

基于区间分位数回归的中国空气质量指数影响因素分析

魏晨飞, 徐登可

杭州电子科技大学经济学院, 浙江 杭州

收稿日期: 2026年9月9日; 录用日期: 2026年9月30日; 发布日期: 2026年10月12日

摘 要

为探究六种污染物浓度对中国空气质量指数的影响, 本文采用区间分位数回归模型, 对168个城市2026年6月1日至2026年6月30日的月区间数据进行实证分析, 结果得到, 从空气质量指数下限数据来看, PM_{10}^l 、 SO_2^l 、 O_3^l 为核心影响因素; 从空气质量指数上限数据来看, $PM_{2.5}^u$ 、 PM_{10}^u 、 O_3^u 为核心影响因素, 且其作用强度随空气质量变化存在显著差异。预测结果表明, 本文提出的基于最小最大法的区间分位数回归方法预测性能较优。

关键词

空气质量指数, 区间数据, 最小最大法, 分位数回归

Analysis of Influencing Factors of Air Quality Index in China Based on Interval Quantile Regression

Chenfei Wei, Dengke Xu

School of Economics, Hangzhou Dianzi University, Hangzhou Zhejiang

Received: September 9, 2026; accepted: September 30, 2026; published: October 12, 2026

Abstract

To explore the impacts of the concentrations of six pollutants on Chinese Air Quality Index (AQI), this paper adopts an interval quantile regression model to conduct an empirical analysis on monthly interval data of 168 cities from June 1 to June 30, 2026. The results show that, in terms of

the lower bound data of the Air Quality Index, PM_{10}^l , SO_2^l , O_3^l serve as the core influencing factors; while for the upper bound data of the Air Quality Index, $PM_{2.5}^u$, PM_{10}^u , O_3^u are the core influencing factors, and their effect intensities differ significantly with changes in air quality. The prediction results indicate that the proposed interval quantile regression method based on the Min-Max method achieves better prediction performance.

Keywords

Air Quality Index, Interval Data, Min-Max Method, Quantile Regression

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着全球气候变化与区域经济的快速发展,我国空气污染形势日益严峻,空气污染不仅危害公众健康,还会对居民收入[1]、科技创新[2]、区域经济增长[3]等造成不可忽视的负面影响,因此社会各界对空气质量的关注度也持续提升。空气质量指数(AQI)作为根据 $PM_{2.5}$ 、 PM_{10} 、 SO_2 、 NO_2 、 O_3 、 CO 等污染物浓度计算得出的综合性指标,能直观、量化地反映全国各区域的空气污染程度,已成为我国大气环境监测、污染防控及环境治理的重要依据。探究空气质量的影响因素并实施相关预测也已成为环境科学领域的研究热点,越来越多学者围绕该主题展开了大量的实证研究,例如吴雪萍等[4]研究了空气质量与经济增长之间的非线性关系;郑元琳[5]通过构建组合预测模型,对济南市 AQI 进行了预测分析。

随着各学者对空气质量指数影响因子及预测分析研究的深入,相应的分析方法也不断革新。黄振庭等[6]利用多元线性回归模型,研究了惠州市 AQI 与各污染物及气象因素的相关关系;雷雨薇等[7]基于 XGBoost-SHAP 方法对福建省进行了空气质量指数影响因素分析;He 等[8]采用分布滞后非线性模型评估了大气污染物与 IHD 死亡风险之间的影响关系;倪洋等[9]基于加权分位数和回归模型,分析了天津市 AQHI 和 AQI 与不同特征人群和季节呼吸系统疾病死亡风险的暴露反应关系;史学良等[10]利用 EEMD 与改进的 LSTM 网络相结合的组合模型对太原市空气质量指数进行了预测;杜艳芳[11]通过分位数回归方法对甘肃省三大城市的 AQI 进行了评价与预测;景伊等[12]基于 ARIMA 模型短期预测了运城市未来空气质量指数变化趋势;方晓萍等[13]利用 PSO-GSA-LSTM 模型对长江经济带空气质量指数进行了预测研究;史佳宜[14]利用 LSTM 网络对南京市空气质量进行了有效的时序预测;Tian 等[15]提出了一种基于混合核函数核主成分分析的空气质量指数预测模型;陈博等[16]利用变分模态分解并采用改进的河马优化算法对兰州市的空气质量指数进行预测分析;Wei 等[17]提出了一种具有 SHAP 可解释性的 Transformer-XGBoost 框架,用以预测临汾市重工业化地区的 AQI 数据;Xiong 等[18]利用融合时间卷积和双向 LSTM 的短期预测方法,对多城市 AQI 进行了预测。

