DOI: 10.19783/j.cnki.pspc.190814

基于集合经验模态分解和套索算法的短期风速组合变权预测模型研究

杨 磊[1]，黄元生[1]，张向荣[2]，董玉琳[2]，高 冲[2]

(1. 华北电力大学(北京)，北京 102206；2. 华北电力大学(保定)，河北 保定 071003)

摘要：准确的风速预测对风电场实现平稳出力具有重要意义。为提高短期风速预测精度，提出一种基于集合经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)、套索算法 (Least Absolute Shrinkage and Selection Operator, LASSO)、遗传算法(Genetic Algorithm, GA)、广义回归神经网络(General Regression Neural Network, GRNN)和长短期记忆模型(Long Short-Term Memory, LSTM)的短期风速变权组合预测模型(Variable Weighted Hybrid Model, VWHM)。首先运用集合经验模态分解技术，将原始风速时间序列分解成多个不同的子序列。然后运用套索算法对各个子序列的数据变量进行筛选，提取代表性变量作为预测输入。最后利用 GA 的全局优化能力，对由 GRNN 和 LSTM 构成的组合预测模型的权重系数进行移动样本自适应变权求解，并加权得到最终预测结果。仿真结果表明，所提的变权组合模型比单一模型以及传统组合模型具有更高的预测精度，且在风速预测中具有优越性。

关键词：短期风速预测；集合经验模态分解；套索算法；广义回归神经网络；长短期记忆；遗传算法

Research on short-term wind speed hybrid variable weight prediction model based on ensemble empirical mode decomposition and LASSO algorithm

YANG Lei[1], HUANG Yuansheng[1], ZHANG Xiangrong[2], DONG Yulin[2], GAO Chong[2]

(1. North China Electric Power University, Beijing 102206, China;
2. North China Electric Power University, Baoding 071003, China)

Abstract: Accurate wind speed prediction is significant for wind farm development and utilization of wind energy. In order to improve the prediction accuracy of short-term wind speed, a combined prediction model with variable weight of short-term wind speed based on Ensemble Empirical Mode Decomposition (EEMD), Least Absolute Shrinkage and Selection Operator (LASSO), Genetic Algorithm (GA), General Regression Neural Network (GRNN) and long-term and short-term memory is proposed. First, the ensemble empirical mode decomposition technique is used to decompose the original wind speed time series into multiple sub-sequences. Then, using the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm, the historical data of each sub-sequence are filtered, and representative variables are extracted as prediction inputs. Finally, using the global optimization ability of a GA, the weight coefficients of the combined prediction model composed of GRNN and LSTM are adaptively solved by moving samples, and the final prediction results are obtained by weighting. The simulation results show that the proposed variable weight combination model has higher prediction accuracy than a single model and a traditional combination model, and has superiority in wind speed prediction.

This work is supported by National Natural Science Foundation of China (No. 61973117).

Key words: short-term wind speed prediction; ensemble empirical mode decomposition; least absolute shrinkage and selection operator; generalized regression neural network; long-term and short-term memory; genetic algorithm

0 引言

随着中国经济的快速发展，化石燃料的消耗也急速增加，由此也带来了诸多环境问题。作为目前最具潜力的可再生能源之一，各个国家都在密切关注风能的开发利用。但在利用风能进行发电的过程中，由于受到温度、地形、纬度等因素的影响，风速会呈现出较强的波动性，从而导致风力发电的不

基金项目：国家自然科学基金面上项目资助(61973117)

确定性增加[1-2]。这种不确定性会对电网的电能质量产生巨大影响，甚至威胁电网系统的安全稳定运行[3]。因此，准确的风速预测是大力发展风力发电的必要前提。

目前，国内外学者已经对风速预测进行了大量的研究工作。主要用到的短期风速预测方法有时间序列预测法、持续预测法、卡尔曼滤波法、神经网络预测法等。由于智能算法具有较好的适应性和拟合能力，基于机器学习的风速预测模型成为众多学者的研究关注点。但是，随着风速预测领域研究的深入，对于预测模型的预测精度和预测性能的要求越来越高，单一的预测方法往往难以取得令人满意的预测效果。基于此，将多种预测方法进行组合，形成预测精度更高的复合预测模型的研究得到广泛关注和应用。文献[4]提出了将人工蜂群算法(Artificial Bee Colony Algorithm, ABC)和 BP(Back Propagation, BP)神经网络相结合的风速预测方法。文献[5]提出了基于 EEMD、随机森林(Random Forest, RF)、极端学习机(Extreme Learning Machine, ELM)的风速预测模型。文献[6]提出了基于小波包变换(Wavelet Packet Decomposition, WPD)、多智能体入侵杂草算法(Multi Agent Invasive Weed Optimization, MAIWO)和人工神经网络(Artificial Neural Network, ANN)组成的混合预测方法。文献[7]提出了基于 Adaboost 算法和 GRNN 的风速组合预测方法。通过以上文献可以看出，组合预测模型相对于单一预测模型更具有优势。

在含有风力发电的电力系统中，预测模型的作用在于提前获取未来时刻风力发电的出力情况，从而为电力调度部门提供可靠信息，用以提前安排发电计划，制定电力调度方案。但由于风速变化具有强烈的随机性和波动性，致使风力发电出力稳定性较差，接入电力系统后增大了制定电力调度方案的难度。因此，有必要对风力发电的风速数据进行多步超前预测，时刻监控风速的变化趋势，从而为电力调度人员及时调整调度方案和发电计划提供有效参考[8-9]。

本文提出一种基于数据预处理和特征选择的短期风速变权组合多步预测模型。首先，使用 EEMD 对原始风速序列进行分解；其次，利用 LASSO 对分解得到的序列进行输入变量筛选；再次，分别利用 GRNN 和 LSTM 对分解序列进行预测，并通过遗传算法根据区间表现优化各个模型权重，得到子序列预测结果；最后，将各子列的预测数据进行累加得到最终的风速预测结果。本文的创新点在于通过遗传算法对权重的自适应优化，结合了 GRNN 和 LSTM 各自处理数据的特点，实现针对性的预测。

1　集合经验模态分解

集合经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)是在经验模态分解(Empirical Mode Decomposition, EMD)方法基础上的一种改进[10]。通过将白噪声添加到初始序列并利用 EMD 对原始信号进行提取，最终对数次提取后的子序列采用集总计算，其具体步骤如下：

