工程与管理科学
Engineering and Management Science
第 8 卷◆第 10 期◆版本 1.0◆2026 年
文章类型：论文|刊号（ISSN）：2705-0637(P) / 2705-0645(O)

新能源场站基础数据治理对负荷预测精度的影响分析

陆振军　余广朋　檀庭方
南京国电南自软件工程有限公司　江苏南京　211100

DOI：10.32629/ems.v8i10.22602

［**摘　要**］本文围绕新能源场站基础数据治理展开研究，探讨其对负荷预测精度所产生影响。经过对数据完整性、准确性、一致性进行治理，进而揭示出数据缺失、异常情况及不一致状态对于预测精度造成的制约与干扰。建立标准化数据治理体系、实时监控数据质量、强化特征提取和模型迭代，乃是提高预测精度的关键途径所在。研究显示，优化数据治理能够提高负荷预测的准确性与可靠性。

［**关键词**］新能源场站；基础数据；治理；负荷预测精度；影响

伴随着新能源迅猛发展，负荷预测于电力系统调度中的重要程度逐渐变得越发突出。然而，新能源场站的数据质量问题严重影响了预测精度。本文聚焦于探究基础数据治理怎样凭借保障数据具备完整性、准确性与一致性，来提高负荷预测精准度，进而为新能源场站高效运行提供支持。

1 新能源场站基础数据治理的主要内容

1.1 数据完整性治理

新能源场站基础数据的完整性治理，是指在数据采集、传输、存储和应用全流程中，通过建立全链路校验方式，保证主要数据要素不遗漏、重要时间节点不缺失。场站运行需要同步采集气象参数，包括风速、辐照度和温度，还要采集设备工况，包括逆变器功率和变流器温度，同时也要采集电网交互数据，包括并网电压和输送功率[1]。如果采集过程中出现采集盲区或者时间断层，会直接影响负荷预测模型输入的完整性。治理工作需要制定数据覆盖清单，并且要说明必采字段和采集频率。对于传感器离线、网络中断和存储故障造成的缺失数据，应当使用插值法或者参考历史同类场站来进行补齐，并且要标记异常时段，以此来避免无效数据对信息的干扰。从数据源头保证负荷预测所需的信息链条完整且没有断点[2]。

1.2 数据准确性治理

数据准确性治理主要针对纠正数据偏差和失真问题，提高负荷预测精度的重要环节。新能源场站的传感器容易受到环境干扰而出现漂移，数据编码也可能因为格式不规范而产生偏差，造成气象数据和实际不符合，设备参数和实际状态不一致。治理需要从设备校准开始，定期对气象传感器和计量装置进行精度校验。在数据处理环节设置阈值过滤，可以剔除超出合理范围的异常值，并且利用交叉比对，比如风速数据和风机出力的关联验证，来识别矛盾数据。准确的数据可以保证负荷预测模型能够捕捉到场站出力同外部条件的精准关联[3]。光伏场站辐照度数据偏差达到 10%，会使负荷预测误差明显增加，所以需要通过多方面校验来使数据真实可靠。

1.3 数据一致性治理

新能源场站基础数据治理中，数据一致性治理是主要环节。该过程主要为了消除多源异构数据之间的冲突和矛盾，使同一测点在时间、空间和语义维度上保持一致。新能源场站数据来源于多个渠道，包括 SCADA 系统、气象观测设备、功率预测系统和调度主站等，各个系统之间经常出现采样频率不同、时标偏移、量纲差异和命名混乱等情况，这会直接对负荷预测模型输入特征的可靠性产生影响。一致性治理通过建立统一的数据字典和编码规范，进行时标对齐、单位归一化、异常值协同修正及跨系统关联校验，能够处理数据冗余和矛盾，给负荷预测提供高质量且可信任的输入基础，进而提高预测精度和模型稳定性。数据一致性治理内容与影响对照如表 1 所示。

表 1　数据一致性治理内容与影响对照

治理维度	主要问题	对负荷预测的影响
时间一致性	时标偏移、采样频率不一	减少时序错位误差
空间一致性	测点命名混乱、位置歧义	提升特征关联准确性
语义一致性	量纲差异、定义冲突	保障模型输入可信度

2 新能源场站基础数据治理对负荷预测精度的影响

2.1 数据缺失对精度的制约

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.

