

基于单分类机器学习的生活垃圾焚烧发电设施选址可行性与影响因素识别：来自全国 1058 个项目的证据

杨静¹, 毛瑾欣², 赵文鑫¹, 赵芳^{1*}, 李琳¹

(1. 生态环境部环境发展中心, 北京 100029; 2. 上海复旦规划建筑设计研究院有限公司, 上海 200433)

【摘要】 如何实现生活垃圾焚烧发电设施的科学高效选址, 是其环境社会风险防范的关键问题。本文以全国 1058 个已建设施为正类样本, 整合环境、人口、社会与经济四维度的 16 项指标, 提出一套基于单分类机器学习的选址可行性评估流程。研究构建树模型算法 (IF、OCC Tree、RFDE、LOF Tree) 用于解释特征变量贡献, 并与 OC-SVM、LOF、EE、GMM、AE 等算法进行跨范式对比, 最终将表现最优的 OCC Tree、LOF 与 AE 算法集成为综合模型以提升稳健性。结果显示, 人口密度与环境投诉为主导因子, 植被指数、学校密度和地区碳排放情况具有中等影响, 运输效率对选址具有成本约束。综合模型表现出了较好预测性能, 在验证集上的 Accuracy=0.987、F1=0.994。敏感性分析表明, 在 $\pm 10\% \sim \pm 20\%$ 扰动情景下, 植被指数和环境投诉的微小变化即可引起选址适宜度得分的较大波动。研究为“邻避”设施在厂址预筛选和冲突预防阶段提供了可解释、可复用的技术路径。

【关键词】 垃圾焚烧发电设施; 单分类机器学习; 选址优化; 环境社会风险

【中图分类号】 X705; TM621.1; TP18

【文章编号】 1674-6252 (2025) 06-0028-11

【文献标识码】 A

【DOI】 10.16868/j.cnki.1674-6252.2025.06.028

引言

在“双碳”与“无废城市”战略推进下, 我国生活垃圾处理正由填埋为主转向减量化、资源化、无害化协同治理, 生活垃圾焚烧发电设施在缩减填埋规模、回收能源等方面发挥重要作用^[1,2]。然而, 垃圾焚烧设施在建设及运行阶段带来的环境排放、景观影响与运输扰动, 叠加信息不对称与风险沟通不足, 容易引发环境社会风险, 表现为“邻避”(NIMBY) 情绪、群体性异议与信任危机^[3-6], 从而推高项目落地的不确定性与治理成本。

随着我国东部及大中城市的生活垃圾焚烧发电

设施布局基本完成, 针对存量设施的选址后评估成为关键, 其评估结果可服务于设施的优化运营、迁建决策^[7]。而在中西部地区, 由于增量需求、老旧设施搬迁改造等, 厂址选择仍是规划实践中的难点。一些研究发现, 垃圾焚烧设施常布局于既有填埋场周边, 以降低经济成本^[8], 却可能加剧环境风险与社会成本的空间不公平, 因此需要在选址规划和可行性分析阶段更系统地考察环境^[9]、人口^[5,10]、社会^[11,12] 多维因素的影响。

传统研究多利用文献研究法、案例分析法、问卷调研法等开展“邻避”冲突研究^[13,14], 为垃圾

资助项目: 国家自然科学基金青年科学基金项目“基于大数据的‘邻避’设施环境社会风险的识别、预测和防控策略研究”(72204274)。

作者简介: 杨静 (1990—), 女, 博士, 高级工程师, 研究方向为生态环境大数据, E-mail: yangjing01@edcmep.org.cn。

*** 责任作者:** 赵芳 (1978—), 女, 硕士, 正高级工程师, 研究方向为环境影响评价, E-mail: zhaofang@edcmep.org.cn。

焚烧设施选址提供基础。近期选址研究广泛采用多准则决策^[15]，结合评价类数学方法（如 AHP、TOPSIS 等）与地理信息系统进行叠加分析，但往往存在权重设定主观性强、指标相关冗余、难以处理非线性耦合等局限。随着大数据与人工智能的快速发展，机器学习正逐渐被应用于国土空间规划和选址的相关研究中^[16]。然而当前实践中往往只有已建成设施点位的正样本，缺乏可验证的“未通过/被否决选址”负样本，使得二分类或有监督学习的机器学习算法难以应用。因此本研究将选址问题转化为单分类范式，在仅有正类样本的前提下学习适宜选址的地区特征，将显著偏离特征的地区判定为不适宜或可行性较低地区。这一范式已在复杂系统监测中展现出较强的能力^[17,18]，适合处理空间规划数据，且能规避负样本缺失带来的偏差风险。

基于此，本文在全国 1058 个既有设施的真实点位样本上，构建环境、人口、社会与经济四维度共 16 项输入指标，系统比较各种单分类算法的选址适宜度预测结果，其中通过树模型解释了影响选址的因素，并通过模型融合提升结果稳健性。本研究并非试图以机器学习替代现有的规划决策程序，而是为垃圾焚烧等“邻避”设施在厂址预筛选、方案比选和风险沟通阶段提供辅助决策工具。基于此，研究的贡献主要体现在：①在缺乏负样本的现实约束下，构建了一套可复用的单分类选址可行性评估流程，解决传统多准则决策方法的不足；②量化并排序环境、人口、社会与经济等因素对选址可行性的影响，将技术判别转化为可操作的规划考量^[1,3,7]。

1 数据与方法

1.1 数据来源

生活垃圾焚烧发电设施数据来自生活垃圾焚烧发电厂自动监测数据公开平台 (<https://ljgk.envsc.cn/manage/index.html>)，研究以具有独立排放监测记录的生活垃圾焚烧发电设施为统计单元。其中同一厂址的扩建分期认为是不同项目单元，仅更名或更换投资方而厂址未发生变化的认为是同一项目单元。最终获得截至 2025 年 3 月的 1058 个全国运行中的生活垃圾焚烧发电项目。

基于选址可行性与环境社会风险理论，本文从环境、人口、社会和经济四大维度构建了 16 项选址适宜度指标作为模型输入变量。指标及数据来源详见表 1，其中环境和人口数据主要来源于 2020 年度。

表 1 指标构建

序号	一级指标	二级指标	三级指标	变量
1	环境因素	空气	平均 PM _{2.5} 浓度	ENV_PM _{2.5}
2			平均 AQI 指数	ENV_AQI
3			年平均风速	ENV_WIND
4		碳排放	年排放总量	ENV_CO ₂
5		水体	河网密度	ENV_RIVD
6		生态	植被归一化指数	ENV_NDVI
7			保护区距离	ENV_PROT
8	人口因素	人口数量	人口密度	POP_DENS
9		人口风险点位密度	居民点密度	POP_SETD
10			学校密度	POP_SCH
11	社会因素	政府信任度	官员腐败指数	SOC_COR
12		环境社会负面舆情度	垃圾焚烧负面舆论比例	SOC_NEG
13			环境投诉负面数量比例	SOC_COM
14		犯罪率	万人被刑事批捕率	SOC_ARR
15			万人被刑事起诉率	SOC_IND
16	经济因素	运输成本	垃圾运输效率	ECO_ROAD

