

基于 EWT 和分位数回归森林的短期风电功率概率密度预测

孙国强¹, 梁 智¹, 俞娜燕², 倪晓宇³, 卫志农¹, 臧海祥¹, 周亦洲¹

(1. 河海大学 能源与电气学院, 江苏 南京 210098; 2. 国网无锡供电公司, 江苏 无锡 214061;

3. 无锡扬晟科技股份有限公司, 江苏 无锡 214106)

摘要: 概率密度预测能够给出未来风电功率可能的波动范围、预测值出现的概率及不确定性等更多信息, 提出基于经验小波变换(EWT)和分位数回归森林的短期风电功率概率密度组合预测模型。首先, 采用新型自适应信号处理方法——经验小波变换, 将原始风电功率序列分解为一系列频率特征互异的经验模式; 然后, 对每一经验模式序列分别构建分位数回归森林预测模型, 得到任意分位点条件下的预测结果, 通过叠加不同经验模式预测结果获得最终的短期风电功率预测值; 最后, 对预测值条件分布采用核密度估计获得任意时刻概率密度预测。仿真结果验证了所提模型的有效性。

关键词: 经验小波变换; 分位数回归森林; 核密度估计; 概率密度; 短期风电功率预测; 模型

中图分类号: TM 761

文献标识码: A

DOI: 10.16081/j.issn.1006-6047.2018.08.023

0 引言

风力发电在电网中的装机比例逐年提升, 有效地缓解了能源紧张、环境污染格局, 但其间歇性和不确定性又严重影响着电网的安全、稳定及经济运行。短期风电功率预测作为自动发电控制和安排电力调度的重要决策依据, 能够有效地提高电力系统的运行可靠性。为此, 需要研究新技术与新方法, 以提高风电功率预测精度, 满足工程应用需求。

目前, 国内外学者对短期风电功率预测进行了大量的研究, 主要有时间序列分析、人工神经网络、支持向量机(SVM)、相关向量机等模型^[1-6]。在现有研究的基础上, 文献[7-9]提出了基于优化算法的改进预测模型。另外, 为了进一步降低风电功率预测误差, 相关学者提出了组合预测模型。实践证明: 组合预测模型相较于单一预测方法能够优势互补, 在提高预测精度的同时, 增强了模型的鲁棒性。组合预测按机理策略的不同主要分为2类。第一类是采用不同原理的预测模型分别进行预测, 然后将预测结果按一定的方式进行优化组合。文献[10]首先采用自回归积分滑动平均拟合风电功率序列, 然后对拟合残差分别建立了人工神经网络、SVM预测模型; 文献[11]通过权重矩阵集成最小二乘SVM、回声状态网络及正则化极限学习机3种模型的预测结果, 并通过优化算法动态更新调整权重系数, 从而有效地发挥各个模型的优势, 提高预测精度。第二类是采用信号处理技术对原始风电功率序列进行分

解处理, 对不同分解量建立预测模型, 最后对各分量的预测结果进行组合。文献[12-14]分别采用小波变换、经验模态分解(EMD)和集成经验模态分解对原始风电功率进行处理, 然后对各子序列分别建立预测模型, 有效地提高了预测精度。

一般的短期风电功率预测方法仅给出确定性点的预测结果, 难以完全描述风能的不确定性、变化规律。因此, 相关学者提出了风电功率的概率预测方法, 如分位数回归、区间预测、密度预测等^[15]。概率预测能更好地描述未来风电功率可能的波动范围、不确定性及面临的风险, 从而更有研究价值^[16-18]。

本文针对EMD方法易出现模态混叠、计算效率低、缺乏理论基础等缺点, 采用新型自适应信号处理方法——经验小波变换EWT(Empirical Wavelet Transform)对原始风电功率序列进行分解处理。该方法通过对信号频谱的自适应分割, 在各个频谱构造合适的正交小波滤波器来提取Fourier频谱的调幅、调频成分, 然后采用Hilbert变换对不同调幅、调频模态进行处理, 获得瞬时频率和瞬时幅值^[19-20]。EWT方法的计算量小且具有较强的鲁棒性。

分位数回归森林(QRF)结合分位数回归和随机森林(RF)的基本原理, 可给出不同分位点的回归预测结果。作为一种非参数集成机器学习方法, QRF具有运算速度快、模型性能受参数影响小、容噪性较强等优点。本文建立基于QRF的风电功率预测模型, 获得不同分位点的预测输出, 然后采用核密度估计实现风电功率概率密度预测。

综上所述, 本文结合EWT和QRF的优点, 建立基于EWT-QRF的短期风电功率概率密度预测模型。首先, 采用EWT将原始风电功率序列分解为一系列频率不同的经验模式, 对每一经验模式分别建立QRF预测模型, 获得不同分位点的回归预测结果, 将各经验模式的预测结果叠加, 得到最终的风电

收稿日期: 2017-07-14; **修回日期:** 2018-06-25

基金项目: 国家自然科学基金资助项目(51507052); 国家电网公司科技项目(J2017089)

Project supported by the National Natural Science Foundation of China(51507052) and the Science and Technology Project of SGCC(J2017089)

功率预测值。最后,采用核密度估计方法给出风电功率概率密度预测。

1 EWT 方法的基本原理

经验小波本质上是根据信号频谱特性选择的一组带通滤波器,能够自适应地从原始信号中筛选调幅、调频成分。为了确定带通滤波器的频率范围,首先对信号的 Fourier 谱进行自适应分割,见图 1。

