

数据驱动下采样技术智能化革新与新污染物精准采样策略研究

李磊

四川中衡检测技术有限公司 四川 德阳 618000

摘要: 新污染物有痕量存在、毒性隐蔽、迁移性强等特点,传统经验采样模式难满足精准监测需求。本文以数据驱动为核心,探索采样技术智能化革新路径,构建精准采样体系。通过整合多源数据搭建全链条智能框架,研发机器学习点位优化、深度学习频率自适应算法。结合新污染物特性与环境行为制定精准策略,配套全流程质控措施。研究实现采样技术从“经验驱动”到“数据驱动”的转变,提升检出率与数据可靠性,为环境风险评估管控提供支撑。

关键词: 数据驱动; 采样技术; 智能化革新; 新污染物; 精准采样策略

引言: 在环境问题日益严峻的当下,新污染物因其独特性质给环境监测带来巨大挑战。传统经验采样模式在应对新污染物时,暴露出诸多不足,难以满足精准监测需求。随着数据时代的到来,数据驱动理念为解决这一问题提供新思路。立足数据驱动理念,探索采样技术智能化路径与适配策略,对提升新污染物监测效能具有重要现实意义。

1 数据驱动的采样技术智能化理论基础

1.1 数据驱动的基本概念

数据驱动是以各类结构化与非结构化数据为核心输入,通过统计分析、机器学习等技术挖掘数据关联规律,进而支撑决策优化的方法论。在环境采样领域,其核心逻辑是打破传统采样的主观局限性,依托海量监测数据构建量化模型。这些数据涵盖新污染物本底浓度、环境介质参数、气象水文动态等多维度信息,通过数据质量评估(完整性 $\geq 95\%$ 、误差 $\leq 5\%$)与标准化处理,转化为可直接应用的特征变量。数据驱动强调“数据-模型-应用”的闭环联动,通过实时数据更新持续优化模型输出,使采样决策能够动态适配新污染物的时空分布变化,为采样技术智能化提供核心逻辑支撑,区别于依赖个人经验的传统模式。

1.2 智能化技术在采样中的应用潜力

智能化技术融合传感器集成、自动控制与数据处理,推动采样技术升级。传感器微型高精度化,能实时采集温度等参数,精度达微克级,满足新污染物痕量监测。自动控制技术可驱动采样设备完成深度调节等操作,响应时间 ≤ 10 秒,减少人工误差。机器学习与深度学习算法能挖掘数据隐含规律,如用历史数据训练模型预判污染物迁移路径^[1]。这些技术的融合应用,能够解决

传统采样中效率低、代表性差、时空覆盖不足等问题,使采样过程具备自适应、自优化能力,为实现新污染物全生命周期监测提供技术可能。

2 数据驱动的采样技术智能化框架构建

2.1 总体框架设计

数据驱动的采样技术智能化框架采用分层架构设计,呈现“纵向贯通、横向协同”的特点,整体分为数据采集层、数据传输层、数据处理与分析层、决策支持层四个核心层级,各层级通过标准化接口实现数据无缝流转。框架以“精准捕捉新污染物时空分布”为目标,建立“数据输入-模型分析-决策输出-执行反馈”的闭环机制。底层数据采集为上层分析提供基础支撑,上层决策结果反向指导采样设备执行,形成动态优化的良性循环。框架充分考虑不同环境介质(水、土壤、大气)的差异性,预留模块扩展接口,可适配不同类型新污染物的监测需求,具备较强的通用性与灵活性。

2.2 数据采集层设计

数据采集层采用“固定站点+移动设备+卫星遥感”的立体采集模式,确保数据来源的全面性与代表性。固定站点部署多参数传感器阵列,同步采集环境介质温度、湿度、污染物浓度等核心参数,采样间隔可设置为1-60分钟,满足不同监测精度需求。移动设备包括搭载便携式传感器的无人机与监测车,针对复杂地形区域开展补充采样,提升空间覆盖密度。卫星遥感数据用于获取大范围气象水文与地表覆盖信息,实现宏观尺度数据支撑。采集数据类型涵盖新污染物理化参数、环境因子动态、设备运行状态等,通过数据标签化处理(标注采集时间、位置、设备编号),为后续分析奠定基础^[2]。

2.3 数据传输层设计

数据传输层构建“5G+物联网+边缘计算”的混合传输模式,保障数据传输的实时性与稳定性。对于实时性要求高的采样设备状态、污染物浓度等数据,通过5G网络实现毫秒级传输,确保决策指令及时下达。对于海量历史数据与卫星遥感影像,采用物联网技术进行批量传输,降低传输成本。边缘计算节点部署在采样现场附近,对采集数据进行初步过滤与压缩,剔除无效数据(如传感器故障数据),减少传输带宽占用。传输过程采用端到端加密技术,通过数据加密算法与访问权限控制,防止数据泄露与篡改,保障监测数据的安全性与完整性,符合环境监测数据管理规范。

