

# 基于预测误差分布优化模型的风电功率 超短期概率区间预测研究

杨 茂, 杨春霖, 董骏城

(现代电力系统仿真控制与绿色电能新技术吉林省重点实验室(东北电力大学), 吉林 132012)

**摘 要:** 提出一种基于预测误差分布优化模型的风电功率概率区间预测方法。由于风功率数据存在显著的时间相依结构,该方法首先对预测功率按出力不同进行划分,以划分区段内的预测误差为统计样本,分别采用多种分布模型拟合误差概率密度,通过拟合指标选择优化模型,进而求解该分布模型的累积概率,并通过计算指定置信水平下的置信区间进行概率区间预测。利用性能指标比较典型单一分布模型和优化模型的预测结果,表明基于优化模型的概率性预测区间覆盖率更高、平均带宽更窄、精度更好、效果更优。

**关键词:** 风电功率; 误差分布; 区间预测; 优化模型

**中图分类号:** TM614

**文献标识码:** A

## 0 引 言

风力发电技术在电力能源领域所扮演的角色日益加重。但风电长期以来因其间歇性、随机性和波动性对电网稳定运行造成不利影响而饱受诟病,因此风电功率的预测环节显得尤为重要<sup>[1-3]</sup>,风电功率的不确定性研究对于电力系统规划和运行决策至关重要。目前风功率短期的确定性的点预测主要包括物理法和统计法<sup>[4-11]</sup>。由于各种不确定因素的影响,实际工程中的确定性点预测通常存在一定误差,且无法给出预测数值出现的概率及波动范围。为满足决策者在电网规划、可靠性评估等方面的要求,需采用风电功率的概率性区间预测法量化不确定性,该量化区间的重要意义主要体现在:1)可在制定电力系统鲁棒调度计划时提供准确的风电波动范围,避免由于优化区间过大、调度计划过于保守造成的经济损失;2)可为概率潮流或最优潮流计算提供更准确的风电区间信息,从而得到更加准确的节点电压、支路功率越限等信息,为调度提供参考<sup>[12]</sup>。

现阶段的风电功率概率区间预测方法主要集

中于单一分布模型。其预测原理通常为假定预测误差服从某种形式的分布模型,如正态分布、贝塔分布、含有尺度参数和位置参数的  $t$  分布( $t$  location-scale)等<sup>[13-16]</sup>,由历史误差数据进行统计拟合确定假设分布模型的参数,然后进行预测。单一分布对统计样本分布特性依赖性强、分布形态单一,且一旦样本本身分布特性超出所采用分布模型的描述范围会对其预测结果的精度产生较大影响。为克服单一分布拟合误差的局限性提高模型的适用性和准确性,将多个单一分布进行组合优化,即成为风电功率概率区间预测的新方法。文献[17]采用分段指数分布描述短期风电功率预测误差分布,较单一分布具有明显效果;文献[18]采用混合偏态分布模型拟合短期风电功率预测误差分布,拟合精度高,但并未将该方法应用于区间预测。

基于上述研究,本文提出一种基于预测误差分布优化模型的风电功率概率区间预测方法,分别对比多种分布模型拟合误差概率密度,通过拟合指标选择优化模型,求解该分布模型的累积概率,通过计算指定置信水平下的置信区间进行概率区间预测。针对模型适用性和有效性进行详细探究,利用

性能指标比较典型单一分布模型和优化模型的预测结果,表明基于优化模型的概率性区间预测具有明显优势。

## 1 风电功率误差分析

风电功率预测误差是通过历史风电功率预测值与实际值之差统计得到,其计算公式可表示为:

$$P_w = \frac{P_R(t) - P(t)}{P_{cap}} \quad (1)$$

式中,  $P_R(t)$  ——风电功率预测值;  $P(t)$  ——风电功率实际值;  $P_{cap}$  ——风电场总装机容量。

预测得到的大规模风电场风电功率预测误差总体分布直方图如图 1 所示。由图 1 可看出,风电场全场预测误差的总体分布呈对称分布特性,然而实际预测中不同预测功率区段内的预测误差概率分布存在较大差异,其误差分布存在多种特性,不同预测功率区段内的预测误差具有有偏性、多峰性、厚尾性等特征。从理论上讲,预测功率区段划分越细,越能清晰反映预测误差的局部分布形态,但预测功率区段划分过细会导致区段内的预测误差样本容量过少,无法充分体现预测误差的分布特性,因此需权衡预测功率区段划分数量和各区段内样本容量二者之间的关系。

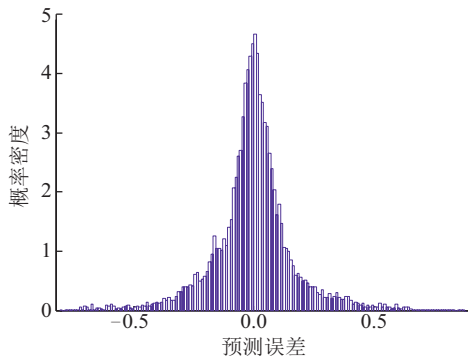


图1 风电功率预测误差总体分布直方图

Fig. 1 Overall distribution histogram of wind power forecasting error

图 2 为以 50 MW 为划分间隔将某风电场全场总出力按预测功率水平不同划分成 5 个区段后,各区段内的预测误差分布情况。由图 2 可看出,不同预测功率区段的预测误差分布情况各不相同,高预测功率区段的预测误差呈厚尾特性,且随预测功率

水平的增加,误差分布的厚尾性越显著,其物理原因在于高预测功率水平本身出力值越高,误差值越高。不同预测功率区段进行划分时,需同时满足预测区段多且各区段内样本容量大,这 2 个条件,预测区段的划分也是影响预测效果的因素之一。又因不同预测功率段内误差分布各不相同,其分布特性很难用总体分布进行统一描述,所以需针对不同预测功率段内的误差分布局部特性进行分析。

