

The Prediction and Research of Air Quality in Chengdu Based on ARMA Model

Binduo Niu, Yuliang Yin

School of Economics, Beijing Technology and Business University, Beijing
Email: nbdjy12@126.com

Received: Nov. 30th, 2016; accepted: Dec. 13th, 2016; published: Dec. 23rd, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

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

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



Open Access

Abstract

In recent years, many cities in China face the problem of smog. The problem of environmental pollution is becoming more and more serious. It has many negative effects on the lives of the residents, so monitoring and predicting the atmospheric pollution has become an important part in the study of air quality. In this article we select ambient air quality composite index per month of Chengdu from January 2013 to December 2015, use the ARMA model (the autoregressive moving average model), establish the air quality forecasting ARMA model of Chengdu through the analysis of the stationarity, correlation and partial correlation of the datum, and test the application of model. The research results show that the ARMA model can effectively predict the ambient air quality composite index in Chengdu, so it provides a basis for accurately monitoring the situation of air pollution and the reasonable use of environmental pollution control measures.

Keywords

ARMA Model, Ambient Air Quality Composite Index, Prediction

基于ARMA模型的成都市空气质量预测研究

牛炳朵, 尹玉良

北京工商大学经济学院, 北京
Email: nbdjy12@126.com

收稿日期: 2016年11月30日; 录用日期: 2016年12月13日; 发布日期: 2016年12月23日

摘要

近年中国许多城市出现雾霾天气, 环境污染日益严重, 监控预测大气污染已成为空气质量研究中的重要组成部分。本文采用2013年1月到2015年12月成都市月环境空气质量综合指数, 通过对数据序列进行平稳性、自相关和偏相关性的分析, 建立成都市空气质量预测ARMA模型, 检验模型的适用性, 并基于实际数据进行预测分析。研究表明, 所建立的ARMA模型能够有效预测成都市空气质量指数, 从而为监控空气污染状况、合理运用环境污染治理措施提供依据。

关键词

ARMA模型, 环境空气质量综合指数, 预测

1. 引言

2016年5月, 第二届联合国环境大会发布的最新报告称, 每年世界各地约有700万人死于空气污染。去年, 美国的一项研究报告显示, 中国的空气污染平均每天会造成约4000人死亡, 占中国总死亡人数的17%, 这些数字都表明中国城市环境污染问题日益突出。环境对居民生活和经济发展产生了巨大影响, 不仅造成生产力下降等看得见的损失, 还会影响投资环境和产品的市场竞争力。因此, 环境的发展变化和大气污染的监控预测已成为空气质量研究的重要组成部分。

2016年1月1日, 环境空气质量标准(Ambient air quality standards) GB 3095-2012代替GB 3095-1996和GB 9137-88实施。新空气质量标准是为了贯彻《中华人民共和国环境保护法》和《中华人民共和国大气污染防治法》, 保护和改善生活、生态环境, 保障人类身体健康而制定的。该标准规定了环境空气功能区分类、标准分级、污染物项目、平均时间及浓度限值、监测方法、数据统计的有效性规定及实施与监督等内容, 并提出了环境空气质量综合指数来全面综合地反应空气质量。随着环境空气质量新标准的实施, 人们对空气质量的关注程度会更高, 因此, 科学地评价、分析和预测环境空气质量综合指数就成一项必不可少的研究工作。

2016年是十三五规划的开局之年, 四川省委提出了把成都市建设成为西部地区门户枢纽、西部金融中心、西部经济核心增长极的规划建议。在此经济规划下, 成都市的环境质量愈发重要。而成都市位于四川盆地的“锅底”, 四川盆地特殊的地理位置和地形特征, 这使得成都市云雾较多, 且对该市污染物的垂直扩散形成抑制作用。这对市民身体健康和当地经济发展产生了较大威胁和不良影响。所以准确预测成都地区的空气质量, 不仅与居民的生活息息相关, 也与成都市的经济发展密不可分。

