

基于多源数据融合的大气环境智慧监管平台建设研究

裴浩言¹, 陈建军²

¹湖南有色金属职业技术学院资源环境系, 湖南 株洲

²郴州市生态环境局, 湖南 郴州

收稿日期: 2025年5月20日; 录用日期: 2025年6月18日; 发布日期: 2025年6月27日

摘要

本研究构建了基于“真、准、全、快、新”五维数据质量保障体系的大气环境智慧监管平台, 通过物联网技术整合环境质量监测网、污染源监控网等12类异构数据(日均处理量10 TB), 建立多源数据中心与AI算法中心双核驱动架构。平台创新性实现“感知预测-精准溯源-动态管控-效果评估”闭环管理; 采用LSTM模型提升污染态势预判能力, 结合PMF源解析技术识别重点污染源; 依托三级垂直管理机制与“智慧调度一张图”, 实现指令10分钟级穿透与跨部门协同, 为破解“精准治污、科学治污、系统治污”的实践难题提供技术支撑。

关键词

多源数据融合, 智慧监管平台, 污染动态溯源, 闭环管理

Research on Constructing a Smart Supervision Platform for Atmospheric Environment Using Multi-Source Data Fusion

Haoyan Pei¹, Jianjun Chen²

¹Resource and Environment Department of Hunan Non-Ferrous Metals Vocational and Technical College, Zhuzhou Hunan

²Chenzhou Ecological Environment Bureau, Chenzhou Hunan

Received: May 20th, 2025; accepted: Jun. 18th, 2025; published: Jun. 27th, 2025

Abstract

This study developed an intelligent atmospheric environment supervision platform based on a five-dimensional data quality assurance framework (“Truthfulness, Accuracy, Completeness, Timeliness, and Novelty”). The platform integrates heterogeneous data from 12 sources, including environmental quality monitoring networks and pollution source monitoring networks, via Internet of Things (IoT) technology, processing an average of 10 terabytes (TB) of data daily. It establishes a dual-core-driven architecture comprising a Multi-source Data Center and an AI Algorithm Center. The platform innovatively achieves closed-loop management encompassing “perception & prediction—precise source identification—dynamic control—effectiveness evaluation”. Specifically, it employs Long Short-Term Memory (LSTM) models to enhance pollution trend forecasting capabilities and combines Positive Matrix Factorization (PMF) source apportionment techniques to identify key pollution sources. Leveraging a three-tier vertical management structure and a “Smart Dispatch Map”, the platform enables directive penetration within 10 minutes and facilitates cross-departmental coordination. This system provides robust technical support for addressing the practical challenges of implementing “targeted, scientific, and systematic pollution control”.

Keywords

Multi-Source Data Fusion, Intelligent Supervision Platform, Dynamic Pollution Source Tracing, Closed-Loop Management

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着大气污染治理向精细化转型,传统监管模式面临三大瓶颈:① 数据维度割裂,环境质量、污染源、气象等多源数据分散于不同系统,难以支撑协同分析[1]-[3];② 溯源精准不足,人工研判依赖经验,难以解析复杂传输过程,如以区域传输为主的污染情景[4];③ 管理闭环缺失,管控指令与效果评估脱节,导致减排措施科学性不足[1]-[3] [5]。

针对上述问题,本研究提出“数据融合驱动智能决策”路径,构建大气环境智慧监管平台。核心创新在于:① 架构设计,基于五维数据质量保障体系,集成天空地一体化监测网络(卫星遥感、超级站、热点网格)、污染源动态数据(5000 余家企业在线监控)及跨部门数据(交通、气象),通过冷热双轨存储(InfluxDB + HDFS)与区块链存证实现高效治理;② 技术突破, AI 算法中心融合 LSTM 时序预测模型(输入 7 天滑动窗口数据, MAE < 10 μg/m³)与 PMF 源解析技术(解析工业源、移动源等 5 类贡献率),攻克动态溯源难题;③ 机制创新,建立“市-区-街道”三级垂直闭环管理,通过可视化分析中心生成“靶向管控”指令(如基于用电关联分析触发偷排预警)。

