基于气象大数据的环评技术复核研究

伯鑫[1]，唐千红[2]，王骏[3]，王成鑫[1,4]，朱荣杰[1,5]，王彤[6]，李涵[7]
（1.生态环境部环境工程评估中心，北京 100012；2.中国气象局公共气象服务中心，北京 10081；
3.天气科技（北京）有限公司，北京 100081；4.四川大学建筑与环境学院，四川 成都 610065；
5.天津大学环境科学与工程学院，天津 300354；6.陕西省环境调查评估中心，
陕西 西安 710000；7.重庆市生态环境工程评估中心，重庆 401121）

摘要：针对环评技术复核工作中存在的一些问题和需求，以中国气象局权威气象数据为基础，建立了一套基于气象大数据的环评技术复核原型系统，对环评业务使用的气象数据进行智能化、自动化复核。典型复核案例结果表明，该系统可定量评估气象数据真实性，识别相关疑似造假行为。

关键词：空气质量模型；环评；气象；复核；大数据
中图分类号：X82　**文献标识码：**A　**文章编号：**2095-672X(2021)02-0031-06
DOI.10.16647/j.cnki.cn15-1369/X.2021.02.005

Research of Environment Impact Assessment technical reviewing based on meteorological big data

Bo Xin[1], Tang Qianhong[2], Wang Jun[3], Wang Chengxin[1,4], Zhu Rongjie[1,5], Wang Tong[6], Li Han[7]
(1.Appraisal Center for Environment and Engineering, Ministry of Ecology and Environment, Beijing 100012, China; 2.Public Meteorological Service Center, China Meteorological Administration, Beijing 100081, China; 3.WeatherTech (Beijing) Co., Ltd. Beijing 100081, China; 4.School of Architecture and Environment, Sichuan University, Chengdu Sichuan 610065, China; 5.School of Environmental Science and Engineering, Tianjin University, Tianjin 300354, China;6.Shaanxi Environmental Investigation and Assessment Center, Xi'an Shaanxi 710000, China; 7.Chongqing Ecological Environment Engineering Evaluation Center, Chongqing 401121, China)

Abstract: With consideration of the existing problems and requirements in Atmospheric Environment Impact Assessment checking, we have established a set of EIA technical review system based on meteorological big data which comes from China Meteorological Administration. This system can review the meteorological data used in EIA business automatically and intelligently. Rely on the analysis of typical cases, the result shows that the system can evaluate the authenticity of meteorological data quantitatively and identify the deceptive behaviours.

Key words: Air quality models; EIA;Meteorology; Technical review;Big data

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2019年，生态环境部发布《建设项目环境影响报告书（表）编制监督管理办法》[1]，要求对全国环评文件开展技术复核工作，并规定“鼓励利用大数据手段开展复核工作”。2020年，生态环境部发布《关于严惩弄虚作假提高环评质量的意见》[2]，提出“生态环境部推进大数据在线自动查重，对各地审批的环评文件及时开展智能校核”。2020年，我国公布了《中华人民共和国刑法修正案（十一）（草案二次审议稿）》[3]，明确提出环境影响评价机构及相关人员的造假行为“入刑”，进一步强化了环评造假行为惩罚力度。

针对环评报告大气环境影响预测章节，本研究团队结合《环境影响评价技术导则 大气环境》等要求，已开展了大量技术复核工作[4~6]，发现一些环评单位在空气质量模型参数设置、数据处理方面存在一些错误问题，例如篡改气象数据、错误使用气象数据等。

大气环境影响技术复核工作主要审查气象、地形、污染源、坐标投影、地表参数、模型输入及输出文件等[7~8]，输入模型的气象文件格式主要是SFC格式文件（AERMOD模型）、MET格式文件（ADMS模型）、DAT格式文件（CALPUFF模型）等，气象要素包括风速、风向、云量、温度、降水、云底高度、相对湿度等，时间分辨率为1h。但由于气象数据存储量大，气象要素的技术复核以人工审查为主，较为繁琐，难以实现复核的自动化、智能化应用。目前，国内气象数据主要应用于公众服务、交通、旅游、农业、水利等领域[9]，尚未应用于环评技术复核业务。

针对上述问题，本研究以中国气象局实时监测、质控、发布的权威气象数据为基础，建立了一套基于气象大数据的环评技术复核原型系统，依托机器学习和大数据分析技术，通过云端服务方式对环评业务使用的气象数据进行智能化、自动化复核，旨在为环评业务的监督管理提供技术支持。