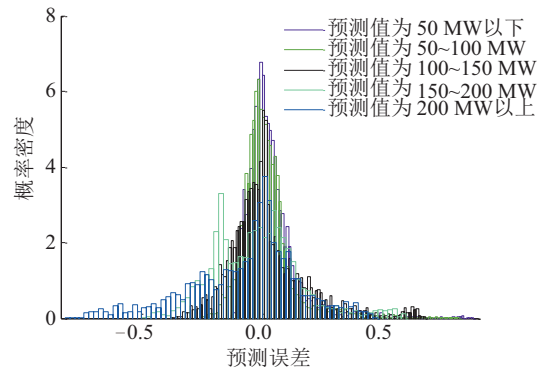


图2 不同预测功率区段预测误差分布图

Fig. 2 Forecasting error distributions for different forecasting value segment

## 2 分布模型参数估计与评价

### 2.1 单一分布模型和优化模型

描述风电功率预测误差分布情况的模型主要集中于单一分布模型,常用单一分布模型有正态分布、逻辑斯特分布、极值分布和含有尺度参数和位置参数的  $t$  分布(简称  $t$  分布)。

正态分布的概率密度函数表达式为:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2\sigma_i^2}(x-\mu_i)^2} \quad (2)$$

式中,  $x$  ——样本点;  $\mu_i$  ——样本均值;  $\sigma_i^2$  ——样本方差。

逻辑斯特分布的概率密度函数表达式为:

$$f(x) = \frac{e^{\left(\frac{x-\mu}{\sigma}\right)}}{\sigma \left[1 + e^{\left(\frac{x-\mu}{\sigma}\right)}\right]^2} \quad (3)$$

式中,  $\sigma$  ——样本标准差。

极值分布的概率密度表达式为:

$$f(x) = \frac{e^{\left(\frac{x-\mu}{\sigma} - e^{\left(\frac{x-\mu}{\sigma}\right)}\right)}}{\sigma} \quad (4)$$

$t$  分布概率密度函数表达式为:

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left[1 + \frac{\left(\frac{x-\mu}{\sigma}\right)^2}{v}\right]^{-\frac{v+1}{2}} \quad (5)$$

式中,  $\sigma$  —— 样本标准差;  $v$  —— 自由度; 若记  $y = \frac{x-\mu}{\sigma}$ , 则  $y$  服从自由度为  $v$  的  $t$  分布。

单一分布对统计样本分布特性依赖性强、分布形态简单、拟合能力有限。优化模型是建立在单一分布模型基础上的线性组合分布模型, 它能克服单一分布拟合误差的局限性, 可根据研究对象分布形态灵活选取组合数, 提高模型的适用性和准确性。本文中优化模型采用混合正态分布和含有尺度参数和位置参数的混合  $t$  分布(简称混合  $t$  分布)。其概率密度函数即为上述单一分布概率密度函数的线性组合, 且线性组合系数之和为 1, 其概率密度函数可表示为:

$$\begin{cases} f(x) = \sum_{i=1}^n w_i f_i(x_i) \\ \sum_{i=1}^n w_i = 1 \end{cases} \quad (6)$$

式中,  $f(x)$  —— 优化模型的概率密度函数,  $x$  为样本点;  $f_i(x_i)$  —— 单一分布模型的概率密度函数;  $x_i$  —— 对应样本点;  $w_i$  —— 第  $i$  个单一分布的权重系数, 当  $n=1$  时, 优化模型的概率密度函数即为单一分布模型的概率密度函数。

## 2.2 分布模型参数估计

对于单一分布模型的概率密度函数参数估计可结合最小二乘法对误差概率密度函数进行拟合, 以概率密度函数与分布直方图的残差平方和最小为目标函数, 即可确定单一分布模型的各项参数。

对于优化模型的概率密度函数参数估计则需将原始风电功率预测误差分布进行聚类分析, 形成多类预测误差分布, 即将预测误差分布按类内相似度最高, 类间相似度最低原则分解为多个预测误差分布, 这样就将问题转化为单一分布的参数估计, 可针对每一类预测误差分布采用单一分布模型拟

合各参数值, 并计算各类预测误差样本占原始预测误差样本的比重即为权重, 由此可见, 各类预测误差样本的权重之和必为 1。

## 2.3 拟合优度评价指标

针对不同预测功率区段的预测误差直方图分别采用单一分布模型中的正态分布、逻辑斯特分布、极值分布、 $t$  分布及优化模型中的混合正态分布、混合  $t$  分布进行概率密度函数拟合, 为比较不同分布模型对风电功率预测误差的拟合优度, 分别采用均方根误差、绝对值平均误差、拟合相关系数来定量评价模型拟合优度。

拟合分布模型概率密度函数的离散化处理公式可表示为:

$$y_i = f(\bar{C}_i) \quad (7)$$

则模型误差为:

$$e_i = y_i - \bar{N}_i \quad (8)$$

式中,  $y_i = f(\bar{C}_i)$  —— 在给定横坐标数值  $\bar{C}_i$  上对应的纵坐标数值, 其中  $f$  为拟合分布的概率密度函数,  $i=1, 2, \dots, M$  ( $M$  为原始预测误差分布直方图的数组);  $\bar{N}_i$ 、 $\bar{C}_i$  —— 分布直方图中第  $i$  个直方柱的纵坐标数值和其中间位置对应的横坐标数值。

均方根误差值越小说明拟合优度越高, 可表示为:

$$e_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (9)$$

绝对值平均误差值越小说明拟合优度越高, 可表示为:

$$e_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (10)$$

拟合相关系数值越接近于 1 说明拟合优度越高, 可表示为:

$$R = \frac{cov(Y_1, Y_2)}{\sqrt{D(Y_1)} \cdot \sqrt{D(Y_2)}} \quad (11)$$

式中,  $n$  —— 数组长度;  $cov(\cdot)$  —— 协方差;  $D(\cdot)$  —— 方差;  $Y_1$ 、 $Y_2$  —— 不同的样本。

## 3 基于预测误差分布优化模型的风电功率区间预测

### 3.1 风功率预测区段的划分

实际预测中不同出力区段的误差分布相差较

大,为准确掌握误差分布规律需对功率进行区间划分。本文采取等间隔划分,功率取值范围是  $[P_1, P_k]$ ,假设功率段长为  $\Delta P$ ,则所分区间为:

$$Q_i = [P_1 + (i-1)\Delta P, P_1 + i\Delta P] \quad (12)$$

式中,  $i=1, 2, \dots, N$ ,  $N$  为区段数。

则:

$$N = \frac{(P_k - P_1)}{\Delta P} + 1 \quad (13)$$

对于预测功率区间的划分,通常划分的区间越小,对误差分布规律的描述越详细,但进行计算的数据总量是固定的,划分的区间越小,区段数量越多,每个区段内的样本数量越少,使得某些区段样本数量较少(高出力段尤其明显),无法由此分析预测误差统计规律。一般情况下,大样本时能得到较好效果,而小样本时可能因为某些点未抽到而效果较差,因此需对功率区间进行二次划分,将区间内点数量较少的相邻区段合并(一般为高出力段),合并后新区间将能得到更好预测效果。