(1) 加入白噪声 $w(t)$ 到风速时间序列中，得到新的序列 $X(t)=x(t)+w(t)$；

(2) 用 EMD 分解带有白噪声的风速时间序列 $X(t)$，得到各 IMF 分量；

(3) 重复步骤(1)和(2)，每次加入相同幅值的新的白噪声序列，将每次得到的 IMF 取均值作为最终的结果。

由于风速序列变化随机性较强，而利用数据分解技术对序列进行分解可以有效地降低原始数据的波动性，从而提高预测精度。同时，选用 EEMD 对风速序列进行分解，可以很好地解决 EMD 在噪音消除过程中产生的频率混合的弊端[11]。因此，本文使用 EEMD 对风速时间序列 $x(t)$ 做分解以达到降低数据波动性的目的。

2　套索算法

套索算法(LASSO)是一种压缩估计的方法，能够有效地处理复杂的高维数据问题[12]。一个满足经典假设的多元线性回归模型式(1) 、式(2)所示。

$$Y=X\beta+\varepsilon \tag{1}$$

$$\varepsilon_{\mathrm{i}}\sim \mathrm{i.i.d}N(0,\varepsilon^2) \tag{2}$$

表示每个残差相互独立且同属于正态分布。故响应变量如式(3)所示。

$$Y\sim N(X\beta,\varepsilon^2 I) \tag{3}$$

其中 I 为单位阵。

$$\min_{\beta}\frac{1}{2}\|Y-X\beta\|_2^2 \tag{4}$$

通过对式(4)目标函数求极值，得到的普通最小二乘估计结果为

$$\hat{\beta}_{\mathrm{ols}}=X'X^{-1}X'Y \tag{5}$$

假设矩阵中的行和列分别记为 s 和 k，如果 $s<k$，说明样本量小于解释变量的个数。此时 $rand(X'X)\le rank(X)\le s$，其中 $(X'X)$ 是一个 k 行 k 列的方阵。由于 $(X'X)$ 的不存在，普通最小二乘估计无法进行计算。因此 Tibshirani 在式(4)的基础上增加了 L_1 正则项，其目标函数变为

$$\min_{\beta}\frac{1}{2}\|Y-X\beta\|_2^2+\frac{\lambda}{2}\|\beta\|_1 \tag{6}$$

式中：λ 为惩罚参数，与惩罚强度正相关；$\|\beta\|_1$ 为第一范数，满足 $\|\beta\|_1=|\beta_1|+|\beta_2|+...+|\beta_k|$。式(6)中，第一项为损失函数，它衡量了回归模型对数据的拟合效果。第二项是惩罚函数，它将一些不显著系数压缩为 0。通过优化式(6)，可以得到回归系数向量估计 $\hat{\beta}_{\text{lasso}}$，即 LASSO 回归。对于式(6)其等价的约束为

$$\begin{aligned}&\min_{\beta}\frac{1}{2}\|Y-X\beta\|_2^2\\&\text{s.t. }\|\beta\|_1\le C\end{aligned} \tag{7}$$

式中：C 为与 λ 对应的约束参数，与惩罚强度呈负相关。利用式(6)和式(7)可以得到最优解。由于惩罚函数是带绝对值的一阶函数，不能用微积分方法求解，回归方程的求解需要非常复杂的数值计算。Efron 等人[13]提出了最小角度回归算法(LARS)来解决规划问题，利用 LARS 能够有效地求解 LASSO 回归问题。

在建立预测模型的过程中，预测输入变量的选择对预测精度和计算法复杂性均会产生一定影响。由于风速预测常采用历史数据作为预测模型输入，数据量较大，难以人为选出合适变量作为预测模型输入。因此，有必要在进行预测步骤前，利用 LASSO 算法对预测输入变量进行筛选，从而选择出相关性最强的代表性变量作为预测模型输入。

3　广义回归神经网络

广义回归神经网络(Generalized Regression Neural Network, GRNN)是一种非线性的前馈式神经网络[14-17]。其结构包括输入层、模式层、求和层和输出层，图 1 所示是 GRNN 的网络结构图。其中输入变量为 $x_1,x_2,\cdots,x_T$，输出值为 $y_1,y_2,\cdots,y_T$，$T=1,2,\cdots,n$。

输入层和模式层神经元数量与训练样本相等。输入层作用是将输入变量传递给模式层。模式层用于传递的函数如式(8)所示。

$$P_i=\exp[-\frac{(X-X_i)^{\mathrm{T}}(X-X_i)}{2\sigma^2}] \tag{8}$$

式中：X 代表输入变量，第 i 个神经元所述的学习样本由 X_i 表示，$i=1,2,\cdots,n$；平滑因子由 σ 表示。

求和层存在两种方式：对计算模式层各神经元输出的加权求和以及直接加和，分别用以下两式进行表示。

$$S_{Nj}=\sum_{i=1}^{n}h_{ij}P_i=\sum_{i=1}^{n}h_{ij}\exp[-\frac{(X-X_i)^{\mathrm{T}}(X-X_i)}{2\sigma^2}] \tag{9}$$

$$S_D=\sum_{i=1}^{n}P_i=\sum_{i=1}^{n}\exp[-\frac{(X-X_i)^{\mathrm{T}}(X-X_i)}{2\sigma^2}] \tag{10}$$

式中：$j=1,2,\cdots,m$；y_{ij} 代表第 i 个因变量的第 j 个元素。

输出层结果可由式(11)进行表示。

$$y_j=\frac{S_{Nj}}{S_D} \tag{11}$$

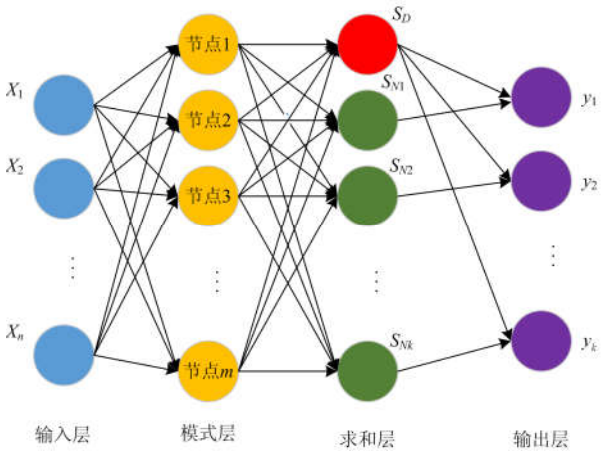

图 1　GRNN 结构图

Fig. 1 GRNN structure diagram

广义回归神经网络是一种径向基神经网络，对于非线性数据具有较好的拟合能力，可以快速捕捉到数据中的信息，在样本数据量较少时也具有较好的预测性能，可以处理波动性较强且不存在明显线性特征的风速序列，所以本文选取 GRNN 作为预测模型之一。

