

# 基于因子分析和神经网络分位数回归的 月度风电功率曲线概率预测

李丹<sup>1</sup>, 任洲洋<sup>1</sup>, 颜伟<sup>1</sup>, 朱继忠<sup>2</sup>, 赵霞<sup>1</sup>, 余娟<sup>1</sup>

(1. 重庆大学电气工程学院, 重庆市 沙坪坝区 400000; 2. 南方电网科学研究院, 广东省 广州市 510080)

## Month-ahead Wind Power Curve Probabilistic Prediction Based on Factor Analysis and Quantile Regression Neural Network

LI Dan<sup>1</sup>, REN Zhouyang<sup>1</sup>, YAN Wei<sup>1</sup>, ZHU Jizhong<sup>2</sup>, ZHAO Xia<sup>1</sup>, YU Juan<sup>1</sup>

(1. School of Electrical Engineering, Chongqing University, Shapingba District, Chongqing 400000, China;

2. Electric Power Research Institute, CSG, Guangzhou 510080, Guangdong Province, China)

**ABSTRACT:** Concerning the existing problems such as large number of highly correlated variables, few available weather information and high uncertainty in month-ahead wind power curve prediction, a new month-ahead wind power curve probabilistic prediction method based on factor analysis and quantile regression neural network (QRNN) was proposed. By means of the factor analysis method to reduce the dimensionality of intraday hourly wind power time series vector, the independent common factors were extracted and used as the predictor variables to separately build QRNN probabilistic forecasting models with the daily weather features as inputs. Based on these probabilistic forecasting models, the probability density functions of wind power common factors of next 30 days were predicted by inputting the daily weather features from medium-range weather forecasts. By simulating common factors and specific factors follow the predicted distributions to recover wind power forecast curves day by day, statistical scenarios of wind power production in next month were generated finally. The accuracy, adaptability and high efficiency of the proposed method have been verified by the forecasting results of two actual wind farms, which can provide a new feasible solution for the probabilistic medium-and long-term wind power forecasting.

**KEY WORDS:** month-ahead wind power prediction; factor analysis; quantile regression neural network; medium-range weather forecasting; probabilistic prediction

**基金项目:** 国家自然科学基金项目(51177178, 51677012); 重庆市科委基础与前沿研究计划项目(cstc2013jcyjA90001)。

The National Natural Science Foundation of China (51177178 & 51677012); The Foundation and Frontier Research Project of Chongqing Science and Technology Commission (cstc2013jcyjA90001).

**摘要:** 针对月度风电曲线预测存在的预测变量多且关系复杂、可利用天气信息少以及不确定性强等问题,提出了一种基于因子分析和神经网络分位数回归的月度风电曲线概率预测方法。采用因子分析对日内小时级风电功率序列向量降维,提取出相互独立的风电公共因子作为预测变量,分别建立以日天气特征为输入的神经网络分位数条件概率模型;利用中期天气预报信息,预测未来30日各公共因子的概率分布;最后通过模拟服从预测分布的风电公共因子和各时刻特殊因子,并代入因子模型逐日还原风电预测曲线,生成未来月风电曲线的随机场景。两个实际风电场的预测结果验证了所提风电曲线概率预测方法的准确性、适应性和高效性,为中长期风电功率概率预测提供了一种可行的解决思路。

**关键词:** 月度风电功率预测; 因子分析; 神经网络分位数回归; 中期天气预报; 概率预测

## 0 引言

随着风电占比不断提高,风电本身的不确定性和波动性使电力系统随机性不断增强,作为应对手段之一,提前对风电做出准确预测尤为重要<sup>[1]</sup>。风电预测按时间尺度分为超短期、短期、中期和长期预测,其中,以月度为代表的中期风电预测能为月度发电调度和交易计划安排提供可靠的依据。然而相对于短期和超短期风电预测,目前中期风电预测研究明显不足,且预测精度较低,预测结果多为月平均或日平均功率,无法提供风电功率的小时级时序波动信息,影响了预测结果在实际电力调度和交易中的应用<sup>[2]</sup>。

目前实现月度风电功率曲线的准确预测面临三大挑战。一是预测变量多且相关性强。相对于未

来 24 小时的日前风电预测, 月度风电预测的预测变量增加了 30 倍, 而且各时刻风电功率之间具有时序相关性, 直接建模会导致“维数灾”而难以求解。针对这一问题, 文献[3]将经济学中在长期预测领域应用效果较好的动态因子模型(dynamic factor model, DFM)应用于风电预测, 利用风功率波动具有的日周期性, 将每日 24 维小时级风电功率序列降维成 2~3 维动态因子线性组合叠加上高斯白噪声的随机过程, 采用时间序列方法预测风电公共因子, 从而实现风电功率曲线的确定性预测。该方法较好的反映了日内各时刻风电功率间的时序相关性, 并将预测提前时间延长到 1~2 个月, 预测误差下降到 18%~20%。但该文献只利用风电功率的历史样本信息, 属于统计预测方法, 预测精度进一步提高的空间有限。

二是可利用天气信息少。众所周知, 风电受天气因素特别是风速和风向影响很大, 如果能获得预测期间的天气信息将大大改善风电预测精度。如文献[4]采用局部回归神经网络以邻近风电场多个地点的风速和风向数值天气预报(numerical weather prediction, NWP)信息为模型输入, 使提前 72 小时的风电预测误差从 16%下降到 12%。然而目前 NWP 只能精确预测未来 48 小时内的风速和风向变化, 无法用于月度风电预测。但日天气特征如平均风速、主导风向、天气类型和日均气温等, 我国中期天气预报已能实现未来 30 天预测<sup>[5]</sup>, 这使得将中期天气预报信息补充到月度风电预测的可利用信息中, 进一步提高预测精度成为可能。

三是风电功率不确定性强。传统确定性预测方法很难全面刻画风电波动的不确定性, 相比之下风电功率预测<sup>[6-7]</sup>不仅能提供预测结果的波动范围, 还能估计各取值出现的概率, 为电力系统调度决策以及运行风险评估提供更丰富的信息<sup>[8]</sup>。目前风电功率预测方法有贝叶斯学习法<sup>[9]</sup>、区间估计法<sup>[10]</sup>、核密度估计法<sup>[11-12]</sup>和分位数回归法<sup>[13]</sup>等。其中, 分位数回归概率预测方法不需要先验分布假设, 而且能提供稳定的预测信息, 在负荷和新能源出力的概率预测中得到了广泛应用<sup>[13-15]</sup>。如文献[13]建立了基于支持向量机和分位数回归的短期风电功率区间预测模型, 实现了日前风功率的波动区间分析。但目前还未见分位数回归方法在中期风电功率预测中的应用。