### 3.2 风功率区间预测模型预测步骤

为保证预测数据和历史数据的完整,需先对数据进行预处理,包括补全缺失和处理异常数据(如功率为负值)等。若数据缺失严重,根据其缺失点前后记录值,用插值法补缺;如果某机组数据缺失严重,则需利用该机组与其周围机组的历史数据关系进行补缺,最终得到所需的风电场完整历史数据<sup>[19]</sup>。对处理后的数据,以历史预测误差序列为样本,本文提出采用预测误差分布优化模型进行概率区间预测,步骤为:

1)先将历史数据按照上述要求进行处理,计算功率预测的误差序列,并按照预测功率等间隔划分成不同区段,得到不同预测功率区段对应的预测误差序列。

2)根据各个区间内的样本点数量进行二次划分,将样本点数量较少的相邻区间进行合并,得到新的不同预测功率区段的预测误差序列。

3)分别采用单一分布中的正态分布、逻辑斯特分布、极值分布、 $t$  分布及优化模型中的混合正态分布、混合  $t$  分布对不同预测功率区段的预测误差序列进行拟合,根据拟合优度评价指标选择模型,确定模型参数。

4)根据 3)选出的 3 种概率密度函数的相关参数求取相应的累积概率分布函数:

$$F(x) = \int_{-\infty}^{+\infty} f(x) dx \quad (14)$$

5)采用估计区间最窄原则,使估计区间  $[\theta'_i, \theta''_i]$  满足:

$$\begin{cases} F(\theta''_i) - F(\theta'_i) = 1 - \alpha \\ \min(\Delta\theta = \theta''_i - \theta'_i) \end{cases} \quad (15)$$

式中,  $\theta'_i$ 、 $\theta''_i$  ——估计区间下限、估计区间上限;  
 $1 - \alpha$  ——置信度;  $i$  ——第  $i$  个样本。

6)将每个时刻的预测值的置信区间上限和下限分别联结即可形成上下 2 条包络线,得到概率性区间预测。

## 4 算例分析

在对已有功率预测模型误差分析的基础上,以吉林省西部某风场为研究对象,验证优化分布模型用于描述不同预测出力区段时预测误差分布的有效性。该风场共有 267 台风电机组,单机容量为 1.5 MW,总装机容量为 400.5 MW,时间分辨率为 15 min,取风场 2014 年 4~10 月份的实际值和预测值为样本,对 4~9 月份的数据样本统计其预测误差相应的概率分布特性,仿真比较基于不同分布模型拟合预测误差分布的适用性和准确性。并采用优化模型对 10 月份的风电功率概率区间进行预测,与单一分布进行对比,验证模型的有效性。

### 4.1 预测误差的概率密度函数和累积概率函数

由于在不同预测功率区段内预测误差分布差异较大,因此将依据不同功率区段对误差进行区间划分。将总体预测误差样本根据预测功率值分段,区间跨度为 50 MW,为避免出现某些区间样本点过少而不能较好反映预测误差统计规律的问题,需将样本容量较少的区间与相邻区间进行合并,故将预测误差按预测出力水平分为 5 个区段,分别为预测介于 0~50 MW、50~100 MW、100~150 MW、150~200 MW、200 MW 以上。

分别采用单一分布模型中正态分布、逻辑斯特分布、极值分布、 $t$  分布及优化模型中混合正态分布、混合  $t$  分布对不同区段下的预测误差进行概率



密度函数拟合,并统计各项拟合优度评价指标,比较上述各分布模型的优劣。

图 3a 为预测值介于 0~50 MW 之间时,分别采用正态分布、逻辑斯特分布、极值分布、 $t$  分布、混合正态分布、混合  $t$  分布 6 种模型对预测误差进行概率密度拟合图,图 3b 为相应的累积概率分布图。由图 3a 可看出,预测值介于 0~50 MW 之间时,相应的预测误差呈对称分布特性,且厚尾性不明显。 $t$  分布、混合正态分布、混合  $t$  分布 3 种模型的概率密度拟合明显优于其他模型,结合表 1 中 3 种拟合优度评价指标可得出优化模型中的混合正态分布、混合  $t$  分布在 3 种评价指标下均优于单一  $t$  分布模型,而优化模型中混合  $t$  分布具有最好的拟合效果,其均方根误差较单一  $t$  分布下降 5.35%,绝对值平均误差下降 2.86%,拟合相关系数提高 0.65%。由图 3b 可看出混合  $t$  分布同样具有最好的效果。

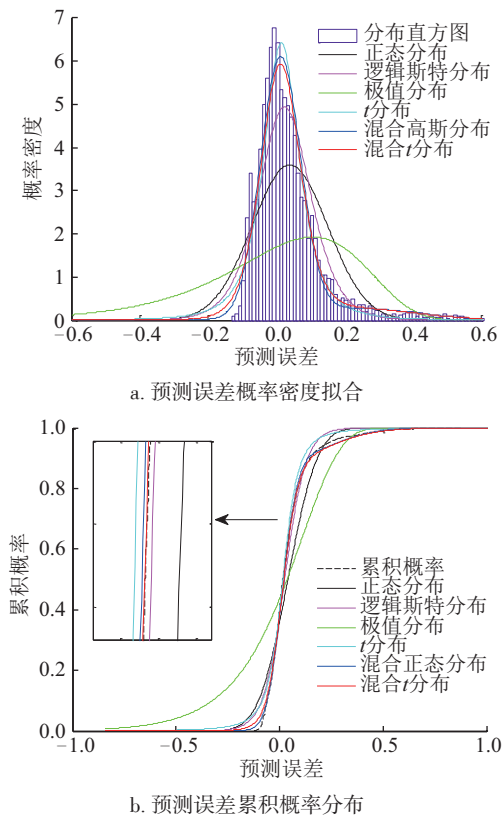


图 3 预测值为 0~50 MW 时预测误差分布图

Fig. 3 Forecasting error distribution of 0-50 MW forecasting value

图 4a 为预测值介于 50~100 MW 之间时,分别采用 6 种模型对预测误差进行概率密度拟合图,图

4b 为相应的累积概率分布图。由图 4a 可看出预测值介于 50~100 MW 之间时,相应预测误差仍具有对称性,但厚尾性较图 3a 明显。从图 4a 可知, $t$  分布、混合正态分布、混合  $t$  分布明显优于其他分布,且从表 1 中量化拟合优度评价指标可看出,混合  $t$  分布较  $t$  分布的均方根误差下降 4.46%,平均绝对值误差降低 3.94%,拟合相关系数提高 0.39%。由图 4b 可看出,混合  $t$  分布仍具有最好的效果。