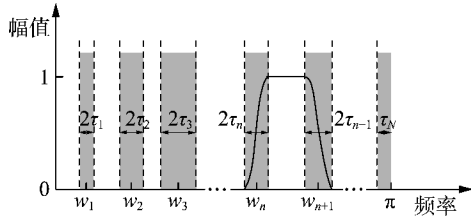


图 1 Fourier 频谱分割

Fig.1 Partitioning of Fourier frequency spectrum

结合图 1 说明 EWT 的自适应分解过程。依据香农准则,定义 Fourier 支撑为 $[0, \pi]$ 并假设其被分割成 N 个连续部分,令 $\Lambda_n = [w_{n-1}, w_n]$ 表示各分割片段的边界,其中 $n=1, 2, \dots, N, w_0=0, w_N=\pi, w_n$ 选取为信号 Fourier 谱相邻 2 个极大值点之间的中点,显而易见 $\cup \Lambda_n = [0, \pi] (n=1, 2, \dots, N)$ 。以每个 w_n 为中心,定义宽度为 $T_n = 2\tau_n$ 的过渡区域,见图 1 中阴影部分。在分割区间 Λ_n 上,定义经验小波为每个 Λ_n 上的带通滤波器,并根据 Meyer 小波的构造方法构造经验小波。Gilles 构造的经验小波函数为^[19]:

$$\hat{\psi}_n(w) = \begin{cases} 1 & w_n + \tau_n \leq |w| \leq w_{n+1} - \tau_{n+1} \\ \cos\left\{\frac{\pi}{2}\beta\left[\frac{1}{2\tau_{n+1}}(|w| - w_{n+1} + \tau_{n+1})\right]\right\} & w_{n+1} - \tau_{n+1} \leq |w| \leq w_n + \tau_n \\ \sin\left\{\frac{\pi}{2}\beta\left[\frac{1}{2\tau_n}(|w| - w_n + \tau_n)\right]\right\} & w_n - \tau_n \leq |w| \leq w_{n+1} + \tau_{n+1} \\ 0 & \text{其他} \end{cases} \quad (1)$$

经验尺度函数为:

$$\hat{\phi}_n(w) = \begin{cases} 1 & |w| \leq w_n - \tau_n \\ \cos\left\{\frac{\pi}{2}\beta\left[\frac{1}{2\tau_n}(|w| - w_n + \tau_n)\right]\right\} & w_n - \tau_n \leq |w| \leq w_n + \tau_n \\ 0 & \text{其他} \end{cases} \quad (2)$$

其中, $\tau_n = \gamma w_n$, 且有 $\gamma < \min\left\{\frac{w_{n+1} - w_n}{w_{n+1} + w_n}\right\}$ 。通常, 函数 $\beta(z)$ 定义为 $\beta(z) = z^4(35 - 84z + 70z^2 - 20z^3)$ 。

然后,原始信号可被重构为:

$$x(t) = W_x^e(0, t) \otimes \phi_1(t) + \sum_{n=1}^N W_x^e(n, t) \otimes \phi_n(t) \quad (3)$$

其中, $\otimes$ 为卷积运算; $W_x^e(0, t)$ 为逼近系数; $W_x^e(n, t)$ 为 $x(t)$ 的经验小波变换。

经验模式 $x_k(t)$ 按式(4)定义。

$$\begin{cases} x_0(t) = W_x^e(0, t) \otimes \phi_1(t) \\ x_k(t) = W_x^e(n, t) \otimes \phi_n(t) \end{cases} \quad (4)$$

2 基于 QRF 的概率密度预测

2.1 分位数回归原理

分位数回归是单因变量 Y 的条件分位数对自变量 X 进行回归,从而获得所有分位点下的回归预测模型。

在给定条件 $X = [x_1, x_2, \dots, x_k]$ 下,条件分布函数是 $Y \leq y$ 的累积概率,即:

$$F(y|X) = P(Y \leq y|X) \quad (5)$$

α 分位数 $Q_\alpha(X)$ 为在给定条件 $X = [x_1, x_2, \dots, x_k]$ 下, $Y \geq Q_\alpha(X)$ 的累积概率恰好为 α , 即:

$$Q_\alpha(X) = \inf\{y: F(y|X) \geq \alpha\} \quad (6)$$

其中, $\inf\{\cdot\}$ 为取最小值运算。

一般的线性条件分位数回归表示为^[21]:

$$Q_Y(\alpha|X) = \beta_0(\alpha) + \beta_1(\alpha)x_1 + \beta_2(\alpha)x_2 + \dots + \beta_k(\alpha)x_k = X'\beta(\alpha) \quad (7)$$

其中, $Q_Y(\alpha|X)$ 为因变量 Y 在自变量 $X = [x_1, x_2, \dots, x_k]$ 下的第 α 个条件分位数, α 为分位点, $\alpha \in (0, 1)$; $X' = [1, x_1, x_2, \dots, x_k]$ 为 X 的扩展形式; $\beta(\alpha)$ 为回归系数向量,它随着分位点 α 的变化而变动; $\equiv$ 表示恒等于。

条件分位数通过最小化损失函数求解参数向量 $\beta(\alpha)$ 的估计值,定义损失函数为:

$$L_\alpha(y, X'\beta) = \begin{cases} \alpha |y - X'\beta| & y > X'\beta \\ (1-\alpha) |y - X'\beta| & y \leq X'\beta \end{cases} \quad (8)$$

从而分位数回归可以转化为如下最优化问题:

$$\min_{\beta} \sum_{i|Y_i \geq X'_i\beta} \alpha |Y_i - X'_i\beta| + \sum_{i|Y_i < X'_i\beta} (1-\alpha) |Y_i - X'_i\beta| \quad (9)$$

其中, Y_i 为第 i 个数据样本因变量; X'_i 为第 i 个数据样本扩展自变量。

给定某个分位点 α , 通过求解对应的参数向量估计值,即可描述此时的自变量对因变量的影响。继而当 α 在可行区间 $(0, 1)$ 内连续取值时,即可得到 Y 的条件分布。

2.2 QRF 的基本过程

QRF 是 RF 算法的改进,通过结合分位数回归的特性,可提供因变量的全部条件分布信息。QRF 作为一种非参数机器学习方法,具有理论基础,同时被证明具有一致性^[22]。

