第 52 卷 第 7 期：3174-3188 高电压技术 Vol.52, No.7: 3174-3188
2026 年 7 月 31 日 High Voltage Engineering July 31, 2026
DOI: 10.13336/j.1003-6520.hve.20251059

基于 XGBoost-KDE 的谐波源不确定性通用模型

王嘉琦[1,2]，陈红坤[1,2]，褚昱麟[1,2]，陈　磊[1,2]
（1. 交直流智能配电网湖北省工程中心(武汉大学)，武汉 430072;
2. 武汉大学电气与自动化学院，武汉 430072）

摘　要：在新型电力系统建设背景下，随着新能源的大规模并网与非线性设备的广泛应用，电网谐波问题呈现出更强的随机性和波动性特征。构建谐波源不确定性通用模型，是评估谐波影响和优化治理策略的基础。然而现有的模型计算复杂度高，难以充分发挥大样本数据的优势，导致模型精度受限。为解决上述问题，提出了一种基于XGBoost-KDE 的谐波源不确定性通用模型建立方法。首先，通过极端梯度提升(eXtreme gradient boosting, XGBoost)算法建立谐波源谐波电流与谐波发射特性相关的物理量之间的映射关系；其次，采用核密度估计方法(kernel density estimation, KDE)得到预测误差的概率密度函数和概率分布函数，进而构造预测误差区间；然后，结合点预测结果求解一定置信水平下谐波电流的区间预测结果；最后，提出一种基于自适应双策略的模型在线更新方法，使得模型更为准确地跟踪谐波源谐波特性的动态变化。经过实测和仿真数据验证，相比于其他的建模方法，所提建模方法在单一谐波源建模和复杂多谐波源场景下具有较高的精度和较强的通用性，所提模型在线更新方法能够利用更多的历史样本数据进行模型动态更新，进而提高了模型的适用性。

关键词：电能质量；极端梯度提升；核密度估计；谐波源模型；不确定性

General and Uncertain Harmonic Source Model Based on XGBoost-KDE

WANG Jiaqi[1,2], CHEN Hongkun[1,2], CHU Yulin[1,2], CHEN Lei[1,2]
(1. Hubei Engineering and Technology Research Center for AC/DC Intelligent Distribution Network (Wuhan University), Wuhan 430072, China; 2. School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China)

Abstract：In the context of the construction of new power systems, with the large-scale grid connection of new energy and the widespread application of nonlinear equipment, the harmonic problems in the power grid exhibit stronger randomness and volatility characteristics. Building a general model for uncertainty of harmonic sources is the foundation for evaluating the impact of harmonics and optimizing governance strategies. However, existing models have high computational complexity and are difficult to fully leverage the advantages of large sample data, resulting in limited model accuracy and difficulty in meeting practical engineering requirements. To solve the above problems, a method for establishing a general model with uncertain harmonic source based on XGBoost-KDE was proposed. Firstly, the mapping relationship between the physical quantities related to the harmonic current and harmonic emission characteristics of the harmonic source was established through the eXtreme gradient boosting (XGBoost) algorithm. Secondly, the kernel density estimation (KDE) method was used to obtain the probability density function and probability distribution function of the prediction error and to construct the prediction error interval. Then, based on the point prediction results, the interval prediction results of harmonic current at a certain confidence level were solved. Finally, a model online updating method based on adaptive dual strategy was proposed, which enabled the model to more accurately track the dynamic changes of harmonic characteristics of harmonic sources. Through actual testing and simulation data verification, compared with other modeling methods, the proposed modeling method has high accuracy and strong universality in single harmonic source modeling and complex multi harmonic source scenarios. The proposed online model updating method can use more historical sample data for dynamic model updates, thereby improving the applicability of the model.

Key words：power quality; eXtreme gradient boosting; kernel density estimation; harmonic source model; uncertainty

基金资助项目：国家自然科学基金联合基金重点项目(U23B20117)；中国南方电网有限责任公司科技项目(000005KK52220039)。
Project supported by the Joint Fund Key Program of National Natural Science Foundation of China (U23B20117), Science and Technology Project of China Southern Power Grid Co., Ltd. (000005KK52220039).

0 引言

在构建以新能源为主体的新型电力系统进程中，随着风电、光伏等可再生能源发电占比的持续提升，以及工业领域中电力电子装置、电弧炉的大规模应用，电网中存在的谐波污染问题日益严峻，威胁着电力系统的安全性和用电设备的正常运行[1-5]。我国新实行的《电能质量管理办法(暂行)》加强了对电能质量的监管力度，对电力系统谐波水平管理提出了更高的要求[6]。为了有效分析与治理电力系统中的谐波问题，必须采用合适的模型来表征谐波源的谐波发射特性[7-9]。然而，电力系统中的谐波源种类众多，各种类型谐波源的谐波产生机理也互不相同[10-13]，如果对各类谐波源逐一深入分析并构建模型，不仅建模效率较低，其适用范围也有限，因此，研究通用模型已成为谐波源建模领域的重要方向。谐波源通用模型可分为确定性模型和不确定性模型。确定性模型以精确数值表征谐波发射特性，无法反映未被考虑的物理因素对谐波源特性的不确定性影响，导致谐波评估与治理方案无法根据系统保守度和冗余度的实际需求进行灵活调整。而不确定性模型通常以概率分布或区间范围表征谐波源谐波发射特性，更符合实际需求。因此，建立谐波源不确定性通用模型具有重要意义。

目前针对谐波源不确定性通用模型的研究主要有基于概率统计方法和基于机器学习方法。基于概率统计的方法通常利用拟合函数得到谐波电流的概率密度[14-16]，然而在电网实际运行中，谐波发射特性受到多种因素的综合影响，导致样本维度较高。在高维空间中，数据点之间的距离会增大，导致局部邻域内有效样本数量减少，从而使得概率统计模型难以准确捕捉数据的分布特征，进而影响概率统计建模的效果。此外，若仅利用一维数据进行建模，则难以充分反映各变量之间的耦合关系。相比之下，基于机器学习的谐波源不确定性建模方法能够更好地挖掘高维数据中复杂的特征关系，从而建立输入变量与输出变量之间的不确定性关系。文献[17]将基于机器学习的概率式预测方法应用到谐波源建模中，将谐波源谐波电流与谐波电压之间的映射关系融入到高斯过程回归模型(Gaussian processes regression, GPR)中，通过求解模型参数对谐波电流进行区间预测，所提方法可以刻画谐波源谐波发射特性的不确定性行为，为谐波源不确定性建模研究提供了一种新颖的技术路线，然而，GPR 模型中核矩阵的求逆计算使得其计算复杂度为 $O(n^3)$，随着训练样本数量 n 增加，计算复杂度呈立方级增长，难以满足快速计算的要求。此外，基于少量样本数据训练的 GPR 模型在实际工程应用中难以适应复杂多变的负荷运行工况，使得模型需要频繁重复更新，从而影响了模型的可靠性。

随着电力系统的数字化和智能化发展，大量的谐波监测设备被广泛应用，海量的谐波量测数据得以采集，这为基于大样本数据的谐波源建模提供了丰富的数据基础。因此，如何充分发挥大样本数据的优势以提升当前谐波源不确定性通用模型的可靠性仍是目前谐波源不确定性建模存在的一大难题。近年来，基于机器学习的混合构造区间法在不确定性分析领域得到了应用[18-20]，该方法利用机器学习模型挖掘输入与输出特征之间的复杂非线性关系，获取模型的点预测结果，再利用核密度估计(kernel density estimation, KDE)对点预测误差进行统计分析，进而构建出区间预测结果，这种方法具有广泛的应用范围和较强的鲁棒性，为谐波源不确定性通用模型建立提供了一种新的思路。

为此，本文提出了一种基于 XGBoost-KDE 的谐波源不确定性通用模型建立方法。该方法基于极端梯度提升(eXtreme gradient boosting, XGBoost)算法建立谐波发射特性相关的物理量与谐波电流之间的映射关系，利用核密度估计方法得到谐波电流点预测误差的概率密度函数和概率分布函数，在给定的置信水平下，计算并得到谐波电流的区间预测范围，从而实现了谐波源不确定性通用模型的建立。相较于准确值，区间预测通过范围而非单一数值呈现结果，能够更全面地反映谐波特性的不确定性行为。在此基础上，提出了一种基于自适应双策略的模型在线更新方法。通过与其他建模方法进行对比，运用光伏电站的实测数据以及电解铝负荷、电弧炉和多谐波源复杂网络的仿真数据，验证了本文所提建模方法在单一谐波源建模和复杂多谐波源场景下具有较高的精度和较强的通用性，同时，验证了所提模型在线更新方法能够使模型更好地适应重复性的工况变化，从而提高了模型的适用性。

1 研究方法的基本原理

1.1 谐波源通用模型建立方法

谐波源通用模型是用来描述谐波电流和相关

物理量之间的映射关系，可通过以下式表示：

$$\boldsymbol{I}_h = F_h\left(\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3,\cdots,\boldsymbol{U}_h,C\right) \tag{1}$$

式中：h 表示为谐波次数；$\boldsymbol{I}_h$ 为谐波源产生的各次谐波电流相量；$\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3,\cdots,\boldsymbol{U}_h$ 为谐波源的基波电压相量和各次谐波电压相量；$F_h(\cdot)$ 表示为从输入变量到输出变量的映射关系；C 为特征参数集合，表示与谐波源谐波发射特性相关的参数，目前特征参数集合 C 的选取暂无统一的规范，需结合实际应用场景与量测数据的可获取性进行灵活选取，例如：基波电流、换流器的直流母线电压与控制模式、以及电弧炉的炉温与炉膛气压等。

谐波源频谱分布广，若直接基于全频域谐波电压特征构建谐波源模型，不仅会提高模型的复杂度，还会增加样本集数据存储的压力。因此，为了更有针对性地提取谐波源的频域特征，本文分析并筛选出主导频次谐波电压作为模型的输入。需要注意的是，虽然在频域中部分非主导频次的谐波信息被忽略，但仍能据此构建出有效的谐波源模型，且可提高模型的建模效率。

