

二氧化硫浓度时间序列特征分析与多元预测

——基于北京市 2015 年至 2023 年空气污染数据的分析

摘要

二氧化硫是影响我国北方城市大气环境质量的首要污染物, 由于其对人体健康的影响已经引起科学界的广泛重视, 是当前大气环境研究的重要对象之一。本文基于北京市 2015-2013 年的空间污染指标数据, 通过时间序列分析, 尝试探索二氧化硫浓度的短期时间依赖性、长期趋势与季节特征以及其与其他空气污染指标的协整关系, 并建立一元与多元 ARIMA 模型对 2023 年 1-5 月二氧化硫浓度进行预测。

在短期分析中, 本文通过定义一天中每小时的二氧化硫相对浓度, 探究二氧化硫浓度是否有一定的短期时间依赖性, 发现一天内午间的二氧化硫相对浓度较高。

在长期一元分析中, 本文通过因素分解乘法模型, 分析序列总体趋势以及季节性波动, 发现在趋势上二氧化硫浓度逐年递减, 在季节效应上呈现“U”型曲线, 夏季浓度最低, 冬春季最高。同时, 建立 ARIMA 季节乘法模型用以拟合与预测未来的二氧化硫浓度。

在长期多元分析中, 本文通过分析二氧化硫浓度与二氧化氮、颗粒物等污染指标的因果关系, 建立 ARIMA 多元协整模型, 并与 ARIMA 一元季节乘法模型进行比较, 发现纳入多元变量后的模型在拟合程度与预测精度上有所提升。

关键词: 二氧化硫浓度; 空气污染指标; 因素分解模型; ARIMA 季节乘法模型; Granger 因果检验; ARIMA 多元协整模型

目录

1 问题背景与研究思路.....	3
1.1 问题背景.....	3
1.2 研究思路.....	3
2 数据说明.....	4
2.1 数据来源.....	4
2.2 预处理.....	4
3 短期波动分析.....	5
4 长期一元分析.....	5
4.1 因素分解乘法模型.....	5
4.2 ARIMA 季节乘法模型.....	7
5 长期多元分析.....	7
5.1 Granger 因果检验.....	7
5.2 协整检验与定阶.....	8
5.3 多元协整模型估计.....	8
5.4 一元模型与多元模型比较.....	9
6 总结.....	10
参考文献.....	11
附录.....	11

1 问题背景与研究思路

1.1 问题背景

二氧化硫是影响我国北方城市大气环境质量的首要污染物，由于其对人体健康的影响已经引起科学界的广泛重视，是当前大气环境研究的重要对象之一。

本文基于时间序列分析，尝试回答以下问题：

- （1）短期来看，一天内二氧化硫浓度是否有明显的时间依赖性？
- （2）长期来看，二氧化硫浓度的变化趋势如何？有无季节效应？
- （3）长期来看，二氧化硫浓度的变化与空气中其他污染物（一氧化碳、二氧化氮、颗粒物等）有无关系？利用协整模型来预测二氧化硫浓度是否会提高预测的准确性？

1.2 研究思路

对于问题（1），由于二氧化硫浓度随着时间有一定的趋势性变化，本文定义了每个小时的**相对浓度**，用以表示某时浓度对于当天平均浓度的偏离程度，对于一年 365 天每个小时的相对浓度进行描述性分析，探索是否有一定的规律性。

对于问题（2），针对存在趋势和季节效应时间序列数据，本文利用**确定性因素分解**的方法将序列分解成趋势效应、季节效应和随机波动，帮助我们探究序列总体趋势以及季节性波动。但是，确定性因素分解法对序列的拟合和预测效果仍有不足，故对于序列的预测，进一步构建 **ARIMA 季节乘法模型**。

对于问题（3），我们在模型中纳入其他空气污染浓度指标，进行协整检验并构建**多元 ARIMA 协整模型**。本文对比了问题（2）中的单变量季节乘法模型与多元协整模型的拟合均方误差，并以 2023 年 1-5 月数据为基准对比模型的预测准确性。



图 1 研究思路

2 数据说明

2.1 数据来源

本文采用数据来源为环保部环境监测总站空气质量实时发布系统¹，数据为北京市自 2015 年 1 月 1 日起，至 2022 年 12 月 31 日止，共 96 周，含有如下六个变量，本文主要的研究对象为二氧化硫浓度。