基于以上分析, 本文提出一种基于因子分析和

神经网络分位数回归的月度风电功率曲线概率预测方法。针对风电功率曲线所具有的波动性和时序相关性, 采用因子分析技术对 24 维的日内小时级风电功率序列向量降维, 用因子荷载矩阵解释不同时刻风电功率的时序相关性, 少数相互独立的公共因子向量解释日内风电功率共性信息; 然后, 将公共因子作为月度风电曲线预测的可预测变量, 基于日天气特征向量与公共因子值概率分位数之间的非线性关系, 建立神经网络分位数回归模型, 以中期天气预报提供的未来 30 日天气特征预测值为输入量, 预测各日的公共因子值分位数序列, 并通过非参数核密度方法估计公共因子的连续概率分布曲线, 体现风电公共因子预测的不确定性; 最后, 依据公共因子预测概率分布逐日对风电公共因子随机抽样, 抽样值经过因子还原并叠加各时刻特殊因子的随机值, 最终实现未来月风电功率曲线的随机模拟。将本文方法应用于两个装机容量不同的实际风电场月度风电曲线概率预测, 预测结果表明, 引入中期天气预报信息有助于改善预测精度; 将因子分析技术和神经网络分位数概率预测方法相结合, 不仅能在保证概率预测精度的条件下获得更高的预测效率, 而且生成的月度风电曲线场景能准确反映实际风电功率的随机性、波动性和时序相关性, 为调度人员提供更准确、全面的风电预测信息。

## 1 日内风电功率曲线的因子分析模型

本节在将原始日风电功率曲线标准化和面板处理的基础上, 通过因子分析降维技术提取出表征日内风电功率共性特征的  $r$  维潜在公共因子向量, 代替 24 维原始风电功率向量, 达到减少预测变量个数、降低预测难度的目的。

### 1.1 24 维标准化风电功率序列向量

采用面板数据方法(panel approach)将风电功率表示为 24 维的风电功率标准化序列向量  $\mathbf{P}^{[3]}$ , 分别对应 24 时刻的风电功率时间序列。

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_{24} \end{bmatrix} = \begin{bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,d} & \cdots & P_{1,D} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,d} & \cdots & P_{2,D} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ P_{i,1} & P_{i,2} & \cdots & P_{i,d} & \cdots & P_{i,D} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ P_{24,1} & P_{24,2} & \cdots & P_{24,d} & \cdots & P_{24,D} \end{bmatrix} \quad (1)$$

式中:  $D$  为样本观测日的总天数;  $\mathbf{P}_i$  为  $i$  时刻标准风电功率逐日变化的时间序列;  $P_{i,d}$  为第  $d$  天第  $i$

时刻的风电功率标准化值,它与风电功率原始值  $\tilde{P}_{i,d}$  间的变换公式如下:

$$\begin{cases} P_{i,d} = \frac{\tilde{P}_{i,d} - \mu_i}{S_i} \\ \mu_i = \frac{1}{D} \sum_{d=1}^D \tilde{P}_{i,d} \\ S_i = \sqrt{\frac{1}{D-1} \sum_{d=1}^D (\tilde{P}_{i,d} - \mu_i)^2} \end{cases} \quad (2)$$

式中  $\mu_i$  和  $S_i$  分别表示  $i$  时刻原始风电功率的样本均值和样本标准差。

## 1.2 因子分析模型

因子分析是一种通过降维技术化简多维向量的多元统计方法,属于主成分分析的推广<sup>[16]</sup>。其基本思想是通过分析多变量数据的相关关系,找到支配这种相关关系的少数几个相关独立的潜在因子,达到简化观测数据、用少数变量解释研究复杂问题的目的。在电力系统中,它常被用于实时电价等具有复杂相关关系多维向量的降维化简<sup>[17-18]</sup>。具有日周期特性的小时级风电功率时间序列,不仅序列变量数多达 24 维,而且不同时刻的风电功率变量之间有很强的时序相关性,采用因子分析可以减少向量的维度,降低问题的复杂度。

设标准化后的 24 维风电功率总体为  $\mathbf{P} = [P_1, P_2, \dots, P_i, \dots, P_{24}]^T$ , 其因子分析的一般模型为

$$\mathbf{P} = \mathbf{A}\mathbf{F} + \boldsymbol{\varepsilon} \quad (3)$$

式中:  $\mathbf{F} = [f_1, f_2, \dots, f_r]^T$  为公共因子向量;  $\mathbf{A} = (a_{ij})_{24 \times r}$  为因子荷载矩阵,解释了变量  $P_i$  之间的相关关系,  $a_{ij}$  为变量  $P_i$  在公共因子  $f_j$  上的荷载,它反映了公共因子  $f_j$  对变量  $P_i$  的重要程度。模型中公共因子向量与因子荷载矩阵的乘积  $\mathbf{A}\mathbf{F}$  称为公共分量,包含了日风电功率曲线的共性信息。 $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{24}]^T$  为特殊因子向量,代表各时刻风电功率不能被公共因子解释的部分,且假设  $E(\boldsymbol{\varepsilon}) = \mathbf{0}_{24 \times 1}$ ,  $\text{var}(\boldsymbol{\varepsilon}) = \mathbf{D} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_{24}^2)$ 。公共因子和特殊因子彼此独立,为不可观测的隐变量。

求解因子模型式(3)的关键是估计因子荷载阵  $\mathbf{A}$  和特殊方差阵  $\mathbf{D}$ , 这里采用主成分估计法:

$$\hat{\mathbf{A}} = [\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1, \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2, \dots, \sqrt{\hat{\lambda}_r} \hat{\mathbf{e}}_r] = (\hat{a}_{ij})_{24 \times r} \quad (4)$$

$$\hat{\mathbf{D}} = \text{diag}(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_{24}^2) \quad (5)$$

$$\hat{\sigma}_i^2 = s_{ii} - \sum_{j=1}^r \hat{a}_{ij}^2, \quad i = 1, 2, \dots, 24 \quad (6)$$

式中:  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_r \geq 0$  为样本协方差矩阵  $\mathbf{S}_{24 \times 24}$  前  $r$  个最大特征值;  $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \dots, \hat{\mathbf{e}}_r$  为相应的正交单位特征向量;  $\hat{\sigma}_i^2$  为  $i$  时刻特殊因子方差估计值;  $s_{ii}$  为样本协方差矩阵  $\mathbf{S}$  中第  $i$  个对角线元素;  $\hat{a}_{ij}$  为第  $i$  行  $j$  列的荷载矩阵元素估计值。

基于历史观测样本,由式(4)~(6)得到因子荷载阵  $\mathbf{A}$  和特殊方差阵  $\mathbf{D}$  估计值后,可计算样本对应的公共因子估计向量  $\hat{\mathbf{F}}$ , 也称为公共因子得分:

$$\hat{\mathbf{F}}_d = \hat{\mathbf{A}}' \mathbf{S}^{-1} \mathbf{P}_d \quad (7)$$

式中:  $\hat{\mathbf{F}}_d$  为  $r \times 1$  的第  $d$  日公共因子估计向量;  $\mathbf{S}$  为  $24 \times 24$  的样本协方差矩阵;  $\mathbf{P}_d$  为  $24 \times 1$  的第  $d$  日风电功率标准值向量。