RF 被看作是一个适应性近邻分类和回归过程,对每一个 $\mathbf{X}=[x_1, x_2, \dots, x_k]$, 可以得到原始 M 个观察值的一个权重集合 $w_i(\mathbf{X}) (i=1, 2, \dots, M)$ 。RF 本质上是利用所有因变量观测值的加权和作为因变量 Y 条件均值 $E(Y|\mathbf{X})$ 的估计。QRF 决策树是以标准 RF 算法产生的, 条件分布是通过观测到的因变量加权估计得到的, 其中每个观测值的权重等于 RF 算法的权重^[23]。

由此, QRF 定义 $E(1_{\{Y \leq y\}}|\mathbf{X})$ 的估计为观测值 $1_{\{Y \leq y\}}$ 的加权平均, 即:

$$\hat{F}(y|\mathbf{X}) = \sum_{i=1}^M w_i(\mathbf{X}) 1_{\{Y \leq y\}} \quad (10)$$

QRF 算法的具体步骤如下。

a. 生成 K_T 棵决策树 $T(\theta_t) (t=1, 2, \dots, K_T)$, 考察每棵决策树每个叶节点的所有观测值。

b. 给定 $\mathbf{X}$, 遍历所有决策树。计算每棵决策树观测值的权重 $w_i(\mathbf{X}, \theta_t) (i=1, 2, \dots, M; t=1, 2, \dots, K_T)$ 。通过对决策树权重 $w_i(\mathbf{X}, \theta_t)$ 取平均得到每个观测值的权重 $w_i(\mathbf{X})$ 。

c. 对于所有 $y \in \mathbf{R}$, 利用步骤 b 得出的权重, 通过式(10)计算分布函数的估计。

对于每棵决策树的每个节点, RF 回归只保留了观测值的均值而忽略了其他信息, 而 QRF 保留了节点中所有观测值, 并在此基础上计算条件分布。

2.3 基于条件分布的概率密度预测

采用核密度估计方法从条件分布中获得概率密度预测结果。核密度估计是通过一组观测的来自同一未知分布函数的随机变量来估计其密度函数的非参数计算方法。设 $X_1, X_2, \dots, X_M$ 是取自一元连续总体的样本, 在任意点 x 处的总体密度函数 $f(x)$ 的核密度估计定义为^[24]:

$$\hat{f}_h(x) = \frac{1}{Mh} \sum_{i=1}^M K\left(\frac{x - X_i}{h}\right) \quad (11)$$

其中, $K(x)$ 为核函数, 本文采用的高斯核函数形式为 $K(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$; h 为带宽系数, 取值范围为 1.8~2.0。

3 基于 EWT-QRF 的风电功率概率密度预测

针对一般风电功率点预测方法难以完全表征功率变化不确定性的缺点, 本文建立了基于 EWT-QRF 的短期风电功率概率密度预测模型, 可以获得任意时刻风电功率的波动范围及概率密度输出。图 2 为本文短期风电功率概率密度预测流程图。首先, 采用 EWT 信号处理技术对原始风电功率序列进行预处理, 将其自适应分解为若干频率互异的经验模式。通过 EWT 分解可以细致把握不同分量的变化规律并针对性地建立预测模型, 从而有效地提高风电功

率预测精度。为了减少建模任务量, 将频率大小相近的经验模式合并为新的分量。然后, 对新的分量选取输入变量集合并建立 QRF 预测模型, 得到不同分位点下的风电功率预测结果, 将不同分量预测结果叠加获得预测值的条件分布。最后, 采用核密度估计方法输出风电功率概率密度预测。

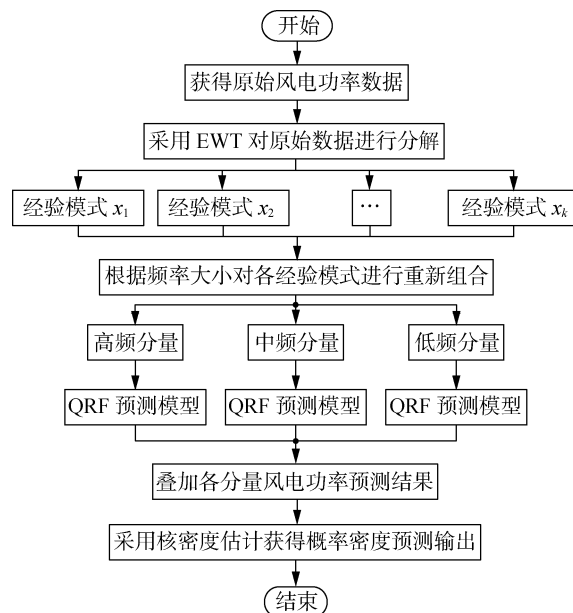


图 2 短期风电功率概率密度预测流程图

Fig.2 Flowchat of short-term wind power probability density forecasting

3.1 风电功率分解结果

采用江苏省某一风电场实测风电功率数据作为研究对象, 验证所提模型的预测性能。该风电场内风机总数为 33 台, 装机容量为 49.5 MW。已知数据采样时间间隔为 30 min, 原始风电功率序列为 2008 年 8 月 13 日至 8 月 22 日共 480 个点的风功率数据。采用 EWT 方法对原始风电功率序列进行分解, 结果如图 3 所示。从图 3 中可以看出, EWT 将原始风电功率分解为 11 个经验模式, 各个模式的频率特征较为明显。按照频率大小将经验模式分为 3 类: 低频、中频和高频, 然后将类中的经验模式进行合并, 分别重构为低频分量、中频分量和高频分量, 如图 4 所示。对各分量分别建立 QRF 预测模型, 从而降低建模工作量、提高效率。

为了说明 EWT 用于风电功率预测的有效性, 同时将 EMD 方法用于功率数据的分解, 并对比预测结果。EMD 结果及重构分量分别见图 5、图 6。

3.2 输入变量选择

输入变量的选取对模型预测性能有直接影响, 采用皮尔森相关系数定量评价变量间的相关性, 并从待预测时刻的前 10 个时刻中选取相关性较大的输入变量集合。给定时间序列, 皮尔森相关系数衡量了 x_t 与 $x_{t-\tau}$ 间的相关关系^[25]。其中, τ 为滞后阶