上述实证研究多基于点值数据,难以捕捉数据的波动差异,与之相比,区间数据能保留更多有效信息,并反映数据的波动性及不确定性。针对区间数据,学者们提出了许多方法,如 Billard 等[19]提出了中心法和最小最大法;Lima Neto 等[20]提出了中心半径法;Souza 等[21]提出了参数化方法。然而,基于区间数据的影响因素及预测研究仍然较少。例如呼延璇[22]通过变分模态分解与和声搜索算法对 AQI 进行了区间预测;Yang 等[23]提出了一种基于模糊时间序列和误差分布的空气质量指数智能区间预测系统,

并验证了该预测系统的良好性能；Suo 等[24]将空气质量指数与 MTIP 方法结合，得出了 $\text{PM}_{2.5}$ 是首要污染物的结论；石逸[25]采用多维标度方法对 AQI 区间数据展开了实证研究。

因此，本文利用区间型分位数回归模型，对中国空气质量指数数据进行影响因素分析及预测，以期相关政策制定及实施提供科学的依据。

2. 实证分析

2.1. 数据来源与选取

本文依据生态环境部逐月公布的《全国环境空气质量状况》¹，选取了 168 个重点城市包括京津冀及周边地区 38 个城市、长三角地区 31 个城市、汾渭平原 13 个城市、成渝地区 16 个城市、长江中游城市群 21 个城市、珠三角地区 9 个城市，以及其他省会城市和计划单列市 40 个城市的 AQI 指数及 $\text{PM}_{2.5}$ 、 PM_{10} 、 SO_2 、 NO_2 、 O_3 、CO 浓度的区间数据，主要为 2026 年 6 月 1 日~2026 年 6 月 30 日的月最小值和最大值，数据来源于中国环境监测总站。

2.2. 描述性分析

表 1 展示了具体变量的描述性统计，其中上下限分别用上标 u 和 l 表示。可以发现，各变量月度上限均值显著大于下限均值，且上限标准差远大于下限标准差，这说明不同城市产业结构、气象条件的差异会拉大城市间月度污染峰值的差异，而月度污染谷值更多由区域大气基础环境决定，波动较小。

Table 1. Descriptive statistics for variables

表 1. 变量的描述性统计

变量	符号	均值	标准差	最小值	最大值
AQI	AQI^l	10.50	4.53	6.00	25.00
	AQI^u	113.50	57.47	45.00	500.00
$\text{PM}_{2.5}$	$\text{PM}_{2.5}^l$	3.00	0.69	2.00	9.00
	$\text{PM}_{2.5}^u$	76.50	31.95	23.00	303.00
PM_{10}	PM_{10}^l	7.00	2.76	4.00	18.00
	PM_{10}^u	113.50	92.94	38.00	697.00
SO_2	SO_2^l	2.00	1.57	1.00	8.00
	SO_2^u	12.00	19.79	6.00	161.00
NO_2	NO_2^l	2.50	1.41	1.00	7.00
	NO_2^u	35.50	15.92	15.00	89.00
O_3	O_3^l	11.50	10.18	1.00	64.00
	O_3^u	216.00	32.82	128.00	295.00
CO	CO^l	0.19	0.09	0.10	0.60
	CO^u	0.86	0.43	0.52	3.17