环境指标反映背景环境承载性和生态敏感性^[19]，是评价选址可行性的基础。环境指标包含空气指标、碳排放指标、水体指标和生态指标。空气指标包括平均 PM_{2.5} 浓度 (ENV_PM_{2.5})、平均空气质量指数 (Air Quality Index) (ENV_AQI) 和风速 (ENV_WIND)，其中空气质量数据来源于生态环境部环境监测总站空气质量实时发布系统 (<https://air.cnemc.cn:18007/>)，风速数据来自中国气象数据网 (<http://www.nmic.cn/>)。碳排放指标使用年排放总量 (ENV_CO₂) 表征，数据来源于欧盟委员会联合研究中心和荷兰环境评估署建立的全球大气研究排放

数据库 (Emissions Database for Global Atmospheric Research) 的年度碳排放总量数据。水体指标使用网格的河网密度 (ENV_RIVD) 表征, 数据来自地理空间数据云 (<https://www.gscloud.cn/>) 河网数据。生态指标使用植被归一化指数 (ENV_NDVI) 和保护区距离 (ENV_PROT) 表征, 其中 NDVI 数据来源于资源环境数据云 (<https://www.resdc.cn/>) 产品数据; 保护区数据来源于地理遥感生态网科学数据注册与出版系统 (www.gisrs.cn)。

人口指标用与健康风险直接相关的潜在暴露人群规模和人口结构来表示^[20]。人口指标包括人口密度 (POP_DENS) 和人口风险点位密度 [含居民点密度 (POP_SETD)、学校密度 (POP_SCH)]。其中人口密度数据来自 World Population 公里网格数据库 (<https://www.worldpop.org/>)。居民点密度数据来自地理空间数据云, 学校点位数据来自 Esri GeoScience 数据。

社会指标用以刻画当地社会信任与环境风险感知程度, 已有研究表明, 公众对政府的信任水平、历史投诉和网络舆情会显著影响“邻避”设施的社会许可与冲突概率^[11, 12]。社会指标包括政府信任度 [官员腐败指数 (SOC_COR)]、环境社会负面舆情 [含垃圾焚烧负面舆论比例 (SOC_NEG)、环境投诉负面数量比例 (SOC_COM)]、犯罪率 [(含万人被刑事批捕率 (SOC_ARR) 和万人被刑事起诉率 (SOC_IND))]。其中政府信任度使用哈佛大学官员腐败数据进行表征, 该数据含 2012—2016 年近 20 000 条腐败信息, 含所属省份、官员/落马/等级等信息。公众对于“垃圾焚烧”的负面评论数据来自网络舆情信息收集和分析; 环境投诉数据来自生态环境部 12369 数据和信访举报数据, 该数据包含了 2016—2020 年 12369 电话/微信/网络/信访举报四种投诉方式。犯罪率数据来自 2007—2016 年《中国检察年鉴》, 主要信息包含总犯罪、少年/妇女/老年人/罪犯/学生犯罪等被刑事批捕数量、被刑事起诉数量。

经济指标通过运输成本进行表征, 是厂址在经济可持续性上的重要约束条件^[11, 21]。指标构建使用到的道路数据, 其来源于开放街道地图 (open street Map) 在 2024 年更新的全国范围的路网数据。

部分垃圾焚烧设施选址实践中的重要因子, 如

与既有生活垃圾填埋场同址或邻近布局^[8], 由于在全国尺度上尚缺乏统一、可核查的空间数据, 难以在本研究中构建量化指标。因此, 本文当前并未纳入“填埋场同址”因子, 而是通过环境质量、人口分布和社会投诉等指标的综合作用间接反映相关影响。

1.2 研究方法

1.2.1 指标构建和空间处理

1.2.1.1 常规因素指标构建和处理

环境指标中, ENV_PM_{2.5}、ENV_AQI 和 ENV_WIND, 均通过对全国监测站点每日监测数据取年均值计算获取; ENV_RIVD 通过网格内河流长度和网格面积进行计算; ENV_PROT 代表到任一保护区 (包括自然保护区、文物保护单位、风景名胜区、文化遗产) 的直线距离。

人口指标中, POP_DENS 进行了重新优化计算, 取栅格周边 8 栅格平均人口密度数据; POP_SCH 通过统计网格内学校数量进行指标计算。

社会指标中, SOC_COR 等于各省份腐败信息等级总和的年均值; SOC_NEG 等于各省份的负面评论占总评论的比例; SOC_COM 等于投诉记录数量占人口数量的比例; SOC_ARR 和 SOC_IND, 结合各省份被刑事批捕数量、被刑事起诉数量、各省份人口数量进行计算。

除 ENV_CO₂、ENV_NDVI、POP_SETD 等部分指标为原始数据未做处理外, 其余变量均通过 IDW 插值获得全国统一分辨率的栅格数据。

1.2.1.2 经济因素指标构建和处理

研究构建了垃圾运输效率指标 (ECO_ROAD) 来表征运输成本, 其基于潜在选址点 p 到垃圾收集点 c 的距离及垃圾产生量 (以人口数量代替) 计算。

研究以县域为服务单元, 选取建设用地栅格的中心作为潜在选址点 p , 对每个县随机采样不超过 30 个点, 并优先选取空间间距大于 5km 的点, 以确保样本点的空间覆盖性和代表性。根据规划规范, 将社区或行政村中心点指定为垃圾收集点 c 。构建指标 ECO_ROAD (p) 如下:

$$ECO_ROAD(p) = \frac{1}{|C_p|} \sum_{c \in C_p} \frac{Pop(c)}{d(p,c)} \quad (1)$$

其中, C_p 是与 p 同一县的收集点集合; $\text{Pop}(c)$ 是与 c 关联的行政村或社区的总人口; $d(p,c)$ 为选址点 p 到垃圾收集点 c 的最短路径, 使用 Dijkstra 算法计算。

为获得覆盖所有合适建设用地栅格的连续运输成本层, 使用 IDW 方法将采样点的运输成本指标值插值到空间每个栅格单元。

1.2.2 机器学习框架和评估

1.2.2.1 预处理与特征筛选

为消除指标量纲不一致和数值范围差异显著问题, 依据分布特征, 对于分布相对均匀连续型指标及呈现偏态分布与零膨胀特征的指标, 研究分别采用 IQR 方法和四分位方法进行归一化, 将指标转换到 $[0,1]$ 范围。随后使用空间连接获取 1058 个已建设点位中所有指标的特征值。在此基础上, 使用皮尔逊相关系数对 16 个候选指标进行共线性分析, 以 $|r| \geq 0.5$ 作为较强相关的阈值, 保留解释性更强或数据质量更高的变量。

