

生态环境在线监测数据异常识别研究

张有福

呼和浩特市生态环境监控中心, 内蒙古 呼和浩特 010020

DOI:10.61369/EAE.2026040002

摘 要 : 生态环境在线监测数据异常既包括设备失真、传输缺陷和量值漂移, 也包括工况变化引起的真异常与伪异常混杂问题。文章围绕生态环境在线监测数据异常识别展开研究, 先梳理异常来源和异常分类, 再从阈值判别、时序分析、多变量关联和智能预警等方面总结识别方法, 最后提出人工复核、场景化校验和模型迭代相结合的闭环路径。研究认为, 在线监测异常识别的关键不在于规则设置得越严越好, 而在于让规则、算法和业务复核能够共同区分设备异常、数据异常和排放异常。

关 键 词 : 生态环境在线监测; 数据异常; 异常识别; 时序分析; 智能预警

Research on Abnormal Identification of Ecological Environment Online Monitoring Data

Zhang Youfu

Hohhot Ecological Environment Monitoring Center, Hohhot, Inner Mongolia 010020

Abstract : Abnormal online monitoring data of ecological environment includes both equipment distortion, transmission defects, and value drift, as well as the mixed problem of true and false anomalies caused by changes in working conditions. The article focuses on the identification of anomalies in online monitoring data of ecological environment. Firstly, the sources and classifications of anomalies are sorted out. Then, identification methods are summarized from the aspects of threshold discrimination, time series analysis, multivariate correlation, and intelligent warning. Finally, a closed-loop path combining manual review, scenario verification, and model iteration is proposed. Research suggests that the key to identifying anomalies in online monitoring is not to set stricter rules, but to enable rules, algorithms, and business reviews to jointly distinguish equipment anomalies, data anomalies, and emission anomalies.

Keywords : online monitoring of ecological environment; data abnormality; abnormal recognition; time series analysis; smart early warning

引言

在线监测系统持续扩容后, 生态环境管理越来越依赖连续数据来识别风险、评估排放和辅助执法。然而, 平台看到的异常并不都来自真实污染过程, 其中既有传感器漂移、通信丢包、停机补传, 也有负荷切换和治理设施调整带来的正常波动。若不区分异常来源, 很多真问题会被当成设备噪声忽略, 很多假问题又会被反复核查消耗监管资源。因此, 异常识别研究的核心, 不是简单提高报警数量, 而是提高异常结论与真实场景之间的一致性。

一、生态环境在线监测数据异常的主要来源

(一) 仪器漂移和量值失真会在长周期内累积成系统偏差

在线监测仪器并不会一直保持理想稳定状态, 零点漂移、量程偏差、采样管路污染和比对维护滞后, 都会使监测值逐步偏离真实水平。这类问题最难识别之处在于曲线往往依然连续, 看上去“很像真的”, 但长期偏差会不断累积并最终误导分析结果。环境在线监测数据质量控制研究指出, 采集端质控一旦松动, 后续平台看到的很多异常其实并不是排放变化, 而是量值失真。异

常识别首先要识别那些由设备状态变化引起的系统偏差, 而不能把所有连续曲线都默认当作真实排放轨迹^[1]。异常识别起点不应只看单个点位超限, 还要同步审视采样链完整性、运维触发记录和时间同步状态, 避免把链路缺陷误判成排放异常。缺少这一层校验, 系统很容易把链路缺陷误读成排放波动, 后续判断基础也会被一起带偏。

(二) 传输缺陷和平台处理错误会制造结构化假异常

数据异常并不只发生在前端仪器, 传输断点、补传逻辑、时间戳错位、字段映射错误和状态码混乱, 同样会制造看似规律的

假异常。例如同一段时间数据被重复入库、补传数据覆盖原始值、离线状态被误写为零值，这些问题在图形上都可能表现为突降、平直或突升。污染源排放数据质量影响因素研究指出，平台规则和数据处理流程本身也是异常生成的重要来源。异常识别需要同时盯住前端数据和中后台处理轨迹，避免把结构化处理错误误判为现场排放异常^[2]。工况波动解释需要结合治理设施响应迟滞、旁路切换和短周期负荷变化，因为很多异常并非数值本身错误，而是语境缺失导致的误识别。语境一旦缺失，原本正常的短时响应就会被错误归为异常，从而推高不必要的核查负担。

（三）工况变化和治污设施切换会形成真假混杂的业务异常

业务异常最复杂的地方，在于它既可能是真实风险，也可能是合理波动。设备启停、负荷切换、旁路切换、冲洗校准和治污设施检修，都会让监测曲线出现阶段性变化。如果系统缺少工况标签和维护记录，这些波动很难被正确解释。污染源在线自动监测数据智能预警研究表明，异常识别如果脱离运行场景，只靠时间序列本身作判断，往往会把部分正常工况变化误判为排放异常。业务异常识别必须把工况语义纳入分析框架，才能在“真异常”和“伪异常”之间建立边界。平台若只保存结果值而不保留补传标识、判废原因和人工修正痕迹，后续训练样本就会被污染，模型学习到的也会是混杂规则。只有把这些过程痕迹完整保留下来，样本库才不会在多轮处理后丢失原始判定依据。