2.4 数据处理与分析层设计

数据处理与分析层是框架的核心,承担数据清洗、特征提取与模型运算功能。数据清洗阶段采用 3σ 准则剔除异常值,通过线性插值与随机森林算法填补缺失数据,确保数据完整性 $\geq 98\%$ 。特征提取模块通过主成分分析,从多维度数据中筛选出与新污染物分布强相关的关键因子(如水温、流速、光照强度),降低模型运算复杂度。分析层集成机器学习与深度学习算法库,包括用于点位优化的K-means聚类算法、用于频率预测的LSTM神经网络等,通过GPU加速技术提升模型运算效率。同时构建数据可视化平台,以热力图、时序曲线等形式直观呈现分析结果,为决策提供清晰参考。

2.5 决策支持层设计

决策支持层基于数据处理结果输出精准采样指令,实现“数据-决策-执行”的无缝衔接。该层核心为智能决策引擎,内置新污染物监测知识库,涵盖不同类型污染物的采样技术规范与环境适配参数。决策引擎通过匹配实时数据与知识库规则,输出采样点位坐标、采样深度、采样频率等具体参数,例如针对水溶性抗生素,自动推荐0.5-1米水层采样。同时开发人机交互界面,支持工作人员对决策结果进行人工校验与调整,兼顾智能化与人性化需求。决策结果通过标准化协议传输至采样设备控制系统,驱动设备自动执行采样操作,并实时反馈执行状态,形成决策闭环。

3 智能采样算法研发与应用

3.1 基于机器学习的采样点位优化算法

基于机器学习的采样点位优化算法以K-means聚类与支持向量机为核心,实现点位布设的精准化。算法首先对历史监测数据进行聚类分析,将监测区域划分为若干污染特征相似的子区域,确保每个子区域布设代表性点位。随后通过支持向量机模型建立污染物浓度与环境因子的关联模型,识别高污染潜在区域,在该区域加密

布设点位。算法引入空间权重因子,对交通便利度、地形复杂度等因素进行量化评估,避免点位布设在难以到达的区域。应用该算法在某流域进行试点,相比传统均匀布点法,高浓度区域检出率提升32%,点位数量减少25%,显著提升采样效率与代表性^[1]。

3.2 基于深度学习的采样频率自适应算法

基于深度学习的采样频率自适应算法采用LSTM神经网络构建,实现采样频率的动态调整。算法以近5年的污染物浓度时序数据、气象水文数据为训练集,将数据划分为平稳期、波动期、高浓度期三种场景,分别对应不同采样频率。通过模型学习污染物浓度变化规律,当预测浓度波动幅度 $\leq 10\%$ 时,设置采样频率为每周1次;波动幅度10%-30%时,调整为每3天1次;波动幅度 $> 30\%$ 或浓度超标时,启动每日采样模式。算法具备实时更新能力,通过新采集数据持续优化模型参数,预测准确率稳定在85%以上。

4 新污染物精准采样策略制定

4.1 新污染物的特性分析

对新污染物特性展开深入分析,主要聚焦于理化性质与环境行为这两个关键维度,以此为精准采样策略的制定提供坚实的科学依据。在理化性质层面,着重剖析溶解度、挥发性、稳定性等关键参数。以微塑料为例,其密度差异显著,范围在0.9-1.4g/cm³之间,这种特性使得微塑料在水体中并非均匀分布,而是依据密度差异分布于不同水层,为准确采集不同水层的微塑料样品,采样策略需充分考虑这一特性。抗生素大多具备良好的水溶性,这使得它们极易随地表径流发生迁移,在制定采样策略时,就要关注其可能迁移到的区域;在环境行为方面,深入研究污染物在水、土壤、大气等环境介质中的迁移转化规律,同时明确温度、降水等环境因子对其产生的影响。例如,在高温条件下,农药类新污染物的降解速率会显著加快,这就要求在采样后需尽快进行检测分析,缩短保存时间。通过构建涵盖微塑料、PPCPs、抗生素等典型类别的新污染物特性数据库,系统梳理各类污染物的关键特征,为制定差异化的采样策略筑牢根基。

4.2 基于新污染物特性的采样方法选择

基于新污染物特性进行采样方法选择,严格遵循“精准捕捉、减少损耗”原则,力求实现采样方法与污染物的精准匹配。对于水体中的微塑料,考虑到其粒径大小不一,采用筛网过滤与密度分离结合法。先选用不同孔径的筛网,如20 μm 和100 μm ,对水体进行初步过滤,分离出不同粒径的微塑料,有效避免漏采情况。对于土壤中的半挥发性有机物,固相萃取法是理想选择。

