

基于时间序列的一线城市空气微粒排放物建模分析与预测

宋婉宁, 覃红霞, 吕 萌, 周影慧, 高婉琪, 陈佳男, 白晓东

大连民族大学理学院, 辽宁 大连

收稿日期: 2024年10月25日; 录用日期: 2024年11月18日; 发布日期: 2024年11月27日

摘 要

大气可吸入颗粒物污染已成为危害人类健康的主要因素之一。对城市空气微粒排放物的分析与预测可以为改善空气质量及制定污染防控措施提供科学依据。本文以2014~2018年北京市、上海市和广州市的空气质量指数(Air Quality Index, 简称: AQI)、PM2.5浓度和PM10浓度的逐日数据为研究样本。数据经过异常值剔除和插值处理后, 采用ARIMA和SARIMA模型对时间序列特征进行分析, 并对未来三期空气质量进行预测。结果显示, 真实值均位于95%置信区间内, 大部分数据的相对精度低于10%, 模型拟合度较高。此外, 研究表明, 北京、上海和广州的AQI、PM2.5浓度及PM10浓度呈下降趋势, 短期内空气质量显著提升。

关键词

PM2.5, PM10, 空气质量指数(AQI), 时间序列分析, ARIMA模型

Time Series-Based Modeling Analysis and Forecasting of Air Particulate Emissions in Tier-1 Cities

Wanning Song, Hongxia Qin, Meng Lyu, Yinghui Zhou, Wanqi Gao, Jianan Chen, Xiaodong Bai

School of Science, Dalian Minzu University, Dalian Liaoning

Received: Oct. 25th, 2024; accepted: Nov. 18th, 2024; published: Nov. 27th, 2024

Abstract

Atmospheric particulate matter pollution has become one of the primary factors endangering human

文章引用: 宋婉宁, 覃红霞, 吕萌, 周影慧, 高婉琪, 陈佳男, 白晓东. 基于时间序列的一线城市空气微粒排放物建模分析与预测[J]. 应用数学进展, 2024, 13(11): 4990-5000. DOI: 10.12677/aam.2024.1311481

health. The analysis and forecasting of urban particulate emissions can provide a scientific basis for improving air quality and formulating pollution control measures. This paper uses daily data on the Air Quality Index (AQI), PM_{2.5} concentration, and PM₁₀ concentration from 2014 to 2018 in Beijing, Shanghai, and Guangzhou as research samples. After removing outliers and performing interpolation, ARIMA and SARIMA models were applied to analyze the time series characteristics and predict air quality for the next three periods. The results indicate that the actual values fall within the 95% confidence interval, with most data exhibiting a relative accuracy of less than 10%, and the models show a high degree of fit. Additionally, the study reveals a downward trend in AQI, PM_{2.5} concentration, and PM₁₀ concentration in Beijing, Shanghai, and Guangzhou, with significant short-term improvements in air quality.

Keywords

PM_{2.5}, PM₁₀, Air Quality Index (AQI), Time Series Analysis, ARIMA Model

Copyright © 2024 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

研究背景及意义

北京、上海和广州作为中国的一线城市，在城市发展、人才吸引、经济实力、教育普及、国际竞争力、交通和科技创新等方面表现突出[1]，这些城市同样也是中国科技创新的中心，集聚了大量高科技企业和研发机构，经济活动、能源消耗、工业规模和机动车数量居全国前列[2]。然而，城市的快速发展也带来了人口密集、交通拥堵和环境污染等问题，尤其是环境污染对城市发展带来了巨大压力。因此，改善环境质量已成为急需解决的问题。

为此，我们开展北京、上海和广州这三个一线城市环境空气中 PM_{2.5} 和 PM₁₀ 细颗粒污染物未来趋势的研究。拟建立恰当的时间序列模型分析各城市细颗粒污染的变化规律和特征，并对未来趋势作出预测分析，有助于人们认识到空气污染对人体健康造成的危害，帮助有关部门提出改善环境质量的有效对策。

2. 预备知识及数据来源

1) 预备知识

a) ARIMA 模型

满足下式的模型为 ARIMA(p, d, q) 模型

$$\Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \quad E(\varepsilon_t) = 0, \quad \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, \quad E(x_s \varepsilon_t) = 0, \quad \forall s < t,$$

$$\nabla^d = (1 - B)^d; \quad \Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

为平稳可逆的 ARMA(p, q) 模型的自回归系数多项式； $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ 为平稳可逆的 ARMA(p, q) 模型的移动平滑系数多项式[3]。

b) 季节模型

对于一些非平稳的时间序列来讲，经过 d 阶差分 and D 阶差分后，可以变为平稳的时间序列 $\{y_t\}$ ，即：

$$y_t = \nabla^d \nabla_s^D x_t$$

若 $\{y_t\}$ 满足季节周期为 s 的 $ARMA(p,q) \times (P,Q)_s$ 模型, 则称 $\{x_t\}$ 为季节周期为 s , 非季节阶数为 p,d,q 、季节阶数为 P,D,Q 的乘积季节求和和回归移动平均模型, 记作 $SARIMA(p,d,q) \times (P,D,Q)_s$ 。

c) AIC 准则与 BIC 准则

赤池信息准则(Akaike Information Criterion, 简称: AIC)是评估统计模型的复杂度和衡量统计模型拟合优良性的一种标准, 是由日本统计学家 Akaike 于 1973 年提出的, 它是基于最小信息量思想的准则, 在实际应用中, 我们应优先考虑 AIC 值最小的模型。

