

适应面向电力现货市场中短期价格预警的统调负荷概率密度预测方法研究

Research on Probability Density Prediction Method of Unified Load Adapted to Short-term Price Warning in Power Spot Market

尹硕¹ 路尧¹ 张元² 赵剑² 姜昊言²

Shuo Yin¹ Yao Lu¹ Yuan Zhang² Jian Zhao² Haoyan Jiang²

1. 国网河南省电力公司经济技术研究院 中国·河南 郑州 450052

2. 北京清能互联科技有限公司 中国·北京 100084

1. State Grid Henan Electric Power Company Economic and Technical Research Institute, Zhengzhou, Henan, 450052, China

2. Beijing Qingneng Interconnection Technology Co., Ltd., Beijing, 100084, China

摘要: 针对电力现货价格预测、预警中需要提供不同场景下统调负荷预测值区间的需求, 结合统调负荷影响因素多、基础负荷明显、存在复杂的周期性等特点, 建立基于模态分解与分位数回归森林的统调负荷概率密度组合预测模型, 一方面提高了预测准确度, 另外一方面也贴合电力现货价格预警业务需要提供区间预测的诉求。首先, 针对原始时序数据中存在的异常值进行修正, 确保高品质的数据基础。其次, 采用 EMD 方法对统调负荷原始时间序列数据进行分解计算。再次, 对分解出的分量分别建立 QRF 负荷预测模型, 叠加不同分量预测结果从而获得预测值的条件分布。最后, 利用核密度估计输出任意时刻统调负荷概率密度预测结果。相对于确定性点预测方法, 概率密度预测具有描述统调负荷未来可能的波动范围及不确定性优势, 且通过算例测试验证了模型的有效性。

Abstract: In response to the need to provide forecast value ranges of unified load under different scenarios in electricity spot price prediction and early warning, combined with the characteristics of multiple influencing factors of unified load, obvious basic load, and complex periodicity, a model based on modal decomposition and quantification was established. The unified load probability density combination forecast model of the numerical regression forest improves the forecast accuracy on the one hand, and also provides interval forecast requirements in line with the needs of the power spot price early warning business. First, correct the outliers existing in the original time series data to ensure a high-quality data foundation. Secondly, the EMD method is used to decompose and calculate the original time series data of the unified load. Thirdly, QRF load prediction models are established for the decomposed components respectively, and the prediction results of different components are superimposed to obtain the conditional distribution of predicted values. Finally, the kernel density estimation is used to output the unified load probability density prediction result at any time. Compared with the deterministic point prediction method, probability density prediction has the advantage of describing the possible future fluctuation range and uncertainty of the regulated load, and the effectiveness of the model has been verified through numerical example testing.

关键词: 电力现货市场; 模态分解; 预测模型

Keywords: power spot market; modal decomposition; forecasting mode

DOI: 10.12346/peti.v5i4.8830

1 引言

在电力现货市场中, 精准预测未来电价信息对于把握市场运行情况、保证市场各参与方有效进行相关决策、推动市场资源合理优化配置及引导电力市场平稳运行具有重要意义。

义。电价受负荷需求、发电能力、市场经济水平等因素的影响, 呈现出较强的波动性和周期性等特点, 尤其自中国现货市场试点展开试运行工作以来, 受多种因素的影响, 源荷双侧同时出现了不确定性明显的现象, 并广泛影响电力现货市

【作者简介】尹硕 (1985-), 男, 中国河南三门峡人, 博士, 高级经济师, 从事能源经济、电力市场研究等研究。

场价格,使电力现货市场价格也一并出现了较为明显的波动,出现极高电价与负电价反复交替出现的现象,给电力现货市场的健康稳定运行带来挑战。为了规避或降低电价波动的影响,平稳推进电力市场改革,释放改革红利。需要开展现货电力市场电价预测尤其是极端电价预警工作。而负荷需求是影响电价波动的一个重要因素,电价变化一定程度上是电负荷水平波动的反馈,因此实现现货市场价格预警对未来负荷预测提出较高要求,具体体现在以下两个方面:

统调负荷面临涉及用电类型多、影响因素复杂的特点。为有效适配电力现货市场价格预警的需求,统调负荷预测根据预测周期不同可以分为中期预测和短期预测。其中中期预测以月度及周为单位,对次月或下周每一天的负荷曲线进行预测,但较长的预测周期为预测带来的不确定因素较多,预测的准确性随之降低,因此需提升中短期负荷预测的准确率,从有效支持价格预警工作的开展^[1]。

传统负荷预测方法以时间序列法为代表,建立一个时间序列的数学模型用来描述电力负荷随时间序列变化的时序性对于负荷预测具备一定的有效性。近年来以人工神经网络法为代表的人工智能方法和机器学习算法也得到了越来越广泛的研究。但上述方法都只能得到确定的点预测结果,无法反映实际电力市场受到随机性扰动对负荷预测造成的影响^[2-5]。

为提高电力系统短期负荷的预测精度,有效反映实际电力市场受到随机扰动时可能出现的波动,区间预测方法研究逐步受到关注,产生的研究成果有 Bootstrap 法、基于 ANN 的上下限估计法、概率密度预测法等。其中,概率密度预测法的建立基于贝叶斯理论,是一种构造预测量的分布模型,可以提供不同置信度区间下的预测结果,既能反映未来的变化趋势,可以计算出现极端场景的概率,因此更适用于电力现货市场中短期价格预警场景的负荷预测。丁美荣等^[6]提出了一种分位数回归模型 MCQRDDC,能够较好地分析负荷数据中的时序信息和内在规律问题。杨宏等^[7]为解决传统的基于正态分布的风电预测误差区间估算问题,提出一种基于 Beta 分布的风电预测误差区间估算方法,将复杂化的等式约束问题转化为无约束的最优问题,并结合 Beta 分布特征,设计出一种快速求解方法。庞传军等^[8]利用变分贝叶斯方估计法,在以往负荷数据信息的基础上,对模型中的参数进行后验概率分布,并利用所建立的模型对电网负荷进行概率区间预测,且所得到的预测区间更窄、覆盖范围更广。

2 统调负荷概率密度预测方法设计

2.1 实验数据预处理

采集某地区某日 24 小时的历史统调负荷数据及对应的气象信息,若日数据缺失严重,则将该日剔除不计。若数据少量缺失或异常则可以通过计算历史日统调负荷点变化率对缺失数据进行修正填补(见表 1)。

表 1 实验数据预处理

公式	参数含义
$\Delta x_{i,t} = \begin{cases} (x_{i,t} - x_{i,T})/x_{i-1,T} & t = 1 \\ (x_{i,t} - x_{i,t-1})/x_{i,t-1} & t = 2, \dots, T \end{cases}$ $\overline{\Delta x}_{i,t} = \frac{\sum_{i=1}^M \Delta x_{i,t}}{M}$ $x_{i,t} = \begin{cases} x_{i-1,T} + \overline{\Delta x}_{i,t} x_{i-1,T} & t = 1 \\ x_{i,t-1} + \overline{\Delta x}_{i,t} x_{i,t-1} & t = 2, \dots, T \end{cases}$	<p>其中, $\Delta x_{i,t}$ 表示第 i 日时刻的日统调负荷变化率, $x_{i,t}$、$x_{i-1,T}$ 分别表示第 i 日及第 $i-1$ 日采样点数的负荷值, $x_{i,t}$、$x_{i,t-1}$ 分别表示第 i 日 t 时刻和 $t-1$ 时刻的统调负荷值, $\overline{\Delta x}_{i,t}$ 为日统调负荷变化率, M 为采集历史数据日数。统调负荷的修正值为 $x_{i,t}$</p>
$A_t = \frac{1}{M} \sum_{i=1}^M x_{i,t}$ $\sigma_t = \sqrt{\frac{1}{M} \sum_{i=1}^M (x_{i,t} - A_t)^2}$	<p>根据文献^[9],通过数据纵向出力的方法对异常数据进行修正, A_t、σ_t 分别为原始统调数据 t 时刻的均值及标准差</p>
$d_{i,t} = \frac{ x_{i,t} - A_t }{\sigma_t}$	<p>$d_{i,t}$ 为负荷点的偏离率</p>