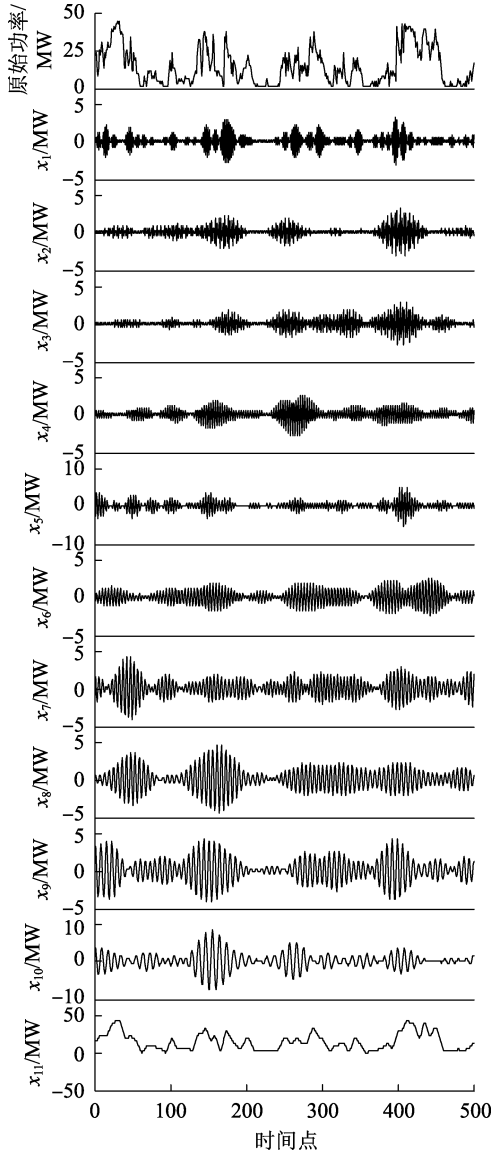


图3 原始风电功率序列及 EWT 分解结果

Fig.3 Original wind power sequence and EWT decomposition results

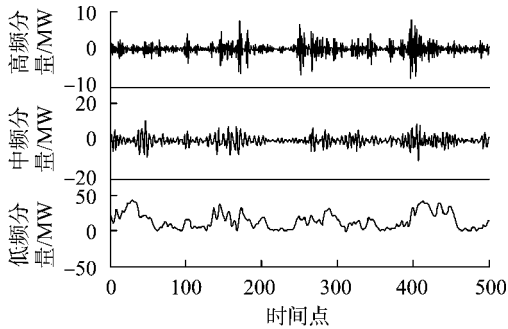


图4 经验模式重构结果

Fig.4 Empirical mode reconstruction results

数。皮尔森相关系数 r 的取值范围为 $[-1, 1]$, 1 表示完全正相关, -1 表示完全负相关, 0 表示变量间无相关性。当 $|r| > 0.3$ 时, 即认为变量间存在较强的相关关系。待预测时刻风电功率与滞后 τ 阶功率间的相关系数计算公式为:

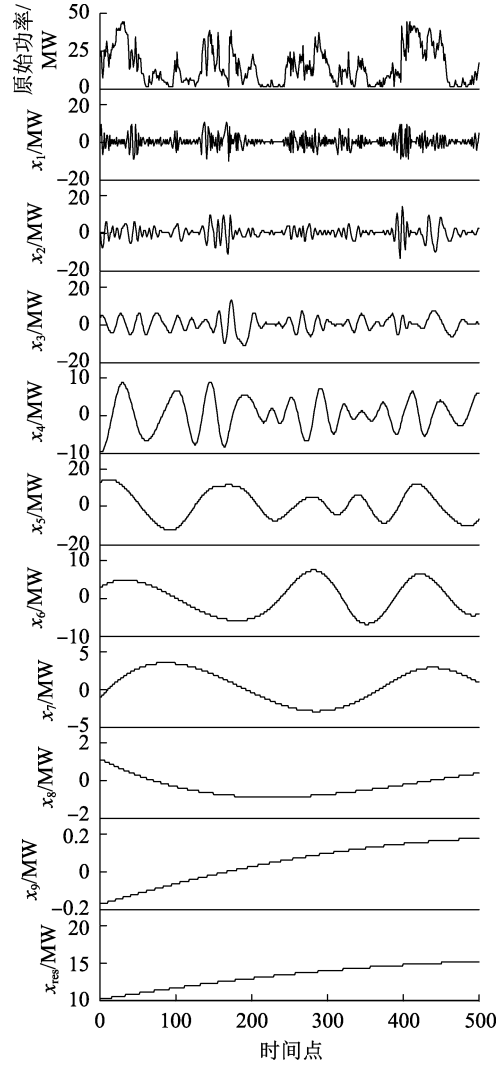


图5 原始风电功率序列及 EMD 结果

Fig.5 Original wind power sequence and EMD results

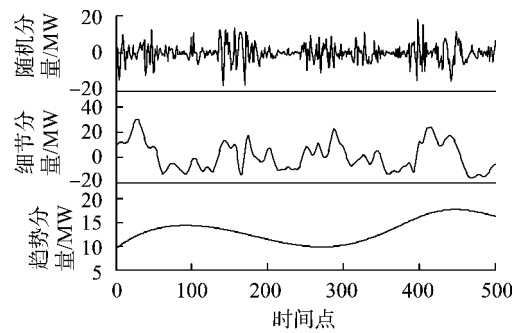


图6 EMD 重构分量

Fig.6 Reconstruction components of EMD

$$r = \frac{\sum_{i=1}^S (x_{t,i} - \bar{x}_t)(x_{t-\tau,i} - \bar{x}_{t-\tau})}{\sqrt{\sum_{i=1}^S (x_{t,i} - \bar{x}_t)^2} \sqrt{\sum_{i=1}^S (x_{t-\tau,i} - \bar{x}_{t-\tau})^2}} \quad (12)$$

