

DOI:10.3876/j.issn.1000-1980.2026.05.019

干散货港区 PM_{2.5} 浓度多变量混合预测模型

陶凯旋¹, 封学军¹, 季皆名¹, 涂傲辉¹, 林志端²

(1. 河海大学港航物流与绿色发展研究所; 2. 中交水运规划设计院有限公司)

摘要: 为实现干散货港区 PM_{2.5} 浓度的精准预测与有效管控,提出了一种融合 XGBoost 特征筛选与粒子群算法(PSO)智能优化的支持向量回归(SVR)多变量混合预测模型(XGBoost-PSO-SVR 模型)。该模型充分考虑了港口 PM_{2.5} 形成机制与气象条件下扩散过程的差异,通过 XGBoost 量化关键影响因子,并利用 PSO 优化后的 SVR 进行精准预测。长江南京段典型港区的实例验证结果表明,XGBoost-PSO-SVR 模型在高浓度区间具有较强的鲁棒性,与单变量 STL-ARIMA-SVR 模型相比,其平均绝对误差、平均绝对百分比误差和均方根误差分别降低 15.6%、14.3%和 24.9%;与传统多变量 SVR 模型相比,分别降低 52.5%、57.1%和 37.7%;XGBoost-PSO-SVR 模型具有较好的预测性能,可为干散货港区智慧化抑尘提供决策支撑。

关键词: 干散货港区; PM_{2.5}; 智慧抑尘; 多变量预测; 混合模型; XGBoost

中图分类号: X736.1

文献标志码: A

文章编号: 1000-1980(2026)05-0164-09

Multivariate hybrid prediction model for PM_{2.5} concentration in dry bulk ports

Tao Kaixuan¹, Feng Xuejun¹, Ji Jieming¹, Tu Aohui¹, Lin Zhiduan²

(1. Institute of Maritime Logistics and Green Development, Hohai University; 2. CCCC Water Transportation Consultants Co., Ltd.)

Abstract: To achieve accurate prediction and effective control of PM_{2.5} concentration in dry bulk port areas, a multivariate hybrid prediction model based on support vector regression (SVR) integrating XGBoost feature selection and particle swarm optimization (PSO) intelligent optimization was proposed, namely XGBoost-PSO-SVR model. The differences in PM_{2.5} formation mechanisms and diffusion processes under meteorological conditions in ports were fully considered by the model. XGBoost was employed to quantify the key influencing factors, and the SVR optimized by PSO was utilized for accurate prediction. The case verification results of typical port areas in the Nanjing section of the Yangtze River show that the XGBoost-PSO-SVR model has strong robustness in the high-concentration range; compared with the univariate STL-ARIMA-SVR model, its MAE, MAPE, and RMSE decrease by 15.6%, 14.3%, and 24.9%, respectively; compared with the conventional multivariate SVR model, they decrease by 52.5%, 57.1%, and 37.7%, respectively. The XGBoost-PSO-SVR model has good prediction performance and can provide decision support for intelligent dust suppression in dry bulk port areas.

Key words: dry bulk port; PM_{2.5}; intelligent dust suppression; multivariate prediction; hybrid model; XGBoost

随着国际贸易的发展和陆海运输通道的完善,铁矿石、煤炭、谷物等大宗商品的流通日益频繁,干散货成为全球贸易和港口服务持续发展的关键动力。根据克拉克森统计,2024 年全球贸易量增长 6.5%,总量达到 126.6 亿 t,其中干散货占比高达 42.3%。然而,干散货装卸、运输和存储作业会产生大量粉尘污染,常规风力条件下其影响半径可达 3~4 km,强风条件下可扩展 10 km 以上,在不同尺度上影响着港区周边空气质量、设备安全,并对作业人员和周边居民的健康构成威胁^[1-2]。厘清干散货港区颗粒物浓度变化的影响因素并实现精准预测,为管理者提前制定针对性的防尘措施提供科学依据,正成为港口绿色智慧化发展的迫切要求。

大量研究指出,颗粒物浓度受多种因素影响,包括风速、风向等气象条件及货种、物料含水率、粒径和作业强度等。Dong 等^[3]利用露天煤尘源估算模型发现,含水率相同时起尘量与风速正相关;Manoli 等^[4]指出 Volos 港内 PM₁₀ 浓度峰值多出现在风速低于 0.5 m/s 的无风天气;李绍武等^[5]通过风洞试验验证了铁矿粉起尘量在自然含水率条件下随风速增大呈幂函数增长;Gupta 等^[6]通过对港区颗粒物及气象监测数据进行

基金项目: 国家重点研发计划项目(2021YFB2600203)

作者简介: 陶凯旋(2002—),男,硕士研究生,主要从事港航运筹与优化算法研究。E-mail: tkx@hhu.edu.cn

通信作者: 封学军(1975—),男,教授,博士,主要从事绿色智慧港航工程研究。E-mail: fxj@hhu.edu.cn

统计分析,发现 TSP 浓度与风速具有较强相关性;Žibert 等^[7]采用聚类方法,证实了风速与降雨是影响颗粒物时空变化的重要因素。在颗粒物时空分布方面,研究表明码头前沿和堆场区域是高浓度的主要分布区^[8]。别淑君^[9]指出青岛港 PM_{2.5} 的主要来源为港口活动与船舶排放;王康等^[10]采用希尔伯特-黄变换(HHT)方法揭示了港口颗粒物浓度的多尺度波动规律。Shen 等^[11]和刁昶皓等^[12]分别从垂直与水平空间维度细化颗粒物分布特征,指出不同作业源和抑尘措施对颗粒物浓度有显著影响。尽管如此,如何在现有基础上构建兼顾多因子影响的高精度颗粒物浓度预测模型,仍是当前研究的热点与难点。

