

# 环境空气数据审核常见问题及对策

张宏伟

山西省忻州生态环境监测中心

DOI:10.32629/eep.v8i11.2966

**[摘要]** 环境空气数据的审核工作是保证数据的有效性、可信性的主要过程。本文对目前审核实务中经常出现的一些典型问题进行了系统的梳理,主要包括监测数据记录不完整、异常数值判定模糊、仪器设备校准存在偏差和操作人员主观误差等,并给出了相应的改进措施。经过对从采集到传输全过程的全流程管理、构建起多维质量控制体系、开展规范化操作人员培训计划和大力引进利用人工智能自动审核工具等来提高数据审核工作整体效率和结果准确率。本文目的在于给环境监测部门改进数据管理、提高质控水平给予有实践意义的技术途径和决策建议。

**[关键词]** 环境空气监测; 数据质控; 异常值辨识; 智能化审核

**中图分类号:** P412.2 **文献标识码:** A

## Common Problems and Countermeasures in Environmental Air Data Audit

Hongwei Zhang

Shanxi Xinzhou Ecological Environment Monitoring Center

**[Abstract]** The review of environmental air quality monitoring data is the main process to ensure the validity and reliability of the data. This article systematically reviews some typical problems that often occur in current auditing practices, including incomplete monitoring data records, fuzzy determination of abnormal values, deviation in instrument and equipment calibration, and subjective errors of operators. Corresponding improvement measures are also provided. Through the full process management from collection to transmission, the establishment of a multidimensional quality control system, the implementation of standardized operator training programs, and the vigorous introduction and utilization of artificial intelligence automatic auditing tools, the overall efficiency and accuracy of data auditing work have been improved. The purpose of this article is to provide practical technical approaches and decision-making recommendations for environmental monitoring departments to improve data management and enhance quality control levels.

**[Key words]** environmental air monitoring; Data quality control; Identification of outliers; Intelligent auditing

环境空气监测数据的审核属于环境管理体系中的重要基础性工作,它的质量好坏直接影响到污染状况评价的准确性、环境政策的科学性和公众健康风险防控的有效性。但是,在实际的业务运作过程中,数据审核环节经常遇到很多的技术和管理方面的问题,比如数据序列的完整性很难保证,异常数据的成因复杂难以准确甄别,监测仪器自身存在的系统性误差,还有就是人在操作过程中的疏忽、不一致等。以上问题或多或少会降低监测数据的公信力以及应用价值。本文目的在于分析环境中空气质量数据审核中经常出现的一些问题,并且提出一个包含技术改进、流程规范和能力培养在内的综合解决办法,从而提高环境监测数据的质量,完善质量控制体系。

### 1 环境空气监测数据审核面临的主要问题

#### 1.1 数据记录不完整与序列中断

监测数据不足或者中断,审核阶段最常见的情况就是数据不全或者断续。造成该问题的主要原因有现场监测设备突然发生故障、外部的电力供应或通信网络不稳定、站点例行维护造成的计划外停机等。数据记录的缺失,尤其是在污染浓度发生快速变化的关键时刻,会影响时间序列的完整性,以及之后的趋势分析、均值统计以及污染事件的发现都会受到影响。以大气颗粒物监测为例,在重污染期间出现几个小时数据空白的话,就会使日均浓度、超标评价等重要的指标计算结果产生较大的误差,从而不能对污染过程做出正确的评判。

#### 1.2 异常监测值的辨识与处置困境

环境空气数据审核时出现的异常监测值是技术上的一个大问题。它的形成原因比较复杂,既可以是真实的极端排放事件或者特殊的气象条件所导致的,也可以是由于监测设备自身漂移、

传感器故障和采样干扰造成的。传统的用固定的阈值来判断的方法,不能准确地区分出真实的污染峰值和仪器的异常信号。二氧化氮浓度突然增加既可能是由于突发性排放事故,也可能是由于仪器光电倍增管老化。如果处理不当,把真实的高值当成异常剔除,就会掩盖污染的事实;反之,如果把故障信号保留,则会产生误差,歪曲污染的特征。本课题暴露的是对智能、准确异常值识别技术的需求。

### 1.3 监测仪器校准与维护管理不到位

监测仪器自身准确性和稳定性的高低决定着数据的质量好坏。校准不规范或者维护不及时,是造成系统性误差、使数据长时间偏离真值的主要原因。部分监测站存在校准周期过长、未严格按照多点校准规程进行校准,或者使用标准气体不规范等问题。例如,用于测定环境臭氧浓度的紫外光度分析仪,如果它的内部光学窗片被污染或者光源强度下降而没有及时进行校准和维护,则所测得的监测值就会出现系统性的偏低或者偏高。另外,采样管路由于颗粒物积累而堵塞或者吸附、温湿度传感器长期暴露在恶劣环境中造成性能下降等问题,在日常巡检过程中很容易被忽略,但是它们却是影响数据准确性的隐性但持续存在的问题<sup>[1]</sup>。

