

## 监测与评价

基于 SARIMA 模型的二氧化氮时间序列  
预测研究王一龙<sup>1</sup>, 申云霞<sup>2</sup>, 陈晓红<sup>1</sup>

(1. 烟台市芝罘区环境保护局, 山东烟台 264000;

2. 烟台市高新区住房和城乡建设管理局, 山东烟台 264670)

**摘要:** 基于烟台主城区 2010 年 1 月至 2018 年 6 月的 NO<sub>2</sub> 浓度数据, 利用 Eviews 统计软件建立了季节自回归移动平均模型(SARIMA), 经过序列平稳化、模型识别及模型诊断, SARIMA(2, 0, 3) (1, 1, 1)<sup>12</sup> 模型的相对误差可控制率在 5% 以内, 拟合效果较为理想。拟合及预测结果表明, 烟台市主城区 NO<sub>2</sub> 浓度具有季节性特征, 呈递增趋势, 随着预测步长的延长, 预测误差逐渐增大。SARIMA 模型目前适合进行短期预测, 今后可结合非线性动力学方法, 对其进行改进。

**关键词:** SARIMA 模型; NO<sub>2</sub>; 预测

**中图分类号:** X823

**文献标识码:** A

**文章编号:** 1006-8759(2019)03-0051-04

Time series prediction of NO<sub>2</sub> based on SARIMA ModelWANG Yi-long<sup>1</sup>, SHEN Yun-xia<sup>2</sup>, CHEN Xiao-hong<sup>1</sup>

(1. Environmental Protection bureau of Zhifu district Yantai city, Yantai 264000, China;

2. Housing and Construction Administration of Yantai High-tech Zone, Yantai 264670, China)

**Abstract:** Based on NO<sub>2</sub> concentrations in main urban area of Yantai from January 2010 to June 2018, a seasonal autoregressive integrated moving average model (SARIMA) was established with the statistical software Eviews. After smooth sequence, model recognition and diagnosis, the relative error of SARIMA (2, 0, 3) (1, 1, 1)<sup>12</sup> model could be controlled within 5% and the fitting results were satisfied. The fitting and predicted results showed that NO<sub>2</sub> concentrations in main urban area of Yantai performed seasonal characteristics and an increasing trend. With the extension of prediction time, the error of SARIMA model increases. Thus, this model was suitable for short-time forecasting. In further study, this model may be improved by combining the theory of non-linear dynamics.

**Key words:** SARIMA model; NO<sub>2</sub>; Prediction

## 0 引言

随着城市化进程的加快, 大气污染已引起越来越多的重视, 关于 AQI、PM<sub>10</sub>、PM<sub>2.5</sub> 建模预测的报道<sup>[1-3]</sup>也日益增多, 但关于 NO<sub>2</sub> 的报道较少。近期研究表明, 大气污染物中 NO<sub>2</sub> 对成人哮喘患者肺功能与生命质量影响显著<sup>[4]</sup>, 空间分布呈现显著的季节性特点<sup>[5,6]</sup>, 具有明显的趋势变动性和周

期性二重特征。近年来, GM(1, 1)模型、神经网络模型、ARIMA 模型在对具有二重特征的时间序列进行预测方面得到广泛应用。GM(1, 1)模型处理趋势性时间序列有较好的效果<sup>[7]</sup>, 但处理周期波动性时间序列, 短期预测结果比不上 ARIMA 模型<sup>[8]</sup>。神经网络预测模型具有强大的学习能力, 可轻松实现非线性拟合, 但易出现局部最优、过拟

收稿日期: 2019-01-22

第一作者简介: 王一龙(1983-), 男, 山东临沂人, 工程师, 硕士。E-mail: wangyilong1983@163.com

引用格式: 王一龙, 申云霞, 陈晓红. 基于 SARIMA 模型的二氧化氮时间序列预测研究[J]. 能源环境保护, 2019, 33(3):51-54.