因子分析模型中的公共因子维数  $r$  通常根据公共因子重要性程度来确定。第  $j$  个公共因子  $f_j$  对 24 时刻风电功率总方差的贡献是衡量公共因子重要性的一个度量,可用样本协方差矩阵  $\mathbf{S}$  的第  $j$  个最大特征值  $\hat{\lambda}_j$  来近似。当排名靠前的  $r$  个主成分的累积贡献率  $\sum_{j=1}^r \hat{\lambda}_j / \sum_{i=1}^{24} \hat{\lambda}_i$  达到一个较高的水平(一般 85%以上)时,即可近似认为前  $r$  个公共因子已经能够基本解释原始数据的变化,按照这个原则选择的  $r$  往往远小于原始向量  $\mathbf{P}$  的维数,一般为 2 或 3。

按照以上方法,以重庆某风电场 2013—2014 年观测样本为例,对日风电功率曲线进行因子分析,前 3 个主成分的贡献率分别为 58.3%、26.6% 和 8.4%,按照累计贡献率大于 85% 的原则选择  $r=3$  (公共因子累计贡献率达到 93.28%)。图 1 给出了 3 个公共因子对应的估计荷载  $a_{ij}(i=1,2,\dots,24, j=1,2,3)$  的图形化表示。

图 1 中公共因子 1 对 24 时刻风电功率的荷载为较稳定的正值;而公共因子 2、3 在不同时段荷

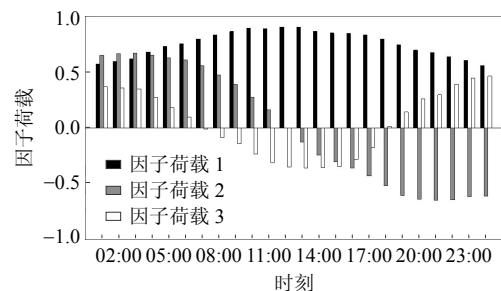


图 1 观测样本的公共因子荷载  
Fig. 1 Loads for three common factors of observed samples

载的正负以及具体值都有较大差别, 其中公共因子 3 的日内风电功率荷载波动频率更高。

为更详细地解释不同公共因子和特殊因子的物理意义, 图 2 以同一实例的 2013 年 1 月 1 日—3 日为例, 分别给出了由各公共因子所解释的公共分量部分和特殊因子部分对比, 从上到下依次为总公共分量曲线与实际风电曲线对比、由公共因子 1 解释的公共分量、由公共因子 2 解释的公共分量、由公共因子 3 解释的公共分量和特殊因子分量曲线。各公共因子解释的公共分量由对应公共因子值乘上对应因子荷载得到。

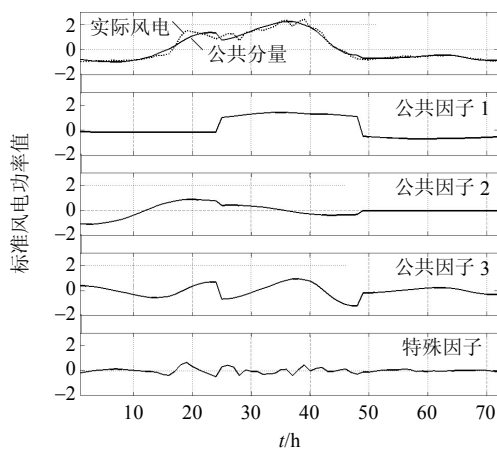


图 2 观测样本的公共分量与特殊分量

Fig. 2 Common components and specific components of observed samples

从图 2 中可以看出, 因子 1 公共分量曲线的值在日内基本保持稳定, 体现了日内风电功率的平均水平; 而因子 2、3 的公共分量在日内具有不同的曲线形状和波动频率, 它们反映出风电功率的日内波动特征。由三个因子的公共分量叠加得到的总公共分量曲线与原始风电曲线非常接近, 它们之间的误差即为特殊分量曲线, 基本在 0 值附近波动, 大部分低于实际风电功率的 10%。经计算各公共因子之间协方差和分析各时刻特殊分量的概率分布, 结果表明 3 维公共因子之间彼此独立, 而各时刻特殊分量近似满足均值为 0 的正态分布。

以上因子分析的实例表明, 用少数相互独立的  $r$  维风电公共因子代替原始 24 维风电功率作为预测变量, 能在保留原始风电功率数据最大共性信息量的条件下, 减少预测变量数, 简化变量间关系。而且, 相比完全随机的特殊分量, 反映日风电曲线共性特征的公共因子与日天气特征之间关系密切, 更具可预测性。

## 2 风电公共因子神经网络分位数回归预测

### 2.1 神经网络分位数回归模型表示

分位数回归依据被解释变量  $Y$  的条件分位数对输入变量  $X$  进行回归, 能更细致地反映出输入变量  $X$  在不同范围内对被解释变量  $Y$  位置、散布、形状影响的信息<sup>[19]</sup>。它具有以下优点: 首先, 它对模型中的随机扰动残差项不做任何分布的假定, 回归模型的稳健性优于最小二乘法回归; 其次, 分位数回归是对所有分位数进行回归, 对数据中出现的异常点或奇异点具有耐抗性, 可以克服非平稳变动, 从而更好地反映响应变量; 最后, 分位数回归估计出来的参数, 具有在大样本理论下的渐进优良性<sup>[20]</sup>。

2000 年, Taylor 提出了神经网络分位数回归模型(quantile regression neural network, QRNN)<sup>[21]</sup>, 利用神经网络的非线性核函数, 分析解释变量  $x_1$ 、 $x_2$ 、 $\dots$ 、 $x_K$  对响应变量  $y_m$  复杂的非线性影响。QRNN 模型使用三层 BP 感知器神经网络, 解释变量为输入层, 响应变量为输出层, 隐含层包含  $J$  个节点, 其模型表示为

$$Q_m(\tau) = f^{(o)}\left(\sum_{j=1}^J \omega_j^{(o)}(\tau) g_j(\tau) + b^{(o)}(\tau)\right) \quad (8)$$

式中:  $\tau$  为分位点;  $Q_m(\tau)$  为  $y_m$  的  $\tau$  条件分位数;  $\omega_j^{(o)}$  为输出层权重;  $b^{(o)}(\tau)$  为输出层偏移;  $f^{(o)}(\cdot)$  为输出层转换函数;  $g_j(\tau)$  为隐含层节点, 满足:

$$g_j(\tau) = f^{(h)}\left(\sum_{k=1}^K \omega_{jk}^{(h)}(\tau) x_k + b_j^{(h)}(\tau)\right) \quad (9)$$

式中:  $\omega_{jk}^{(h)}$  为隐含层权重;  $b_j^{(h)}(\tau)$  为隐含层偏移;  $f^{(h)}(\cdot)$  为隐含层转换函数。

特别地, 当隐含层节点数  $J=1$ , 转换函数  $f^{(h)}(\cdot)$  和  $f^{(o)}(\cdot)$  均为等值函数时, QRNN 模型退化为线性分位数回归模型<sup>[20]</sup>。而神经网络模型就是神经网络分位数回归模型在  $\tau=0.5$  时的特例形式。

