

考虑风电场储能容量配置的风电功率预测误差 估计算法对比研究

蒋艾町,李小雨,夏雪,李嘉逸

(中国电力工程顾问集团西南电力设计院有限公司,四川成都 610021)

摘要:针对风电功率预测误差估计方法中混合高斯分布拟合法和特征值提取估计法这两种适用范围较广的风电功率预测误差估计方法,详细介绍其原理和误差估计流程,利用实际风电场数据对两种方法进行算例验证,并根据计算结果,针对两种方法下的估计区间对储能容量配置的影响进行对比研究,为工程应用时的方法选取提供参考。同时,为了兼顾误差估计区间的有效性和经济性,有效指导风电场储能系统的容量配置,在高斯混合模型的基础上对风电功率预测误差进行状态划分,结合马尔可夫模型,提出一种MM-GMM优化预测误差区间估计算法并对其进行算例验证。

关键词:风电功率预测误差估计;储能容量配置;混合高斯分布;特征值提取;马尔可夫模型

中图分类号:TM614 **文献标志码:**A **文章编号:**1003-6954(2021)02-0043-05

DOI:10.16527/j.issn.1003-6954.20210209

Research and Comparative Study on Wind Power Forecast Error Estimation Algorithms Considering Wind Farm Energy Storage Capacity Configuration

Jiang Aiting, Li Xiaoyu, Xia Xue, Li Jiayi

(CPECC Southwest Electric Power Design Institute Co., Ltd., Chengdu 610021, Sichuan, China)

Abstract: Aiming at two common wind power forecast error estimation methods, that is, the Gaussian mixture distribution fitting method and the eigenvalue extraction method, the principle and process of these two methods are introduced in detail. Moreover, these two methods are verified by the actual wind farm data, and the estimation intervals of two methods are compared according to the impact on energy storage capacity configuration. In order to give consideration to the accuracy and validity of error estimation interval and guide the configuration of energy storage system effectively, based on Gaussian mixture model (GMM) and combined with Markov model (MM), an MM-GMM based optimized forecast error interval estimation algorithm is proposed and verified.

Key words: wind power forecast error estimation; energy storage capacity configuration; Gaussian mixture distribution; eigenvalue extraction; Markov model

0 引言

随着化石能源等不可再生能源的不断消耗,资源紧张、环境污染和气候变化等问题日益加剧。风力发电以其清洁安全、可持续、技术成熟、开发价值高等优势,迅速在各国的能源战略中占据重要位置^[1-3]。但由于风力资源固有的随机和波动特性,不论采用何种风电功率预测方法,风功率预测都具有

一定的不确定性,对风电系统的电网调度、弃风限电、备用容量配置以及安全运行都会产生影响^[4-5],因此,中国各省陆续推出风电功率预测误差考核政策。储能作为改善风电功率预测误差的一种有效手段在工程中被广泛应用,有效估计风电功率预测误差,不仅能为风电并网系统的调度和运行提供参考,还能对风电场储能系统容量的合理配置进行指导,因此有必要对风电功率预测误差估计方法进行研究。

目前国内研究最为广泛的是采用概率分布拟合的方法对风电功率预测误差进行估计计算。该方法通过统计风电功率预测误差历史数据,对统计数据进行拟合,建立其概率密度分布模型,进而求取预测误差的估计区间^[6-13]。文献[6]利用混合高斯分布对风电功率预测误差进行拟合研究分析,结果表明不同预测方法下产生的功率预测误差的分布特性不同,但混合高斯分布对于应用不同预测方法的风电功率预测误差均具有高精度的适应性。文献[7-8]基于贝塔分布对风电场实际风电功率进行分析,得到其贝塔分布的概率密度函数,然后建立相应的最优解模型,求取风功率预测误差的最小概率区间。文献[9-12]采用非参数概率区间估计方法,不对总体分布进行先验假设,通过滑动分块百分位数 Bootstrap 法、核密度估计等方法直接求取预测误差的置信区间。文献[13]首先采用聚类算法将预测误差采样数据进行分类,然后针对每类数据计算其概率密度函数,求取分布模型,进而得到风电预测误差估计区间。

此外,特征值提取法也经常用于风功率预测误差估计模型的建立^[14-18]。文献[14-16]根据风电场的历史运行数据和日前预测数据提取数据中的有效特征值,如当日功率波动、近期功率波动及预测方法精度等,通过多元线性回归法建立回归方程,进而求取预测误差的估计区间。文献[17-18]在直接提取数据特征值的基础上,加入小波分析、粗糙集理论及字典学习等算法,对风电功率预测数据进行处理后采用多元线性回归方程,建立风电功率预测误差的估计模型。然而,这些方法虽然能够改善预测误差估计效果,但增加了算法的复杂程度,工程实用性较低。