一般情况下, AIC 可以表示为:

$$AIC = 2\ln(\text{模型的极大似然函数值}) + 2(\text{模型中未知参数的个数})$$

AIC 准则为模型选择提供了重要依据, 但当样本量接近 n 时, AIC 基准确定的模型不收敛于实际模型, 模型的阶数高于模型的实际阶数。此时, Akaike 提出贝叶斯信息准则(Bayesian Information Criterion, 简称: BIC)以弥补 AIC 准则的不足。定义如下式:

$$BIC = 2\ln(\text{模型的极大似然估计值}) + \ln(n)(\text{模型中未知参数的个数})$$

BIC 准则考虑了样本个数, 样本数量多, 可以防止模型精度过高造成的模型复杂度过高。

2) 数据来源

本文数据来源真气网(<https://www.aqistudy.cn/historydata/>)。选取了北京、上海、广州三个一线城市的 AQI、PM2.5、PM10 2014 年 1 月 1 日-2018 年 12 月 31 日的逐日数据, 共 1826 个样本。

3. 模型建立与预测

1) 数据预处理

a) 异常值处理

通过绘制北京市、上海市和广州市的 AQI、PM2.5、PM10 时序图, 我们发现样本数据存在异常值, 有极个别天数的数据比总体平均值高, 为避免后续对建模产生不良影响, 我们需要对数据进行标准化, 剔除异常数据。

首先, 绘制箱形图观察序列中异常值的大致分布情况。在箱线图中, 数据的上四分位数(Q3)与下四分位数(Q1)分别位于箱子的上下底, 小于 $Q1-1.5IQR$ 或大于 $Q3+1.5IQR$ 的值被定义为异常值。见图 1。

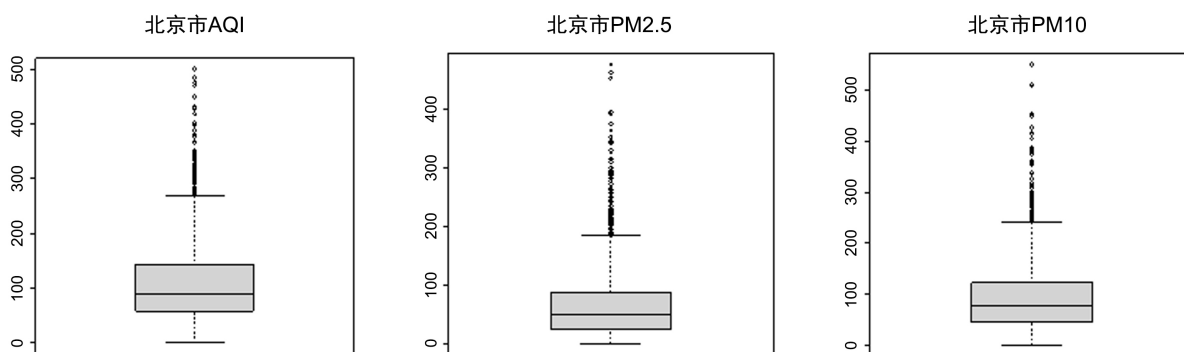


Figure 1. Box diagram before pretreatment of AQI, PM2.5 and PM10 in Beijing

图 1. 北京市 AQI、PM2.5、PM10 预处理前箱线图

根据箱型图的特性, 我们对数据中的异常值进行处理。首先, 分别选取 $\text{coef} = 1.1$ 、 $\text{coef} = 1.5$ 、 $\text{coef} = 1.2$ 的盒子长度, 使处理前的盒子长度必须为处理后的 1.1 倍、1.5 倍、1.2 倍, 在异常值处理后, 可以

看到处理后的箱型图没有明显的特异点,使得后续数据不会对进一步的建模造成较大的影响。然后,我们对数据中的缺失部分进行插值处理,处理后的数据可以进行后续分析。见图 2。

