

团 体 标 准

T/HZDJXH XXX—2023

计及海量分布式新能源的地区发电智能预测技术导则

Guidelines for intelligent prediction technology of regional power generation
considering massive distributed new energy

(征求意见稿)

2023-XX-XX 发布

2023-XX-XX 实施

惠州市电机工程学会 发布

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由惠州市电机工程学会提出并归口。

本文件起草单位：广东电网有限责任公司惠州供电局，惠州市电机工程学会

本文件主要起草人：

计及海量分布式新能源的地区发电智能预测技术导则

1 范围

本文件规定了计及海量分布式新能源接入地区的分布式光伏发电功率预测的基本技术方法。

本文件适用于海量分布式新能源接入场景下的分布式光伏电站日前发电功率预测,以及相应产品的设计、研发、检测及运行。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 30149 电网通用模型描述规范

GB/T 31464 电网运行准则

GB/T 33590.1-2017 智能电网调度控制系统技术规范第1部分:总体架构

GB/T 33590.2-2017 智能电网调度控制系统技术规范第2部分:术语

GB/T 33602-2017 电力系统通用服务协议

GB/T 33602-2017 电力系统简单服务接口规范

GB/T 40604 新能源场站调度运行信息交换技术要求

GB/T 40607 调度侧风电或光伏功率预测系统技术要求

DL/T 516 电力调度自动化系统运行管理规程

DL/T 1170 电力调度工作流程描述规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

分布式光伏电站 distributed photovoltaic (PV) station

利用分散式资源,装机规模较小的使用光伏电池的光生伏特效应,将太阳辐射能直接转换成电能的发电系统。

3.2

风电场 wind farm

由一批风力发电机组或风力发电机组群(包括机组单元变压器)、汇集线路、主升压变压器及其他设备组成的发电站。

3.3

数值天气预报 numerical weather prediction

根据大气实际情况,在一定的初值和边值条件下,通过大型计算机进行数值计算,求解描写天气演变过程的流体力学和热力学方程组,预测未来一定时段的大气运动状态和天气现象,通过数值的形式给

出不同气象要素的预报值。

3.4

光伏总辐射 total photovoltaic radiation

使用太阳总辐射表水平安装后测试到水平面上太阳的总辐射。

3.5

光伏散射辐射 photovoltaic scattered radiation

使用太阳散射辐射表（总辐射表+散射装置）测试到被遮光环遮挡掉太阳直接辐射后的辐射值。

3.6

光伏直接辐射 photovoltaic direct radiation

使用直接辐射表正对着太阳实时跟踪太阳测试的太阳辐射。

3.7

光伏组件温度 photovoltaic modules temperature

光伏组件中所有电池连接点温度的平均值。

3.8

环境温度 ambient temperature

光伏组件安装环境的冷热程度。

3.9

聚类簇 cluster

由聚类所生成的一组样本的集合。同一簇内样本彼此相似，与其他簇中的样本相异。

3.10

风电场有功功率 active power of wind farm

风电场通过其并网点输出到电网的有功功率。

3.11

数据清洗 data cleaning

对数据进行重新审查和校验的过程，目的在于删除重复信息、纠正存在的错误，并提供数据一致性。

4 分布式新能源发电预测

4.1 分布式光伏发电预测

4.1.1 数据要求

分布式光伏发电功率预测使用的数据应包括分布式光伏电站的功率数据、环境监测站数据、数值天

气预报、光伏电站装机容量及其地理位置信息等。具体内容包括：

- a) 分布式光伏电站的历史运行数据应不少于1年，功率数据的时间分辨率应为15分钟；
- b) 分布式光伏电站装机容量及其地理位置信息；
- c) 分布式中压光伏电站气象环境监测设备应部署在分布式中压光伏电站厂区内，周围空旷无遮挡，且不影响光伏组件，监测信息至少应包括总辐射、直接辐射、散射辐射、组件温度、环境温度以及风速、湿度等信息，数据的测量时间分辨率应不小于15分钟；
- d) 分布式低压光伏电站气象环境检测设备应部署在各个分布式低压光伏电站聚类簇质心的位置，周围空旷无遮挡，且不影响光伏组件，监测信息至少应包括总辐射、直接辐射、散射辐射、组件温度、环境温度以及风速、湿度等信息，数据的测量时间分辨率宜不小于15分钟；
- e) 数值天气预报数据的时间分辨率宜为15分钟，应至少包括短波辐射、温度、湿度、风速等参数。

