

基于混合自适应核密度估计的风电波动特性分析

李立新¹, 田旭², 刘飞², 张君², 张祥成², 张鑫¹, 李媛媛¹

(1. 电网安全与节能国家重点实验室(中国电力科学研究院有限公司), 北京 100192; 2. 国网青海省电力公司经济技术研究院, 青海 西宁 810000)

摘要: 分析掌握风电出力波动特性的内在规律有利于提高风电出力的预测精度, 进而指导电网调度部门合理安排发电计划, 提高系统运行的经济性。为描述风电出力波动的概率密度分布特性, 文章首先分别修正由经验法及无偏交叉验证法得到的固定带宽, 建立两种自适应带宽核密度估计模型; 然后, 将上述两种模型进行组合优化; 最后, 建立基于混合自适应核密度估计(HAKDE)的风电出力波动量概率密度分布模型。采用多种概率密度分布模型对华北某省风电场的不同时空尺度下的风电出力波动量进行拟合, 算例结果表明, HAKDE模型的拟合效果最优, 从而验证了HAKDE模型的有效性。

关键词: 核密度估计; 风电; 波动特性; 概率密度分布; 多时空尺度

中图分类号: TK81 **文献标志码:** A **文章编号:** 1671-5292(2024)07-0923-06

0 引言

随着全球化石能源的枯竭以及双碳战略目标的提出^[1,2], 可再生能源的利用受到了各界的关注。风能凭借其低成本、清洁等优点, 在世界各国得到了广泛的应用与发展。然而, 由于风电具有波动性和间歇性等特点, 随着风电接入容量的持续增长, 导致电网对风电的调度困难加剧, 因此, 对风电出力波动性的准确描述显得尤为重要。

基于统计学的风电出力波动量的概率密度分布研究得到了广泛应用。根据是否依赖于参数估计模型的选择, 拟合风电出力波动量概率分布的方法主要分为参数法和非参数法两种。

参数法具体做法: 假设风电出力波动量符合某种特定概率分布, 根据极大似然估计法、最小二乘法等估计方法确定分布参数, 得到最终风电出力波动量概率密度分布模型。文献[3]首先分析了风电出力骤降的分布规律, 然后采用广义 Pareto 分布模型对其概率分布特性进行拟合。文献[4,5]使用贝塔分布拟合风电功率预测误差。为解决单一分布拟合不够精确的问题, 文献[6]分别采用混合 t 分布和混合高斯分布描述了风电出力的波动特性。

参数法通常假定风电出力波动量服从某个已知概率分布模型, 并通过一定的算法得到模型参

数。但是当预先假定模型与实际分布匹配度不高时, 误差可能会很大, 基于数据驱动的非参数法为这个问题提供了新的思路和解决方案。在描述数据分布特性时, 非参数核密度估计(KDE)不依赖于参数估计模型的选择^[7], 无需对风电出力波动量分布的形式做任何假设, 并能够挖掘历史数据中的统计信息。文献[8]提出了利用 KDE 法和 GA-SVM 的风电出力长期波动特性的预测方法。文献[9]通过提取风电出力波动量, 结合 KDE 法提出了改进非参数核密度估计的风功率波动性概率建模方法。

基于上述对于风电出力波动量概率密度分布特性的相关研究, 本文采用经验法及无偏交叉验证法确定核密度估计的固定带宽, 为改善局部拟合优度较低的情况, 有针对性地修正带宽, 建立两种自适应带宽核密度估计模型, 然后将上述两种模型进行组合, 最终提出混合自适应核密度估计(HAKDE)模型。分别对不同场群规模、不同采样时间尺度下的风电出力波动量概率密度分布进行拟合, 并通过拟合指标的比较, 验证了 HAKDE 模型的精确性。