其中, $x_{t,i}$ 、 $x_{t-\tau,i}$ 分别为变量 x_t 、 $x_{t-\tau}$ 的第 i 个样本值; $\bar{x}_t$ 、 $\bar{x}_{t-\tau}$ 分别为变量 x_t 、 $x_{t-\tau}$ 的平均值; S 为样本数量。

分别对 EWT 和 EMD 的分量选取输入变量集

合,结果如表 1 和表 2 所示。

表 1 EWT 分量输入变量选择结果

Table 1 Selection results of input variables for EWT method

分量	输入变量
原始功率	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8}, x_{t-9}, x_{t-10}$
高频分量	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-7}$
中频分量	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-8}, x_{t-10}$
低频分量	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8}, x_{t-9}, x_{t-10}$

表 2 EMD 分量输入变量选择结果

Table 2 Selection results of input variables for EMD method

分量	输入变量
随机分量	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$
细节分量	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-9}$
趋势分量	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8}, x_{t-9}, x_{t-10}$

3.3 评价指标

采用平均绝对百分比误差 (MAPE) 和均方根误差 (RMSE) 作为确定性点预测模型效果评价指标。

$$\text{MAPE} = \frac{1}{n_f} \sum_{i=1}^{n_f} \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{n_f} \sum_{i=1}^{n_f} (y_i - \hat{y}_i)^2} \quad (14)$$

其中, n_f 为预测点个数; $y_i, \hat{y}_i$ 分别为第 i 个预测点功率真实值和预测值。

可靠性和清晰度是风电功率概率区间预测模型的 2 个重要评价准则^[26]。可靠性为实际观测值落入预测区间的概率, 此值越大, 表明预测结果更准确; 清晰度用预测区间宽度表征, 预测区间宽度越窄, 表明模型预测性能越好。高可靠性及尽可能小的区间宽度是风电功率概率预测追求的目标, 但实际上两者是矛盾的。引入预测区间覆盖率 (FICP) 指标计算预测区间可靠性, 计算公式为:

$$\text{FICP} = \frac{1}{n_f} \sum_{i=1}^{n_f} \xi_i^{(1-\alpha)} \quad (15)$$

其中, ξ 在给定置信水平 $1-\alpha$ 条件下的取值为 0 和 1, 若实际值落于预测区间上下限范围内, 则 $\xi^{(1-\alpha)} = 1$, 否则 $\xi^{(1-\alpha)} = 0$ 。

引入预测区间平均宽度 (FIAW) 指标计算模型清晰度, 计算公式为:

$$\text{FIAW}_{1-\alpha} = \frac{1}{n_f} \sum_{i=1}^{n_f} \frac{u_i - l_i}{y_i} \quad (16)$$

其中, u_i 为第 i 个预测点置信区间上限; l_i 为第 i 个预测点置信区间下限。

4 算例分析

以 2008 年 8 月 13 日至 8 月 22 日共 480 个点的风功率数据为训练数据, 对 8 月 23 日 48 个实测风电功率进行提前 30 min 预测。本文中 EWT 分解、QRF

算法及 BP 神经网络、SVM 算法均在 MATLAB R2014a 编程平台实现, CPU 主频为 3.3 GHz, RAM 为 4 GB。

a. 直接对原始功率数据进行预测。QRF 参数设置如下: 决策树数目为 500, 节点最小尺寸为 5, 每棵决策树从输入变量集合中随机选取 $m_{\text{try}} = 2D/3$ 个变量进行权重学习, D 参照表 1 选取。为了获得条件分布, 设置分位点范围为 0.01~0.99, 步长为 0.01, 对每个预测点即可获得 99 个预测结果。同时, 建立 BP 神经网络、SVM 算法的对比模型。BP 神经网络的学习率为 0.001, 学习目标为 0.01, 迭代 10 000 次, 模型结构参数经多次试验比较后设置为: 9-15-1 (输入层变量个数为 9, 隐含层神经元个数为 15, 输出层神经元个数为 1)。SVM 模型的学习参数 C 和 ε 通过网格搜索法优化选取, 参数范围为 $[-8, 8]$, 迭代步长为 1。图 7 为原始功率曲线及 BP 神经网络、SVM 和 QRF 模型的预测结果, QRF 模型取 0.5 分位点条件下的预测值。不同模型风电功率预测结果统计如表 3 所示。从图 7、表 3 可以看出, 采用单一预测模型时, 预测值落后于真实值的变化趋势, 存在滞后, 有较大的预测误差。

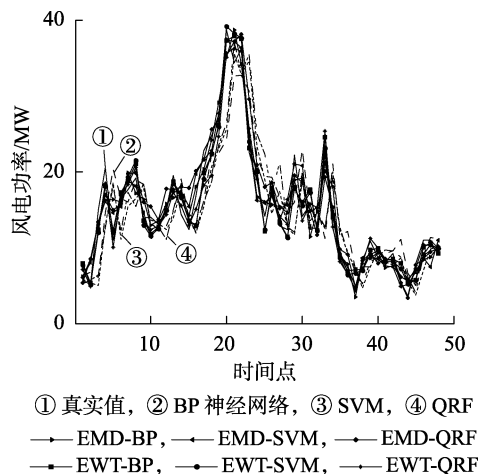


图 7 原始功率曲线及提前 30 min 预测结果

Fig.7 Original wind power curve and forecasting results ahead 30 minutes

表 3 不同模型风电功率预测结果

Table 3 Wind power forecasting results of different models

预测模型	MAPE/%	RMSE/MW	运行时间/s
BP 神经网络	27.44	5.16	7
SVM	27.27	5.12	32
QRF	26.81	4.90	15
EMD-BP	16.14	2.50	37
EMD-SVM	13.77	2.22	121
EMD-QRF	13.82	2.43	69
EWT-BP	12.11	1.82	48
EWT-SVM	11.90	1.79	136
EWT-QRF	10.77	1.71	74

