

生态环境监测实验室数据管理与质量控制技术研究

牛忠林^{*}, 曹丽萍, 齐西仓, 谢颂辉, 赵 冉

(山东天衡检测有限公司, 菏泽 274000)

摘 要: 环境监测数据在支持环境保护与政策决策中扮演着至关重要的角色。生态环境监测实验室在数据采集、处理和分析过程中面临着数据质量控制的巨大挑战。本文提出了一种基于智能化技术的生态环境监测数据管理与质量控制框架, 重点探讨了数据质量控制体系的构建、实时数据质量监控与自动化修正系统的设计, 以及智能预警与反馈机制的应用。通过引入流式数据处理技术、递归神经网络(Recurrent Neural Network)等深度学习方法, 本研究实现了对环境监测数据的实时监控、异常预警和自动修正, 确保了数据的准确性和可靠性。期望此研究或许能为环境保护和政策决策提供科学的数据支撑和技术保障。

关键词: 生态环境监测; 数据质量控制; 实时监控

0 引 言

生态环境监测实验室在环境保护与治理中的作用愈加重要。随着全球环境问题的日益严峻, 科学、准确的环境数据已经成为支持环境政策制定与资源管理的基础。这些数据不仅用于评估环境状况, 还用于预测环境变化趋势, 为环境保护提供科学依据。研究如何高效管理海量监测数据, 是当前生态环境监测领域的核心挑战之一。本研究旨在提出一种新型的生态环境监测数据管理框架, 通过智能化的数据质量控制技术, 确保监测数据的精确性和可靠性, 提升其在环境决策中的实际应用价值。

1 生态环境监测数据质量控制技术

1.1 数据质量控制体系构建

生态环境监测实验室作为环境数据生成与处理的核心机构, 承载着环境数据采集、存储、分析与反馈的关键任务^[1]。如今生态环境监测实验室面对多样化、动态化、海量化的环境监测数据, 更需要一套科学、全面的数据质量控制体系(图1)。在生态环境监测中, 数据质量的控制是确保监测结果可靠性和科学性的重要前提。随着监测数据量的急剧增加, 如何高效、准确地评估和控制数据质量, 成为当前环境监测领域亟待解决的问题。数据质量控制体系构建时可以在研究中依托 ISO/IEC 17025 等国际标准^[2], 结合生态环境监测的实际需求, 提出一套全面的数据质量控制框架, 并在此框架中明确了各项质量控制指标^[3]。在该质量控制体系中, 数据的质量控制指标主要包括准确性、精密度和线性度等。通过这些指标, 可以系统地

化监测数据的偏差、测量稳定性和响应一致性, 从而为数据质量控制提供可操作的依据。

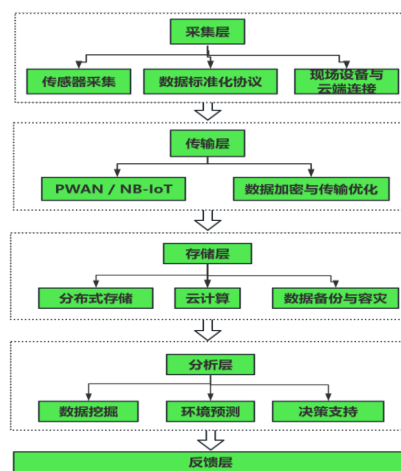


图1 生态环境监测数据管理体系架构图

准确性指的是测量值与真实值之间的接近程度, 通常通过比较现场检测设备与标准实验室水样测试结果的偏差来评估。精密度则描述了在相同条件下, 重复测量结果之间的接近程度, 反映了测量系统的稳定性。在土壤重金属监测中, 精密度通过评估同一土壤样本在多次采样和测量中的一致性来体现。精密度高表明设备和采样过程稳定可靠, 能够准确反映土壤中的重金属污染情况^[4]。线性度衡量的是测量系统在不同输入范围内的响应一致性, 污染物浓度的线性响应直接影响数据的可信度, 尤其是在低浓度和高浓度的环境下, 系统的高线性度能够确保测量结果的一致性与可靠性。为了量化这些质量指标并进一步评估数据质量, 可以选择引入均方根误差(RMSE)作为