大气污染物浓度预测模型主要可分为数值模式和统计机器学习两类,数值模式基于大气动力学原理,借助 CMAQ^[13]、WRF-Chem^[14]等模型模拟污染物扩散过程,需建立复杂的扩散模型,且预测结果易受初始场数据影响,预报过程较为复杂,结果往往具有不确定性。相比之下,统计机器学习方法通过历史数据构建回归模型,采用多元线性回归、人工神经网络(ANN)、支持向量机(SVM)及支持向量回归(SVR)等模型进行预测,在特定时空条件下可实现较高的预测精度。目前已有多项研究选取 PM_{2.5} 浓度、温度、湿度、气压、风速、水汽压等影响因子进行研究^[15-20],但仍未有效融合 PM_{2.5} 的形成机制与气象条件下的扩散过程,存在一定的局限性。

针对时间序列数据的非线性与波动性特征,自回归积分滑动平均(ARIMA)模型与 SVM 模型被广泛应用并展现出良好的适应性。ARIMA 模型形式相对简单,依赖内生变量即可实现中长期趋势预测,而 SVM 模型则在处理小样本、非线性、高维空间问题中表现出较高精度^[21]。彭斯俊等^[22]研究表明,ARIMA 模型在分时段 PM_{2.5} 预测中优于灰色模型和全时段 ARIMA 模型;宋国君等^[23]发现,ARIMA-SVM 组合模型在峰谷值识别方面优于未进行气象类型划分的传统机器学习模型;戴李杰等^[24]通过引入粒子群优化(PSO)算法提升了 SVM 在 PM_{2.5} 小时级预报中的准确性;Shahriar 等^[25]对 ARIMA-ANN、ARIMA-SVM、决策树和 CatBoost 等多种模型在孟加拉国 PM_{2.5} 浓度预测中的表现进行了系统评估,结果表明组合模型在预测精度方面具有明显优势。此外,ADMS、AERMOD 等大气扩散模型也被用于污染物空间分布模拟^[26-28]。然而,当前研究多集中于单一港区及其周边区域,较少关注颗粒物浓度的多尺度(年、季、日、小时)动态特征及混沌特性,如非平稳性与长期持续性,而基于海量实时监测数据及多变量因素耦合影响探讨港区颗粒物浓度变化的研究亦相对不足。

为了应对多元时序数据和复杂气象及多变量因素的耦合影响,本文系统整合 PM_{2.5} 浓度变化的关键影响因素,提出一种融合特征选择与智能优化的多变量混合预测模型(XGBoost-PSO-SVR 模型),以期为干散货港区 PM_{2.5} 精准防控提供新的技术方案和决策参考。

1 模型构建

PM_{2.5} 浓度变化受气象因子(如温度、湿度、风速等)显著影响,且与 PM₁₀ 具有较强相关性^[29];总悬浮颗粒物(TSP)作为港区颗粒物的重要指标,其变化能够反映作业扬尘特征^[11],仅依赖 PM_{2.5} 浓度单一历史序列进行建模(如 ARIMA 模型)往往受限于其对历史气象条件的“记忆”,难以反映当前或未来气象背景下污染物扩散机制的变化,从而限制了预测的适应性和准确性。为探究复杂多因素交互作用下干散货港区 PM_{2.5} 浓度的预测方法,本文构建了融合特征选择与智能优化的 XGBoost-PSO-SVR 模型,其结构如图 1 所示。

本文建模时,在模型输入层引入气象与颗粒物等多因子数据,通过 XGBoost 算法识别相关变量并剔除冗余特征;为提高 SVR 模型的泛化能力与参数适配性,采用 PSO 算法对其核心参数(惩罚因子 C、核函数核系数 σ)进行全局寻优;最后将筛选后的特征变量输入参数优化后的 SVR 模型,实现 PM_{2.5} 浓度回归预测。该“特征筛选—参数优化—回归预测”三阶段建模策略,通过特征筛选与参数优化的协同作用,可提升模型对复杂多元时序预测问题的适应能力。此外,为验证模型有效性,选取单变量和多变量模型作比较,以突出模型在 PM_{2.5} 形成机制与气象条件下扩散过程耦合预测方面的机制优势。

1.1 XGBoost 重要性指标排序方法

XGBoost 是基于梯度提升的高效机器学习算法,支持多种学习器形式,包括线性分类器和树模型。本文采用树模型进行特征选择,通过量化各特征变量在模型中的重要性,实现以下两个目标:①识别与 PM_{2.5} 高度相关的特征变量;②在保证预测精度的前提下,筛选数量较少的关键特征变量。在模型构建过程中,

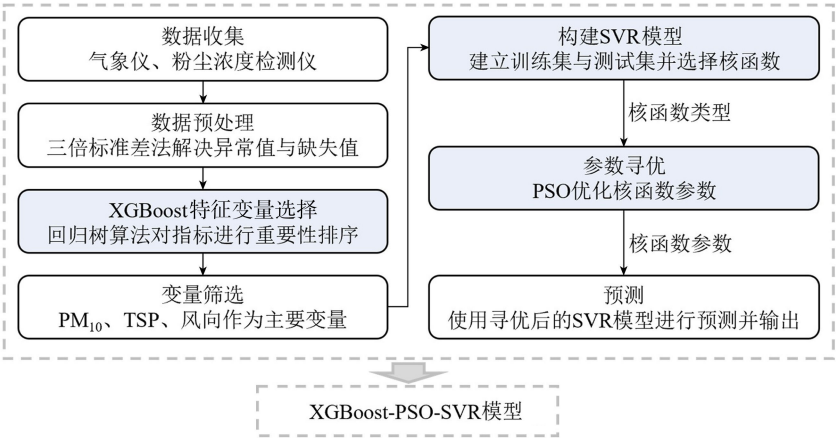


图 1 XGBoost-PSO-SVR 模型结构示意图

Fig. 1 Schematic diagram of XGBoost-PSO-SVR model structure

XGBoost 通过最小化分裂代价、最大化增益来优化树结构,特征变量分裂次数越多、平均增益越大,说明该特征变量的重要性越高。重要性度量指标主要包括特征变量分裂次数、平均增益和平均覆盖率,计算流程为:遍历各节点的所有特征变量→按照样本值对特征变量进行排序→确定增益最大的分裂点并选择相应特征变量进行分裂→更新样本与树节点的映射关系→按照最大深度迭代构建树模型→集成各树完成模型构建→计算特征变量重要性度量指标。