1 研究方法

1.1 基于气象大数据的环评技术复核原型系统

本研究团队总结了环评气象复核的工作难点：(1)气象数据存储量大，通过人工方式逐条复核耗时长、易出错。(2)针对气象数据的弄虚作假行为较为隐蔽，如通过篡改少数时段的气象数据来实现模拟浓度达标。(3)复核人员需要掌握一定气象专业知识，否则难以满足地方复核业务化需求。

本研究建立了基于气象大数据的环评技术复核原型系统（以下简称原型系统），依托机器学习和大数据分析技术开展环评气象复核工作，快速识别环评气象数据中存在的弄虚作假行为，降低复核审查成本，有效提升监督管理效率。该系统主要分为以下几个模块（图1）。

1.1.1 气象大数据平台

气象大数据平台存储了自1949年以来经过中国气象局实时监测、质控和发布的各类权威数据集。这些数据包含全国2400多个国家级地面气象观测站、6万多个区域自动气象站、近200部天气雷达、2000多个土壤水分观测站、1000多个交通气象观测站、300多个雷电观测站、120个探空气象观测站、7颗在轨风云卫星的气象监测数据等。每年新增数据存储量600TB左右，涵盖降水、温度、风力风向等30余种气象要素，并均已实现观测自动化，观测频率达到分钟级，平均气象观测站间距20公里，乡镇覆盖率达到98%。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

气象大数据平台作为环评技术复核原型系统的基础支撑，可支持对任何种类、任意气象要素、任意空间范围、任意时间范围、任意精度的环评气象数据比对分析，能够有效满足不同环评业务的气象数据审查场景，实现“一站式”复核。

1.1.2 环评气象数据预处理模块

环评气象数据预处理模块用于对上传的环评气象文件进行大数据分析前的预处理。由于环评业务中不同环境空气质量模型对输入的气象数据文件在数据结构、特征值、数据处理方式等方面有不同的要求，因此，通过脱密脱敏、坐标转换、格式转换、特征值转换、异常值检测、数据质量控制和可用性检测等多个预处理流程，可以将不同种类的环评气象文件处理为标准的、统一的、结构化的气象数据文件，进而输入大数据分析模型进行分析评分。

1.1.3 基于机器学习的大数据分析模块

大数据分析模块承担着对输入的环评气象数据文件进行比对分析和评分的工作。大数据分析模块在接收到环评气象数据后，将从气象大数据平台中获取对应种类、气象要素、空间范围、时间范围、时空精度的权威数据，并对两类数据进行初步的比对和残差运算。

然后对两类数据的残差进行特征提取，进而采用多种方法开展多维度的统计分析工作，如时序分析、统计分析、聚合分析、空间分析等，最终由评价模型输出环评气象数据文件的复核评分。为便于理解，本文将对数据分析环节中涉及到的部分关键统计指标进行展示和介绍，具体详见本文1.2统计指标。

采用基于机器学习的评价模型对环评气象数据复核的优势在于：随着复核数据文件的增多，系统会对存在造假行为的环评气象数据文件进行持续性标注，从而推动模型对其造假行为特征不断迭代，进一步提升环评气象数据复核的准确性。

1.1.4 交互页面

为了提升环评气象数据复核工作的效率，原型系统提供了便利易用的前端交互界面，用户根据页面指引可快速对须审查的环评气象文件进行复核。具体业务流程（图2）如下：在用户登录系统中，上传审查的环评气象文件到原型系统，原型系统自动将审查各气象要素，并与中国气象局权威气象数据集中的相应数据对比分析，进行智能复核并自动生成复核报告。