2. 文献综述

在已有研究中, 许多学者通过不同方法对空气质量进行预测预报。孙柏峰利用灰色理论预测模型对吉林省未来十年的污染物排放量进行预测与分析, 并且计算了该预测模型的平均相对误差、后验算比以及小误差概率, 来衡量模型的准确度[1]。宋宇辰和甄莎运用BP神经网络法和时间序列法, 对包头市的二氧化硫、二氧化氮和可吸入颗粒物的年份浓度值和月份浓度值进行预测, 进而对包头市日后的空气质量进行分析[2]。朱悦, 郑洪波等运用具有适用性广, 预测准确率高等优点的灰色系统预测模型对大连市未来五年空气中二氧化硫的浓度变化趋势进行预测, 对其模型的精度和可行性进行了分析和检验, 并与目前常用的指数平滑法的预测结果进行了对比, 最终得出了灰色系统预测模型精度高的结论[3]。

上述文献所涉及到的预测模型中,对空气质量预测常用的有时间序列法、灰色理论预测法、回归分析法、专家系统法、神经网络法等。其中, BP 神经网络模型适用于中长期的预测,优点是逼近效果好,计算速度快,不需要建立数学模型。缺点是无法表达和分析被预测系统的输入和输出间的关系,预测人员无法参与预测过程,收敛速度慢,难以处理海量数据,算法不完备。灰色模型,即对原始数据作累加生成得到近似的指数规律再进行建模的处理方法。优点是不需要很多的数据,能解决历史数据少、序列的完整性及可靠性低的问题;运算简便,易于检验。但缺点是只适用与中长期的预测,和指数增长的预测,数据波动性对预测精度有较大影响,预测精度较差。时间序列预测法(ARMA 模型预测)适应于短期或中期预测。时间序列分析预测法的前提是假定事物的过去会延续到未来,事物的现实是历史发展的结果,而事物的未来又是现实的延伸,事物的过去和未来是有联系的。其优点是简单易行,便于掌握,能够充分运用原时间序列的各项数据,计算速度快,采用组合的时间序列或者把时间序列和其他模型组合效果更好。缺点是不能反映事物的内在联系,不能分析因素之间的相关关系。

在空气质量预测方面,可将城市空气质量的月度变化看成是一组时间序列,基于过去和未来空气质量之间的关联性,利用时间序列的方法研究城市空气污染指数在时间尺度上的演变。综合比较各种方法的优缺点,并结合所收集的数据特征, ARMA 模型与其他预测模型相比,具有模型清晰、操作简单、便于执行、所需指标数据简单、预测精度高等优点,因此更加适用于本文的研究。

3. ARMA 模型相关理论

3.1. 理论背景

博克斯-詹金斯法,简称 B-J 法或 ARMA 法,是以美国统计学家 Geogre E.P. Box 和英国统计学家 Gwilym M. Jenkins 的名字命名的一种时间序列预测方法。博克斯-詹金斯法依据的基本思想是:将预测对象随时间推移而形成的数据序列视为一个随机序列,即除去个别由于偶然原因引起的观测值外,时间序列是一组依赖于时间的随机变量。这组随机变量所具有的依存关系或自相关性表征了预测对象发展的延续性,而这种自相关性一旦被相应的数学模型描述出来,就可以从时间序列的过去值及现在值预测其未来值。

Box-Jenkins 预测模型包括自回归模型 AR、综合模型 I 和移动平均模型 MA。通过模型设定、参数估计、诊断检验以及预测四个步骤,最终确立最优的时间序列模型 ARIMA。

3.2. 模型定义

ARMA 模型(自回归滑动平均模型)是研究时间序列的重要方法,由 AR 模型(自回归模型)与 MA (滑动平均模型)为基础“混合”构成,其目的是由时间序列的过去值及现在值预测其未来的值[4]。

自回归模型(AR):反映经济变量的当前值与其过去值的关系。AR(p)模型是回归模型的一种形式,其一般形式为:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + u_t$$

移动平均模型(MA):反映经济变量当前值与当前及过去误差项的关系。一般形式的 MA(q)模型可以表示为:

$$y_t = u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} \cdots - \theta_q u_{t-q}$$

