for the one-step-ahead forecasting.[3]

The one-step-ahead forecasting results during the period from June 2011 to May 2012 are listed in Table 2. Note the difference between the weights of WA and those of TSSE and HM-TWA: the former ones are the same as the in-sample constant weights which are estimated according to the historical performance during the whole in-sample period, but the latter ones are given by the formula shown in footnote 3 or extrapolated from the in-sample time-varying weights by the high-order Markov chain model.

With the above predetermined parameters for BPN, LSSVR and SARIMA in this paper, Table 2 shows that there are noticeable improvements for HM-TWA compared with BPN, LSSVR, SARIMA, PSTM, PTTM, HWM, SA, WA and TSSE, and the MAPE has been reduced by 76.98%, 68.87%, 18.77%, 71.24%, 23.01%, 25.91%, 28.47%, 24.55% and 16.62% respectively. Precisely, noting all of the weights of WA on PSTM being 0, one can imagine that the historical performance of PSTM during the in-sample period is inferior to the other two individual models. This is because the constant weights of WA can only reflect the overall in-sample performance and neglect the detailed time-varying property of the individual models. However, PSTM can be more accurate than others and redounds to obtaining the more precise combined forecasts at some special time points. This leads to the non-zero weights of TSSE and HM-TWA during the out-of-sample period. These non-zero weights lead PSTM to play a role to obtain more accurate combined forecasts through using its advantage that compensate for the shortage of the other component models, which proves the superiority of time-varying weights.

To perform the forecasting for horizons up to eleven-step-ahead, i.e., procure all of the one-year forecasts directly in May 2011, the individual forecasts should be first gotten as in Fig. 2. Tables 3 and 4 show these forecasts in the forecast period from June 2011 to May 2012. From these tables, the experimental process can be clearly observed. There are three columns of h-step-ahead ($h=1,2,\ldots,11$) individual forecasts. The number of h-step-ahead individual forecasts is just h, and the shaded values are the last forecasts. According to Fig. 2, after the corresponding combining weights for the shaded values have been predicted by the high-order Markov chain model, the h-step-ahead combined forecast is just the inner product of the shaded values and these weights. Table 5 shows these forecasting results. Since the WA also uses the same multi-step-ahead forecasting scheme, there are different weights for each forecasting horizon. This does not mean that the weights are time-varying.

Table 5 shows that the MAPE of HM-TWA is 2.415%. Although this accuracy is lower than the short-term forecasting (one-step-ahead forecasting), it falls within the acceptable level and has different degrees of improvement compared with other eight models. Furthermore, considering the importance of each horizon's stability for multi-step-ahead forecasting, Table 6 lists the minimum and maximum relative errors (R.E.) along with their months of occurrence, for each model during the period from June 2011 through May 2012. By observing the difference between the minimum and maximum relative errors of each model, it can be seen that HM-TWA is the most stable with a minimum difference (4.823%).

6. Conclusions

This paper proposes a novel monthly electricity consumption combining forecasting model called HM-TWA. To overcome the limitation of traditional combining methods that only consider the overall in-sample performance and neglect the detailed time-varying property of the individual models, HM-TWA makes the combining weights change over time using a high-order Markov chain model. This matches the instability, caused by various uncertainties, of China's electricity consumption. In addition, for the sake of guaranteeing the forecast principle that the sample

[3] The time-varying weights of TSSE is determined by $w_{j,t}=(1/\sigma_{j,t}^2)/\sum_{k=1}^{m}(1/\sigma_{k,t}^2)$ where $\sigma_{j,t}^2$ depends on v (equal to 11 in this paper) previous forecast errors, so it cannot perform multi-step-ahead forecasting. For instance, if we have the original data up to epoch T, since we do not know the actual value at epoch $T+1$, the forecast error at epoch $T+1$ is also unknown and the combined forecast at epoch $T+2$ cannot be gotten.

Table 4
The (6–11)-step-ahead individual forecasts during the forecast period.