综上所述,下面选取适用范围较广的基于混合高斯分布拟合法和特征值提取法的风电功率预测误差估计算法,针对这两种方法进行了详细的对比研究。同时提出了一种结合高斯混合模型(Gaussian mixture model, GMM)和马尔可夫模型(Markov model, MM)的 MM-GMM 优化预测误差区间估计算法并验证其有效性。

1 混合高斯分布拟合法

基于混合高斯分布拟合法的风电功率预测误差估计方法首先假设误差数据的概率密度函数符合混

合高斯分布,然后通过对历史风电功率预测误差数据的统计分析,进行数据拟合,得到样本数据的混合高斯模型,进而设置合理的置信水平求得预测误差的估计区间。该方法适用范围广、精度高^[6]。

一维混合高斯分布模型由多个正态分布线性组成,其概率密度函数表达式为

$$f(x; \mu_1, \dots, \mu_N, \sigma_1, \dots, \sigma_k) = a_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2} + \dots + a_k \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2\sigma_k^2}(x-\mu_k)^2} \quad (1)$$

式中: k 为正态分布的个数; $a_i (1 \leq i \leq k)$ 为各正态分布的权值; μ_i 为各正态分布的均值; σ_i 为各正态分布的标准差。

根据风电场历史误差数据,利用期望最大化(expectation maximization, EM)算法对混合高斯模型进行参数估计,求解式(1)中的 a_i 、 μ_i 和 σ_i ^[6,19]。

设 $\theta_j = (a_j, \mu_j, \sigma_j)$, $j = 1, 2, \dots, k$ 为 GMM 中各正态分布模型的参数, $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$ 为 GMM 中的所有参数。EM 算法主要包含以下两个步骤:

1) 计算样本数据 X 的对数似然函数的最大似然估计值,其对数似然函数如式(2)。

$$L(X | \Theta^{(q)}) = \sum_{i=1}^T \ln \sum_{j=1}^k a_j f_j(X_i; \mu_j, \sigma_j) \quad (2)$$

式中: T 为样本总数; k 为正态分布的个数; $\Theta^{(q)}$ 为第 q 次迭代时计算得到的参数的最大似然估计值。

2) 最大化上一步得到的最大似然估计值,求解参数。

$$\Theta^{(q+1)} = \arg \max L(\Theta^{(q)}) \quad (3)$$

重复迭代上述步骤至 $\Theta^{(q+1)}$ 和 $\Theta^{(q)}$ 之间的差值无穷小,即收敛时,则为最优的模型参数。

根据上述步骤建立基于混合高斯分布拟合法的风电功率预测误差估计模型,然后设置合理的置信水平,即可求得风电功率预测误差估计区间。

2 特征值提取法

基于特征值提取法的风电功率预测误差估计是通过分析与预测误差相关的特征值,将这些特征值与风电功率预测误差联合建立多元线性回归方程,进而求取预测误差的估计区间。

获得风电场的历史运行数据,包括历史预测功率和实际功率以及风电场日前预测功率,从统计学角度出发,利用各种统计学参数提炼出与预测误差

相关的特征值,如日前预测功率波动性、近期风电功率波动性、风功率幅值及预测方法的预测精度等,进行相关性分析^[14-15]。选取相关性较强的特征值,利用多元线性回归方程建立预测误差估计模型。其中,相关性计算使用卡尔·皮尔逊对相关系数 r 的表述如式(4)所示^[14,17]。

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (4)$$

式中: X 、 Y 为研究相关性的两组样本数据; n 为样本总数; X_i 为样本 X 中的样本数据; \bar{X} 为样本 X 的平均值; Y_i 为样本 Y 中的样本数据; \bar{Y} 为样本 Y 的平均值。 r 的取值范围为 $[-1, 1]$, $|r|$ 越大,两者相关性越强。当 $|r|=1$ 时, X 和 Y 线性相关; $r=0$ 时, X 和 Y 不相关。

风电功率预测误差估计模型为

$$E = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n \quad (5)$$

式中: E 为当日风电功率的平均绝对误差; x_1, \dots, x_n 为与 E 强相关的各个特征值; b_0, \dots, b_n 为估计模型的回归系数。

通常来讲,与风电预测功率的平均绝对误差相关性较强的因素有当日风电功率波动、当日风电功率水平和预测方法的预测精度等^[14-16]。其中,当日风电功率波动和当日风电功率水平对预测误差的影响较大^[13]。