合等问题<sup>[9,10]</sup>,在处理具有线性特征的时间序列效果上,一般不如 ARIMA 模型<sup>[11]</sup>。通过构建季节性自回归移动平均模型(SARIMA),可揭示二重特征时间序列本身的结构与规律,挖掘系统的固有特性<sup>[12]</sup>,使预测分析更加准确。王坤,阮金梅,邓妮根据曲靖市 2014 年 1 月-2017 年 12 月 AQI 数据,建立了 SARIMA 模型,利用模型进行预测,并取得较好的效果<sup>[13]</sup>。本研究选取烟台市主城区 2010 年 1 月至 2018 年 6 月 102 个 NO<sub>2</sub> 月均浓度值构建 SARIMA 模型,利用该模型预测 2018 年下半年烟台市主城区 NO<sub>2</sub> 月均浓度值并进行预测效果评价,为有效地治理和控制烟台市主城区空气污染提供科学依据。

## 1 基于 SARIMA 模型构建

### 1.1 数据来源

本文所选数据为烟台市主城区 2010 年 1 月至 2018 年 6 月 102 个 NO<sub>2</sub> 月均浓度值数据,数据来源于山东省环保厅空气质量自动监控运行管理系统。其中将 2010 年 1 月至 2018 年 6 月的 NO<sub>2</sub> 月均浓度值数据作为实验数据集,2018 年 1 月至 12 月的数据作为模型预测准确率的验证数据集。

### 1.2 数据的平稳化处理与检验

将 2010 年 1 月至 2018 年 6 月的月度 NO<sub>2</sub> 月均浓度值数据记为 X<sub>t</sub> 序列,做序列 X<sub>t</sub> 的时序图,如图 1 所示。由图 1 可知,X<sub>t</sub> 序列呈下降趋势,并自带周期 k=12 的季节波动,故对该序列进行一阶十二步差分消除时间序列的季节性,所得序列记为 Y<sub>t</sub>,在此基础上采用 ADF 检验进行平稳性检验,结果如表 1 所示。由表 1 可知,Y<sub>t</sub> 序列为平稳序列(t=-5.8422,p=0.0000),满足 Box-Jenkins 模型平稳的建模要求,因此对 Y<sub>t</sub> 序列可以建 SARIMA(p, d, q) (P, D, Q)<sup>s</sup> 模型。

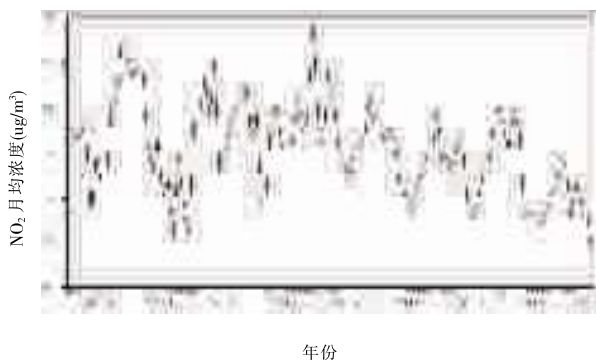


图 1 NO<sub>2</sub> 月均浓度值时序图

表 1 ADF 检验结果

	t 统计量	p 值
ADF 检验统计量	-5.842157	0.0000
测试临界值	1 % level	-3.505595
	5 % level	-2.894332
	10 % level	-2.584325

### 1.3 模型的定阶与识别

对于 SARIMA(p, d, q) (P, D, Q)<sup>s</sup> 模型, X<sub>t</sub> 经过一阶季节性差分消除了季节性特征变为平稳序列,所以差分阶数 d=0,季节差分阶数 D=1,模型中的其他参数的阶数,利用自相关函数和偏自相关函数的统计特性,可识别模型中的 p, d 阶数,由图 2 可以看出,Y<sub>t</sub> 序列的自相关系数在 2 阶截尾,偏自相关系数在 3 阶截尾,其余大多数落在随机区间以内。综合考虑以 p=1, 2 与 q=1, 2, 3 进行参数估计,另外由于 k=12 时,样本的自相关系数和偏相关系数都不显著为 0,所以 P=Q=1。