图 1 分别展示了区间变量下限及上限的 Pearson 相关性系数热力图。可以看出，对于区间变量的下限数据，AQI 与 PM_{10} 、 O_3 之间存在着较高的相关性，对于区间变量的上限数据，AQI 与 $\text{PM}_{2.5}$ 、 PM_{10} 之间

¹<https://www.mee.gov.cn/hjzl/dqhj>

存在着较高的相关性。另外, 不管是下限还是上限数据, 六个变量间的相关性都较小, 说明多重共线性风险较低, 后续可以将其作为解释变量构建回归模型。

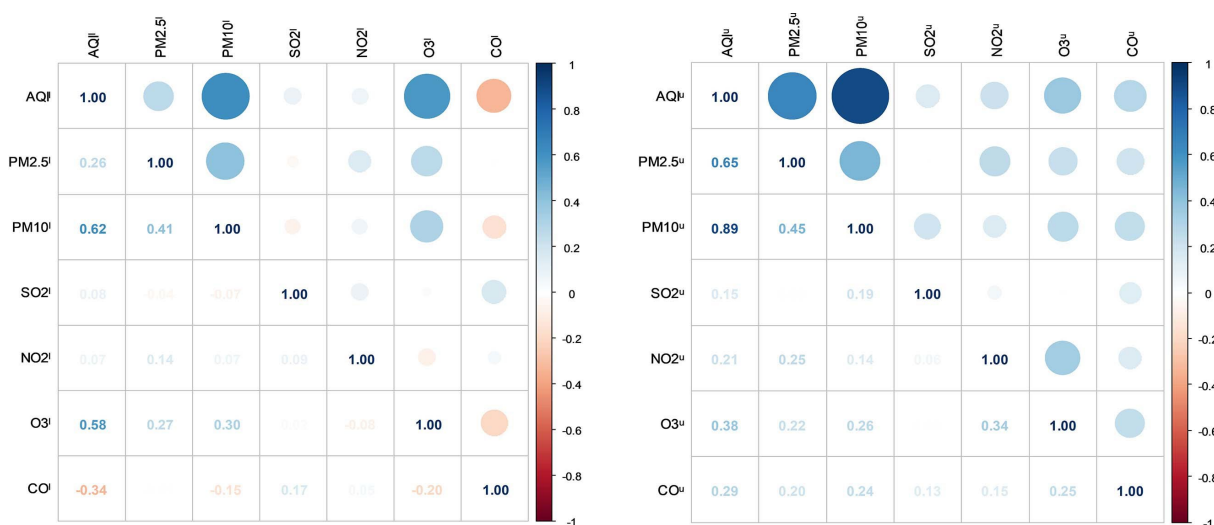


Figure 1. Correlation heatmap of variables

图 1. 各变量相关性热力图

3. 区间分位数回归模型

3.1. 模型介绍

相较于普通的回归方程模型, 分位数回归对异常值和非正态分布的误差项不敏感, 同时能清晰识别自变量对不同分位水平因变量的差异化影响, 估计结果更稳健、更丰富。针对区间数据, 现有研究常用方法主要包括中心法(利用自变量每个区间值的中点建立线性回归模型), 最小最大法(利用自变量每个区间值的上下限建立线性回归模型)以及中心半径法(利用自变量每个区间的中点与半径建立线性回归模型), 本文根据数据的结构特点, 选用最小最大法进行下面的分析, 具体的区间分位数回归模型表示如下:

$$\begin{cases} Q_{y_i^l}(\tau|x_i^l) = x_i^l \beta_\tau^l, \\ Q_{y_i^u}(\tau|x_i^u) = x_i^u \beta_\tau^u, \end{cases} \quad (1)$$