二、在线监测数据异常识别的基础规则方法

（一）阈值超限与突跳判别适合筛查显性异常

显性异常最适合用规则方法快速识别，如突跳、断崖、长时间恒值、超量程和数据缺测后异常回填等。这类规则透明、可解释，适合作为异常识别第一层筛查，把大量明显问题尽快拦截出来。规则方法的关键不在于条目越多越好，而在于阈值是否分因子、分行业、分场景设置。如果所有数据都套用同一波动阈值，在高波动行业中会制造大量误报，在低波动因子上又可能放过慢性异常。规则方法要与监测对象特征保持一致，才能让第一层筛查既灵敏又不过度噪声化。采集端质控还应增加维护前后对照、零跨复检和异常点回看，使同一设备在不同维护阶段的稳定性差异能够被持续追踪。这样才能分清是设备状态变化，还是维护动作本身引入了新的扰动因素。

（二）时序连续性与周期性分析有助于识别缓慢漂移异常

有些异常并不会以一次性突变出现，而是表现为缓慢抬升、缓慢回落、周期被打断或峰谷结构异常变化。对这类问题，仅用瞬时阈值很难发现，需要通过时序连续性和周期性分析来判断其是否偏离历史规律。若同一监测点在相似工况下长期高于历史中位水平，或原本稳定的昼夜节律突然消失，就需要进一步排查是否存在仪器漂移、采样异常或治理设施状态变化。时序分析的价值，在于把异常从“单点高低”提升到“整体走势是否失常”的层面，使一些不显眼但影响很大的慢性问题不被忽略。对传输链异常的识别，可增加字段漂移监测、频率缺口扫描和时序错位校核，先排除链路问题，再进入排放逻辑层面的判断。先排除链路

故障，再做排放解释，识别结果才不会在前置环节上反复失真。

（三）关联参数比对能识别单因子正常但逻辑失衡的异常

在线监测数据很多时候不是数值离谱，而是逻辑不通。例如浓度曲线大幅波动但流量和负荷完全静止，或排放值稳定不变但氧量、温度和运行状态明显切换，这些现象都提示单因子虽然看着正常，整体逻辑却已经失衡。污染源排放数据质量研究说明，多参数关联是区分“真排放变化”和“假数据变化”的重要方法^[3]。异常识别不能停留在单点数据是否超线，而应把浓度、流量、温度、设备状态、维护记录等一起纳入比较，使异常判断具备更强的业务解释力。语义层建模时要把停机、低负荷、切料、冲洗、标定和故障旁路分别编码，只有状态刻画足够细，模型输出才不会过度泛化。状态划分越粗，模型越容易把不同来源的问题压成同一种异常表型。

三、复杂场景下的智能识别方法

（一）基于特征提取的模型更适合处理多维时序混杂问题

当监测因子增多、采样频率提高、异常模式变复杂后，单纯人工制定规则会迅速遇到边界不足的问题。此时更适合先对时间窗口、波动幅度、变化速率、峰谷分布和关联参数差值等特征进行提取，再由分类模型或聚类模型识别异常模式。这样做的优势在于，它能看似无序的曲线变化转成可比较的特征结构，从而更高效地识别重复出现的异常形态。智能模型的意义不是替代监管经验，而是让高维数据先完成初步分组和优先级排序，使人工复核更聚焦。显性异常规则库应定期用历史误报案例回标，校正连续平值、极端跳变和量程顶格的识别阈值，避免规则长期僵化后命中率下降。规则若长期不校正，典型异常也会逐渐失去应有的识别灵敏度和稳定性。

（二）机器学习辅助预警能提高隐性异常的提前发现能力

隐性异常往往出现在短时高值之前或设备失效之前，其前兆可能只是多个变量之间的弱相关变化。机器学习模型在这类场景下的优势，在于能够从大量历史数据中捕捉“异常前的异常”，即在显著问题出现前发现模式偏移。人工智能辅助污染源追踪研究指出，算法更适合完成复杂场景下的高频筛查和候选异常定位，而最终结论仍需回到业务核验。在生态环境在线监测场景下，机器学习更适合承担预警前置角色，把需要人工关注的时间窗提前挑出来，而不是直接替代监管判定。隐性异常更适合采用多变量联动判别，例如把浓度、流量、氧含量、负压和电流共同纳入窗口分析，识别整体逻辑是否自洽。这种整体验证比单点判断更能发现数值表面正常但机理关系已经断裂的情况^[4]。