4　长短期记忆网络

长短期记忆网络(Long Short－Term Memory, LSTM)是一种改进的循环神经网络(Recurrent Neural Network, RNN)模型[18-20]。LSTM 的结构包括遗忘门、输入门、输出门。如图 2 所示为 LSTM 结构图。

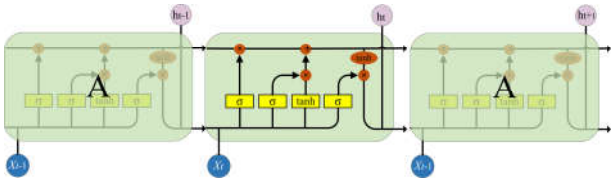

图 2　LSTM 结构图

Fig. 2 LSTM structure diagram

假设 $X=x_1,x_2,\cdots,x_T$ 代表风速序列，则 LSTM 网络的输出为 $y=y_1,y_2,\cdots,y_T$，LSTM 通过遗忘门 f_t 来决定记忆单元状态 c_{t-1} 中的应删除信息，遗忘门的激活状态由激活函数 $\sigma(\cdot)$ 来决定。

$$f_t=\sigma(\boldsymbol{W}_{fx}x_t+\boldsymbol{W}_{fh}h_{t-1}+\boldsymbol{W}_{fc}c_{t-1}+b_f) \tag{12}$$

式(12)的输出 f_t 是对应于最后一个单元状态 c_{t-1} 的 0 到 1 之间的值。

其次，LSTM 存储到新的单元状态 c_t 的新信息由输入门 i_t 来决定，通过下式进行计算。

$$i_t=\sigma(\boldsymbol{W}_{ix}x_t+\boldsymbol{W}_{ih}h_{t-1}+\boldsymbol{W}_{ic}c_{t-1}+b_i) \tag{13}$$

$$u_t=g(\boldsymbol{W}_{cx}x_t+\boldsymbol{W}_{ch}h_{t-1}+b_c) \tag{14}$$

式(14)中：u_t 为添加到新单元状态 c_t 的候选值；$g(\cdot)$ 为激活函数。

旧记忆单元状态 c_{t-1} 更新为新状态 c_t 的过程如式(15)所示。

$$c_t=c_{t-1}f_t+u_ti_t \tag{15}$$

式中：$c_{t-1}f_t$ 决定有多少信息将从 c_{t-1} 中遗忘；u_ti_t 确定新单元状态 c_t 会增添哪些新信息。

最后使用输出门 o_t 计算 h_t 和 y_t 的过程如下式所示。

$$o_t=\sigma(\boldsymbol{W}_{ox}x_t+\boldsymbol{W}_{oh}h_{t-1}+\boldsymbol{W}_{oc}c_{t-1}+b_o) \tag{16}$$

$$h_t=o_t\ell(c_t) \tag{17}$$

式中：激活函数 σ 为 sigmoid 函数；ℓ 和 g 为 tanh 函数；$\boldsymbol{W}_{ix}$、$\boldsymbol{W}_{fx}$、$\boldsymbol{W}_{ox}$、$\boldsymbol{W}_{cx}$ 为连接输入信息 x_t 的权值矩阵；$\boldsymbol{W}_{ih}$、$\boldsymbol{W}_{fh}$、$\boldsymbol{W}_{oh}$、W_{ch} 为连接隐含层输出信号 h_t 的权值矩阵；$\boldsymbol{W}_{ic}$、$\boldsymbol{W}_{fc}$、$\boldsymbol{W}_{oc}$ 为连接神经元激活函数输出矢量 c_t 和门函数的对角矩阵；b_i、b_f、b_o、b_c 为对应的偏置[21]。

长短期记忆神经网络是时间递归神经网络的一种，这种神经网络通过输入门、输出门和遗忘门的设计，解决了递归神经网络中出现的“梯度消失”问题，能够记忆长期的信息，在处理和预测序列中间隔和延迟相对较长的事件时具有较好的性能；由于风速训练数据中存在间隔较长的数据点，所以本文选取 LSTM 作为另一预测模型。

5　变权组合模型的权重确定

尽管运用单个的预测模型进行短期风速预测时能够得到较好的预测结果，但其预测精度难以满足实际的精度需求。本文提出的 EEMD-LASSO-VWHM 模型可以将单个的预测模型组合起来并发挥出单个预测方法的特点，解决了单个模型的局限性问题，从而得到更为精确的预测结果。遗传算法(GA)是模拟自然选择机制和遗传学机理的生物进化过程的计算模型，具有良好的全局搜索能力，可以快速地将解空间中的全体解搜索出来，而且不会陷入局部最优解的快速下降陷阱中[22]。因此，本文选择 GA 来优化 VWHM 中的动态权重。

5.1 遗传算法过程

遗传算法[23-24]的具体步骤如下。

(1) 种群初始化：初始种群包含 N 个个体，由算法随机生成，所求问题的初始解则一一对应于编码后的 N 个个体。

(2) 求适应度值：适应度值用来评价每个个体的好坏。

(3) 选择：根据初始种群中每个个体的适应度表现，选择出适应度高的 M 个个体作为下一代群体。

(4) 交叉变异：交叉操作在 M 个个体中随机选出的两个个体之间进行，按杂交概率 PC 进行选取。操作完成后得到两个全新的个体，重复进行交叉操作直至所有需要杂交的个体都两两交叉杂交完成。按概率 PM 以及变异规则从 M 个个体中选择出一些个体进行个体变异操作。