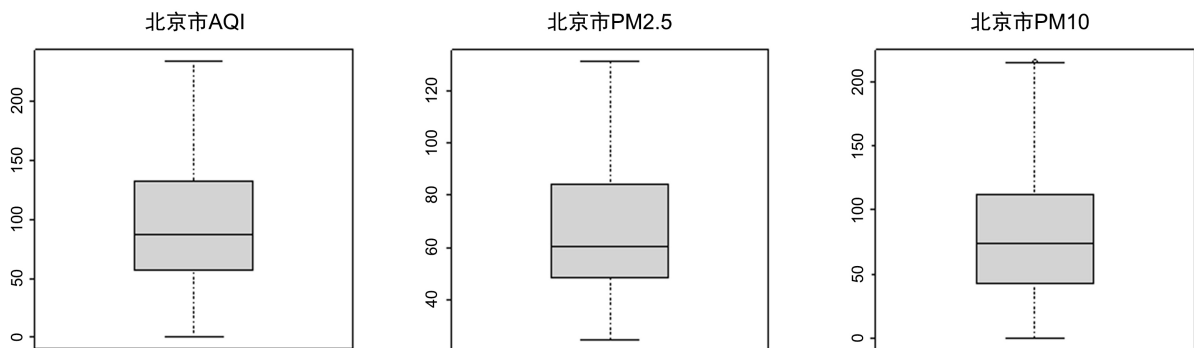
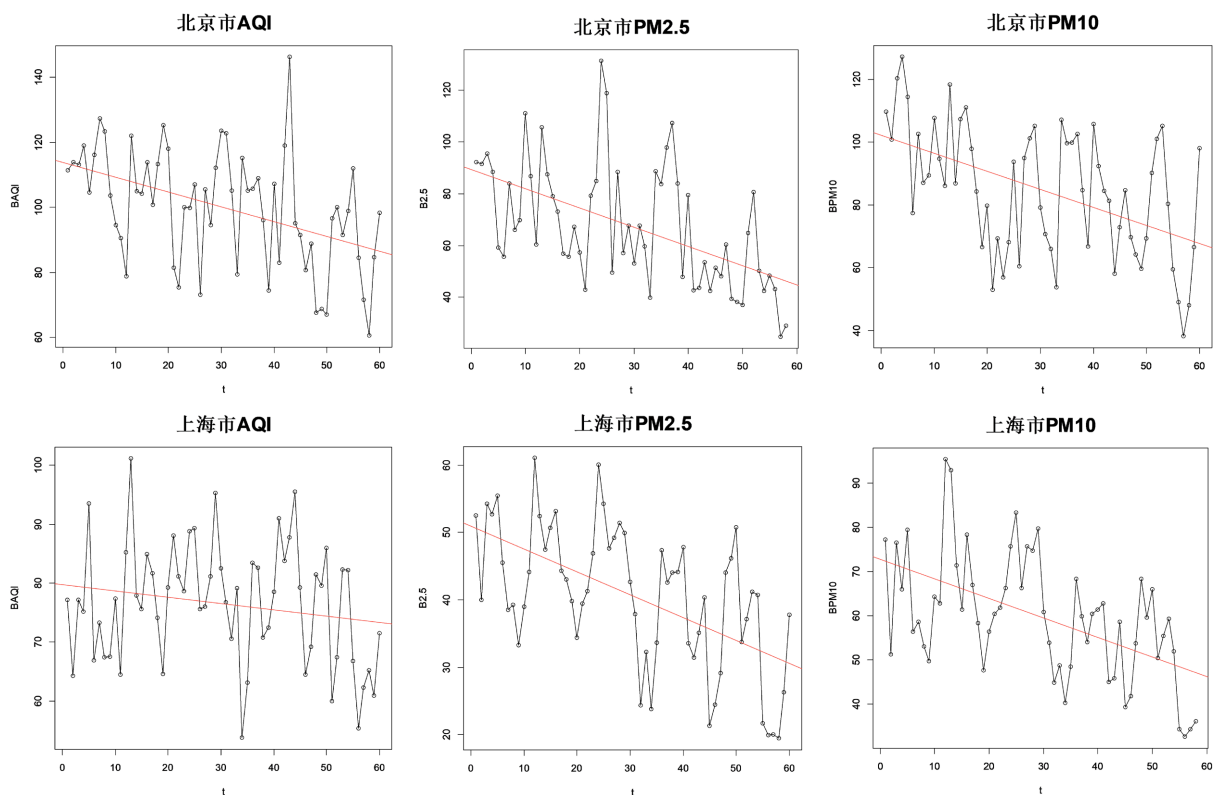


Figure 2. Box diagram of pretreatment of AQI, PM2.5 and PM10 in Beijing
图 2. 北京市 AQI、PM2.5、PM10 预处理后箱线图

b) 平稳性检验

为了方便观察大量数据的序列趋势,我们将数据进行分组,利用平均值进行模型的建立,在均值处理后我们使用这 60 组数据进行时序图的绘制。

由线性拟合图可以看出大部分时序图存在明显的线性趋势,需要进行平稳化处理,我们首先对存在明显趋势的序列进行一阶差分操作,发现部分处理后的序列仍存在明显的季节性周期,进而进行一阶十二步的季节差分处理使序列平稳。见图 3。



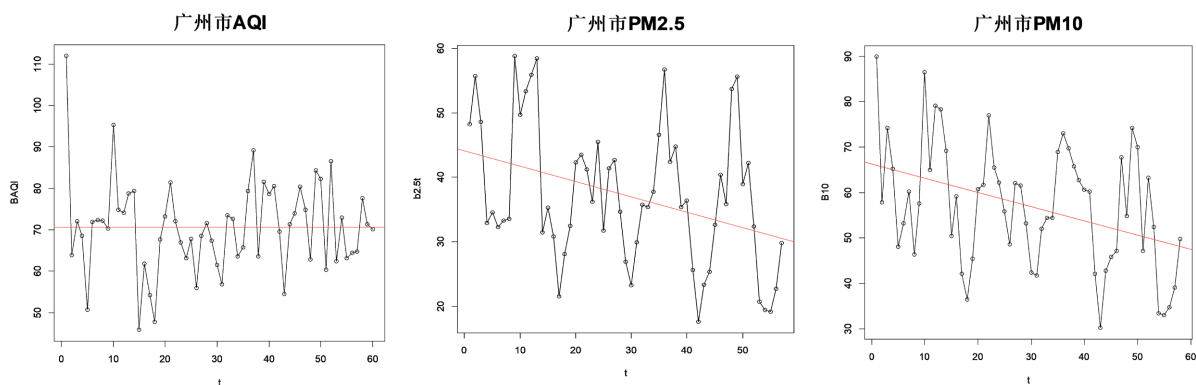


Figure 3. Linear fitting of AQI, PM2.5 and PM10 in Beijing, Shanghai and Guangzhou