b. 分别采用 EMD 和 EWT 方法对原始功率数据进行分解处理, 然后对每个分量分别建立预测模型,

结果如图 7、表 3 所示。从图 7 及表 3 可以看出组合模型有效地缓解了滞后效应。此外,EWT 方法的性能表现优于 EMD 方法,相较于 EMD-BP、EMD-SVM、EMD-QRF 模型,EWT-BP、EWT-SVM 和 EWT-QRF 模型的 MAPE 指标分别降低了 24.97%、13.58% 和 22.07%,RMSE 指标分别降低了 27.20%、19.37% 和 29.63%。其中,EMD-BP 模型随机分量、细节分量和趋势分量的结构参数分别为 4-11-1、8-16-1 以及 10-17-1;EWT-BP 模型高频分量、中频分量和低频分量结构参数分别为 5-9-1、7-15-1 和 10-15-1;SVM 和 QRF 参数设置同步步骤 a。

图 8 为 70%、80%、90%、95% 置信水平下功率区间预测结果。由图 8 知,EWT-QRF 模型的区间宽度更窄,优势明显,从而更有利于科学决策。

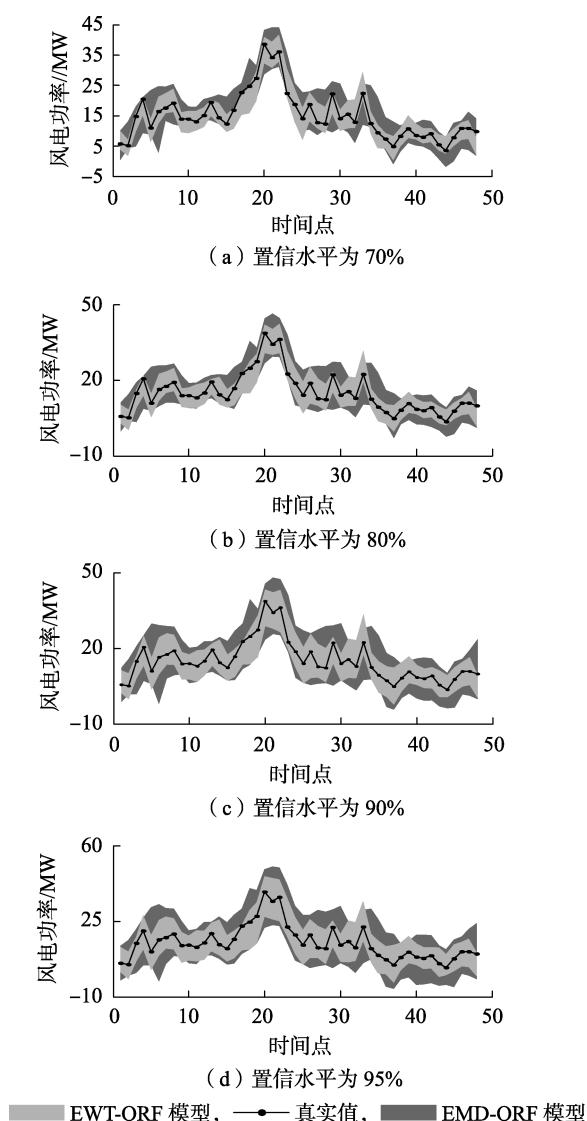


图 8 不同置信水平下功率区间概率预测结果

Fig.8 Probability forecasting results of power interval under different confidence levels

图 9 为不同置信水平下的 FICP、FIWA 指标变化曲线。由图 9 可看出,虽然 QRF 模型比 EMD-

QRF 模型有更窄的区间宽度,但由于 QRF 模型的滞后性,导致其可靠性较差,较难反映风电功率真实值;EWT-QRF 模型有最窄的区间宽度,在 40%~90% 置信区间范围内,其与 EMD-QRF 模型间的 FIAW 差值逐渐增大,导致 EWT-QRF 模型的 FICP 指标较差,这符合 FIAW、FICP 指标相互矛盾的规律。

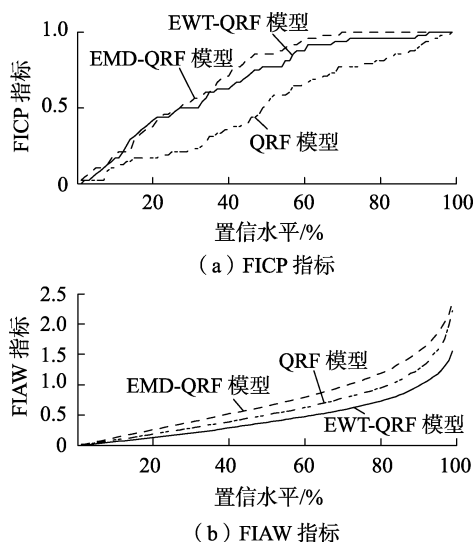
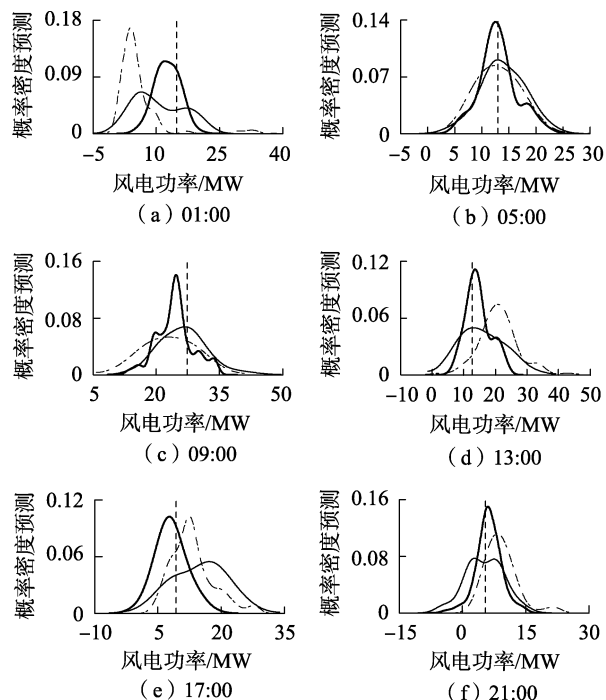