2. 多源数据融合平台构建

2.1. 平台架构设计

本研究基于“真、准、全、快、新”五维数据质量保障体系,构建了大气多源数据融合智能平台。采用物联网技术集成环境质量网、污染源监控网的各类数据,以空气质量模型、溯源模型、机器深度学习

和数据融合挖掘技术为途径, 实现“感知预测 - 精准溯源 - 动态管控 - 效果评估”闭环管理, 为环境管理的“精准治污、科学治污、依法治污、系统治污”提供技术支撑(见图 1)。

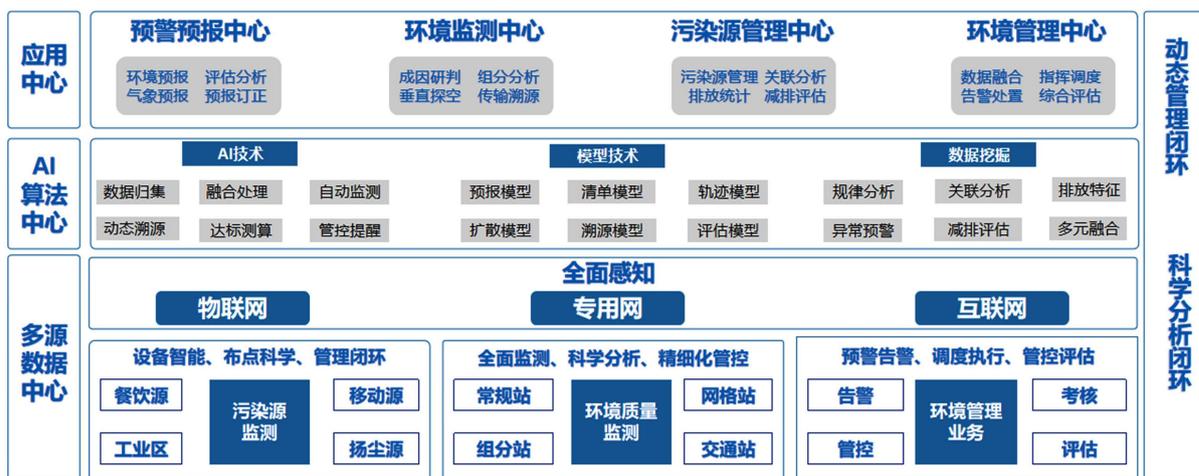


Figure 1. Intelligent Multi-Source Data Fusion Platform Architecture for Atmospheric Applications
图 1. 大气多源数据融合智能平台架构

AI 算法中心作为核心引擎, 采用机器学习(如 LSTM、随机森林)与深度学习模型(CNN、图神经网络), 结合空气质量数值模型(如 CMAQ、WRF-Chem), 实现三项关键技术突破: 一是感知预测, 基于环境预报与气象预报模块, 实现 72 小时高精度污染态势模拟; 二是精准溯源, 通过动态溯源、需量探空技术解析污染传输路径, 识别重点贡献源(如工业区排放与移动源); 三是动态管控, 利用达标测算与排放特征分析生成差异化减排方案, 通过管控提醒模块推送至管理部门。

预警预报中心与环境管理中心形成决策闭环。预警模块基于阈值规则与概率模型触发告警, 指挥调度模块生成“一源一策”管控指令(如企业限产、道路降尘), 效果评估环节则依托减排评估、综合评估模块量化措施成效, 结合考核机制实现“动态复阅环”(即动态复核 - 反馈优化)。

平台创新性在于将传统环境管理的“单向流程”重构为“智能闭环”: 以数据融合挖掘驱动污染源管理(排放统计、关联分析)、以 AI 替代人工研判(典型研判、重点分析)、以动态评估支撑科学决策。其技术架构有效响应了“精准治污”(靶向溯源)、“科学治污”(模型仿真)、“系统治污”(多源协同)的核心诉求, 为区域大气污染防治提供可验证的数字化范式。