2.2 历史负荷数据信号分解

为了分析统调负荷时间序列的随机性与波动性,采用 EMD 将其分解为若干模态分量 (IMF), 对干扰因素进行解耦处理,降低预测模型误差。分解步骤为:

找出负荷时序中全部的极值,采用插值法分别拟合得到包络线 $y(t)$ 、 $z(t)$, 其平均值 $m(t)$ 为:

$$m(t) = \frac{1}{2}(y(t) + z(t))$$

判断原始负荷时间序列与该平均值的差 $n(t)$ 是否满足 IMF 分量条件。如果满足则得到第一个 IMF 分量 n_0 ; 相反,则将 m_0 记作新时间序列并重复上述操作直至满足 IMF 条件。

从原始时间序列 X'_t 分离出 n_0 后,将剩余的分量作为新的时间序列,并重复上述操作,直至得到全部 r 个 IMF 分量 $n_{i(t)}$ 及一个剩余分量 $p_{n(t)}$, 即:

$$X'_t = \sum_{i=1}^n n_{i(t)} + p_{n(t)}$$

2.3 基于 EMD-QRF 的概率密度预测模型

2.3.1 分位数回归森林基础原理

分位数回归森林法 (QRF) 衍生于分位数回归法和 Breiman 随机森林法,该方法不仅能输出不同分位点的回归预测结果,还可以提供因变量的全部条件分布信息。该算法的步骤为:

①生成 k 棵决策树 $T(\vartheta_t)$, $t = 1, 2, \dots, n$, 构成随机森林,考察每棵树每个叶节点的所有观测值, 即:

$$\hat{F}(y|X=x) = \sum_{i=1}^n w_i(x) I(y_i < y)$$

②给定 $X=x$, 遍历全部决策树, 计算每棵决策树的观测值权重 $\omega_i(x, \vartheta_t)$, $i \in \{1, 2, \dots, n\}$, 并取平均得到 $w_i(x)$ 。

③对所有 $y \in R$, 根据步骤②得到的权重, 计算分布函数的估计。

2.3.2 核密度估计方法

为了实现对未来负荷概率分布的预测及分析, 引入核密度估计方法^[10]对其进行估计, 设 X_1, X_2, \dots, X_M 是取自一元连续总体的样本, 在任意点处的总体密度函数 $f(x)$ 的核密度估计定义为:

$$\hat{f}_h(x) = \frac{1}{Mh} \sum_{i=1}^M K\left(\frac{x - X_i}{h}\right)$$

其中, $K(x)$ 为核函数, 常用的高斯函数形式为 $K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$; h 为带框系数, 取值为 1.8~2.0。

2.4 预测模型流程

首先, 对参与模型训练的历史数据进行预处理, 对异常数据进行纠偏、补全。其次, 利用 EMD 将校正后的统调负荷原始数据展开分解, 获得多个具有互异性的 IMF 分量。随之构建 QRF 预测模型, 获得各分位数的回归预测结果, 并将不一致分量得到预测结果累加, 得到预测值的条件分布情况。最后, 通过核密度估计法得到概率密度预测结果。

算例分析: 以华中地区某省 2020 年 1 月 1 日 00:00 时至 2020 年 12 月 31 日 23:45 时采样间隔为 15 分钟共 34464 个统调负荷数据作为研究对象, 分别建立 EMD-ARIMA、EMD-BP 及 EMD-QRF 的统调负荷预测模型, 并对比不同模型的表现, 以验证模型的有效性。

2.5 统调负荷原始数据的预处理

通过对统调负荷原始数据进行扫描, 计算, 定位原始数据中存在的异常数据问题, 如: 在 2020 年 3 月 20 日 18:00 至 18:45 间出现了异常的 0 值, 经过研判, 确定为因通讯异常导致数据丢失, 该时间段内电网无其他故障。