1 风电出力波动量的概率分布

风电场分布范围广, 风电出力具有不可预期的波动性, 其波动量可较直观地表征风电场发电

收稿日期: 2022-10-31。

基金项目: 国网青海省电力公司经济技术研究院科技项目(SGQHJY00GHJS220052)。

作者简介: 李立新(1989-), 男, 工程师, 主要从事新能源并网及大电网安全稳定分析工作。E-mail: lilixin@epri.sgcc.com.cn

功率的平稳情况,该值越小表征风电出力越稳定,并网运行对系统的影响越小,反之影响越大。

图 1 为华北某省风电出力时间序列图。

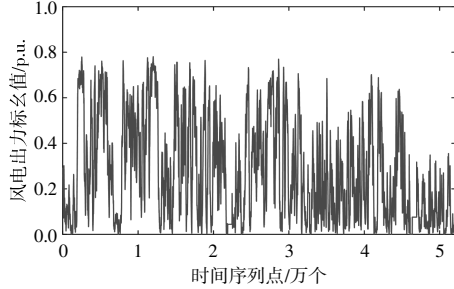


图 1 华北某省风电出力时间序列图

Fig.1 Time series of wind power output of a province in North China

对图 1 进行一阶差分变换,得到风电出力一阶差分波动量时间序列(图 2)。

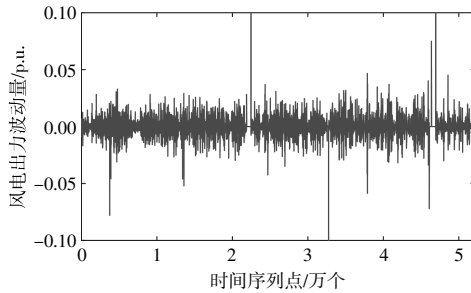


图 2 风电出力一阶差分波动量时间序列图

Fig.2 Time series of first-order differential fluctuations of wind power output

一阶差分波动量表达式为

$$\Delta P_t = P_t - P_{t-1}, t=2, 3, 4, \dots, n \quad (1)$$

式中: ΔP_t 为 t 时刻风电出力一阶差分波动量; P_t 为 t 时刻风电出力。

对风电出力波动量进行概率分布统计,绘制概率分布直方图,进而利用概率密度分布函数描述风电出力波动量变化规律,为最大限度利用风电提供理论依据。图 3 为风电出力一阶差分波动量概率分布直方图。

2 基于 HAKDE 的风电出力波动量概率密度分布

2.1 自适应带宽核密度估计模型

设 x_1, x_2, \dots, x_n 为风电出力波动量的 n 个样本,采用高斯函数作为核函数,得到风电出力波动量核密度估计函数表达式:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{m=1}^n K\left(\frac{x-x_m}{h}\right) =$$

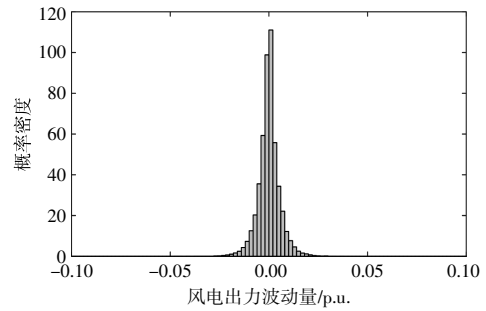


图 3 风电出力一阶差分波动量概率分布直方图

Fig.3 Histogram of the probability distribution of the first-order differential fluctuations of wind power output

$$\frac{1}{\sqrt{2\pi} nh} \sum_{m=1}^n \exp\left[-\frac{1}{2} \left(\frac{x-x_m}{h}\right)^2\right] \quad (2)$$

式中: $\hat{f}_h(x)$ 为风电出力波动量的核密度估计函数; $K(\cdot)$ 为核函数; h 为带宽; x_m 为风电出力波动量的第 m 个样本值; n 为样本总数。