表 1 空气污染指标变量

字段	字段说明	单位
SO ₂	二氧化硫浓度	ug/m3
NO ₂	二氧化氮浓度	ug/m3
CO	一氧化碳浓度	mg/m3
O ₃	臭氧浓度	ug/m3
PM2.5	颗粒物（粒径小于等于 2.5μm）浓度	ug/m3

2.2 预处理

在短期波动分析中，我们采用 2020 年共 365 天每天 24 小时的二氧化碳浓度数据。

在长期分析中，我们将每月每天的空气污染物浓度（24 小时滑动平均）数据进行平均作为当月浓度，最终获取 2015 年 1 月至 2022 年 12 月共 96 周的 5 个污染物的月平均指标数据。

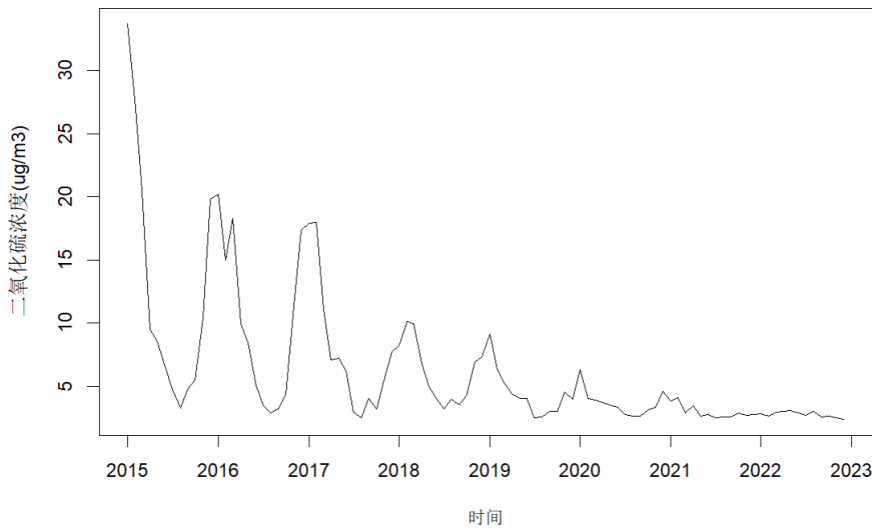


图 2 二氧化硫月平均浓度变化

¹ 来源为青悦数据，根来源为环保部环境监测总站空气质量实时发布系统（<https://air.cnemc.cn:18007/>）

3 短期波动分析

在短期的分析中，我们想要探索在一天内二氧化硫浓度是否有明显的时间依赖性。采用 2020 年共 365 天每天 24 小时的二氧化硫浓度数据，对其进行处理如下，定义第*i*天第*t*小时的**相对浓度**为其对于当天浓度均值的偏离除以当天浓度变化标准差，即

$$\widetilde{a}_{it} = \frac{a_{it} - \bar{a}_i}{sd(a_i)}$$

$$i = 1, 2, \dots, 365, \quad t = 1, 2, \dots, 24$$

其中 \bar{a}_i 代表第*i*天平均浓度， $sd(a_i)$ 代表第*i*天 24 小时浓度的标准差。

对若干天的相对浓度变化曲线进行描述性分析，如图 3 所示，我们可以看出虽然有部分天数的浓度变化无明显规律，但一些天数的浓度变化呈现“在下午 14 时左右达到峰值”的趋势。平均来看，在白天上午 11 时到下午 16 时，二氧化硫浓度处于一天的高值，而在凌晨与夜晚中空气的二氧化硫浓度较低。这可能是由于午间风速一般较缓，使得二氧化硫浓度形成暂时的堆积效应。