由于出现连续数据缺失, 将采取纵向修补的方式, 利用 3 月 19 日及 3 月 21 日对应时刻的数据进行计算, 计算结果见表 2。

表 2 连续型异常数据修补示例表

类型	日期	18:00	18:15	18:30	18:45
原始数据	2020/3/20	0.00	0.00	0.00	0.00
对应时刻均值		20238.91	19999.01	19750.47	19538.72
对应时刻标准差		576.43	542.13	540.95	547.04
对应时刻偏离率		35.11	36.89	36.51	35.72
t-1 数据	2020/3/19	21015.60	20719.40	20370.50	20107.80
t+1 数据	2020/3/21	20104.90	19916.80	19673.50	19489.40
纵向修补值		20560.25	20318.10	20022.00	19798.60
修补值偏离率		0.56	0.59	0.50	0.48
最终采纳修补值	2020/3/20	20238.91	19999.01	19750.47	19798.60

而对于部分离散的单一性缺失或异常值, 则采取论文介绍的横向修补方式进行修补, 较为简单, 在此不再赘述。

2.6 统调负荷原始数据的分解

图 1 为统调负荷时间序列经 EMD 分解得到的部分结果。可以看出该时序数据由 EMD 分解后得到 6 个 IMF 分量和一个余量序列 r , 且 IMF 函数具有周期性, 从而反映出用户负荷在不同周期的用电特征与规律。

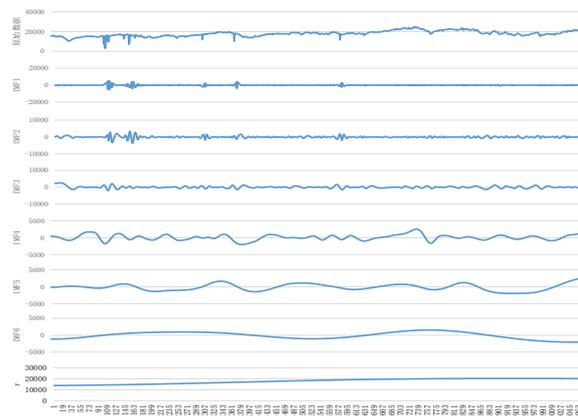


图 1 EMD 算法分解结果

2.7 评价指标

为了准确的评估概率密度预测方法的有效性, 采用均方根误差 (RMSE)、平均绝对百分比误差 (MAPE)、预测区间覆盖率 (FICP)、预测区间平均宽度 (FIAW) 四种预测评价指标来对概率密度预测模型的结果进行评价。其中 FICP 描述的是实际值在预测范围内的可能性, 该数值较高, 表明预测结果更为精准; FIAW 值越宽则更不利于做出相应的决策。

2.8 预测结果比较

为检验模型的可靠性, 针对 2021 年 1 月 1 日时间节点, 对该省统调负荷在该节点下分别展开提前 30 天和提前 24 小时的预测。QRF 模型的分位点区间在 0.01~0.99 之间, 步长取 0.01。

①提前 30 日预测结果。根据图 2 与表 3 中的信息, 可以看出统调负荷的基础负荷水平较高, 变化相对平缓但依然存在较为明显的昼夜特性。不同的预测模型可以较好的跟踪统调负荷的变化趋势, 但 ARIMA 方法在负荷变化的拐点处常出现滞后现象。QRF 模型由于具有较小的 MAPE 和 RMSE 指标且能够给出一个区间内的预测值比其他算法具有较强的优越性。但其所需的预测区间较宽, 表现出 FICP 与 FIAW 评价指标之间的矛盾性。