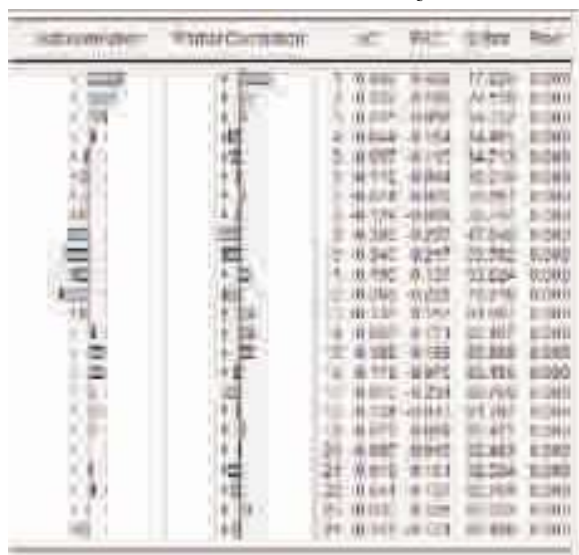


图 2 NO<sub>2</sub> 月均浓度值序列相关图、偏自相关图

对上述组合进行参数估计和检验,部分模型的部分参数不显著 ( $p > 0.05$ ),对冗余参数进行剔除,存在以下待选模型:(2, 0, 3) (1, 1, 1)<sup>12</sup>、(1, 0, 1) (1, 1, 1)<sup>12</sup>、(2, 0, 2) (1, 1, 1)<sup>12</sup> 进行比较评价。在选择模型时,一般根据 AIC 值以及 SC 值最小的准则和修正的拟合优度 R<sup>2</sup> 最大的原则作出选择。

表 2 为各预选模型建模和检验结果,根据模型的 AIC、SC 最小以及拟合度最优准则,最终确定 SARIMA(2, 0, 3) (1, 1, 1)<sup>12</sup> 为最优模型。

表 2 模型检验结果

预选模型	调整的 R 平方	AIC	SC
(2, 0, 3) (1, 1, 1) <sup>12</sup>	0.761871	6.391724	6.637064
(1, 0, 1) (1, 1, 1) <sup>12</sup>	0.626270	6.791690	6.913446
(2, 0, 2) (1, 1, 1) <sup>12</sup>	0.743675	6.453638	6.668311

## 2 模型参数估计及检验

对预选模型 SARIMA(2, 0, 3) (1, 1, 1)<sup>12</sup> 模型进行参数估计和检验,结果如表 3 所示。经检验,该模型的残差、实际值、拟合值初步证明了模型的拟合程度较高,如图 3 所示。模型残差序列的相关图、偏自相关图不存在序列相关,说明残差序列中的信息已被充分提取,如图 4 所示。综上所述,预测模型 SARIMA(2, 0, 3) (1, 1, 1)<sup>12</sup> 拟合程度好,适合解释当前时间序列和用来预测。

表 3 模型参数估计

变量	系数	标准误差	t 统计量	p 值
C	-3.434079	0.380997	-9.013404	0.0000
AR(1)	0.804378	0.078112	10.29780	0.0000
AR(2)	-0.548250	0.052569	-10.42911	0.0000
SAR(12)	-0.202436	0.081860	-2.472971	0.0159
MA(1)	-0.665455	0.130309	-5.106738	0.0000
MA(2)	0.674718	0.122325	5.515763	0.0000
MA(3)	0.311952	0.122422	2.548178	0.0131
SMA(12)	-0.987727	0.025380	-38.91731	0.0000

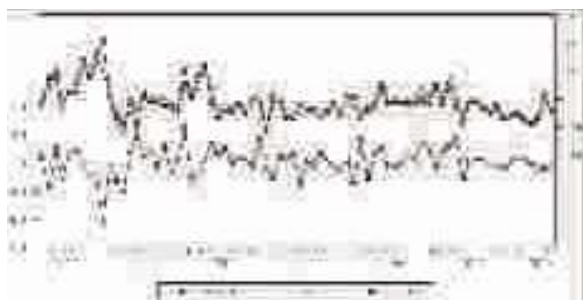


图 3 SARIMA(2, 0, 3) (1, 1, 1)<sup>12</sup> 模型真实值、拟合值、残差序列图



图 4 SARIMA(2, 0, 3) (1, 1, 1)<sup>12</sup> 残差序列的相关图、偏自相关图

## 3 SARIMA 模型预测

根据建立的 SARIMA(2, 0, 3) (1, 1, 1)<sup>12</sup> 模型,

对 2018 年上半年的 NO<sub>2</sub> 月均浓度值进行静态预测,静态预测结果见表 4,预测值与实际值比较发现,相对误差较小,说明模型具有很高的预测精度,能很好的拟合效果。预测值与真实值之间的偏差,可能是拟合模型建立过程中不可避免的误差所致。