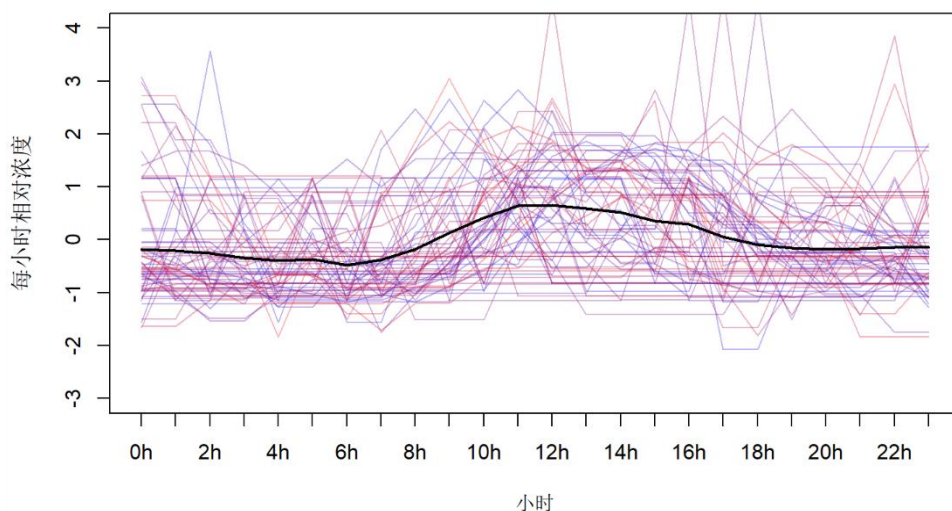


图 3 每小时相对浓度变化趋势（黑线上的点代表每小时相对浓度均值）

4 长期一元分析

4.1 因素分解乘法模型

观察时间序列可以发现，随着趋势的递减，每个季节的振幅也在减少，这说明季节效应受到趋势的影响，故尝试因素分解乘法模型。将二氧化硫浓度序列分解为趋势效应、季节效应与随机波动，结果为图 4 所示。