1.2.2.2 机器模型构建

本研究将 1058 个样本按 7:3 随机划分为训练集和验证集。对于模型输入, 单个点位的特征向量为 $\mathbf{x} \in R^m$, 其中 m 为基于表 1 筛选的核心指标。样本集合 $\mathbf{D} = \{\mathbf{x}_i\}_{i=1}^N$, $N=1058$, 且 $y_i=1$ 。模型输出为适宜度分数, $S(\mathbf{x}) \in [0,1]$, 数值越大, 选址适宜度和可行性越高, 表示该点位的特征越接近已建厂平均水平。 $S(\mathbf{x})$ 可作为厂址预筛选和风险分级的辅助量化参考, 不直接决策是否允许建设。

(1) 树模型构建。为实现因子解释, 研究首先使用基于树的单分类模型进行选址适宜度的特征重要性分析, 包括孤立森林 (Isolation Forest, IF)、一类分类树 (One-Class Classification Tree, OCC Tree)、随机密度森林 (Random Forest Density Estimation, RFDE) 和局部异常因子树 (Local Outlier Factor Tree, LOF Tree), 这些模型通过分裂纯度和信息增益来量化特征贡献度。

(2) 其他单分类模型构建。为系统对比不同单分类范式, 本文选取其他五种不同机理的代表性算法, 包括一类支持向量机 (One-Class Support Vector Machine, OC SVM) [22, 23]、局部异

常因子 (Local Outlier Factor, LOF) [23,24]、椭圆包络 (Elliptic Envelope, EE) [25,26]、高斯混合模型 (Gaussian Mixture Model, GMM) [27,28]、自编码器 (Autoencoder, AE) 开展分析 [23,25]。OC SVM 在核特征空间中学习将样本与原点分离的超平面, 以最大化间隔估计正类支持域 [22,23]。

(3) 模型验证评估。研究分别以 0.3、0.5、0.7 的适宜度得分为阈值标准, 以判断某点位是否适宜选址或具备选址可行性。以阈值 0.5 为例, 当验证集点位预测的适宜度得分大于 0.5 时, 认为其适宜选址。基于该点位是否适宜选址的判断和实际已选址结果, 计算模型的准确率 (Accuracy)、精确率 (Precision) 和召回率 (Recall) 和 F1 分数等指标, 对模型性能进行验证评估。

(4) 综合模型构建。为提高稳健性和泛化能力, 减少单一模型的偏差, 选择以上性能较优的若干模型, 将单模型输出的适宜分数通过加权计算生成综合分数, 计算如下:

$$S_{\text{集成}}(\mathbf{x}) = \sum_{i=1}^n [i_n S_n(\mathbf{x})] \quad (2)$$

其中, i 为各模型的权重, 其依据 F1、Recall 及预测适宜性分数的标准差 (Std) 三个指标进行量化, 初始加权评分 q_n 计算如下, q_n 归一化得到最终权重 ω_i :

$$q_n = 0.3 \times F1_n + 0.4 \times \text{Recall}_n + 0.3 \times \text{Std}_n \quad (3)$$

研究的技术路线, 即选址影响因素识别和可行性预测的流程如图 1 所示。

2 结果与讨论

2.1 数据的描述性统计

基于 1058 个样本的 16 项指标描述性统计如表 2 所示, 结果呈现不同的分布特征, 归一化后的各特征分布如图 2 所示。

环境因素方面, $\text{ENV_PM}_{2.5}$ 均值为 $33.76 \mu\text{g}/\text{m}^3$, ENV_AQI 均值为 61.30, 两者均表现出明显左偏; 气象条件相对稳定, 年平均风速均值 4.83 m/s 。部分变量因自然条件分布不均, 存在零膨胀特征, 如 ENV_CO_2 均值为每 $0.1^\circ \times 0.1^\circ$ 网格 461.91t , 但四分位距为 $0 \sim 11.65\text{t}$, 最大值高达 627 万 t, 表现出明显右偏; 如 ENV_RIVD 均值 $0.23 \text{ m}/\text{km}^2$, 四

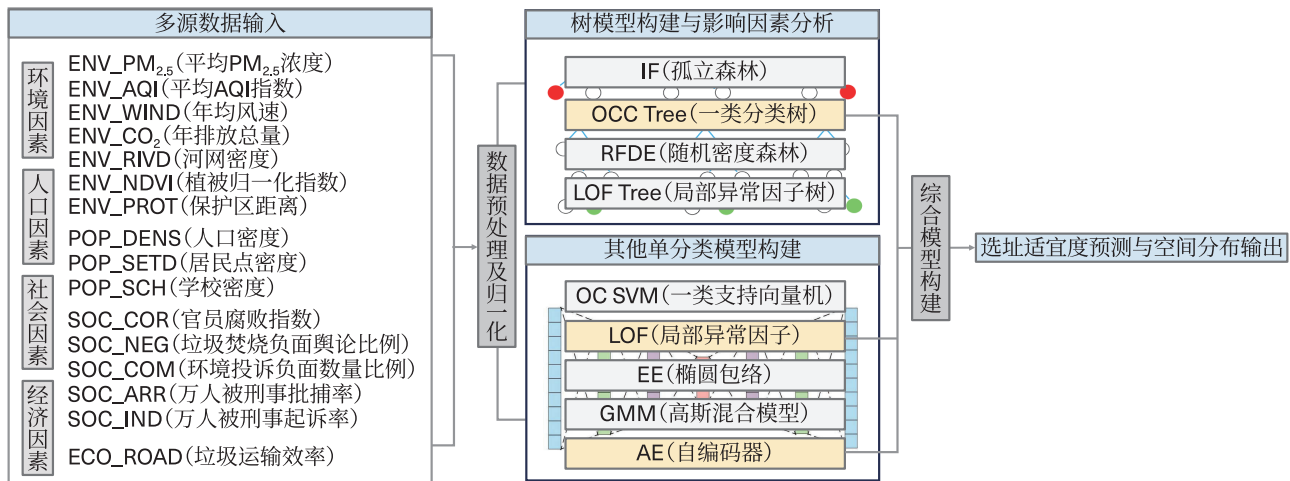


图1 选址影响因素识别和可行性预测流程

表2 原始变量的描述性统计

变量	平均值	标准差	最小值	25%分位	50%分位	75%分位	最大值
ENV_PM _{2.5}	33.76	19.30	4.96	21.70	29.88	40.80	186.05
ENV_AQI	61.30	29.54	22.49	40.26	54.08	71.45	228.00
ENV_WIND	4.83	0.75	2.72	4.31	4.85	5.35	8.34
ENV_CO ₂	461.91	30 036.20	0	0	2.99	11.65	6 274 350
ENV_RIVD	0.23	0.27	0	0	0.14	0.31	1.19
ENV_NDVI	0.55	0.33	0	0.19	0.69	0.85	0.92
ENV_PROT	33.21	31.00	0	11.46	22.97	45.12	245.31
POP_DENS	396.54	704.26	0	85.48	215.25	462.26	75 371.78
POP_SETD	0.01	0.02	0	0	0	0	1.00
POP_SCH	0.05	0.36	0	0	0	0	28.92
SOC_COR	110.80	91.34	0	60.90	75.00	162.00	356.50
SOC_NEG	0.05	0.05	0	0.03	0.04	0.04	0.18
SOC_COM	166.87	227.62	0	55.17	85.40	208.54	1 418.06
SOC_ARR	6.57	1.85	0	5.19	5.91	8.34	13.38
SOC_IND	9.00	2.67	0	6.86	8.47	10.73	20.33
ECO_ROAD	201.46	667.57	0.16	36.57	97.19	221.93	526 581.44