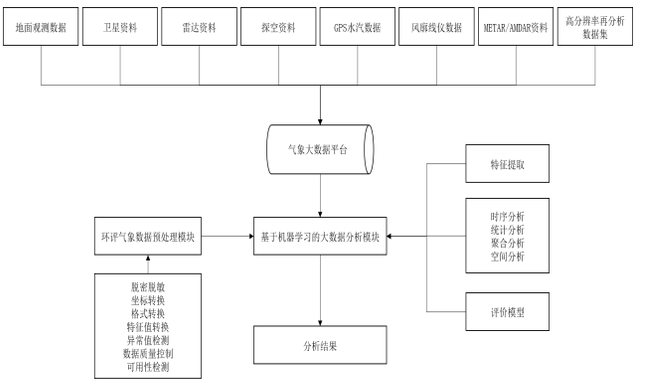


图1 基于气象大数据的环评技术复核原型系统技术路线图

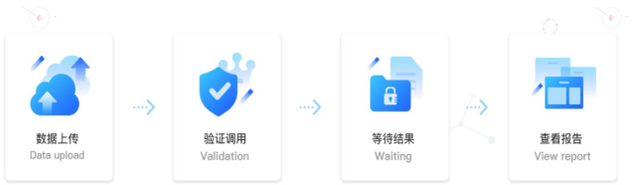


图2 基于气象大数据的环评技术复核原型系统业务流程图

1.2 统计指标

本研究选取平均偏差，平均绝对误差，均方根误差，相关系数等作为统计指标，其计算方法见公式(1)-(4)。

平均偏差（Bias），指气象要素观测平均值

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

和模型分析平均值的差值。

$$\mathrm{Bias}_j = \frac{1}{N}\left(\sum_{i=1}^{N} x_{ij}^{a} - \sum_{i=1}^{N} x_{ij}^{o}\right) \quad (1)$$

平均绝对误差(AE),指对气象要素观测值与模型分析值的差值绝对值进行平均。

$$\mathrm{AE}_j = \frac{1}{N}\sum_{i=1}^{N}\left|x_{ij}^{a} - x_{ij}^{o}\right| \quad (2)$$

均方根误差(RMSE),指对气象要素观测值与模型分析值差值平方和的均值 开方。

$$\mathrm{RMSE}_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_{ij}^{a} - x_{ij}^{o}\right)^2} \quad (3)$$

相关系数(Corr): 衡量气象要素观测值与模型分析值的线性相关程度。

$$\mathrm{Corr}_j = \frac{\sum_{i=1}^{N}\left(x_{ij}^{a} - \overline{x_j^{a}}\right)\left(x_{ij}^{0} - \overline{x_j^{0}}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_{ij}^{a} - \overline{x_j^{a}}\right)^2}\sqrt{\sum_{i=1}^{N}\left(x_{ij}^{o} - \overline{x_j^{o}}\right)^2}} \quad (4)$$

式中,N为统计时次的个数;x_{ij}^{a}为中国气象局数据中气象台站 j 在第 i 个时段的观测值;x_{ij}^{o}为模式输出数据提取的气象台站 j 在第 i 个时段的分析值;$\overline{x_j^{a}}$为分析时段内中国气象局数据中气象台站 j 观测值的平均值;$\overline{x_j^{o}}$为分析时段内模式输出数据提取的气象台站 j 分析值的平均值。

2 复核案例分析

为验证原型系统可靠性,本研究选取典型环评气象数据案例,定量评估了气象数据有效性,识别疑似造假行为。

2.1 风速、温度典型案例分析

该环评案例预测文件中气象数据整体错位 1h,并且每日 21-24 时风速、气温数据存在异常。其中风速存在 1785 个时次的偏差,占总数据的 20.4%,风速偏差的时间分布存在明显规律性,77.6% 的风速偏差出现在 21-24 时;温度存在 1739 个时次的偏差,占总数据的 19.9%,温度偏差的时间分布也存在明显规律性,64.1% 的温度偏差出现在 21-24 时。复核结论为该环评预测文件中每天 21-24 时的风速、温度数据存在人为调整的痕迹。统计指标结果见表 1。

表 1 模型气象数据和地面气象观测站实测数据对比表

因子	平均偏差	平均绝对偏差	均方根误差	相关系数
风速	-0.0005	0.2106	0.7828	0.8268
气温	-0.0542	0.6237	1.7159	0.9889