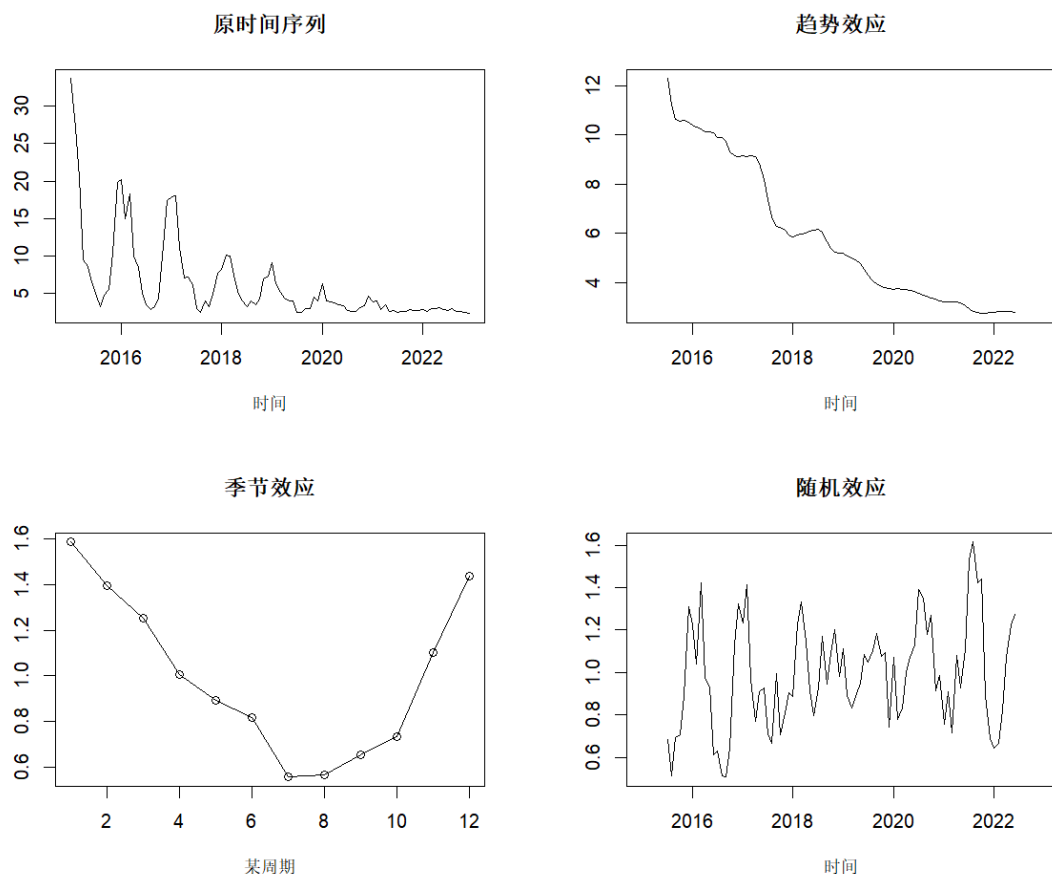


图 4 因素分解乘法模型

从趋势效应我们可以看出，二氧化硫浓度呈现明显的**逐年减小**的趋势，表明整体来看我国环境治理成效显著，空气质量在逐年改善。

从季节效应我们可以看出，月度指数变化呈**“U”形曲线**，夏季是二氧化硫浓度最低的季节，其中 7 月份浓度最低，冬春季二氧化硫浓度相对较高，1 月份是浓度最高峰。本文对 2015-2017 年的煤炭销量与二氧化硫浓度变化进行描述性分析，结果如图 5 所示。随着冬季取暖需求增加，煤炭消耗量增长，是二氧化硫浓度升高的主要原因。

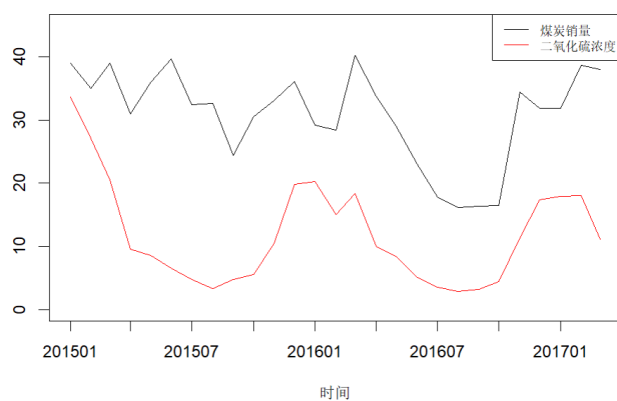


图 5 2015-2017 年煤炭销量与二氧化硫浓度变化描述性分析

4.2 ARIMA 季节乘法模型

对序列进行一阶十二步差分提取趋势和季节效应，模型定阶如下：

首先考虑一阶十二步差分之后序列，单位根检验结果表明差分后序列平稳，由于序列通过 BJ 检验，即为白噪声序列，故短期序列利用 ARMA(0,0)模型进行拟合。

其次考虑季节自相关特征，主要考察以周期长度为单位的自相关系数和偏自相关系数的特征，如图 6 可以认为季节特征是自相关系数二阶截尾，偏自相关系数拖尾，这时用以十二步为周期的 ARMA(0,2)[12]模型提取差分后序列的季节自相关信息。

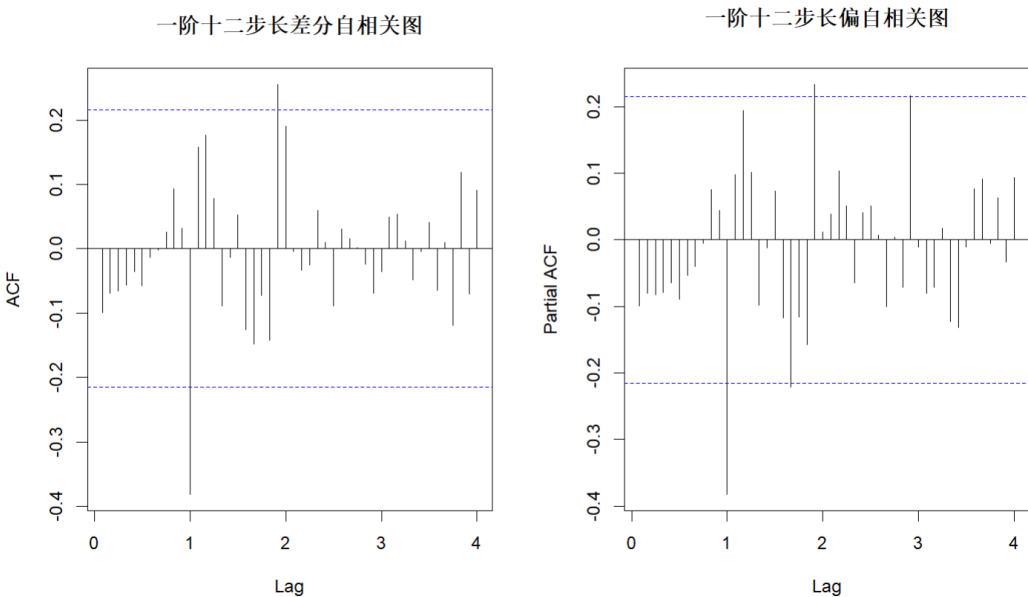


图 6 一阶十二步长差分序列相关图

最终拟合 ARIMA(0,1,0)(0,1,2)[12]模型，估计结果如表所示，模型系数显著，检验结果显示拟合效果较好（见附录），模型表达式为：

$$\nabla \nabla^{12} \text{SO}_2_t = (1 - 0.3864B^{12})(1 + 0.2824B^{24})\varepsilon_t$$