分位距为 0~0.31m/km²。ENV_NDVI 均值 0.55；ENV_PROT 均值 33.21 m，最大 245.31 m，呈长尾分布。人口因素方面，零膨胀特征明显，且人口在大城市周边显著升高，POP_SETD 与 POP_SCH 呈高度聚集，数据分布呈现极端右偏特征。其中 POP_DENS 均值为 396.54 人/km²，最大 7.54 万人/km²；POP_SETD 与 POP_SCH 中位数均为 0，但最大分别达 1.00 与 28.92。社会因素方面，

SOC_COR 分布较为均匀，均值 110.80。SOC_NEG 和 SOC_COM 在部分地区明显偏高，存在明显的空间集聚，其中 SOC_NEG 均值 0.05，SOC_COM 均值 166.87 例/万人，二者左偏明显。万人被刑事批捕率 SOC_ARR 与起诉率 SOC_IND 分别为 6.57 例/万人与 9.00 例/万人，波动相对较小。经济因素方面，垃圾运输效率 ECO_ROAD 呈现明显右偏，均值 201 人/km，最大 52.66 万人/km。

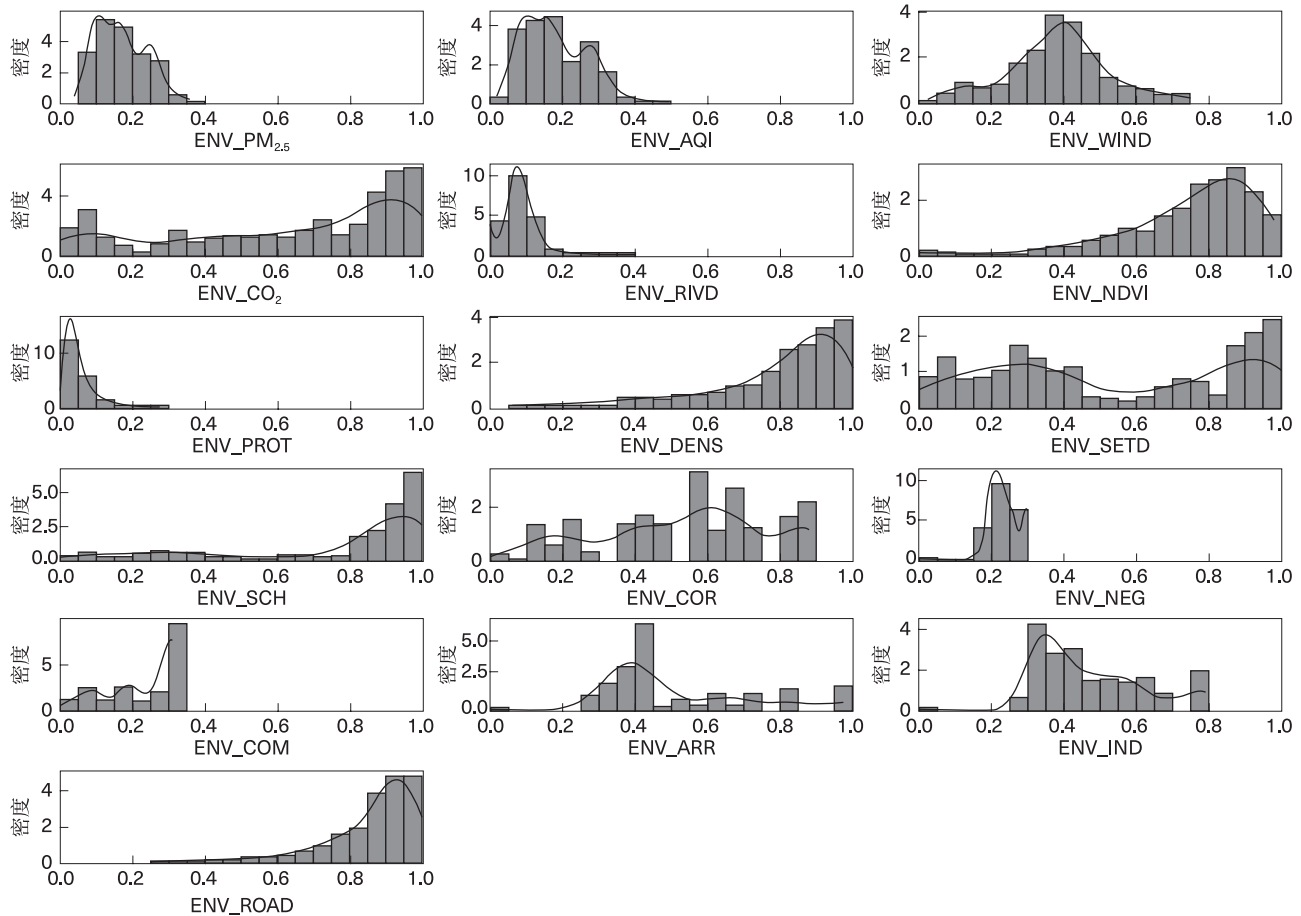


图 2 归一化的特征因素分布情况

2.2 特征选择和特征重要度评估

特征筛选使用的皮尔逊系数热图如图 3 所示，结果表明，ENV_PM_{2.5} 与 ENV_AQI 相关性极高 ($r \approx 0.98$)，研究保留了综合性更高的 ENV_AQI；SOC_ARR 与 SOC_IND 均表征犯罪率，相关性较高 ($r \approx 0.92$)，为降低冗余予以同时剔除；SOC_NEG 与 SOC_COM 均刻画社会负面反应，考虑数据可核验性，保留信访投诉类的 SOC_COM。最终从 16 项变量中移除 4 个冗余特征 (ENV_PM_{2.5}、SOC_NEG、SOC_ARR、SOC_IND)，形成 12 维核心建模特征集。

基于保留的 12 项指标，对四类树模型在验证集 ($n=318$) 上的预测性能进行比较。结果表明 (表 3)，OCC Tree 在各阈值下整体优于 IF、RFDE 和 LOF Tree，Accuracy、Precision、Recall 和 F1 均保持在较高水平。