1.2 PSO-SVR 自动寻优模型

SVR 算法作为一种基于统计学习理论的机器学习方法,通过引入核函数将非线性问题映射到高维空间实现线性回归,具有较强的鲁棒性,能够解决非线性、高维、小样本等实际问题,因此常用于非线性回归预测。SVR 模型性能受惩罚因子、不敏感系数与核函数影响显著^[30-31]。惩罚因子 C 控制模型对训练误差的容忍度, C 值过大会导致过拟合,反之则模型泛化不足;不敏感系数 ϵ 则影响支持向量的数量;核函数核系数 σ 决定数据映射的复杂度,本文选择径向基核函数(RBF)。高斯核函数是最常用的 RBF^[31-32],该函数具有局部性质,其参数可决定函数作用范围,同时 RBF 对降低样本数据内的噪声干扰有较好的抑制效果,其中核系数 σ 对模型的拟合情况起关键作用,当 σ 过大时,会出现拟合不足的情况,过小时则会出现过拟合现象。

传统参数选择方法依赖经验试错,易陷入局部最优。为此,本文采用 PSO 算法对 SVR 参数进行全局寻优,构建 PSO-SVR 自动寻优模型,以提升预测精度与泛化能力。PSO 是一种基于群体协同搜索的智能优化算法,通过粒子在参数空间中的迭代搜索确定较优参数组合。本文将每个粒子的位置表示为一组 SVR 参数组合,并以验证集的均方根误差(RMSE)作为适应度指标,RMSE 越小,表明该参数组合对应的模型预测性能越好。

PSO-SVR 自动寻优模型的具体构建流程如图 2 所示,具体步骤如下:

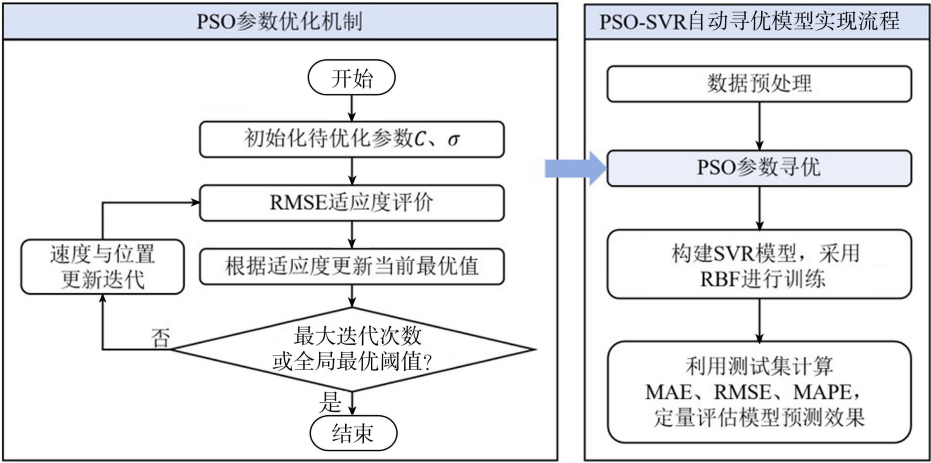


图 2 PSO-SVR 自动寻优模型构建流程

Fig. 2 Construction flowchart of PSO-SVR automatic optimization model

- a. 数据预处理。对输入数据归一化处理,消除量纲影响,并按比例划分为训练集、验证集与测试集。
- b. 参数寻优。设定 PSO 参数(种群规模、迭代次数、搜索范围),执行算法获取最优参数 C^* 、 σ^* 。
- c. 模型训练与验证。基于最优参数构建 SVR 模型,采用 RBF 进行训练。
- d. 性能评估。利用测试集计算平均绝对误差(MAE)、平均绝对百分比误差(MAPE)和 RMSE,定量评估模型预测效果。

2 实例验证

以长江南京段南京西坝码头有限公司(以下简称“西坝码头”)、南京大厂港务有限公司(以下简称“大厂港务”)、南京大胜关港港口有限公司(以下简称“大胜关港口”)、南京明州码头有限公司(以下简称“明州码头”)和南京霞瑞实业有限公司(以下简称“霞瑞码头”)5 个主要干散货港区为案例分析对象,其主要营运货种、年通过能力、港区内抑尘措施与设备等如表 1 所示。

表 1 长江南京段港区简况

Table 1 Overview of port areas in the Nanjing section of the Yangtze River			
名称	运营货种	抑尘措施与设备	年通过能力/万 t
西坝码头	煤炭、砂石、石油焦及铜精矿等	防风抑尘网,封闭式皮带输送机,除尘喷淋设备	1 911
大厂港务	砂石、钢铁、石化、煤炭	封闭式皮带输送机	3 436
大胜关港口	砂石料、液体化工品	防风抑尘网,堆场覆盖,封闭式皮带输送机	350
霞瑞码头	煤、石油焦、砂石	封闭式皮带输送机,封闭式堆场	150
明州码头	煤炭、矿石、建材	U 形斗、防风抑尘网	649

颗粒物数据来源于江苏省港口粉尘在线监测监管平台,该平台涵盖了全省主要干散货港区 PM_{2.5}、PM₁₀、TSP 和大气压力、风力、风向、温度、湿度分钟级实时监测数据,每个港区至少配置 1 台 β 射线颗粒物监测仪(港界处)和覆盖码头前沿、堆场及卡口的多台光散射法颗粒物监测仪,总数超过 5 000 台套。粉尘在线监测系统 24 h 连续工作,记录每分钟各检测对象浓度,数据结构如图 3 所示。