(5) 若算法运行结束，则可解码得到最优解，否则转到步骤(2)继续执行迭代运算，直至得到最优解[25-28]。

5.2 权重确定过程

组合预测模型结合了不同的个体模型，降低了运用单个预测方法进行预测的误差[29]。GA 可以动态更新单个模型的权重。根据历史数据的预测误差，优化下一次预测时单个模型的权重。变权组合预测模型在每次预测的时候都会更新单个模型的权重。具体过程如下：每当有新的加载数据进入，时间窗口便会向右移动，然后删除最早的数据，组合预测模型接着对包含最新加载数据的序列进行预测，同时利用 GA 优化单个模型的权重[30-31]。重复这个过程直到获得最终的加权预测结果。假设 $P(t)$ 为时间 t 风速的实际值，$\hat{P}_{\mathrm{A}}(t)$ 和 $\hat{P}_{\mathrm{B}}(t)$ 分别为 GRNN 模型和 LSTM 模型的预测值。于是，组合模型的预测相对误差为

$$e_t=\frac{\alpha\cdot\hat{P}_{\mathrm{A}}(t)+\beta\cdot\hat{P}_{\mathrm{B}}-P(t)}{P(t)} \tag{18}$$

式中，α 和 β 表示单个模型相应的权重。

对于实际的可变权重优化问题，优化目标如式(19)所示。

$$\min_{\alpha,\beta}\sum_{i=1}^{j}e^2_{t-i}\Big/j \tag{19}$$

式中，j 表示移动窗口内样本个数，约束条件是

$$\alpha+\beta=1 \tag{20}$$

$$0\le\alpha,\ \beta\le1 \tag{21}$$

由于权重需要在每个预测时间更新，而且为了获取准确的权重，我们需要一个计算速度快，精度高的优化方法。 本文提出运用遗传算法(GA)来优化组合模型中单个模型的权重。

6 基于 EEMD 和 LASSO 的短期风速组合变权预测模型

6.1 预测评价指标的选取

为了保证能够对各个模型的预测准确率进行有效且全面地评价，本文采用了三种不同的评价指标：平均绝对误差(Mean Absolute Error, MAE)、平均绝对百分比误差(Mean Absolute Percentage Error, MAPE))和均方根误差(Root of the Mean Squared Error, RMSE)来对涉及模型的预测结果进行评价。

$$E_{\mathrm{MAE}}=\frac{1}{N}\sum_{i=1}^{N}\left|y_t-\hat{y}_t\right| \tag{22}$$

$$E_{\mathrm{MAPE}}=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_t-\hat{y}_t}{y_t}\right|\times100\% \tag{23}$$

$$E_{\mathrm{RMSE}}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_t-\hat{y}_t\right)^2} \tag{24}$$

式中：N 为测试样本数；y_t 为 t 时刻实际风速值；$\hat{y}_t$ 为 t 时刻预测风速值[32]。

6.2 建模过程

建模流程图如图 3 所示。具体建模步骤如下：

(1) 运用 EEMD 技术降低原始风速时间序列的非线性，得到多个不同的子序列。

(2) 运用套索算法对分解后的子序列进行变量筛选，将筛选出的相关变量作为 GRNN 和 LSTM 的输入变量进行预测，同时根据 5.2 节权重确定过程步骤，运用 GA 对组合预测模型的权重系数进行移动样本自适应变权求解，从而得到各子序列预测结果。

(3) 最后叠加全部子序列的预测值得到实际预测结果。

(4) 对预测结果进行误差分析。

7 案例分析

本文验证案例为河北省张家口某风电场 2018 年 5 月所采集的实际风速数据。验证样本集共 900 个风速数据，每两个样本点之间时间间隔为 5 min。其中，选取样本集中的第 1 个至第 800 个风速数据

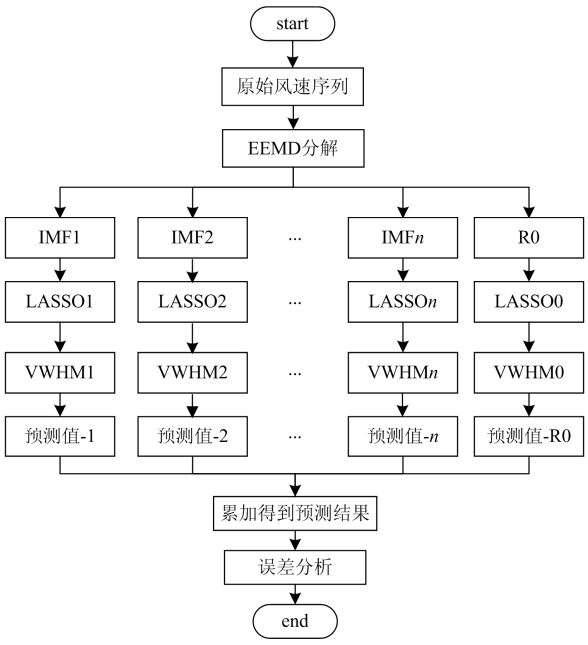

图 3 基于 EEMD-LASSO-VWHM 的短期风速预测模型结构

Fig. 3 Short-term wind speed prediction model structure diagram based on EEMD-LASSO-VWHM

形成预测模型的训练样集合，第 801 个至第 900 个数据则作为预测模型的验证集合。风速观测值数据如图 4 所示，从图中曲线可以看出风速变化随机性强，变化幅度较大。风速观测值特性统计信息如表 1 所示。

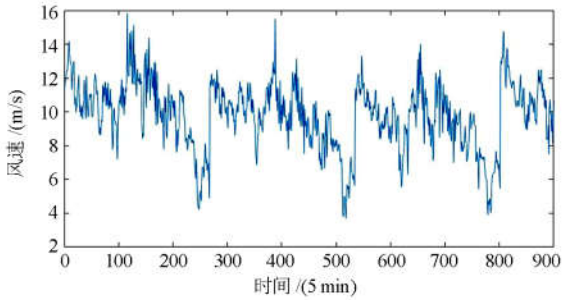

图 4 初始风速数据

Fig. 4 Initial wind speed data

表 1 训练数据集和验证数据集的统计信息

Table 1 Statistical information of the training data set and the testing data

评价指标	最大值	最小值	均值	方差
数值	15.80	3.69	9.79	3.97

本文所涉及模型的训练和仿真过程均在 Matlab R2018a 以及 Windows 操作系统环境下完成，计算运行设备为 RAM 为 8GB 的 PC。各模型参数如表 2 所示。