据此选择 OCC Tree 开展特征重要度分析，结

果如表 4 所示。可以看出，人口暴露在模型决策中的权重远高于其他指标，人口密度 (POP_DENS) 权重最高 (0.373)，表明减少人口暴露是选址可行性的最核心约束。环境投诉 (SOC_COM, 0.185) 次之，表明其对选址成败也具有决定性作用，这与学者“技术合规—社会许可”机制相呼应^[6]，即在技术合规之外，通过透明、共享与公平程序方能获得持续认可。植被指数 (ENV_NDVI, 0.136)、学校密度 (POP_SCH, 0.094) 具有中等影响，分别对应生态敏感性、易感人群保护的选址约束。碳排放 (ENV_CO₂, 0.072) 也具有中等影响，代表碳排放高值区域的监管可能更严。交通运输效率 (ECO_ROAD, 0.064) 贡献适中，体现成本对选址具有一定的限制。空气质量 (ENV_AQI, 0.050) 与保护区距离 (ENV_PROT, 0.026) 影响相对较弱但不可忽视。其余指标在 OCC Tree 下的重要性接近于 0，表明在控制上述关键因子后其他变量的边际解释力有限。

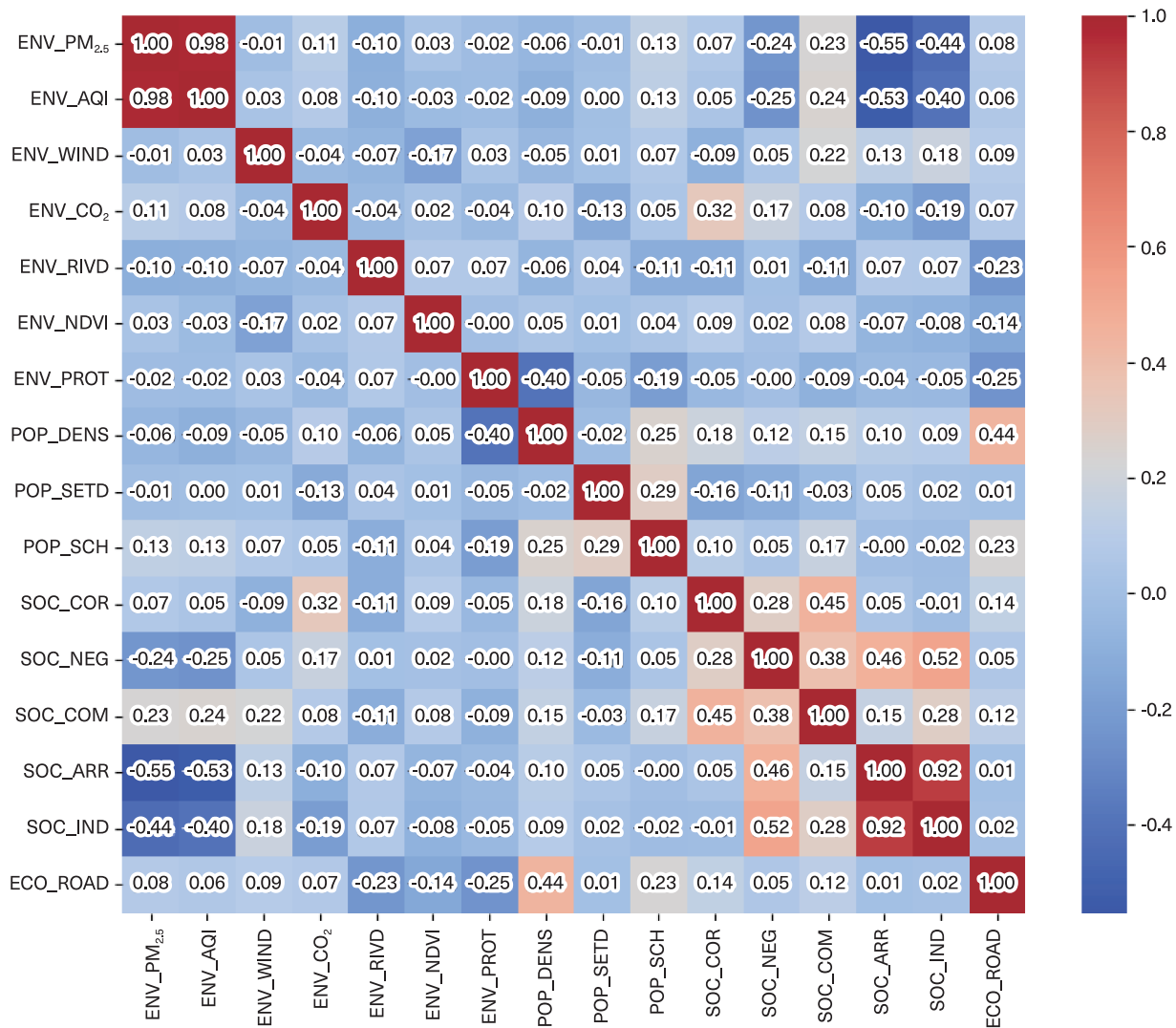


图3 特征相关性热图

表3 各树模型的评价结果

國值	模型名称	准确率 Accuracy	精确率 Precision	召回率 Recall	F1
0.3	IF	0.97	1.00	0.97	0.99
	OCC Tree	1.00	1.00	1.00	1.00
	RFDE	0.63	1.00	0.63	0.77
	LOF Tree	0.98	1.00	0.98	0.99
0.5	IF	0.90	1.00	0.90	0.95
	OCC Tree	0.99	1.00	0.99	1.00
	RFDE	0.54	1.00	0.54	0.70
	LOF Tree	0.92	1.00	0.92	0.96
0.7	IF	0.60	1.00	0.60	0.75
	OCC Tree	0.85	1.00	0.85	0.92
	RFDE	0.47	1.00	0.47	0.64
	LOF Tree	0.68	1.00	0.68	0.81

表4 OCC Tree的特征相关分析结果

排名	特征名称	重要性得分
1	POP_DENS	0.373
2	SOC_COM	0.185
3	ENV_NDVI	0.136
4	POP_SCH	0.094
5	ENV_CO ₂	0.072
6	ECO_ROAD	0.064
7	ENV_AQI	0.050
8	ENV_PROT	0.026

2.3 综合预测模型结果

其他五种非树模型单分类算法的评估结果如

表 5 所示。在给定的三档适宜度阈值 0.3/0.5/0.7 下，AE 与 LOF 综合表现最佳，显著优于 OC SVM、EE 与 GMM。以阈值 0.5 为例，AE 的 Accuracy/Precision/Recall/F1 分别为 0.99/1.00/0.99/1.00，LOF 为 0.94/1.00/0.94/0.97，均显著优于 OC SVM (0.82/1.00/0.82/0.90)，此外 EE 与 GMM 表现则更弱。