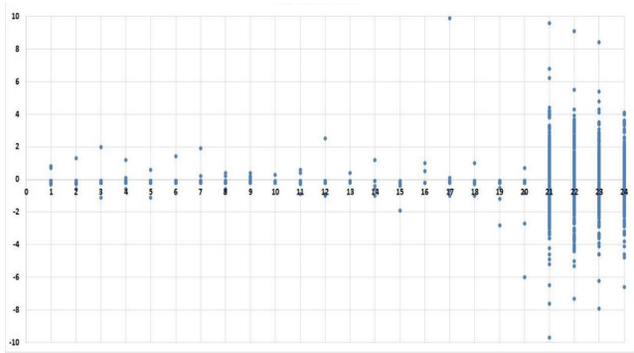


图 3 模型气象数据和地面气象观测站实测数据偏差的时间分布图

2.2 风向典型案例分析

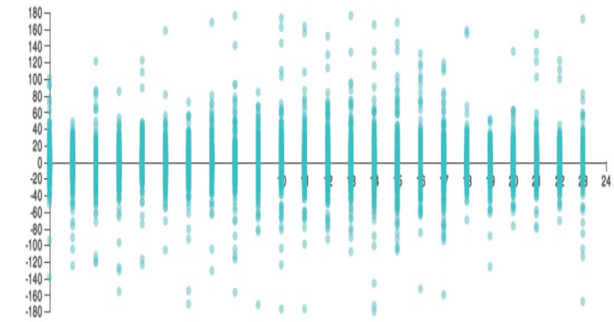


图 4 系统生成风向偏差小时分布散点图

该环评案例预测文件经系统分析,8291 个时次数据中共有 3675 个时次的风向数据出现偏差,占总数据量的 44.3%。并且两者在部分月份的风玫瑰图存在明显的差异。复核结论