其中 ε_t 为零均值白噪声序列，其方差估计为 4.621.

5 长期多元分析

5.1 Granger 因果检验

对于二氧化硫浓度，首先我们进行因果检验，探究其余空气污染变量是否是二氧化硫浓度变化的原因，或互为因果关系。检验结果显示，二氧化氮、一氧化碳与颗粒物检验均显著，可以视

为使得二氧化硫预测精度有显著提高的 Granger 原因，检验关系如下图所示。

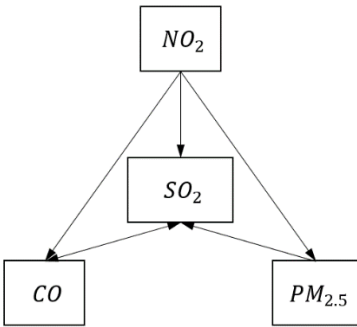


图 7 Granger 因果检验

5.2 协整检验与定阶

各空气污染指标时序如图 8 所示。通过 EG 两步法对协整假设进行检验，首先建立 SO₂ 浓度序列与 NO₂、CO、PM_{2.5} 序列之间的回归模型，之后对回归残差序列进行平稳性检验，结果显示回归残差序列为平稳非白噪声序列。根据残差序列的相关图属性，判断其自相关系数与偏自相关系数均不截尾，故尝试对其拟合 ARMA(1,1,1)(0,1,0)[12]季节加法模型，进一步提取相关信息。

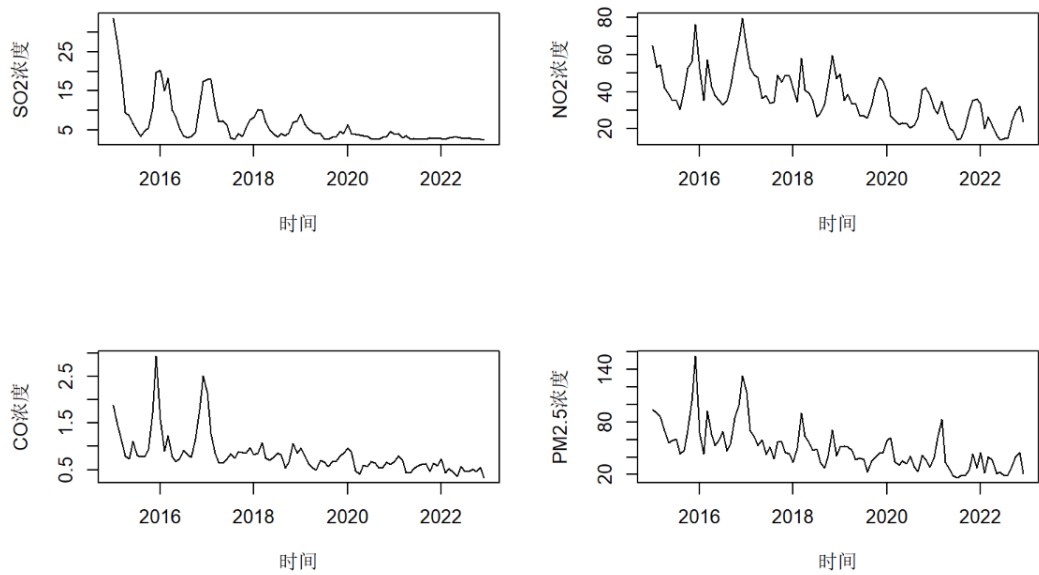


图 8 空气污染指标时间序列

5.3 多元协整模型估计

最终拟合模型结果如表 2 所示，模型系数显著，检验结果显示拟合效果较好（见附录），模型表达式为：

表 2 多元协整模型估计结果

Coefficients	AR1	MA1	NO2	CO	PM2.5
Estimate	0.6417	-0.9156	0.1519	5.2815	-0.048
S.E.	0.1298	0.0612	0.0346	1.5385	0.020