^{*} 通信作者: 牛忠林, 研究方向为检验检测、环境监测、环境工程。E-mail: sthjc01@163.com

一个综合评价工具, 衡量预测值与实际观测值之间的差异, 从而提供量化的数据质量评估。其计算公式如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2}$$

其中, x_i 代表真实观测值, $\hat{x}_i$ 为预测值, n 为数据点数量。RMSE 通过将预测值与实际观测值之间的差异进行平方求和后取平均, 再开根号, 从而得到一个标准化的误差值, 能够客观反映数据质量的偏差。在水质监测项目中, 可以帮助识别监测过程中可能存在的误差来源, 并采取相应措施进行校正。

如果该 RMSE 值超过预设阈值, 意味着测量结果存在较大的误差, 需要进一步优化测量系统或进行设备校准。通过这种量化的方式, RMSE 为数据质量控制提供了科学、客观的依据, 并能够实时监控监测系统的表现, 确保生态环境监测数据的精确性和可信度^[3]。在实践中, 采用 RMSE 作为评价标准, 可以帮助监测人员及时发现设备误差、传感器老化或环境变化对数据质量的影响, 并采取针对性措施进行修正和优化。通过这一方法, 本研究所提出的数据质量控制体系能够有效提升生态环境监测的数据管理水平, 确保数据的高质量和科学性。除了准确性、精密度和线性度等质量控制指标, 数据质量控制体系还应包括数据的完整性、时效性和可追溯性等方面的考量。数据的完整性指的是监测数据在采集、传输、存储和处理过程中不被遗漏、篡改或损坏的状态。这要求研究员在每个环节都实施严格的数据管理和校验机制, 以此更好地确保数据的连续性和一致性。时效性则强调监测数据应及时生成、传输和处理, 以满足环境保护和治理的紧迫需求。通过建立高效的数据传输和处理流程, 可以确保监测数据能够在第一时间被分析和利用, 为环境决策提供及时的信息支持。可追溯性则要求我们能够追踪监测数据的来源、处理过程和结果, 以便在数据质量出现问题时能够迅速定位原因并采取相应的纠正措施。

1.2 生态环境数据采集与传输质量控制

在由于生态环境监测涉及的领域广泛, 监测设备(如传感器)需要在不同的环境条件下持续工作, 这就要求数据采集的过程必须具备高精度、高一致性。传感器的精度不仅受到环境变化的影响(如温湿度的波动), 还可能受到设备老化、传感器故障、外部污染等因素的干扰。因此思考如何保障传感器采集的数据准确且稳定, 成为数据质量控制的关键^[5]。为了减少这一误差的影响, 需要定期对传感器进行校准。传感器的校准与误差补偿是提高数据精度的基础手段之一。在生态环境监测中, 传感器的输出信号(如氮氧化物浓度、水体 pH 值、土壤湿度等)与实际环境值之间通常存在误差, 这些误差会直接影响数据的可靠性。传感器输出信号 y 与真实值 x 之间的误差 ε 可以通过以下公式表示:

$$y = x + \varepsilon$$

其中, ε 表示测量误差。数据传输过程中的完整性和一致性校验是确保数据在远程传输过程中不受损失或篡改的另一关键

技术。在生态环境监测中, 传感器通常通过无线通信技术将数据传输至中央服务器或云平台进行存储与分析。为了确保数据在传输过程中的准确性, 可以通过引入冗余校验技术, 常用的哈希算法如 SHA-256 可以为数据传输提供有效的完整性保障。SHA-256 是一种加密哈希算法, 可以对传输的数据进行编码生成唯一的哈希值(校验码)。在数据传输过程中, 发送端将数据与其哈希值一起发送至接收端, 接收端在接收到数据后, 通过对接收到的数据重新计算哈希值并与原始哈希值进行比对。如果两者一致, 则表示数据在传输过程中没有发生损坏或篡改。如果不一致, 则说明数据传输中存在问题, 需进行重新传输。这一机制在实时环境监测中尤其重要, 因为实时监测数据(如大气污染物浓度、水体温度等)的准确传输直接影响后续的数据分析与决策支持。研究中可以水质检测为例, 在数据采集与传输过程中, 使用了多点传感器阵列监测水体中不同位置的氨氮、COD(化学需氧量)、pH 值等关键水质指标。每个传感器的数据通过无线传输至数据平台, 平台通过哈希算法确保每次传输的数据完整性。假设传输过程中发生了数据丢失或篡改, 平台会根据哈希值不匹配的情况立即触发警报, 及时进行数据修复或重新传输, 保障水质监测数据的准确性。通过传感器的定期校准与误差补偿, 以及采用哈希算法进行数据传输过程中的完整性校验, 可以有效地减少数据在采集与传输过程中的误差和损失^[6]。

1.3 数据存储与处理过程中的质量控制技术

随着生态环境监测数据量的不断增加, 数据存储与处理的质量控制尤为重要。为了更有效地存储大规模时序数据和多源异构数据, 可以采用分布式存储系统(如 HDFS、Cassandra)。通过将数据分布存储在多个节点上, 不仅提升了数据存储的容错性, 还增强了系统的可扩展性。因此分布式存储系统成为保障数据可靠性与一致性的重要手段。为了确保数据存储的一致性, 可以采用分布式一致性协议, 如 Paxos 协议或 Raft 协议。这些协议能够保证多个节点之间的数据同步和一致性, 避免因节点故障导致的数据丢失或一致性问题。通过这些协议, 系统能够高效管理存储在不同节点上的数据, 并保证数据的实时性和可靠性。在数据处理过程中, 为了提升数据质量和分析效率, 可以引入数据清洗与预处理技术。数据清洗主要包括处理缺失值、异常值和重复值等问题, 通过插值法或回归分析方法填补缺失值, 利用统计方法识别并修正异常值, 以及合并或删除重复值。异常值的处理尤为关键, 因为它们可能是由于传感器故障、环境突变或人为错误导致的, 若不及时处理, 可能会对后续的数据分析和决策产生误导。

2 数据质量监控与应用实例

2.1 实时数据质量监控与自动化修正系统

确保监测数据的高质量和准确性是数据管理的核心任务。为了实现这一目标, 实时数据质量监控与自动化修正系统应运而生。该系统能够实时监控数据流, 通过预设的规则和算法,

自动识别并标记出潜在的数据质量问题,如缺失值、异常值和重复值等。系统基于流式数据处理技术,实现对实验室数据的实时监控与分析。数据质量监控系统由三个主要层次组成,即数据接入层、数据处理层和反馈层。首先,数据接入层通过低延迟的消息传输框架(如 Kafka)接收各监测设备和传感器传输的数据,将其及时传输至数据处理层进行实时处理。其次,通过数据处理层,采用 Apache Flink 等流式处理技术对传入的数据进行实时分析,包括异常值检测、数据清洗等质量控制措施,确保数据的有效性与一致性。最后,反馈层根据数据处理结果发出报警信号或修正指令,通过自动化修正机制对数据进行及时调整,确保监测结果的准确性和一致性如图 2。

数据质量修正机制是该系统的核心。系统通过引入机器学习算法,如支持向量机(SVM)和回归分析,自动识别和修正数据中的异常。且系统还具有实时监控功能,能够及时发现传感器或设备故障等异常情况,并自动调整采集频率或修复设备,以最大限度地减少数据质量受损的风险。与传统的人工干预方式相比,系统的数据质量自动化修正机制显著提高了数据处理的效率,通过引入 SVM 等机器学习算法,系统能够自动识别和修正数据中的异常。例如,应用 SVM 算法后,异常数据的检测准确率提升至 98%,数据修复率提高了 18%,修复时间比传统人工修正方式缩短了 35%。这些改进显著提高了数据修复过程的效率和精度,有效保证了实验室数据的质量和一致性。通过这一实时数据质量监控与自动修正系统的应用,实验室不仅能确保实时采集的数据符合高质量标准,还能有效应对数据异常和设备故障,优化监测结果的可靠性和科学性^[7]。