本文以 PM_{2.5} 日均浓度作为预测目标,预测时间尺度为天。数据集同时涵盖颗粒物与气象因子两个维度,构成典型的复杂多元时序结构,具有显著的多变量特性。这一特征要求在模型构建中充分考虑污染物形成与气象因子的耦合影响,从而提升模型的预测精度与泛化能力。

2.1 数据预处理

粉尘在线监测监管平台数据由于天气、设备等原因会导致数据丢失,造成某时间段颗粒物检测数据的某个特征值为“0”,或者出现风向范围不在 0°~360°之间、颗粒物质量浓度达到仪器极值(300 $\mu\text{g}/\text{m}^3$)等明显不符合实际的数据,预处理时将该时间段数据删除。针对其他异常数据,采用三倍标准差法进行异常值识别与处理^[33]。

2.2 特征变量选择与智能优化

为简化模型结构并提升预测性能,本文从多个影响因子中筛选出与 PM_{2.5} 浓度关联度最高的核心特征变量作为原始输入,涵盖颗粒物因子(PM_{2.5}、PM₁₀、TSP)和气象因子(大气压力、风速、风向、温度、湿度)。采用 XGBoost 中的回归树算法,对所有特征变量的影响程度进行排序,计算各特征变量的重要性得分。剔除影响较小的特征变量,以简化计算并提高预测精度。PM_{2.5} 浓度的预测中,颗粒物因子的重要性高于气象因子,重要性排序为 PM₁₀、TSP、风向、温度、湿度和风速。在颗粒物因子中,由于 PM_{2.5} 与 PM₁₀ 和 TSP 存在较强的线性关系,且为向下包含关系,因此 PM₁₀ 和 TSP 的重要性得分较高。尽管 PM₁₀、TSP 均包含 PM_{2.5},但基于 XGBoost 的重要性量化指标结果,可以认为 PM₁₀ 和 TSP 并非冗余特征,而是预测 PM_{2.5} 浓度最关键的



图 3 港口粉尘在线监测监管平台数据结构示例
Fig. 3 Example of data structure of port dust online detection and supervision platform

核心特征变量。因此,本文选择保留 PM_{10} 和 TSP 两个特征变量作为模型的输入。

气象因子中,风向重要性较高,温度次之,风速和湿度的重要性相对较低,这与康俊锋等^[34]研究结果相近。因此,本文选取 PM_{10} 、TSP 和风向作为影响预测 $\text{PM}_{2.5}$ 浓度的特征变量。PSO 算法优化后,确定惩罚因子 $C=20$ 、核函数核系数 $\sigma=0.046$ 。PSO 算法参数设置如下:粒子数量为 100,粒子维度为 2,最大迭代次数为 500,局部和全局学习因子均为 2,惯性因子为 0.8,参数最大值为 25、最小值为 0.001。

2.3 模型预测结果对比与分析

港区颗粒物浓度在不同时间尺度(小时、日、月、季和年)下的平均水平存在差异,但整体变化趋势基本一致。大厂港务作为典型的多货种、高通量港区,具备较强代表性,故以其作为案例分析对象。本文采用大厂港务 2020 年 1 月 1 日至 12 月 27 日的 362 个 $\text{PM}_{2.5}$ 日均质量浓度数据,基于前 352 天(1 月 1 日至 12 月 17 日)数据进行拟合,后 10 天(12 月 18 日至 12 月 27 日)的数据进行预测,结果见图 4 和表 2。由图 4 可见,日均 $\text{PM}_{2.5}$ 质量浓度的拟合值与实测值变化趋势基本一致,拟合效果较好。结合表 2 预测效果可见,模型对数据的解释能力强($R^2=0.89$),已经较为接近理想水平,说明该模型能够有效捕捉 $\text{PM}_{2.5}$ 浓度的变化规律;MAE 为 $6.46\text{ }\mu\text{g}/\text{m}^3$,RMSE 为 $8.91\text{ }\mu\text{g}/\text{m}^3$,误差均较小;而 MAPE 为 6%,表明相对误差非常小,进一步证明了该模型在实际预测中的高准确性。因此,XGBoost-PSO-SVR 模型在 $\text{PM}_{2.5}$ 浓度预测中展现出了良好性能,其预测精度和泛化能力可满足港区污染防控需求,可为类似研究和实际应用提供有力支持。

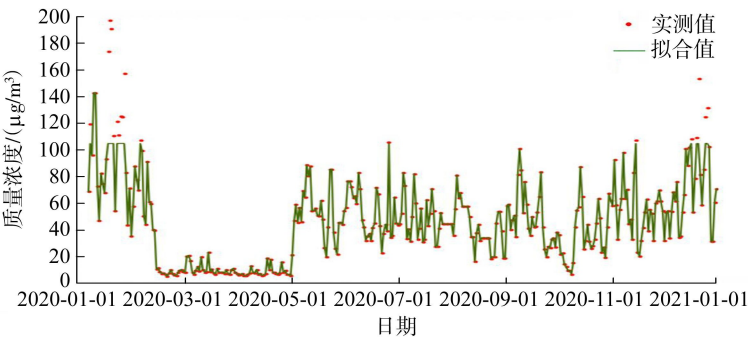


图 4 XGBoost-PSO-SVR 模型拟合效果

Fig. 4 Fitting effect of XGBoost-PSO-SVR model

2.4 与其他模型预测精度对比

为验证多特征变量选择对 $\text{PM}_{2.5}$ 日均质量浓度预测结果的影响,选取 ARIMA 模型^[22]、STL-ARIMA-SVR 模型^[32,35]、使用全部特征变量作为输入的传统 SVR 模型,与本文 XGBoost-PSO-SVR 模型进行对比分析,以综合评估模型性能。

2.4.1 ARIMA 模型

ARIMA 模型是一种经典的时间序列预测模型,通过自回归、差分和移动平均等过程刻画时间序列中的趋势性和随机波动特征,广泛应用于具有明显时间相关性的预测问题。本文采用 ARIMA 模型作为单变量预测模型的对比模型,用于分析 $\text{PM}_{2.5}$ 历史浓度序列的变化规律。