据此，结合前述树模型评估中表现最优的 OCC Tree，将 AE、LOF 与 OCC Tree 进行集成形成综合模型。综合模型在阈值 0.5 下取得 Accuracy=0.99、Precision=1.00、Recall=0.99、F1=0.99，即识别为适宜选址的比例为 0.99；当阈值降至 0.3 时，四项指标进一步提升至 1.00/1.00/1.00/1.00；当阈值提高到 0.7 时，仍保持较高的综合性能，四项指标为 0.94/1.00/0.94/0.97。图 4 比较了单一模型与综合模型的适宜度评分分布，横轴为选址适宜度评分 $S(x)$ ，纵轴为不同评分对应的预测项目频数。图 4a 显示，AE、LOF 与 OCC Tree 三种模型在大部分样

表 5 各模型的评价结果

阈值	模型	准确率 Accuracy	精确率 Precision	召回率 Recall	F1
0.3	OC SVM	0.95	1.00	0.95	0.98
	LOF	0.99	1.00	0.99	0.99
	EE	0.19	1.00	0.19	0.32
	GMM	0.00	0.00	0.00	0.00
	AE	1.00	1.00	1.00	1.00
	OCC Tree	1.00	1.00	1.00	1.00
0.5	OC SVM	0.82	1.00	0.82	0.90
	LOF	0.94	1.00	0.94	0.97
	EE	0.08	1.00	0.08	0.16
	GMM	0.00	0.00	0.00	0.00
	AE	0.99	1.00	0.99	1.00
	OCC Tree	0.99	1.00	0.99	1.00
0.7	OC SVM	0.34	1.00	0.34	0.50
	LOF	0.81	1.00	0.81	0.89
	EE	0.03	1.00	0.03	0.06
	GMM	0.00	0.00	0.00	0.00
	AE	0.97	1.00	0.97	0.98
	OCC Tree	0.85	1.00	0.85	0.92

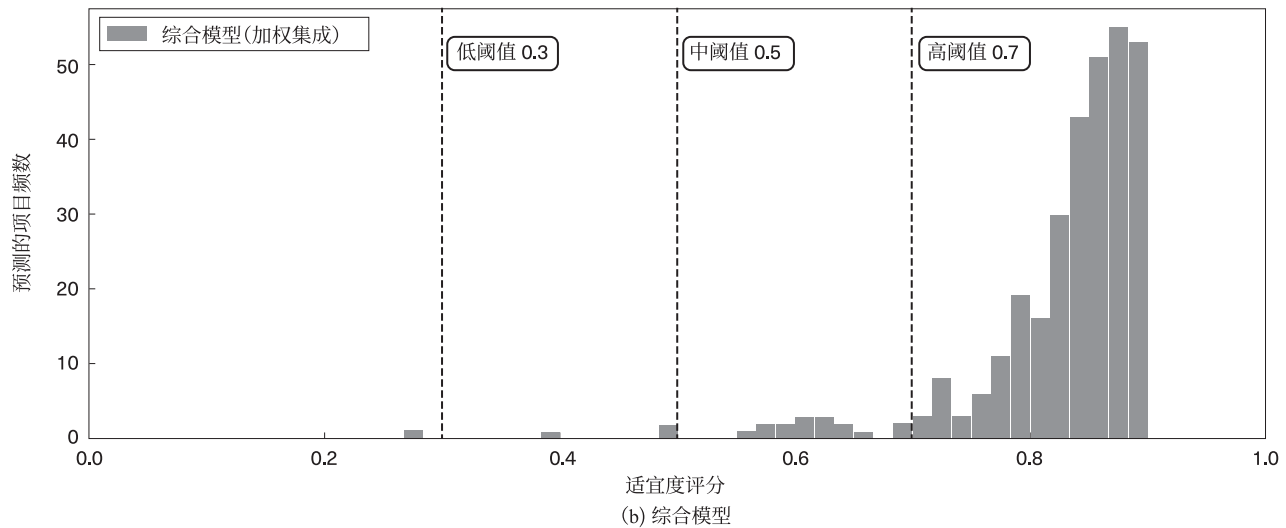
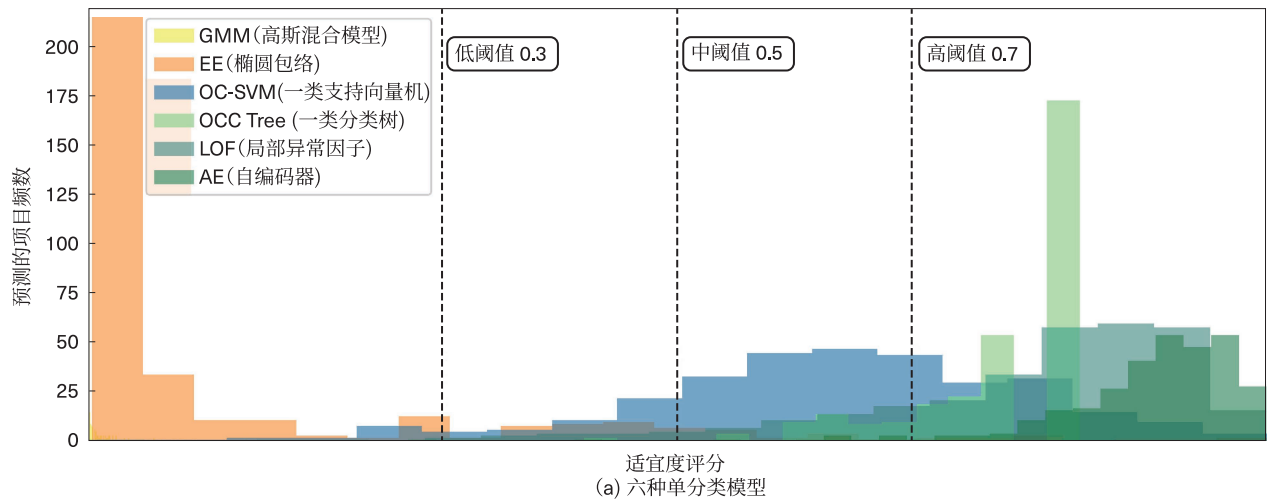


图 4 不同模型的适宜度评分直方图分布

本上均给出偏高适宜度的评分。图 4b 所示的综合模型，主体样本选址适宜度评分预测值主要分布在 0.6~0.9，相较单一模型，综合模型具有更好的分数分布稳定性与判别一致性，为后续的选址适宜度与可行性分析提供了可靠的评分基准。

2.4 模型敏感性分析

为评估关键因子变化对选址可行性评估结果的影响，本研究开展了单因子敏感性分析，即在保持其他特征不变的前提下，分别对每个特征因子施加 ±10% 和 ±20% 的扰动比例，并通过比较扰动前后平均适宜度评分的变化量 $\Delta S(x)$ ，计算相对扰动幅度。

如图 5 所示，植被指数 (ENV_NDVI) 对模型输出最为敏感，其 ±20% 扰动引起的平均适宜度变化在 0.005 量级，影响的相对幅度约 -0.68%~0.57%。环境投诉 (SOC_COM) 次之，影响的相对幅度约 -0.57%~0.42%。保护区距离 (ENV_PROT) 和空气质量 (ENV_AQI) 均具有一定影响。人口密度 (POP_DENS) 虽然在特征重要度中排名第一 (表 4)，但其 ±20% 扰动带来的平均适宜度变化略低于上述