表 4 模型静态预测结果

月份	实际值	预测值	相对误差(RPE)
1	38	37	-0.0263
2	26	25	-0.0385
3	33	33	0.0000
4	29	30	0.0345
5	28	29	0.0357
6	18	18	0.0000

根据建立的 SARIMA(2, 0, 3) (1, 1, 1)<sup>12</sup> 模型,对 2018 年下半年的 NO<sub>2</sub> 月均浓度值进行动态预测,预测结果见表 5,由于随之预测长度的增大,移动平均项逐步淡出预测模型。因此,长期预测效果不是很理想,动态预测误差较静态预测误差大。

表 5 模型动态预测结果

月份	实际值	预测值	相对误差(RPE)
7	14	14	0.0000
8	16	17	0.0625
9	22	24	0.0909
10	26	29	0.1153
11	38	32	-0.1578
12	40	35	-0.1428

## 4 结论

本文通过建立 SARIMA(2, 0, 3) (1, 1, 1)<sup>12</sup> 模型,对烟台市主城区 2018 年上半年的 NO<sub>2</sub> 月均浓度值进行样本内静态预测,对 2018 年下半年的 NO<sub>2</sub> 月均浓度值进行样本外动态预测,得出如下结论:

(1) SARIMA(2, 0, 3) (1, 1, 1)<sup>12</sup> 模型的样本内静态预测表明,2018 年 1 月至 6 月的样本内预测与真实值差距较小,相对误差控制率均在 5% 内,表明模型拟合效果比较理想。

(2) SARIMA(2, 0, 3) (1, 1, 1)<sup>12</sup> 模型的样本外动态预测表明,下半年烟台市主城区 NO<sub>2</sub> 月均浓度值呈现递增的趋势,随着预测步长的延长,预测误差将逐渐增大。

(3) 该模型虽然可以较为精确的预测烟台市主城区 NO<sub>2</sub> 月均浓度变化趋势,但无法揭示导致该地区 NO<sub>2</sub> 浓度变化的其他影响因素所起到的作用,加之影响 NO<sub>2</sub> 浓度的因素错综复杂,SARIMA