本文在 python3.8 平台中通过编写程序调用 statsmodels. api 数据包构建 ARIMA 模型,经由遍历可得,当自回归阶数 $p=2$ 、差分阶数 $d=0$ 、移动平均阶数 $q=1$ 时,AIC、BIC 以及 HQIC 值均为最小,故选取 ARIMA(2, 0, 1)作为最优模型进行预测。

2.4.2 STL-ARIMA-SVR 模型

STL-ARIMA-SVR 模型是一种单变量混合预测模型,该模型将 STL(seasonal-trend decomposition procedure based on loess)分解方法与 ARIMA、SVR 模型相结合,擅长处理具有季节性和趋势性的时间序列数据。尽管该方法最早由 Cleveland 等在 1990 年提出^[36],但其基于分解的核心思想依然具有很强的时效性^[35-36],近年

表 2 XGBoost-PSO-SVR 模型预测效果

Table 2 Prediction performance of XGBoost-PSO-SVR model

日期	实测值/($\mu\text{g}/\text{m}^3$)	预测值/($\mu\text{g}/\text{m}^3$)	相对误差/%
2020-12-18	110.25	109.15	1
2020-12-19	82.59	87.49	6
2020-12-20	58.83	60.73	3
2020-12-21	90.23	87.13	3
2020-12-22	91.43	90.33	1
2020-12-23	124.28	118.03	5
2020-12-24	145.25	125.81	13
2020-12-25	99.72	92.62	7
2020-12-26	69.14	66.04	4
2020-12-27	132.52	115.88	13

来仍有学者将 STL 分解方法应用于 PM_{2.5} 浓度预测并取得了较好的效果^[32]。本文选用 STL-ARIMA-SVR 单变量预测模型作为对比基准,旨在突出 XGBoost-PSO-SVR 多变量预测模型在考虑多因素耦合影响方面的优势。STL-ARIMA-SVR 模型预测步骤如下:

- a. 分解。采用 STL 分解方法将原始 PM_{2.5} 浓度数据分解为长期趋势项、季节变动项和余项。
- b. 预测。采用 ARIMA 模型对长期趋势项进行预测,采用 SVR 模型对季节变动项和余项进行预测,得到长期趋势项、季节变动项和余项对应的预测结果。
- c. 重构。将长期趋势项、季节变动项和余项的预测结果进行叠加重构,获得最终的 PM_{2.5} 浓度预测值。
- d. 评估。采用 MAE、MAPE 和 RMSE 作为评价指标,对模型预测性能进行评价。

2.4.3 结果对比与分析

特征选择和智能优化能显著提高 SVR 模型的拟合和预测精度,与本文 XGBoost-PSO-SVR 模型(图 4 和表 2)相比,未经 XGBoost 特征筛选和 PSO 参数优化的 SVR 模型拟合效果及预测性能均较差(图 5 和表 3),PM_{2.5} 日均质量浓度最大绝对误差为 21.26 μg/m³,MAE、MAPE、RMSE 和 R² 分别为 13.60 μg/m³、14%、14.30 μg/m³ 和 0.70。虽然 SVR 模型在低质量浓度区内拟合效果较好(尤其在第 50~100 天),但在需要重点关注的高质量浓度区,特别是尖峰值的预测上存在较大偏差。本文 XGBoost-PSO-SVR 模型在 3—11 月的拟合效果明显优于 SVR 模型,同时预测结果的 R² 值由 0.70 提升至 0.89。因此,采用 XGBoost 筛选后的特征变量作为输入,能够显著提升预测精度,验证了本文 XGBoost-PSO-SVR 模型的优势。

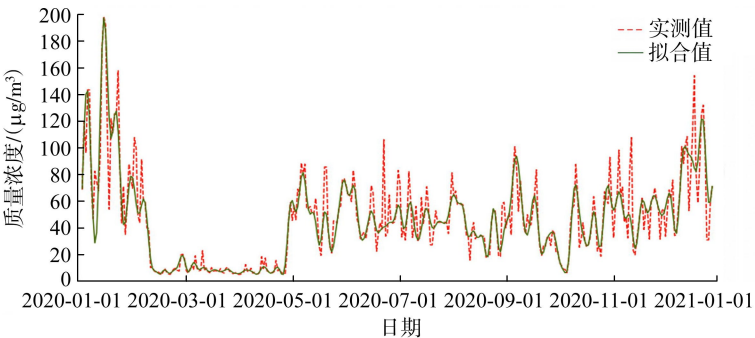


图 5 SVR 模型拟合效果
Fig.5 Fitting effect of SVR model

3 种对比模型与本文模型的预测结果对比见图 6 和表 4。相较于 ARIMA、STL-ARIMA-SVR 和传统 SVR 模型,本文 XGBoost-PSO-SVR 模型在各项指标上均表现出更优的预测精度。在单变量预测模型中,经 STL 分解后的 STL-ARIMA-SVR 模型精度明显优于 ARIMA 模型,MAE、MAPE 和 RMSE 分别降低了 66.8%、68.2%和 57.0%。在多变量预测模型中,经特征选择后的 XGBoost-PSO-SVR 模型精度明显优于 SVR 模型,MAE、MAPE 和 RMSE 分别降低了 52.5%、57.1%和 37.7%。与单变量 STL-ARIMA-SVR 模型相比,多变量 XGBoost-PSO-SVR 模型预测结果总体更加准确,MAE、MAPE 和 RMSE 分别降低了 15.6%、14.3%和 24.9%。

在表现较好的两个混合模型中,XGBoost-PSO-SVR 模型在前 3 天的预测表现更加突出,且全周期内保持较低的误差波动,整体稳定性更强。STL-ARIMA-SVR 模型通过 STL 分解能有效提取趋势、季节和残差项,改善了 ARIMA 对非平稳数据的适应能力,尤其在第 4 天和第 5 天 PM_{2.5} 质量浓度波动时,STL-ARIMA-SVR 模型能够在局部时间窗口内提供较高精度的预测。然而,从整体预测性能看,XGBoost-PSO-SVR 模型的最大绝对误差为 19.44 μg/m³,小于 STL-ARIMA-SVR 模型的 25.89 μg/m³。XGBoost-PSO-SVR 模型在多变量输入基