变量。

敏感性分析结果表明，人口密度主要决定区域整体适合度，但其对整体扰动较为稳健。生态背景和社会投诉的变化则具有更明显的“放大效应”。

3 结论

3.1 主要结论

本研究在全国 1058 个已建生活垃圾焚烧发电项目的基础上，构建了适用于负样本缺失情景的单分类选址可行性评估与影响因素识别流程。研究通过多源数据整合和相关性筛选形成了 12 项核心特征，通过系统比较各典型单分类算法，将 OCC Tree、LOF 与 AE 进行加权集成，形成综合适宜度评分，综合模型表现出良好的稳健性和判别一致性。

影响机制上，人口密度与环境投诉为主导因子，植被指数、学校密度和碳排放具有中等影响；运输效率具有一定成本约束作用；空气质量与保护区距离影响相对较弱但不可忽视。敏感性分析进一步表明，在 ±10%~±20% 扰动下，植被指数和

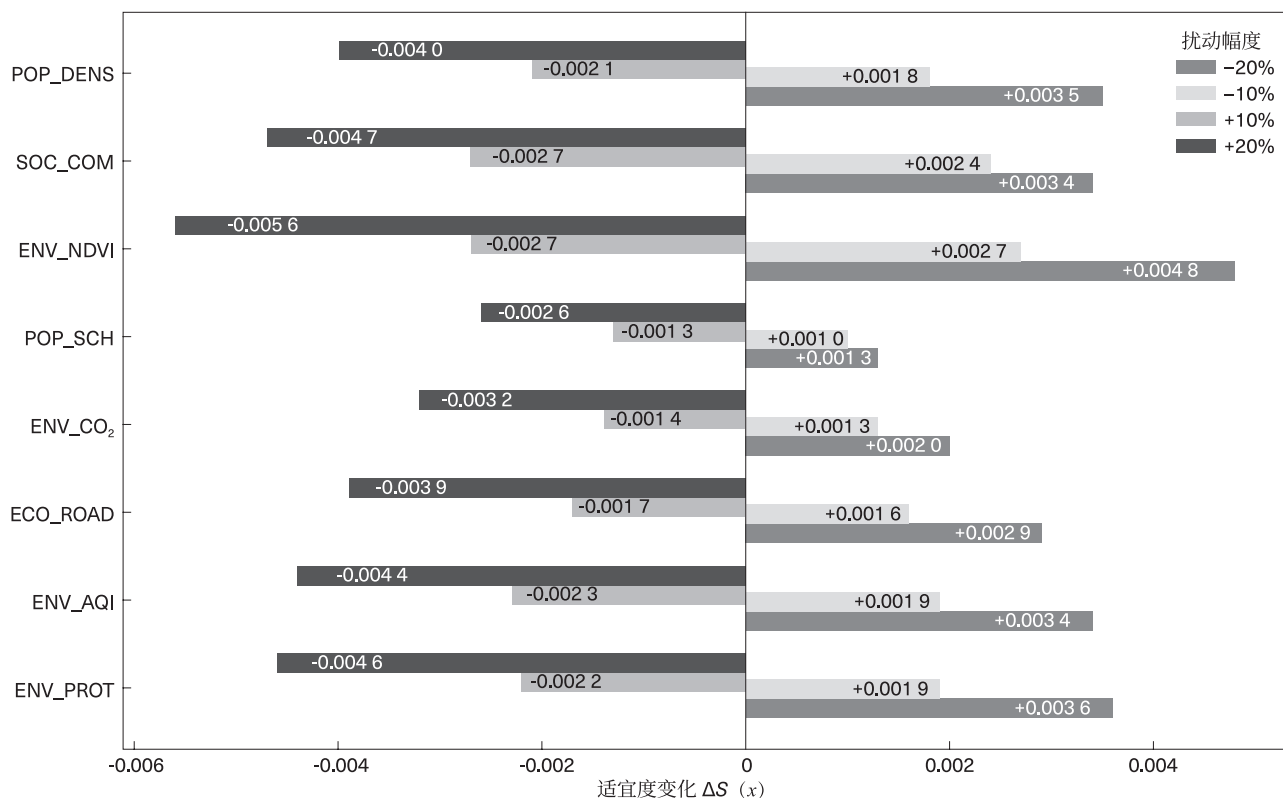


图 5 不同因子扰动的适宜度变化情况

环境投诉的微小变化可能引起适宜度评分的较大波动，而人口密度虽最重要但对选址扰动更为平稳。

3.2 政策启示

一是方法与流程的应用。一方面本方法流程可在城市垃圾焚烧设施规划布局中，快速生成适宜度和可行性分布图，用于预筛选和早期比选。另一方面对于存量设施的搬迁改造，可用于识别现有厂址的环境社会风险程度，并为替代厂址提供定量参考。最后研究可为类似“邻避”环境设施提供可迁移的方法框架。

二是影响因素的选址启示。人口密度的重要性表明减少人口暴露应为刚性约束。环境投诉的高权重强调风险沟通应该前置，可在预选址阶段接入信访与舆情的早期信号，针对性开展防范化解措施以保证选址落地。植被指数等生态背景虽非首要影响因素，其扰动却对选址适宜度具有放大效应，在退耕还林、生态修复等导致植被环境快速变化的地区，应结合新数据开展风险的滚动评估。

3.3 研究展望

一是数据时效与代表性。部分社会指标存在时间滞后问题（如官员腐败指数、犯罪率数据等），可能低估快速变化区域的社会风险。后续需引入最新年度的统一口径数据，并开展更新分析。

二是负样本缺失与外推风险。单分类范式虽适配现状，但其学习的“可行性特征”隐含了既有决策和制度环境，在向条件差异较大地区或未来情景外推时需谨慎。后续可采集被否决/搁置、发生群体性事件的选址案例作为反向样本，开展二分类/多分类的监督学习。

三是选址影响因子的扩展。目前模型尚未纳入“与既有填埋场同址/邻近布局”等情景因子，未来可尝试获取历史填埋场空间数据，将其纳入指标体系。

四是成果应用。受限于最新指标数据缺失，本文未开展全国栅格化的适宜度预测。未来在数据更新后，可开展全国尺度的适宜度栅格化评估和生成分级清单，服务垃圾焚烧设施的增量布局、老旧设施迁改等。