本文以日天气特征因子为解释变量  $\mathbf{X}=[x_1, x_2, \dots, x_K]$ , 分别对  $r$  个公共因子  $\mathbf{Y}=[y_1, y_2, \dots, y_r]$  做条件分位数回归。由于不同天气因子有不同的量纲和取值范围, 本文先采用指标映射法将天气特征  $x_1$ 、 $x_2$ 、 $\dots$ 、 $x_K$  标准化到  $[0,1]$  区间, 标准化方法详见文献[22]。回归模型的隐含层转换函数  $f^{(h)}(\cdot)$  选择双曲正切的  $\tanh$  函数, 输出层转换函数  $f^{(o)}(\cdot)$  选择等值函数, 因此从输入层到输出层之间表现出如下的非线性关系:

$$Q_{y_m}(\tau|X) = f\left(\sum_{j=1}^J \tanh\left(\sum_{k=1}^K \omega_{jk}^{(h)}(\tau)x_k + b_j^{(h)}(\tau)\right) \cdot \omega_j^{(o)}(\tau) + b_0^{(o)}(\tau)\right), \quad m=1,2,\dots,r \quad (10)$$

模型中待估计参数统一用参数向量  $\theta(\tau)$  表示:

$$\theta(\tau) = \{\omega_{11}^{(h)}(\tau), \dots, \omega_{JK}^{(h)}(\tau), b_1^{(h)}(\tau), \dots, b_J^{(h)}(\tau), \omega_1^{(o)}(\tau), \dots, \omega_J^{(o)}(\tau), b_0^{(o)}(\tau)\}$$

由于不同公共因子所表征的日风电功率特征不同,其公共因子值所呈现的取值区间也有区别:如表征日风电功率平均水平的公共因子1具有下边界  $l$ ,对应整日风电功率为0时公共因子1的值;而表征日风电功率形状特征的其它公共因子,取值区间则呈现出以0为中心的对称分布。Cannon指出:对具有下边界的响应变量和无边界的响应变量,应采用不同的 QRNN<sup>[23]</sup>。其差别在于模型式(10)中的  $f(\cdot)$  函数不同,当响应变量  $y_m$  无边界时,  $f(\cdot)$  为等值函数,而响应变量  $y_1$  有下边界  $l$  时,  $f(\cdot)$  替换为下面的激活函数  $r(\cdot)$ 。

$$r(u) = \begin{cases} u, & u \geq l \\ l, & u < l \end{cases} \quad (11)$$

## 2.2 模型参数估计

对 QRNN 模型式(10)中参数  $\theta(\tau)$  的估计,旨在最小化如下形式的损失函数:

$$E_{y_m\tau} = \frac{1}{T} \sum_{t=1}^T [\tau - I(y_{mt} - Q_{y_{mt}}(\tau))](y_{mt} - Q_{y_{mt}}(\tau)) \quad (12)$$

式中:  $T$  为样本量;  $y_{mt}$  表示第  $t$  个样本日的公共因子  $m$  值;  $Q_{y_{mt}}(\tau)$  表示将第  $t$  个样本日的天气因子向量  $X_t$  带入模型(10)得到的公共因子  $m$  的  $\tau$  条件分位数;  $I(\cdot)$  为指示函数:

$$I(u) = \begin{cases} 0, & u \geq 0 \\ 1, & u < 0 \end{cases} \quad (13)$$

为避免 QRNN 模型陷入过度拟合,可在目标函数中增加一个惩罚项<sup>[23]</sup>。此时,QRNN 模型的参数估计转化为下列优化问题:

$$\hat{\theta}(\tau; \lambda, J) = \arg \min_{\theta} \{E_{y_m\tau} + \lambda \frac{1}{KJ} \|\omega^{(h)}(\tau)\|_2\} \quad (14)$$

式中:  $\omega^{(h)}(\tau) = [\omega_{11}^{(h)}(\tau), \omega_{12}^{(h)}(\tau), \dots, \omega_{JK}^{(h)}(\tau)]^T$  为隐含层权重向量;  $\|\cdot\|_2$  为范数;  $\lambda$  为惩罚参数。

实例中,惩罚参数  $\lambda$  和隐含层节点数目  $J$  的最优取值可以通过如下的 AIC 准则来实现:

$$AIC(\lambda, J) = \min(\ln(E_{y_m}(\tau; \theta(\tau; \lambda, J))) + \frac{K(J+1)}{T}) \quad (15)$$

式中  $E_{y_m}(\tau; \theta(\tau; \lambda, J))$  为选择不同  $\lambda$  和  $J$  时  $\tau$  分位数下的最小损失函数。

根据 AIC 准则筛选出惩罚参数  $\lambda$  和隐含层节点数  $J$  的最优取值,代入式(14)计算使损失函数最小的最优估计参数  $\hat{\theta}(\tau; \lambda, J)$ ,由于损失函数式(12)中的指示函数  $I(\cdot)$  和有界模型中的激活函数  $r(\cdot)$  在原点处不可微,传统基于梯度的非线性优化算法并不适合直接优化式(14)。文献[23]提出先采用处处可微的 Huber 范数平滑近似函数来近似代替  $I(\cdot)$  和  $r(\cdot)$ ,再用非线性优化方法求解式(14)。

## 2.3 公共因子的概率密度预测

将求得的最优估计参数  $\hat{\theta}(\tau; \lambda, J)$  代入式(10),输入中期天气预报提供的未来30日天气特征预测值  $X_n = [x_1, x_2, \dots, x_K]_n$  ( $n=1,2,\dots,30$ ),即可得到对应日中公共因子  $y_m$  ( $m=1,2,\dots,r$ ) 的条件分位数预测值  $\hat{Q}_{y_m}(\tau|X_n)$ 。当分位点  $\tau$  在(0,1)区间上连续取值时,条件分位数曲线  $\hat{Q}_{y_m}(\tau|X_n)$  即为第  $n$  待预测日公共因子  $y_m$  的条件累积概率分布曲线  $F(y_m|X_n)$ <sup>[19]</sup>。

如果分位点  $\tau$  在(0,1)区间上按等间隔连续取值,得到的条件分位数  $\hat{Q}_{y_m}(\tau|X_n)$  可看作与  $y_m$  服从同分布的一组随机抽样的样本,采用 Matlab 的非参数核密度估计函数 `ksdensity()` 可估计出公共因子  $y_m$  在输入天气特征  $X_n$  条件下的概率密度函数  $f(y_m|X_n)$ 。核密度估计常用的核函数有均匀核、三角核、Epanechnikov 核、高斯核和余弦核等。由于 Epanechnikov 核在均方误差意义下最优<sup>[24]</sup>,本文选择 Epanechnikov 核函数,窗宽采用 `ksdensity()` 函数的默认窗宽。

## 3 月度风电功率曲线随机场景模拟

由于因子分析提取的公共因子之间相互独立,可基于上章预测得到的未来30日公共因子概率分布  $f(y_m|X_n)$  ( $m=1,2,\dots,r$ ,  $n=1,2,\dots,30$ ) 进行各公共因子的独立随机抽样和生成指定个数  $L$  的月度风电功率曲线随机模拟场景。具体模拟步骤如下:

第1步: 设  $n=1,2,\dots,30$  表示待预测日的编号,重复下面步骤2—3依次生成未来月各日的风电功率曲线随机模拟场景。

第2步: 依据第  $n$  日第  $m$  个公共因子的概率密度函数  $f(y_m|X_n)$  ( $m=1,2,\dots,r$ ),依次对公共因子  $1-r$  进行随机抽样,生成  $L$  组公共因子向量样本。

第3步: 将  $L$  组公共因子向量样本代入因子分

析模型(3), 分别叠加上服从  $N(0, \hat{\sigma}_i^2)$  ( $i=1, 2, \dots, 24$ ) 正态分布的 24 时刻特殊因子随机值, 还原第  $n$  日中各时刻的标准化风电功率样本  $\mathbf{P}_n = \{p_{n,1}, p_{n,2}, \dots, p_{n,24}\}_l$  ( $l=1, 2, \dots, L$ ), 并基于各时刻风电功率的历史样本均值  $\mu_i$  和样本方差  $S_i$ , 进行反标准化变换, 获得第  $n$  日的日风电功率样本曲线  $\tilde{\mathbf{P}}_n = \{\tilde{p}_{n,1}, \tilde{p}_{n,2}, \dots, \tilde{p}_{n,24}\}_l$  ( $l=1, 2, \dots, L$ )。

第 4 步: 分别将 30 个待预测日的  $L$  组日风电功率曲线样本按照时间顺序逐日连接, 得到月度风电功率预测曲线的  $L$  个随机场景。

月度风电功率曲线随机模拟场景结果可直接用于月度发电调度和电力市场交易决策制定时的蒙特卡洛模拟, 以评估由风电随机性引发的系统安全风险和发购电成本概率分布。

## 4 算例分析

### 4.1 算例描述

选择重庆和辽宁某两个不同装机容量实际风电场数据, 对本文预测方法进行验证。其中算例 1 为装机容量 49.3MW 的重庆某风电场, 预测时段为 2015 年 1 月, 观测样本为 2013—2014 年每日天气特征和小时级风电曲线; 算例 2 为装机容量 116MW 的辽宁某风电场, 预测时段为 2013 年 3 月, 观测样本为 2012 年 3 月—2013 年 2 月每日天气特征和小时级风电曲线。经因子分析, 两算例均选择公共因子维数  $r=3$ , 以各公共因子值为响应变量依次建立 QRNN 模型, 其中公共因子 1 建立有下边界的 QRNN 模型。利用 QRNN 模型进行预测时, 输入量为来自公共天气预测网站的未来 30 日天气特征预测值<sup>[5]</sup>(包含天气类型、日均气温、风速和主导风向)。所有天气特征值和风电功率都事先进行量化和标准化处理, 其中风电功率值表示为占风电总装机容量的比值。

在对 3 个公共因子的 QRNN 模型进行训练时, 选择全部观测日样本, BP 神经网络训练迭代次数设为 1000, 选取分位点区间为 0.0001 到 0.9999, 间隔为 0.05, 通过训练样本确定各分位点下 QRNN 模型的最优结构和最优参数。其中, 根据 AIC 准则确定 3 个 BP 神经网络结构为 4-1-1, 即输入层节点数为 4, 隐含层为 1, 输出层为 1, 惩罚参数  $\lambda$  为 0.1。通过训练好的神经网络结构, 分别求出各预测日 3 个公共因子在所有分位点上的分位数, 将其代入 Epanechnikov 核密度估计函数, 估计各预测日

3 个公共因子连续的概率密度曲线(共计 90 条)。预测时考虑了部分时段风电场停机检修的影响。

### 4.2 预测精度评价指标

本文采用平均绝对误差评价月度风电功率曲线的确定性预测结果; 采用可信度、区间平均宽度和技巧得分评价概率预测结果<sup>[6]</sup>。

1) 可靠性(reliability, R)。

$$R^{(1-\alpha)} = \left[ \frac{\xi^{(1-\alpha)}}{N} - (1-\alpha) \right] \times 100\% \quad (16)$$

式中:  $R^{(1-\alpha)}$  为置信度  $1-\alpha$  下的可靠性值;  $N$  为测试样本的个数;  $\xi^{(1-\alpha)}$  为在置信度  $1-\alpha$  下实际风电功率值落入预测置信区间的个数。

2) 区间平均宽度(interval average width, IAW)。

$$I_{AW}^{(1-\alpha)} = \frac{1}{N} \sum_{n=1}^N \theta_n^{(1-\alpha)} \quad (17)$$

式中:  $I_{AW}^{(1-\alpha)}$  为置信度  $1-\alpha$  下的功率区间平均宽度;  $\theta_n^{(1-\alpha)}$  为第  $n$  个测试样本在置信度  $1-\alpha$  下功率预测区间的宽度。

3) 技能分数(skill score, SC)。

$$S_C = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^{N_\tau} (\xi^{\tau_i} - \tau_i)(y_n - \hat{q}^{\tau_i}) \quad (18)$$

$$\xi^{\tau_i} = \begin{cases} 1, & y_n < \hat{q}^{\tau_i} \\ 0, & \text{其他} \end{cases} \quad (19)$$

式中:  $S_C$  为技能分数;  $N_\tau$  为设定的分位点数目;  $\tau_i$  为分位点;  $y_n$  为第  $n$  个测试样本的实际值;  $\hat{q}^{\tau_i}$  为分位点  $\tau_i$  对应的分位数。

4) 确定性预测值的平均绝对百分比误差(normalized mean absolute error, NMAE)。

$$NMAE = \frac{1}{N} \sum_{n=1}^N \left| \frac{y_n^* - y_n}{C} \right| \times 100\% \quad (20)$$

式中:  $y_n^*$  和  $y_n$  分别为第  $n$  个测试样本的确定性预测值和实际值;  $C$  为总装机容量。

上述 4 个评价指标中, 可靠性用于评价置信区间的可信程度, 其绝对值越小, 可信程度越高; 区间平均宽度评价预测结果聚集不确定信息的能力, 其值越小越好; 技能分数是二者的综合评估, 其值越大越好; 确定性预测值的平均绝对百分比误差评价确定性预测值与实际值的偏差, 其值越小越好。

### 4.3 月度风电功率曲线的确定性预测结果

基于 QRNN 预测得到的风电公共因子概率分布, 按照文献[14]中选择密度函数的中位数为点预

测结果的建议,将公共因子预测概率分布的中位数作为未来月各日的公共因子确定性预测值,代入因子分析模型(3)并进行反标准化处理,可得到确定性的月度风电功率预测曲线。图 3 分别给出了两算例采用本文方法得到的确定性风电功率预测曲线与实际风电曲线对比。表 1 对比了文献[3]的 DFM 动态因子方法与本文方法预测风电功率曲线的 NMAE 预测误差。