7.1 数据预处理及案例设计

本文通过采用 EEMD 对原始风速数据进行分解，提取出一系列具有不同带宽的子序列。初始风速数据分解结果如图 5 所示。

运用 LASSO 回归法对风速预测影响较大的变量进行筛选，筛选结果如表 3 所示。

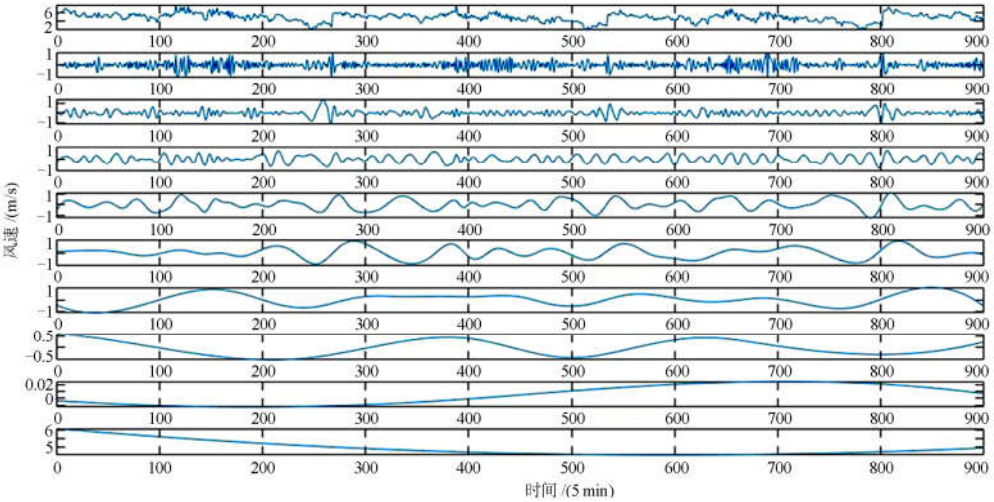

图 5 原始风速序列分解结果

Fig. 5 Decomposition results for the original wind speed sequence

表 2 各算法参数设置

Table 2 Parameter setting for each algorithm

算法	参数	数值
EEMD	集合数量	100
	白噪声标准差	0.01
LASSO	绝对误差容限	1×10^{-4}
	α	1
GA	种群数量	50
	交叉率	0.95
	变异率	0.05
GRNN	平滑因子	1
LSTM	最大次数	100
	梯度阈值	1
	学习率下降期	125
	学习率下降因子	0.2

表 3 LASSO 算法筛选变量结果

Table 3 LASSO algorithm filter variable results

序列	筛选变量
IMF1	X_{t-1}，X_{t-2}，X_{t-3}
IMF2	X_{t-1}，X_{t-2}，X_{t-3}，X_{t-4}，X_{t-5}
IMF3	X_{t-1}，X_{t-2}，X_{t-3}，X_{t-4}
IMF4	X_{t-1}，X_{t-2}，X_{t-4}
IMF5	X_{t-1}，X_{t-4}，X_{t-5}
IMF6	X_{t-1}，X_{t-4}，X_{t-5}，X_{t-10}
IMF7	X_{t-1}，X_{t-3}，X_{t-8}，X_{t-10}
IMF8	X_{t-1}，X_{t-5}，X_{t-6}，X_{t-10}
IMF9	X_{t-6}，X_{t-7}，X_{t-8}，X_{t-9}，X_{t-10}
R0	X_{t-2}，X_{t-3}，X_{t-4}，X_{t-5}，X_{t-6}

将筛选出的影响变量作为预测的输入变量对未来风速进行 1 步至 3 步的预测。

为了验证本文所提出的基于 EEMD 和变权组合模型的预测能力，本文在此设置两种验证方案进行对比分析。

实验 1：将本文所提出的 EEMD-LASSO-VWHM 模型与不同预测模型进行对比，以验证所提预测模型的优越性。所采用的对比模型包括：ARIMA、BPNN、VWHM、EEMD-GRNN 和 EEMD-LSTM。

实验 2：将本文所提出的 EEMD-LASSO-VWHM 模型和传统加权的组合模型(Traditional Weighted Hybrid Model, TWHM)进行对比分析。其中，传统变权方法采取平分权重的思想来对每个模型进行赋权。通过两种模型预测结果的对比，来验证变权组合模型的优越性。

7.2 基于 EEMD-GA-GRNN-LSTM 的预测结果分析

利用本文所提到的模型，对 EEMD 从原始序列中提取出的所有子序列分别进行预测。实验 1 中各模型的预测误差分析如表 4 所示，实验 2 中各个模型的预测表现如表 5 所示。此外，本文所提出的模型预测结果如图 7 至图 9 所示。

从表 4 可以看出，在所有预测模型中本文所提出的 EEMD-LASSO-VWHM 预测方法具有最高的预测精度。在 1 至 3 步预测中，EEMD-LASSO-VWHM 与 ARIMA 相比，E_{MAE} 分别提高了 37.37%、

表 4 实验 1 中模型预测误差分析

Table 4 Analysis of model prediction errors in experiment 1

模型	步长	E_{MAE}	E_{MAPE}	E_{RMSE}
ARIMA	1	1.015 7	0.094 0	1.508 8
	2	1.512 2	0.136 7	2.243 6
	3	1.851 3	0.168 7	2.744 2
BPNN	1	0.763 3	0.072 5	1.050 6
	2	0.928 1	0.084 1	1.344 1
	3	1.108 5	0.092 7	1.497 7
VWHM	1	0.749 7	0.069 6	1.057 2
	2	0.888 8	0.079 2	1.339 2
	3	0.956 3	0.085 4	1.444 6
EEMD-GRNN	1	0.937 5	0.084 6	1.239 6
	2	1.027 0	0.090 0	1.386 5
	3	1.089 2	0.095 4	1.458 4
EEMD-LSTM	1	0.666 8	0.068 4	0.928 9
	2	0.874 3	0.090 8	1.201 0
	3	1.406 2	0.131 2	1.665 0
EEMD-LASSO-VWHM	1	0.636 1	0.059 2	0.909 6
	2	0.770 2	0.069 6	1.129 2
	3	0.882 1	0.078 6	1.269 2

49.07%和 52.35%，证明本文所提方法预测效果优于单一的统计分析模型；EEMD-LASSO-VWHM 与 GRNN 相比，E_{MAE} 分别提高了 16.67%、17.02%和 20.42%，说明本文所提方法预测精度高于单一的智能算法；EEMD-LASSO-VWHM 与 VWHM 相比，E_{MAE} 分别提高了 15.15%、13.34%和 7.75%，从而可以看出数据预处理方法 EEMD 和 LASSO 在促进预测精度提升方面具有积极作用。此外，通过 EEMD-LASSO-VWHM 与 EEMD-GRNN 和 EEMD-LSTM 的对比，可以看出本文所提及的变权组合预测方法通过发挥不同预测模型各自的优点，根据不同模型在移动区间上的预测表现动态调整模型赋权从而得到更为准确的预测结果。预测结果表明，其预测表现要好于一般的组合预测模型。