4.1.2 数据预处理

分布式光伏发电功率预测使用的数据需进行数据处理。具体包括：

- a) 数据应进行数据清洗，清洗后的数据应筛选大部分异常值；
- b) 清洗后的数据应进行数据填补，填补后的数据无空值与负值；
- c) 光伏发电数据应进行归一化或标准化处理，处理后的数据间应不存在不同尺度和幅度之间的差异等问题；

4.1.3 发电智能预测

分布式光伏发电功率预测应具备以下功能：

- a) 应能覆盖整个调度管辖区域内所有分布式光伏电站；
- b) 短期光伏功率预测结果为整域分布式光伏电站出力功率总量且时间分辨率应为15分钟；
- c) 短期预测应为次日96点有功功率曲线，并具备预测从次日0时起至72小时有功功率曲线的能力；
- d) 短期功率预测应能手动设置预测的启动时间及次数；
- e) 应能支持对预测结果的人工修正，并设置响应的权限管理；
- f) 能够对预测曲线进行误差统计，估算一定置信度的误差范围。

4.2 风光联合功率预测

4.2.1 风力发电数据要求

风力发电功率预测使用的数据应包括风电场的功率数据、测风塔测风数据、风电机组运行状态数据、风电机组特性、数值天气预报、风电机组故障数据、风电机组的运行限制条件等。具体内容包括：

- a) 风电场的历史运行数据应不少于1年，实际功率数据的时间分辨率应为15分钟；
- b) 测风塔应在风电场外1~5千米范围内，至少应包括10米、50米及轮毂高程的风速和风向，以及某一层高的气温、气压、湿度等信息，测风数据的时间分辨率宜不小于5分钟；
- c) 风电机组信息应包括机组型号、单机容量、轮毂高度、叶轮直径、功率曲线、推力系数、并网时间、风电机组位置（经、纬度）、海拔高度等；
- d) 数值天气预报数据的时间分辨率宜为15分钟，应至少包括10米、50米、70米三个层高的风速和风向以及温度、湿度、气压、云量等参数；
- e) 应添加风电场地理位置、地形和风资源分布图，以及周边地区的气象特征和季节变化等信息；
- f) 风电机组故障数据应包括风电机组故障报告和维护记录，以及故障类型和故障发生的频率等信息；
- g) 应提供风电机组的运行限制条件，如最小切入风速、最大容量限制、最大风机转速等。

4.2.2 风力发电数据预处理

风力发电功率预测使用的数据需进行数据处理。具体包括：

- a) 风力发电数据应进行数据清洗，清洗后的数据应剔除大部分异常值；
- b) 清洗后的数据应选择合适的数据插补方法填补数据中的缺失值，填补后的数据无空值；
- c) 风力发电数据应进行平滑处理，处理后的数据中应无明显噪声数据；
- d) 风力发电数据应进行特征提取，应能得到时间特征、季节特征、风速和风向的统计特征等；
- e) 风力发电数据应进行归一化或标准化处理，处理后的数据间应不存在尺度和幅度的差异等问题；
- f) 风力发电数据应进行聚合处理，处理后的数据应能满足不同预测模型的要求。

4.2.3 风光发电特征分析

风光联合功率预测需要对风力发电和光伏发电数据进行特征分析，具体包括以下内容：

- a) 进行风光发电相关性分析，在分析过程中应考虑季节性的影响，构建风光联合出力 Copula 模型；
- b) 基于风光发电相关性分析的结果，确定风力和光伏发电的季节性和年度变化模式；
- c) 风光数据应进行数据清洗，清洗后的数据应筛选大部分异常值，且满足 Copula 模型范围需求；
- d) 应考虑使用插值技术、回归模型等方法进行风光数据填补，应能确保填补后的数据满足 Copula 模型的要求；
- e) Copula 模型选择和参数估计，应能基于风光数据特点选择适合的 Copula 模型，并确定其参数估计方法；
- f) 应讨论风光联合预测中的不确定性和潜在风险因素。考虑引入不确定性分析方法、考虑不同的风力和光伏发电场景、评估预测误差的可靠性等内容。