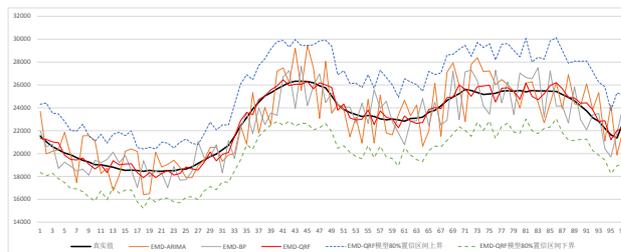


图 2 提前 30 日统调负荷预测结果

表 3 提前 30 日统调负荷预测误差

预测模型	评价指标			
	MAPE(%)	RMSE(MW)	FICP(%)	FIAW(%)
EMD-ARIMA	13.01	1669.92	—	—
EMD-BP	10.55	1543.2	—	—
EMD-QRF	10.08	1416.72	83.88	0.4452

②提前 24h 预测结果。由图 3 及表 4 可以看出，在追加较近时间的输入变量后，各方法的预测精度得到提升，表现为 MAPE 及 RMSE 指标的进一步减少，由于在 EMD-QRF 模型获得条件分布中使用了核密度估计，所以可获得任意时刻统调负荷概况密度预测结果，该模型以更高的概率接近真实值，还能得到统调负荷的波动幅度区间和预测值出现的概率，可为后续的电价预警提供更加贴合需求的边界数据基础，具有实际应用价值。

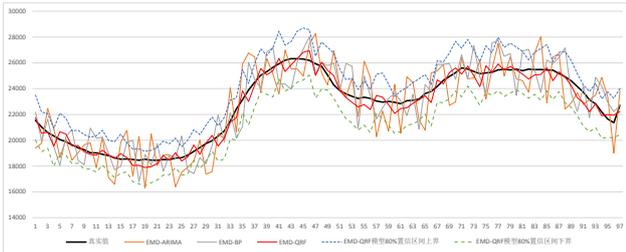


图 3 提前 24h 统调负荷预测结果

表 4 提前 24h 统调负荷预测误差

预测模型	评价指标			
	MAPE(%)	RMSE(MW)	FICP(%)	FIAW(%)
EMD-ARIMA	10.58	1530.48	—	—
EMD-BP	9.68	1313.28	—	—
EMD-QRF	9.32	1177.2	94.38	0.6594

3 结语

论文针对电力现货价格预测、预警中需要提供不同场景下统调负荷预测值区间的需求，结合统调负荷影响因素多、基础负荷明显、存在复杂的周期性等特点，建立基于模态分解与分位数回归森林的统调负荷概率密度组合预测模型，一方面提高了预测准确度，另外一方面也贴合电力现货价格预警业务需要提供区间预测的诉求。

参考文献

- [1] 李瑾,刘金朋,王建军.采用支持向量机和模拟退火算法的中长期负荷预测方法[J].中国电机工程学报,2011,31(16):63-66.
- [2] 马文晓,白晓民,沐连顺.基于神经网络和模糊推理的短期负荷预测方法[J].电网技术,2003(5):29-32.
- [3] 张颖,高中文.基于时间序列和神经网络的电力系统负荷预测[J].哈尔滨理工大学学报,2003(1):30-32+36.
- [4] 彭显刚,王洪森.基于竞争ISPO双胞支持向量回归短期负荷预测[J].电力系统及其自动化学报,2014,26(10):46-50+68.
- [5] 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.
- [6] 丁美荣,张航,蔡高琰,等.基于MCQRDDC的负荷概率预测模型[J].计算机系统应用,2023,32(2):281-287.
- [7] 杨宏,苑津莎,张铁峰.一种基于Beta分布的风电功率预测误差最小概率区间的模型和算法[J].中国电机工程学报,2015,35(9):2135-2142.
- [8] 庞传军,张波,余建明,等.基于结构化负荷模型的电力负荷概率区间预测[J].中国电力,2021,54(9):89-95.
- [9] 李杰,靳孟宇,马士豪.基于粒子群算法的极限学习机短期电力负荷预测[J].制造业自动化,2019,41(1):154-157.
- [10] 周松林,茆美琴,苏建徽.风电功率短期预测及非参数区间估计[J].中国电机工程学报,2011,31(25):10-16.