### 1.4 人为操作与主观判断引入的误差

环境监测数据审核工作中人的因素属于系统误差的主要来源。操作人员由于对设备原理和质控程序的理解不够,或者在长期的重复性工作当中产生疏忽,造成数据录入错误、误标有效/无效数据,在预处理过程中使用不恰当的修正方法。审核一氧化碳数据的时候,如果不知道它的浓度在交通高峰时段的规律性变化,就会把正常的日变化当作异常波动来判断。尤其是对于边界模糊的数据判定,常常依靠审核人员的经验来决定。由于没有明确统一的标准操作程序(SOP),不同的审核人员对于同一个数据或者异常现象作出不同的处理决定,从而大大降低审核结果的准确性以及客观性,也会给不同时期、不同站点的数据对比带来困难。

## 2 提升数据审核质量的核心对策

### 2.1 优化数据采集链路的可靠性与冗余度

提高数据完整性,必须从源头上加强采集和传输系统建设。首要措施就是建立关键监测参数的冗余采集能力,在主监测设备之外设置备用传感器或者采用不同的原理配置比对监测设备,形成数据互补和互校。其次要对数据传输系统进行升级改造,使用具有断点续传和本地大容量缓存功能的数采仪,可保证在网络瞬间中断时,高频数据(如分钟级)能被完整保存,通信恢复后自动补传。另外,创建实时运行状况监控系统,对监测设备的电压、内部温度、流量、重要部件运转时间等参数进行在线监测和预警,可将维修模式从原来的“事后维修”转变为“事前维护”,从源头上防止设备故障导致数据丢失。

### 2.2 发展多维融合的异常值智能辨识方法

为了克服传统阈值法的不足,应构建由多源数据融合和智能算法组成的异常值检测系统。该方法的关键在于不孤立地只

看一个监测参数的时序变化,而是将其与相关的协变量一起进行联动分析。在判断某个站点PM<sub>2.5</sub>浓度峰值是不是真的时候,可同时同一站点其它污染物(如SO<sub>2</sub>、NO<sub>x</sub>等)、气象参数(如风速、风向、相对湿度)、邻近站点的浓度空间分布进行分析,并结合历史同期的浓度分布模式。从技术角度而言,可将统计过程控制(SPC)模型、基于距离的聚类算法(LOF)、用于时间序列的深度神经网络(如LSTM等)相结合,构建一个动态、自适应的异常检测系统。系统自动筛选出的可疑异常点,还需要由有专业知识的人士进行最后的复核,形成机器初筛、人工确认的高效可靠的流程。

### 2.3 实施严格且精细化的仪器全生命周期质量管理

保证仪器数据的长期准确可靠,必须建立包含仪器购置、安装、验收、日常运营、期间核查、维护、报废的全生命周期质量管理体系。重点就是制定并强制执行比国家规范更严格、更详细的内部校准和维护计划<sup>[2]</sup>。对核心分析设备,比如气体污染物分析仪等,需要做高频次的零点/跨度检查(每天自动进行),并且定期用有证的标准物质做多点校准和线性检验。第二,建立健全仪器设备档案和运行履历,记载每一个校准、维护、部件更换、性能测试的结果,达到状态可追溯的目的。第三,采用远程诊断和标定技术,在远端对仪器的状态参数进行检查、部分校准,可以加快技术支持人员对仪器故障情况的了解,提高维修效率,适合于地理位置偏僻或者人工难以到达的地方监测站。

### 2.4 推行标准化操作规范与系统化人员能力建设

最大限度地减少人为误差,需要从制度与人两方面一起着手。制度上要制定出详细的、可操作性强的《环境空气质量监测数据审核标准作业程序》(SOP),明确从原始数据采集、有效性判断、异常值处理、数据修正、缺失值标识到最后审核确认的全过程的操作规范、判断标准及记录要求,用客观的标准取代主观的判断。在人的方面,应该创建起一个周期性、分层次分类的人员培训和资格认证制度。培训的内容不但是仪器的操作、质控的流程,还要包含基础的环境科学知识、污染物的变化规律、常见的数据问题的诊断和案例分析。另外定期对不同的人、不同的站点进行数据的交叉审核、盲样考核和能力验证活动,来保证审核团队整体的技术水平以及判断的一致性。

## 3 智能化技术在数据审核流程中的应用与探索

### 3.1 自动化数据审核平台的建设与效益

创建起自动化的数据审核平台,是实现审核工作高效化、标准化、规模化的重要途径。这类平台利用内置的审核规则引擎以及算法模型来对大量的监测数据实施实时或者准实时的自动筛选、预判与标记。平台可以自动完成时序连续性检测、阈值超限报警、气象条件合理性校验、多站点数据空间一致性比对等很多重复工作。对于疑似异常的数据,平台会产生一个带有相关协变量数据图表的问题报告,并发送给指定的审核人员。自动化审核的实行,把审核人员从大量的机械、重复的初步筛查工作中解脱出来,使他们可以更多地投入到对复杂、疑难案件