新能源场站基础数据的完整性，是形成高精度负荷预测模型必须依靠的基础，数据缺失会对其造成限制，影响比较明显。风电场在风速和风机状态这些重要特征出现缺失的情况下，模型就很难完整捕捉新能源出力的动态变化规律。光伏电站中，辐照度和组件温度等主要参数如果缺失，会大大减弱模型描述出力波动特点的能力。当重要数据处于缺失状态时，模型不能精确地还原历史相似场景的特点，也难以有效地识别影响出力的隐蔽变量，只能依靠插补或拟合来推测缺失区间的信息[4]。但是，该推测通常偏离实测值，使预测曲线出现断层或失真。特别是在出力发生突变时段，缺失数据会令模型失去修正能力，无法及时响应功率的剧烈变化，这会明显降低预测精度，扩大预测误差范围，最终影响新能源场站的调度决策和运行安全。

2.2 数据异常对精度的干扰

新能源场站基础数据的一致性，是保障预测结果可靠性的主要条件。数据不一致会导致预测精度降低。此类问题通常表现为多源数据间的时间标签错位、计量单位不统一和数据格式相互冲突。例如，气象数据和出力数据在同一时刻的采样频率上存在差异，或者同一设备在不同系统中记录的数据精度不一致。上述不一致会使数据在时空维度上的关联逻辑被割裂，造成模型难以建立准确的关联映射，并干扰特征提取的可靠性。经过学习冲突信息，模型输出的输出功率和相关特征之间的数学关系容易失真，在动态预测场景中会出现逻辑混乱和预测跳变，使预测曲线的平滑度和可信度降低，最终严重削弱负荷预测模型的实用价值[5]。

2.3 数据不一致对精度的冲击

新能源场站运行环境中有许多复杂且不确定的因素，像传感器出现故障、通信受到干扰或者遇到极端气象条件等情况，这些因素都有可能产生异常数据。异常数据会对负荷预测精度造成干扰，影响比较突出。异常值会破坏模型所依赖的均衡分布状态，使得模型在学习过程中捕捉到错误的特征关联，最终导致模型拟合偏离真实出力的异常数值。要是没能及时进行识别并剔除，会使模型的预测逻辑出现严重偏差，特别是在异常值对应的时间点，预测结果常常表现为离谱的波动情况[6]。更为严重的是，这种干扰并不局限于异常点本身，还会让相关特征的映射关系产生偏差，致使模型在正常工况下的预测出现持续性偏移现象。在极端天气或者设备故障引发的异常数据场景中，模型会失去预测能力，整体预测误差也会因此大幅提高。

3 新能源场站基础数据治理对负荷预测精度提升路径

3.1 构建标准化数据治理体系

新能源场站基础数据的采集一直存在设备接口协议不兼容、字段命名随意、采集频次不统一和历史数据缺失等问题，这使负荷预测模型无法对多场站和多个时段的数据进行统一适配，成为精度提高的主要障碍。建立标准化数据治理模式的主要目的是把负荷预测的业务需求当作依据，把数据采集到入库的全过程标准化。需要先制定统一的场站元数据规范，明确逆变器、风力发电机和储能单元等主要设备的数据字段定义、采样周期、数据单位，并且强制要求新建场站和存量场站改造后都遵循统一协议接入数据中台。其次，需要建立数据资产目录，对已有的历史运行数据、气象数据和电网交互数据进行分级分类标注，完成分散脏数据的清洗和归一化预处理，把非结构化的运维日志和设备台账转化为可以被模型直接调用的结构化数据资源。最后需要建立全生命周期数据管理方式，明确运维方的数据上报责任和校验规则，从制度上保证数据不会产生新的偏差。为衡量治理成效，可引入数据标准化覆盖率公式（1）：

$$C=\frac{w_m S_m+w_s S_s+w_h S_h}{w_m+w_s+w_h}\times 100\% \quad (1)$$

例如，当场站元数据规范覆盖率 $S_m=85\%$、数据资产目录与结构化覆盖率 $S_s=85\%$、全生命周期数据管理机制覆盖率 $S_h=85\%$，权重分别为0.4,0.35,0.25时，计算得：

$$C=\frac{0.4\times 85\%+0.35\times 70\%+0.25\times 90\%}{1}\times 100\%=81\%$$

即该区域数据标准化覆盖率达到 81%,说明其已能为后续负荷预测模型提供较稳定、合规、高兼容性的输入基础，从源头上降低因数据不统一带来的预测误差。

3.2 开展实时的数据质量监控

传统负荷预测工作主要依靠周期性的人工数据核验，不能及时识别采集中断、设备漂移、异常值等隐性数据问题，错误的异常数据会直接误导模型参数，造成预测结果出现大幅偏差，所以进行实时数据质量监控是提高预测精度的主要保障环节。因此，需要建立包括全流程的监控指标模式，针对新能源场站数据出现的重点异常情况，设置包括采集连续性、数值合理性、关联一致性这三个方面的监控阈值，比如单个逆变器连续 10 分钟没有数据上报、功率数值超出设备额

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.