4.3 分析管理

4.3.1 数据统计

应对历史出力数据进行统计，主要功能为：

- a) 能对分布式低压光伏电站进行聚类处理；
- b) 能对历史功率数据进行统计，内容应包括数据完整性统计、频率分布统计、变化率统计等；
- c) 应能对历史辐射数据、历史测风数据、数值天气预报数据进行统计，内容包括完整性统计、风速频率分布统计、风向频率统计、辐射分布统计等；
- d) 应能对新能源电场运行参数进行统计，内容包括发电功率、有效发电时间、电站实际投入容量、最大出力及其发生时间、利用小时数及平均负荷率等。

4.3.2 相关性检验

相关性检验应满足以下要求：

- a) 能对历史功率数据，辐射数据和数值天气预报数据进行相关性检验；
- b) 能对相邻光伏电站的功率数据进行相关性检验，对异常数据进行标识。

4.3.3 误差统计

误差统计应满足以下要求：

- a) 应能对任意时段的预测结果进行误差统计；
- b) 应能对各新能源电场上报的预测曲线进行误差统计；
- c) 误差统计指标至少应包括均方根误差、平均绝对误差、最大预测误差等。

4.3.4 预测评价指标

功率预测结果评价是评价预测功率与实际功率的偏离程度，可以科学有效地评价光伏发电功率预测

结果。常用的光伏发电功率预测结果评价指标包括均方根误差、准确率、平均绝对误差、相关性系数、合格率、极大误差率、95%分位数偏差率、预测数上报率等。

- 1) 均方根误差 E_{rmse} : 均方根误差用于评价预测误差的分散程度, 可从整体上对光伏发电功率的预测效果进行评价, 其计算方法为:

$$E_{rmse} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{P_{Pi} - P_{Mi}}{C_i} \right)^2} \quad (1)$$

式中: P_{Pi} —— i 时刻的预测功率; P_{Mi} —— i 时刻的实际功率; C_i —— i 时刻的开机容量; n ——统计时段内的总样本数。

- 2) 准确率 C_R : 准确率可以根据均方根误差结果计算得到, 数值越大, 对应的预测精度越高。准确率较为直观, 因此在电力调度中得到普遍应用, 其计算方法为:

$$C_R = 1 - \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{P_{Mi} - P_{Pi}}{C_i} \right)^2} \quad (2)$$

- 3) 平均绝对误差 E_{mae} : 平均绝对误差用于表征预测结果与实际结果的平均偏离程度, 计算方法为:

$$E_{mae} = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_{Pi} - P_{Mi}}{C_i} \right| \quad (3)$$

- 4) 相关性系数 r : 相关性系数用于表征两组曲线趋势的一致性, 取值范围为 $[-1, 1]$, 其值接近 1 时表明预测功率与实际功率的线性正相关程度越大, 理想的预测结果为 $r = 1$ 。当 r 接近 0 时表明预测功率与实际率线性不相关, 其值小于 0 时表明预测功率与实际功率线性负相关, 计算方法为:

$$r = \frac{\sum_{i=1}^n [(P_{Mi} - \bar{P}_M)(P_{Pi} - \bar{P}_P)]}{\sqrt{\sum_{i=1}^n (P_{Mi} - \bar{P}_M)^2 \sum_{i=1}^n (P_{Pi} - \bar{P}_P)^2}} \quad (4)$$

- 5) 合格率 Q_R : 合格率指预测合格点数占评价时段总点数的百分比合格点数指预测绝对偏差小于给定阈值的点数, 计算公式为:

$$Q_R = \frac{1}{n} \sum_{i=1}^n B_i \times 100\% \quad (5)$$

$$B_i = \begin{cases} 1 & \left| \frac{P_{pi} - P_{Mi}}{C_i} \right| \leq T \\ 0 & \left| \frac{P_{pi} - P_{Mi}}{C_i} \right| > T \end{cases} \quad (6)$$

其中, B_i 代表 i 时刻预测绝对误差是否合格, 若合格为 1, 不合格为 0; T 为判定阈值, 根据电力调度部门的实际情况确定, T 一般不大于 0.25。