两者结合的模型自回归移动平均模型 ARMA(p,q)模型。如果时间序列 y_t 是它的当期和前期的随机误差项以及前期值的线性函数,即可表示为:

$$y_t = c + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

3.3. 模型识别

对于一个平稳时间序列的预测问题, 首先要考虑的是寻求与它拟合最好的预测模型, 而模型的识别与阶数的确定则是选择模型的关键。

AR(p)模型的识别。若序列的偏自相关函数在 p 以后截尾, 而且自相关系数是拖尾的, 则此序列是自回归 AR(p)序列。

MA(q)模型的识别。若序列的自相关函数在 q 以后截尾, 而且偏自相关系数是拖尾的, 则此序列是移动平均 MA(q)序列。

ARMA(p,q)模型的识别。若序列的自相关函数和偏自相关系数都是拖尾的, 则此序列是自回归移动平均 ARMA(p,q)序列。至于模型中 p 和 q 的识别, 则要从低阶开始逐步试探, 直到定出合适的模型为止。模型分类如表 1 所示。

3.4. 模型估计

经过模型识别, 确定了时间序列模型的结构和阶数后, 需要对模型进行估计。

估计模型的方法较多, 最常用的有最小二乘法、矩估计和利用自相关系数直接估计。

4. 成都市空气质量预测研究

4.1. 数据来源与处理

所选数据为中国环境监测总站(<http://www.cnemc.cn>)环境质量报告中 2013 年 1 月到 2015 年 12 月的环境空气质量综合指数(下文均简称综合指数)。环境空气质量综合指数是新的空气质量标准中所提出的一个衡量空气质量的综合指数, 它是一个描述城市环境空气质量综合状况的无量纲指数, 综合考虑了 SO_2 、 NO_2 、 PM_{10} 、 $\text{PM}_{2.5}$ 、 CO 、 O_3 等六项污染物的污染程度, 环境空气质量综合指数数值越大表明综合污染程度越重。

一般情况下, 人们通过所在城市的 AQI 数据, 可以了解该城市的空气质量情况。在之前的研究文献中, 也多采取空气质量指数 AQI (Air Quality Index)来进行分析。而在本文中, 我们选用环境空气质量综合指数进行研究, 这是因为, 环境空气质量综合指数可以更加全面、综合地反映地区空气质量。

4.2. 模型建立

4.2.1. 平稳性检验

对综合指数序列(CI)画时间序列图, 并做 ADF 平稳性检验。结果分别如图 1 和表 2 所示。

由图 1 可以看出, CI 始终在一个常数附近随机波动, 而且波动的范围有界, 无明显趋势和周期特征。初步判断 CI 序列是平稳序列。由图 2 所示的单位根检验可以看出, t 统计量的值小于临界值, p 值小于 0.05, 进一步验证了 CI 是平稳序列。

Table 1. Model classification

表 1. 模型分类

自相关系数(ACF)	偏自相关系数(PACF)	模型
拖尾	p 阶截尾	AR(p)
p 阶拖尾	截尾	MA(q)
拖尾	拖尾	ARMA(p,q)

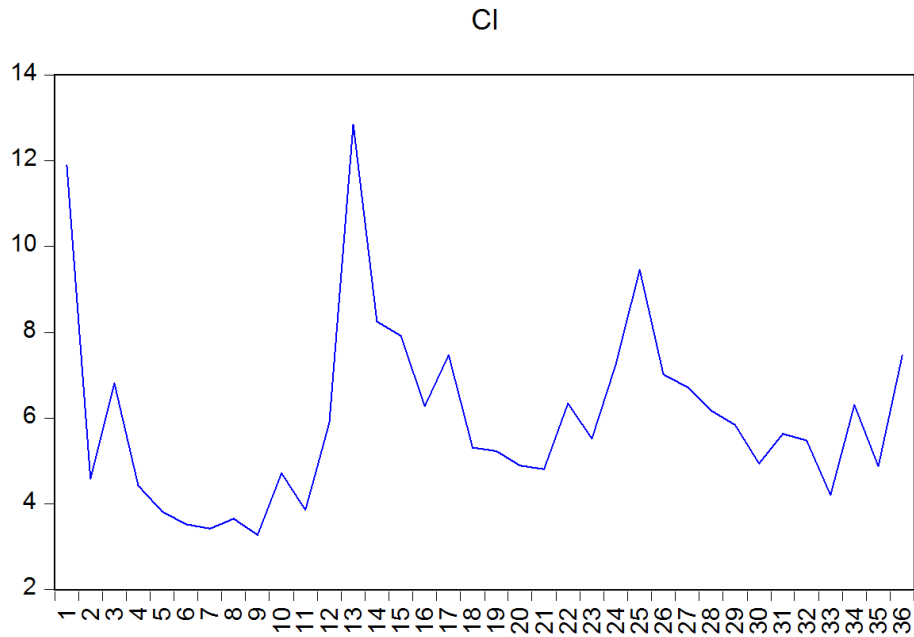


Figure 1. 2013.1-2015.12 CI series timing chart
图 1. 2013 年 1 月~2015 年 12 月 CI 时序图

