

环境监测数据审核及异常数据分析

周亚武

中卫市生态环境监测站 宁夏 中卫 755000

摘要：文章聚焦环境监测数据审核，阐述其重要性，包括对环境科学研究、管理与决策的关键作用。详细介绍审核流程，涵盖多源数据融合初筛、动态校准溯源、时空关联性挖掘及智能补全校验。分析异常数据成因，如传感设备隐性失效、环境要素耦合干扰、数据传输噪声污染。探讨异常数据分析方法，如动态阈值自适应模型、生成对抗网络异常检测、知识图谱关联分析，为保障环境监测数据质量提供参考。

关键词：环境监测数据审核；异常数据成因；异常数据分析

引言

环境监测数据是洞察环境状况、制定环保策略的重要依据。在环境问题日益严峻的当下，其质量直接关系到环境管理的成效与决策的科学性。环境监测数据在采集、传输等过程中易受多种因素干扰，出现异常情况。对环境监测数据进行严格审核至关重要。有效的数据审核流程和科学的异常数据分析方法，能够及时发现并纠正数据问题，确保数据的准确性和可靠性，为环境保护工作提供坚实的数据支撑。

1 环境监测数据审核的重要性

环境监测数据作为环境科学研究、环境管理与决策的重要基础，其质量直接关系到对环境状况评估的准确性和有效性，而环境监测数据审核是保障数据质量的关键步骤，具有不可忽视的重要性。从环境科学研究角度看，高质量的监测数据是深入探究环境问题本质、揭示环境变化规律的基础。环境科学领域的研究往往需要长期、连续且准确的数据支持，以分析环境要素之间的相互关系和动态变化。若监测数据未经严格审核，存在误差或虚假信息，将导致研究结果偏离实际，误导研究方向，使科研人员难以准确把握环境问题的核心，阻碍环境科学理论的创新与发展。在环境管理方面，准确可靠的监测数据是制定科学合理环境管理措施的重要依据。环境管理部门需依据监测数据评估环境质量现状、预测环境变化趋势，确定污染控制目标、制定排放标准以及规划环境治理项目。若数据审核不严，出现数据失真情况，导致环境管理决策失误，无法有效解决环境问题，甚至引发新的环境风险。对于环境决策而言，环境监测数据审核同样至关重要。决策者需要基于准确的数据来权衡不同方案的利弊，做出最优决策。严格的环境监测数据审核能够为环境决策提供坚实的数据支撑，确保决策的科学性和合理性，推动环境与经济的协调发展。

2 环境监测数据审核流程

2.1 多源数据融合与初筛

现代环境监测体系通过部署大气、水体、土壤等多类型传感器，构建起庞大的异构数据采集网络。这些数据蕴含着复杂的物理关联性，需借助跨源数据校验机制实现异常初筛。以城市群空气质量监测网络为例，PM_{2.5}浓度与风速、湿度等气象参数间存在明确的扩散迁移规律。当静风环境下PM_{2.5}浓度出现骤降，或在高风速条件下浓度未呈现预期稀释效果，系统会依据预设的物理模型，自动标记此类违背传输规律的数据组合。通过动态阈值设定，如将PM_{2.5}浓度的常规波动系数界定在±30%，高效识别并剔除约20%的明显异常数据。该过程基于多维数据间的内在逻辑关系，借助算法快速过滤极端值与矛盾数据，明显降低后续审核压力，为精准数据处理奠定基础。

2.2 动态校准与溯源审核

（1）数据可靠性的维护依托实时校准与全流程溯源的联合机制。以水质监测为例，在线TOC分析仪通过构建与实验室离线检测数据的回归关系，达成测量值的动态修正。当在线监测数据偏离校准曲线阈值超15%时，设备自动触发自检程序，对传感器漂移、试剂失效等因素进行排查。（2）区块链技术在数据管理中具关键作用。其通过哈希加密算法处理采样设备编码、人员资质等核心信息，构建不可篡改的数据溯源体系，精准记录数据采集至传输全流程，保障全生命周期的完整性与真实性。数据上链后，非授权修改会被实时识别，杜绝造假。分布式存储特性使数据多节点备份，设备故障不影响数据完整性与可用性，为环境监测数据可信度筑牢基础。（3）这种双重审核模式意义重大。它既保障了数据采集的精确性，又实现了从样本采集到数据分析全流程的透明化管控。任何数据变更都能追溯到具体操作环