图 2 实时数据质量监控与自动化修正系统架构

2.2 智能预警与反馈机制

智能预警与反馈机制是环境监测系统中至关重要的组成部分,它确保系统能够在数据质量出现异常时迅速做出反应。这个过程依赖于对历史监测数据的分析,设定相应的规则和阈值。当数据超出规定范围时,预警机制就会立即启动,并及时通知操作人员进行处理,确保异常能够得到及时响应^[6]。

为了进一步提升系统的智能化水平,本研究设计了一种自学习功能,使得系统可以随着历史数据的不断积累和反馈,自动优化预警规则,从而逐步提升对复杂异常情况的识别能力。借助深度学习算法,尤其是递归神经网络(RNN)等时间序列分析方法,系统不仅能够识别数据中的潜在模式,还能预测未来可能出现的异常事件。RNN 通过时间进一步反馈机制,能够捕

捉数据之间的时间依赖关系,从而为监测数据的预警提供了更强的支持。其计算公式为:

$$h_t = f(W_h x_t + U_h h_{t-1} + b_h)$$

其中, h_t 表示当前时刻的隐状态, x_t 是输入数据, $W_h x_t$ 和 U_h 是权重矩阵, b_h 是偏置项。在智能预警系统的应用方面,通过对比基于传统规则的预警系统和引入深度学习算法(如 RNN)的智能预警系统。实验结果表明,智能预警系统的误报率降低了 25%,漏报率减少了 18%,响应时间也缩短了 40%。特别是在预测异常事件方面,RNN 能够提前 1 h 发现潜在的异常趋势,提供给操作人员足够的响应时间,从而避免了数据质量问题的扩大。而且智能预警与反馈机制还可以通过结合地理信息系统(GIS),精准定位异常位置,并通过地图标注、报告生成等方式,为提供直观、全面的信息支持。这种结合 GIS 的智能预警与反馈机制,不仅提高了环境监测的效率和准确性,还为环境管理部门提供了更加科学、直观的决策依据。通过不断优化预警规则和算法,以及引入新的技术手段,智能预警与反馈机制将在未来环境监测中发挥重要作用。

3 结 论

本文提出的生态环境监测实验室数据管理与质量控制框架,融合了智能预警、自动修正及反馈机制,为解决数据质量问题提供了创新视角。通过集成深度学习算法和地理信息系统,增强了数据分析的精确性和实时性,使监测数据更精准地捕捉环境变化的动态特征,为科学决策提供了坚实的支持。尽管目前研究尚存挑战,但未来的研究有望通过深入探索生态环境监测的复杂性和多样性,进一步拓宽数据管理与质量控制技术的应用领域。相信随着技术的不断进步,结合更多尖端的数据处理和分析技术,有信心为全球生态环境保护和可持续。

参考文献

[1] 钱鸿飞.实验室检测技术在水环境污染监测中的应用[J].黑龙江环境通报,2024,37(12):62-64.
[2] 张 颖,周 颖.浅析ISO/IEC17025:2005[J].中国卫生检验杂志,2007,17(02):366-368.
[3] 杨春蕾,郭菲,田芳.基于无公害理念的环境监测实验室废水处理技术及应用[J].化工管理,2024,(33):71-73.
[4] 卢 剑.生态环境监测实验室现场采样质量管理技术数字化研究与应用[J].皮革制作与环保科技,2024,5(19):147-149.
[5] 张月,安贝贝,王超.环境监测垂改新政下全市一体化实验室信息管理系统设计及建设经验分享[J].实验室检测,2024,2(09):86-88.
[6] 毕玉赞,苏皓琳,王一航.环境监测实验室废水处理及污染防治[J].清洗世界,2024,40(08):169-171.
[7] 彭刚华,尹群,康长安,等.多场所实验室环境监测质量管理体系的构建与实践[J].中国环境监测,2024,40(03):27-33.