

# 基于 Python 的 SCADA 数据分析诊断

王佳楠

国华能源投资有限公司 北京 100000

**【摘要】**随着风力发电的迅速发展, SCADA系统作为风电机组数据采集与监控的重要工具, 提供了大量的实时数据支持, 为风电场的高效运行与决策支持提供了可靠依据。本文深入探讨了风电机组SCADA数据的特点、数据采集与传输方式, 并分析了数据的质量、存储与隐私安全等方面的挑战。结合实际案例, 介绍了基于Python的风电数据分析过程, 强调了数据清洗、预处理及功率曲线等分析方法的应用。最终, 研究指出了风电SCADA数据在提升风电机组运行效率、优化决策及保障数据安全方面的广阔前景。

**【关键词】** SCADA数据、Python; 风电机组、数据存储、智能化管理

Analysis and diagnosis of SCADA data based on Python

Wang Jianan

Guohua Energy Investment Co., LTD., Beijing 100000

**【Abstract】** With the rapid development of wind power generation, SCADA system, as an important tool for data collection and monitoring of wind turbines, provides a large number of real-time data support, which provides a reliable basis for the efficient operation and decision support of wind farms. This paper deeply discusses the characteristics of wind turbine data of SCADA, data collection and transmission mode, and analyzes the challenges of data quality, storage and privacy security. Combined with practical cases, the wind power data analysis process based on Python is introduced, and the application of analysis methods such as data cleaning, pretreatment and power curve is emphasized. Finally, the research points out the broad prospects of wind power SCADA data in improving the operation efficiency of wind turbine, optimizing decision-making and ensuring data security.

**【Key words】** SCADA data, Python; wind turbine, data storage, intelligent management

## 1. 引言

随着全球能源转型的加速, 风力发电作为清洁能源的重要形式得到了广泛应用。SCADA系统能够实时采集风机的运行数据, 包括工况信息、环境条件及电气参数等, 为风电场的高效运行与决策支持提供了重要依据。然而, 由于风电场规模的扩大及SCADA系统的复杂性, 如何充分挖掘数据的潜力, 同时克服数据质量、隐私安全及算法适用性等多重挑战, 成为当前研究的重点。

## 2. 风电机组 SCADA 数据的特点

### 2.1 数据来源与采集方式

风电机组的 SCADA 系统通过对风机运行状态的实时监测与控制, 提供了大量的数据支持。SCADA 数据的采集主要依赖于传感器、控制器和数据通信设备的协同工作。风机本体上分布着多种类型的传感器, 包括测量风速、风向的气象传感器, 监测叶片角度的桨距传感器, 以及用于轴承温度、振动的运行状态传感器。每个传感器实时采集数据, 并通过

嵌入式控制器进行预处理<sup>[1]</sup>。

采集到的数据通过通信网络传输至集中控制系统, 这一过程通常采用光纤、无线或有线网络。现代风电场普遍采用工业级通信协议, 如 Modbus、IEC 61850 等, 确保数据传输的稳定性与高效性。此外, SCADA 系统还通过历史数据存储模块记录数据, 以便后续分析与建模<sup>[2]</sup>。

### 2.2 数据类型与结构

SCADA 系统生成的数据种类繁多, 涵盖了风机运行的多个方面, 主要可分为工况数据、环境数据和电气参数数据三类。

#### (1) 工况数据

工况数据记录了风电机组的运行状态, 包括转速、功率输出、叶片角度、偏航状态等。这些数据反映了风机运行的动态特性, 是性能优化与故障诊断的核心依据。例如, 通过分析功率输出数据, 可以评估风机的发电效率; 而转速与叶片角度的数据结合, 则可用于监测风机是否处于最佳工况<sup>[3]</sup>。

#### (2) 环境数据

环境数据主要涉及风资源与外部条件, 包括风速、风向、气温、湿度等。风速与风向是影响风机功率输出的关键变量, 而气温与湿度则影响风机部件的运行稳定性和寿命。这些数

据通常由独立的气象站或机组自带的气象传感器提供,并实时更新。

### (3) 电气参数数据

电气参数数据涵盖电机的电压、电流、功率因数等,这些数据反映了电能传输的效率和机组电气系统的健康状态。通过分析电气参数,可以发现潜在的电气故障,如发电机过载或电流波动异常,从而避免重大损失<sup>[4]</sup>。

### 2.3 数据的时间分辨率与实时性

SCADA 系统的时间分辨率与实时性是其数据特点中的重要维度。通常,SCADA 系统的数据采集频率在秒级或分钟级,例如风速、功率输出等关键数据每秒更新一次,而一些非关键数据则可能以分钟或小时为单位更新。这种高时间分辨率的数据采集方式确保了对风机运行状态的实时感知。