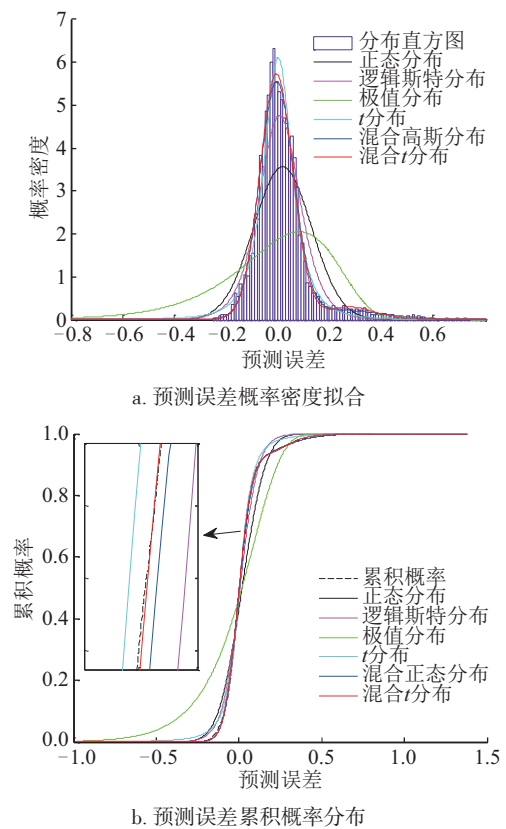


图 4 预测值为 50~100 MW 时预测误差分布图

Fig. 4 Forecasting error distribution of 50-100 MW forecasting value

图 5a 为预测值介于 100~150 MW 之间时,分别为 6 种模型下的预测误差概率密度拟合图,图 5b 为相应的累积概率分布图。由图 5a 可看出预测值介于 100~150 MW 之间时,相应预测误差则具有偏态性且厚尾性明显。图 5a 中  $t$  分布、混合正态分布、混合  $t$  分布优于其他 3 种分布,且优化模型优于单一分布模型,结合表 1 可看出混合  $t$  分布较  $t$  分布的均方根误差下降 1.26%,平均绝对值误差降低 2.37%,拟合相关系数提高 0.42%。由图 5b 可看出混合  $t$  分布较其他分布具有最好的效果。

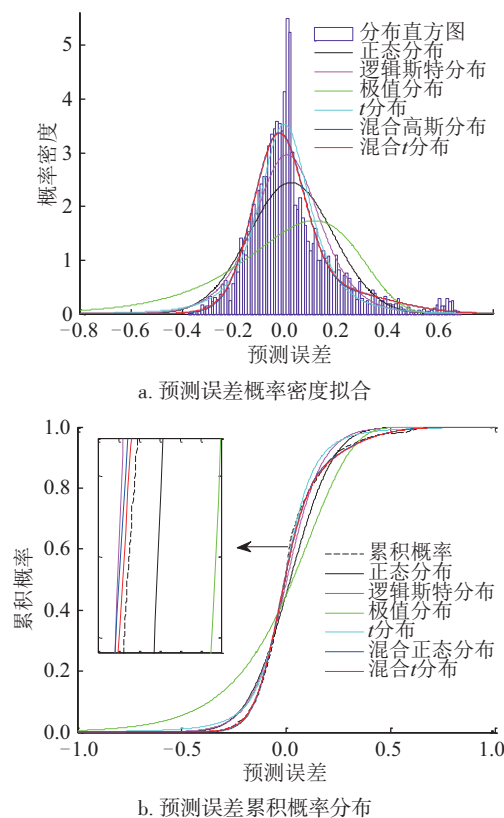


图5 预测值为100~150 MW时预测误差分布图  
Fig. 5 Forecasting error distribution of 100~150 MW forecasting value

图6a为预测值介于150~200 MW之间时,分别为6种模型下的预测误差概率密度拟合图,图6b为相应的累积概率分布图。由图6a可知预测值介于150~200 MW之间时,相应预测误差则具有偏态性,厚尾性和多峰性。图6a中优化模型中混合正态分布和混合 $t$ 分布优于其他分布,尤其在尾部拟合效果具有明显优势,结合表1可看出混合 $t$ 分布均方根误差为0.3464,较 $t$ 分布降低1.43%;平均绝对值误差为0.1966,下降2.59%;拟合相关系数为

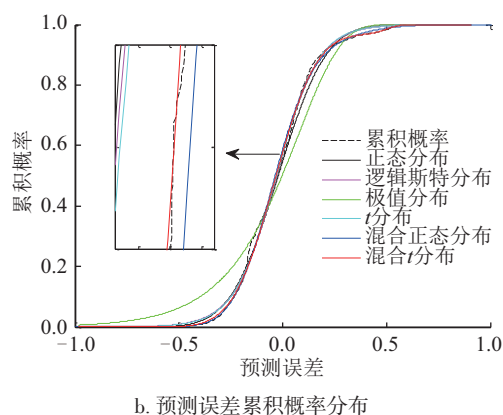
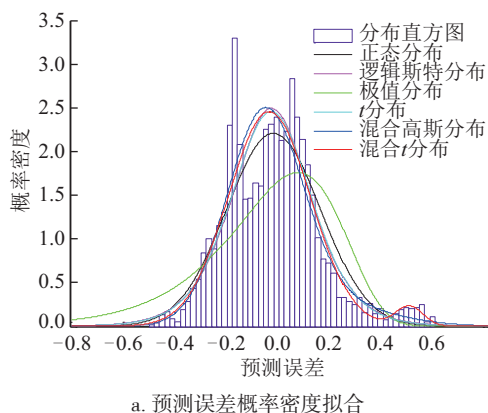