目前，多数谐波在线监测装置的谐波频次采集范围为 2~50 次，且在大部分工业与民用负荷中，谐波源通常选取前 50 次谐波分量作为主要研究对象，故本文仅确定前 50 次谐波中的主导谐波频次。根据谐波监测装置的采样数据，通过计算各次谐波电压有效值与基波电压有效值的百分比，即可求得 h 次谐波电压含有率 η_h。为减少谐波电压含有率较低的特征对模型的影响，选取谐波电压含有率大于 1%的 h 次谐波电压 $\boldsymbol{U}_h$ 作为主导频次谐波电压，并将其纳入主导频次谐波电压集合 U_{main} 中，具体选取准则如下式所示：

$$\begin{cases}\boldsymbol{U}_h \in U_{\text{main}}, & \eta_h > 1\% \\ \boldsymbol{U}_h \notin U_{\text{main}}, & \eta_h \le 1\%\end{cases} \tag{2}$$

将式(2)中各次谐波电压向量集合替换为主导频次谐波电压集合，则通用谐波源模型可表示为：

$$\boldsymbol{I}_h = F_h\left(\boldsymbol{U}_1,U_{\text{main}},C\right) \tag{3}$$

需要说明的是，在动态谐波源模型的在线更新过程中，由于谐波源的运行状态变化复杂多样，主导频次谐波电压集合可能发生改变，因此其选取需根据具体工况进行分析。

1.2 极端梯度提升算法

极端梯度提升是一种基于决策树的机器学习算法。XGBoost 继承了传统梯度提升树的优点，并在此基础上进行改进和优化，使得其在计算精度和训练时间等方面具有优势，因此，它被广泛应用于解决回归预测问题。

XGBoost 算法采用的决策树是分类与回归树(classification and regression tree, CART)，该算法通过对 K 个决策树的预测值进行求和，最终得到预测输出值。假设训练数据集包含 n 个样本$(x_i, y_i)(i=1, 2,\cdots, n)$，XGBoost 算法的预测结果可以通过以下式表示：

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F \tag{4}$$

式中：x_i 为输入的第 i 个样本数据；$\hat{y}_i$ 为第 i 个样本的预测值；$f_k(x_i)$ 表示为第 k 个决策树的预测值；f_k 为函数空间 F 中的一个函数；F 表示所有可能的决策树集合。

XGBoost 算法的目标函数 L_{Obj} 由两部分构成，分别是损失函数与正则化惩罚项，其表达式为：

$$L_{\text{Obj}} = \sum_{i=1}^{n} l(y_i,\hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{5}$$

$$\Omega\left(f_k\right) = \gamma T_{\text{num}} + \frac{1}{2}\lambda \| \omega \|^2 \tag{6}$$

式中：$l(y_i,\hat{y}_i)$ 为损失函数，其值为预测值 $\hat{y}_i$ 与实际值 y_i 之间的误差；$\Omega(f_k)$ 为正则化项，用于防止模型过拟合；T_{num} 为叶子节点的个数；ω 为叶子节点的权重；γ 和 λ 为惩罚系数。

XGBoost 算法的基本思想是不断拆分要素以添加树，添加树是为了添加新的加法函数 $f_r\left(x_i\right)$，用于拟合上一次预测的残差，相关表达式如下：

$$\hat{y}_i^{(r)} = \hat{y}_i^{(r-1)} + f_r\left(x_i\right) \tag{7}$$

式中：$\hat{y}_i^{(r)}$ 为在第 r 轮迭代的预测值；$\hat{y}_i^{(r-1)}$ 为上一轮迭代的预测值。

基于式(7)，可将目标函数改写为式(8)。随后，利用快速优化的二阶近似方法对目标函数进行推导，最终得到如式(9)所示的表达式，相关式如下：

$$L_{\text{Obj}}^{(r)} = \sum_{i=1}^{n} l\left(y_i,\hat{y}_i^{(r-1)} + f_r\left(x_i\right)\right) + \Omega\left(f_r\right) \tag{8}$$

$$L_{\text{Obj}}^{(r)} \approx \sum_{i=1}^{n}\left(l\left(y_i,\hat{y}_i^{(r-1)}\right) + g_i f_r\left(x_i\right) + \frac{1}{2}h_i f_r^2\left(x_i\right)\right) + \Omega\left(f_r\right) \tag{9}$$

$$g_i = \partial_{\hat{y}_i^{(r-1)}} l\left(y_i,\hat{y}_i^{(r-1)}\right) \tag{10}$$

$$h_i = \partial^2_{\hat{y}_i^{(r-1)}} l\left(y_i,\hat{y}_i^{(r-1)}\right) \tag{11}$$

式中：g_i 和 h_i 分别为损失函数的一阶梯度和二阶梯度；$\partial_{\hat{y}_i^{(r-1)}}$ 表示为对预测值 $\hat{y}_i^{(r-1)}$ 求一阶偏导；$\partial^2_{\hat{y}_i^{(r-1)}}$ 表

示为对预测值 $\hat{y}_i^{(r-1)}$ 求二阶偏导。

因此，XGBoost 在训练过程中通过不断向现有的决策树中添加新的决策树来提高模型精度，并利用贝叶斯优化调整超参数，训练完成后，模型的结构和参数便得到确定。当需要对新样本进行回归预测时，模型中的决策树将输入特征映射到输出结果，从而得到该样本的回归预测值。

1.3 核密度估计方法

核密度估计是一种基于数据样本本身的非参数估计方法，无需对样本数据的分布做出先验假设，能够直接估计样本的概率密度函数。对于随机变量 t 的样本集$\{t_1, t_2, \cdots, t_n\}$，则通过 KDE 方法估计的概率密度函数(probability density function, PDF)的表达式为：

$$\hat{f}_{\mathrm{T}}(t)=\frac{1}{nl_{\mathrm{bw}}}\sum_{i=1}^{n}K_{\mathrm{s}}\left(\frac{t-t_i}{l_{\mathrm{bw}}}\right) \tag{12}$$

式中：$\hat{f}_{\mathrm{T}}(t)$ 为估计的 PDF；n 为样本数量；l_{bw} 为核函数的带宽；$K_{\mathrm{s}}(\cdot)$ 为核函数；t_i 为第 i 个样本值。

核密度估计的好坏取决于核函数和带宽的选择。常用的核函数包括三角核函数、正态核函数以及 Epanechnikov 核函数等。文献[21]研究表明，不同核函数的选择对核密度估计的准确性影响较小。本文选取常用的正态核函数，核函数式为：

$$K_{\mathrm{s}}(u)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{u^2}{2}} \tag{13}$$

式中：u 为自变量。

选择合适的带宽是核函数密度估计的关键，它会直接影响到曲线的拟合程度。常用的优选带宽确定原则是使估计 PDF 与真实 PDF 之间的平均积分平方误差(mean integral squared error, MISE)最小，MISE 的表达式为：

$$E_{\mathrm{MISE}}\left(\hat{f}_{\mathrm{T}}(t)\right)=E\left(\int\left(\hat{f}_{\mathrm{T}}(t)-f_{\mathrm{T}}(t)\right)^2\mathrm{d}t\right) \tag{14}$$

式中：$f_{\mathrm{T}}(t)$ 为真实的 PDF；$E(\cdot)$ 为数学期望值。

对式(14)进行泰勒展开并省略无穷小量得到渐进平均平方积分误差(asymptotic mean integral square error, AMISE)，然后根据求解函数极值的思路，将 AMISE 对带宽进行求导，得到 AMISE 取最小值时所对应的最优带宽 $l_{\mathrm{bw}}^{\mathrm{best}}$，其表达式为：

$$l_{\mathrm{bw}}^{\mathrm{best}}=\left(\frac{\int K_{\mathrm{s}}^{\,2}(t)\mathrm{d}t}{n\left(\int t^2K_{\mathrm{s}}(t)\mathrm{d}t\right)^2\int f_{\mathrm{T}}^{''}(t)^2\,\mathrm{d}t}\right)^{\frac{1}{5}} \tag{15}$$

式中：$f_{\mathrm{T}}''(t)$ 表示为 $f_{\mathrm{T}}(t)$ 的二阶导数。

当核函数为正态核函数时，通常利用 Silverman 的拇指法则[22]，进一步简化式(15)为：

$$l_{\mathrm{bw}}^{\mathrm{best}}=\left(\frac{4}{3n}\right)^{\frac{1}{5}}\sigma \tag{16}$$

式中：σ 为样本数据的标准差。

2 基于 XGBoost-KDE 的谐波源不确定性通用模型

2.1 建模流程

为了反映谐波源谐波特性的不确定性行为，本文提出一种基于 XGBoost-KDE 的谐波源不确定性通用模型，首先利用基于 XGBoost 的谐波源确定性模型对谐波电流进行点预测，进而结合 KDE 算法实现谐波电流的区间预测。以预测某次谐波电流的幅值和相角为例，具体建模流程如下。

步骤 1：获取谐波源端口处的谐波电流、基波电压、主导频次谐波电压以及特征参数作为样本数据，按 7:2:1 的比例划分为训练集、验证集和测试集，并采用最小-最大值归一化方法对数据进行归一化处理。

步骤 2：将谐波电流作为模型的输出变量，与谐波特性相关的物理量作为模型的输入变量，通过 XGBoost 算法挖掘输入变量与输出变量之间的映射关系，并利用训练集和验证集对模型进行训练和评估，进而完成模型的建立，并计算验证集的预测值与实际谐波电流的点预测误差，计算式如下：

$$\begin{cases}\overline{I}_i=\breve{I}_i-\hat{I}_i\\ \overline{\theta}_i=\breve{\theta}_i-\hat{\theta}_i\end{cases} \tag{17}$$

式中：$\overline{I}_i$ 和 $\overline{\theta}_i$ 分别为第 i 个样本谐波电流幅值和相角的点预测误差；$\hat{I}_i$ 和 $\hat{\theta}_i$ 分别为验证集中第 i 个样本幅值和相角的点预测结果；$\breve{I}_i$ 和 $\breve{\theta}_i$ 分别为第 i 个样本幅值和相角的真实值。