（三）智能识别必须与质量标注和样本更新机制同步建设

很多智能识别系统效果不稳定，并不是算法本身能力不足，而是缺少高质量标注样本和持续更新机制。若训练样本中将设备异常、补传异常和真实排放异常混在一起，模型学习到的就是模糊甚至混乱的判断边界，后续预警自然容易误判。若模型上线后不持续回收误报、漏报和人工复核案例，识别能力也会随着工况变化、设备更新和监管场景变化而逐渐老化。污染源在线自动监

测数据智能预警研究表明，系统效果依赖数据分析与业务反馈形成闭环。因此，智能识别建设必须同步规划样本标注、争议复盘、案例入库和模型迭代，让模型在真实应用中不断校正，而不是依赖固定样本一次训练后长期不变。机器学习模块输出预警结果时，还应同步提供关键特征贡献、相似案例索引和历史处置记录，帮助人工复核快速判断命中原因。解释信息足够充分后，复核人员就不必再从头翻找同类工况和过往记录，预警处置效率也会更高。

四、异常识别结果的复核与应用闭环

（一）人工复核是区分设备异常和排放异常的必要环节

无论规则还是算法，都只能给出“值得怀疑”的对象，而不能完全替代人工复核。真正的复核应包括设备状态检查、工况台账核对、维护记录回看和必要的现场比对监测。只有这样，系统识别出来的异常才能被分流为设备问题、数据问题或排放问题，不至于把所有报警都推向现场执法。人工复核并不是降低智能化程度，而是在关键节点上为异常判定增加事实支撑，使系统输出真正能进入管理动作。异常处置闭环不能止于标记完成，还要记录是否复测、是否到场、是否修正规则以及是否更新样本库，才能支持持续迭代。闭环记录越完整，后续做版本优化时越能看清哪一类动作真正降低了误判。

（二）按异常等级分流处置有助于降低无效核查成本

不是所有异常都需要同样强度的响应。突跳和超量程值适合快速远程核查，长期漂移适合安排维护比对，多变量逻辑失衡则更适合结合工况台账深度复盘。通过设置异常等级和分流处置路径，可以把有限核查资源优先投向风险更高、后果更大的异常对象。这样既能避免平台天天报警、现场无从下手，也能提高真正

高风险问题的发现率和处置时效。异常识别系统只有与分级处置机制相连，才不会陷入“发现很多、解决很少”的低效循环。数据治理回写阶段应比较误报率、漏报率和人工复核负荷变化，若只追求命中数量上升，往往会把系统推向低解释性的极端。只有三项指标一起观察，系统优化才不会只追求命中数量而忽略业务可用性^[5]。

（三）把识别结果回写平台规则和运维考核才能形成持续改进

异常识别工作最怕每次都从头开始。更有效的做法，是把每次复核后的结论回写到规则库、模型样本库和运维考核台账中，形成持续学习机制。某类设备反复出现同样异常，就应追溯其维护机制是否需要调整；某类规则频繁误报，就应修订其阈值和适用条件；某类企业总在同一工况下触发异常，就应把该工况列入重点核查场景。通过这种回写机制，异常识别系统会越来越贴近真实业务，而不是停留在静态预警工具层面。模型更新必须设置灰度验证和回溯评估，先检查新旧版本对典型异常样本的判定差异，再决定是否正式替换现有识别策略。灰度校验能提前暴露版本替换后的偏差扩散，避免整个平台上线后被动返修。

五、结语

生态环境在线监测数据异常识别的核心，不是把更多数据判成异常，而是更准确地区分设备异常、处理异常和排放异常。规则筛查、时序分析、多变量交叉和智能预警各有边界，只有与人工复核、分级处置和样本回写机制结合起来，异常识别体系才能持续稳定运行。异常识别做得越精确，在线监测数据在监管和治理中的可信度就越高。

参考文献

- [1] 全吉昌. 污染源在线自动监测数据的智能分析与预警系统研究 [J]. 资源与环境科学进展, 2024, 3(1): 12-15.
- [2] 张飞杰. 环境在线监测数据质量控制方法及其有效性 [J]. 低碳世界, 2026, 16(1): 28-30. DOI: 10.16844/j.cnki.cn10-1007/tk.2026.01.011.
- [3] 邓春娟, 杨震, 李亚杰, 等. 污染源排放数据质量影响因素分析及对策 [J]. 四川环境, 2022, 41(3): 162-168.
- [4] 郭鹏. 人工智能辅助下的 VOCs 排放源追踪与监测技术研究 [J]. 低碳世界, 2026, 16(1): 34-36. DOI: 10.16844/j.cnki.cn10-1007/tk.2026.01.005.
- [5] 申凌霄. 挥发性有机物在线监测与溯源技术研究 [J]. 皮革制作与环保科技, 2025, 6(18): 56-58. DOI: 10.20025/j.cnki.CN10-1679.2025-18-19.