图6 预测值为150~200 MW时预测误差分布图  
Fig. 6 Forecasting error distribution of 150~200 MW forecasting value

0.9222, 提高0.69%。由图6b可看出,混合 $t$ 分布较其他分布具有最好的效果。

图7a为预测值介200 MW以上时,分别为6种模型下的预测误差概率密度拟合图,图7b为相应的累积概率分布图。由图7a可知预测值为200 MW

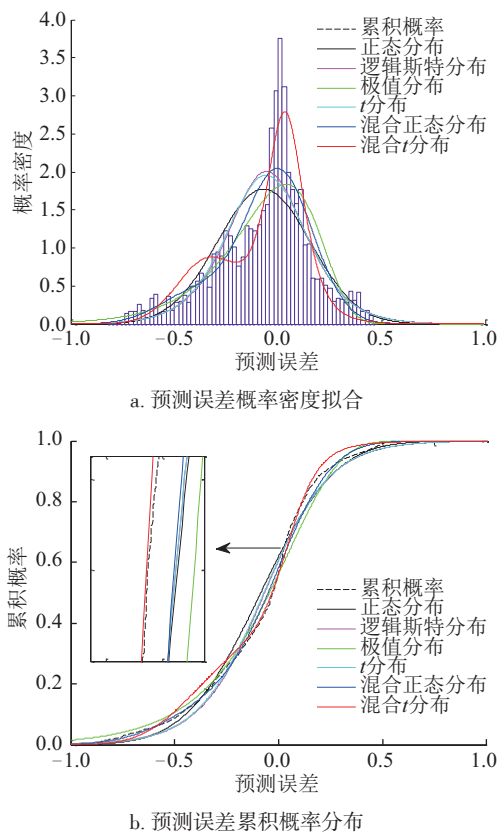


图7 预测值为200 MW以上时预测误差分布图  
Fig. 7 Forecasting error distribution of above 200 MW forecasting value

以上时,相应预测误差则具有偏态性、厚尾性、多峰性、不对称性。结合表 1 可看出混合  $t$  分布均方根误差为 0.2968,平均绝对值误差为 0.2185,拟合相关系数为 0.9276。均方根误差较  $t$  分布降低 12.54%,平均绝对值误差较  $t$  分布降低 3.24%,拟合相关系数提高 8.3%。且由图 7b 可看出混合  $t$  分布

具有最好的效果。  
综上所述,优化分布模型对于不同形态的预测误差均有很好的实用性,且误差分布形态越复杂,其描述效果越优于单一分布,在上述不同预测值区段下,混合  $t$  分布对预测误差分布情况的描述效果最佳。

表 1  拟合优度评价指标结果统计  
Table 1  Evaluation indicators of fitting goodness

预测值区段	评价指标	正态	逻辑斯特	极值	$t$ 分布	混合正态	混合 $t$ 分布
0~50 MW	均方根误差	0.8702	0.5411	1.3675	0.3889	0.3366	0.3354
	绝对值平均误差	0.4793	0.2811	0.7535	0.2131	0.1959	0.1845
	拟合相关系数	0.8764	0.9526	0.6662	0.9750	0.9810	0.9815
50~100 MW	均方根误差	0.7599	0.4069	1.2411	0.2452	0.2022	0.2006
	绝对值平均误差	0.4394	0.2348	0.7200	0.1460	0.1069	0.1066
	拟合相关系数	0.8981	0.9710	0.6769	0.9886	0.9922	0.9925
100~150 MW	均方根误差	0.6222	0.4912	0.8966	0.3905	0.3813	0.3779
	绝对值平均误差	0.3868	0.2887	0.5764	0.2210	0.2006	0.1937
	拟合相关系数	0.8514	0.9079	0.6560	0.9422	0.9450	0.9464
150~200 MW	均方根误差	0.3664	0.3667	0.5590	0.3607	0.3638	0.3464
	绝对值平均误差	0.2370	0.2275	0.4113	0.2225	0.2147	0.1966
	拟合相关系数	0.9134	0.9124	0.7937	0.9153	0.9143	0.9222
200 MW 以上	均方根误差	0.4520	0.4159	0.4151	0.4222	0.3645	0.2968
	绝对值平均误差	0.2699	0.2491	0.2444	0.2509	0.2312	0.2185
	拟合相关系数	0.8201	0.8496	0.8521	0.8446	0.8877	0.9276

## 4.2  风电功率概率区间预测

利用 4.1 节中的累积分布函数,分别采用单一分布模型中的  $t$  分布和优化模型中的混合  $t$  分布对 2014 年 10 月份数据进行概率区间预测,以估计区间最窄原则获取各功率区段下误差区间上下限,获取满足误差分布规律的置信区间,将区间上下边界点连接,得到的 2 条包络线之间的区间为某置信水平下的概率性区间预测结果。以区间覆盖率(coverage probability,  $CP$ )和区间平均带宽(normalized average width,  $NAW$ )为区间质量性能指标<sup>[20]</sup>。

区间覆盖率计算公式为:

$$CP=\frac{M-M_{out}}{M} \tag{16}$$

式中,  $M$  ——预测样本数量;  $M_{out}$  ——预测目标值

未落入预测区间内的样本数量。

区间平均带宽计算公式为:

$$NAW=\frac{\sum_{i=1}^M P_{up}^{(\alpha)}(i)-P_{down}^{(\alpha)}(i)}{M} \tag{17}$$

式中,  $P_{up}^{(\alpha)}(i)$ 、 $P_{down}^{(\alpha)}(i)$  ——设定置信度下对应预测区间的上下限。

当  $CP$  一定时,  $NAW$  值越小表明区间范围越窄,预测精度越高。图 8 和图 9 分别为 90%置信水平下,基于单一  $t$  分布模型和混合  $t$  分布模型的风功率概率预测区间。从图 8、图 9 和表 2 可看出,在相同置信水平下,2 种模型进行概率区间预测时覆盖率均为 100%,但平均带宽不同,混合  $t$  模型进行概率区间预测时平均带宽为 188.05 MW,较单一  $t$  分布模型减小 22.61 MW,可见其效果更好。