的审核中去,大大提高了整个审核效率,也有效地减少了由于人工疲劳或者疏忽而导致的漏判、误判,提高了审核结果的一致性。

### 3.2 大数据与人工智能模型的深度赋能

大数据、人工智能技术给数据审核带来范式创新的可能。从长远的角度出发,以多个站点、多种参数(污染物、气象、源清单、交通流量等)为背景的数据汇集在一起而形成的高维环境空气质量数据库。在此基础上,用机器学习、深度学习的方法来训练出一个预测模型,可以对某一地点、某一时间段内污染物浓度的正常波动范围做出准确的预测。在实际的审核过程中,可以把实时监测数据同模型的预测区间做动态比对,超出合理预测范围的数据点可以被迅速高效地标识出可能的异常。另外可以采用人工智能技术去挖掘各个污染物浓度之间的隐含联系规律,进而建立起一个污染指纹库。当某一站点污染物浓度比例关系明显与该站的历史“指纹”不符的时候,系统会发出警告信号,提示有仪器干扰或者特殊污染源。因为区块链技术具有不可篡改、可追溯的特点,在保证监测数据采集、传输、发布过程真实性的同时,也可以起到重要的作用。

### 3.3 技术应用的当前局限与人工复核的必要性

虽然智能化技术有很好的发展前景,但是目前处在发展的初期阶段,在这个时期还必须清醒地认识到它的局限性。第一,任何一种算法模型都是建立在历史数据的基础上进行训练的,在没有出现过新的新型污染事件、复杂的极端气象条件或者多种因素耦合而产生的异常情况的时候,模型的识别能力就会变得很弱,甚至会出现误判的情况。比如由于区域大面积山火造成的PM10、CO浓度的特殊关联变化,如果模型训练数据中没有这类案例的话,就会把它误认为是异常。其次,算法的决策逻辑大多为“黑箱”或者“灰箱”,它给出的异常判断没有直观的物理解释,不能满足环境管理中需要数据异常原因可以追溯、可以说明的要求<sup>[3]</sup>。因此在未来的一段时间里,智能审核系统应该以“高级辅助工具”的目标来定位自己,并不能取代人。系统设计要保留便捷、灵活的人工复核和最终裁决接口,形成“机器智能筛查、专家知识研判、人机协同决策”的优化审核模式,并把人工复核中发现的新知识、新规则及时反馈给模型,使算法得到不断的迭代更新。

## 4 结论与展望

### 4.1 结论

保证环境空气监测数据准确可靠属于一个系统的工程,数据审核就是其中的核心质控环节。本文对目前审核中遇到的数据不完整、异常值判断困难、仪器系统偏差和人为操作不一致等问题进行了系统的分析。因此,从技术、管理、人员三个方面提出综合性的对策:提高数据链路的可靠性,使用智能检测手段来应对技术上的挑战,推行仪器全生命周期精细化管理,建立标准审核程序来加强管理,开展系统化的培训来提高人员素质。自动化审核和人工智能技术的应用,给提高审核的效率及一致性提供新的途径。

### 4.2 展望

未来环境空气数据审核会向着更智能化、标准化、一体化的方向发展。一方面,在物联网、边缘计算、数字孪生等技术不断发展的情况下,监测数据的采集、质量控制、审核流程会越来越紧密地结合起来,并且一部分的质量控制审核功能会被嵌入到感知端。另外一方面就是人工智能算法的可解释性、对复杂的污染过程的适应性、快速学习新型污染物的能力等将成为下一代智能审核系统要解决的主要技术难题。另外,要创建起国家或者区域层面统一的数据审核规范和技术指导方针,促进各地、各机构之间审核标准的一致性,对创建起全国统一、可比的高质量环境空气质量监测网络有重大的意义。最后形成的技术先进、标准统一、人机协同的高效数据审核体系,会成为精准治污、科学决策、保护公众健康的强大数据基础。

### [参考文献]

- [1]周晶晶.环境空气自动监测数据审核中异常数据处理办法探讨[J].皮革制作与环保科技,2023,4(22):81-83.
- [2]王慎阳,徐建阁,马红磊.环境空气自动监测数据审核中异常数据处理办法探讨[J].环境与发展,2020,32(09):174-175.
- [3]崔静.环境空气自动监测系统智能化发展展望[J].清洗世界,2025,41(06):95-97.

### 作者简介:

张宏伟(1984--),男,汉族,山西原平人,本科,中级十级,研究方向:环境空气预警预报、环境空气数据审核、环境空气监测仪器等。