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

为该环评预测文件中部分月份的风向数据存在人为调整的痕迹。

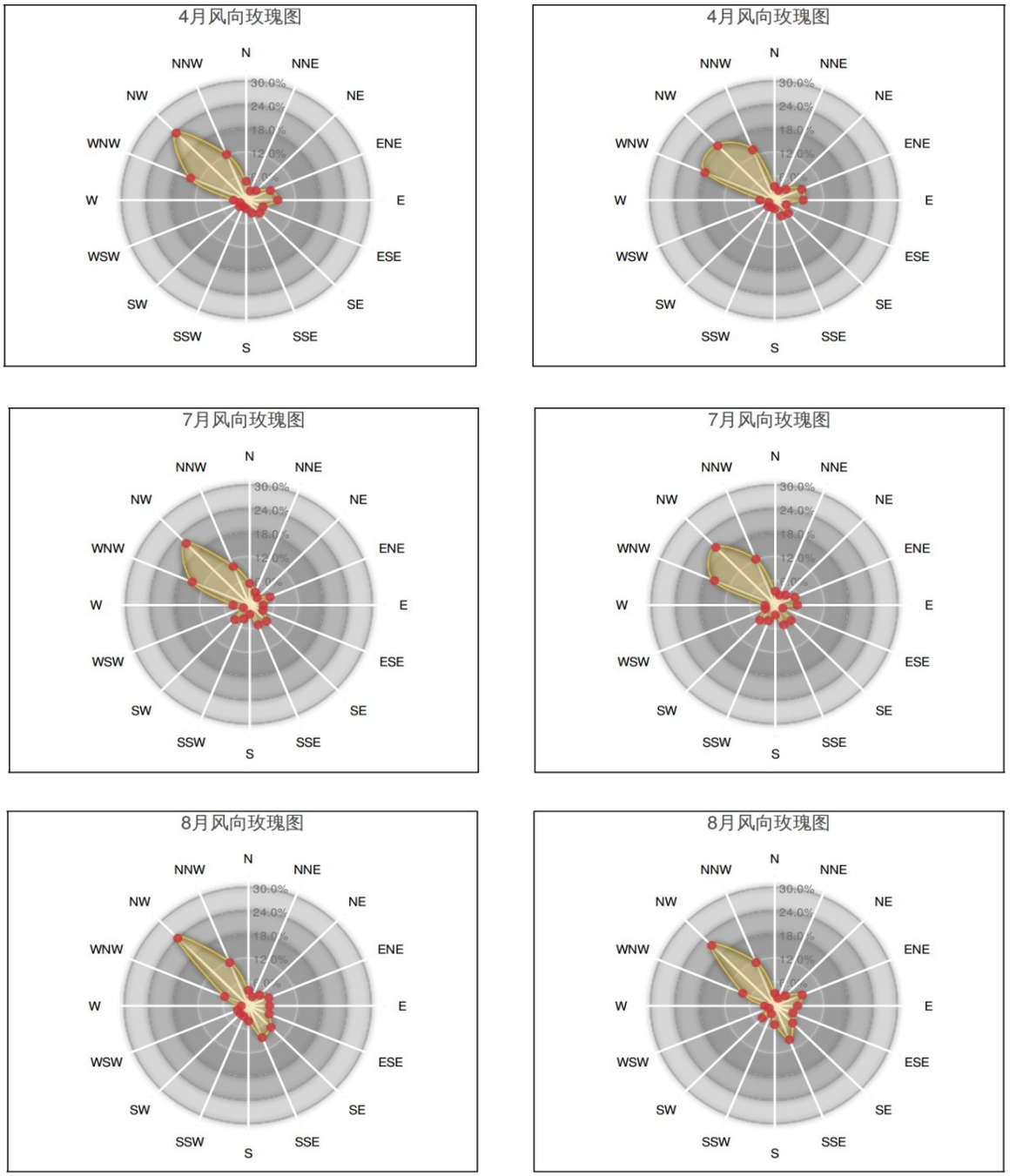


图5 模型气象数据和地面气象观测站实测数据的风玫瑰对比图

（左为模型气象数据的月风玫瑰图，右为气象观测站实测数据的月风玫瑰图）

2.3 云量、云高典型案例分析

该环评案例预测文件中云量数据经原型系统分析，1808个时次数据中共有1784个时次的低云量数据出现偏差，占总数据量的98.673%。其中，中国气象局地面气象站实测云量数据大于模型云量数据的时次有1657个，占比为92.9%；中国气象局地面气象站实测云量数据小于模型云量数据的时次有127个，占

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

比为7.1%。系统生成的低云量偏差小时分布散点图(见图6)。此外,该预测文件中的云高数据人为设定为定值808米,与实际情况不符,存在较大问题,导致结果失真。

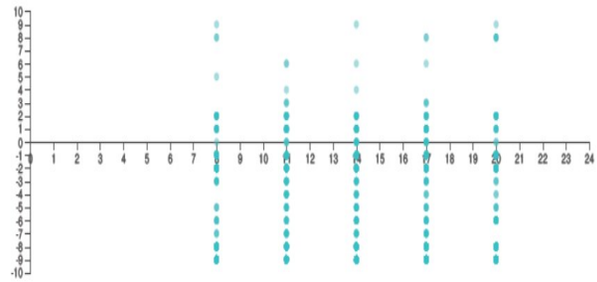


图6 云量数据复核结果

3 结论

本研究建立了基于气象大数据的环评技术复核原型系统,实际复核案例应用表明,该系统依托机器学习和大数据分析技术,突破了人工复核方法的瓶颈,提高了大气技术复核工作的效率,为环评智能复核工作提供了新的技术方法。下一步建议管理部门要求环评单位将大气预测模式输入文本文件、气象数据、地形数据、地表参数等作为四级联网上传附件内容,以备抽查、复核。

参考文献

[1]生态环境部.建设项目环境影响报告书(表)编制监督管理办法[EB/OL]. http://www.mee.gov.cn/xxgk2018/xxgk/xxgk02/201909/t20190925_735606.html, 2019-9-20.

[2]生态环境部.关于严惩弄虚作假提高环评质量的意见[EB/OL]. http://www.mee.gov.cn/xxgk2018/xxgk/xxgk03/202009/t20200923_800011.html, 2020-09-22.

[3]全国人民代表大会常务委员会.中华人民共和国刑法修正案(十一)(草案二次审议稿)[EB/OL]. http://www.npc.gov.cn/npc/xfxzasyca/xfxzasyca.shtml,2020-10-22

[4]张尚宣,伯鑫,周甜,等.AERMOD模式在我国环境影响评价应用中的标准化研究[J].环境影响评价,2018,40(02):51-55.

[5]伯鑫,王刚,田军,等.AERMOD模型地表参数标准化集成系统研究[J].中国环境科学,2015,35(09):2570-2575.

[6]伯鑫,吴忠祥,王刚,等.CALPUFF模式的标准化应用技术研究[J].环境科学与技术,2014,37(S2):530-534.

[7]伯鑫.空气质量模型:技术、方法及案例研究[M].中国环境出版集团:北京,2018:104-105.

[8]伯鑫.CALPUFF模型技术方法与应用[M].中国环境出版社:北京,2016:126-131.

[9]沈文海.气象数据的"大数据应用"浅析——《大数据时代》思维变革的适用性探讨[J].中国信息化,2014(11):20-31.

收稿日期:2021-01-31

作者简介:伯鑫(1983-),男,山东烟台人,博士,主要研究方向为大气环境质量模拟、污染源清单、环评技术复核,发表论文80余篇。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net