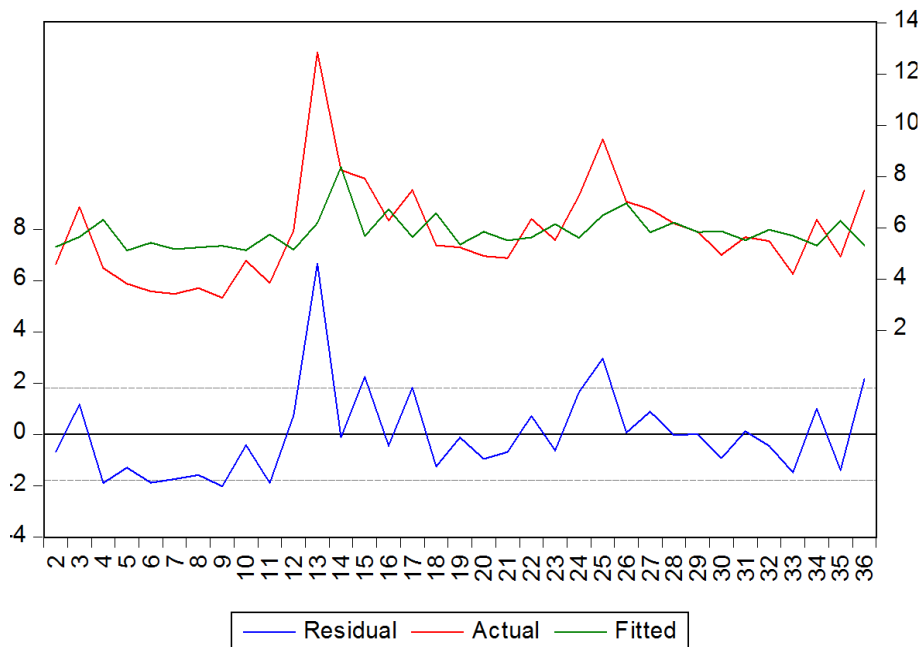


Figure 2. Effect of model fitting
图 2. 模型拟合效果图

4.2.2. 模型选定及定阶

根据 CI 序列的自相关系数和偏相关系数以及模型分类表(表 1), 可确定模型为 ARMA(p,q)模型。

在选定模型的基础上, 确定 p 和 q 的值。参考 CI 数据以及一阶差分后自相关系数和偏相关系数图, 根据 AIC 值越小、相关系数 R_2 值越大越好的原则, 比较 AIC 值和 R_2 值, 根据 AIC 信息准则判别确定模型的阶数(表 3), 可知较理想的模型是 ARMA(1,1)。

4.2.3. 模型参数估计

通过 Eviews 对模型 ARMA (1,1)进行参数估计, 结果见表 4。

由此确定 ARMA (1,1)模型如下:

$$CI_t = 5.8585 - 0.0550CI_{t-1} + \varepsilon_t + 0.4339\varepsilon_{t-1}$$

4.2.4. 模型诊断

通过 Eviews 软件做出模型拟合效果图, 可以得出拟合效果较好的结论。为检验回归方程残差的序列相关性, 可计算如图 2 所示的 Q 统计量, 可以看出, 自相关和偏相关值都接近于零, Q 统计量的 P 值大于 5%, 接受原假设, 即残差序列不相关, $\{\varepsilon_t\}$ 是一个白噪声序列。

4.3. 模型预测

Eviews 软件对 CI 预测有两种方法: 动态预测和静态预测。本文采用静态预测方法对第 37 个 CI 的数据进行预测(如图 3)。

经过静态预测得到 2016 年 1 月的预测值 $CI_{37} = 6.62$ 。与真实值 6.75 相比, 绝对误差 0.13, 相对误差 1.93%。再根据图 3, 拟合值与真实值、预测值与真实值之间的误差大部分都在 2%之内, 认为预测效果较好。通过对未来几期空气质量的预测, 发现 CI 数据变化不大, 维持在 6 左右, 这也意味着成都市的

Table 2. ADF test

表 2. CI 序列 ADF 检验

Null Hypothesis: CI has a unit root			
Exogenous: Constant			
Lag Length: 0 (Automatic—based on SIC, max lag = 9)			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.702396	0.0006
Test critical values:	1% level	-3.632900	
	5% level	-2.948404	
	10% level	-2.612874	

*MacKinnon (1996) one-sided p-values.

