

数据分析在大气污染源识别与追踪中的应用

张瑞丽

河南速创环保科技有限公司, 中国·河南 郑州 450000

摘要: 本研究针对全球大气污染严峻形势, 深入探究数据分析在大气污染源识别与追踪中的应用。详细阐述大气污染源按工业、交通、生活源分类的排放特征, 指出传统识别与追踪方法的数据获取难、时效性差、监测范围有限等局限。通过整合大气监测、气象、地理信息等多元数据, 运用机器学习、深度学习、统计分析等技术及相关工具, 构建了精准的污染源识别与追踪体系。成功识别出主要污染源并揭示其扩散规律, 但研究在数据质量、模型精度和应用范围方面存在不足。未来需借助技术创新、多元数据融合及跨区域研究, 提升大气污染治理水平, 应对全球大气污染挑战。

关键词: 大气污染; 数据分析; 污染源识别与追踪

Application of Data Analysis in Identification and Tracking of Air Pollution Sources

Ruili Zhang

Henan Suchuang Environmental Protection Technology Co., Ltd., Zhengzhou, Henan, 450000, China

Abstract: This study aims to explore the application of data analysis in identifying and tracking air pollution sources in response to the severe global situation of air pollution. Elaborate on the emission characteristics of air pollution sources classified by industrial, transportation, and domestic sources, and point out the limitations of traditional identification and tracking methods such as difficult data acquisition, poor timeliness, and limited monitoring scope. By integrating multi-source data such as atmospheric monitoring, meteorology, and geographic information, and using machine learning, deep learning, statistical analysis, and related tools, an accurate pollution source identification and tracking system has been constructed. Successfully identified the main sources of pollution and revealed their diffusion patterns, but there are shortcomings in data quality, model accuracy, and application scope in the research. In the future, technological innovation, multi-source data fusion, and cross regional research are needed to improve the level of air pollution control and address global air pollution challenges.

Keywords: air pollution; data analysis; identification and tracking of pollution sources

0 前言

在全球工业化与城市化快速发展下, 大气污染已严重威胁人类健康与生态环境。世界卫生组织数据显示, 2016 年约 700 万人因空气污染过早死亡。中国京津冀等发达地区雾霾频发, 如 2013 年多地空气质量指数爆表, $PM_{2.5}$ 浓度远超安全标准, 引发多种疾病。大气污染危害多样复杂, 既损害人体健康, 导致呼吸道、心血管疾病等, 又影响气候环境, 加剧全球变暖、引发酸雨。污染源识别与追踪是治污关键, 但传统监测治理方式因监测范围窄、时空分辨率低等难以满足需求。此时, 数据分析作用凸显, 它能整合大气监测、气象、地理等多源数据, 深度挖掘分析, 精准识别污染源, 追踪污染物传输扩散, 还能预测污染趋势, 为大气污染治理提供新思路与科学依据。

1 大气污染源识别与追踪概述

1.1 大气污染源分类及特征

大气污染源按人类活动分为工业源、交通源、生活源等, 各有独特排放特征。

工业源涉及钢铁、化工等多行业, 生产中排放大量污染物。例如, 钢铁冶炼产生颗粒物, 会危害人体呼吸; 还排放 SO_2 、 NO_x 、VOCs 等, 分别能致酸雨、参与光化学烟雾形成、损害人体神经系统。其排放浓度高, 部分企业远超国标, 排放规律与生产工艺、计划相关, 有连续或集中排放情况。

交通源含汽车等交通工具, 是城市大气污染主因之一。汽车尾气有 CO、HC、 NO_x 、PM 等污染物, CO 会致人体缺氧, HC 与 NO_x 参与光化学烟雾形成。排放浓度与车辆类型、行驶状态、道路条件有关, 柴油车颗粒物和 NO_x 浓度高, 车辆不同行驶状态排放有别, 早晚高峰排放集中, 城市中心和交通枢纽贡献突出。

生活源涵盖取暖、烹饪、垃圾焚烧等。北方冬季燃煤取暖排放 SO_2 、颗粒物; 烹饪油烟含 VOCs 和致癌物; 不规范垃圾焚烧产生剧毒二噁英。生活源排放浓度低但分布广, 总量可观, 排放规律与居民生活习惯相关, 老旧小区贡献大。

1.2 传统大气污染源识别与追踪方法及局限性

传统方法有一定作用但存在局限。源清单法通过调查

各类污染源排放情况建清单识别主污染源,能直观展示排放,有助于治理。但数据获取难且不准,排放系数与实际有偏差,小型企业数据难获,且时效性差,难反映动态变化,更新慢,也难考虑地理、气象对污染物扩散的影响。现场监测法在固定站点设备监测污染物浓度,还可用移动设备补充,能得实时数据。然而监测站点分布有限有盲区,数据代表性受周边环境的影响,且只能知浓度,无法直接确定来源。气象示踪法依气象与污染物传输关系追踪,借风向等参数和浓度变化推断来源方向。但气象复杂多变,特殊条件下准确性受挑战,且只能给大致方向,难精确定位。