步骤 3：考虑在不同运行状态下，谐波源所发射的谐波电流大小可能存在较大差异，谐波电流的预测误差分布也可能有所不同。因此需要根据谐波电流大小的实际情况建立分段 KDE 误差模型，以实现更精准的不确定性量化。具体而言，根据谐波电流幅值和相角的大小，分别将其划分为 Z 和 W 个区间段，其中，幅值区间段编号为 z=1, 2,…, Z, 相角区间段编号为 w=1, 2,…, W。理论上，分段越多，KDE 拟合精度越高，但同时也会增加计算复杂性，

因此在保证拟合优度检验的有效性前提下，区间段的划分通常需要在精度和效率之间进行平衡[14]。

步骤 4：对每个区间段的幅值和相角的预测误差数据进行核密度估计，得到各区间段幅值和相角的概率密度分布函数 $f_{\bar{\mathrm{I}}}^{z}\left(\bar{I}\right)$ 和 $f_{\bar{\theta}}^{w}\left(\bar{\theta}\right)$，进而通过积分计算，获得各区间段幅值和相角的累计分布函数 $F_{\bar{\mathrm{I}}}^{z}\left(\bar{I}\right)$ 和 $F_{\bar{\theta}}^{w}\left(\bar{\theta}\right)$。

步骤 5：通过计算累计分布函数的反函数，可以得到在置信水平 $(1-\alpha)$ 下各个区间段的幅值和相角的预测误差区间，计算式如下：

$$\begin{cases}\bar{I}_{\mathrm{up}}^{z}=G_{\bar{\mathrm{I}}}^{z}\left(1-\alpha/2\right)\\ \bar{I}_{\mathrm{low}}^{z}=G_{\bar{\mathrm{I}}}^{z}\left(\alpha/2\right)\end{cases}\tag{18}$$

$$\begin{cases}\bar{\theta}_{\mathrm{up}}^{w}=G_{\bar{\theta}}^{w}\left(1-\alpha/2\right)\\ \bar{\theta}_{\mathrm{low}}^{w}=G_{\bar{\theta}}^{w}\left(\alpha/2\right)\end{cases}\tag{19}$$

式中：$G_{\bar{\mathrm{I}}}^{z}(\cdot)$ 和 $G_{\bar{\theta}}^{w}(\cdot)$ 分别为 $F_{\bar{\mathrm{I}}}^{z}(\cdot)$ 和 $F_{\bar{\theta}}^{w}(\cdot)$ 的反函数；$\bar{I}_{\mathrm{up}}^{z}$ 和 $\bar{I}_{\mathrm{low}}^{z}$ 分别为第 z 个区间段的幅值预测误差区间的上限值和下限值；$\bar{\theta}_{\mathrm{up}}^{w}$ 和 $\bar{\theta}_{\mathrm{low}}^{w}$ 分别为第 w 个区间段的相角预测误差区间的上限值和下限值。

步骤 6：将测试集输入到训练好的 XGBoost 模型中，对谐波电流进行确定性点预测，将确定性点预测结果与相对应区间段的预测误差区间相叠加，得到预定置信水平下的谐波电流预测区间。具体计算式如下：

$$\begin{cases}I_i^{\mathrm{up}}=\tilde{I}_i+\bar{I}_{\mathrm{up}}^{z}\\ I_i^{\mathrm{low}}=\tilde{I}_i+\bar{I}_{\mathrm{low}}^{z}\end{cases}\tag{20}$$

$$\begin{cases}\theta_i^{\mathrm{up}}=\tilde{\theta}_i+\bar{\theta}_{\mathrm{up}}^{w}\\ \theta_i^{\mathrm{low}}=\tilde{\theta}_i+\bar{\theta}_{\mathrm{low}}^{w}\end{cases}\tag{21}$$

式中：I_i^{up} 和 I_i^{low} 分别为第 i 个样本谐波电流幅值的区间预测范围的上界和下界；$\tilde{I}_i$ 为第 i 个样本谐波电流幅值的点预测值；θ_i^{up} 和 θ_i^{low} 分别为第 i 个样本谐波电流相角的区间预测范围的上界和下界；$\tilde{\theta}_i$ 为第 i 个样本谐波电流相角的点预测值。

2.2 模型评价指标

评估谐波源不确定性通用模型的性能应从两方面考虑。首先，点预测结果应尽可能接近真实值。其次，预测区间的覆盖率应尽量大，而预测区间的宽度则应尽可能小。基于上述考虑，本文采用 3 种常用指标来评价模型的预测性能[23-24]。此外，为了评估核密度估计方法的拟合效果，本文选用应用较为广泛的 Kolmogorov-Smirnov(K-S)拟合优度检验作为评价指标[25-26]。

1）平均绝对百分比误差

平均绝对百分比误差(mean absolute percentage error, MAPE)是一种衡量点预测结果准确性的指标，用于反映点预测值偏离实际值的程度，其值越小越好。MAPE 的计算表达式如下：

$$E_{\mathrm{MAPE}}=\frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{y}_i-y_i\right|}{y_i}\times 100\%\tag{22}$$

式中：y_i 为真实值；$\hat{y}_i$ 为点预测值。

2）预测区间覆盖率

预测区间覆盖率(prediction intervals coverage probability, PICP)表示为预测区间能够覆盖真实值的比例，用于评价预测区间的可靠性和准确性，其值越大越好。PICP 的计算表达式如下：

$$R_{\mathrm{PICP}}=\frac{1}{n}\sum_{i=1}^{n}c_i\tag{23}$$

$$c_i=\begin{cases}1, & I_i\in\left[R_i^{\mathrm{low}},R_i^{\mathrm{up}}\right]\\ 0, & I_i\notin\left[R_i^{\mathrm{low}},R_i^{\mathrm{up}}\right]\end{cases}\tag{24}$$

式中：R_i^{up} 和 R_i^{low} 分别是第 i 个样本的预测区间上界和下界；c_i 表示第 i 个谐波电流真实值 I_i 是否位于模型的预测区间内。

3）预测区间归一化平均宽度

预测区间归一化平均宽度(prediction interval normalized average width, PINAW)表示为预测区间的宽窄程度，用于评价预测结果聚焦不确定信息的能力。PINAW 越窄，不确定信息越集中，其值越小越好。PINAW 的计算表达式为：

$$E_{\mathrm{PINAW}}=\frac{1}{nR}\sum_{i=1}^{n}(R_i^{\mathrm{up}}-R_i^{\mathrm{low}})\tag{25}$$

式中：R 为真实值的最大值与最小值之差，用于归一化。

4）Kolmogorov-Smirnov 拟合优度检验

K-S 检验以概率分布估计 $F_{\mathrm{o}}\left(\bar{I}_i\right)$ 与经验分布 $F_n\left(\bar{I}_i\right)$ 之间的差异性作为检验统计量。以某次谐波电流幅值的预测误差为例，将点预测误差 $\bar{I}$ 按照从小到大排序，记作 $\bar{I}_1\le\bar{I}_2\le\cdots\le\bar{I}_n$，则经验分布函数定义为：

$$F_n\left(\bar{I}\right)=\begin{cases}0, & \bar{I}\le\bar{I}_1\\ \dfrac{q}{n}, & \bar{I}_q\le\bar{I}<\bar{I}_{q+1}\\ 1, & \bar{I}\ge\bar{I}_n\end{cases}\tag{26}$$

式中：q=1, 2, …, n−1。

K-S 检验的统计量 H_n 定义为：

$$H_n = \max_{1\le i\le n}\left|F_n\left(\bar{I}_i\right) - F_0\left(\bar{I}_i\right)\right| \tag{27}$$

K-S 检验的临界值可利用经验公式 $1.36/\sqrt{n}$ (置信水平为 95%)计算，当 H_n 小于临界值时，认为通过拟合优度检验。

3 基于自适应双策略的模型在线更新方法

在实际电力系统中，谐波源的谐波发射特性会随着运行工况变化而改变，而静态的谐波源模型在离线模式下无法跟踪谐波源谐波特性的动态变化，从而导致模型精度的下降。因此，需基于实时的在线监测数据对模型进行更新。此外，具备高实时性的谐波源模型能够实现有效的实时仿真[27]。针对上述问题，本文提出一种基于自适应双策略的模型在线更新方法。

该方法通过新增和删除两种更新策略对模型进行更新，当模型满足一定条件时，自适应地执行相应的更新策略修正模型参数，从而确保模型重新满足精度要求，有效应对因谐波源运行工况变化引起的模型精度下降的问题，实现模型的实时精准映射。两种更新策略具体如下。

1）新增策略

新增策略能够使谐波源模型学习新的映射关系，从而适应新的运行工况。新增策略的具体操作为：从实时数据集中随机选取 70%的样本加入原样本集的训练集，其余 30%加入原样本集的验证集，然后基于新构建的样本集重新训练模型，并更新模型参数。

2）删除策略

采用新增策略后，随着模型的不断更新，建模样本数量逐渐增多，导致计算量显著增大，从而影响在线更新的速度。为了满足实时性要求[17]，依据 DL/T1297—2013《电能质量监测系统技术规范》，谐波监测数据的记录周期设定为 3 s，如果建模计算时间超过 3 s，则执行删除策略更新模型，以减少数据存储空间的消耗和提高模型的运算速度。

删除策略的具体操作为：将新样本加入样本集后，计算当前新样本与原样本集中每个样本的欧氏距离，找到与新样本最为相似的旧样本，然后从样本集中删除旧样本。

值得注意的是，谐波源运行工况的变化可能导致谐波电压的主导频次发生改变，若未对实时谐波监测数据中的主导频次谐波电压特征与样本集中已有特征进行匹配，而直接将新的主导频次特征输入模型，不仅导致模型无法有效预测，还将造成计算资源的大量浪费。因此，需要检测实时谐波监测数据中的主导频次是否与样本集匹配。此外，当将谐波源工况变化后的数据输入至基于工况变化前数据建立的模型时，XGBoost 模型的点预测误差特别大，使得 KDE 生成的预测误差区间宽度很大，导致实测数据可能落在模型输出的预测区间内，但在一定程度上失去了实际工程应用的意义。因此，根据现有的研究[28-29]，若模型的点预测指标 MAPE 大于 20%，则通常认为模型劣质，预测结果无效，视为未能覆盖真实值。