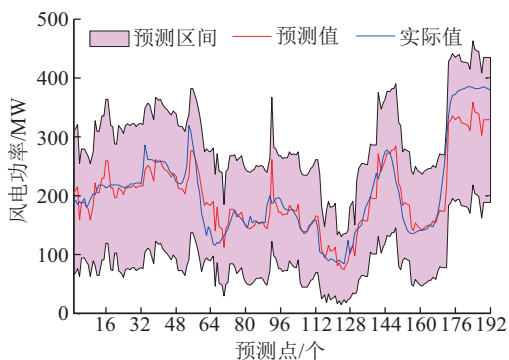
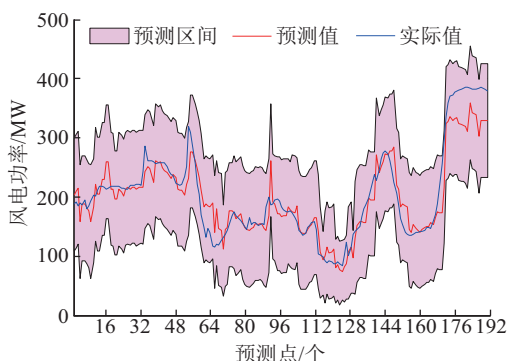
图8 基于单一 $t$ 分布模型的预测区间Fig. 8 Forecasting interval based on a single  $t$  distribution model图9 基于混合 $t$ 分布模型的预测区间Fig. 9 Forecasting interval based on the mixed  $t$  distribution model

表2 预测性能指标比较

Table 2 Comparison of the forecasting performance indicators

置信度/%	模型	CPI/%	NAW/MW
90	$t$ 分布	100	210.66
	混合 $t$ 分布	100	188.05

#### 4.3 区段划分数量对预测效果的影响分析

从理论上讲,划分间隔越细区段越多,越能反映预测误差的局部特性,预测效果越好,但划分间隔过细及区段过多会导致各区段内样本容量减少,直接影响拟合效果,因此需在划分区段数量和拟合优度之间寻求最优划分方法。为进一步探究区段划分数量对预测效果的影响,将总体预测误差样本按预测功率值进行更加细致的划分,以不同间隔将总体预测误差样本划分为不同区段,统计不同划分方法下平均拟合相关系数的变化情况,如图 10 所示。

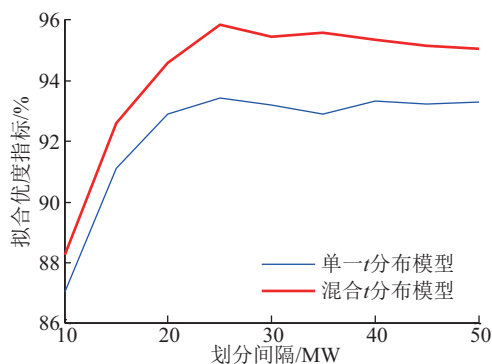


图10 划分间隔与拟合优度关系图

Fig. 10 Relationship between partition interval and fitting goodness

由图 10 可看出当划分间隔为 25 MW 时,拟合优度指标中的平均拟合相关系数最高,故以 25 MW 为间隔将总体预测误差样本进行划分,同时合并样本点较少的相邻区段,得到 0~25 MW、25~50 MW、50~75 MW、75~100 MW、100~125 MW、125~150 MW、150~175 MW、175~200 MW、200~225 MW、225 MW 以上 10 个区段的预测误差序列,对各区内预测误差段分别采用拟合效果较好的  $t$  分布和混合  $t$  分布进行概率密度拟合,相应拟合优度评价指标见表 3。从表 3 可看出,混合  $t$  分布模型各项拟合优度评价指标均优于单一  $t$  分布模型。通过不同区段的概率密度函数求取相应的累积概率分布函数,并基于估计区间最窄原则进行概率区间预测,在 90%置信水平下,进行风电功率概率区间预测,预测结果如图 11 和图 12 所示。由图 11 和图 12 可看出,基于混合  $t$  分布模型的预测效果优于单一  $t$  分布模型。

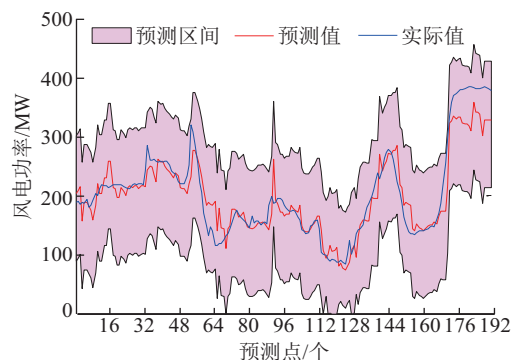
图11 基于单一 $t$ 分布模型和最优区段划分方式的预测区间Fig. 11 Forecasting interval based on a single  $t$  distribution model and optimal segment division



表3 最优区段划分方式下拟合优度评价指标结果统计

Table 3 Evaluation indicators of fitting goodness for optimal segment division

预测值区段	评价指标	$t$ 分布	混合 $t$ 分布	预测值区段	评价指标	$t$ 分布	混合 $t$ 分布
0~25 MW	均方根误差	0.5812	0.5604	125~150 MW	均方根误差	0.3497	0.2737
	绝对值平均误差	0.2430	0.2351		绝对值平均误差	0.2663	0.2155
	拟合相关系数	0.9512	0.9553		拟合相关系数	0.9207	0.9515
25~50 MW	均方根误差	0.3836	0.3075	150~175 MW	均方根误差	0.3163	0.3147
	绝对值平均误差	0.2097	0.1826		绝对值平均误差	0.2337	0.2173
	拟合相关系数	0.9762	0.9854		拟合相关系数	0.9395	0.9418
50~75 MW	均方根误差	0.4208	0.3436	175~200 MW	均方根误差	0.3549	0.3324
	绝对值平均误差	0.2189	0.1753		绝对值平均误差	0.1231	0.1046
	拟合相关系数	0.9703	0.9801		拟合相关系数	0.9565	0.9631
75~100 MW	均方根误差	0.2659	0.2559	200~225 MW	均方根误差	0.4764	0.3074
	绝对值平均误差	0.1612	0.1458		绝对值平均误差	0.3257	0.2376
	拟合相关系数	0.9884	0.9891		拟合相关系数	0.8433	0.9428
100~125 MW	均方根误差	0.4080	0.4053	225 MW 以上	均方根误差	0.4146	0.3105
	绝对值平均误差	0.2338	0.2135		绝对值平均误差	0.2454	0.2252
	拟合相关系数	0.9522	0.9592		拟合相关系数	0.8447	0.9175