在 KDE 中,带宽的选取决定了拟合曲线的平滑程度,带宽越大越平滑,其拟合效果则越差。目前,确定固定带宽的方法主要有经验法则和无偏交叉验证法^[10]。

经验法则的表达式为

$$h_1 = 1.06\sigma n^{-1/5} \quad (3)$$

式中: h_1 为采用经验法则求得的固定带宽; σ 为风电出力样本正态分布标准差。

无偏交叉验证法的表达式为

$$h_2 = \operatorname{argmin} \left[\frac{1}{nh^2 \sum_{i=1}^n \sum_{j=1}^n \bar{K}(X_i - X_j)} - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K(X_i - X_j) \right] \quad (4)$$

式中: h_2 为采用无偏交叉验证法求得的固定带宽;

argmin 为求取函数值最小的自变量; $K(v) = \int K(u) \cdot K(v-u) du$,为由 $K(\cdot)$ 导出的卷积核函数。

通过两个固定带宽 h_1, h_2 得到的 KDE 模型均存在局部拟合优度较低的情况,故求出两种固定带宽后,再有针对性地修正带宽 h_1, h_2 ,分别求解出两种方法的自适应带宽,改进策略如下。

①在得到固定带宽 h_0 后,将整个样本区间分为互不重叠的 k 个子样本区间 l_1, l_2, \dots, l_k ,对于样本区间的模型误差进行判别,对于任意子样本区间如果不满足式(5),则称该子样本区间存在拟合优度较低的问题。

$$\chi_{l,i}^2 \leq \chi_{m-1}^2(\alpha) \quad (5)$$

式中: $\chi_{l,i}^2$ 为 l_i 子区间的 χ^2 检验统计数; χ_{m-1}^2 为显著水平下 α 下自由度为 $m-1$ 的 χ^2 分布, 本文中 α 取 0.05。

将 l_i 子区间分为 s 组, $\chi_{l,i}^2$ 的数学表达式为

$$\chi_{l,i}^2 = \sum_{j=1}^s \frac{(t_{ij} - d_{ij} F_{ij})^2}{d_{ij} F_{ij}} \quad (6)$$

式中: t_{ij} 为 l_i 子区间第 j 组的样本实际频数; d_{ij} 为该组的样本个数; F_{ij} 为该组的理论概率值。

②为提高拟合优度, 在拟合优度较低的区间内引入局部带宽因子 $\lambda_{l,i}$, 通过基于固定带宽得到的各个样本点处的概率密度可以得到 $\lambda_{l,i}$ [11]。

$$\lambda_{l,i} = \sum_{i=1}^{d_i} \hat{f}_h(x_i)^{-\frac{1}{2}} \quad (7)$$

式中: d_i 为 l_i 子区间的样本个数。

将 $\lambda_{l,i}$ 与 h_0 相乘可得自适应带宽 $h_{l,i}$ 。

$$h_{l,i} = h_0 \times \lambda_{l,i} = h_0 \sum_{i=1}^{d_i} \hat{f}_h(x_i)^{-\frac{1}{2}} \quad (8)$$

③将式(2)修改为式(9), 从而得到自适应核密度估计模型。

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi} d h_{l,i}} \sum_{x_m \in l_i} \exp \left[-\frac{1}{2} \left(\frac{x - x_m}{h_{l,i}} \right)^2 \right], \quad x_m \in l_i, i=1, 2, \dots, k \quad (9)$$

2.2 混合自适应核密度估计模型

综合考虑上述两种带宽求解方法的优点, 将上述得到的两个自适应核密度估计函数进行组合, 建立混合自适应核密度估计优化模型。因不同带宽估计出的概率密度函数具有不同的误差, 而这些误差可以通过加权组合相互抵消, 所以, 混合自适应核密度估计 $f_{\text{HAKDE}}(x)$ 可以更为准确地拟合风电出力波动量概率密度分布。

$$\hat{f}_{\text{HAKDE}}(x) = \beta_1 \hat{f}_1(x) + \beta_2 \hat{f}_2(x) \quad (10)$$