节,明显增强了数据的可信度与可验证性,为后续工作奠定坚实基础^[1]。

2.3 时空关联性深度挖掘

基于时空数据立方体的构建,可实现环境监测数据的多维一致性审核。以河流断面监测为例,沿程溶解氧浓度受水流扩散、生物耗氧等因素影响,呈现特定的衰减规律。环境监测数据审核中,时空关联性分析是捕捉异常的关键路径。当河流断面溶解氧数据较上下游站点骤降40%,且无水利工程扰动、突发污染等常规诱因时,监测系统会即刻启动异常预警程序。借助GIS空间分析技术,将土地利用类型数据与监测数据进行空间叠加,可深度剖析数据合理性。以氨氮浓度监测为例,在农业活动密集区域,因化肥流失、畜禽养殖废水排放等因素,氨氮浓度通常显著高于非农业区域。若监测数据与此规律不符,则需全面排查采样环节,包括点位布设科学性、采样频次合理性及器具清洁程度;同时校验仪器测量精度,检查传感器是否存在漂移、校准参数是否过期等问题。这种融合空间拓扑结构与环境要素分布特征的审核方法,精准识别数据空间变异异常,确保监测数据真实反映地理环境客观状况^[2]。

2.4 智能补全与完整性校验

(1)为应对环境监测数据缺失难题,采用时空插值与机器学习相融合的技术路径实现数据修复。其中,空间维度上借助克里金插值法,依据空间自相关原理完成缺失点位数据的初步估算;时间维度上运用LSTM神经网络,通过挖掘历史数据时间序列特征,对插值结果进行动态校验与预测修正,保障数据估算的可靠性。(2)在数据完整性评估层面,通过构建量化评价体系实现科学审核。该体系聚焦三大核心指标:监测频次达标率,即实际监测次数与计划次数的达标比例;项目覆盖度,评估监测指标是否完整覆盖环境要素;数据连续度,衡量监测数据在时间序列上的连贯程度。各指标均采用数值化评分,结合加权计算得出综合得分,直观反映数据完整性水平,为后续数据有效性判定提供可靠依据。(3)基于该评估体系设置预警阈值,当监测单元评分低于85分时,即刻触发加密监测机制,通过增加采样频率或增设监测点位等手段,确保环境监测数据在时间序列与空间分布上的完整连续,为后续环境数据分析与决策提供坚实的数据支撑。

3 异常数据成因分析

3.1 传感设备的隐性失效

环境监测传感器的隐性失效是异常数据的重要诱因,其中传感器漂移问题尤为突出。以某城市气象站为

例,长期暴露于户外环境的温湿度传感器会因粉尘附着、元件老化等因素产生非线性误差,在夏季高温高湿时段,其测量误差可达 $\pm 5^{\circ}\text{C}$,显著偏离实际环境参数。对传感器历史校准数据的统计分析发现,漂移速率与设备服役时长呈现二次函数关联,服役超过3年的传感器,其异常数据发生率达到新设备的2.3倍。这种隐性失效具有渐进性特征,初期误差较小易被忽视,但随着时间推移,会对数据准确性产生累积性影响。压力传感器的膜片形变、电化学传感器的电解液消耗等,也会导致测量值偏离真实环境指标,形成隐蔽性异常数据^[3]。

3.2 环境要素的耦合干扰

(1)环境要素的耦合干扰常导致监测数据呈现非线性异常。在近海水质监测场景中,潮汐与降雨的协同作用尤为显著。当潮差超过2米且降雨量突破50mm时,陆源径流携带的悬浮物与潮汐搅动的底泥剧烈混合,使得水体浊度数据波动幅度可达正常水平的3-5倍,异常发生概率激增67%。这种复合干扰打破了传统阈值的适用性,使基于历史数据的预警体系失去效能,凸显环境变量交互对数据准确性的复杂影响。(2)在大气监测领域,逆温层与静风条件的耦合作用同样不容忽视。二者相互配合会抑制污染物的扩散,使得 $\text{PM}_{2.5}$ 浓度出现区域性异常升高现象。而仅依据单一的风速或温度指标,难以对这类现象进行预判。(3)环境要素的时空耦合具有随机性,其引发的异常数据往往呈现出非典型分布特征。这种非典型分布增加了数据审核的难度,审核人员难以依据常规的数据审核方法和经验,准确识别和判断这类异常数据,给环境监测数据的准确性和可靠性带来挑战。