综上，本文所提的模型在线更新方法如图 1 所示，具体执行步骤如下。

步骤 1：从历史谐波监测数据或测试数据中筛选出主导频次谐波电压集合，将主导频次谐波电压和其他物理量共同构成模型输入的样本集。将样本集按 7:3 的比例随机划分为训练集和验证集，建立初始的 XGBoost-KDE 模型。

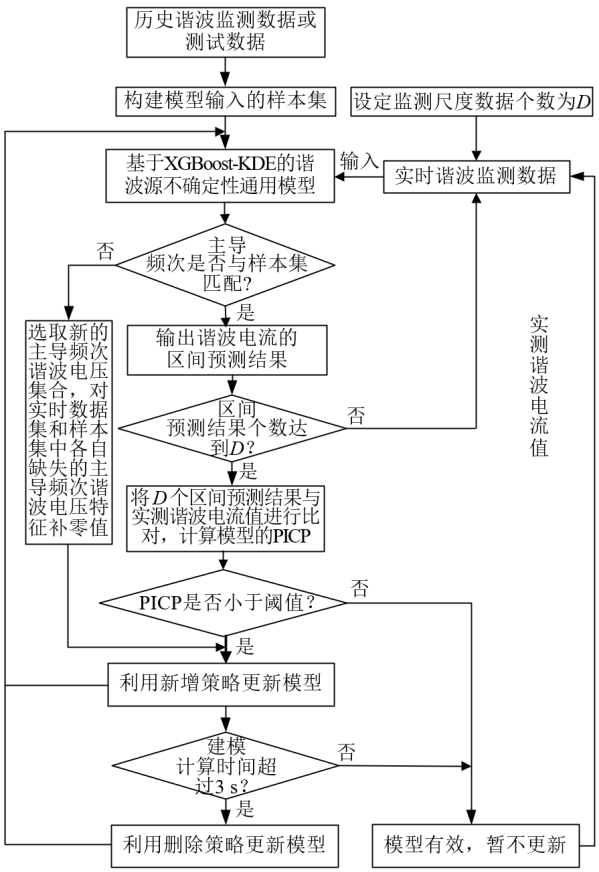


图 1 基于自适应双策略的模型在线更新方法流程图

Fig.1 Flowchart of online model updating method based on adaptive dual strategy

步骤 2：输入实时监测数据，判断其中的主导频次谐波电压特征是否与样本集中已有的主导频次谐波电压特征相匹配。如果不匹配，则采集 D 个实时谐波监测数据构建实时数据集，将实时数据集中主导频次谐波电压特征和样本集中已有的特征进行合集选取，以此合集作为新的主导频次谐波电压特征集合，并对实时数据集与样本集中各自缺失的主导频次谐波电压特征以 0 值填充以保持特征维度一致，再利用新增策略更新模型，并跳转到步骤 5；如果匹配，则直接利用现有模型输出谐波电流的区间预测结果。

步骤 3：判断模型输出的区间结果个数是否达到 D 个，如果达到 D 个，则利用实时谐波电流值计算 D 个区间预测结果的 PICP。

步骤 4：如果模型的 PICP 小于阈值，则利用新增策略更新模型，并跳转到步骤 5；否则模型无需更新。

步骤 5，如果原有的训练样本集新增 D 个样本后，建模计算时间超过 3 s，则利用删除策略更新模型，否则模型无需更新。

4 算例分析

为了验证所提的谐波源不确定性通用建模方法在单一谐波源建模和复杂多谐波源场景中的有效性，本文选取电力电子型谐波源、电弧炉、多谐波源复杂网络作为研究对象，分别基于实测数据和仿真数据对所提建模方法进行测试。此外，通过改变多谐波源复杂网络中部分设备的运行工况来模拟谐波特性的变化，进而验证所提出模型在线更新方法的可行性。本文采用的仿真软件为 MATLAB 2024a，借助 Parallel Computing Toolbox 启用并行计算模式以提升计算效率，处理器配置为 Intel Core i9-14900HX，RAM 为 32 GB。

4.1 电力电子型谐波源实测和仿真数据验证

4.1.1 逆变型电力电子谐波源实测数据验证

为了验证本文所提方法用于逆变型电力电子谐波源的有效性，以光伏电站作为建模对象。选取贵州省某 150 MW 光伏电站并网处的宽频测量装置所记录的实测数据进行分析验证，该宽频测量装置的谐波数据记录周期为 1 s，记录频率范围为 50~2 500 Hz。从 2024 年 4 月 25 日 9 时至 12 时采集的数据样本中选出 2 000 个 A 相宽频谐波数据样本用于建模，将数据集按照 7:2:1 的比例划分为训练集、验证集和测试集。电力电子型谐波源的谐波发射特性与基波电流存在较强的相关性，利用基波电流能够较好地表征该类谐波源的谐波特性[30]，因此，将基波电流作为电力电子型谐波源的特征参数。为了简化模型结构并提高计算效率，根据第 1.1 节所提方法，选取谐波电压含有率较高的谐波分量作为主导频次，因此，模型的输入变量为 5、7、11、13 次谐波电压的幅值和相角、基波电压和基波电流的幅值和相角。本节以幅值较大的 5 次和 7 次谐波电流作为预测目标。本文设定置信水平为 95%，并采用 XGBoost-KDE 方法进行建模，以 5 次谐波电流的幅值 I_5 和相角 θ_5 为例，附录 A 中给出了 KDE 拟合的概率密度结果及拟合优度检验结果。为了验证本文所提模型的优越性，建立另外两种模型进行对比，第 1 种为现有的基于 GPR 的谐波源不确定性通用模型[17]，第 2 种是在传统的基于支持向量机(support vector machine, SVM)的谐波源确定性通用模型上[31]，结合本文所提的 KDE 方法，构建基于 SVM-KDE 的谐波源不确定性通用模型。表 1 列出了 3 种逆变型电力电子谐波源模型的 PICP 和 PINAW 指标对比。图 2 的 MAPE 对比结果清晰地展示了各个模型在性能上的差异。

从附录 A 可以看出，本节将 5 次谐波电流的幅值和相角分别划分为 3 和 8 个区间段，核密度估计模型能够有效地刻画各个区间段内谐波电流预测误差的分布规律，且通过 K-S 拟合优度检验。此外，为了确保对比实验的公平性和可靠性，本文构建的 XGBoost-KDE 和 SVM-KDE 模型均通过 K-S 拟合优度检验。

从表 1 和图 2 中可以得出，与 SVM-KDE 和 GPR 模型相比，XGBoost-KDE 模型在 PICP 和 PINAW 两个指标上表现较为优异，能够在保证较高预测区间覆盖率的同时，维持较窄的预测区间宽度。相比之下，SVM-KDE 模型的 PICP 指标较低且 PINAW 较大，这是由于 SVM 模型的点预测性能较差，制约了基于其构建的不确定性模型性能。而 XGBoost-KDE 模型的 MAPE 指标均小于其他两种模型，表明本文所提模型不仅在不确定性建模方面表现最佳，在需要精确值估计的场合同样具有良好的精度。此外，GPR 建模和 SVM-KDE 建模平均耗时分别为 4.45 s 和 4.62 s，而 XGBoost-KDE 建模平均耗时为 0.28 s，具有最佳的建模效率。

表 1　3 种逆变型电力电子谐波源模型的区间预测评价指标对比

Table 1　Comparison of interval prediction evaluation indicators for three inverter-type power electronic harmonic source models

预测对象	评价指标	SVM-KDE	GPR	XGBoost-KDE
I_5	PICP	0.965	0.975	0.980
	PINAW	0.173	0.136	0.079
θ_5	PICP	0.965	0.970	0.970
	PINAW	0.040	0.028	0.013
I_7	PICP	0.945	0.955	0.965
	PINAW	0.179	0.171	0.099
θ_7	PICP	0.970	0.975	0.985
	PINAW	0.035	0.032	0.029

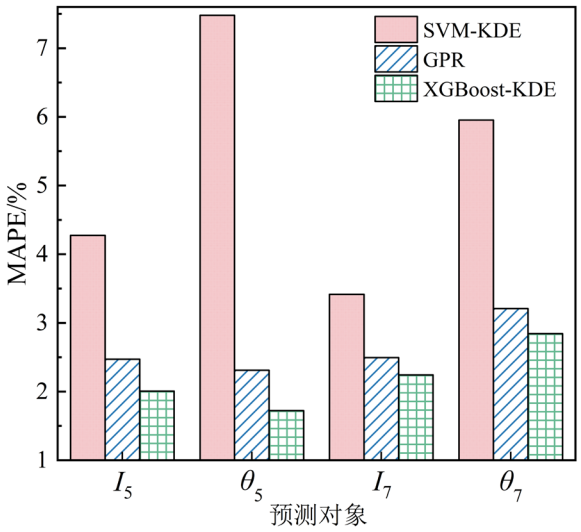


图 2　3 种逆变型电力电子谐波源模型的点预测评价指标对比

Fig.2　Comparison of point prediction evaluation indicators for three inverter-type power electronic harmonic source models

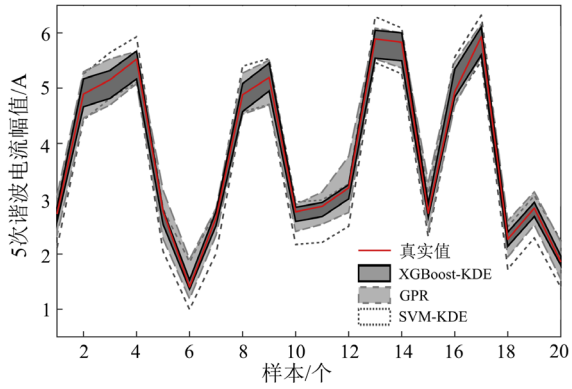


图 3　5 次谐波电流幅值的区间预测结果对比

Fig.3　Comparison of interval prediction results for the 5th harmonic current amplitude