2.2. 多源数据设计

多源数据融合中心的构建是实现环境治理现代化的重要技术基础。其核心架构包括数据采集、清洗、存储、处理与共享五大模块, 通过整合环境监测感知数据(如空气质量传感器、卫星遥感数据)、生态环境局业务数据(如污染排放记录)、跨部门协同数据(如气象、交通信息)及第三方数据(如企业自测报告), 形成标准化、高可信度的数据资源池。构建过程中需依托分布式存储技术与云计算平台, 实现海量异构数据的高效管理; 同时, 通过数据清洗与标准化流程, 消除数据格式与语义差异, 确保多源数据的兼容性与可比性。此外, 基于区块链技术的共享机制可增强数据流转的安全性与透明度, 推动跨部门数据协同(见图 2)。

存储与处理层依托分布式技术实现“全、快、新”要求。一是存储架构采用“冷热双轨”策略, 实时高频数据存入时序数据库(如 InfluxDB)满足秒级响应, 历史低频数据置于分布式文件系统(如 HDFS)保障

海量扩展性。二是处理引擎引入 Flink 流式计算框架, 实现污染事件动态溯源、传输通道模拟的毫秒级响应。三是区块链存证技术为跨部门数据共享建立信任锚点, 通过智能合约自动记录数据流转路径, 确保气象、工信等部门协同数据的不可篡改性。

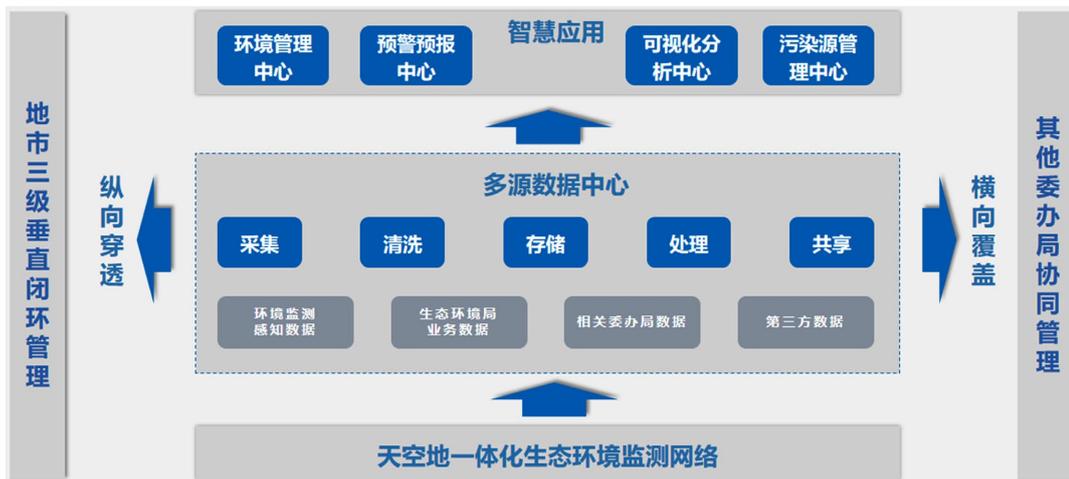


Figure 2. Data Processing Workflow for Multi-Source Data Fusion Centers
图 2. 多源数据融合中心数据处理流程

共享模块的创新性在于构建了“三级垂直闭环管理”体系: 纵向穿透, 打通“市-区-街道”三级环境管理链条, 预警预报中心生成的管控指令可直达基层执行单元(如网格员移动终端); 横向协同, 通过政务云平台与相关委办局(住建、交管)建立数据交换接口, 例如融合交通卡口数据优化移动源排放清单; 智慧应用闭环, 可视化分析中心将多源数据转化为动态知识图谱, 支撑污染源管理中心的“靶向管控”, 如基于企业用电量、视频监控的实时关联分析, 自动触发偷排嫌疑预警。