我们可以发现，二氧化氮、一氧化碳与二氧化硫有着正相关关系，这可能是由于相同的能源燃烧排放所致；粉尘颗粒物与二氧化硫有负相关关系，这与相关研究结论一致，因为粉尘对二氧化硫具有吸附性，颗粒物浓度越高，对空气中二氧化硫的吸附就越多。

$$\nabla \nabla^{12}(SO2_t - 0.1519NO2_t - 5.2815CO_t + 0.048PM2.5_t) = \frac{1 - 0.9156B}{1 - 0.6417B} \varepsilon_t$$

5.4 一元模型与多元模型比较

比较纳入二氧化氮、一氧化碳等其他空气污染指标前后建立的季节乘法模型与多元协整模型，如表 3 和图 9 所示，多元情况下的模型在拟合程度有所提升；如图 10 所示，建立多元模型对于二氧化硫的预测准确度与精度有改善作用。

表 3 一元与多元模型拟合效果比较

模型指标	AIC	BIC	MSE
ARIMA 一元季节模型	369.7	376.95	3.899057
ARIMA 多元协整模型	330.11	347.49	2.378356

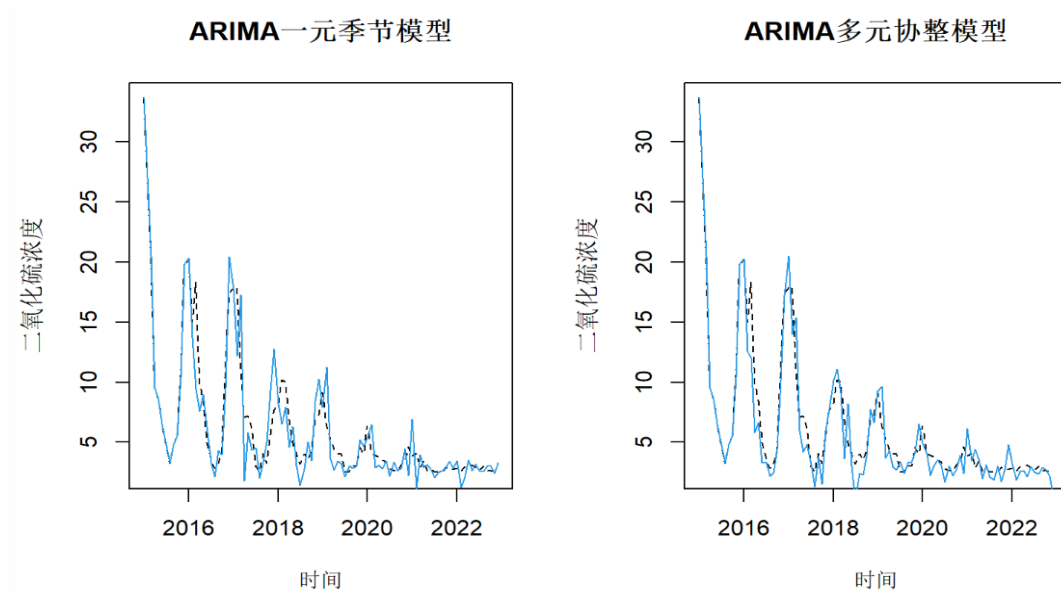


图 9 两模型拟合效果比较（其中虚线为真实时间序列，蓝线为拟合值）

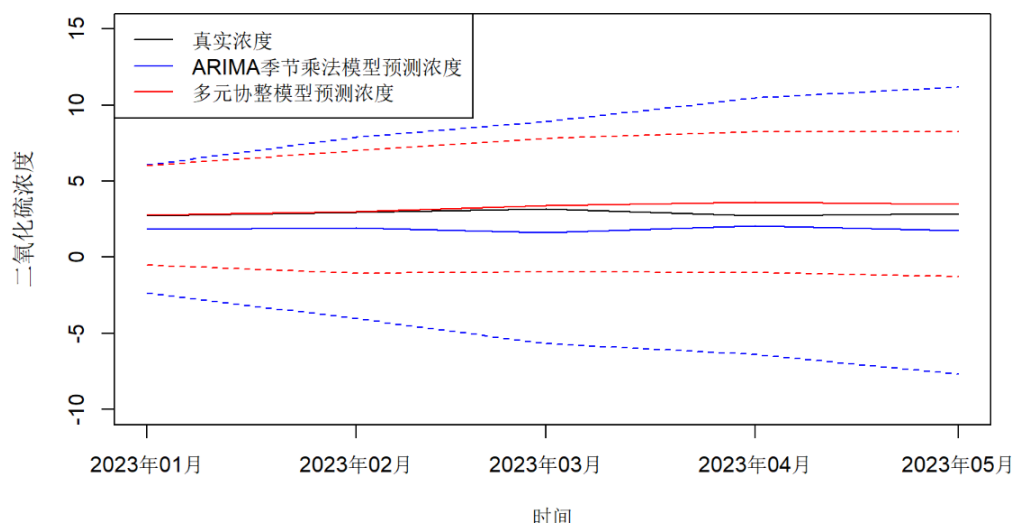


图 10 两模型预测效果比较（虚线为 95%置信区间）

6 总结

本文基于空气污染指标数据，对二氧化硫浓度的短期与长期时间序列特征进行探究，基于因素分解模型、季节乘法模型以及多元协整模型，得出以下结论：

（1）从短期波动来看，二氧化硫浓度有一定的时间依赖性，一天内 12 时-14 时的二氧化硫浓度较高，这可能与午间风速较低等气象因素相关。

（2）从长期波动来看，在趋势上二氧化硫浓度逐年递减，在季节效应上二氧化硫浓度呈现“U”型曲线，夏季浓度最低，冬春季最高，这可能与冬季燃煤量增加有关。

（3）从长期波动来看，二氧化硫与其他空气污染物存在协整关系，其中二氧化氮、一氧化碳与其为正相关关系，而颗粒物与其有负相关关系；从模型序列拟合与预测上看，纳入其他因素后的多元协整模型的在预测效果上有一定程度的改善。