Month	The 6-step-ahead forecast			The 7-step-ahead forecast			The 8-step-ahead forecast			The 9-step-ahead forecast			The 10-step-ahead forecast			The 11-step-ahead forecast		
	PSTM	PTTM	HWM	PSTM	PTTM	HWM	PSTM	PTTM	HWM	PSTM	PTTM	HWM	PSTM	PTTM	HWM	PSTM	PTTM	HWM
June-11	367.47	370.02	376.52	370.70	369.92	376.69	374.32	370.20	378.24	377.67	369.45	378.42	380.96	366.85	375.83	384.45	364.53	374.66
July-11	399.54	403.05	408.47	403.09	401.61	408.79	406.61	401.51	408.70	410.58	401.80	410.47	414.25	401.00	410.69	417.85	398.22	407.76
August-11	404.10	405.40	416.05	407.34	406.60	413.15	410.97	405.19	413.45	414.58	405.09	413.35	418.65	405.37	415.25	422.38	404.59	415.49
September-11	373.69	374.76	386.30	376.43	373.65	383.23	379.46	374.75	381.05	382.86	373.47	381.35	386.23	373.37	381.27	390.03	373.63	383.11
October-11	364.23	364.58	377.62	367.20	364.51	376.91	369.93	363.44	373.52	372.94	364.50	371.30	376.34	363.26	371.62	379.71	363.17	371.54
November-11	368.65	368.79	385.46	371.75	368.39	384.39	374.81	368.32	382.62	377.63	367.27	379.01	380.74	368.32	376.66	384.24	367.09	377.01
December- 11	–	–	–	397.70	393.71	413.49	401.04	393.30	411.54	404.33	393.23	409.55	407.34	392.13	405.52	410.68	393.22	402.89
1&2–12	–	–	–	–	–	–	357.45	350.26	367.64	360.43	349.90	365.83	363.37	349.84	363.98	366.06	348.88	360.24
March-12	–	–	–	–	–	–	–	–	–	399.95	387.95	407.09	403.24	387.57	404.99	406.49	387.50	402.85
April-12	–	–	–	–	–	–	–	–	–	–	–	–	392.85	377.30	396.82	396.10	376.93	394.67
May-12	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	407.03	387.09	407.26

Table 5
The 11-step-ahead forecasting performance comparisons during the period from June 2011 to May 2012.

Month	Actual value (TW h)	Predicted value (TW h) of									Weight of WA on			Weight of HM-TWA on		
		BPN	LSSVR	SARIMA	PSTM	PTTM	HWM	SA	WA	HM-TWA	PSTM	PTTM	HWM	PSTM	PTTM	HWM
June-11	396.82	368.57	346.31	384.08	352.92	371.89	384.11	369.64	384.10	383.05	0.000000	0.000804	0.999200	0.006166	0.070731	0.923100
July-11	425.15	370.60	363.30	415.34	387.18	403.61	416.55	402.44	416.01	414.63	0.000000	0.041302	0.958700	0.028757	0.082664	0.888580
August-11	426.04	400.73	366.47	428.28	394.69	407.15	423.23	408.36	422.02	420.47	0.000000	0.075620	0.924380	0.049526	0.084120	0.866350
September-11	386.06	401.91	363.66	386.82	367.67	375.25	389.77	377.56	388.23	386.94	0.000000	0.105720	0.894280	0.050272	0.117910	0.831820
October-11	364.04	398.02	358.11	371.25	361.21	364.98	378.86	368.35	377.10	375.95	0.000000	0.126690	0.873310	0.064100	0.127930	0.807970
November-11	371.30	416.96	367.49	383.91	368.65	368.79	385.46	374.30	383.07	382.30	0.000000	0.143450	0.856550	0.073128	0.115860	0.811010
December-11	403.81	403.81	373.24	406.54	397.70	393.71	413.49	401.63	410.90	410.91	0.163870	0.000000	0.836130	0.091521	0.057450	0.851030
1&2–12	359.38	428.08	356.45	377.60	357.45	350.26	367.64	358.45	365.40	364.72	0.219980	0.000000	0.780020	0.177030	0.064486	0.758490
March-12	405.99	450.08	369.53	422.14	399.95	387.95	407.09	398.33	405.11	398.18	0.277610	0.000000	0.722390	0.536380	0.265740	0.197880
April-12	371.83	456.85	366.40	405.69	392.85	377.30	396.82	388.99	395.50	390.61	0.331190	0.000000	0.668810	0.410530	0.234690	0.354780
May- 12	389.86	470.60	374.92	417.03	407.03	387.09	407.26	400.46	407.17	400.12	0.378590	0.000000	0.621410	0.509150	0.348100	0.142750
RMSE		50.736	34.379	16.404	22.400	13.865	12.676	13.941	11.766	10.413						
MAE		43.831	26.763	13.045	17.215	11.374	10.749	11.082	9.8057	9.3588						
MAPE (%)		11.380	6.579	3.401	4.291	2.834	2.808	2.777	2.552	2.415						