图 3. 北京市、上海市、广州市 AQI、PM2.5、PM10 线性拟合图

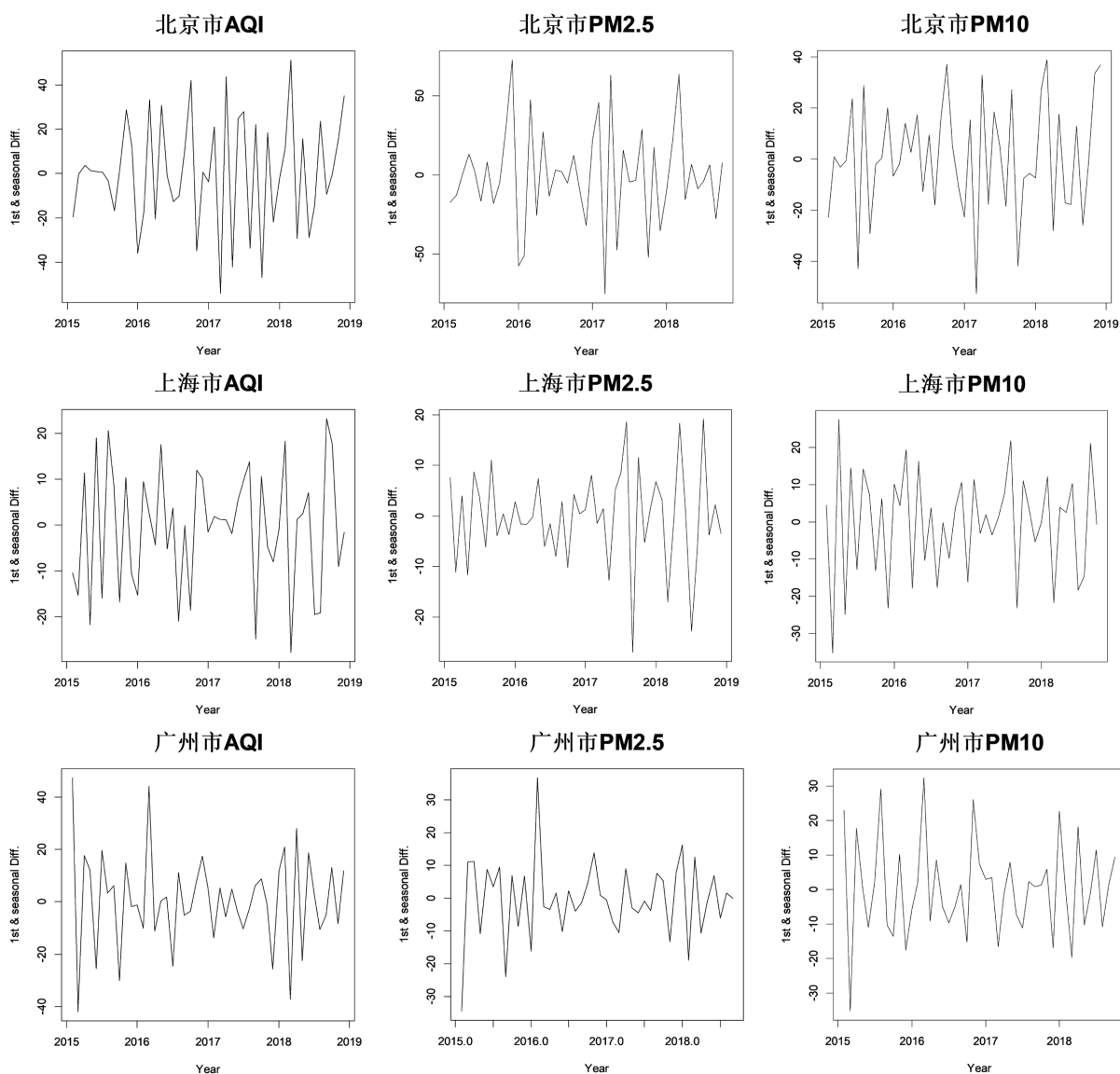


Figure 4. Sequence diagram after smoothing

图 4. 经过平稳化处理后的时序图

由图 4 可以看到, 经过差分处理后的序列呈现出平稳的性质。但图检验的方法仍存在很强的主观色彩, 为了使结果更加准确, 我们进一步对差分处理后的序列进行单位根扩展的迪克富勒检验(Augmented Dickey-Fuller Tested, 简称 ADF 检验加以辅助判别。根据 ADF 检验结果可知所有序列 P 值均小于 0.05, 因此可判断序列全部平稳。

c) 序列白噪声检验

进行平稳性判别后, 我们还需要对序列进行白噪声检验, 确保序列具有统计研究意义。我们对判别为平稳的序列进行 Ljung-Box 检验[4]。检验结果显示, 所有 P 值均小于显著性水平 0.05, 因此可以判断序列不是白噪声序列, 序列波动具有统计规律, 可以进行统计分析。见图 5。