图 9 不同置信水平下指标变化曲线

Fig.9 Change curves of indices under different confidence levels

对 QRF、EMD-QRF 和 EWT-QRF 模型获得的条件分布采用核密度估计可给出任意时刻风电功率概率密度预测,取一天中不同时刻预测结果见图 10。



— QRF 模型, — EMD-QRF 模型, — EWT-QRF 模型, - - 真实值

图 10 不同时刻风电功率的概率密度预测

Fig.10 Probability density forecasting of wind power at different moments

可以看出,EWT-QRF模型以较高的概率接近真实值,概率密度曲线更瘦高,波动范围更集中,有利于在更窄范围内做出可靠决策。

5 结论

针对一般风电功率点预测方法输出结果单一的缺点,本文建立了基于EWT-QRF的概率密度组合预测模型。

a. 采用EWT这一新型自适应信号处理方法对原始风电功率序列进行分解处理,将其分解为多个频率特征互异的经验模式。相较于EMD方法,EWT分解的分量更具有解释意义。所建立的EWT-QRF预测模型相较于EMD-QRF模型具有更好的预测精度,MAPE指标降低了22.07%,RMSE指标降低了29.63%;同一置信水平下,EWT-QRF模型的区间宽度更窄,更有利于做出科学决策。

b. QRF能够给出任意分位点下的预测结果,从而获得预测值的条件分布。作为一种非参数估计方法,其具有受模型参数影响小、鲁棒性强、计算量较少的优点,适用于短期风电功率概率密度预测。

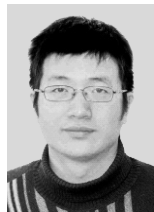
概率区间的可靠性和清晰度是相互矛盾的评价指标,不可能有同时获得高可靠性和高清晰度的预测模型。下一步将研究评价指标优化模型,使得概率区间输出在具有较好可靠性的条件下,具有更窄的区间宽度。

参考文献:

- [1] SOMAN S S, ZAREIPOUR H, MALIK O, et al. A review of wind power and wind speed forecasting methods with different time horizons [C] // North American Power Symposium. Arlington, TX, USA: IEEE, 2010: 1-8.
- [2] BLONBOU R. Very short-term wind power forecasting with neural networks and adaptive Bayesian learning [J]. Renewable Energy, 2011, 36(3): 1118-1124.
- [3] 甘迪, 柯德平, 孙元章, 等. 考虑爬坡特性的短期风电功率概率预测[J]. 电力自动化设备, 2016, 36(4): 145-150.
GAN Di, KE Deping, SUN Yuanzhang, et al. Short-term probabilistic wind power forecast considering ramp characteristics [J]. Electric Power Automation Equipment, 2016, 36(4): 145-150.
- [4] 朱思萌, 杨明, 韩学山, 等. 多风电场短期输出功率的联合概率密度预测方法[J]. 电力系统自动化, 2014, 38(19): 8-15.
ZHU Simeng, YANG Ming, HAN Xueshan, et al. Joint probability density forecast of short-term multiple wind farms output power [J]. Automation of Electric Power Systems, 2014, 38(19): 8-15.
- [5] 凌武能, 杭乃善, 李如琦. 基于云支持向量机模型的短期风电功率预测[J]. 电力自动化设备, 2013, 33(7): 34-38.
LING Wuneng, HANG Naishan, LI Ruqi. Short-term wind power forecasting based on cloud SVM model [J]. Electric Power Automation Equipment, 2013, 33(7): 34-38.
- [6] YAN J, LIU Y, HAN S, et al. Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine [J]. Renewable & Sustainable Energy Reviews, 2013, 27(6): 613-621.
- [7] 范磊, 卫志农, 李慧杰, 等. 基于变分模态分解和蝙蝠算法-相关向量机的短期风速区间预测[J]. 电力自动化设备, 2017, 37(1): 93-100.
FAN Lei, WEI Zhinong, LI Huijie, et al. Short-term wind speed interval prediction based on VMD and BA-RVM algorithm [J]. Electric Power Automation Equipment, 2017, 37(1): 93-100.
- [8] 江岳春, 张丙江, 邢方方, 等. 基于混沌时间序列 GA-VNN 模型的超短期风功率多步预测[J]. 电网技术, 2015, 39(8): 2160-2166.
JIANG Yuechun, ZHANG Bingjiang, XING Fangfang, et al. Super-short-term multi-step prediction of wind power based on GA-VNN model of chaotic time series [J]. Power System Technology, 2015, 39(8): 2160-2166.
- [9] OSÓRIO G J, MATIAS J C O, CATALÃO J P S. Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information [J]. Renewable Energy, 2015, 75: 301-307.
- [10] SHI J, GUO J, ZHENG S. Evaluation of hybrid forecasting approaches for wind speed and power generation time series [J]. Renewable & Sustainable Energy Reviews, 2012, 16(5): 3471-3480.
- [11] 张亚超, 刘开培, 秦亮. 基于VMD-SE和机器学习算法的短期风电功率多层次综合预测模型[J]. 电网技术, 2016, 40(5): 1334-1340.
ZHANG Yachao, LIU Kaipei, QIN Liang. Short-term wind power multi-leveled combined forecasting model based on variational mode decomposition-sample entropy and machine learning algorithms [J]. Power System Technology, 2016, 40(5): 1334-1340.
- [12] LI S, WANG P, GOEL L. Wind power forecasting using neural network ensembles with feature selection [J]. IEEE Transactions on Sustainable Energy, 2015, 6(4): 1447-1456.
- [13] 叶林, 刘鹏. 基于经验模态分解和支持向量机的短期风电功率组合预测模型[J]. 中国电机工程学报, 2011, 31(31): 102-108.
YE Lin, LIU Peng. Combined model based on EMD-SVM for short-term wind power prediction [J]. Proceedings of the CSEE, 2011, 31(31): 102-108.
- [14] ZANG H, FAN L, GUO M, et al. Short-term wind power interval forecasting based on an EEMD-RT-RVM model [J]. Advances in Meteorology, 2016(3): 1-10.
- [15] MORALES J M, CONEJO A J, MADSEN H, et al. Integrating renewables in electricity markets [M]. [S.l.]: Springer, 2014: 20-26.
- [16] 李智, 韩学山, 杨明, 等. 基于分位点回归的风电功率波动区间分析[J]. 电力系统自动化, 2011, 35(3): 83-87.
LI Zhi, HAN Xueshan, YANG Ming, et al. Wind power fluctuation interval analysis based on quantile regression [J]. Automation of Electric Power Systems, 2011, 35(3): 83-87.
- [17] 杨锡运, 关文渊, 刘玉奇, 等. 基于粒子群优化的核极限学习机模型的风电功率区间预测方法[J]. 中国电机工程学报, 2015, 35(S1): 146-153.
YANG Xiyun, GUAN Wenyuan, LIU Yuqi, et al. Prediction intervals forecasts of wind power based on PSO-KELM [J]. Proceedings of the CSEE, 2015, 35(S1): 146-153.
- [18] 杨明, 朱思萌, 韩学山, 等. 风电场输出功率的多时段联合概率密度预测[J]. 电力系统自动化, 2013, 37(10): 23-28.
YANG Ming, ZHU Simeng, HAN Xueshan, et al. Joint probability density forecast for wind farm output in multi-time-interval [J]. Automation of Electric Power Systems, 2013, 37(10): 23-28.
- [19] GILLES J. Empirical wavelet transform [J]. IEEE Transactions on Signal Processing, 2013, 61(16): 3999-4010.
- [20] 陈浩, 郭军海, 齐巍. 基于经验小波变换的目标加速度估计算法[J]. 北京航空航天大学学报, 2015, 41(1): 154-159.