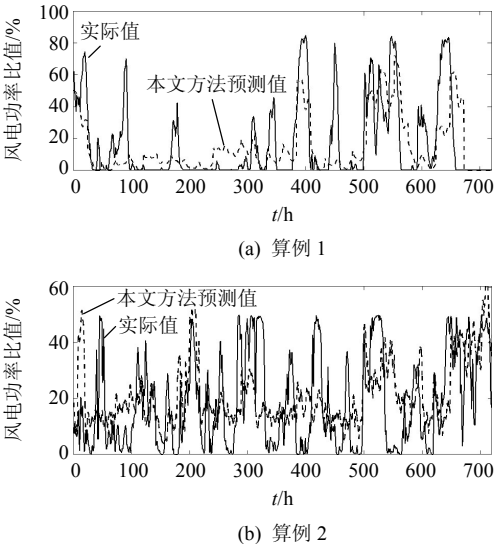


图 3 风电功率曲线的确定性预测结果  
Fig. 3 Point prediction results of wind power curve

表 1 不同预测方法的 NMAE 误差

Tab. 1 NMAE of different prediction methods %		
算例	本文方法	DFM 方法
1	13.08	20.80
2	12.91	15.98

从图 3 可以看出,本文方法得到的确定性风电预测曲线与实际曲线整体变化趋势基本一致。由于利用了未来月各日天气特征预测信息,表 1 中本文方法对两算例确定性预测结果的 NMAE 误差指标比仅利用历史风电样本信息的 DFM 方法降低了 3%~8%。

4.4 月度风电功率曲线的随机场景模拟结果

按照第 3 节方法对两个算例分别模拟生成  $8 \times 10^3$  组预测月的风电曲线随机场景,并根据各小时风电模拟样本值分布情况,统计不同置信度下的风电功率置信区间。图 4 给出了其中 20%、70%和 90%置信度下两个算例预测结果的风电功率置信区间、确定性预测值与实际值的对比。

从图 4 可以直观看出:随着概率预测置信度的增大,两算例中风电功率置信区间的范围变大,实

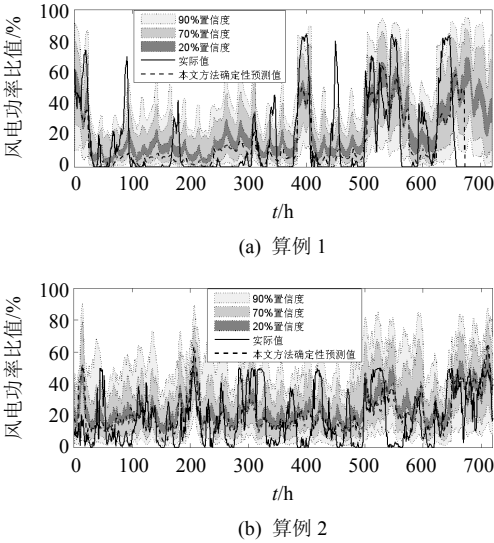
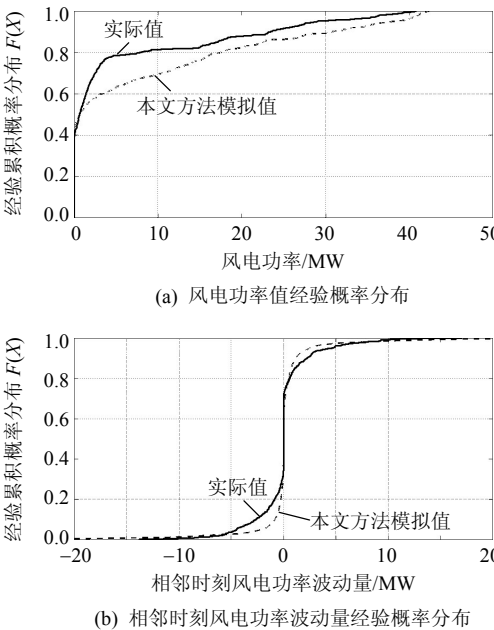


图 4 不同置信度下的风电功率曲线概率预测结果  
Fig. 4 Probabilistic forecasting results of wind power curve with different confidence levels

际值有更大的可能落在置信区间内;无论是概率置信区间还是中位数确定性预测结果,预测结果与实际风电功率曲线基本保持相似的变化趋势,实际风电功率曲线大部分落在 90%的置信区间内,体现了本文概率预测方法的有效性;另外,风电功率变化越剧烈,确定性预测结果与实际值误差也越大,概率预测置信区间的宽度也相应变大,说明本文方法在风电功率变化相对平缓时段的预测效果优于风电功率变化剧烈时段。

为验证模拟风电曲线的随机性、波动性和时序相关性与实际风电功率曲线是否一致,图 5 以算例 1 为例,给出了本文方法模拟风电功率曲线中各



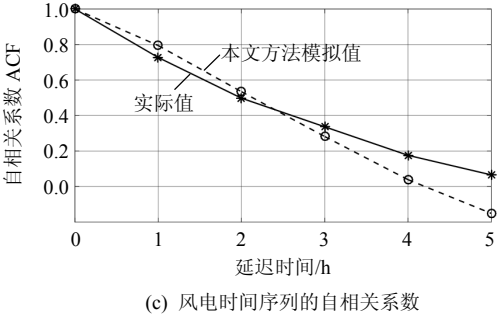


图 5 两种方法模拟风电曲线的特性对比

Fig. 5 Features comparison of the simulated wind power curves produced from two methods

时刻风电功率值和风电功率波动量(相邻时刻风电功率之差)的经验累积概率分布、以及风电时间序列的自相关函数值(auto correlation function, ACF)曲线和实际值的对比。

从图 5 来看,在各时刻风电功率值概率分布和相邻时刻风电功率波动量概率分布方面,两种方法模拟结果与实际风电功率值概率分布基本一致。在模拟风电功率时间序列的自相关系数方面,风电曲线只存在 2 小时内的显著时序相关,当延迟时间不超过 3 小时,本文方法模拟的风电功率时间序列自相关系数与实际风电曲线的自相关系数一致。以上结果说明本文方法能准确模拟实际风电功率的随机性、相邻时刻功率的波动性和短时时序相关性。

4.5 与分时刻风电功率概率预测方法对比分析

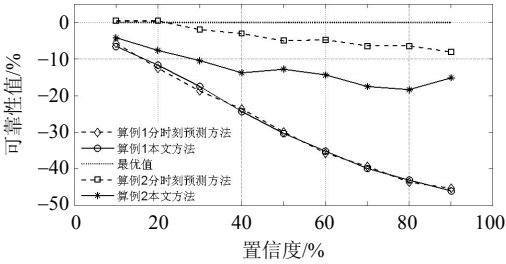
为进一步评价因子分析降维化简技术在本文方法中的作用,将本文方法与直接对各时刻风电功率进行独立概率预测的分时刻预测方法进行对比研究。

为具有可比性,分时刻预测方法的输入量同样选择预测日天气特征,QRNN 模型结构和参数估计方法也与本文方法相同,并采用文献[25]提出的正态分布反变换法在风电功率模拟曲线中体现不同时刻风电功率间的相关关系。表 2 列出了两种方法在预测模拟时间、技能分数和确定性预测值 NMAE 误差方面的对比。其中预测模拟时间采用 Matlab 程序中定义时间函数的方法进行监测(使用酷睿 i5