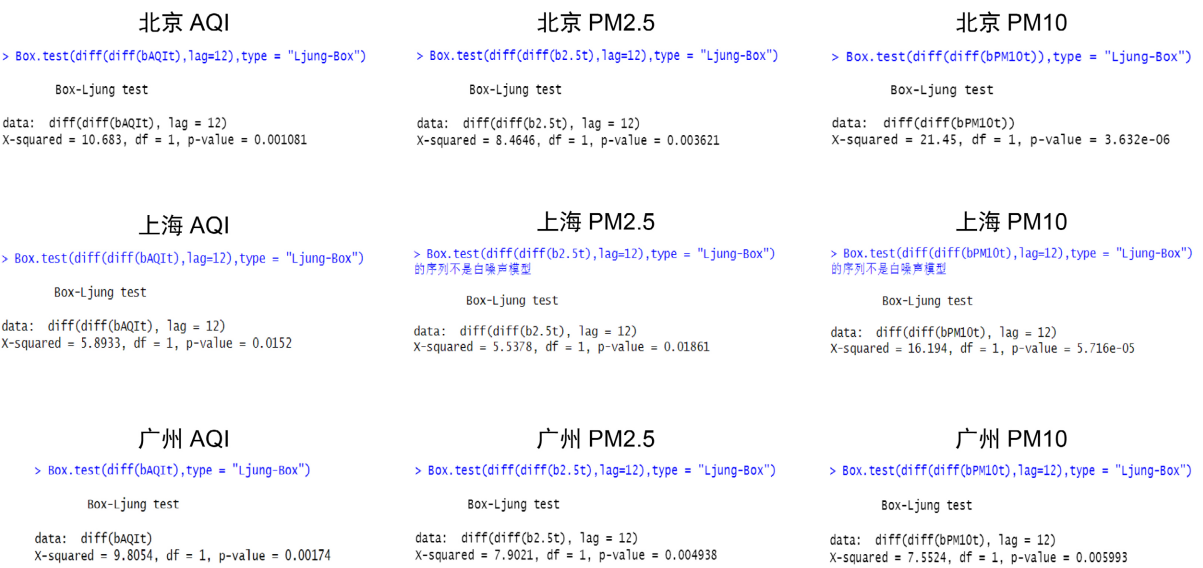


Figure 5. White noise test for nine sequences
图 5. 九个序列的白噪声检验

2) 模型识别与参数估计

我们通过绘制序列的 ACF 和 PACF 图来确定模型的自回归系数和移动平均系数, 但由于此种方法具有较强的主观性, 模型参数确定不唯一, 存在着多个模型均通过检验的情况, 此时我们引入 AIC 和 BIC 准则对不同参数的模型进行多次测试并辅助 auto.arima()函数自动定阶, 以确定最优拟合模型如下[5]。九个模型全部通过残差检验, 检验结果 P 值显著大于显著性水平 0.05, 表明序列中的大部分信息已成功提取, 模型拟合成功, 进一步检验参数的显著性, 剔除冗余的参数。最终确定模型见表 1:

Table 1. AQI index, PM2.5 mass concentration, PM10 mass concentration model formula, AIC and BIC values of the three cities
表 1. 三个城市的 AQI 指数、PM2.5 质量浓度、PM10 质量浓度模型公式及 AIC、BIC 值

模型		公式	AIC	BIC
AQI	$SARIMA(1,1,1) \times (1,0,0)_{12}$	$(1 - 0.2806B)(1 - B)y_t = (1 + 0.9268B)(1 - 0.3038B^{12})\varepsilon_t$	500.93	509.24
北京 PM2.5	$SARIMA(1,1,1) \times (1,0,0)_{12}$	$(1 - 0.2849B)(1 - B)y_t = (1 + 0.8926B)(1 - 0.2692B^{12})\varepsilon_t$	509.49	517.67
PM10	$ARIMA(1,1,1)$	$(1 - 0.4868B)(1 - B)y_t = (1 + 0.923B)\varepsilon_t$	510.38	516.62