3. 核心技术体系与实现路径

3.1. 多源异构数据融合与标准化处理技术

与传统的环境监管平台不同, 基于数据融合的智慧监管平台通过系统性数据治理流程, 实现了环境数据的全生命周期管理, 解决了数据碎片化与质量参差问题。

3.1.1. 异构数据归集与分类

平台整合 12 大类环境数据, 涵盖空气质量监测、污染源排放、气象观测、遥感监测等。一是环境质量数据, 如空气质量监测站(6 项常规污染物: SO_2 、 NO_2 、 PM_{10} 、 CO 、 O_3 、 $PM_{2.5}$); 大气超级站(25 类精细化监测设备, 如 VOCs 分析仪、颗粒物粒径谱仪); 热点网格监测($PM_{2.5}$ 、 PM_{10} 实时浓度, 覆盖全域)。二是污染源数据, 如工业企业排污许可证信息(含产能、排放限值等); 企业在线监控数据(NO_x 、 SO_2 、用电工况等实时数据流, 采样频率 5 分钟/次); 机动车尾气遥感监测数据(日均检测车辆超 5000 辆次, 黑烟车识别准确率 $>95\%$)。

通过物联网传感器(如工地扬尘监测仪)、API 接口(气象局、美国 NOAA 数据接口)及非结构化数据处理技术(OCR 识别历史纸质数据), 平台日均处理数据量达 10TB, 支撑实时分析与动态决策。

3.1.2. 数据清洗与关联规则

数据清洗分为四阶段: 一是原始数据规范化, 统一单位, 如 $PM_{2.5}$ 单位统一为 $\mu g/m^3$ 。修正异常值,

剔除传感器故障导致的离群值。二是实体关联匹配,以“统一社会信用代码”为核心标识符,关联企业排污许可、在线监控、二污普数据,构建企业全生命周期档案。三是冗余数据取舍,对文本信息如企业地址等,采用“最新优先”规则;数值型数据如排放量等,保留多来源数据以支持交叉验证。三是规则化识别,对数值偏差 >30%的数据添加标签,如在线数据与历史数据偏差大于临界值,即触发人工复核流程。

3.1.3. 资源目录与元数据管理

依据《环境信息分类与代码》(HJ721-2014),平台构建了三级数据资源目录。一级目录,按业务域分类,如环境质量、污染源、气象等。二级目录,细化至数据类别,如PM_{2.5}监测、VOCs排放清单。三级目录,定义数据字段,如监测时间、浓度值、设备编号。

通过可视化元数据管理工具,实现数据血缘分析(即追踪数据来源与流转路径)和版本控制,确保数据可追溯性。

3.2. 智能算法与动态溯源技术

平台基于深度学习与空间分析技术,构建了多场景AI模型,显著提升污染溯源与预测精度。

3.2.1. 时序数据分析模型

一是LSTM空气质量预测,模型输入包含历史空气质量数据(PM_{2.5}、O₃)、气象参数(风速、湿度)、污染源排放量等,通过时间窗口滑动(窗口长度7天)与特征工程,预测未来7天污染物浓度,平均绝对误差(MAE) < 10 μg/m³。二是动态污染溯源模型,融合气象场(WRF模型输出)、污染源排放清单(包含5000余家工业企业)与监测数据,采用动态权重分配算法(气象条件权重占比30%~50%),识别对空气质量影响排名前10%的重点污染源。

3.2.2. 图像识别与空间组网分析

一是非现场监管算法,工地违规识别通过视频监控与YOLOv5模型,实时检测未覆盖防尘网的裸露土方,识别准确率达92%;生物质燃烧识别通过结合热红外遥感数据与卷积神经网络(CNN),定位焚烧火点,响应时间 < 10分钟。二是激光雷达组网技术,部署2台颗粒物激光雷达,反演气溶胶光学厚度(AOD)与边界层高度(PBL),结合卫星遥感数据,构建污染物传输通道模型,解析区域传输贡献率。