其中, 分位水平 τ 的取值范围为 $(0,1)$, y_i^l 为第 i 个响应变量的下限, y_i^u 为第 i 个响应变量的上限。解释变量 $x_i^l = (x_{i1}^l, \dots, x_{ip}^l)$, $x_i^u = (x_{i1}^u, \dots, x_{ip}^u)$ 为 $1 \times p$ 维的向量。回归系数 $\beta_\tau^l = (\beta_1^l, \dots, \beta_p^l)^T$, $\beta_\tau^u = (\beta_1^u, \dots, \beta_p^u)^T$ 为 $p \times 1$ 维的向量, 且依赖于分位水平 τ , 可以通过极小化得到 β_τ^l 和 β_τ^u 的估计值:

$$\begin{cases} \hat{\beta}_\tau^l = \arg \min_{\beta_\tau^l} \sum_{i=1}^n \rho_\tau(y_i^l - x_i^l \beta_\tau^l), \\ \hat{\beta}_\tau^u = \arg \min_{\beta_\tau^u} \sum_{i=1}^n \rho_\tau(y_i^u - x_i^u \beta_\tau^u), \end{cases} \quad (2)$$

其中 $\rho_\tau(u) = u(\tau - I_{(-\infty, 0)}(u))$ 为损失函数, $I(\cdot)$ 为指示函数。

3.2. 区间分位数回归的参数估计

本文将数据集的前 80% 数据作为训练集, 后 20% 数据作为测试集进行具体分析。表 2 展示了利用基

于最小最大法的区间分位数回归估计方法得到的各参数具体估计结果。

Table 2. Regression results for influencing factors of AQI

表 2. 空气质量指数影响因素回归结果

分位点	0.25	0.5	0.75
$PM_{2.5}^l$	-0.314	0.191	2.018
PM_{10}^l	0.910***	0.810***	0.618***
SO_2^l	0.513***	0.371**	0.559*
NO_2^l	0.344*	0.322	0.386
O_3^l	0.208***	0.233***	0.180***
CO^l	-4.446	-0.687	-7.042
$PM_{2.5}^u$	0.588***	0.561***	0.511***
PM_{10}^u	0.347***	0.347***	0.468***
SO_2^u	-0.034	-0.078	-0.134
NO_2^u	-0.090*	-0.136	-0.234
O_3^u	0.129***	0.167*	0.223***
CO^u	1.343	5.738	2.782

注：***、**、*分别表示在 1%、5%和 10%显著水平下通过显著性检验。

由上述估计结果可知, 各类污染物对 AQI 上下限的边际效应存在显著差异化。从 AQI 下限数据来看, PM_{10}^l 、 SO_2^l 、 O_3^l 为核心影响因素, 且均为正值, 这意味着 PM_{10} 、 SO_2 、 O_3 浓度下限与 AQI 下限呈正向关联, 且 PM_{10}^l 数值随着污染加重不断变小, 这可能是因为当月扩散条件最好时粗颗粒物浓度较高, 随着扬尘、沙尘等外源持续输入粗颗粒污染物, 且污染物难以被降水、风力等彻底清除, 当月空气污染水平下限被不断拉高, 而随着污染的加重, 在多种污染物的共同作用下, PM_{10} 浓度下限对 AQI 下限的影响会有所减弱。 SO_2 浓度下限高意味着当地工业源排放量较大时, SO_2 基础排放水平较高, 即使在扩散条件好, 大部分污染物被清除的情况下, SO_2 也很难降解至极低值, 从而拉高当月空气质量下限值; 臭氧浓度下限高意味着区域大气光化学污染基础水平高, 即使日照不足, 大气中仍存留着一定浓度臭氧, 从而拉高当月空气质量底线。从 AQI 上限数据来看, $PM_{2.5}^u$ 、 PM_{10}^u 、 O_3^u 为核心影响因素, 且均为正值, 这意味着 $PM_{2.5}^u$ 、 PM_{10}^u 、 O_3^u 浓度上限与 AQI 上限呈正向关联, 这可能是因为受当地大风、施工等事件干扰, $PM_{2.5}$ 、 PM_{10} 等颗粒物浓度急速上升, 短期内极易拉高当月空气污染上限; 另外 O_3^u 为正值且随着空气污染的加重, 数值不断增大, 这可能是因为轻度污染时, 臭氧前体物较少, 光化学反应受限, 臭氧浓度对整体空气污染影响较弱, 而随着污染的加重, 大气中臭氧前体物本底浓度升高, 一旦出现高温、强日照天气, 光化学反应被充分激活, 臭氧浓度快速上升, 极易拉高当月 AQI 峰值。