通过精心优化萃取剂配比，例如采用正己烷-丙酮混合液，能显著提升提取效率。大气中挥发性新污染物，吸附管采样法较为适用，选用Tenax-TA吸附剂，可确保吸附效率达到90%及以上。针对痕量污染物，在线富集技术能发挥重要作用，可将样品浓缩100-1000倍，极大提升检出能力^[4]。同时，为每种方法明确适用范围与操作规范，并形成方法选择对照表，为现场采样人员提供直接、清晰的指导。

4.3 采样时间与频率的精准确定

采样时间与频率的精准确定，需紧密结合污染物特性与环境动态，以实现“捕捉关键时段、避免干扰因素”的目标。在时间选择上，针对农业面源排放的新污染物，由于灌溉后24小时内污染物浓度处于峰值，此时采样能获得最具代表性的数据。对于大气污染物，选择无风（风速 $\leq 1.5\text{m/s}$ ）、无降水时段采样，可有效减少气流扰动对采样结果的影响。频率确定采用“基础频率+动态调整”模式，在常规监测中，微塑料每月采样1次，抗生素每两周采样1次。当遇到气象条件突变，如暴雨、极端高温，或者模型预测浓度异常时，立即启动应急采样，将频率提升至每日1次。此外，通过整合短期气象预报与浓度预测结果，提前3天确定采样计划，确保采样时间精准无误，为准确监测污染物浓度变化提供有力保障。

4.4 采样点位的精准布设

采样点位的精准布设采用“全域覆盖、重点加密”的空间布局策略，充分结合智能算法与人工校验来确定最终点位。在流域监测中，沿水流方向科学布设控制断面、消减断面，同时在入河排污口下游500米处增设重点点位，以便精准捕捉污染物排放峰值。土壤采样采用“梅花形布点法”，在每1000平方米区域内合理布设5个点位，涵盖区域中心与边缘，确保所采集样品具有代表性。大气采样时，在工业区上风向、下风向以及周边敏感点（如居民区）均布设点位，全面分析污染物扩散范围。点位确定后，通过高精度的GPS定位（精度 $\leq 5\text{米}$ ）记录坐标，并建立详细的点位信息档案，档案内容包含周边环境特征、布设依据等，为后续监测工作提供可靠的溯源支撑，保证采样点位布设的科学性和合理性。

4.5 采样质量控制与保障措施

采样质量控制与保障措施贯穿采样全流程，是确保监测数据可靠的关键。在设备质控方面，采样前对传感器与分析仪器进行严格校准，将误差控制在2%以内；采样容器需经过酸液浸泡、超纯水清洗等处理，保证空白样污染率 $\leq 3\%$ 。操作质控方面，制定标准化作业流程，明确样品采集、封装、保存条件，例如有机污染物样品采用棕色瓶避光保存，并在4℃环境下冷藏运输，防止样品变质。数据质控方面，采用平行样采集（每批样品平行样比例 $\geq 10\%$ ）与标准物质比对的方法，确保数据相对偏差 $\leq 10\%$ 。同时，建立完善的质量追溯体系，为每个样品赋予唯一编码，详细记录采样人员、设备、时间等信息，实现从采样到分析的全流程追溯，严格符合环境监测质量规范，为监测数据的准确性和可靠性提供坚实保障。

结束语

本文围绕数据驱动与新污染物精准采样核心命题，系统构建采样技术智能化框架，研发两类核心智能算法，制定全流程精准采样策略。研究明确数据驱动在打破传统采样局限、提升监测精度中的核心价值，通过分层框架实现多源数据的高效利用，算法应用使采样效率与代表性显著提升，精准策略为不同类型新污染物采样提供明确指导。未来需进一步拓展新污染物类型覆盖范围，融合卫星遥感与无人机技术提升采样全域性，推动研究成果转化为行业标准，为新污染物风险管控提供更坚实的技术保障。

参考文献

- [1]陈艳,胡玉玲.城市区域大气颗粒物解析与污染防治策略[J].清洗世界,2023,39(03):140-142.
- [2]杨槟豪,舒洲,辛浩鑫.广安市大气环境细颗粒物(PM_{2.5})污染特征研究[J].山西化工,2023,43(02):200-202.
- [3]苏轩,刘聪,张帅鹏,等.面向日志完备性的事件日志采样方法[J].计算机集成制造系统,2022,28(10):3156-3165.
- [4]王同林.污水即时采样能否作为判定排污行为超标的依据[J].中国环境监察,2022(005):84-85.