Table 6
Forecasting stability comparisons during the period from June 2011 to May 2012.

Model	Minimum R.E.		Maximum R.E.		Difference between min. and max. R.E. (%)
	Month	Value (%)	Month	Value (%)	
BPN	December-11	0.000	April-12	22.865	22.865
LSSVR	1&2–12	0.815	July-11	14.548	13.733
SARIMA	September-11	0.197	April-12	9.106	8.910
PSTM	1&2–12	0.537	June-11	11.063	10.526
PTTM	October-11	0.258	June-11	6.282	6.024
HWM	March-12	0.271	April-12	6.721	6.450
SA	1&2–12	0.259	June-11	6.850	6.591
WA	March-12	0.217	April-12	6.366	6.149
HM-TWA	September-11	0.228	April-12	5.051	4.823

data should have the same properties as the required forecasts, a reasonable multi-step-ahead forecasting strategy is proposed in which the out-of-sample combined forecast for one horizon is predicted by using individual forecasts for the same horizon as the sample data. Out-of-sample tests of forecasting accuracy show the effectiveness of the proposed HM-TWA method to perform forecasting for one through eleven months (i.e., one year) ahead. Furthermore, to ensure its applicability of one-year-ahead forecasting in annual electricity planning and policy-making, the forecasting performance of HM-TWA has been compared with eight models, including three common methods (BPN, LSSVR and SARIMA), the individual models used in this paper (PSTM, PTTM and HWM), two traditional combining models (SA and WA) and one existing time-varying combining method (TSSE). According to the three criteria (RMSE, MAE and MAPE), with the predetermined parameters for BPN, LSSVR and SARIMA in this paper, HM-TWA has different degrees of improvement compared with other models for both one-month-ahead and one-year-ahead forecasting.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No. 71171102). The authors are grateful to Mrs. Jan Pope for her assistance in polishing the writing of this paper, and would like to gratefully acknowledge the editors and reviewers for their valuable suggestions.

Appendix A. Predetermined seasonal term method

Denote $\{x_t, t=1,2,\ldots,N\}$ as $\{x_{11}, x_{12}, \ldots, x_{1s}; x_{21}, x_{22}, \ldots, x_{2s}; \ldots; x_{m1}, x_{m2}, \ldots, x_{ms}\}$. Let $\overline{x}_k = (x_{k1}+x_{k2}+\cdots+x_{ks})/s$ $(k=1,2,\ldots,m)$, then the jth seasonal index is defined as follows [5,55]:

$$I_j = \frac{I_{1j}+I_{2j}+\cdots+I_{mj}}{m}, \quad j=1,2,\ldots,s \tag{A.1}$$

where $I_{kj} = x_{kj}/\overline{x}_k$ $(k=1,2,\ldots,m;\ j=1,2,\ldots,s)$. Then the seasonal effect can be eliminated by

$$x'_{kj} = \frac{x_{kj}}{I_j}, \quad k=1,2,\ldots,m, \ j=1,2,\ldots,s \tag{A.2}$$

The obtained series $\{x'_{11}, x'_{12}, \ldots, x'_{1s};\ x'_{21}, x'_{22}, \ldots, x'_{2s}; \ldots; x'_{m1}, x'_{m2}, \ldots, x'_{ms}\}$ can be rewritten as $x'_t (t=1,2,\ldots,N)$. After computing its h-step-ahead forecast $\hat{x}'_{N+h}$ using linear regression or two-order moving average, the h-step-ahead forecast of the original series x_t can be attained as follows:

$$\hat{x}_{N+h} = \hat{x}'_{N+h} I_{h-s\lfloor \frac{h-1}{s} \rfloor} \tag{A.3}$$