实时性是 SCADA 系统的核心优势之一。通过对传感器数据的快速采集与处理,SCADA 系统可以及时捕获风机运行中的异常现象。例如,当风速突然增加或振动传感器检测到异常震动时,SCADA 系统能在秒级时间内发出报警信号,提醒运维人员采取措施。与此同时,这种实时性也为风电场的调度优化提供了支持,使得风能利用率得到最大化。

### 2.4 数据量及存储挑战

以一个典型的风电场为例,单台风机每天可生成数百 MB 的数据,而一个百台规模的风电场每年将产生数百 TB 的数据量。如此庞大的数据规模对存储设备和数据管理提出了严峻挑战。

数据存储的首要难题是容量需求的快速增长。传统的本地存储设备难以满足现代风电场的存储需求,因此云存储与分布式存储逐渐成为主流选择。这些技术不仅能提供海量存储空间,还支持数据的快速检索与共享。此外,数据的冗余备份与容灾策略也是存储系统的重要组成部分,以保障数据安全与完整性。

另一个关键挑战是数据管理与处理效率。高频率的实时数据和低频率的历史数据需要分类存储和处理,而大数据技术在这一领域显示出巨大的潜力。例如,通过分布式数据库和数据湖技术,可以高效地存储、索引和分析 SCADA 数据,从而支持风电场的智能化决策。

## 3 案例分析

### 3.1 项目背景

风力发电和光伏发电作为可持续能源发展的关键领域,近年来得到了迅猛发展。随着国华投资公司装机规模的不断增加,对能效监督、控制与保护监督的要求也在不断增加。在各公司的运营过程中积累了大量的数据,如发电设备运行参数、气象数据、发电量数据等。有效分析这些数据能够帮助各公司深入了解发电系统的性能,优化发电计划,降低运营成本,提高发电效率和可靠性,以及更好地应对市场竞争和政策要求。Python 作为一种功能强大、易于学习且拥有

丰富数据处理和分析库的编程语言,在各公司数据分析中具有广阔的应用前景。

### 3.2 实施过程

#### 3.2.1 数据特点

新能源发电数据具有多源性、实时性、海量性和复杂性等特点,这些数据实时产生且数据量庞大,并且数据之间存在复杂的关联关系。

#### 3.2.2 数据清洗与预处理

由于数据量庞大,需要先对数据进行必要的清洗。使用 Python 中的数据处理库(如 Pandas)可以方便地进行数据清洗操作。例如,通过数据筛选和条件判断去除明显的错误数据,利用统计方法识别和处理异常值,采用插值法或其他填充策略填充缺失值,以提高数据的质量和可用性。数据预处理包括数据标准化、归一化等操作。对于不同量级和量纲的数据,如发电功率与气象数据,进行标准化处理可以使它们在同一尺度上进行分析,避免因数据量级差异导致的分析偏差。

#### 3.2.3 数据分析与可视化

##### (1) 功率曲线分析

通过绘制风速与功率输出的散点图(使用 Matplotlib 库),发现两者呈现出明显的非线性关系。在低风速区域,功率输出随着风速的增加而缓慢上升;当风速达到额定风速后,功率输出趋于稳定;当风速超过切出风速时,风电机组为保护自身安全停止运行,功率输出降为零。利用 Scikit-learn 库中的线性回归和多项式回归模型对风速与功率关系进行拟合。结果表明,多项式回归模型能够更好地拟合数据,其拟合优度(R 值)较高,说明该模型能够较为准确地描述风速与功率之间的关系。通过模型可以预测不同风速下的理论功率输出,与实际功率输出进行对比,从而评估风电机组的发电效率。通过与叶片角度数据进行关联分析,发现当叶片角度趋近于 0 时,变桨电机的扭矩也会增加,但不会超过经验阈值。

### 3.3 完成情况

现已完成华锐、金风机组的运行参数数据分析的编程工作,并收集了大量经验数据,制定分析模型,快速协助场站做好数据分析工作。未来将根据场站的需求,不断改进、完善数据分析代码,并根据需求定制化开发其他相关功能,以满足现场的使用情况,提高数据分析的可靠性,并作为各子分公司决策的重要依据。

## 4. 面临的挑战与对策

### 4.1 数据质量与可靠性

挑战: SCADA 数据的质量是数据分析与模型开发的基础,但在实际采集过程中,可能受到多种因素的影响,导致数据质量问题。首先,传感器的故障或精度不足会引入噪声或错误数据。其次,数据缺失是常见问题,可能由通信故障

或采集设备的不稳定性引起。此外，环境条件的变化，例如极端天气，也可能导致数据异常。这些问题直接影响了后续分析结果的准确性与可靠性。

对策：提高数据质量需要多层次的措施。首先，应加强传感器的维护与校准，确保其长期稳定运行。其次，在数据处理阶段引入数据清洗技术，对缺失数据进行合理补全，例如利用插值法、时间序列模型等方法。此外，可以通过异常检测算法自动识别和剔除错误数据。例如，基于统计学的箱线图法或基于机器学习的孤立森林算法可用于检测异常点。同时，通过实时监控系统的建设，可以在数据采集过程中快速发现和纠正问题，进一步提高数据可靠性。