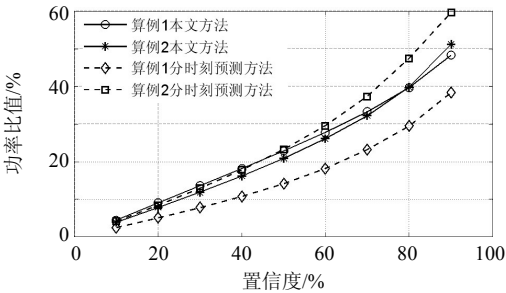
表 2 两种概率预测方法的预测结果

Tab. 2 Results of two probabilistic prediction methods				
算例	方法	预测时间/s	技能分数 $S_C$	NMAE/%
1	本文方法	464	-1.191	13.08
	分时刻方法	8132	-1.038	15.78
2	本文方法	318	-1.016	12.91
	分时刻方法	5044	-1.285	15.72

双核 2.4GHz, 内存 4G 计算机)。图 6 为两种方法在各置信度下的可靠性和区间平均宽度对比。



(a) 可靠性结果



(b) 区间平均宽度结果

图 6 两种方法的概率预测评价指标对比

Fig. 6 Probability indexes comparison of two methods

从图 6 中可以看出,在可靠性指标方面,算例 1 中两种方法基本相同,但算例 2 中本文方法比分时刻预测方法略差;而且随着置信度增加,两种方法预测可靠程度均有不同程度的下降。在区间平均宽度方面,本文方法在算例 1 中的预测结果比分时刻预测方法略差,而算例 2 的预测结果则略优。表 2 中综合技能分数方面,本文方法在算例 1 中略差于分时刻预测方法,但算例 2 比分时刻预测方法略优;确定性预测结果的 NMAE 指标方面,两算例中本文方法均优于分时刻预测方法。整体上看,两方法在概率预测精度方面差别不大,但表 2 中本文方法预测程序耗费时间仅为分时刻预测方法的 1/18~1/16。这是因为本文方法只需要对因子分析化简后的 3 个风电公共因子建立 QRNN 模型,而分时刻预测方法则需要分别建立 24 个对应时刻的风电功率 QRNN 模型,建模和模型参数估计时间都大大增加。

5 结论

本文引入因子分析技术对复杂的多维风电功率变量进行降维化简,提取出相互独立的日风电公共因子作为预测变量,建立神经网络分位数回归概率模型,并基于中期天气预测信息,对未来月风电曲线进行概率预测和随机场景模拟分析。以某两个



装机容量不同的实际风电场为算例,验证了所提概率预测方法的有效性和适应性,得出结论如下:

1) 本文提出的月度风电曲线概率预测方法,不仅能获得较高精度的确定性预测结果,而且能够模拟未来月可能出现的风电曲线随机场景,准确反映实际风电功率的随机性、波动性和时序相关性,提供更准确、全面的月度风电预测信息。

2) 由于引入了中期天气预测信息,本文以概率中位数为确定性预测结果的预测误差,比仅利用历史风电信息的DFM动态因子确定性预测方法预测误差降低了约3%~8%。

3) 与不经过因子分析的分时刻风电功率概率预测方法相比,本文方法可以在不明显降低预测精度的条件下,将预测效率提高16~18倍。

本文利用中期天气预测信息,采用因子分析与QRNN概率预测相结合的方法,不仅能用于风电功率预测,亦能为其他受天气支配的新能源出力和负荷中长期时间尺度下的概率预测提供新的思路,其预测结果可用于含风电电力系统的概率优化调度、概率风险分析和决策等诸多方面。

## 参考文献

- [1] 张丽英,叶廷路,辛耀中,等.大规模风电接入电网的相关问题及措施[J].中国电机工程学报,2010,30(25): 1-9.  
Zhang Liying, Ye Tinglu, Xin Yaozhong, et al. Problems and measures of power grid accommodating large scale wind power[J]. Proceedings of the CSEE, 2010, 30(25): 1-9(in Chinese).
- [2] Foley A M, Leahy P G, Marvuglia A, et al. Current methods and advances in forecasting of wind power generation[J]. Renewable Energy, 2012, 37(1): 1-8.
- [3] García-Martos C, Sánchez M J. A dynamic factor model for mid-term forecasting of wind power generation[C]// Proceedings of the 2013 10th International Conference on the European Energy Market (EEM). Stockholm, Sweden: IEEE, 2013: 1-7.
- [4] Barbounis T G, Theocharis J B, Alexiadis M C, et al. Long-term wind speed and power forecasting using local recurrent neural network models[J]. IEEE Transactions on Energy Conversion, 2006, 21(1): 273-284.
- [5] 中国天气网:城市天气预报 30 天[DB/OL].海口:中国气象局,2009[2015-01-01].<http://www.tianqi.com/>.  
China Weather Network: 30-day forecast for city [DB/OL]. Haikou: China Meteorological Administration, 2009[2015-01-01]. [http://www.tianqi.com/\(in Chinese\)](http://www.tianqi.com/(in Chinese)).
- [6] Pinson P, Nielsen H A, Møller J K, et al. Non-parametric probabilistic forecasts of wind power: required properties and evaluation[J]. Wind Energy, 2007, 10(6): 497-516.
- [7] Sideratos G, Hatzigiargyriou N D. Probabilistic wind power forecasting using radial basis function neural networks[J]. IEEE Transactions on Power Systems, 2012, 27(4): 1788-1796.
- [8] 雷宇,杨明,韩学山.基于场景分析的含风电系统机组组合的两阶段随机优化[J].电力系统保护与控制,2012, 40(23): 58-67.  
Lei Yu, Yang Ming, Han Xueshan. A two-stage stochastic optimization of unit commitment considering wind power based on scenario analysis[J]. Power System Protection and Control, 2012, 40(23): 58-67(in Chinese).
- [9] 杨明,范澍,韩学山,等.基于分量稀疏贝叶斯学习的风电场输出功率概率预测方法[J].电力系统自动化, 2012, 36(14): 125-130, 142.  
Yang Ming, Fan Shu, Han Xueshan, et al. Wind farm generation forecast based on componential sparse Bayesian learning[J]. Automation of Electric Power Systems, 2012, 36(14): 125-130, 142(in Chinese).
- [10] Khosravi A, Nahavandi S. Combined nonparametric prediction intervals for wind power generation[J]. IEEE Transactions on Sustainable Energy, 2013, 4(4): 849-856.
- [11] 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(in Chinese).
- [12] Juban J, Fugon L, Kariniotakis G. Probabilistic short-term wind power forecasting based on kernel density estimators[C]//European Wind Energy Conference. Milan, Italy: European Wind Energy Association, 2007: 7-10.
- [13] 李智,韩学山,杨明,等.基于分位数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(in Chinese).
- [14] 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(in Chinese).
- [15] 何耀耀,闻才喜,许启发,等.考虑温度因素的中期电力负荷概率密度预测方法[J].电网技术,2015, 39(1):