参考文献

- [1] KURNIAWAN T A, LIANG X, O' CALLAGHAN E, et al. Transformation of solid waste management in China: moving towards sustainability through digitalization-based circular economy[J]. Sustainability, 2022, 14(4): 2374.
- [2] LEE R P, MEYER B, HUANG Q L, et al. Sustainable waste management for zero waste cities in China: potential, challenges and opportunities[J]. Clean energy, 2020, 4(3): 169-201.
- [3] BRUNNER P H, MORF L S. Waste to energy, indispensable cornerstone for circular economy: a mini-review[J]. Waste management & research, 2025, 43(1): 26-38.
- [4] CHEN J B, HE X Y, SHEN Y, et al. Demographic differences in public acceptance of waste-to-energy incinerators in China: high perceived stress group vs. low perceived stress group[J]. Frontiers in psychology, 2022, 13: 948653.
- [5] HUANG Y S, ZHANG Z Q, ZHANG Y B, et al. Perceptual differences in the factors of local acceptance of waste incineration plant[J]. Frontiers in psychology, 2022, 13: 1067886.
- [6] XU M, LIU Y, CUI C Y, et al. Social acceptance of NIMBY facilities: a comparative study between public acceptance and the social license to operate analytical frameworks[J]. Land use policy, 2023, 124: 106453.
- [7] TRAVEN L. Sustainable energy generation from municipal solid waste: a brief overview of existing technologies[J]. Case studies in chemical and environmental engineering, 2023, 8: 100491.
- [8] 高军波, 乔伟峰, 刘彦随, 等. 超越困境: 转型期中国城市邻避设施供给模式重构——基于番禺垃圾焚烧发电厂选址反思[J]. 中国软科学, 2016(1): 98-108.
- [9] 赵曦, 李娟, 张文忠, 等. 垃圾焚烧厂污染控制水平评价指标体系研究[J]. 环境科学与技术, 2016, 39(5): 162-168.
- [10] LIU Y, CUI C Y, ZHANG C Q, et al. Effects of economic compensation on public acceptance of waste-to-energy incineration projects: an attribution theory perspective[J]. Journal of environmental planning and management, 2021, 64(9): 1515-1535.
- [11] 高建伟, 孟琪琛, 刘江涛. 基于概率多值中智集和 GIS 的大型环境类邻避设施两阶段规划选址[J]. 环境工程, 2023, 41(12): 256-263, 295-295.
- [12] 郭少青. 环境邻避的冲突原理及其超越——以双重博弈结构为分析框架[J]. 城市规划, 2019, 43(2): 109-118.
- [13] 唐祖君. 生活垃圾焚烧发电厂选址方法研究——以福建省长汀县为例[D]. 南京: 南京大学, 2019.
- [14] 张利周. 邻避设施选址公众接受度及影响因素研究——以东莞和深圳的垃圾焚烧厂为例[J]. 洛阳师范学院学报, 2020, 39(5): 86-94.

- [15] 陈君丽, 岳隽, 单樑. 跨地市垃圾焚烧发电设施规划选址决策方法及应用 [J]. 城乡规划, 2023(6): 64-73.
- [16] 钟振远, 张晴, 田湘攸, 等. 基于机器学习的现代城市生态空间规划预测研究 [J]. 粘接, 2025, 52(11): 190-193.
- [17] PANG G S, SHEN C H, CAO L B, et al. Deep learning for anomaly detection: a review[J]. ACM computing surveys (CSUR), 2021, 54(2): 38.
- [18] RUFF L, KAUFFMANN J, VANDERMEULEN R A, et al. A unifying review of deep and shallow anomaly detection[J]. Proceedings of the IEEE, 2021, 109(5): 756-795.
- [19] 何晶晶, 陈淼, 杨娜, 等. 我国生活垃圾焚烧发电过程中温室气体排放及影响因素——以上海某城市生活垃圾焚烧发电厂为例 [J]. 中国环境科学, 2011, 31(3): 402-407.
- [20] 刘思君, 靖常峰, 杜明义, 等. GIS 多准则的城市垃圾楼空间布局优化选址 [J]. 测绘科学, 2018, 43(8): 45-49.
- [21] 张卫凌, 刘涛. 基于改进遗传算法的生活垃圾转运站选址问题 [J]. 现代城市研究, 2022(8): 92-98, 105-105.
- [22] LENZ O U, PERALTA D, CORNELIS C. Optimised one-class classification performance[J]. Machine learning, 2022, 111(8): 2863-2883.
- [23] STRANI L, COCCHI M, TANZILLI D, et al. One class classification (class modelling): state of the art and perspectives[J]. TrAC trends in analytical chemistry, 2025, 183: 118117.
- [24] PAOLINI D, DINI P, SOLDAINI E, et al. One-class anomaly detection for industrial applications: a comparative survey and experimental study[J]. Computers, 2025, 14(7): 281.
- [25] AGYEMANG E. Anomaly detection using unsupervised machine learning algorithms: a simulation study[J]. Scientific African, 2024, 26: e02386.
- [26] CHEN Y L, WANG S C, ZHAO Q Y, et al. Detection of multivariate geochemical anomalies using the bat-optimized isolation forest and bat-optimized elliptic envelope models[J]. Journal of earth science, 2021, 32(2): 415-426.
- [27] HUANG J Z, YANG J, LI D S, et al. A novel damage detection method based on sequential iteration and Gaussian mixture model for structural health monitoring under environmental effects[J]. Scientific reports, 2025, 15(1): 23823.
- [28] LI X, MILLER D J, XIANG Z, et al. BIC-based mixture model defense against data poisoning attacks on classifiers: a comprehensive study[J]. IEEE transactions on knowledge and data engineering, 2024, 36(8): 3697-3711.

One-Class Machine Learning-Based Feasibility Assessment and Factor Identification for Siting Waste-to-Energy Incineration Facilities: Evidence from 1058 Projects in China

YANG Jing¹, MAO Jinxin², ZHAO Wenxin¹, ZHAO Fang^{1*}, LI Lin¹

(1.Environmental Development Center of the Ministry of Ecology and Environment, Beijing,China;
2.Shanghai Fudan Planning and Architectural Design Institute Co., Ltd., Shanghai, China)

Abstract The scientific and efficient siting of waste-to-energy (WtE) incineration facilities has become a key issue for preventing related environmental and social risks. Using 1058 existing projects nationwide as positive samples, this study integrates 16 indicators across four dimensions—environment, population, society, and economy—and proposes a siting feasibility assessment workflow based on one-class machine learning. Tree-based algorithms (IF, OCC Tree, RFDE, LOF Tree) are constructed to interpret feature contributions and are compared across paradigms with OC-SVM, LOF, EE, GMM, and AE. The best-performing models, OCC Tree, LOF, and AE, are ultimately ensembled into a composite model to enhance robustness. The results show that population density and environmental complaints are the dominant factors, while vegetation index, school density, and regional carbon emissions have moderate impacts; transportation efficiency imposes cost constraints on siting. The ensemble model exhibits strong predictive performance, achieving an accuracy of 0.987 and an F1 score of 0.994 on the validation set. Sensitivity analysis indicates that under $\pm 10\% \sim \pm 20\%$ perturbation scenarios, small changes in vegetation index and environmental complaints can lead to substantial fluctuations in siting feasibility scores. The study provides an interpretable and reusable technical pathway for early-stage site screening and conflict prevention of NIMBY-type facilities.

Keywords: waste-to-energy incineration facilities; one-class machine learning; site selection optimization; environmental and social risk