式中: $\hat{f}_1(x)$ 为修正经验法固定带宽 h_1 得到的自适应带宽核密度估计模型; $\hat{f}_2(x)$ 为修正无偏交叉验证法固定带宽 h_2 得到的自适应带宽核密度估计模型; β_1, β_2 为权重系数, 且满足 $\beta_1 + \beta_2 = 1$ 。

3 基于 HAKDE 的风电出力波动量概率密度分布建模流程

本文提出的风电出力波动量概率密度分布建

模方法流程如图 4 所示。

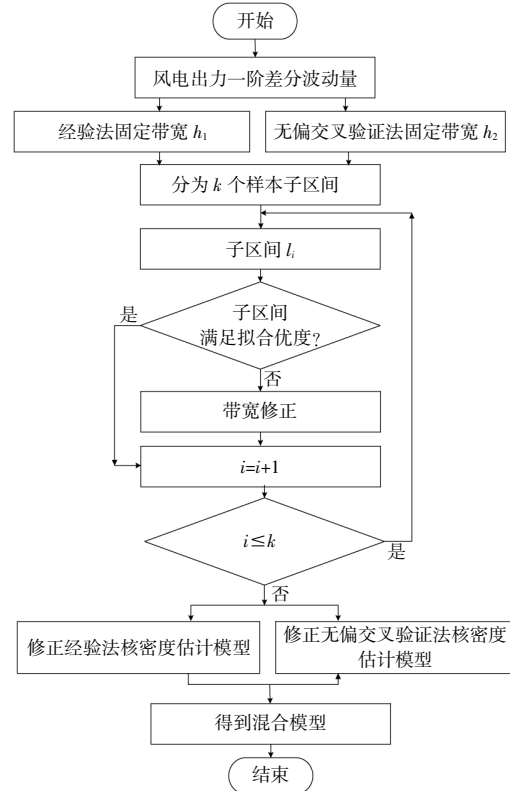


图 4 风电出力波动概率密度分布建模流程

Fig.4 Modeling process for probability density distribution of wind power output fluctuation

首先, 分别采用经验法和无偏交叉验证法确定核密度估计模型的固定带宽 h_1 和 h_2 , 再在子区间 l_i 内进行带宽的修正, 即求出两种方法的自适应带宽, 最后, 通过加权组合最终得到 HAKDE 模型。

4 算例分析

4.1 数据说明

为验证所提基于 HAKDE 的风电出力波动量概率密度分布模型的有效性和精确性, 以华北某省风力发电场的风电出力数据为样本进行分析, 并做归一化处理, 该风电场的总装机容量为 4 764 MW。

4.2 拟合评价指标

为了评价概率密度分布函数的拟合优度, 采用平均绝对误差(MAE)、均方根误差(RMSE)和确定系数 (R^2) 3 种评价指标验证模型的有效性。RMSE 和 MAE 越小说明模型越精确, 而 R^2 越接近于 1, 表明模型拟合程度越优。

3 种评价指标的计算式分别为

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{f}(x_i) - y(x_i)| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [\hat{f}(x_i) - y(x_i)]^2} \quad (12)$$

$$R^2 = \sqrt{\frac{\sum_{i=1}^n [\hat{f}(x_i) - \bar{y}(x_i)]^2}{\sum_{i=1}^n [y(x_i) - \bar{y}(x_i)]^2}} \quad (13)$$

式中: $i=1, 2, \dots, n$, n 为样本序列的个数; $y(x_i)$ 为第 i 个直方图的纵坐标; $\hat{f}(x_i)$ 为拟合的概率密度函数对应的函数值; $y(x)$ 为直方图纵坐标数值的平均值。

4.3 不同时间尺度下风电出力波动量概率密度分布特性

为分析不同时间尺度下风电出力波动量的概率密度分布特性, 以全省风电场的总出力为研究对象, 分别统计 5 min, 15 min 采样时间间隔的全省风电出力波动量的概率密度分布直方图。利用本文所提建模方法和正态分布、Logistic 分布、混合高斯分布、传统核密度估计分别对不同采样时间间隔下全省的风电出力波动量进行建模。