表 3 SVR 模型预测效果

Table 3 Prediction performance of SVR model

日期	实测值/(μg/m ³)	预测值/(μg/m ³)	相对误差/%
2020-12-18	110.25	101.15	8
2020-12-19	82.59	96.69	17
2020-12-20	58.83	43.93	25
2020-12-21	90.23	108.33	20
2020-12-22	91.43	84.53	8
2020-12-23	124.28	135.18	9
2020-12-24	145.25	123.99	15
2020-12-25	99.72	113.62	14
2020-12-26	69.14	60.24	13
2020-12-27	132.52	114.54	14

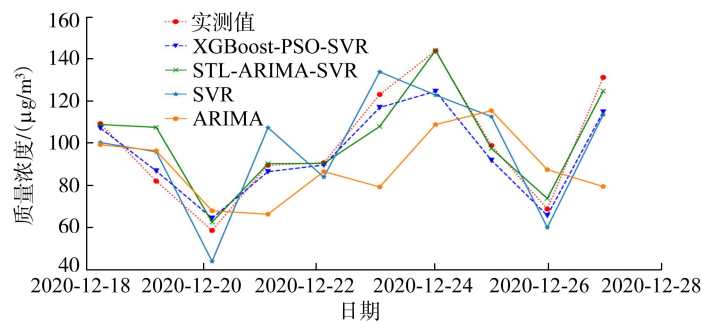


图 6 不同模型未来 10 天 $\text{PM}_{2.5}$ 质量浓度预测结果对比

Fig. 6 Comparison of $\text{PM}_{2.5}$ mass concentration prediction results of different models for next 10 days

表 4 $\text{PM}_{2.5}$ 日均质量浓度预测精度比较

Table 4 Comparison of prediction accuracy of $\text{PM}_{2.5}$ daily average mass concentration

变量类型	模型	MAE/($\mu\text{g}/\text{m}^3$)	MAPE/%	RMSE/($\mu\text{g}/\text{m}^3$)	R^2
单变量	ARIMA	23.01	22	27.60	0.34
	STL-ARIMA-SVR	7.65	7	11.86	0.80
多变量	SVR	13.60	14	14.30	0.70
	XGBoost-PSO-SVR	6.46	6	8.91	0.89

础上融合了特征选择与 PSO 参数优化机制,减少了冗余变量干扰,增强了 SVR 模型的泛化能力,性能表现更佳。此外,当 $\text{PM}_{2.5}$ 质量浓度处于 $55\sim95\mu\text{g}/\text{m}^3$ 时,所有模型的预测精度均较高;当质量浓度超过 $95\mu\text{g}/\text{m}^3$ 时,预测精度有所下降,但 XGBoost-PSO-SVR 模型波动幅度相对较小,表现出更强的鲁棒性与适应性。

3 结 语

本文针对干散货港区 $\text{PM}_{2.5}$ 浓度受多因素影响、时序变化复杂的问题,提出了融合 XGBoost 特征筛选与 PSO 参数优化的 SVR 多变量混合预测模型。该模型利用 XGBoost 筛选 PM_{10} 、TSP 和风向等关键影响因子,并采用 PSO 对 SVR 模型参数进行优化。长江南京段典型干散货港区监测数据验证结果表明,XGBoost-PSO-SVR 模型的 MAE、MAPE、RMSE 和 R^2 分别为 $6.46\mu\text{g}/\text{m}^3$ 、6%、 $8.91\mu\text{g}/\text{m}^3$ 和 0.89,整体预测性能优于 ARIMA、STL-ARIMA-SVR 和传统 SVR 模型,且在 $\text{PM}_{2.5}$ 高质量浓度区间具有较好的稳定性和适应性。研究表明,特征筛选与参数优化能够有效提升 SVR 模型对港区 $\text{PM}_{2.5}$ 浓度变化的预测能力,可为干散货港区污染预测及智慧抑尘决策提供参考。后续可进一步考虑极端天气条件下气象变量突变对预测性能的影响,以提高模型在复杂环境条件下的鲁棒性。

参考文献:

[1] Shen Jinxing, Liu Qinxin, Feng Xuejun. Prediction of PM_{10} concentration in dry bulk ports using a combined deep learning model considering feature meteorological factors[J]. Environmental Science, 2024, 45(9): 5179-5187.

[2] Shen Jinxing, Liu Qinxin, Feng Xuejun. Hourly $\text{PM}_{2.5}$ concentration prediction for dry bulk port clusters considering spatiotemporal correlation: a novel deep learning blending ensemble model[J]. Journal of Environmental Management, 2024, 370: 122703.

[3] Dong Shipai, Liu Bin. Analysis of coal dust source strength in open storage yard of coal port area of Huanghua Port[J]. E3S Web of Conferences, 2020, 206: 01009.

[4] Manoli E, Chelioti-Chatzidimitriou A, Karageorgou K, et al. Polycyclic aromatic hydrocarbons and trace elements bounded to airborne PM_{10} in the harbor of Volos, Greece: implications for the impact of harbor activities[J]. Atmospheric Environment, 2017, 167: 61-72.

[5] 李绍武, 李文善, 雷鹏, 等. 港口散货堆场铁矿粉起尘风洞试验研究[J]. 水道港口, 2016, 37(5): 558-562. (Li Shaowu, Li Wenshan, Lei Peng, et al. Research on dust emission of iron ore for bulk-cargo yard of ore terminal by using wind tunnel test[J]. Port & Waterway Engineering, 2016, 37(5): 558-562. (in Chinese))

[6] Gupta A K, Patil R S, Gupta S K. A Statistical analysis of particulate data sets for jawaharlal nehru port and surrounding harbour

region in India[J]. Environmental Monitoring and Assessment,2004,95(1/2/3):295-309.