Appendix B. Predetermined trend term method

The predetermined trend term method (PTTM) is similar to the above PFTM. They only differ in the consideration order for seasonal and trend terms. PTTM first uses linear regression or two-order moving average to model the original series x_t and acquires its trend series $\{\overline{x}_t, t=1,2,\ldots,N\}$ and the h-step-ahead forecast $\overline{x}_{N+h}$. The trend series can be used to calculate the seasonal index. Let $p_t = x_t/\overline{x_t}$ and $\overline{I}_j = 1/m(p_j + p_{s+j} + p_{2s+j} + \cdots + p_{(m-1)s+j})$ $(j=1,2,\ldots,s)$, then the seasonal index can be gotten according to the following normalization process [55,56]:

$$I_j = \frac{s\overline{I}_j}{\overline{I}_1+\overline{I}_2+\cdots+\overline{I}_s} \tag{B.1}$$

which guarantees $I_1+I_2+\cdots+I_s = s$. Finally, the h-step-ahead forecast of the original series x_t can be obtained as follows:

$$\hat{x}_{N+h} = \overline{x}_{N+h} I_{h-s\lfloor \frac{h-1}{s} \rfloor} \tag{B.2}$$

Appendix C. Holt–Winters method

The Holt–Winters method (HWM) (also known as the triple exponential smoothing) takes into account both seasonal changes and trends. Its h-step-ahead forecast is given by [57]

$$\hat{x}_{N+h} = (S_N + b_N h) I_{N-s+(h-s\lfloor \frac{h-1}{s} \rfloor)} \tag{C.1}$$

where S_N and b_N are the smoothed observation and trend factor at time N respectively. The subscript $N-s+(h-s\lfloor \frac{h-1}{s} \rfloor)$ of I means forecasts of more than one full season beyond the end of the data will reuse the last season's seasonal indexes. These parameters can be acquired by computing the following three equations at $t=1$ through $t=N$ successively [58]:

$$\begin{aligned} S_t &= \alpha \frac{x_t}{I_{t-s}} + (1-\alpha)(S_{t-1}+b_{t-1}) \\ b_t &= \gamma(S_t - S_{t-1}) + (1-\gamma) b_{t-1} \\ I_t &= \beta \frac{x_t}{S_t} + (1-\beta) I_{t-s} \end{aligned} \tag{C.2}$$

where α, β and γ are constants that are constrained to be 0–1.

References

[1] Önkal D, Sayım KZ, Lawrence M. Wisdom of group forecasts: does role-playing play a role?. OMEGA – The International Journal of Management Science 2012;40(6):693–702.
[2] Boylan JE, Syntetos AA. Forecasting in management science. OMEGA – The International Journal of Management Science 2012;40(6):681.
[3] Thury G, Witt SF. Forecasting industrial production using structural time series models. OMEGA – The International Journal of Management Science 1998;26(6):751–767.
[4] Bacchetti A, Saccani N. Spare parts classification and demand forecasting for stock control: investigating the gap between research and practice. OMEGA – The International Journal of Management Science 2012;40(6):722–737.
[5] An N, Zhao W, Wang J, Shang D, Zhao E. Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. Energy 2013;49(1):279–288.
[6] Li D-C, Chang C-J, Chen C-C, Chen W-C. Forecasting short-term electricity consumption using the adaptive grey-based approach – an Asian case. OMEGA – The International Journal of Management Science 2012;40(6):767–773.
[7] Hahn H, Meyer-Nieberg S, Pickl S. Electric load forecasting methods: tools for decision making. European Journal of Operational Research 2009;199(3):902–907.
[8] Zhou D. The truth of the electricity shortage (in Chinese), stock market trend analysis weekly 21; 2011. ⟨http://www.gsdtfx.com/magazine.php?mod=view&id=13711⟩.