为进一步说明本文所提变权组合模型的优势，以 1 步风速变权预测时各模型的权重变化情况为例对权重分配情况进行展示，权重分配情况如图 6 所示。

从图 6 可以看出 IMF1 至 IMF3 中权重偏向于 GRNN，而在后面的 IMF4 至 IMF9 中权重则偏重于 LSTM；这说明对于频率较高、波动性较强的序列，GRNN 的非线性拟合能力可以更高效地对数据进行预测，而对于频率相对较低、波动性较弱的序列，LSTM 则可发挥其对于长时间间隔序列具有的记忆能力，实现更高的预测精度。

由表 5 中对比数据可以看出，本文所提出的 EEMD-LASSO-VWHM 方法预测精度更高。在 1 步至 3 步预测中，EEMD-LASSO-VWHM 相对于传统加权预测模型，E_{MAE} 分别提高了 22.33%，11.40%和 12.68%；E_{MAPE} 分别提高了 20.89%，10.21%和 12.20%；E_{RMSE} 分别提高了 17.23%，4.14%和 3.78%。由此说明，相对于传统加权方法，本文所提出的 EEMD-LASSO-VWHM 方法，通过充分发挥 GA 的全局搜索能力，根据历史数据的预测表现来进行动态赋权，可以更准确地捕捉每个模型预测效果的变化，从而对各个预测模型权重进行及时调整，最终得到更为准确的预测结果。

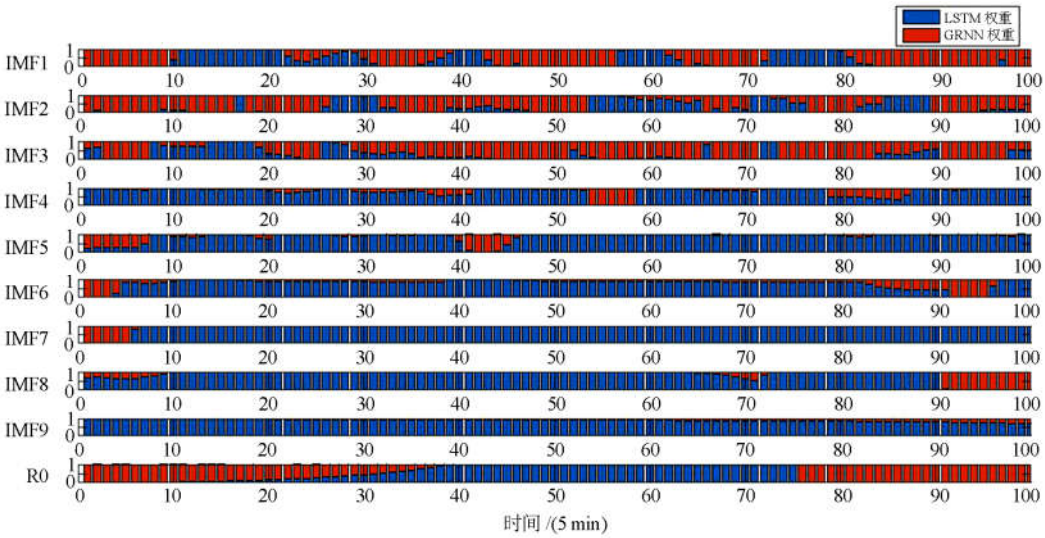

图 6 组合变权模型权重

Fig. 6 Weights for each model in VWHM

表 5 变权预测模型与传统加权预测模型误差分析

Table 5 Error analysis of variable weight prediction model and traditional weighted prediction model

模型	步长	E_{MAE}	E_{MAPE}	E_{RMSE}
EEMD-LASSO-TWHM	1	0.8189	0.0748	1.0990
	2	0.8693	0.0775	1.1779
	3	1.0103	0.0895	1.3191
EEMD-LASSO-VWHM	1	0.6361	0.0592	0.9096
	2	0.7702	0.0696	1.1292
	3	0.8821	0.0786	1.2692

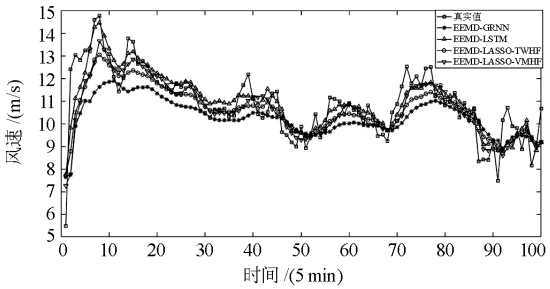

图 7 不同对比模型的 1 步预测结果

Fig. 7 One-step prediction results of different contrast models

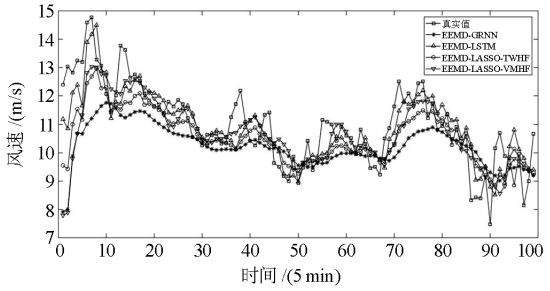

图 8 不同对比模型的 2 步预测结果

Fig. 8 Two-step prediction results of different contrast models

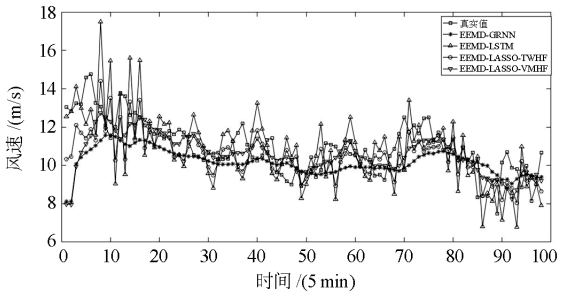

图 9 不同对比模型的 3 步预测结果

Fig. 9 Three-step prediction results of different contrast models

7.3 模型运行时间对比

本文所涉及的模型运行时间如表 6 所示。

通过表 6 可以看出，本文提出的组合变权模型相较于传统的回归模型以及其他对比模型运行时间更长，这是由于组合模型计算复杂性更高，需要更多的时间进行训练和建模。

表 6 各模型运行时间比对表

Table 6 Running time of each model

预测模型	1步	2步	3步
ARIMA	32.96	29.04	29.3
BPNN	9.32	9.24	9.43
VWHM	91.92	89.29	90.48
EEMD-GRNN	19.88	19.83	19.79
EEMD-LSTM	9.53	9.46	9.47
EEMD-LASSO-TWHM	22.91	22.69	22.46
EEMD-LASSO-VWHM	98.49	95.64	97.56