5 min 采样间隔下全省风电场的风电出力波动量概率密度曲线对比如图 5 所示。

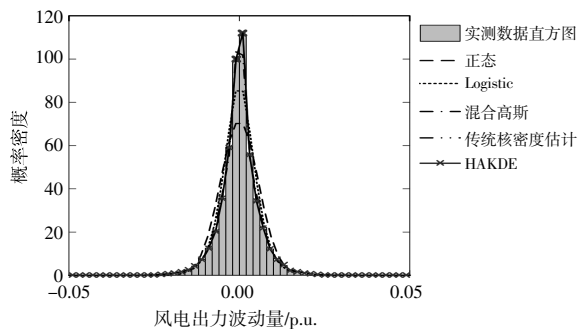


图 5 5 min 采样间隔下不同概率密度曲线拟合对比
Fig.5 Comparison of different probability density curve fits at 5 min sampling interval

由图 5 可知, 5 min 级风电出力波动量集中于数值较小的范围, 风电出力波动量大多分布在 $-0.25 \sim 0.25$ p.u., 其概率分布呈现“厚尾”特点, 整体具有对称性。

风电出力波动量概率密度误差运算结果如表 1 所示。

表 1 5 min 采样间隔下概率密度分布模型评价指标对比
Table 1 Comparison of evaluation indexes of probability density distribution models under 5 min sampling interval

模型	MAE	RMSE	R^2
正态分布	1.548 1	5.685 9	0.880 3
Logistic	0.977 1	3.540 3	0.949 2
混合高斯	0.309 7	1.201 9	0.978 9
传统核密度估计	0.131 8	0.852 5	0.981 9
HAKDE	0.036 5	0.152 8	0.999 4

由表 1 可知: HAKDE 模型的拟合效果最好, 其 3 项指标均为最优; 相比于其他两种分布模型, 正态分布、Logistic 分布以及混合高斯分布的 MAE 和 RMSE 较高, 而且 R^2 较低。由此可见, 如果先验分布选择错误, 参数估计法很难取得较好的建模精度。

15 min 采样间隔下全省风电场的风电出力波动量概率密度直方图如图 6 所示。

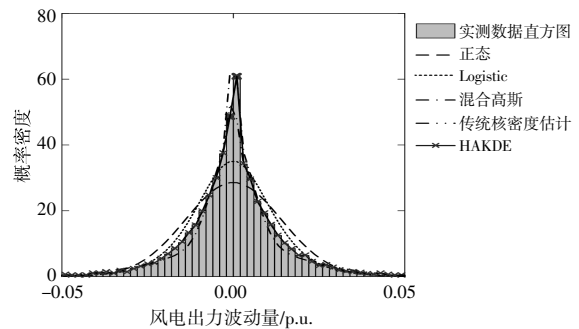


图 6 15 min 采样间隔下不同概率密度曲线拟合对比
Fig.6 Comparison of different probability density curve fits at 15 min sampling interval

由图 6 可知, 风电出力波动量以 0 为对称轴对称分布, 波动量标幺值最大约为 0.05。拟合精度评价结果如表 2 所示。

表 2 15 min 采样间隔下概率密度分布模型评价指标对比
Table 2 Comparison of evaluation indexes of probability density distribution model under 15 min sampling interval

模型	MAE	RMSE	R^2
正态分布	1.686 3	4.431 7	0.855 2
Logistic	1.103 6	3.268 9	0.920 5
混合高斯	0.640 9	1.638 2	0.947 5
传统核密度估计	0.240 8	1.348 4	0.961 3
HAKDE	0.055 2	0.086 8	0.999 8

由表 2 可知, 相较于其他 4 种分布模型, HAKDE 模型对风电出力波动量概率密度直方图的

拟合程度更高。

4.4 不同空间尺度下风电出力波动量概率密度分布特性

为分析不同空间尺度下风电出力波动量的概率密度分布特性,分别统计 15 min 采样时间间隔的单个风电场、风电场群的风电出力波动量,绘制其概率密度分布直方图,并分别用 HAKDE 模型、正态分布、Logistic 分布、混合高斯分布、传统核密度估计进行拟合。