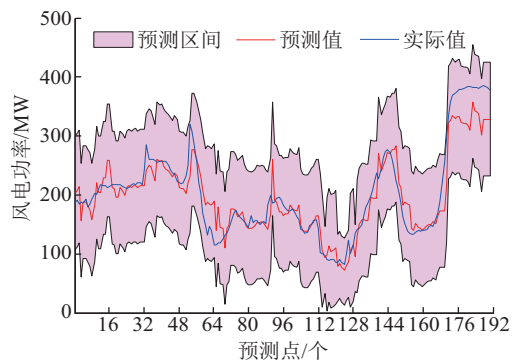


图12 基于混合 $t$ 分布模型和最优区段划分方式的预测区间

Fig. 12 Forecasting interval based on the mixed  $t$  distribution model and optimal segment division

将以上概率区间预测效果和基于总体预测误差分析下的预测效果以及以 50 MW 为间隔将总体预测误差分成 5 个区段分析下的预测性能进行对比,得到表 4 预测性能指标。由表 4 可看出,在同一置信度下,混合  $t$  分布模型在最优区段划分方式下的预测效果最好,当置信度为 90%时,相应区间覆盖率为 100%,区间平均带宽为 183.86 MW,区间平均带宽越小意味着预测精度越高,不确定性程度越小。在置信度和所用模型相同的情况下,不同区段划

表4 不同模型预测性能指标比较

Table 4 Comparison of performance indicators for different models

置信水平/%	划分间隔	分布模型	$CPI/\%$	$NAW/\text{MW}$
90	总体误差	$t$	100	215.51
		混合 $t$	100	190.86
	50 MW 间隔	$t$	100	210.66
	划分 5 区段	混合 $t$	100	188.05
	25 MW 间隔	$t$	100	209.27
	划分 10 区段	混合 $t$	100	183.86
80	总体误差	$t$	96.88	180.87
		混合 $t$	96.88	177.92
	50 MW 间隔	$t$	97.40	170.56
	划分 5 区段	混合 $t$	97.40	160.22
	25 MW 间隔	$t$	97.40	161.23
	划分 10 区段	混合 $t$	97.40	156.39

分方式得到的预测效果也不同,划分区段越细,其预测效果越好,但在细化区段的同时需要保证各区段内样本数量充足,才能保证模型的拟合优度,因此需在二者之间进行最优权衡。根据表 4 还可看

出,在置信度为 90%时,采用混合  $t$  分布模型所得到的最差预测效果亦优于  $t$  分布模型的所得到的最好预测效果,可见在所用模型和区段划分方式这二者对预测结果影响的重要程度来看,所用模型具有更加重要的地位。

## 5 结 论

本文提出一种基于预测误差分布优化模型的风电功率概率区间预测方法,具有以下特点:

1)随着预测出力水平的提高,风电功率预测误差分布厚尾性越显著,为此,有必要对不同预测区段的预测误差分布进行具体分析,以得到预测误差的局部分布特性;

2)在置信度和所用模型相同的情况下,不同区段划分方式得到的预测效果不同,预测出力区间划分越细,进行概率预测时效果越好,因此需在二者之间进行最优权衡,在保证个体样本容量充足的前提下,细化预测出力区间;

3)优化模型中的混合  $t$  分布对不同预测区段的预测误差分布的概率密度拟合精度均最高,且预测效果亦最优,明显优于单一分布模型,其具有很好的适用性和灵活性;

4)从所用模型和预测区段划分方式二者对预测结果影响的重要程度来看,所用模型具有更加重要的地位,但预测区段划分方式的不同,对进一步提高预测效果具有重要意义。

与单一的预测模型相比优化预测模型能提高预测精度,但同时在实际的操作过程中也会不可避免地增加操作的复杂度和计算的冗余度。如何在保证预测精度的同时兼顾运算的简易性将是后续研究的重点。

### [参考文献]

- [1] 杨 茂, 杨春霖. 考虑风向的风电功率预测不确定性研究[J]. 东北电力大学学报, 2018, 38(5): 9—15.
- [1] Yang Mao, Yang Chunlin. Uncertainty study on wind power forecasting considering wind direction[J]. Journal of Northeast Electric Power University, 2018, 38(5): 9—15.
- [2] 王丽婕, 冬 雷, 高 爽. 基于多位置 NWP 与主成分分析的风电功率短期预测[J]. 电工技术学报, 2015, 30(5): 79—84.
- [2] Wang Lijie, Lei Lei, Gao Shuang. Short term prediction of wind power based on multi position NWP and principal component analysis[J]. Transactions of China Electrotechnical Society, 2015, 30(5): 79—84.
- [3] 张颖超, 郭晓杰, 邓 华. 一种基于改进 GPR 和 Bagging 的短期风电功率组合预测方法[J]. 电力系统保护与控制, 2016, 44(23): 46—51.
- [3] Zhang Yingchao, Guo Xiaojie, Deng Hua. A short term wind power forecasting method based on improved GPR and Bagging[J]. Power System Protection and Control, 2016, 44(23): 46—51.
- [4] 陈 昊, 万秋兰, 王玉荣. 基于厚尾均值广义自回归条件异方差族模型的短期风电功率预测[J]. 电工技术学报, 2016, 31(5): 91—98.
- [4] Chen Hao, Wan Qiulan, Wang Yurong. Short term wind power forecasting based on the generalized autoregressive conditional heteroskedasticity model with thick tail[J]. Transactions of China Electrotechnical Society, 2016, 31(5): 91—98.
- [5] 马 斌, 张丽艳, 郭 成. 一种变权重风电功率最优组合预测模型[J]. 电力系统保护与控制, 2016, 44(5): 117—121.
- [5] Ma Bin, Zhang Liyan, Guo Cheng. An optimal combination forecasting model of wind power with variable weight[J]. Power System Protection and Control, 2016, 44(5): 117—121.
- [6] 杨 茂, 陈新鑫. 基于支持向量机的短期风速预测研究综述[J]. 东北电力大学学报, 2017, 37(4): 1—7.
- [6] Yang Mao, Chen Xinxin. Summary of short-term wind speed prediction research based on support vector machine[J]. Journal of Northeast Electric Power University, 2017, 37(4): 1—7.
- [7] 刘 波, 贺志佳, 金 昊. 风力发电现状与发展趋势[J]. 东北电力大学学报, 2016, 36(2): 7—13.
- [7] Liu Bo, He Zhijia, Jin Hao. Wind power status and development trends[J]. Journal of Northeast Electric Power University, 2016, 36(2): 7—13.
- [8] 杨 茂, 张 强. 风电功率超短期预测误差的非参数估计分布研究[J]. 东北电力大学学报, 2018, 38(1): 15—20.
- [8] Yang Mao, Zhang Qiang. Nonparametric estimation distribution of ultra-short-term prediction errors of wind power[J]. Journal of Northeast Electric Power University, 2018, 38(1): 15—20.
- [9] 刘 纯, 范高峰, 王伟胜, 等. 风电场输出功率的组合预测模型[J]. 电网技术, 2009, 33(13): 74—79.