2 数据分析技术基础

2.1 大数据技术简介

大数据以 4V 特性重塑数据应用。其数据量庞大,全球众多大气监测站点每日生成海量污染物及气象数据,如中国空气质量监测网络日收集数百万条记录。数据产生与处理速度快,需秒级处理大气监测数据,应对突发污染事件。数据类型多样,涵盖结构化监测数据、非结构化卫星影像及社交媒体信息等。虽价值密度低,但经深度分析,能挖掘出大气污染源位置、传输路径等关键信息。

2.2 数据分析在大气污染研究中的应用原理

在大气污染研究中,数据分析借助多源数据挖掘与分析发挥关键效能。数据挖掘层面,关联规则挖掘能探寻污染物与气象状况间的潜在联系,聚类分析可对繁杂污染数据加以分类梳理,时间序列分析则用于预测污染物浓度走势。机器学习算法不可或缺,监督学习利用训练数据构建模型,像支持向量机、决策树等实现污染源识别;无监督学习里,K 均值聚类挖掘污染模式,主成分分析降低数据维度。深度学习领域,卷积神经网络处理图像识别污染源,循环神经网络及 LSTM 分析时间序列数据预测污染物变化,助力大气污染防控。

3 基于数据分析的大气污染源识别与追踪方法

3.1 多源数据收集与整合

在大气污染源研究里,多源数据收集整合极为关键,涵盖大气监测、气象、地理信息等数据,为分析大气污染提供全面信息。

大气监测数据直接反映污染状况,包含 $PM_{2.5}$ 、 SO_2 等多种污染物浓度。这些污染物危害各异,像 $PM_{2.5}$ 会严重威胁人体健康, SO_2 能破坏生态平衡。获取准确数据需合理布局监测站点,城市依人口、工业、交通设点,农村也适当布局。站点靠先进设备实时采集数据并传输至数据中心。

气象数据影响污染物扩散。气温、湿度、风速等气象要素与污染物浓度分布紧密相关。高温低湿利于扩散,低温高湿易积聚。风速风向决定传输方向速度。气象站分布各地,通过观测设备获取实时及历史数据,历史数据能反映气象变化规律,助力分析污染长期趋势。地理信息数据提供空间背

景。地形地貌影响污染物扩散,山区易积聚,平原相对均匀但受建筑影响。土地利用类型不同,污染源类型分布有别,工业用地排放多,农业用地也有污染物。城市布局影响污染分布,交通繁忙处及商业区、居民区污染突出。地理信息靠卫星遥感、GIS 获取,GIS 可叠加分析,展示污染空间分布及与地理要素关系。

多源数据收集后需整合预处理。要解决格式不一致、缺失、冗余问题,将不同格式数据统一,用合适方法插补缺失值,去重筛选冗余数据。预处理还包括清洗噪声异常值,标准化不同量级数据,方便后续分析。

3.2 基于数据分析的大气污染源识别方法

统计分析方法揭示大气污染数据背后信息,助力识别污染源。

主成分分析(PCA)将多个相关变量转化为少数不相关综合变量。大气污染数据变量多且相关,PCA 通过线性变换使新变量方差最大且不相关,按方差排序,累计贡献率达 85% 以上的主成分可代表原始数据。分析主成分与原始变量关系,能识别污染主要因素与潜在污染源。PCA 可降低冗余,但主成分物理意义不明确,对非线性数据处理欠佳。

因子分析将观测变量分解为公共因子和独特因子线性组合。公共因子解释变量相关性,独特因子为变量特有。大气污染识别中,因子分析找出公共因子代表不同污染源或污染过程。与 PCA 相比,因子分析更注重挖掘潜在结构,但公共因子提取命名需经验,旋转方法不同结果有别。

4 数据分析在大气污染源追踪中的应用

4.1 大气污染物扩散模型与数据分析结合

利用数据分析优化扩散模型,需深入分析气象、地形、污染源排放等多源数据,并融入模型参数确定与调整。

气象数据影响污染物扩散,分析其要素变化规律可优化模型。风速、风向、温度等改变大气状态,影响扩散路径速率。地形数据影响污染物在复杂地形下的扩散。地形起伏、坡度等改变大气流场,影响扩散路径。利用 GIS 技术处理地形数据,生成数字高程模型,分析地形特征输入模型,调整扩散参数。还可分析地形对气象影响,完善模型对气象条件的模拟。污染源排放数据的准确性影响模型结果。需详细分析污染源位置、排放量、排放方式等。不同类型污染源排放特征差异大,工业源排放量大且集中,交通源流动性和间歇性,生活源分散。通过物料衡算、交通流量监测等方法准确计算排放量,输入模型,真实模拟排放扩散过程。