15 min 采样间隔下单个风电场风电出力波动量概率密度曲线对比如图 7 所示。

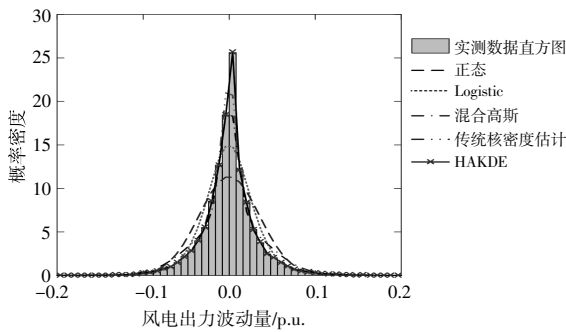


图 7 单个风电场不同概率密度曲线拟合对比

Fig.7 Comparison of different probability density curve fits for a single wind farm

15 min 采样间隔下单个风电场风电出力波动量概率密度误差运算结果如表 3 所示。

表 3 单个风电场概率密度分布模型评价指标对比
Table 3 Comparison of evaluation indexes of probability density distribution model for single wind farm

模型	MAE	RMSE	R^2
正态分布	0.535 8	1.789 1	0.799 3
Logistic	0.345 8	1.272 6	0.893 3
混合高斯	0.206 6	0.790 9	0.895 9
传统核密度估计	0.104 8	0.559 8	0.970 9
HAKDE	0.010 2	0.017 8	0.989 5

结合图 7 和表 3 可知,HAKDE 模型的 RMSE 和 MAE 指标数值均最小,几乎接近于 0,说明 HAKDE 模型拟合的单个风电场概率密度最为精确,而 HAKDE 模型的 R^2 指标较传统核密度估计提升了 1.9%,且最接近于 1,表明 HAKDE 模型拟合程度最优。

15 min 采样间隔下风电场群风电出力波动量概率密度曲线对比如图 8 所示。

15 min 采样间隔下风电场群风电出力波动量概率密度误差运算结果如表 4 所示。

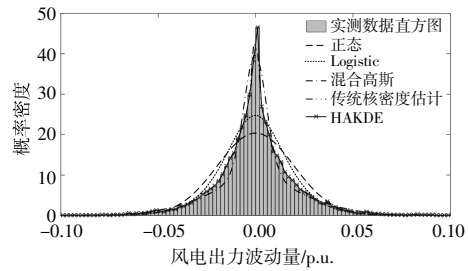


图 8 风电场群不同概率密度曲线拟合对比

Fig.8 Comparison of different probability density curve fits for wind farm clusters

表 4 风电场群概率密度分布模型评价指标对比
Table 4 Comparison of evaluation indexes of probability density distribution model of wind farm group

模型	MAE	RMSE	R^2
正态分布	1.657 8	3.792 8	0.857 5
Logistic	1.117 9	2.920 2	0.917 9
混合高斯	0.883 2	1.825 7	0.966 1
传统核密度估计	0.184 1	0.754 9	0.993 5
HAKDE	0.042 2	0.076 9	0.998 2

由图 8 和表 4 可知,HAKDE 模型的 RMSE 和 MAE 较传统核密度估计模型大幅度减小, R^2 较 1 仅相差 0.001 8,相差无几。综合上述分析可知,HAKDE 模型拟合风电场群概率密度的精确程度仍为最优。