模型只能进行短期预测,今后可结合非线性动力学方法等对其进行改进。

### 参考文献:

- [1] 杨蒙,叶馨,刘文.基于灰色动态模型簇的 AQI 预测及研究[J].科技创新导报,2016,13(36):18+20.
- [2] 尹安琪,林愿仪,林伟俊,等.基于 PSO-BP 神经网络预测广州市日均 PM<sub>10</sub> 浓度[J].中国卫生统计,2016,33(5):763-766.
- [3] 王勘之,曾沛,刘永辉.上海市 PM<sub>2.5</sub> 浓度的分析与预测[J].数学的实践与认识,2017,47(15):210-217.
- [4] 王俊.空气污染对成人哮喘患者肺功能与生命质量的影响[D].北京协和医学院;中国医学科学院;清华大学医学部;北京协和医学院中国医学科学院,2017.
- [5] 杨帆.降雨对大气颗粒物和气态污染物的清除效率及机制[D].南昌大学,2015.
- [6] 张良.石家庄市"十一五"期间环境空气质量变化趋势分析及预测研究[D].河北科技大学,2013.
- [7] 杨克磊,张振宇,和美.应用灰色 GM(1,1)模型的粮食产量预测研究[J].重庆理工大学学报(自然科学版),2015(4):124-127.
- [8] 敖希琴.基于 SARIMA 模型的安徽省 CPI 预测[J].蚌埠学院学报,2017,6(3):83-86.
- [9] 马翠红,赵士超.遗传神经网络结合 LIBS 技术对钢液 Mn 元素定量分析[J].现代电子技术,2018,41(15):169-173.
- [10] Zhang J, Ding W. Prediction of Air Pollutants Concentration Based on an Extreme Learning Machine: The Case of Hong Kong[J]. Int J Environ Res Public Health, 2017, 14(2):114.
- [11] 熊志斌.基于 ARIMA 与神经网络集成的 GDP 时间序列预测研究[J].数理统计与管理,2011, 30(2):306-314.
- [12] 严宙宁,牟敬锋,赵星,等.基于 ARIMA 模型的深圳市大气 PM<sub>2.5</sub> 浓度时间序列预测分析[J].现代预防医学,2018,45(2):220-223,242.
- [13] 王坤,阮金梅,邓妮.基于 SARIMA 模型的曲靖市空气质量指数预测[J].曲靖师范学院学报,2018,37(3):25-29.
- [14] 张莹,邵毅,王式功,等.北京市空气污染物对呼吸系统疾病门诊人数的影响[J].中国环境科学,2014,34(9):2401-2407.
- [15] 洪新如,孙庆华.大气污染对心血管疾病影响及其机制的研究进展[J].中国心血管杂志,2008,13(3):223-226.
- [16] 穆泉,张世秋.2013年1月中国大面积雾霾事件直接社会经济损失评估[J].中国环境科学,2013,33(11):2087-2094.
- [17] 廖玉学,彭朝琼,余淑苑,等.2006-2016年深圳市大气污染物趋势变化研究[J].环境卫生学杂志,2017,7(6):439-444.
- [18] 陶燕,羊德容,兰岚,等.兰州市空气污染对呼吸系统疾病入院人数的影响[J].中国环境科学,2013,33(1):175-180.
- [19] 张秉玲,牛静萍,曹娟,等.兰州市大气污染与居民健康效应的时间序列研究[J].环境卫生学杂志,2011,1(2):1-6.
- [20] 王宇红,梁婷婷,李盛,等.兰州市医院儿科呼吸系统疾病日门诊量与空气污染的相关性研究[J].环境卫生学杂志,2017,7(5):356-362.
- [21] 中华人民共和国环境保护部,国家质量监督检验检疫总局.环境空气质量标准:GB3095-2012.[S].北京:中国环境科学出版社,2012.
- [22] Wang Y, McElroy M B, Boersma K F, et al. Traffic restrictions associated with the Sino-African summit: reductions of NOx detected from space [J]. Geophysical Research Letters, 2007,34,L08814,doi: 10.1029/2007GL029326.
- [23] 周丹.东莞市大气环境污染来源分析与模拟研究[D].北京:北京化工大学,2008.
- [24] 程念亮,李云婷,张大伟,等.2013-2014年北京市 NO<sub>2</sub> 时空分布研究[J].中国环境科学,2016,36(1):18-26.
- [25] 平凉市统计局,国家统计局平凉调查队.2015年平凉市国民经济和社会发展统计公报[N].平凉日报,2016-03-23(002).

(上接第 59 页)

由于能源开采运输生产造成大气污染。本文只对 2015 年 3 市的演变特征加以分析,缺乏大尺度演变规律的研究。此外,城市空气质量还受气象因素及人口、经济发展水平等影响,这些将在今后的工作中加以研究。

### 参考文献:

- [1] 蒯雪芹,王岱.中国城市空气质量时空演化特征及社会经济驱动力[J].地理学报,2016,71(8):1357-1371.
- [2] Guan D B, Su X, Zhang Q, et al. The socioeconomic drivers of China's primary PM<sub>2.5</sub> emissions[J]. Environmental Research Letters, 2014, 9:1-9.
- [3] Jacob D J, Winner D A. Effect of climate change on air quality [J]. Atmospheric Environment, 2009, 43(1):51-63.
- [4] Parrish D D, Singh H B, Molina L, et al. Air quality progress in North American megacities: A review [J]. Atmospheric Environment, 2011, 45(39):7015-7025.
- [5] 赵荣仙,刘传姚,张兵,等.2003-2010年西南地区省会城市空气污染指数的动态变化[J].环境与健康杂志,2013,30(5):422-425.
- [6] 黄德生,张世秋.京津冀地区控制 PM<sub>2.5</sub> 污染的健康效益评估[J].中国环境科学,2013,33(1):166-174.
- [7] 张江华,郭常义,许慧慧,等.上海市大气污染与某医院呼吸系统疾病门诊量关系的时间序列研究[J].环境与职业医学,2014,31(11):846-851.
- [8] 吴彧,张衍桑,刘玲,等.郑州市空气污染对每日门诊人次影响的滞后效应和收获效应[J].环境与健康杂志,2011,28(3):233-237.