[7] Žibert J,Pražnikar J. Cluster analysis of particulate matter (PM₁₀) and black carbon (BC) concentrations[J]. Atmospheric Environment,2012,57:1-12.

[8] 彭宜嵩,封学军,顾建,等. 基于发动机活动的集装箱港作机械排放清单研究[J]. 安全与环境学报,2017,17(3):1136-1141. (Peng Yiqiang,Feng Xuejun,Gu Jian,et al. On the air pollutant emission inventory of the cargo handling equipment based on the activities of the container delivery ports[J]. Journal of Safety and Environment,2017,17(3):1136-1141. (in Chinese))

[9] 别淑君. 港口地区大气 PM_{2.5} 及其化学成分的组成特征及来源解析[D]. 济南:山东大学, 2020.

[10] 王康,张艳,王启明,等. 基于 HHT 的干散货港区大气 TSP 浓度时频特征分析[J]. 环境科学与技术,2020,43(9):125-134. (Wang Kang,Zhang Yan,Wang Qiming,et al. Time-frequency analysis of TSP in dry bulk port based on HHT[J]. Environmental Science & Technology,2020,43(9):125-134. (in Chinese))

[11] Shen Jinxing, Feng Xuejun,Zhuang Kai,et al. Vertical distribution of particulates within the near-surface layer of dry bulk port and influence mechanism;a case study in China[J]. Sustainability,2019,11(24):7135.

[12] 刁昶皓,封学军,张艳,等. 基于无人机平台的干散货港区 PM_{2.5} 分布规律研究[J]. 环境污染与防治,2020,42(12):1502-1506. (Diao Changhao,Feng Xuejun,Zhang Yan,et al. Study on PM_{2.5} distribution in dry bulk port based on unmanned aerial vehicle platform [J]. Environmental Pollution & Control,2020,42(12):1502-1506. (in Chinese))

[13] Chen Jianjun, Lu Jin, Avise J C, et al. Seasonal modeling of PM_{2.5} in California's San Joaquin Valley [J]. Atmospheric Environment,2014,92:182-190.

[14] Grell G A, Peckham S E, Schmitz R, et al. Fully coupled "online" chemistry within the WRF model [J]. Atmospheric Environment,2005,39(37):6957-6975.

[15] Mohan M,Gupta M. Sensitivity of PBL parameterizations on PM₁₀ and ozone simulation using chemical transport model WRF-Chem over a sub-tropical urban airshed in India[J]. Atmospheric Environment,2018,185:53-63.

[16] Lu Weizhen,Wang Wenjian,Wang Xiekang,et al. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air,Hong Kong[J]. Environmental Research,2004,96(1):79-87.

[17] Zhou Qingping,Jiang Haiyan,Wang Jianzhou,et al. A hybrid model for PM_{2.5} forecasting based on ensemble empirical mode decomposition and a general regression neural network[J]. Science of the Total Environment,2014,496:264-274.

[18] Narut S,Tassanee P,Kraichat T. Relationship between PM₁₀ and PM_{2.5} levels in high-traffic area determined using path analysis and linear regression[J]. Journal of Environmental Sciences,2018,69:105-114.

[19] 陈亚玲,赵智杰. 基于小波变换与传统时间序列模型的臭氧浓度多步预测[J]. 环境科学学报,2013,33(2):339-345. (Chen Yaling,Zhao Zhijie. A multi-step-ahead prediction of ozone concentration using wavelet transform and traditional time series model[J]. Acta Scientiae Circumstantiae,2013,33(2):339-345. (in Chinese))

[20] 陈军,高岩,张烨培,等. PM_{2.5} 扩散模型及预测研究[J]. 数学的实践与认识,2014,44(15):16-27. (Chen Jun,Gao Yan,Zhang Yepei,et al. Study on the diffusion model and forecast of PM_{2.5} pollution[J]. Mathematics in Practice and Theory,2014,44(15):16-27. (in Chinese))

[21] 武骁,宗蓓华. 基于支持向量机回归的港口吞吐量非线性组合预测[J]. 河海大学学报(自然科学版),2005,33(4):482-484. (Wu Xiao,Zong Beihua. Nonlinear combined prediction of port throughput based on support vector machine regression[J]. Journal of Hohai University (Natural Sciences),2005,33(4):482-484. (in Chinese)).

[22] 彭斯俊,沈加超,朱雪. 基于 ARIMA 模型的 PM_{2.5} 预测[J]. 安全与环境工程,2014,21(6):125-128. (Peng Sijun,Shen Jiachao,Zhu Xue. PM_{2.5} prediction based on ARIMA model[J]. Safety and Environmental Engineering,2014,21(6):125-128. (in Chinese))

[23] 宋国君,国潇丹,杨啸,等. 沈阳市 PM_{2.5} 浓度 ARIMA-SVM 组合预测研究[J]. 中国环境科学,2018,38(11):4031-4039. (Song Guojun,Guo Xiaodan,Yang Xiao,et al. Forecasting of PM_{2.5} concentration in Shenyang City with ARIMA-SVM combined model[J]. China Environmental Science,2018,38(11):4031-4039. (in Chinese))

[24] 戴李杰,张长江,马雷鸣. 基于机器学习的 PM_{2.5} 短期浓度动态预报模型[J]. 计算机应用,2017,37(11):3057-3063. (Dai Lijie,Zhang Changjiang,Ma Leiming. Dynamic forecasting model of short-term PM_{2.5} concentration based on machine learning [J]. Journal of Computer Applications,2017,37(11):3057-3063. (in Chinese))

[25] Shahriar S A, Kayes I, Hasan K, et al. Potential of ARIMA-ANN, ARIMA-SVM, DT and CatBoost for atmospheric PM_{2.5} forecasting in Bangladesh[J]. Atmosphere,2021,12(1):100.