- CHEN Hao, GUO Junhai, QI Wei. Estimation of target's acceleration based on empirical wavelet transform[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(1): 154-159.
- [21] 何耀耀, 许启发, 杨善林, 等. 基于 RBF 神经网络分位数回归的电力负荷概率密度预测方法[J]. 中国电机工程学报, 2013, 33(1): 93-98.
- HE Yaoyao, XU Qifa, YANG Shanlin, et al. A power load probability density forecasting method based on RBF neural network quantile regression[J]. Proceedings of the CSEE, 2013, 33(1): 93-98.
- [22] NICOLAI M. Quantile regression forests[J]. Journal of Machine Learning Research, 2006, 7(2): 983-999.
- [23] 方匡南. 随机森林组合预测理论及其在金融中的应用[M]. 厦门: 厦门大学出版社, 2012: 89-99.
- [24] 周松林, 茆美琴, 苏建徽. 风电功率短期预测及非参数区间估计[J]. 中国电机工程学报, 2011, 31(25): 10-16.
- ZHOU Songlin, MAO Meiqin, SU Jianhui. Short-term forecasting of wind power and non-parametric confidence interval estimation[J]. Proceedings of the CSEE, 2011, 31(25): 10-16.
- [25] 刘瑞叶, 黄磊. 基于动态神经网络的风电场输出功率预测[J]. 电力系统自动化, 2012, 36(11): 19-22.
- LIU Ruiye, HUANG Lei. Wind power forecasting based on dynamic neural networks[J]. Automation of Electric Power Systems, 2012, 36(11): 19-22.
- [26] WAN C, XU Z, PINSON P, et al. Optimal prediction intervals of wind power generation[J]. IEEE Transactions on Power Systems, 2014, 29(3): 1166-1174.

作者简介:



孙国强

孙国强(1978—),男,江苏江阴人,副教授,博士,主要研究方向为电力系统运行分析与控制、可再生能源发电系统(**E-mail**: hhusunguoqiang@163.com);

梁 智(1991—),男,江苏睢宁人,硕士研究生,主要研究方向为电力系统负荷预测、新能源发电预测(**E-mail**: liangzhi_hhu@163.com);

俞娜燕(1981—),女,江苏无锡人,工程师,硕士,主要研究方向为售电及能源管理(**E-mail**: 475881673@qq.com);

倪晓宇(1980—),男,江苏无锡人,高级工程师,硕士,主要研究方向为售电及能源管理(**E-mail**: nixiaoyu@smartwuxi.com)。

Short-term wind power probability density forecasting based on EWT and quantile regression forest

SUN Guoqiang¹, LIANG Zhi¹, YU Nayan², NI Xiaoyu³, WEI Zhinong¹, ZANG Haixiang¹, ZHOU Yizhou¹

(1. College of Energy and Electrical Engineering, Hohai University, Nanjing 210098, China;

2. State Grid Wuxi Power Supply Company, Wuxi 214061, China;

3. Wuxi Yang Sheng Technology Co., Ltd., Wuxi 214106, China)

Abstract: Probability density forecasting can give more information about future wind power output such as possible fluctuation ranges, probability and uncertainty of predicted values and so on. A combined probability density forecasting model for short-term wind power is proposed based on EWT(Empirical Wavelet Transform) and quantile regression forest. Firstly, a new adaptive signal processing method namely empirical wavelet transform is adopted to decompose the original wind power sequence into a series of empirical modes, whose frequency characteristics are different from each other. Then, the quantile regression forest forecasting model is established respectively for each empirical mode, obtaining the predictive results under arbitrary quantiles, and the final wind power forecasting values are obtained by summing up the predictive results of different empirical modes. Finally, the kernel density estimation is applied for the conditional distribution to obtain the probability density forecasting results at any moment. Simulative results verify the effectiveness of the proposed model.

Key words: empirical wavelet transform; quantile regression forest; kernel density estimation; probability density; short-term wind power forecasting; models