随着场群规模逐渐扩大,风电出力波动量会随之减小。产生这种现象的主要原因是,当同一阵风经过多个风电场的时间有所不同,各个风电场的出力在时序上存在互补的现象。

5 结论

本文针对目前风电出力预测精度方面存在的问题,提出了一种基于混合自适应核密度估计的多时空尺度风电波动特性分析模型,得到如下结论。

①针对多时空尺度下风电出力波动量的分布特性,HAKDE 模型能够有效提高拟合优度,具有更高的适用性和精确性。

②随着采样时间间隔增长,风电出力波动量呈现增大的趋势,采用 HAKDE 模型拟合不同采样时间间隔下的风电出力波动量概率密度分布直方图效果更好。

③随着场群规模扩大,风电出力波动量呈现减小的趋势,HAKDE 模型对不同场群规模的风电场风电出力波动量的拟合精度更佳。

参考文献:

- [1] 陈国平,董昱,梁志峰.能源转型中的中国特色新能源高质量发展分析与思考[J].中国电机工程学报,2020,40(17):5493-5506.
- [2] 习近平.在第七十五届联合国大会一般性辩论上的讲话[N].人民日报海外版,2020-09-23.
- [3] 夏添,查晓明,秦亮,等.风电功率短时骤降的极值统计分析[J].电力系统保护与控制,2015,43(7):8-15.
- [4] Bludszuweit H, Domínguez -Navarro J, Llombart A. Statistical analysis of wind power forecast error[J]. IEEE Transactions on Power Systems, 2008, 23 (3): 983 - 991.
- [5] Bludszuweit H, Domínguez -Navarro J. A probabilistic method for energy storage sizing [J]. IEEE Transactions on Power Systems, 2011, 26(3): 1651-1658.
- [6] 杨茂,董骏城.基于混合分布模型的风电功率波动特性研究[J].中国电机工程学报,2016,36(S1):69-78.
- [7] 杨楠,崔家展,周崢,等.基于模糊序优化的风功率概率模型非参数核密度估计方法[J].电网技术,2016,40(2):335-340.
- [8] 肖白,邢世亨,王茂春,等.基于改进 KDE 法和 GA-SVM 的多风电场聚合后输出功率长期波动特性预测方法[J].电力自动化设备,2022,42(2):77-84.
- [9] 杨楠,黄禹,叶迪,等.基于 NACEMD 和改进非参数核密度估计的风功率波动性概率分布研究 [J]. 电网技术,2019,43(3):910-917.
- [10] 何耀耀,闻才喜,许启发.基于 Epanechnikov 核与最优窗宽组合的中期电力负荷概率密度预测方法[J].电力自动化设备,2016,36(11):120-126.
- [11] 张卫贞,石慧,曾建潮,等.基于核微分同胚变换的实时剩余寿命预测 [J]. 计算机集成制造系统,2020,26(10):2781-2791.

Analysis of wind power fluctuation characteristics based on hybrid adaptive kernel density estimation

Li Lixin¹, Tian Xu², Liu Fei², Zhang Jun², Zhang Xiangcheng², Zhang Xin¹, Li Yuanyuan¹
(1.State Key Laboratory of Grid Security and Energy Conservation (China Electric Power Research Institute), Beijing 100192, China; 2.State Grid Economic and Technological Research Institute of Qinghai Electric Power Company, Xining 810000, China)

Abstract: The analysis and mastery of the inner law of the fluctuation characteristics of wind power output is conducive to improving the prediction accuracy of wind power output, thus guiding the power grid scheduling department to reasonably arrange the power generation plan and improve the economy of system operation. To characterize the probability density distribution of wind power output fluctuations, two adaptive bandwidth kernel density estimation models are developed by modifying the fixed bandwidths obtained from the empirical method and the unbiased cross-validation method. Then, the above two models are combined and optimized, and finally the probability density distribution model of wind power output fluctuation based on hybrid adaptive kernel density estimation (HAKDE) is established. A variety of probability density distribution models were used to fit the fluctuations of wind power output at different spatial and temporal scales in a province in North China. The results show that the fitting effect of the HAKDE model is the best, which verifies the effectiveness of the HAKDE model.

Keywords: kernel density estimation; wind power; fluctuation characteristics; probability density distribution; multi-temporal scales