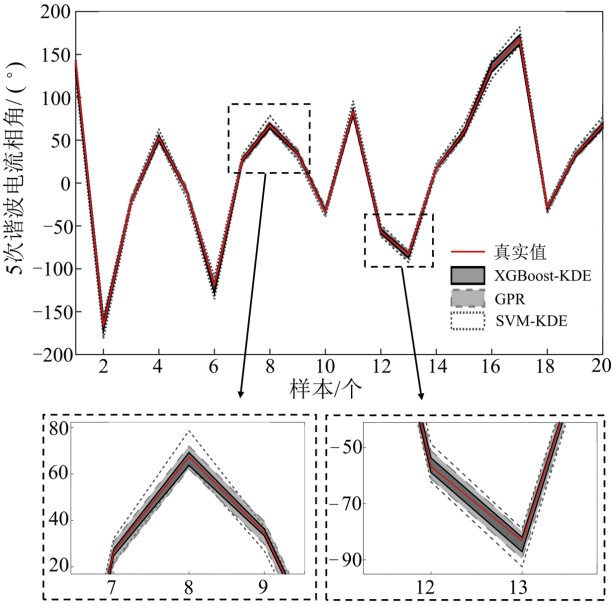


图 4　5 次谐波电流相角的区间预测结果对比

Fig.4　Comparison of interval prediction results for 5th harmonic current phase angle

为了更加直观地对比 3 种逆变型电力电子谐波源模型的不确定性建模效果，从 3 种模型的 5 次谐波电流幅值和相角的测试集中随机选取 20 个样本进行展示，3 种模型的区间预测结果与谐波电流相角真实值之间的比对结果如图 3 和图 4 所示。可以看出，虽然 5 次谐波电流和相角的真实值均能够落在 3 种模型的预测区间内，但是本文所提模型的区间预测宽度最窄，能够更为准确地反应预测结果的潜在不确定性。

4.1.2　整流型电力电子谐波源仿真数据验证

为了评估所提方法应用于整流型电力电子谐波源的有效性，本文基于 MATLAB/Simulink 仿真平台，依据参考文献[32]搭建工业电解铝整流负荷模型，该模型由调压变压器、移相变压器、整流变压器、24 脉波整流器和直流电解槽负荷组成，属于整流型电力电子负荷设备。数据采样点设置在负荷的公共连接点处，采样频率为 10 kHz，采样数据为三相电压和电流，采用快速傅里叶变换(fast Fourier transform, FFT)对 A 相采样数据进行变换分析，选取 2 000 组变换后的谐波样本数据用于建模，按照 7:2:1 的比例将数据集随机划分为训练集、验证集和测试集。为了模拟实际测量设备的测量误差，向采集到的所有仿真数据中添加最大±1%的随机误差。根据第 1.1 节所提方法，工业电解铝整流负荷的主导频次谐波电压为 23、25、47 和 49 次谐波电压，

因此，模型的输入为 23、25、47 和 49 次谐波电压的幅值和相角、基波电压和基波电流的幅值和相角。采用 3 种模型对幅值较大的 23、25 次谐波电流进行区间预测，不同模型的区间预测评价指标对比结果如表 2 所示，点预测评价指标对比结果如图 5 所示。

表 2　3 种整流型电力电子谐波源模型的区间预测评价指标对比

Table 2　Comparison of interval prediction evaluation indicators for three rectifier-type power electronic harmonic source models

预测对象	评价指标	SVM-KDE	GPR	XGBoost-KDE
I_{23}	PICP	0.955	0.990	1.000
	PINAW	0.131	0.093	0.079
θ_{23}	PICP	0.940	0.965	0.990
	PINAW	0.108	0.037	0.032
I_{25}	PICP	0.950	0.995	0.995
	PINAW	0.230	0.154	0.098
θ_{25}	PICP	0.960	0.980	0.990
	PINAW	0.033	0.023	0.010

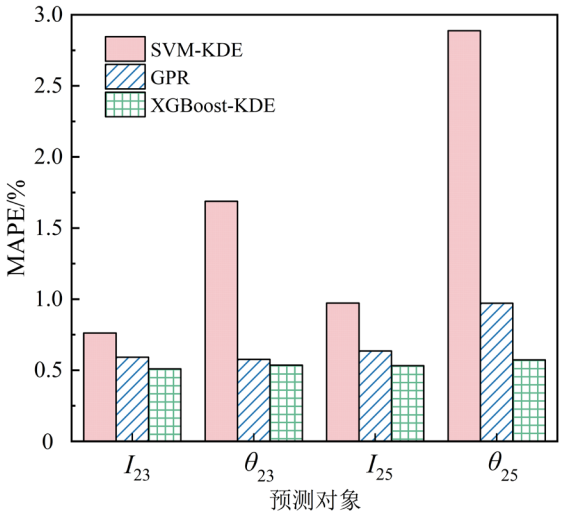


图 5　3 种整流型电力电子谐波源模型的点预测评价指标对比

Fig.5　Comparison of point prediction evaluation indicators for three rectifier-type power electronic harmonic source models

从表 2 可以得出，XGBoost-KDE 模型和 GPR 模型在 PICP 和 PINAW 指标上均优于 SVM-KDE 模型。虽然 XGBoost-KDE 与 GPR 模型的 PICP 指标接近，均能实现较高的预测区间覆盖率，但 XGBoost-KDE 模型在 PINAW 指标上表现更为优异，表明本文所提模型不仅能保证预测区间的高可靠性，还能以更窄的区间宽度反映谐波电流的不确定性。此外，结合图 5 的平均绝对百分比误差分析结果可知，XGBoost-KDE 模型在点预测精度上同样表现较好，其 MAPE 值均小于另外 2 种模型。综合区间预测和点预测评价指标的对比，证明本文所提模型具有最优的建模效果。

4.2　电弧炉谐波源仿真数据验证

为了验证所提方法在电弧炉谐波源建模中的有效性，本文基于 MATLAB/Simulink 仿真平台，建立了 12.5 MVA 的三相交流电弧炉负荷，电弧炉负荷模型基于能量平衡方程和混沌随机理论动态数学模型建立，模型参数为湖北某地区典型交流电石炉厂电压和电流数据的辨识结果[33]。样本的采样和选取方式与 4.1.2 节一致。电弧炉谐波源的谐波电流发射量与基波电流关联性强，故将基波电流作为电弧炉谐波源的特征参数。根据所提的主导频次谐波电压选取方法，电弧炉的主导频次谐波电压为 3~15 次的奇次谐波电压，因此，将这些主导频次谐波电压的幅值和相角、基波电压和基波电流的幅值和相角作为模型的输入。采用 3 种模型对 5、7 次谐波电流的幅值和相角进行区间预测，不同模型的区间预测评价指标对比结果见表 3，点预测评价指标的对比结果如图 6 所示。

由表 3 可知，从 PICP 指标来看，XGBoost-KDE 模型的 PICP 均超过 97%，高于 SVM-KDE 模型和 GPR 模型，这表明本文所提模型具有更高的可靠性。进一步通过分析 PINAW 指标发现，本文所提模型在所有情况下均能保持最窄的预测区间宽度，此外，MAPE 指标的对比结果显示，XGBoost-KDE 模型在不同情况下均具有最优的点预测性能。通过综合分析 PICP、PINAW 和 MAPE3 项评价指标，充分验证了本文所提模型在电弧炉谐波源建模中具有较高的准确性。

表 3　3 种电弧炉谐波源模型的区间预测评价指标对比

Table 3　Comparison of interval prediction evaluation indicators for three harmonic source models of electric arc furnaces

预测对象	评价指标	SVM-KDE	GPR	XGBoost-KDE
I_5	PICP	0.940	0.970	0.990
	PINAW	0.135	0.098	0.038
θ_5	PICP	0.920	0.985	0.990
	PINAW	0.232	0.183	0.101
I_7	PICP	0.950	0.950	0.975
	PINAW	0.189	0.073	0.047
θ_7	PICP	0.955	0.965	0.980
	PINAW	0.210	0.188	0.088

4.3 复杂多谐波源网络仿真数据验证

为了评估所提方法在复杂多谐波源网络建模中的有效性，本文根据附录 B 中图 B1 的电路结构，利用 MATLAB/Simulink 仿真平台建立了交流电弧炉、风电站和线性负荷组成的复杂多谐波源网络模型。其中，风电站配备了 10 MW 的永磁直驱风力机组，包括风力机、永磁同步电机、全功率变流器等组成部分[34]，属于电力电子型谐波源，所发射的谐波特征主要为 5、7、11、13 次。样本的采样和选取方式与 4.1.2 节一致。由于复杂多谐波源网络中包括电力电子型谐波源与电弧炉谐波源，同理将基波电流作为特征参数。根据所提的主导频次谐波电压选取方法，复杂多谐波源网络的主导频次谐波电压为 3 至 15 次奇次谐波电压。因此，模型的输入为主导频次谐波电压的幅值与相角、基波电压和基波电流的幅值与相角。基于 XGBoost-KDE、GPR 和 SVM-KDE 分别构建三种复杂多谐波源网络模型，对 5 次和 7 次谐波电流的幅值和相角进行区间预测。各个模型的区间预测评价指标对比结果见表 4，图 7 给出了不同模型的点预测评价指标对比结果。

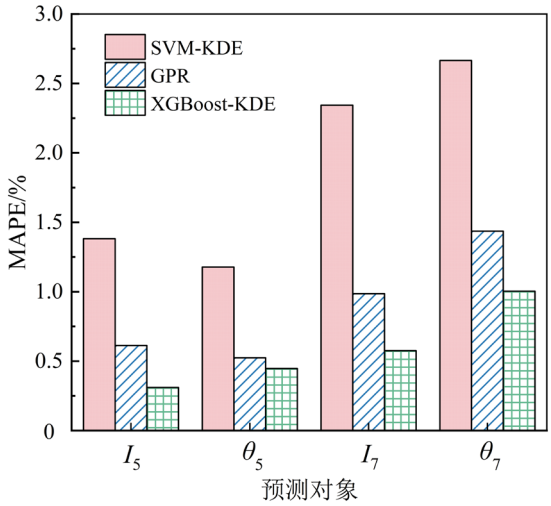


图 6 3 种电弧炉谐波源模型的点预测评价指标对比

Fig.6 Comparison of point prediction evaluation indicators for three harmonic source models of electric arc furnaces