- [9] Liu Chun, Fan Gaofeng, Wang Weisheng, et al. A combination forecasting models for wind farm output power[J]. Power System Technology, 2009, 33(13): 74—79.
- [10] 杨 茂, 杨春霖. 基于模糊粒计算的风电功率实时预测研究[J]. 东北电力大学学报, 2017, 37(5): 1—7.
- [10] Yang Mao, Yang Chunlin. Research on real-time prediction of wind power based on fuzzy granular computing[J]. Journal of Northeast Electric Power University, 2017, 37(5): 1—7.
- [11] 杨秀媛, 肖 洋, 陈树勇. 风电场风速和发电功率预测研究[J]. 中国电机工程学报, 2005, 25(11): 1—5.
- [11] Yang Xiuyuan, Xiao Yang, Chen Shuyong. Wind speed and generated power forecasting in wind farm [J]. Proceedings of the CSEE, 2005, 25(11): 1—5.
- [12] 兰 飞, 桑川川, 梁浚杰, 等. 基于条件Copula函数的风电功率区间预测[J]. 中国电机工程学报, 2016, 36(增刊1): 79—86.
- [12] Lan Fei, Sang Chuanchuan, Liang Junjie, et al. Interval prediction of wind power conditions based on Copula function [J]. Proceedings of the CSEE, 2016, 36(Sup 1): 79—86.
- [13] 刘兴杰, 谢春雨. 基于贝塔分布的风电功率波动区间估计[J]. 电力自动化设备, 2014, 29(12): 26—30.
- [13] Liu Xingjie, Xie Chunyu. Prediction of wind power fluctuation intervals based on Beta distribution[J]. Electric Power Automation Equipment, 2014, 29(12): 26—30.
- [14] 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.
- [14] 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.
- [15] 徐 曼, 乔 颖, 鲁宗相. 短期风电功率预测误差综合评价方法[J]. 电力系统自动化, 2011, 35(12): 20—26.
- [15] Xu Man, Qiao Ying, Lu Zongxiang. A comprehensive error evaluation method for short-term wind power prediction [J]. Automation of Electric Power Systems, 2011, 35(12): 20—26.
- [16] Bessa R J, Miranda V, Botterud A, et al. Time adaptive conditional kernel density estimation for wind power forecasting [J]. IEEE Transactions on Sustainable Energy, 2012, 3(4): 660—669.
- [17] 刘 芳, 潘 毅, 刘 辉, 等. 风电功率预测误差分段指数分布模型[J]. 电力系统自动化, 2013, 37(18): 14—19.
- [17] Liu Fang, Pan Yi, Liu Hui, et al. Piecewise exponential distribution model of wind power forecasting error[J]. Automation of Electric Power Systems, 2013, 37(18): 14—19.
- [18] 刘燕华, 李伟花, 刘 冲, 等. 短期风电功率预测误差的混合偏态分布模型[J]. 中国电机工程学报, 2015, 35(10): 2375—2382.
- [18] Liu Yanhua, Li Weihua, Liu Chong, et al. Mixed skew distribution model of short-term wind power prediction error [J]. Proceedings of the CSEE, 2015, 35(10): 2375—2382.
- [19] 赵永宁, 叶 林, 朱倩雯. 风电场弃风异常数据簇的特征及处理方法[J]. 电力系统自动化, 2014, 38(10): 39—46.
- [19] Zhao Yongning, Ye Lin, Zhu Qianwen. Characteristics and processing method of abnormal data clusters caused by wind curtailments in wind farms [J]. Automation of Electric Power Systems, 2014, 38(10): 39—46.
- [20] 王 铮, 王伟胜, 刘 纯, 等. 基于风过程方法的风电功率预测结果不确定性估计[J]. 电网技术, 2013, 37(1): 242—247.
- [20] Wang Zheng, Wang Weisheng, Liu Chun, et al. Based on the uncertainty of the results of wind power prediction process approach [J]. Power System Technology, 2013, 37(1): 242—247.

## ULTRA-SHORT TERM PROBABILISTIC INTERVALS FORECASTING OF WIND POWER BASED ON OPTIMIZATION MODEL OF FORECASTING ERROR DISTRIBUTION

Yang Mao, Yang Chunlin, Dong Juncheng

*(Modern Power System Simulation Control & Renewable Energy Tech. Jilin Province Key Laboratory,  
Northeast Electric Power University, Jilin 132012, China)*

**Abstract:** Study of uncertainties associated with wind power generation is essential for power system operation and decision-making. In this paper, a probabilistic wind generation forecasting intervals was established based on optimization model of forecasting error distribution. Due to temporal dependence characteristic inside power data. Firstly, forecasting power is divided equidistantly to compute the forecasting error within the segment as a statistical sample. Then using a variety of distribution models to fit the error probability density, and selecting the optimization model by evaluation indicators of fitting goodness. The cumulative probability function of the distribution model can be calculated from the probability density function. Finally, the probabilistic forecasting interval of corresponding power section can be achieved at the specified confidence level. By comparing the results of the typical single distribution model with the optimization model, it is shown that the probability coverage based on the optimization model indicates a more satisfactory performance.

**Keywords:** wind power; forecasting error; interval forecasting; optimization model