4 异常数据分析方法

4.1 动态阈值自适应模型

(1)鉴于环境监测数据呈现出强烈的时空动态特征,传统固定阈值难以满足复杂环境变化的监测需求,由此催生出动态阈值自适应模型。该模型借助滑动窗口技术,通过实时更新统计参数,能精确界定数据正常波动区间。以某工业园区废气排放监测为例,系统选取过去30天的小时浓度数据构建滑动窗口,持续计算均值与标准差,以此生成动态阈值区间。当实时监测数据超出均值 ± 2.5 倍标准差,且异常状态持续3个监测周期,即可判定为有效异常数据。此机制可有效剔除设备瞬时波动或环境短时变化引发的伪异常,相比固定阈值法,对生产负荷变化导致的浓度波动响应准确率提升35%。(2)动态阈值自适应模型的核心优势在于自适应性。在大气污染物监测场景中,季节更替会导致污染物背景浓度显著变化,固定阈值极易引发误判,而动态阈值模型能够

依据季节变化自动调整判定标准,始终维持对异常数据的高敏感度。(3)通过设置异常持续周期,该模型可有效规避数据随机波动造成的误报,保障异常判定的准确性。模型具备良好的参数优化能力,可灵活适配水环境、土壤监测等不同领域的应用需求,展现出广泛的适用性。

4.2 生成对抗网络(GAN)异常检测

生成对抗网络(GAN)作为深度学习领域的创新算法,在环境监测异常数据检测中展现出独特优势。GAN由生成器和判别器组成,通过两者的对抗博弈学习数据分布特征。在水质COD数据监测应用中,生成器通过学习历史正常数据,生成模拟的COD浓度序列,判别器则负责区分真实数据与生成数据。当真实数据被判别器误判为模拟数据的概率超过70%时,表明该数据偏离正常分布模式,被认定为异常数据。实际应用中,GAN模型展现出优异的检测性能。在某流域水质监测项目中,与传统Isolation Forest算法相比,GAN模型对异常数据的识别准确率达91.2%,明显高于Isolation Forest算法的82.7%。其核心优势在于能够自动学习数据的复杂分布特征,无需人工预设数据分布模型。GAN模型擅于处理高维非线性数据,化解环境监测多参数耦合异常难题,优化结构策略可增强其检测精准度与适用性。

4.3 知识图谱关联分析

(1)环境监测数据异常常与多元因素存在复杂耦合关系,知识图谱关联分析通过构建涵盖污染源、气象参数、监测设备工况等要素的语义网络,实现异常数据的深度溯源。该方法通过整合不同维度数据形成结构化知识网络,可有效揭示数据间的潜在关联。(2)知识图谱的核心价值在于其强大的语义推理能力,通过图算法深度挖掘环境要素间的因果关系链条。以雾霾成因解析为

例,该技术能够有效关联工业源排放强度、交通尾气特征及气象边界层条件等多维度数据。在此基础上,利用路径搜索算法,可精准识别对雾霾形成具有关键影响的核心因子,为环境异常数据的深度分析和科学治理提供重要支撑。(3)其动态扩展性具备明显优势,能持续纳入新监测数据与环境变量,对环境系统认知模型进行动态优化。面对复杂异常事件,知识图谱可实现多源信息的系统整合与逻辑推演,通过构建关联关系网络,挖掘异常数据背后的深层逻辑。这种技术手段为异常诊断提供科学依据,也能辅助制定精准的应对策略,在环境数据分析领域展现出独特的技术价值与应用潜力^[4]。

结束语

环境监测数据审核是保障数据质量、推动环境保护工作的重要环节。通过构建完善的数据审核流程,从多源数据融合初筛到智能补全校验,层层把关,可有效提升数据的准确性和完整性。同时,运用动态阈值自适应模型、生成对抗网络异常检测、知识图谱关联分析等先进方法,深入剖析异常数据成因,实现精准识别与溯源。优化数据审核分析技术,服务精准可靠环保监测

参考文献

- [1]汪威.环境监测大数据分析 with 生态保护决策支持[J].中文科技期刊数据库(全文版)自然科学,2025(3):141-144.
- [2]吕承凯.环境监测中异常数据识别与修复[J].中文科技期刊数据库(全文版)自然科学,2025(1):026-030.
- [3]周禹池,郭婷婷,张莲红,陈惠.环境监测数据综合分析 with 评价技术分析[J].皮革制作与环保科技,2025,6(2):169-171.
- [4]黄成,张杰.环境空气质量自动监测数据 with 异常数据相关性分析[J].皮革制作与环保科技,2025,6(2):9-11.