续表

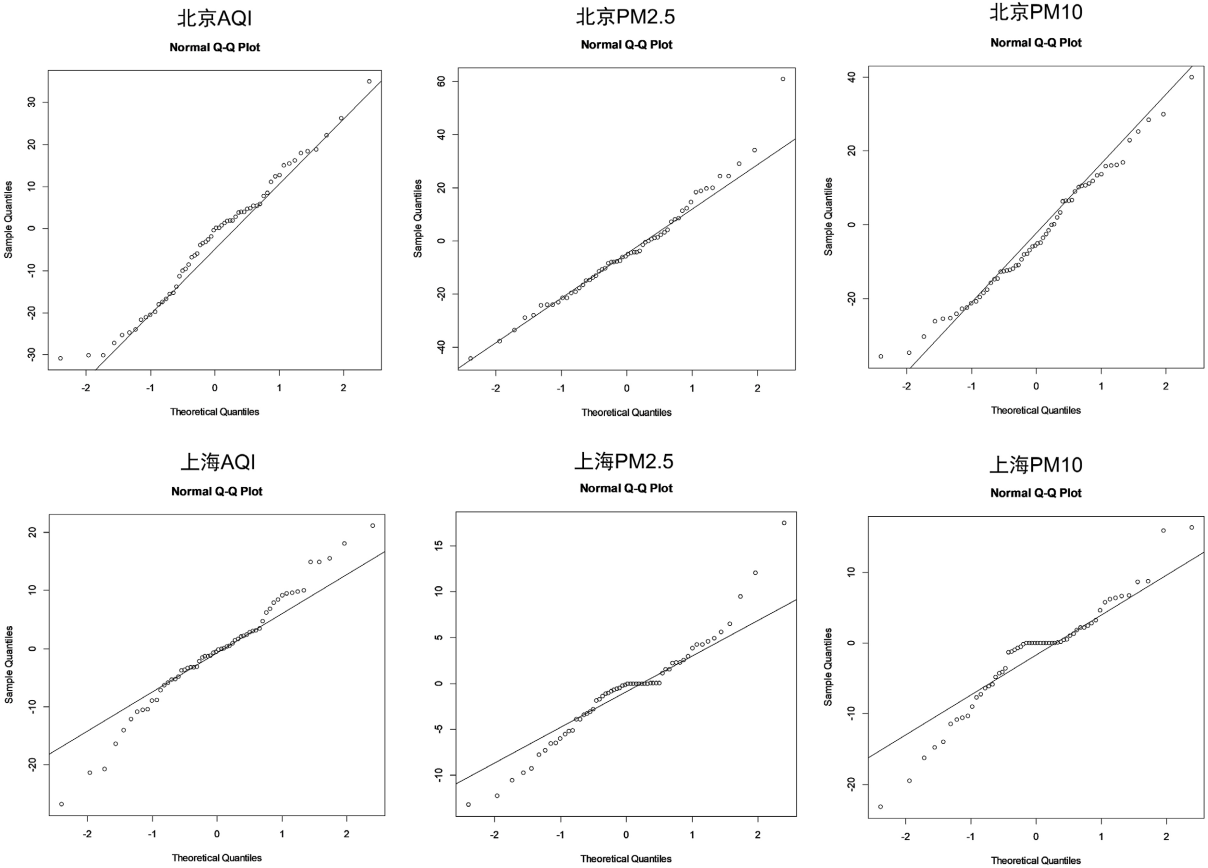
	AQI	$SARIMA(0,1,2) \times (1,0,0)_{12}$	$(1+0.5748B)(1-B)y_t = (1+0.3177B)(1-0.2494B^{12})\varepsilon_t$	442.02	450.32
上海	PM2.5	$SARIMA(0,1,1) \times (1,1,0)_{12}$	$(1+0.8404B)(1-B)y_t = (1+0.5993B^{12})\varepsilon_t$	316.46	322.01
	PM10	$SARIMA(0,1,1) \times (1,1,1)_{12}$	$(1+0.6783B)(1-B)y_t = (1+0.3341B)(1+0.4976B)\varepsilon_t$	339.1	246.33
广州	AQI	$SARIMA(2,1,0) \times (0,0,1)_{12}$	$(1+0.7772B)(1-B)y_t = (1+0.3565B)(1-0.4268B^{12})\varepsilon_t$	461.83	470.14
	PM10	$SARIMA(2,1,0) \times (0,1,1)_{12}$	$(1+0.5915B)(1-B)y_t = (1+0.3732B)(1+0.3882B^{12})\varepsilon_t$	355.79	363.02
	PM2.5	$SARIMA(0,1,1) \times (0,1,1)_{12}$	$(1+0.8858B)(1-B)y_t = (1+0.7793B^{12})\varepsilon_t$	317.74	323.09

3) 模型诊断性检验

a) Q-Q 图检验

b) 残差白噪声检验

利用 Ljung-Box 检验, 对上述城市和指标的残差进行白噪声检验。根据所得结果可知各个指标的 P 值大于 0.05。因此可以判断该序列是随机波动的白噪声, 说明模型拟合效果好。见图 6 和图 7。