Table 3. The AIC value of different models

表 3. 不同模型 AIC 值

ARMA 模型	ARMA(0,1)	ARMA(1,0)	ARMA(1,1)
AIC 值	4.3287	4.0960	4.1056
拟合优度	0.1180	0.1310	0.1714

Table 4. Estimate model

表 4. 模型估计

变量	系数	T 值	P 值
C	5.8585	14.0695	0.0000
AR (1)	-0.0550	-0.2404	0.0030
MA (1)	0.4339	1.7670	0.0068

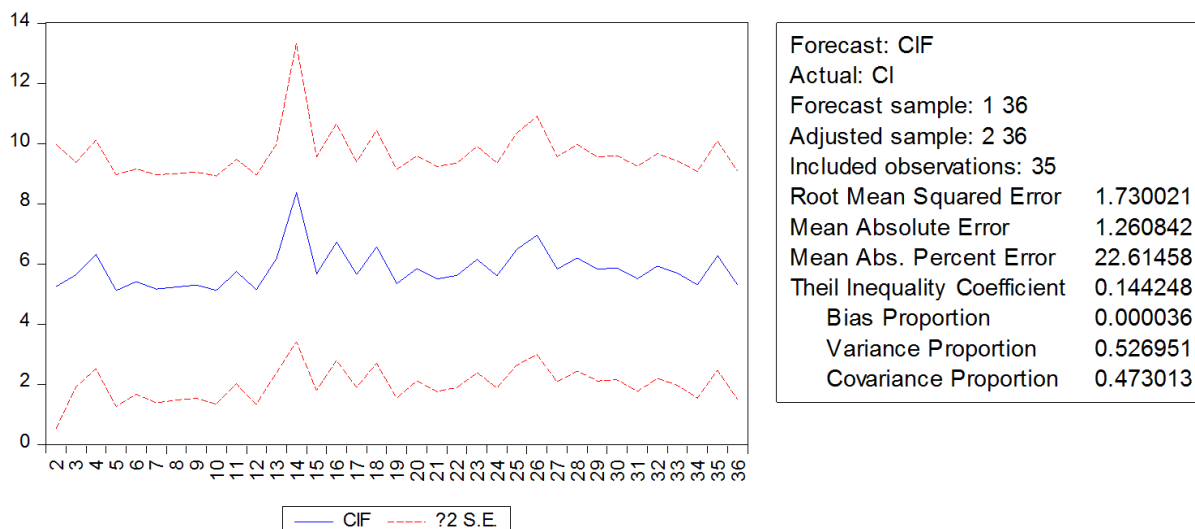


Figure 3. Static prediction

图 3. 静态预测图

空气质量在未来一段时间内不会有所改善, 仍将位于空气质量较差城市之列。因此, 从长远来看, 为改善空气质量状况, 防止出现环境恶化的可能性, 成都市有必要采取适当措施, 以应对大气污染现状, 改善生态环境。

4.4. 模型结论及比较研究

根据 ARMA 模型预测结果及其变化趋势进行分析可以看出, ARMA 模型能够很好的根据空气质量历史数据进行短期预测, 能够较好的预测出成都未来的空气质量, 精确度较高。根据预测结果, 成都市空气质量与去年同期相比有所改善, 但长远来看, 好转趋势并不明显, 基本上是维持现有较高的污染水平上下波动。为此, 仍需要把保护环境、改善空气质量作为政府工作的重心, 仍需采取相关措施加以改善。