从表 4 和图 7 中可以看出，通过 PICP、PINAW 和 MAPE3 种指标的综合对比分析，本文所提模型的综合性能最优，GPR 模型次之，SVM-KDE 模型表现最差。这一结果表明，本文所提模型的点预测精度最高，且在满足高覆盖率的条件下，预测区间的宽度更窄，能够减小谐波电流区间预测的不确定性，从而为决策者提供更可信的参考依据。综合分析结果证明了本文所提模型在复杂多谐波源网络建模中的优越性。

表 4 3 种复杂多谐波源网络模型的区间预测评价指标对比

Table 4 Comparison of interval prediction evaluation indicators for three complex harmonic source models

预测对象	评价指标	SVM-KDE	GPR	XGBoost-KDE
I_5	PICP	0.940	1.000	1.000
	PINAW	0.267	0.214	0.112
θ_5	PICP	0.920	0.990	0.990
	PINAW	0.213	0.186	0.124
I_7	PICP	0.920	0.980	0.990
	PINAW	0.085	0.055	0.038
θ_7	PICP	0.935	0.975	0.980
	PINAW	0.021	0.018	0.013

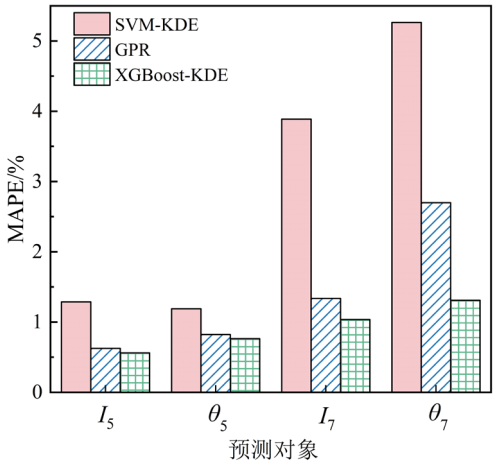


图 7 3 种复杂多谐波源网络模型的点预测评价指标对比

Fig.7 Comparison of point prediction evaluation indicators for three complex harmonic source models

4.4 模型在线更新策略的仿真数据验证

为了评估本文所提模型在线更新方法的有效性，基于 4.3 节的仿真模型进行算例分析，通过改变不同设备的运行状态，模拟复杂多谐波源网络的重复性变化。

在本节算例中，数据长度 D 设定为 200 个，模型的 PICP 阈值设定为 0.9，按照设备的运行状态，将其分为 6 个时段(T_1—T_6)。以经过数据采集并通过 FFT 处理后的样本数作为时间尺度，T_1 时段持续 800 个样本，其余每个时段(T_2—T_6 时段)持续 400 个样本。假定谐波源模型已经基于 T_1 时段的样本集建立完成。按照数据长度 D 进行划分，将 T_2—T_6 时段样本数据分为 D_2—D_{11} 共 10 个区间。不同时段中各个设备的运行状态见表 5，表 5 中各个符号解释如下：●表示电弧炉、风电站和线性负荷分别按照额定功

率正常运行；×表示切除电弧炉设备；◖表示风电站输出功率降至额定功率的一半。将现有的谐波源模型在线更新方法[17]和本文所提方法进行对比，两种方法的评价指标对比如表 6 所示。

表 5 不同时段中各个设备的运行状态

Table 5 Operating status of various devices in different time periods

设备	T_1	T_2	T_3	T_4	T_5	T_6
风电站	●	◖	●	◖	●	●
电弧炉	●	×	×	×	×	●
线性负荷	●	●	●	●	●	●

表 6 两种模型在线更新方法的对比

Table 6 Comparison of two online updating methods for models

时段	区间	现有的方法				本文所提方法			
		PICP	PINAW	MAPE	是否更新	PICP	PINAW	MAPE	是否更新
T_1	D_1	—	—	—	—	—	—	—	—
T_2	D_2	—	—	—	是	—	—	—	是
	D_3	0.995	0.097	1.112%	否	0.995	0.049	0.788%	否
T_3	D_4	0	—	—	是	0	—	—	是
	D_5	0.990	0.233	0.770%	否	0.995	0.105	0.618%	否
T_4	D_6	0	—	—	是	1.000	0.059	0.759%	否
	D_7	0.995	0.111	1.003%	否	0.995	0.058	0.810%	否
T_5	D_8	0	—	—	是	0.980	0.108	0.662%	否
	D_9	0.990	0.191	0.765%	否	0.995	0.104	0.668%	否
T_6	D_{10}	—	—	—	是	0.995	0.119	0.548%	否
	D_{11}	1.000	0.205	0.558%	否	1.000	0.111	0.526%	否

当设备运行状态发生变化但模型尚未更新时(例如在 D_2 和 D_4 区间)，模型往往难以实现有效预测。在 D_2 区间时，由于 D_1 区间谐波源的主导频次谐波电压为 3~15 次奇次谐波电压，而 D_2 区间的主导频次谐波电压仅为 5、7、11 和 13 次谐波电压，此时主导频次谐波电压发生改变，且与样本集中已有的特征不匹配，导致模型无法有效预测，使得模型需要重新选取主导频次谐波电压集合。将 D_2 区间采集 200 个实时谐波检测数据构成实时数据集，由于 D_2 区间实时数据集中的主导频次谐波电压集合是 D_1 区间样本集对应集合的子集，为保证特征维度的一致性，只需对 D_2 区间实时数据集中缺失的特征值(即 3 次、9 次、15 次谐波电压对应的特征)以 0 值填充，随后再通过新增策略完成模型更新。而在 D_4 区间时，谐波源的主导频次谐波电压虽然未发生变化，但两种模型的区间覆盖率均为 0，满足了模型更新条件，故利用新增策略更新模型。经过更新后，两种模型又能跟踪设备的谐波特性，其中本文所提模型表现出更优的预测性能。此外，本文所提在线更新方法的更新次数较少，能够在大多数时间段内反映谐波源谐波特性的不确定性行为，从而提高了模型的可靠性。

两种模型在线更新方式的具体对比如下：由于 GPR 模型在样本量较大时建模效率较低，为了满足实时性要求，现有的模型在线更新方法基于滑动窗口策略进行更新，更新时仅利用前一时段存储的 D 个数据进行重新建模，共需更新 5 次。相比之下，XGBoost-KDE 模型采用基于自适应双策略的模型在线更新方法，使得模型能够利用较多的历史样本数据进行建模，仅需在 T_2 和 T_3 两个时段进行更新，而 T_4—T_6 3 个时段无需更新，更新次数减少为 2 次。例如，在 D_2 区间检测到模型需要更新后，GPR 模型在 D_3 区间起始时刻利用 D_2 区间的 200 个样本建模完成更新；类似地，在 D_5 区间起始时刻使用 D_4 区间的 200 个样本更新模型。而 XGBoost-KDE 模型在区间 D_3 起始时刻时利用 D_1 和 D_2 区间共 1 000 个样本建模完成更新，在区间 D_5 起始时刻时利用 D_1、D_2 和 D_4 区间共 1 200 个样本更新模型。

为了对比两种模型在大样本数据下的建模速度，利用两种模型对 5 万个相同的训练样本(14 维)进行建模训练时，本文所提模型建模用时 2.51 s，GPR 模型建模用时超过 5 min。这一差异主要源于两种模型的计算复杂度不同，GPR 模型的计算复杂度为 $O(n^3)$，而本文所提模型的计算复杂度为 $O(T_t\times d\times m\times\log_2(n))+O(n^2)$，其中，前一部分为 XGBoost 的复杂度，后一部分为 KDE 的复杂度，T_t 为模型的决策树数量，通常为 100 至 500，d 为模型的最大树深度，通常为 1 至 10，m 为输入特征的维度，n 为样本数量[35]。当样本数量较多时，GPR 模型的高计算复杂度使其无法满足实时性要求，故不适用本文所提的模型在线更新方法。而本文所提模型在满足实时性要求的前提下，能够利用较多的历史样本数据进行建模，从而提高模型的性能。

在实际的工程应用中，由于风力发电机组易受阵风扰动、光伏电站输出功率受云层流动干扰以及谐波源设备启停等因素，谐波源的谐波发射特性呈现重复性变化特征，本文所提的模型在线更新方法能够更为准确地动态跟踪谐波发射特性变化，并充分利用大样本数据的优势，减少模型的更新次数，

提高模型在重复性工况下的适用性。

5 结论

1）与其他的谐波源不确定性通用模型进行对比，基于 XGBoost-KDE 的谐波源不确定性通用模型在单一谐波源和复杂多谐波源场景中都展现出较优的模型精度，表明所提模型具有更强的通用性，能够更好地反映谐波源谐波发射特性的不确定性。

2）所提模型在线更新方法能够利用更多的历史样本数据进行建模，减少了谐波源不确定性通用模型的更新次数，进而让模型可以更为准确地动态跟踪谐波源的谐波发射特性，提高了模型在重复性工况下的适用性。

附录见本刊网络版(http://hve.epri.sgcc.com.cn)。

参考文献 References

[1] 张蓬鹤，宋如楠，李兴文，等. 新型电力系统供用电安全技术研究与发展[J]. 新型电力系统，2025，3(4)：384-396.
ZHANG Penghe, SONG Runan, LI Xingwen. Research and development of safety technologies for power supply and consumption in new electric power systems[J]. New Type Power Systems, 2025, 3(4): 384-396.

[2] 张逸，孟祥亮，刘必杰. 谐波源建模现状及新型电力系统下的发展趋势[J]. 电力系统及其自动化学报，2022，34(8)：139-149.
ZHANG Yi, MENG Xiangliang, LIU Bijie. Current situation of harmonic source modeling and development trend under novel power system[J]. Proceedings of the CSU-EPSA, 2022, 34(8): 139-149.