3.3. 模型性能评价

为验证模型的预测性能, 本文选用区间下限的均方根误差 $RMSE_L$, 区间上限的均方根误差 $RMSE_U$, 基于 Hausdorff 的均方根误差 $RMSE_H$ 以及重叠率 RI 四个指标来衡量预测精确度, 其表达式如下所示:

$$RMSE_L = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^l - \hat{y}_i^l)^2}, \quad (3)$$

$$RMSE_U = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^u - \hat{y}_i^u)^2}, \tag{4}$$

$$RMSE_H = \sqrt{\frac{1}{n} \sum_{i=1}^n (|y_i^l - \hat{y}_i^l| + |y_i^u - \hat{y}_i^u|)^2}, \tag{5}$$

$$RI = \frac{1}{n} \sum_{i=1}^n \frac{w(y_{it} \cap \hat{y}_{it})}{w(y_{it} \cup \hat{y}_{it})}. \tag{6}$$

表 3 分别展示了基于最小最大法, 中心半径法以及中心法的区间分位数回归模型(以下简化为 QR-MMM, QR-CRM, QR-CM)在不同分位点的预测结果, 可见 QR-MMM 方法的 $RMSE_L$ 、 $RMSE_U$ 、 $RMSE_H$ 基本都是最小的, RI 都是最大的, 说明 QR-MMM 方法整体上预测效果更好, 且在不同分位水平上预测性能稳定得好, 综上, QR-MMM 方法对于 AQI 上下限具有更优秀、更稳定的拟合与预测能力。

Table 3. Forecasting results of different models
表 3. 不同模型的预测结果

分位数	方法	$RMSE_L$	$RMSE_U$	$RMSE_H$	RI
0.25	QR-MMM	3.264	15.534	17.587	0.858
	QR-CRM	3.543	16.001	18.052	0.851
	QR-CM	17.872	17.408	33.966	0.771
0.5	QR-MMM	2.715	13.995	15.690	0.866
	QR-CRM	3.589	13.583	15.961	0.862
	QR-CM	16.111	18.463	32.797	0.785
0.75	QR-MMM	3.443	24.126	26.492	0.815
	QR-CRM	4.456	24.410	27.881	0.804
	QR-CM	19.471	20.674	35.589	0.765

4. 结论与建议

本文建立区间分位数回归模型, 利用全国空气质量指数与 $PM_{2.5}$ 、 PM_{10} 、 SO_2 、 NO_2 、 O_3 、 CO 浓度数据分析各污染物对空气质量指数的影响。结果表明, 从 AQI 下限数据来看, PM_{10}^l 、 SO_2^l 、 O_3^l 为核心影响因素; 从 AQI 上限数据来看, $PM_{2.5}^u$ 、 PM_{10}^u 、 O_3^u 为核心影响因素。随后, 我们利用基于最小最大法, 中心半径法以及中心法的区间分位数回归模型对 AQI 进行预测, 发现 QR-MMM 方法在预测时的综合性能最优, 这表明本文所提出的方法在 AQI 预测方面具有一定的优势, 也可为空气质量治理提供可靠的参考依据。