[9] Wang Y, Wang J, Zhao G, Dong Y. Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: a case study of China. Energy Policy 2012;48:284–294.
[10] Mohamed Z, Bodger P. Forecasting electricity consumption in New Zealand using economic and demographic variables. Energy 2005;30(10):1833–1843.
[11] Saab S, Badr E, Nasr G. Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon. Energy 2001;26(1):1–14.
[12] Ohtsuka Y, Oga T, Kakamu K. Forecasting electricity demand in Japan: a Bayesian spatial autoregressive ARMA approach. Computational Statistics and Data Analysis 2010;54(11):2721–2735.
[13] Vilar JM, Cao R, Aneiros G. Forecasting next-day electricity demand and price using nonparametric functional methods. International Journal of Electrical Power and Energy Systems 2012;39(1):48–55.
[14] Abdel-Aal RE, Al-Garni AZ, Al-Nassar YN. Modelling and forecasting monthly electric energy consumption in eastern Saudi Arabia using abductive networks. Energy 1997;22(9):911–921.
[15] Kucukali S, Baris K. Turkey's short-term gross annual electricity demand forecast by fuzzy logic approach. Energy Policy 2010;38(5):2438–2445.
[16] Kandananond K. Forecasting electricity demand in Thailand with an artificial neural network approach. Energies 2011;4(8):1246–1257.
[17] Çunkaş M, Altun AA. Long term electricity demand forecasting in Turkey using artificial neural networks. Energy Sources, Part B: Economics, Planning, and Policy 2010;5(3):279–289.
[18] Kavaklioglu K. Modeling and prediction of Turkey's electricity consumption using support vector regression. Applied Energy 2011;88(1):368–375.
[19] Lee DG, Lee BW, Chang SH. Genetic programming model for long-term forecasting of electric power demand. Electric Power Systems Research 1997;40(1):17–22.
[20] Çunkaş M, Taşkiran U. Turkey's electricity consumption forecasting using genetic programming. Energy Sources, Part B: Economics, Planning, and Policy 2011;6(4):406–416.
[21] Ozturk HK, Ceylan H. Forecasting total and industrial sector electricity demand based on genetic algorithm approach: Turkey case study. International Journal of Energy Research 2005;29(9):829–840.
[22] Azadeh A, Tarverdian S. Integration of genetic algorithm, computer simulation and design of experiments for forecasting electrical energy consumption. Energy Policy 2007;35(10):5229–5241.
[23] Ringwood JV, Bofelli D, Murray FT. Forecasting electricity demand on short medium and long time scales using neural networks. Journal of Intelligent and Robotic Systems 2001;31(1–3):129–147.
[24] Nasr GE, Badr EA, Younes MR. Neural networks in forecasting electrical energy consumption: univariate and multivariate approaches. International Journal of Energy Research 2002;26(1):67–78.
[25] Azadeh A, Ghaderi SF, Sohrabkhani S. A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran. Energy Policy 2008;36(7):2637–2644.
[26] Kavaklioglu K, Ceylan H, Ozturk HK, Canyurt OE. Modeling and prediction of Turkey's electricity consumption using artificial neural networks. Energy Conversion and Management 2009;50(11):2719–2727.
[27] Sözen A, Isikan O, Menlik T, Arcaklioglu E. The forecasting of net electricity consumption of the consumer groups in Turkey. Energy Sources, Part B: Economics, Planning, and Policy 2011;6(1):20–46.
[28] Yu L, Wang S, Lai KK. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Economics 2008;30(5):2623–2635.
[29] Guo Z, Zhao W, Lu H, Wang J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renewable Energy 2012;37(1):241–249.
[30] Huang M, He Y, Cen H. Predictive analysis on electric-power supply and demand in China. Renewable Energy 2007;32(7):1165–1174.
[31] Azadeh A, Ghaderi SF, Sohrabkhani S. Forecasting electrical consumption by integration of neural network, time series and ANOVA. Applied Mathematics and Computation 2007;186(2):1753–1761.
[32] Nguyen HT, Nabney IT. Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. Energy 2010;35(9):3674–3685.
[33] González-Romera E, Jaramillo-Morán MÁ, Carmona-Fernández D. Monthly electric energy demand forecasting based on trend extraction. IEEE Transactions on Power Systems 2006;21(4):1946–1953.