[3] 王子辉，贾燕冰，韩肖清，等. 计及谐波静态电压稳定的配电网光储容量优化[J]. 高电压技术，2025，51(2)：864-875.
WANG Zihui, JIA Yanbing, HAN Xiaoqing, et al. Optimal capacity configuration of photovoltaic and storage in distribution network considering the influence of harmonic static voltage stability[J]. High Voltage Engineering, 2025, 51(2): 864-875.

[4] 丁同，陈红坤，吴斌，等. 多谐波源定位及谐波责任量化区分方法综述[J]. 电力自动化设备，2020，40(1)：19-30.
DING Tong, CHEN Hongkun, WU Bin, et al. Overview on location and harmonic responsibility quantitative determination methods of multiple harmonic sources[J]. Electric Power Automation Equipment, 2020, 40(1): 19-30.

[5] 杨涛，路晓庆，董旭柱，等. 交流微电网谐波功率分布式共享控制策略[J]. 高电压技术，2024，50(7)：3060-3069.
YANG Tao, LU Xiaoqing, DONG Xuzhu, et al. Distributed control strategy for harmonic power sharing in AC microgrids[J]. High Voltage Engineering, 2024, 50(7): 3060-3069.

[6] 国家发展改革委. 电能质量管理办法(暂行)[EB/OL]. 国家发展改革委网站，2023[2025-07-11]. https://www.gov.cn/zhengce/202401/content_6925326. htm.
National Development and Reform Commission. Management measures for power quality (provisional)[EB/OL]. Website of the national development and reform commission, 2023[2025-07-11]. https://www. gov. cn/zhengce/202401/content_6925326. htm.

[7] 赵海波，张万起，沈迪，等. 动车组避雷器谐波特性分析及建模[J]. 高电压技术，2022，48(10)：3891-3899.
ZHAO Haibo, ZHANG Wanqi, SHEN Di, et al. Harmonic characteristics analysis and modeling of EMU arrester[J]. High Voltage Engineering, 2022, 48(10): 3891-3899.

[8] 杨明臻，陈柏超，田翠华，等. 曲形磁阀式可控电抗器谐波最小化研究[J]. 高电压技术，2023，49(12)：5042-5051.
YANG Mingzhen, CHEN Baichao, TIAN Cuihua, et al. Study on harmonics minimization of curvilinear magnetic-valve controllable reactor[J]. High Voltage Engineering, 2023, 49(12): 5042-5051.

[9] 褚昱麟，陈红坤，王嘉琦，等. 基于多元高斯过程回归和空间聚类的复信号谐波源建模[J]. 电工技术学报，2025，40(23)：7621-7634.
CHU Yulin, CHEN Hongkun, WANG Jiaqi, et al. Complex signal modeling method for harmonic source based on multivariate Gaussian process regression and DBSCAN algorithm[J]. Transactions of China Electrotechnical Society, 2025, 40(23): 7621-7634.

[10] 黄存强，王玉，米金梁，等. 基于有源功率因数校正的电动汽车直流充电系统接入对电网谐波的影响研究[J]. 高电压技术，2023，49(增刊 1)：239-246.
HUANG Cunqiang, WANG Yu, MI Jinliang, et al. Research on the impact of DC charging system connection for electric vehicles based on active power factor correction on grid harmonics[J]. High Voltage Engineering, 2023, 49(Supplement1): 239-246.

[11] SHAN P B, SUN Y Y, SONG Y Z, et al. Adaptive Parameter tuning and virtual impedance injection control for coupled harmonic mitigation of photovoltaic converter[J]. IEEE Transactions on Power Electronics, 2025, 40(1): 162-175.

[12] 张逸，欧杰宇，陈书畅，等. 基于时-频特征联合提取的谐波源数据驱动建模方法[J]. 中国电机工程学报，2025，45(22)：8832-8844.
ZHANG Yi, OU Jieyu, CHEN Shuchang, et al. A data-driven harmonic source modeling method based on joint time-frequency feature ex-traction[J]. Proceedings of the CSEE, 2025, 45(22): 8832-8844.

[13] 杨苓，黄俊滔，朱涤凡. 基于复合虚拟阻抗的混合储能系统功率分频协调与谐波电流抑制策略[J]. 高电压技术，2023，49(5)：2017-2028.
YANG Ling, HUANG Juntao, ZHU Difan. Power division coordination and harmonic current suppression strategy for hybrid energy storage system based on composite virtual impedance[J]. High Voltage Engineering, 2023, 49(5): 2017-2028.

[14] CHEN M W, ROBERTS C, WESTON P, et al. Harmonic modelling and prediction of high speed electric train based on non-parametric confidence interval estimation method[J]. International Journal of Electrical Power & Energy Systems, 2017, 87: 176-186.

[15] 陈民武，刘洋，韩旭东，等. 高速铁路牵引负荷谐波分布的非参数估计模型与预测评估[J]. 电网技术，2017，41(8)：2598-2603.
CHEN Minwu, LIU Yang, HAN Xudong, et al. Harmonic distribution modeling based on non-parametric estimation and harmonic prediction evaluation for high speed railway traction load[J]. Power System Technology, 2017, 41(8): 2598-2603.

[16] 解绍锋，李群湛，赵丽平. 电气化铁道牵引负载谐波分布特征与概率模型研究[J]. 中国电机工程学报，2005，25(16)：79-83.
XIE Shaofeng, LI Qunzhan, ZHAO Liping. Study on harmonic distribution characteristic and probability model of the traction load of electrified railway[J]. Proceedings of the CSEE, 2005, 25(16): 79-83.

[17] 张逸，刘必杰，邵振国，等. 基于高斯过程回归的谐波源不确定性通用模型[J]. 中国电机工程学报，2022，42(3)：992-1001.
ZHANG Yi, LIU Bijie, SHAO Zhenguo, et al. General and uncertain harmonic source model based on Gaussian process regression[J]. Proceedings of the CSEE, 2022, 42(3): 992-1001.

[18] WANG C, LIN H, YANG M, et al. A novel chaotic time series wind power point and interval prediction method based on data denoising

strategy and improved coati optimization algorithm[J]. Chaos, Solitons & Fractals, 2024, 187: 115442.

[19] YANG F, HUANG G Q, JIAO X. A novel hybrid model based on dual-layer decomposition and kernel density estimation for VOCs concentration forecasting considering influencing factors[J]. Atmospheric Pollution Research, 2025, 16(4): 102439.

[20] LAN H, GAO J J, HONG Y Y, et al. Interval forecasting of photovoltaic power generation on green ship under multi-factors cou-pling[J]. Sustainable Energy Technologies and Assessments, 2023, 56: 103088.

[21] WANG Y, TAN R, ZHENG Z X, et al. Method to estimate sag frequency in doubly fed induction generator integrated power system based on adaptive kernel density estimation[J]. IET Generation, Transmission & Distribution, 2020, 14(7): 1261-1270.

[22] 石慧，宋仁旺，张岩，等. 基于核密度估计和随机滤波理论的齿轮箱剩余寿命预测方法[J]. 计算机集成制造系统，2020，26(3)：632-640.
SHI Hui, SONG Renwang, ZHANG Yan, et al. Prediction method for remaining useful life of gearbox based on kernel estimation and stochastic filtering theory[J]. Computer Integrated Manufacturing Systems, 2020, 26(3): 632-640.

[23] 宗文婷，卫志农，孙国强，等. 基于改进高斯过程回归模型的短期负荷区间预测[J]. 电力系统及其自动化学报，2017，29(8)：22-28.
ZONG Wenting, WEI Zhinong, SUN Guoqiang, et al. Short-term load interval prediction based on improved Gaussian process regression model[J]. Proceedings of the CSU-EPSA, 2017, 29(8): 22-28.

[24] 韩宇超，同向前，邓亚平. 基于概率密度估计与时序 Transformer 网络的风功率日前区间预测[J]. 中国电机工程学报，2024，44(23)：9285-9295.
HAN Yuchao, TONG Xiangqian, DENG Yaping. Probabilistic distribution estimation and temporal Transformer-based interval prediction in day-ahead wind power prediction[J]. Proceedings of the CSEE, 2024, 44(23): 9285-9295.

[25] 徐询，谢丽蓉，叶林，等. 基于非参数核密度估计的风电场有功功率双层优化模型[J]. 电力系统自动化，2022，46(2)：43-55.
XU Xun, XIE Lirong, YE Lin, et al. Bi-level optimization model of active power for wind farm based on nonparametric kernel density estimation[J]. Automation of Electric Power Systems, 2022, 46(2): 43-55.

[26] 杨瑞君，张楚，杨评，等. 非参数核密度估计模型预测双酚 A 的物种敏感度分布规律[J]. 生态毒理学报，2024，19(4)：120-130.
YANG Ruijun, ZHANG Chu, YANG Ping, et al. Application of non-parametric kernel density estimation model in prediction of species sensitivity distribution of bisphenol A[J]. Asian Journal of Ecotoxicology, 2024, 19(4): 120-130.

[27] 王运达. 基于数字孪生的城轨柔性牵引供电系统多时间尺度建模及控制方法研究[D]. 北京：北京交通大学，2022.
WANG Yunda. Research on multi-time scale modeling and control method of urban rail flexible traction power supply system based on digital twin[D]. Beijing, China: Beijing Jiaotong University, 2022.

[28] 张紫玥，刘晓，杜丽丽，等. 基于随机蛙跳降维和 PSO-BPNN 的高光谱水体总氮遥感反演[J]. 激光与光电子学进展，2025，62(17)：476-485.
ZHANG Ziyue, LIU Xiao, DU Lili, et al. Remote sensing inversion of total nitrogen in hyperspectral water bodies based on random frog dimensionality reduction and PSO-BPNN[J]. Laser & Optoelectronics Progress, 2025, 62(17): 476-485.