用当日的日前预测风电功率的标准方差 S_T 表示当日风电功率波动, S_T 越大,表明当日风电功率波动越大,风电功率预测难度增加,预测精度下降,风电功率预测误差则会增大^[14-16]。 S_T 的具体表达式为^[14-16]

$$S_T = \sqrt{\frac{1}{N_1} \sum_{i=1}^{N_1} (P_{pt} - \bar{P}_{pt})^2} \quad (6)$$

式中: P_{pt} 为每一时刻的预测功率; \bar{P}_{pt} 为当日预测功率的平均值; N_1 为当日预测时刻数,通常风电预测系统15 min预测一次,故 $N_1=96$ 。

用当日预测功率平均值表示当日风电功率水平 P_{ave} ,若当日风电功率水平较大,则风电功率容易变化,增大风电功率预测误差^[14-16]。当日风电功率水平 P_{ave} 如式(7)所示^[14-16]。

$$P_{ave} = \frac{1}{N_1} \sum_{i=1}^{N_1} P_{pt} \quad (7)$$

同样地,若风电功率预测系统的预测精度较低,

风电功率预测误差也会较大^[14-16]。风电功率近期预测精度 P_{re} 如式(8)所示^[14]。

$$P_{re} = \frac{1}{N_2} \sum_{i=1}^{N_2} |P_{mt} - P_{pt}| \quad (8)$$

式中: P_{mt} 为每一时刻的实际功率; N_2 表示近几日预测时刻数,取96的正整数倍。

根据风电场历史运行数据计算得到 S_T 、 P_{ave} 和 P_{re} ,并利用式(4)计算各特征值与平均绝对误差的相关性,选取相关性强的因素和平均绝对误差带入式(5),利用Matlab进行求解,即可得到风电功率预测误差估计模型并求解误差估计区间。

3 算例分析

为了验证评估上述两种方法的有效性,利用西北地区某风电场2018年6月1日至2018年7月30日的风电预测功率数据和实际功率数据,根据前50日的数据样本,分别采用两种方法对2018年7月21日至2018年7月30日进行风电功率预测误差估计。

3.1 混合高斯分布拟合法

基于得到的风电场功率数据,对风电场的功率预测误差数据进行归一化,利用Matlab软件,根据第1节的误差估计模型建立步骤,对预测误差进行混合高斯分布拟合,分别建立二权值和三权值的混合高斯模型,得到的结果如表1所示。

表1 二权值和三权值的混合高斯模型参数值

参数	二权值	三权值
a	$a_1 = 0.933\ 1$	$a_1 = 0.683\ 0$
	$a_2 = 0.066\ 9$	$a_2 = 0.066\ 5$
		$a_3 = 0.250\ 5$
μ	$\mu_1 = -0.104\ 1$	$\mu_1 = -0.096\ 3$
	$\mu_2 = 0.469\ 8$	$\mu_2 = -0.464\ 1$
		$\mu_3 = 0.123\ 6$
σ	$\sigma_1 = 0.164\ 5$	$\sigma_1 = 0.092\ 2$
	$\sigma_2 = 0.164\ 9$	$\sigma_2 = 0.098\ 7$
		$\sigma_3 = 0.289\ 6$

根据表1分别绘制二权值和三权值的混合高斯模型曲线,对比其概率密度分布图(如图1所示)。可以看出,对于该风电场的功率预测误差分布,较二权值混合高斯模型而言,三权值混合高斯模型拟合的概率分布曲线更加准确。

因此,风电功率预测误差模型采用三权值的混合高斯模型,设置95%的置信水平,得到风电功率预测误差估计区间为 $[-25.235\ 1, 24.507\ 5]$,其与实际风电运行数据的对比如图2所示。