- 6) 极大误差率 E_{ex} : 极大误差率主要反映单点的极端偏差情况, 可为运行人员提供直观的极端误差参考, 为系统的运行提供依据, 其计算表达式为:

$$E_{ex} = \max \left(\frac{|P_{Pi} - P_{Mi}|}{C_i} \right) \quad (7)$$

考虑到正负误差在系统运行控制中的不同影响,给出了正极大误差率 E_{exp} 和负极大误差率 E_{exn} 的计算公式为:

$$E_{exp} = \max \left(\frac{P_{Pi} - P_{Mi}}{C_i} \right) \quad (8)$$

$$E_{exn} = \min \left(\frac{P_{Pi} - P_{Mi}}{C_i} \right) \quad (9)$$

- 7) 95%分位数偏差率: 95%分位数偏差率包括 95%分位数正偏差率和 95%分位数负偏差率。95%分位数正偏差率指将评价时段内单点预测正偏差率由小到大排列, 选取位于第 95%位置处的单点预测正偏差率, 计算表达式为:

$$\begin{cases} E_i = \frac{P_{Pi} - P_{Mi}}{C_i} \dots 0 & i = 1, 2, \dots, n \\ E_j = \text{sortp}(E_i) & j = 1, 2, \dots, n \\ P_{95p} = E_j & j = \text{INT}(0.95n) \end{cases} \quad (10)$$

95%分位数负偏差率指将评价时段内单点预测负偏差率由大到小排列, 选取位于第 95%位置处的单点预测负偏差率, 计算表达式为:

$$\begin{cases} E_i = \frac{P_{Pi} - P_{Mi}}{C_i} \dots 0 & i = 1, 2, \dots, n' \\ E_j = \text{sortn}(E_i) & j = 1, 2, \dots, n' \\ P_{95n} = E_j & j = \text{INT}(0.95n') \end{cases} \quad (11)$$

式中 P_{95p} 、 P_{95n} ——分别为 95%分位数正偏差率和负偏差率, 其取值步长根据具体情况而定, 无量纲; E_i —— i 时刻预测偏差率, 无量纲; E_j ——排序后的单点预测偏差率, 无量纲; $\text{sortp}(\)$ ——由小到大排序函数; $\text{sortn}(\)$ ——由大到小排序函数; $\text{INT}(\)$ ——取整函数; n 、 n' ——分别为评价时段内的正偏差样本数和负偏差样本数量, 应不少于 1 年的同期数据样本。

- 8) 预测数据上报率 R : 调度机构为考核光伏电站上传的预测数据完整率, 提出了预测数据上报率的指标, 其计算公式为:

$$R_r = \frac{R}{N} \times 100\% \quad (12)$$

式中 R ——评价时段内数据上报成功次数; N ——评价时段内应上报次数。

- 9) 数据完整率 D_r : 考虑到同一时刻下预测功率数据或实测功率数据出现异常或缺失时无法计算对应的误差指标, 提出了数据完整率的统计指标, 计算方法为:

$$D_r = \frac{n_{all} - n_{fault}}{n_{all}} \quad (13)$$

式中 n_{all} ——统计时段内的样本总量； n_{fault} ——不合格数据数量。

由于单一评价指标存在不同程度的局限性，为反映预测的整体状态，在实际应用中，应能考虑不同的预测指标，构建综合评价指标，更加全面地对预测功率的优劣做出综合性评价。

4.4 接口要求

4.4.1 数据输入

分布式新能源发电预测模型各类数据需求如下：

- a) 定时获取至少次日 24 小时的专业数值天气预报数据，时间分辨率为 15 分钟；
- b) 各台区分布式新能源电站的实际有功功率数据和机组状态数据，时间分辨率应短于 15 分钟；
- c) 气象站网格化实测数据，包括不同层高的风向、风速、辐照度、湿度、温度等，时间分辨率不宜超过 15 分钟；
- d) 集中式新能源场站的短期发电功率预测曲线、限电记录，及次日 96 点开机容量；

4.4.2 数据输出

通过分布式新能源发电预测向其它应用提供预测分析结果如下：

- a) 提供区域分布式新能源发电功率次日 96 点预测结果及置信度为 80% 的误差范围；
- b) 提供区域分布式新能源发电功率 4 小时内的预测结果及置信度为 80% 的误差范围；
- c) 提供月误差统计报表；
- d) 提供区域分布式新能源与各类气象因素的相关性分析结果。