基于上述结论, 本文提出以下建议: (1) 常态化守底线。针对 PM_{10}^l 、 SO_2^l 、 O_3^l , 应聚焦常态化污染源长效治理, 抓好扬尘、燃煤、臭氧前体物减排, 加快能源消耗结构转型升级。一方面, 积极发展核能、太阳能等清洁能源, 提高能源利用效率, 扩大清洁能源应用规模, 从源头减少因化石燃料燃烧产生的污染物排放; 另一方面, 通过技术革新提高工业生产污染物的控制及收集率, 有效管控 PM_{10} 、 SO_2 、 O_3 的排放。(2) 应急期降峰值。针对 $PM_{2.5}^u$ 、 PM_{10}^u 、 O_3^u , 应做好短期重污染应急预案, 开展分季节差异化防控。夏季高温, 强日照时段, 加强臭氧监测与预警, 鼓励企业和市政工程中涉 VOCs 排放施工实施精细化管理; 春秋冬光照较弱时, 大风过境、沙尘过境等气象活动频发, 可将治理重心适当倾斜至颗粒物防

控, 停止开采、山体爆破等大扬尘工程, 降低颗粒物污染对空气质量峰值的影响。

本研究仍存在若干不足。一、研究所采用的数据观测周期较短, 难以充分捕捉变量之间的长期动态演变关系。二、模型设定中存在一定的变量遗漏问题, 可能会对实证估计结果产生干扰。三、研究尚未考虑样本的空间效应, 忽略了区域间可能存在的异质性影响。针对以上不足, 未来研究可以进一步拓展样本时间跨度, 纳入更多变量以完善模型, 引入空间计量方法, 以此丰富相关领域的研究结果。