分析多源数据后,将结果融入模型参数调整。高斯模型依气象、地形确定扩散系数;AERMOD 模型依气象、地形确定边界层等参数,依污染源排放确定源强;CALPUFF 模型依预处理生成的微尺度风场等和污染源排放,调整 Calpuff 模块参数,提高模型准确性和可靠性。

4.2 基于数据分析的污染源追踪技术实现

4.2.1 时间序列分析在追踪中的应用

时间序列分析在大气污染源追踪中发挥着关键作用。它主要对污染物浓度随时间变化的数据进行深入剖析,从而揭示污染物排放的时间规律。例如在城市里,我们通过观察 $PM_{2.5}$ 浓度随时间的变化,能发现冬季供暖期以及早晚高峰时, $PM_{2.5}$ 浓度会升高。这是因为供暖时燃烧燃料,还有早晚交通繁忙,汽车尾气排放增多,所以能初步推测这些时段的污染源。常见的时间序列分析手段有移动平均法和指数平滑法。移动平均法就是算出一段固定时间内污染物浓度的平均值,这样能把短期内浓度的波动消除,让长期变化趋势更明显。指数平滑法则是对不同时间的数据给予不同的重视程度,更看重近期的数据。通过调整一个参数,能灵活决定对近期数据的关注程度。在实际操作中,时间序列分析还会和空间分析、机器学习算法一起使用,从空间分布、数据规律挖掘等方面共同发力,大幅提升追踪大气污染源的准确性和效率。

4.2.2 空间分析技术助力污染源定位

GIS空间分析在大气污染源定位中作用突出,具有显著优势。缓冲区分析,简单来说,就是围绕污染源等地理要素创建一定范围的区域。就像以工厂为中心,画出一个圆形区域,这个区域就是缓冲区。通过它,我们能清晰地知道污染源可能影响的范围,判断哪些地方可能成为潜在的污染区域,也能借此评估治理措施有没有起到缩小影响范围的效果。空间插值则是把那些零散分布的监测点数据,转化成连续的空间分布情况。例如,在一个城市里有几个监测站的数据,但中间很多地方没有数据,空间差值就可以根据这些零散数据,估算出中间区域的污染状况,填补数据空白,让我们对整个区域的污染情况有更全面的了解,为精准定位污染源提供更完整的空间信息。网络分析主要是研究污染物

在类似道路网络等地理网络中的移动情况。例如,污染物可能顺着风向、河流等类似网络的路径扩散。通过网络分析,我们能弄清楚污染物的传输路线和扩散规律,进而确定污染源究竟在哪里,还能评估不同污染源对特定区域的污染贡献大小。

5 结论与展望

本研究聚焦数据分析在大气污染源识别与追踪中的应用,成果颇丰。在多源数据处理上,全面收集大气监测、气象及地理信息数据,经清洗、预处理与集成,构建高质量数据集,解决了数据诸多问题,为后续分析筑牢基础。然而,研究仍存不足。数据质量受监测设备精度、环境及传输干扰影响,不同源数据分辨率差异也带来融合难题。模型精度上,缺乏通用模型,机器学习易在小样本数据出现拟合问题,深度学习计算资源需求大且解释性差。应用范围局限于特定区域和事件,对不同地理、气候条件普适性研究不足。未来,技术创新有望借新兴技术提升分析方法与模型性能,多源数据融合需纳入更多数据源挖掘关联,跨区域研究要加强区域协同及国际合作,以应对大气污染的区域性与全球性挑战。

参考文献:

- [1] 张晓郁,陈振飞,徐文哲.ArcGIS在大气污染源排放清单建立中的应用研究[J].环境科学与管理,2017(1).
- [2] 杨一帆,张凯山.突发型大气污染源位置识别反演问题的数值模拟[J].环境科学学报,2013,33(9):7.
- [3] 杨斌.GIS在大气污染扩散模拟与评价系统中的应用[J].国土资源遥感,2010(3):6.
- [4] 李引迪,张明.环境监测在大气污染治理中的作用及应用[J].生态环境与保护,2021,4(2):3-4.
- [5] ZHANG Xiaxia.大气污染物监测数据不确定度评估方法体系建立及其对PMF源解析的影响分析[EB/OL].第五届经济快速发展地区空气质量改善国际学术研讨会,2017.