[29] 郑欣雨，任泽华，周利，等. 基于晶体图卷积神经网络的晶格能回归模型[J]. 化工学报，2025，76(3)：1084-1092.
ZHENG Xinyu, REN Zehua, ZHOU Li, et al. Lattice energy regression model based on crystal graph convolutional neural networks[J]. CIESC Journal, 2025, 76(3): 1084-1092.

[30] 倪俊贤，史明明，姜锋，等. 基于 EV 参数随机性的改进 LHS 概率谐波潮流算法[J]. 电子设计工程，2025，33(2)：139-144.
NI Junxian, SHI Mingming, JIANG Feng, et al. Improved LHS probabilistic harmonic current algorithm based on EV parameter stochasticity[J]. Electronic Design Engineering, 2025, 33(2): 139-144.

[31] 邵振国，许昊铂，肖颂勇，等. 新能源电网中的谐波问题[J]. 电力系统保护与控制，2021，49(4)：178-187.
SHAO Zhenguo, XU Haobo, XIAO Songyong, et al. Harmonic problems in a new energy power grid[J]. Power System Protection and Control, 2021, 49(4): 178-187.

[32] 朱润林. 电力系统冲击负荷建模及其对电压稳定性的影响分析[D]. 成都：西南交通大学，2021.
ZHU Runlin. Modeling of power system shock load and analysis of its effect on voltage stability[D]. Chengdu, China: Southwest Jiaotong University, 2021.

[33] 胡畔，陈红坤，孙志达，等. 一种交流电弧炉通用性模型[J]. 电工技术学报，2016，31(8)：172-180.
HU Pan, CHEN Hongkun, SUN Zhida, et al. A versatility model of ac electric arc furnace[J]. Transactions of China Electrotechnical Society, 2016, 31(8): 172-180.

[34] 董桐宇. 直驱式风力发电机的建模与并网仿真分析[D]. 太原：太原理工大学，2011.
DONG Tongyu. Modeling and simulation analysis on grid-connected directly driven wind generator[D]. Taiyuan, China: Taiyuan University of Technology, 2011.

[35] CHEN T Q, GUESTRIN C. XGBoost: A scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA: ACM, 2016: 785-794.



王嘉琦
1997—，男，博士生
主要从事电能质量分析、数据及人工智能驱动等研究
E-mail: wangjiaqi@whu.edu.cn

WANG Jiaqi
Ph.D. candidate

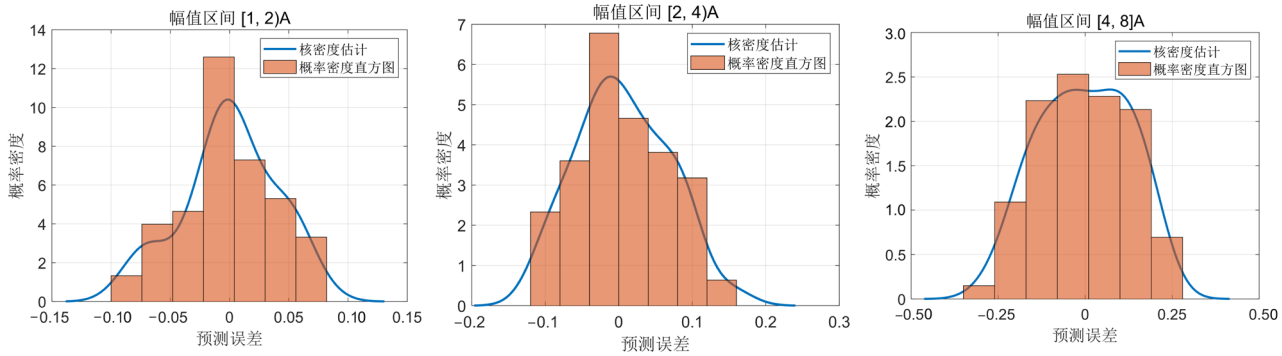


陈红坤(通信作者)
1967—，男，博士，教授，博导
主要从事电能质量分析、电力系统规划等研究工作
E-mail: chkinsz@163.com

CHEN Hongkun
Ph.D., Professor
Corresponding author

收稿日期 2025-07-15 在线出版日期 2025-10-13 编辑 陈 蔓

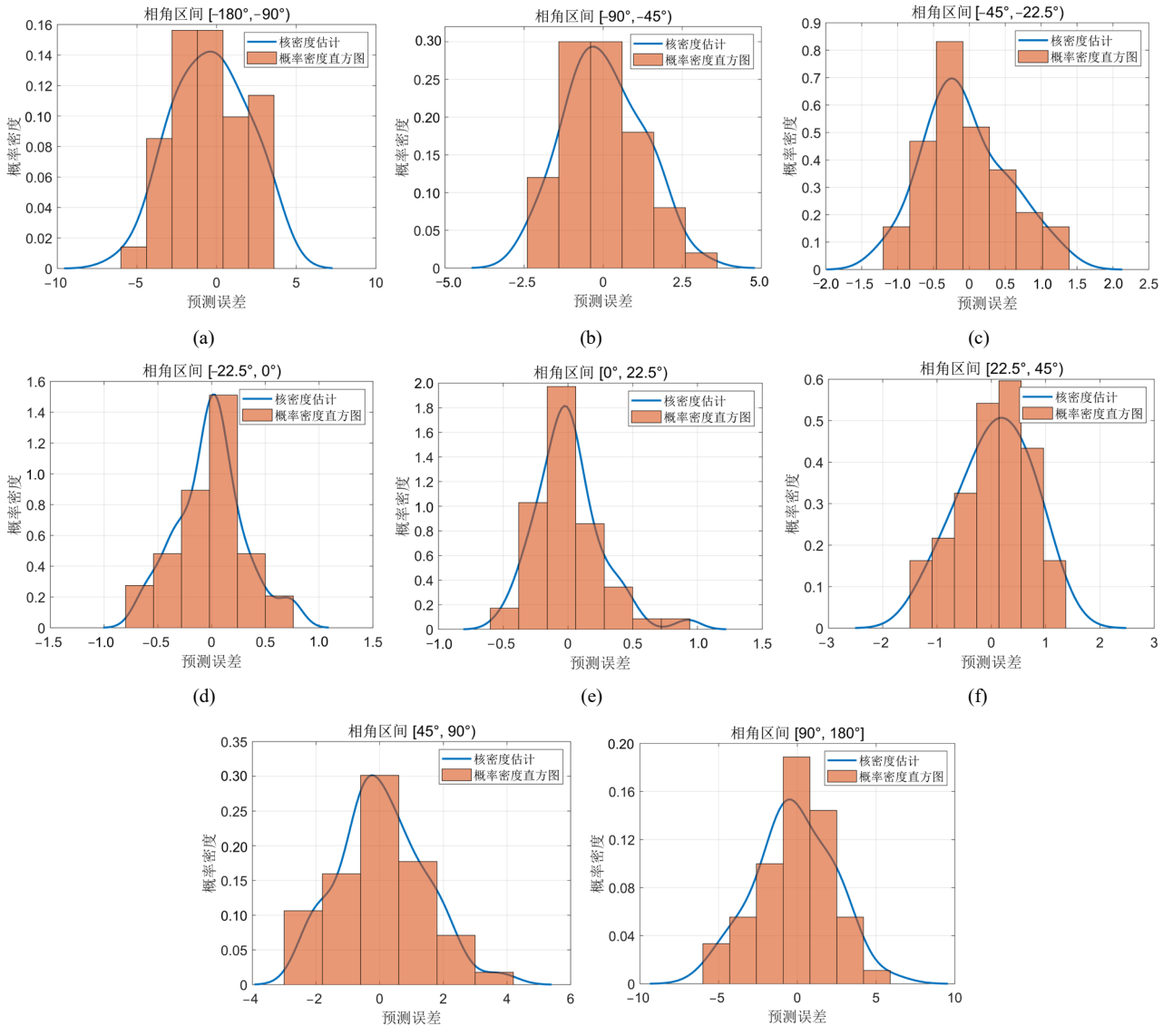
附录 A



(a) (b) (c)

图 A1　不同谐波电流幅值区间的预测误差概率密度拟合图

Fig.A1　Prediction error probability density fitting diagram for different harmonic current amplitude intervals



(a) (b) (c)

(d) (e) (f)

(g) (h)

图 A2　不同谐波电流相角区间的预测误差概率密度拟合图

Fig.A2　Prediction error probability density fitting diagram for different harmonic current phase angle intervals

表 A1　不同谐波电流幅值区间的 KDE 拟合优度检验结果

Table A1　KDE goodness of fit test results for different harmonic current amplitude intervals

幅值区间/A	K-S 检验 统计量	K-S 检验 临界值	拟合优度 检验结果
[1, 2)	0.061 5	0.178 5	通过
[2, 4)	0.053 9	0.125 1	通过
[4, 8]	0.042 5	0.090 8	通过

表 A2　不同谐波电流相角区间的 KDE 拟合优度检验结果

Table A2　KDE goodness of fit test results for different harmonic current phase angle intervals

相角区间/(°)	K-S 检验 统计量	K-S 检验 临界值	拟合优度 检验结果
[−180, −90)	0.070 8	0.205 0	通过
[−90, −45)	0.072 9	0.192 3	通过
[−45, −22.5)	0.090 7	0.188 5	通过
[−22.5, 0)	0.078 3	0.181 7	通过
[0, 22.5)	0.065 3	0.186 8	通过
[22.5, 45)	0.060 9	0.202 7	通过
[45, 90)	0.066 6	0.198 3	通过
[90, 180]	0.084 3	0.186 8	通过

附录 B

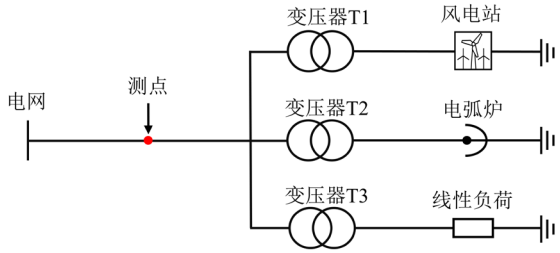


图 B1　复杂多谐波源网络供电拓扑示意图

Fig.B1　Schematic diagram of complex multi harmonic source network power supply topology