BP 网络模型是根据神经网络的特点设计, 以解决因果关系较为复杂的判断、预测等问题为目的的技术模型。BP 网络包括输入层、输出层以及一层或多层的隐含层。运用此种方法对 CI 序列进行预测时, 需要合理有效的选择隐层数和隐层的节点数。运用 SPSS 建立 BP 神经网络预测模型, 对第 37 期的 CI 进行预测, 预测结果为 6.43, 绝对误差 0.32, 相对误差 4.7%。BP 神经网络的预测方法需要将原始数据分为训练集和测试集来建立最优的模型, 因此建立出的模型会损失原始数据的部分信息[5]。针对本文只有一个变量 CI 以及数据量的具体情况, 运用 BP 神经网络来进行预测, 并不能显示出其优势, 并且预测过程较为繁琐, 预测结果误差也较大。经验证, 此方法误差比灰色理论模型和时间序列模型误差都大。

灰色理论预测的建模方法实质是对原始数据序列作一次累加生成, 使生成序列呈一定规律, 并用典型曲线拟合, 建立数学模型。对 CI 序列的预测采取 GM(1,1)模型, 这是最常用的一种灰色模型, 它是由一个只包含单变量的一阶微分方程构成的模型。运用 Excel 建立灰色预测模型, 求出辨识参数, 并对第 37 期的 CI 进行预测, 预测结果为 6.52, 绝对误差 0.23, 相对误差 3.4% [6]。由此可见, 灰色理论模型的预测精度较差, 序列完整性、可靠性低。此方法误差较 BP 神经网络误差小, 较时间序列模型误差大。

上述比较研究的结果表明, 本文对 CI 序列数据的预测, 运用 ARMA 时间序列分析方法进行预测, 操作简单, 运算快, 预测误差在 2%之内, 预测结果较好, 在短期预测中较为实用。

5. 结论和建议

本文采用成都市 36 个月的空气质量综合指数, 利用 ARMA 模型对成都市空气质量进行预测。ARMA

模型仅利用环境质量综合指数这一变量的历史数据,即可对未来的空气质量进行短期预测,操作简单,预测结果精确度也较高,研究表明,成都市空气质量在未来一段时间内无好转趋势。与BP神经网络和灰色理论模型预测法相比预测效果较好,预测误差相对较小。

值得注意的是,ARMA虽然可以较为精确地预测成都地区空气质量的变化趋势,但无法揭示导致该地区空气质量变化的影响因素和影响机制。关于这方面的不足,还有待于进一步的深入研究和分析。

成都市位于我国四川盆地,又是西部地区经济发展的领头羊,保持良好生态环境的重要性不言而喻。针对目前空气质量差、环境污染重的现状,提出以下两点建议。

第一,做好改善环境的长期规划,把环境工作放在重心位置。2016年6月5日是第45个世界环境日,环境保护部发布了今年环境日的主题:“改善环境质量,推动绿色发展”。根据模型预测结果,短期内成都市空气质量并不会有所改善。据此,成都市应立足现在,着眼未来,采取一系列环境保护措施,短期和长期规划并重,切实把保护环境放在政府工作的重心位置。

第二,加强城市综合治理,大力开展植树种草,扩大绿化面积。成都市位于四川盆地“锅底”,这一不利地理位置的事实虽然无法改变,但控制环境污染仍有许多行之有效的具体措施。比如,增大市区内绿化面积以提高整座城市的绿化覆盖率,严格控制易产生粉尘的建设项目以最大程度减少扬尘产生的污染。这些措施,都可以有效保护市区环境,改善空气质量。

参考文献 (References)

- [1] 孙柏峰. 吉林省某市空气质量评价及预测[J]. 吉林: 吉林大学环境与资源学院, 2013.
- [2] 宋宇辰, 甄莎. BP神经网络和时间序列模型在包头市空气质量预测中的应用[J]. 干旱区资源与环境, 2013, 27(7): 65-70.
- [3] 朱悦, 郑洪波, 张树深, 张芸. 基于GM(1,1)模型的空气质量变化趋势预测及分析[J]. 辽宁工程技术大学学报, 2006(s2): 295-297.
- [4] 沃尔特·恩德斯. 应用计量经济学——时间序列分析[M]. 北京: 高等教育出版社, 2012.
- [5] 王俭, 胡筱敏, 郑龙熙, 等. 基于BP模型的大气污染预报方法的研究[J]. 环境科学研究, 2002, 15(5): 62-64.
- [6] 赵金环, 陈建强, 郭立智. Excel在动态GM(1,1)模型预测及检验中的应用[J]. 青岛理工大学学报, 2007, 28(4): 69-73.

期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: sa@hanspub.org