8 结论

本文提出了一种基于数据预处理和变量选择的变权组合风速预测模型。经案例分析表明：通过 EEMD 能够实现对原始风速数据的有效分解，为预测模型提供更多输入信息，降低了数据的波动性。同时进一步通过应用 LASSO 对输入变量进行筛选，为预测模型提供了更为准确的输入变量。案例分析表明，基于 EEMD 的预处理方法以及 LASSO 的变量筛选方法可以很好地提高组合预测模型的预测效果。相对于传统加权预测模型，EEMD-LASSO-VWHM 通过根据历史预测表现动态赋权的方法，更为有效地利用不同模型的优势特征，提高了预测精度。本文所提出的 EEMD-LASSO-VWHM 模型能够敏锐地识别风速的变化趋势，为决策者提供更为准确的风速预测数据，是一种具有较好预测表现的预测模型。

参考文献

[1] PINSON P, NIELSEN N, MOLLER J, et al. Nonparametric probabilistic forecasts of wind power: required properties and evaluation[J]. Wind Energy, 2007, 10(6): 497-516.

[2] 廖旎焕，胡智宏，马莹莹，等. 电力系统短期负荷预测方法综述[J]. 电力系统保护与控制，2011，39(1): 147-152.
LIAO Nihuan, HU Zhihong, MA Yingying, et al. Review of the short-term load forecasting methods of electric power system[J]. Power System Protection and Control, 2011, 39(1): 147-152.

[3] 章伟，邓院昌. 基于灰色-马尔可夫链的短期风速及风电功率预测[J]. 中国电力, 2013, 46(2):103-107.
ZHANG Wei, DENG Yuanchang. Short-term wind speed and wind power prediction based on the grey-Markov chain[J]. Electric Power, 2013, 46(2): 103-107.

[4] 张旸，颜七笙. 基于 ABC-BP 的短期风速预测研究[J].

江西科学, 2019, 37(1): 26-31.
ZHANG Yang, YAN Qisheng. A short-term wind speed prediction based on ABC-BP[J]. Jiangxi Science, 2019, 37(1): 26-31.

[5] 冯义, 刘慧文, 张宝平, 等. 基于集合经验模态分解和特征选择极端学习机的风速预测[J]. 智慧电力, 2018, 46(12): 30-37.
FENG Yi, LIU Huiwen, ZHANG Baoping, et al. Wind speed prediction based on set empirical mode decomposition and feature selection extreme learning machine[J]. Smart Power, 2018, 46(12): 30-37.

[6] 卢海明, 刘建鑫, 刘向东. 基于 WPD-MAIWO-NN 的短期风速多步预测方法[J]. 华北电力技术, 2016(11): 25-30.
LU Haiming, LIU Jianxin, LIU Xiangdong. A short-term wind speed multi-step prediction method based on WPD-MAIWO-NN[J]. North China Electric Power, 2016(11): 25-30.

[7] 芦婧, 曾明. 短期风速的 Adaboost-GRNN 组合预测模型[J]. 电力系统及其自动化学报, 2019, 31(11): 1-6.
LU Jing, ZENG Ming. Adaboost-GRNN combined forecasting model for short-term wind speed[J]. Proceedings of the CSU-EPSA, 2019, 31(11): 1-6.

[8] 刘建华, 李锦程, 杨龙月, 等. 基于 EMD-SLSTM 的家庭短期负荷预测[J]. 电力系统保护与控制, 2019, 47(6): 40-47.
LIU Jianhua, LI Jincheng, YANG Longyue, et al. Short-term load forecasting based on EMD-SLSTM[J]. Power System Protection and Control, 2019, 47(6): 40-47.

[9] ZANG H, LIANG Z, GUO M, et al. Short-term wind speed forecasting based on an EEMD-CAPSO-RVM model[C] // 2016 IEEE Power and Energy Engineering Conference, October 25-28, 2016, Xi'an, China: 439-443.

[10] 田松峰, 胥佳瑞, 王美俊, 等. 基于 EEMD 云模型与 SVM 的汽轮机转子故障诊断方法[J]. 热力发电, 2017, 46(4): 111-114.
TIAN Songfeng, XU Jiarui, WANG Meijun, et al. A rotor fault diagnosis method based on EEMD cloud model and SVM[J]. Thermal Power Generation, 2017, 46(4): 111-114.

[11] 贺坤, 李星野. 基于经验模态分解和遗传算法改进的神经网络模型的风速时间序列预测[J]. 电子测量技术, 2018, 41(8): 146-149.
HE Kun, LI Xingye. Wind-speed time series prediction based on improved neural network model based on empirical mode decomposition and genetic algorithm[J]. Electronic Measurement Technology, 2018, 41(8): 146-149.

[12] TIBSHIRANI R J. The LASSO problem and uniqueness[J]. Electronic Journal of Statistics, 2013, 7: 1456-1490.

[13] EFRON B, HASTIE T, JOHNSTONE I, et al. Least angle regression[J]. AnnStat, 2004, 32(2): 407-451.

[14] 谷志红, 牛东晓, 王会青. 广义回归神经网络模型在短期电力负荷预测中的应用研究[J]. 中国电力, 2006, 39(4): 11-14.
GU Zhihong, NIU Dongxiao, WANG Huiqing. Research on application of general regression neural network in short-term load forecasting[J]. Electric Power, 2006, 39(4): 11-14.

[15] LIU H, MI X W, LI Y F. Smart deep learning based on wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network[J]. Energy Conversion and Management, 2018, 166: 120-131.