3.2.3. 污染源解析技术

将颗粒物、光化学组分数据与常规数据、气象数据结合进行深度应用,通过时序分析、相关性分析、组分重构、规律分析、PMF溯源模型、OBM模型等方式,分析掌握主要组分浓度及占比结构变化,对比分析不同场景下组分浓度特征差异性,分析不同组分因子之间的同源性、关联性等,依据掌握的源谱分布的信息来决定解析出的污染源类型,辅助判断颗粒物或光化学污染组分的主要来源。

3.3. 动态监管与实时预警技术

平台通过多维度监控与智能告警机制,实现污染源全流程闭环管理。

3.3.1. 污染源动态画像

一是工业源监管,建立多源数据融合模型,将烟气排放浓度、厂界污染物监测、企业分表记电、生产设施与治理设施工况监测等众多数据融合为工业源动态画像,及时发现工业企业异常生产与排放状况,来有效应对环保监管过程中的种种问题,实现在线联防联控,深化污染源的非现场监管。二是移动源监管,通过机动车尾气固定遥感监测识别超标车辆,利用多源数据进行数据抽取、关联、异构、挖掘和表达,分析结果自动生成凭证,为执法部门提供现场检查依据,大幅提高移动源执法的时效性和准确率。

3.3.2. 多级告警与“智慧调度一张图”

一是告警规则设定, 如空气质量同比超标(阈值 15%)、环比突变(阈值 25%); 企业排放浓度超过许可限值, 如 SO_2 浓度 $> 100 \text{ mg/m}^3$; 组分数据异常告警, 如 CO 浓度环比超标(阈值 15%)。二是“智慧调度一张图”, 协调指挥机构和各执行单位实现跨平台异地协同, 基于现场图片、多图层地理信息进行专业标绘, 提供环境问题事件的联动分析、任务生成、领导签批、任务下发、任务办理、任务跟踪及任务反馈的全流程可视化指挥调度。

4. 结论

本研究构建的多源数据融合大气环境智慧监管平台, 实现了环境治理从“经验驱动”向“数据驱动”的范式转变。在技术体系层面, 通过“采集-清洗-存储-处理-共享”五步流程整合 12 类异构数据(含企业工况、激光雷达组网及机动车遥感), 依托区块链存证技术保障跨部门数据可信共享; 算法层融合深度学习(LSTM 预测模型、YOLOv5 违规识别)与空间分析技术(WRF 气象场耦合传输通道模型), 显著提升污染溯源精度; 应用层基于污染源动态画像(排放浓度与分表记电数据融合)及阈值-概率双规则告警机制, 实现污染事件分钟级响应。管理模式层面, 纵向贯通“市-区-街道”三级管理链条使预警指令直达网格终端, 横向构建委办局协同网络(如集成交通卡口数据优化移动源清单), 支撑“一源一策”精准管控体系。

未来研究将聚焦模型智能升级与跨域技术融合。一方面通过大语言模型(LLM)增强污染成因推理能力, 构建“碳污协同”监测模块支撑减污降碳联动; 另一方面深化边缘智能部署, 将轻量化算法嵌入物联网终端实现秒级响应, 同时耦合 WRF-Chem 模型量化跨省污染传输贡献率, 拓展区域联防联控技术路径。

参考文献

- [1] 张昺榴, 王思敏, 宗蕾雯, 等. 金沙江白鹤滩水电站生态环境智慧管理平台的应用[J]. 计算机应用, 2024, 44(S1): 374-378.
- [2] 王超, 安贝贝, 张秀. 重庆市生态环境智慧监测管理体系研究[J]. 环境监测管理与技术, 2023, 35(1): 1-3+58.
- [3] 尹华政, 薛联芳, 唐际政, 等. 基于多技术多数据融合的水电环境监测管理平台建设[J]. 水力发电, 2023, 49(2): 6-11+16.
- [4] 刘柏音, 王维, 刘孝富, 等. 大数据技术在我国生态环境领域的应用情况与思考[J]. 环境保护, 2022, 50(14): 57-61.
- [5] 刘柏音, 王维, 刘孝富, 等. 长江流域水环境监测与智慧化管理策略[J]. 中国环境监测, 2022, 38(1): 222-229.