- 176-181.
- He Yaoyao, Wen Caixi, Xu Qifa, et al. A method to predict probability density of medium-term power load considering temperature factor[J]. Power System Technology, 2015, 39(1): 176-181(in Chinese).
- [16] Basilevsky A. Statistical factor analysis and related methods: theory and applications[M]. New York: Wiley, 1994: 351-352.
- [17] García-Martos C, Rodríguez J, Sánchez M J. Forecasting electricity prices and their volatilities using Unobserved Components[J]. Energy Economics, 2011, 33(6): 1227-1239.
- [18] García-Martos C, Rodríguez J, Sánchez M J. Forecasting electricity prices by extracting dynamic common factors: application to the Iberian Market[J]. IET Generation, Transmission & Distribution, 2012, 6(1): 11-20.
- [19] 许启发, 蒋翠侠. 分位数局部调整模型及应用[J]. 数量经济技术经济研究, 2011(8): 115-133.
- Xu Qifa, Jiang Cuixia. Quantile partial adjustment model and its application[J]. The Journal of Quantitative & Technical Economics, 2011(8): 115-133(in Chinese).
- [20] Koenker R, Bassett Jr G. Regression quantiles[J]. Econometrica, 1978, 46(1): 33-50.
- [21] Taylor J W. A quantile regression neural network approach to estimating the conditional density of multi-period returns[J]. Journal of Forecasting, 2000, 19(4): 299-311.
- [22] Li Dan, Yan Wei, Li Wenyuan, et al. A two-tier wind power time series model considering day-to-day weather transition and intraday wind power fluctuations[J]. IEEE Transactions on Power Systems, 2016, 31(6): 4330-4339.
- [23] Cannon A J. Quantile regression neural networks: implementation in R and application to precipitation downscaling[J]. Computers & Geosciences, 2010, 37(9): 1277-1284.
- [24] 叶阿忠. 非参数计量经济学[M]. 天津: 南开大学出版社, 2003: 14-37.
- Ye Azhong. Nonparametric econometrics[M]. Tianjin: Nankai University Press, 2003: 14-37(in Chinese).
- [25] Pinson P, Madsen H, Nielsen H A, et al. From probabilistic forecasts to statistical scenarios of short-term wind power production[J]. Wind Energy, 2009, 12(1): 51-62.



李丹

收稿日期: 2016-06-09。

作者简介:

李丹(1980), 女, 博士研究生, 工程师, 主要从事电力系统概率分析、新能源发电和电力系统优化运行方面的研究, lucy2140@163.com;

任洲洋(1986), 男, 博士, 讲师, 主要从事电力系统规划及概率分析方面的研究, 本文通讯作者, rzhouyang1108@163.com;

颜伟(1968), 男, 博士, 教授, 博士生导师, 主要从事电力系统优化运行与控制方面的研究, cquyanwei@cqu.edu.cn;

朱继忠(1966), 男, 博士, 国家“千人计划”特聘专家, 教授, 博士生导师, 主要从事电力系统分析运行与控制、电力市场、智能电网和可再生能源应用等方向的研究, zhujz@csg.cn。

(责任编辑 乔宝榆)

# Month-ahead Wind Power Curve Probabilistic Prediction Based on Factor Analysis and Quantile Regression Neural Network

LI Dan<sup>1</sup>, REN Zhouyang<sup>1</sup>, YAN Wei<sup>1</sup>, ZHU Jizhong<sup>2</sup>, ZHAO Xia<sup>1</sup>, YU Juan<sup>1</sup>

(1. Chongqing University; 2. China Southern Power Grid Company Limited)

**KEY WORDS:** month-ahead wind power prediction; factor analysis; quantile regression neural network; medium-range weather forecasting; probabilistic prediction

With the increase of wind power penetration in power systems, it is important to accurately forecast wind power production for the system operator. Concerning the existing problems such as the large number of highly correlated variables, few available weather information and high uncertainty in month-ahead wind power curve prediction, a new month-ahead wind power curve probabilistic prediction method based on factor analysis and quantile regression neural network (QRNN) is proposed.

First of all, a factor analysis method (FAM) is adopted to reduce the dimensionality of 24-dimensional vector of normalized wind power time series  $\mathbf{P}$ . Based on the following factor analysis model,  $r$  ( $r \leq 24$ ) unobserved independent common factors  $\mathbf{F} = [f_1, f_2, \dots, f_r]$  are extracted and used as the predictor variables of wind power time series, which contain the common features of the original hourly wind power series.

$$\mathbf{P} = \mathbf{A}\mathbf{F} + \boldsymbol{\varepsilon} \quad (1)$$

In order to obtain more information about the future wind power common factors, probability quantile forecasting models for  $r$  wind power common factors using QRNN are built with the daily weather features  $\mathbf{X}$  as inputs as follows:

$$Q_{y_m}(\tau | \mathbf{X}) = f\left(\sum_{j=1}^J \tanh\left(\sum_{k=1}^K \omega_{jk}^{(h)}(\tau)x_k + b_j^{(h)}(\tau)\right) \cdot \omega_j^{(o)}(\tau) + b_0^{(o)}(\tau)\right), \quad m = 1, 2, \dots, r \quad (2)$$

The choice of the output-layer transfer function  $f(\cdot)$  determines whether censored regression quantiles are to be estimated or not. With left censoring as the first common factor,  $f(\cdot)$  can be replaced by the ramp function as:

$$r(u) = \begin{cases} u, & u \geq l \\ l, & u < l \end{cases} \quad (3)$$

The identity function is used for estimating uncensored regression quantiles of other common factors.

The quantile regression parameters in models (2) are estimated by optimizing the following approximate error function as:

$$\hat{\theta}(\tau; \lambda, J) = \arg \min_{\theta} \{E_{y_m \tau} + \lambda \frac{1}{KJ} \|\omega^{(h)}(\tau)\|_2\} \quad (4)$$

By inputting the future daily weather features from available medium-range weather forecasts, a finite number of conditional quantiles of future wind power common factors can be predicted based on these QRNN models. Given predictions of a non-decreasing set of conditional quantiles, the probability density functions of future wind power common factors  $f(y_m | X_n)$  are estimated by nonparametric kernel density estimation.

By simulating common factors and specific factors following the predicted distributions, a variety of statistical scenarios of wind power production in the next month have been generated based on (1).

The effectiveness and adaptability of the proposed method are verified by the prediction results of two actual cases with different installed capacity. Tab. 1 shows the error index of the point prediction results with the proposed method is reduced by 3~8% compared to that with the dynamic factor model (DFM) method using historical wind power data simply. Tab. 2 shows the proposed method can greatly improve efficiency without sacrificing accuracy compared with the each-hour probabilistic prediction (EHPP) method of wind power. The month wind power curves simulated by the proposed method can accurately reflect the randomness, fluctuation and autocorrelation of actual wind power.

**Tab. 1 NMAE of different prediction methods** %

Case	The proposed method	DFM
1	13.08	20.80
2	12.91	15.98

**Tab. 2 Results of two probabilistic prediction methods**

Case	Method	Time consumption/s	Skill score	NMAE/%
1	The proposed method	464	-1.191	13.08
	EHPP	8132	-1.038	15.78
2	The proposed method	318	-1.016	12.91
	EHPP	5044	-1.285	15.72

The proposed method can provide a new feasible solution for the probabilistic medium- and long-term wind power forecasting.