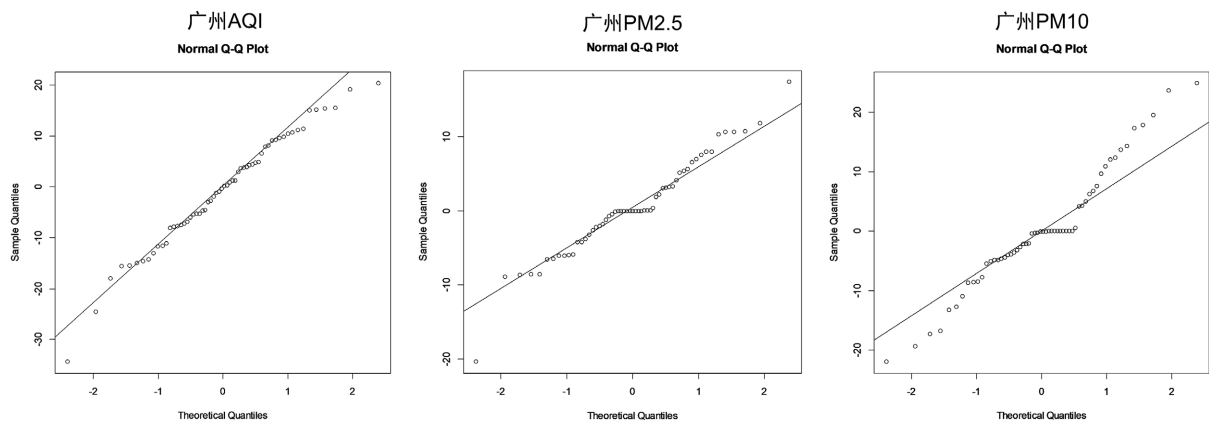


Figure 6. AQI, PM2.5 and PM10 Q-Q of Beijing, Shanghai and Guangzhou

图 6. 北京市、上海市、广州市 AQI、PM2.5、PM10 Q-Q 图



Figure 7. Residual tests for nine models

图 7. 九个模型的残差检验

4) 模型预测及精度分析

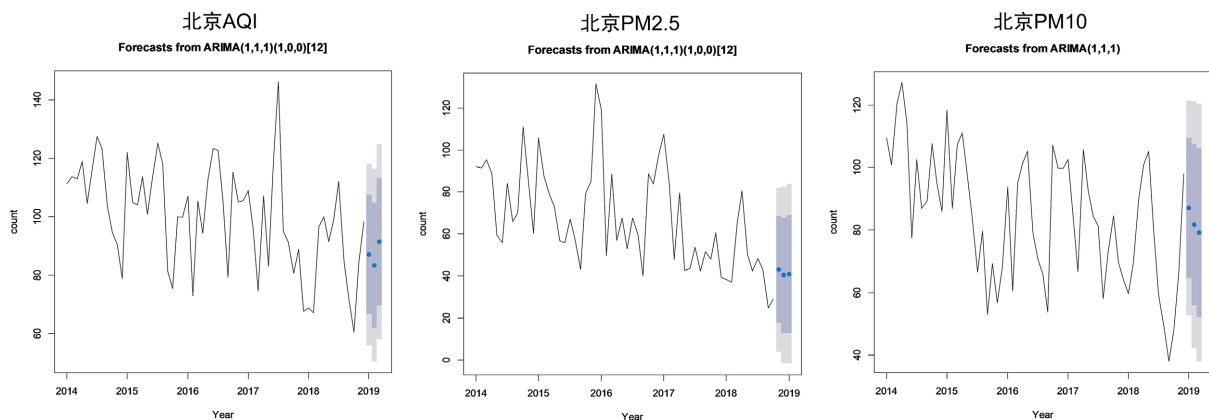


Figure 8. Fitting of the three forecast trends of AQI, PM2.5 and PM10 in Beijing

图 8. 北京市 AQI、PM2.5、PM10 三期预测趋势拟合图

Table 2. Accuracy analysis of AQI, PM2.5 and PM10 models in Beijing
表 2. 北京市 AQI、PM2.5、PM10 模型精度分析

	北京市 AQI				北京市 PM2.5				北京市 PM10			
	真实值	预测值	95%置信区间	相对精度	真实值	预测值	95%置信区间	相对精度	真实值	预测值	95%置信区间	相对精度
第一期	80.55	87.24	[56.12, 118.35]	0.08	71.32	43.10	[4.09, 82.11]	0.39	79.61	87.07	[52.69, 121.44]	0.09
第二期	80.04	83.49	[50.49, 116.50]	0.04	38.10	40.47	[-1.44, 82.38]	0.06	71.96	81.75	[42.30, 121.21]	0.14
第三期	85.29	91.56	[58.13, 124.99]	0.07	51.42	41.05	[-1.72, 83.82]	0.20	80.45	79.16	[37.90, 120.43]	0.02

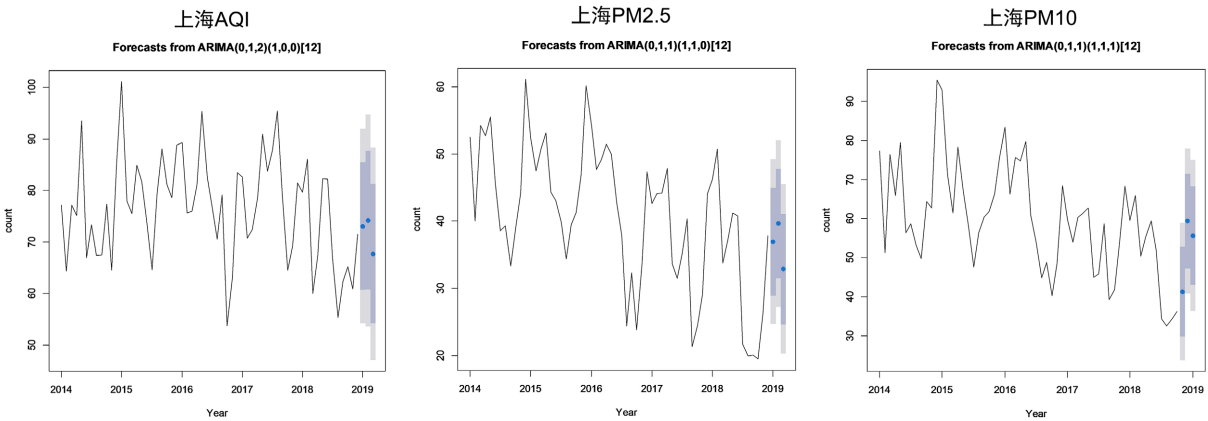


Figure 9. Fitting chart of three forecast trends of AQI, PM2.5 and PM10 in Shanghai
图 9. 上海市 AQI、PM2.5、PM10 三期预测趋势拟合图

Table 3. Accuracy analysis of AQI, PM2.5 and PM10 models in Shanghai
表 3. 上海市 AQI、PM2.5、PM10 模型精度分析

	上海市 AQI				上海市 PM2.5				上海市 PM10			
	真实值	预测值	置信区间	相对精度	真实值	预测值	置信区间	相对精度	真实值	预测值	置信区间	相对精度
第一期	79.68	67.71	[47.03, 88.38]	0.15	50.68	32.87	[20.29, 45.45]	0.35	61.00	55.63	[36.29, 74.97]	0.09
第二期	71.07	69.56	[48.78, 90.33]	0.02	39.63	36.46	[23.73, 49.19]	0.08	50.54	46.19	[26.04, 66.35]	0.09
第三期	85.87	73.27	[52.40, 94.15]	0.15	32.94	29.56	[16.68, 42.44]	0.10	65.00	48.75	[27.82, 69.68]	0.25