[16] MORENO S R, COELHO L D S. Wind speed forecasting approach based on singular spectrum analysis and adaptive neuro fuzzy inference system[J]. Renewable Energy, 2018, 126: 736-754.

[17] 孔雪卉, 张慧芬. 基于优化广义回归神经网络的变电站设备温度预测[J]. 中国电力, 2016, 49(7): 54-59.
KONG Xuehui, ZHANG Huifen. Temperature forecast for substation equipment based on optimized general regression neural network[J]. Electric Power, 2016, 49(7): 54-59.

[18] 魏昱洲, 许西宁. 基于 LSTM 长短期记忆网络的超短期风速预测[J]. 电子测量与仪器学报, 2019, 33(2): 64-71.
WEI Yuzhou, XU Xining. The ultra-short-term wind speed prediction based on LSTM long-term and short-term memory network[J]. Journal of Electronic Measurement and Instrument, 2019, 33(2): 64-71.

[19] 朱乔木, 李弘毅, 王子琪, 等. 基于长短期记忆网络的风电场发电功率超短期预测[J]. 电网技术, 2017, 41(12): 3797-3802.
ZHU Qiaomu, LI Hongyi, WANG Ziqi, et al. Short-term wind power forecasting based on LSTM[J]. Power System Technology, 2017, 41(12): 3797-3802.

[20] 杨国田, 张涛, 王英男, 等. 基于长短期记忆神经网络的火电厂 NOx 排放预测模型[J]. 热力发电, 2018, 47(10): 12-17.
YANG Guotian, ZHANG Tao, WANG Yingnan, et al. Prediction model for NOx emissions from thermal power plants based on long-short-term memory neural network[J]. Thermal Power Generation, 2018, 47(10): 12-17.

[21] 李冰, 张妍, 刘石. 基于 LSTM 的短期风速预测研究[J]. 计算机仿真, 2018, 35(11): 456-461.

LI Bing, ZHANG Wei, LIU Shi. Study on short-term wind speed prediction based on LSTM[J]. Computer Simulation, 2018, 35(11): 456-461.

[22] 梁涛, 孙天一, 邹继行, 等. 基于 GA 优化的加权 LSSVM 短期风速预测[J]. 高技术通讯, 2019, 29(2): 142-148.
LIANG Tao, SUN Tianyi, ZOU Jixing, et al. Short-term wind speed prediction based on GA optimization for weighted LSSVM[J]. High Technology Letters, 2019, 29(2): 142-148.

[23] 陈红, 黎盛鸣, 耿向瑾, 等. 基于元素分析的煤粉工业分析 GA-SVM 预测模型[J]. 广东电力, 2018, 31(1): 30-35.
CHEN Hong, LI Shengming, GENG Xiangjin, et al. GA-SVM prediction model for pulverized coal proximate analysis based on ultimate analysis[J]. Guangdong Electric Power, 2018, 31(1): 30-35.

[24] 彭海应, 陈伟. 基于遗传算法的高压直流断路器振荡回路参数计算[J]. 广东电力, 2013, 26(5): 40-44, 51.
PENG Haiying, CHEN Wei. Calculation for high voltage direct current breaker vibration circuit parameters based on genetic algorithm[J]. Guangdong Electric Power, 2013, 26(5): 40-44, 51.

[25] LANG H S, LIN S C, FANG W H. Subcarrier pairing and power allocation with interference management in cognitive relay networks based on genetic algorithms[J]. IEEE Transactions on Vehicular Technology, 2016, 65(9): 7051-7063.

[26] LI Zhi, YE Lin, ZHAO Yongning, et al. Short-term wind power prediction based on extreme learning machine with error correction[J]. Protection and Control of Modern Power Systems, 2016, 1(1): 9-16. DOI: 10.1186/s41601-016-0016-y.

[27] 李啸骢, 李春涛, 从兰美, 等. 基于动态权值相似日选取算法的短期负荷预测[J]. 电力系统保护与控制, 2018, 45(6): 1-8.
LI Xiaocong, LI Chuntao, CONG Lanmei, et al. Short-term load forecasting based on dynamic weight similarity selection algorithm[J]. Power System Protection and Control, 2018, 45(6): 1-8.

[28] SHAO H J, WEI H K, DENG X, et al. Short-term wind speed forecasting using wavelet transformation and Ada boosting neural networks in Yunnan wind farm[J]. IET Renewable Power Generation, 2017, 11(4): 374-381.

[29] ZHENG D H, ESEYE A T, ZHANG J H, et al. Short-term wind power forecasting using a double-stage hierarchical ANFIS approach for energy management in microgrids[J]. Protection and Control of Modern Power Systems, 2017, 2(2): 136-145. DOI: 10.1186/s41601-017-0041-5.

[30] 颜晓娟, 龚仁喜, 张千锋. 优化遗传算法寻优的 SVM 在短期风速预测中的应用[J]. 电力系统保护与控制, 2018, 44(9): 38-42.
YAN Xiaojuan, GONG Renxi, ZHANG Qianfeng. Application of optimization SVM based on improved genetic algorithm in short-time wind speed prediction[J]. Power System Protection and Control, 2018, 44(9): 38-42.

[31] 王高红, 陈潇一. 基于遗传算法的智能电网非侵入式电器监控策略[J]. 智慧电力, 2017, 45(11): 42-46, 71.
WANG Gaohong, CHEN Xiaoyi. Non-intrusive electrical monitoring strategy for smart grid based on genetic algorithm[J]. Smart Power, 2017, 45(11): 42-46, 71.

[32] 殷豪, 董朕, 马留洋. 基于可变模式分解和 NWCSO 优化极限学习机的短期风速预测[J]. 电力建设, 2017, 38(6): 36-43.
YIN Hao, DONG Wei, MA Liuyang. Short-term wind speed prediction based on variable mode decomposition and NWCSO optimized extreme learning machine[J]. Electric Power Construction, 2017, 38(6): 36-43.

收稿日期：2019-07-12；　　修回日期：2019-12-29

作者简介：

杨　磊(1988—)，男，博士在读，主要从事电力市场理论与方法方面的研究工作；E-mail: 344954864@qq.com

黄元生(1958—)，男，博士，教授，博士生导师，主要从事电力市场理论与方法方面的研究工作；E-mail: hys2656@aliyun.com;

张向荣(1995—)，男，硕士在读，主要从事电力市场理论与方法方面的研究工作。E-mail: 1346565944@qq.com

(编辑　张爱琴)