定范围、风速和功率的对应关系偏离理论区间等情况，系统会自动触发预警。并且要形成异常数据的自动处理闭环，对设备抖动产生的噪声数据进行滤波算法处理，对缺失的短时段数据使用邻近时段的时空关联数据进行插值补全，对无法判定的异常数据直接将其隔离到训练集中，防止脏数据进入模型训练。同时也要同步建立数据质量溯源方式，每次异常处理过程都会留下痕迹可以查询，一旦预测结果出现偏差可以快速追溯到对应数据源，并且对监控规则进行针对性调整，使输入模型的数据始终处于高可用状态，在数据输入端将误差控制在最小范围[7]。

3.3 强化特征提取与模型迭代

提高负荷预测精度主要依靠高质量的特征，依靠负荷历史时序数据无法完全捕捉新能源出力的多个影响因素，所以应以治理后的优质数据作为基础，增强特征提取和模型迭代的能力，把分散的数据价值转变为预测精度的提高。一方面，主要围绕负荷预测的多粒度和多场景需求进行特征工程，提取场站出力的周期性特征、气象关联特征、电网调度特征和储能充放电特征等多方面特征，利用特征重要性评估来选出对负荷波动影响最大的主要特征，并消除干扰特征，从而减少模型的计算负担并提高特征的针对性[8]。另一方面，要建立模型的持续迭代方式，针对不同场景的负荷预测需求建立合适的模型组合，例如针对晴好天气的短期负荷使用时序模型，针对极端天气的小时级预测使用融合气象的特征模型，同时按照季度和月度的周期把最新的运行数据纳入训练集来完成模型动态更新，使模型能够及时适应季节性出力规律的变化和设备老化引起的特性漂移。通过下述表格 2 对比不同迭代阶段的精度表现，使模型迭代效果更加明显。依靠该技术创新和效果验证，能够不断优化模型以适应新的运行场景，持续减少预测误差，提高负荷预测的实用性。

表 2　新能源场站负荷预测模型不同迭代阶段精度指标对比

模型迭代阶段	核心输入特征	平均预测误差率	极端场景误差率
初始基线版本	仅负荷时序数据	15.7%	28.3%
特征优化版本	负荷+气象+场站属性	10.2%	19.6%
动态迭代版本	全量特征+历史偏差修正	6.8%	12.4%

4 结束语

总之，新能源场站的基础数据治理对于负荷预测精度有着比较关键的影响。通过搭建标准化的数据治理模式，能够把数据采集、清洗、存储、共享这些流程进行统一，从一开始就能消除数据缺失、异常、不一致的问题。进行实时的数据质量监控，可以动态地识别并修正异常波动，保证数据的完整性和可靠性。未来应不断优化数据治理策略，以适应新能源场站复杂多变的运行环境，进一步提高电力系统的运行效率与安全性。

[参考文献]

[1]张彦伦,黄曦,王锐,石泽宇,陶守元.新能源场站接地变保护技术研究与实践[J].中国设备工程,2026,(18):226-228.

[2]徐晓伟,翁帅,杨小品,丁云川,张莲月.某地区电网中新能源场站调压能力分析[J].工业控制计算机,2026,39(9):124-126.

[3]施成林,刘英东,吴绍安,曹博禹,余建华,李正志.基于无线自组网的新能源智慧场站平台设计研究[J].中国设备工程,2026,(17):269-271.

[4]李伟.新能源场站供配电系统风险辨识与运行管控策略研究[J].石化技术,2026,33(9):373-375.

[5]陈鑫跃,马志远,孟杰,单小雨,冯旭.基于多源数据融合的新能源场站一次调频性能综合评估方法[J].电气技术与经济,2026,(8):78-82.

[6]涂强.新能源场站变电设备智能巡检与运维效率提升研究[J].应用能源技术,2026,(8):109-111.

[7]施政欣,李琦.大数据驱动新能源集控平台智能化技术研究[J].能源科技,2026,24(4):63-66.

[8]林甲增,李笑竹,王维庆,刘双硕,王波.考虑贡献均衡的新能源场站群共享储能两阶段容量-功率协同分配策略[J].电网技术,1-15.

Copyright © This work is licensed under a Commons Attibution-Non Commercial 4.0 International License.