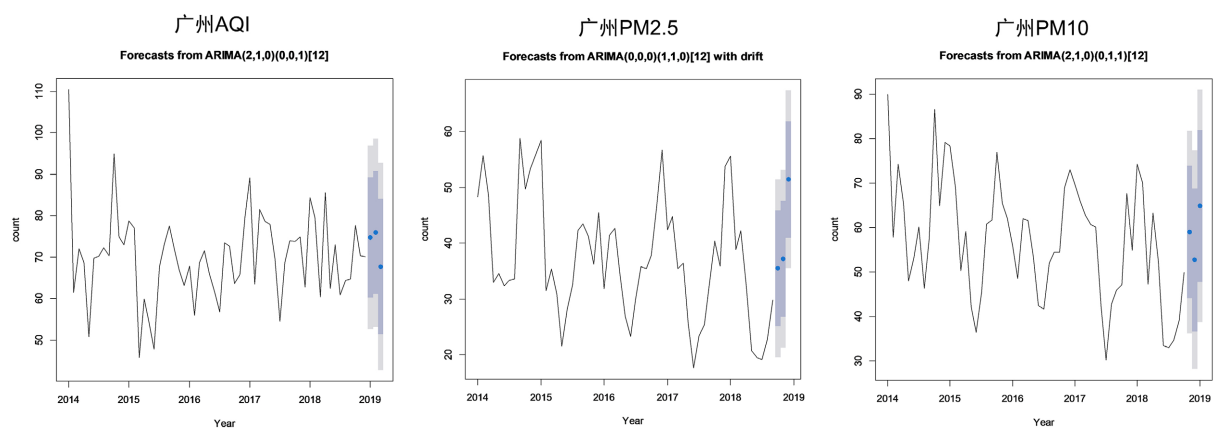


Figure 10. Fitting chart of the three forecast trends of AQI, PM2.5 and PM10 in Guangzhou
图 10. 广州市 AQI、PM2.5、PM10 三期预测趋势拟合图

Table 4. Accuracy analysis of AQI, PM2.5 and PM10 models in Guangzhou
表 4. 广州市 AQI、PM2.5、PM10 模型精度分析

	广州市 AQI				广州市 PM2.5				广州市 PM10			
	真实值	预测值	置信区间	相对精度	真实值	预测值	95%置信区间	相对精度	真实值	预测值	置信区间	相对精度
第一期	72.65	74.77	[52.60, 96.95]	0.03	38.97	35.70	[20.57, 50.83]	0.08	59.20	59.01	[36.24, 81.79]	0.01
第二期	48.68	75.92	[53.21, 98.64]	0.56	31.97	36.26	[21.036, 51.49]	0.13	48.65	52.73	[28.13, 77.33]	0.08
第三期	67.19	67.71	[42.71, 92.71]	0.01	47.90	46.88	[31.56, 62.21]	0.02	69.35	64.88	[38.76, 91.00]	0.06

模型预测趋势拟合图见图 8~10 所示，模型精度分析见表 2~4。

4. 结论

城市大气污染治理是急需解决的问题。本文使用 ARIMA、SARIMA 模型，对 2014~2018 年北京市、上海市、广州市三座城市的 AQI、PM2.5、PM10 等指标进行了时间序列分析和预测[6]，并给出三期预测值。预测结果显示，真实值在 95%置信区间内，可知模型预测精度较好。此外，预测结果也显示北京市、上海市、广州市的 AQI、PM2.5、PM10 浓度有逐渐下降的趋势，空气质量逐渐提高。大气污染不仅会破坏我们赖以生存的大自然，也会对我们的身体造成危害[7]，城市大城市大气污染治理需要全社会共同努力[8]，相信在未来我国空气质量趋势越来越好。

基金项目

大连民族大学创新创业训练计划国家级大创项目(项目编号：20241202643)。

参考文献

- [1] 李颖. 发达城市城乡结合地带居住环境规划设计研究[D]: [硕士学位论文]. 淄博: 山东轻工学院, 2012.

-
- [2] 龚良俊. 城市大气可吸入颗粒污染物对车内空气质量的影响研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2022.
 - [3] 白晓东. 应用时间序列分析[M]. 北京: 清华大学出版社, 2017.
 - [4] 聂巧平, 孙淇铭. 时间序列模型的 Ljung-Box 检验的发展与应用综述[J]. 统计与管理, 2022, 37(4): 101-107.
 - [5] 刘璋温. 赤池信息量准则 AIC 及其意义[J]. 数学的实践与认识, 1980(3): 64-72.
 - [6] 陈佩弟, 周明璋, 肖婷婷, 等. 基于 ARIMA 模型的乌鲁木齐市颗粒物 PM_{10} 污染特征及趋势分析[J]. 职业与健康, 2024, 40(15): 2086-2090.
 - [7] 史英丽. 对城市经济快速增长下环境污染问题的分析[J]. 市场研究, 2017(6): 7-8.
 - [8] 唐湘博, 曹利珍, 马志文, 等. 不同排放控制策略的城市减污降碳协同效应及空气质量改善研究[J/OL]. 环境科学研究, 2024: 1-15. <https://doi.org/10.13198/j.issn.1001-6929.2024.08.23>, 2024-09-22.