参考文献

- [1] 张园. 空气污染、居民收入与健康资本[J]. 统计与决策, 2022, 38(18): 82-86.
- [2] 丁晨辉, 王威, 宋晓明. 空气污染对区域科技创新的双边影响效应分析[J]. 统计与决策, 2024, 40(6): 84-88.
- [3] 王永红, 邢艳春, 郝小娇. 空气污染与经济增长的空间效应及其关系检验[J]. 统计与决策, 2022, 38(15): 77-81.
- [4] 吴雪萍, 高明, 曾岚婷. 基于半参数空间模型的空气污染与经济增长关系再检验[J]. 统计研究, 2018, 35(8): 82-93.
- [5] 郑元琳. 基于多因素分析与多尺度分解的空气质量指数预测[D]: [硕士学位论文]. 济南: 山东财经大学, 2026.
- [6] 黄振庭, 黄仁博, 林霆柏, 等. 基于多元线性回归模型对惠州市空气质量指数实证分析[J]. 东莞理工学院学报, 2026, 33(1): 96-100.
- [7] 雷雨薇, 陈江滢, 尤添革, 等. 基于 XGBoost-SHAP 方法的福建省空气质量指数影响因素分析[J/OL]. 福建农林大学学报(自然科学版): 1-13. [https://doi.org/10.13323/j.cnki.j.fafu\(nat.sci.\).202506003](https://doi.org/10.13323/j.cnki.j.fafu(nat.sci.).202506003), 2026-09-30.
- [8] He, D.Y., Ding, J., Liu, Y.Z., Dong, Y.L., Ding, H.Y., Yang, X.T. and Liu, X.N. (2026) A Multi-Pollutant Air Quality Health Index for Assessing Mortality Risk of Ischemic Heart Disease in Jinan, China. *Scientific Reports*, **16**, Article No. 27015. <https://doi.org/10.1038/s41598-026-56942-3>
- [9] 倪洋, 张继勉, 曾强. 大气污染物复合暴露对呼吸系统疾病死亡效应的空气质量健康指数构建: 基于加权分位数和回归模型[J]. 环境与职业医学, 2026, 43(8): 944-950.
- [10] 史学良, 李梁, 赵清华. 基于改进 LSTM 网络的空气质量指数预测[J]. 统计与决策, 2021, 37(16): 57-60.
- [11] 杜艳芳. 基于分位数回归的空气质量指数分析[D]: [硕士学位论文]. 兰州: 兰州大学, 2016.
- [12] 景伊, 韩鑫钰. 基于 ARIMA 模型的运城市空气质量指数分析与预测[J]. 计算机时代, 2026(5): 53-57.
- [13] 方晓萍, 陈秀鑫, 褚琦, 等. 基于 PSO-GSA LSTM 模型的长江经济带空气质量指数预测研究[J]. 数理统计与管理, 2023, 42(1): 14-25.
- [14] 史佳宜. 基于 LSTM 网络的南京市空气质量指数预测[J]. 现代信息科技, 2026, 10(12): 140-145.
- [15] Tian, Z.D. and Yu, X.Y. (2026) Air Quality Index Prediction Based on Dimensionality Deduction and Optimized Echo State Network. *Chemometrics and Intelligent Laboratory Systems*, **276**, Article ID: 105796. <https://doi.org/10.1016/j.chemolab.2026.105796>
- [16] 陈博, 王志文. 变分模态分解融合改进河马优化算法的空气质量指数预测[J/OL]. 控制工程: 1-12. <https://doi.org/10.14107/j.cnki.kzgc.20250628>, 2026-09-21.
- [17] Wei, J.F., Zhao, T.C., Wei, R., Fan, X., Ji, P. and Zhang, X. (2026) Application of Stacked Transformer-XGBoost Model and SHAP Interpretability in Air Quality Index Prediction in Linfen City. *iScience*, **29**, Article ID: 116396. <https://doi.org/10.1016/j.isci.2026.116396>
- [18] Xiong, J.J., Liang, C.B., Wang, D.G. and Yang, M. (2026) A Novel Deep Learning Framework Integrating Temporal Convolution and Bidirectional LSTM for Multi-City AQI Forecasting. *Information Sciences*, **757**, Article ID: 123941. <https://doi.org/10.1016/j.ins.2026.123941>
- [19] Billard, L. and Diday, E. (2000) Regression Analysis for Interval Valued Data. Springer, 369-374.
- [20] Lima Neto, E.D.A. and de Carvalho, F.D.A.T. (2008) Centre and Range Method for Fitting a Linear Regression Model to Symbolic Interval Data. *Computational Statistics & Data Analysis*, **52**, 1500-1515. <https://doi.org/10.1016/j.csda.2007.04.014>
- [21] Souza, L.C., Souza, R.M.C.R., Amaral, G.J.A. and Silva Filho, T.M. (2017) A Parametrized Approach for Linear Regression of Interval Data. *Knowledge-Based Systems*, **131**, 149-159. <https://doi.org/10.1016/j.knosys.2017.06.012>
- [22] 呼延璇. 基于变分模态分解及和声搜索算法的空气质量指数区间预测[D]: [硕士学位论文]. 兰州: 兰州大学, 2022.

- [23] Yang, H.F., GAO, Y.Y., Zhao, F.S. and Wang, J.Z. (2024) An Intelligent Interval Forecasting System Based on Fuzzy Time Series and Error Distribution Characteristics for Air Quality Index. *Environmental Research*, **251**, Article ID: 118577. <https://doi.org/10.1016/j.envres.2024.118577>
- [24] Suo, C., Li, Y.P., Sun, J. and Yin, S. (2018) An Air Quality Index-Based Multistage Type-2-Fuzzy Interval-Stochastic Programming Model for Energy and Environmental Systems Management under Multiple Uncertainties. *Environmental Research*, **167**, 98-114. <https://doi.org/10.1016/j.envres.2018.07.001>
- [25] 石逸. 区间型数据多维标度分析方法研究及应用[D]: [硕士学位论文]. 厦门: 厦门大学, 2021.