[26] 石柳,刘陈鸿,马继华,等. ADMS 模型在港口规划大气环境影响评价中的应用[J]. 西部交通科技,2019(6):175-205. (Shi Liu,Liu Chenhong,Ma Jihua,et al. Application of ADMS model in atmospheric environmental impact assessment of port

planning[J]. Western China Communications Science & Technology, 2019(6):175-205. (in Chinese))

[27] 何泽慧,彭士涛. AERMOD-EIA 模型参数在港口粉尘污染预测中的敏感性分析[J]. 水道港口, 2012, 33(4):342-347. (He Zehui, Peng Shitao. AERMOD-EIA model parameters sensitivity analysis of port dust pollution prediction[J]. Port & Waterway Engineering, 2012, 33(4):342-347. (in Chinese))

[28] 刘建昌,李兴华,徐洪磊,等. 内河多点分散码头大气污染叠加影响特征[J]. 环境科学, 2013, 34(5):2044-2050. (Liu Jianchang, Li Xinghua, Xu Honglei, et al. Superposition impact character of air pollution from decentralization docks in a freshwater port[J]. Environmental Science, 2013, 34(5):2044-2050. (in Chinese))

[29] 张玮,郭胜利,申付振,等. 南京地区 PM_{2.5} 和 PM₁₀ 浓度分布特征及与相关气象条件的关系[J]. 科学技术与工程, 2016, 16(4):124-129. (Zhang Wei, Guo Shengli, Shen Fuzhen, et al. The relationship between the distribution characteristics of PM₁₀ and PM_{2.5} in Nanjing area and meteorological conditions[J]. Science Technology and Engineering, 2016, 16(4):124-129. (in Chinese))

[30] 谷艳昌,吴云星,黄海兵,等. 基于遗传算法优化支持向量机的大坝安全性态预测模型[J]. 河海大学学报(自然科学版), 2020, 48(5):419-425. (Gu Yanchang, Wu Yunxing, Huang Haibing, et al. Prediction model of dam safety behavior based on genetic algorithm optimized support vector machine[J]. Journal of Hohai University (Natural Sciences), 2020, 48(5):419-425. (in Chinese))

[31] 王红瑞,魏豪杉,胡立堂,等. 基于遗传算法的 SVM-AR 改进模型与应用[J]. 河海大学学报(自然科学版), 2020, 48(6):488-497. (Wang Hongrui, Wei Haoshan, Hu Litang, et al. Improved model and application of SVM-AR based on genetic algorithm [J]. Journal of Hohai University (Natural Sciences), 2020, 48(6):488-497. (in Chinese))

[32] 柳建菲. 基于时间序列分解的 PM_{2.5} 浓度预测[D]. 兰州:兰州大学, 2020.

[33] Bolmanis E, Uhlendorff S, Pein-Hackelbusch M, et al. Anomaly detection and removal strategies for in-line permittivity sensor signal used in bioprocesses[J]. Frontiers in Bioengineering and Biotechnology, 2025, 13:1609369.

[34] 康俊锋,谭建林,方雷,等. XGBoost-LSTM 变权组合模型支持下短期 PM_{2.5} 浓度预测:以上海为例[J]. 中国环境科学, 2021, 41(9):4016-4025. (Kang Junfeng, Tan Jianlin, Fang Lei, et al. Short-term PM_{2.5} concentration prediction based on XGBoost and LSTM variable weight combination model: a case study of Shanghai[J]. China Environmental Science, 2021, 41(9):4016-4025. (in Chinese)).

[35] 吕蓬,岳莉莉,赵晓丽. 小波分解和 ARIMA 模型相结合的短期风速预测[J]. 科技信息, 2011(14):401-402. (Lyu Peng, Yue Lili, Zhao Xiaoli. Short-term wind speed prediction combined with wavelet decomposition and ARIMA model[J]. Science and Technology Information, 2011(14):401-402. (in Chinese))

[36] Cleveland R B, Cleveland W S, McRae J E, et al. STL: a seasonal-trend decomposition procedure based on LOESS[J]. Journal of Official Statistics, 1990, 6(1):3-73.

(收稿日期:2025-07-08 编辑:熊水斌)

(上接第 38 页)

[30] 陈爱军,陶子淳. 多样性地形条件下 IMERG 精度综合评估[J]. 科学技术与工程, 2025, 25(16):6608-6618. (Chen Aijun, Tao Zichun. Comprehensive assessment of IMERG accuracy under diverse topographic conditions[J]. Science Technology and Engineering, 2025, 25(16):6608-6618. (in Chinese))

[31] 雍斌,周振杨,陶新,等. 黄河源区多源融合降水产品综合评价及其水文效应评估[J]. 水资源保护, 2025, 41(2):47-57. (Yong Bin, Zhou Zhenyang, Tao Xin, et al. Comprehensive evaluation and hydrological effect evaluation of multi-source fusion precipitation product in the Yellow River source area[J]. Water Resources Protection, 2025, 41(2):47-57. (in Chinese))

[32] 李致家,龚俊超,孙明坤. 中国半湿润流域 WRF-Hydro 模型参数敏感性分析[J]. 河海大学学报(自然科学版), 2023, 51(5):1-8. (Li Zhijia, Gong Junchao, Sun Mingkun. Parametric sensitivity analysis of WRF-Hydro model in a semi-humid watershed of China[J]. Journal of Hohai University (Natural Sciences), 2023, 51(5):1-8. (in Chinese))

[33] 齐晟慰,谷黄河,王玉明,等. 不同融合降水数据驱动的 WRF-Hydro 模型对中小流域山洪模拟的适用性[J]. 河海大学学报(自然科学版), 2026, 54(2):45-52. (Qi Shengwei, Gu Huanghe, Wang Yuming, et al. Applicability of WRF-Hydro model driven by different precipitation fusion data in mountain torrent simulation of medium-and small-sized watersheds[J]. Journal of Hohai University (Natural Sciences), 2026, 54(2):45-52. (in Chinese))

(收稿日期:2025-08-26 编辑:刘晓艳)