#### 4.2 数据隐私与安全性

挑战：随着风电场规模的扩大和数据量的增长，SCADA 系统的数据隐私与安全问题逐渐受到重视。SCADA 数据不仅包括风电机组的运行信息，还可能涉及风电场调度策略与运维记录，这些数据具有较高的敏感性。一旦数据泄露或被恶意篡改，将对风电场的运行带来严重威胁。此外，风电场与电网之间的数据交互也增加了潜在的网络安全风险。

对策：为保障数据隐私与安全，应从技术与管理两方面入手。在技术上，需加强数据加密与访问控制。例如，采用传输层安全协议（TLS）对数据传输过程进行加密，并引入身份认证机制，确保只有授权用户能访问敏感数据。同时，可采用区块链技术实现数据存储与共享的透明化与不可篡改性。在管理上，应建立严格的数据安全管理制度，明确数据访问权限与责任分工。此外，定期进行网络安全评估和漏洞检测，及时更新防火墙和入侵检测系统，以应对潜在的网络攻击。

#### 4.3 算法性能与模型泛化能力

挑战：SCADA 数据的分析与应用需要依赖算法模型的支持。然而，由于 SCADA 数据具有非线性、非平稳性等特点，许多传统算法在处理时存在适应性不足的问题。此外，不同风电场的地理位置、风资源条件和设备参数差异较大，导致同一模型难以在不同场景中实现泛化。算法的性能瓶颈和泛化能力不足直接影响了分析结果的准确性与模型的推广应用。

对策：针对算法性能的问题，可通过引入深度学习技术提升数据处理能力。例如，循环神经网络（RNN）和长短期

记忆网络（LSTM）在时间序列预测中表现出色，适用于处理 SCADA 数据的动态特性。此外，可采用迁移学习技术，在一个风电场训练的模型基础上，对其他风电场进行二次训练，从而提高模型的适用性与泛化能力。同时，进行多场景、多维度的数据融合分析，构建具有更强鲁棒性的通用模型。最后，通过模型的动态更新与自学习机制，确保算法能够持续适应不同的运行场景。

#### 4.4 数据应用的标准化与行业规范

挑战：SCADA 数据的标准化程度直接影响数据的共享与互操作性。然而，目前风电行业中不同厂家、不同地区的 SCADA 系统在数据格式、通信协议和分析方法上缺乏统一标准，导致数据整合与比较分析的难度较大。此外，由于行业内对数据使用的规范性要求不明确，可能出现数据误用或数据价值未能充分挖掘的问题。

对策：推进数据标准化需要行业内外的协同努力。首先，应制定统一的数据格式与通信协议标准，例如采用 IEC 61400 等国际标准，规范风电机组的 SCADA 数据采集、存储与传输。同时，鼓励设备制造商和风电场运营商开放数据接口，推动跨平台的数据共享与协作。在此基础上，建立行业性的数据库与数据管理平台，促进不同风电场之间的数据比较与经验交流。此外，应加强政策引导与监管，明确数据使用的法律规范，确保数据应用的合法性与规范性，从而提升风电行业整体的智能化水平。

### 5. 结束语

SCADA 数据是风电机组高效运行与智能管理的重要支撑，其来源的多样性和数据结构的复杂性既为风电场的优化提供了广阔的可能性，也提出了数据处理和应用的严峻挑战。在全面分析了 SCADA 数据的特点及应用中的难点后，可以发现，通过提升数据质量、保障数据安全、优化算法性能和推进行业标准化，可以有效释放 SCADA 数据的潜力。未来，随着云计算、大数据和人工智能技术的深度融合，SCADA 数据的应用前景将更加广阔。这不仅有助于提升风电场的运行效率，也将推动风电行业向智能化和可持续发展的方向迈进。

### 参考文献

- [1]胡龙舟,李韬睿,吴頔,等.基于 SCADA 系统的风电机组 KNN 故障状态监测研究[J].机械设计与制造工程,2025,54(01):91-94.
- [2]岳子桐,李艳婷,赵宇.基于神经网络和稳健估计的风电机组状态监测[J/OL].中国机械工程,1-13[2025-01-24].
- [3]岳健,史秉帅,范寒,等.基于多级特征提取框架的风电机组载荷预测方法[J].太阳能学报,2024,45(12):350-359.
- [4]于华楠,李靖雨,王鹤,等.基于动态集群的风电机组异常状态检测方法[J/OL].电力自动化设备,1-15[2025-01-24].