[34] Zhu S, Wang J, Zhao W, Wang J. A seasonal hybrid procedure for electricity demand forecasting in China. Applied Energy 2011;88(11):3807–3815.
[35] Kim Y-O, Jeong D, Ko IH. Combining rainfall-runoff model outputs for improving ensemble streamflow prediction. Journal of Hydrologic Engineering 2006;11(6):578–588.
[36] Chan CK, Kingsman BG, Wong H. Determining when to update the weights in combined forecasts for product demand-an application of the CUSUM technique. European Journal of Operational Research 2004;153(3):757–768.
[37] Moghram I, Rahman S. Analysis and evaluation of five short-term load forecasting techniques. IEEE Transactions on Power Systems 1989;4(4):1484–1491.
[38] Loredo EN. Annual electrical peak load forecasting methods with measures of prediction error. Industrial Engineering, Arizona State University; 2002.
[39] Stock JH, Watson MW. Combination forecasts of output growth in a seven-country data set. Journal of Forecasting 2004;23(6):405–430.
[40] Geweke J, Amisano G. Optimal prediction pools. Journal of Econometrics 2011;164(1):130–141.
[41] Granger CWJ, Ramanathan R. Improved methods of combining forecasts. Journal of Forecasting 1984;3(2):197–204.
[42] Jeong DI, Kim Y-O. Combining single-value streamflow forecasts – a review and guidelines for selecting techniques. Journal of Hydrology 2009;377(3–4):284–299.
[43] Wang J-J, Wang J-Z, Zhang Z-G, Guo S-P. Stock index forecasting based on a hybrid model. OMEGA – The International Journal of Management Science 2012;40(6):758–766.
[44] Chen H-Y. The effectiveness theory of combination forecasting methods and its application. Beijing: Science Press; 2008 [in Chinese].
[45] Wang J, Zhu S, Zhang W, Lu H. Combined modeling for electric load forecasting with adaptive particle swarm optimization. Energy 2010;35(4):1671–1678.
[46] Ching WK, Fung ES, Ng MK. Higher-order Markov chain models for categorical data sequences. Naval Research Logistics 2004;51(4):557–574.
[47] Raftery AE. A model for high-order Markov chains. Journal of the Royal Statistical Society: Series B (Methodology) 1985;47(3):528–539.
[48] Ching W-K, Ng MK, Fung ES. Higher-order multivariate Markov chains and their applications. Linear Algebra and its Applications 2008;428(2–3):492–507.
[49] Wang Z, Gong Z, Zhao W, Zhu W. Higher-order multivariate Markov chains based on particle swarm optimization algorithm for air pollution forecasting. In: Proceedings of the 2009 Asia-Pacific conference on information processing, vol. 1. Shenzhen, China; 2009. p. 42–6.
[50] Zhang J, Li S, Ma C. A research on linear combined forecasting model for instrument precision. In: Proceedings of the 8th international symposium on test and measurement, vol. 1. Chongqing, China; 2009. p. 219–22.
[51] Figlewski S, Urich T. Optimal aggregation of money supply forecasts: accuracy, profitability and market efficiency. Journal of Finance 1983;38(3):695–710.
[52] Kang H. Unstable weights in the combination of forecasts. Management Science 1986;32(6):683–695.
[53] Zhu B, Wei Y. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. OMEGA – The International Journal of Management Science 2013;41(3):517–524.
[54] Trapero JR, Kourentzes N, Fildes R. Impact of information exchange on supplier forecasting performance. OMEGA – The International Journal of Management Science 2012;40(6):738–747.
[55] Niu D, Cao S, Zhao L, Zhang W. Power load forecasting technology and its applications. Beijing: China Electric Power Press; 1998 [in Chinese].
[56] Wang J, Zhu W, Zhang W, Sun D. A trend fixed on firstly and seasonal adjustment model combined with the ε-SVR for short-term forecasting of electricity demand. Energy Policy 2009;37(11):4901–4909.
[57] Rasmussen R. On time series data and optimal parameters. OMEGA – The International Journal of Management Science 2004;32(2):111–120.
[58] Taylor JW, Snyder RD. Forecasting intraday time series with multiple seasonal cycles using parsimonious seasonal exponential smoothing. OMEGA – The International Journal of Management Science 2012;40(6):748–757.