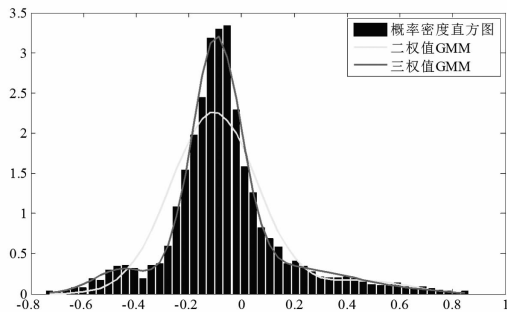


图 1 不同权值混合高斯模型分布拟合图

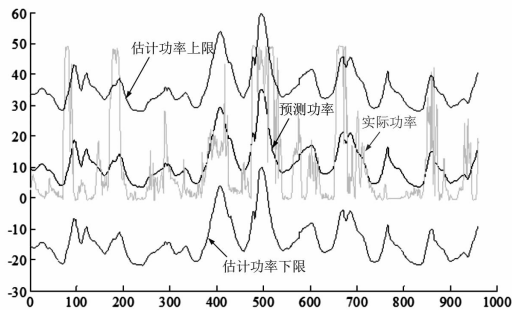


图 2 基于混合高斯分布拟合法的风电功率预测误差区间估计结果

3.2 特征值提取法

同样地,基于得到的风电场功率数据,根据式(6)、式(7)、式(8)分别计算得到特征值 S_T 、 P_{ave} 和 P_{re} ,通过式(4)计算得到 3 个特征值与日平均绝对误差 E 的相关性有: $r_{S_T-E} = 0.5169$ 、 $r_{P_{ave}-E} = 0.7850$ 、 $r_{P_{re}-E} = 0.1264$ 。

可以看出,对于该风电场,当日风电功率波动 S_T 和当日风功率幅值 P_{ave} 与误差的相关性较高。因此,以当日风电功率波动和当日风功率幅值两个特征值作为线性回归方程的自变量来建立如式(5)所示的风电功率预测误差估计模型。

基于以上分析,利用 Matlab 软件,计算得到风电功率预测误差估计模型的线性回归系数及其 95% 的置信区间,如表 2 所示。

表 2 基于特征值提取的风电功率预测误差估计模型回归系数

系数	b_0	b_1	b_2
取值	1.953 8	0.062 6	0.519 1
95% 置信区间	[1.798 6 2.109 0]	[0.028 6 0.096 5]	[0.504 8 0.533 4]

根据表 2 和式(5),求得风电场 2018 年 7 月 21 日至 2018 年 7 月 30 日这 10 天的风电功率预测误差估计区间,结果如图 3 所示。

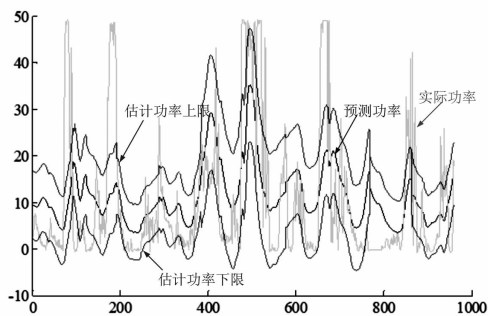


图 3 基于特征值提取的风电功率预测误差区间估计结果

3.3 误差估计区间评价

风电功率预测误差估计区间是决定风电场储能容量配置的关键。利用两个指标对得到的误差估计区间进行评估:区间覆盖率 R_{PICP} 和区间平均带宽 R_{PINAW} ,这两个指标的数学表达式如式(9)^[9,12]所示。

$$\begin{cases} R_{PICP} = \frac{1}{N} \sum_{t=1}^N k_t \\ R_{PINAW} = \frac{1}{NR} \sum_{t=1}^N [U_t(x_t) - L_t(x_t)] \end{cases} \quad (9)$$

式中: N 为样本总数; k_t 为布尔量,当 t 时刻的误差值位于误差估计区间内, k_t 为 1,否则为 0; R 为实际值的变化范围; $U_t(x_t)$ 为 t 时刻误差估计区间的上限; $L_t(x_t)$ 为 t 时刻误差估计区间的下限。

可以看出,对于风电场储能的容量配置, R_{PICP} 决定了容量配置的有效性, R_{PINAW} 决定了容量配置的经济性。对于一个理想的误差估计区间, R_{PICP} 越大,覆盖率越高,储能容量配置则越有效; R_{PINAW} 越小,区间精度越高,储能容量配置则越经济。根据前两个小节分析计算,两种方法的误差估计区间的 R_{PICP} 和 R_{PINAW} 如表 3 所示。

表 3 两种方法的误差估计区间的指标结果

估计方法	$R_{PICP}/\%$	$R_{PINAW}/\%$
混合高斯分布拟合法	92.08	80.37
特征值提取法	58.85	25.43

由表 3 可知,基于特征值提取法的风电功率预测区间估计的区间覆盖率和基于混合高斯分布拟合方法的风电功率预测区间估计的区间覆盖率分别为 58.85% 和 92.08%,后者的区间覆盖率远高于前者。但是基于混合高斯分布拟合方法得到的估计区间的带宽较大,为 80.37%,因此根据该区间配置储能容量的经济性较差,而基于特征值提取法得到的

估计区间的 R_{PINAW} 则较小,为 25.43%。

3.4 MM - GMM 优化预测误差区间估计算法

根据前面的分析可以看出,若根据混合高斯分布拟合得到的误差估计区间进行风电场储能容量配置,其有效性较高但经济性较差;若根据特征值提取法得到的误差估计区间进行储能容量配置,其经济性较好但有效性较低。为了解决上述问题,兼顾储能容量配置时的有效性和经济性,基于混合高斯模型,结合马尔可夫模型,提出了一种 MM - GMM 优化预测误差区间估计算法。

由于风本身固有的随机性和波动性,风电功率预测误差也具有较强的不确定性,风电功率预测通常只与相邻时段有关,而与过去所处的状态无关,因此,考虑在风电预测误差区间估计中引入马尔可夫模型。

若随机过程具有“无后效性”,即在已知现在状态的前提下,其未来的状态与过去的状态无关,则称其为马尔可夫过程^[20-21]。

定义 P 为马尔可夫过程的一步状态转移概率矩阵, P_{ij} 为状态转移矩阵 P 中由状态 S_i 一步转移到状态 S_j 的概率,根据其无后效性有^[20-21]:

$$P_{ij}(t) = P\{S(t+1) = j | S(t) = i\} \quad (10)$$

式中: $S(t) = i$ 表示 t 时刻的状态为 i ; $S(t+1) = j$ 表示 $t+1$ 时刻的状态为 j 。

根据上述分析,结合第 1 节的混合高斯模型的求解过程,MM - GMM 优化预测误差区间估计算法的具体步骤如下:

- 1) 利用 EM 算法求出样本数据中每个时刻的功率预测误差处于混合高斯分布中的具体分布及其均值和标准差;
- 2) 根据计算得到的每个时刻误差所处的分布状态,计算马尔可夫过程的一阶状态转移矩阵 P_E ;
- 3) 根据最近时刻所处的状态,利用计算得到的一阶状态转移矩阵 P_E ,预测未来一天各时刻误差所处的分布状态;
- 4) 设置置信水平,计算预测状态所处分布的对应置信区间即为该时刻的预测误差估计区间。

对所提的优化算法进行算例验证,同样地,利用西北地区某风电场的的数据样本,采用所提的优化算法对 2018 年 7 月 21 日至 2018 年 7 月 30 日进行风电功率预测误差估计。每个状态采用 95% 的置信区间,得到的结果如图 4 所示。

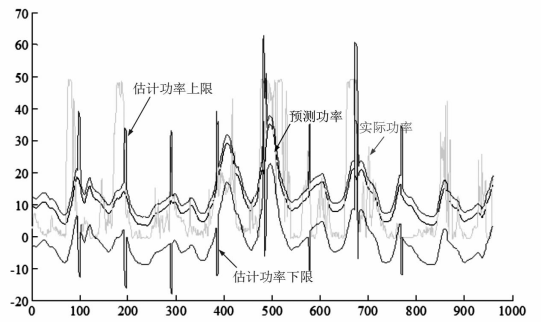


图 4 MM - GMM 优化预测误差区间估计算法的区间估计结果

根据如图 4 所示的估计区间,计算该区间 R_{PICP} 为 71.56%, R_{PINAW} 为 26.01%。3 种方法的误差估计区间指标结果的对比如表 4 所示。可以看出,与前两种方法相比,所提的优化算法不仅有效性较高,且区间较窄,能够为风电场的储能容量配置提供有效的指导信息,经济性较好。

表 4 3 种方法的误差估计区间的指标结果

估计方法	$R_{PICP}/\%$	$R_{PINAW}/\%$
混合高斯分布拟合法	92.08	80.37
特征值提取法	58.85	25.43
MM - GMM 优化估计算法	71.56	26.01

4 结 语

对混合高斯分布拟合法和特征值提取法这两种典型的风电功率预测误差估计算法进行的详细研究表明:基于混合高斯分布拟合的误差估计方法得到的误差估计区间进行储能容量配置的有效性高但经济性较差;基于特征值提取的误差估计方法得到的误差估计区间进行储能容量配置的经济性好但有效性较低。

为了有效估计风电功率预测误差,上面给出对于风电场储能配置具有指导意义的预测误差估计区间,在混合高斯模型的基础上,对风电功率预测误差进行状态划分,结合马尔可夫模型,提出了一种 MM - GMM 优化误差区间估计算法并进行了算例验证。结果表明,基于所提的优化算法得到的估计区间不仅覆盖率较高,而且带宽较小,兼顾了风电场储能容量配置时的有效性和经济性,具有较好的综合性能,能够为改善风电功率预测误差的储能系统容量配置提供理论依据。

(下转第 94 页)