同时本文也存在一些不足之处，并启发了一些可以继续研究的方向。其一是对于二氧化硫的影响因素选取不充分，本文只纳入了空气污染指标，较多文献指出二氧化硫浓度受气象因素影响，这部分的讨论可以将气温、风速、风向等因素纳入来进一步探究；其二是对于影响因素的探究方面，由于数据缺失，本文只对 2015-2017 两年的燃煤量与二氧化硫浓度关系进行了描述性分析，可以进一步探索煤炭消耗量与二氧化硫浓度变化的因果关系，将其纳入模型，提升预测精度。

参考文献

[1] 于淑秋, 林学椿, 徐祥德.北京市区大气污染的时空特征[J].应用气象学报,2002,: 92-99

[2] 刘炳江, 郝吉明, 贺克斌, 柴发合, 薛志刚, 樊元生, 刘孜, 赵维钧.中国酸雨和二氧化硫污染控制区区划及实施政策研究[J].中国环境科学,1998,(1): 1-7

[3] 蔡旭晖, 张睿, 宋宇, 谢绍东.北京地区大气 PM10 和 SO2 的背景浓度分析[J].气候与环境研究,2004,第 9 卷(3): 445-453

[4] 任建光, 张焯, 石美风, 黄继忠.云冈石窟大气中二氧化硫浓度的变化特征分析[J].中国文物科学研究,2015,(1): 85-90

[5] 马昭.近年来大连市大气中二氧化硫浓度变化特征分析[J].科学技术创新,2021,(7): 37-38

[6] 李会娟, 于文博, 刘永泉.城市二氧化氮、悬浮颗粒物、二氧化硫健康危险度评价[J].国外医学（医学地理分册）,2007,(3): 133-135,144

[7] HJ 633-2012.环境空气质量指数(AQI)技术规定(试行)[S].行业标准-环保(CN-HJ),2012.02.29

附录

表 4 ARIMA 季节乘法模型系数

Coefficients	sma1	sma2
Estimate	-0.3864	0.2824
S.E.	0.1143	0.1243

表 5 ARIMA 多元协整模型系数

Coefficients	AR1	MA1	NO2	CO	PM2.5
Estimate	0.6417	-0.9156	0.1519	5.2815	-0.048
S.E.	0.1298	0.0612	0.0346	1.5385	0.020

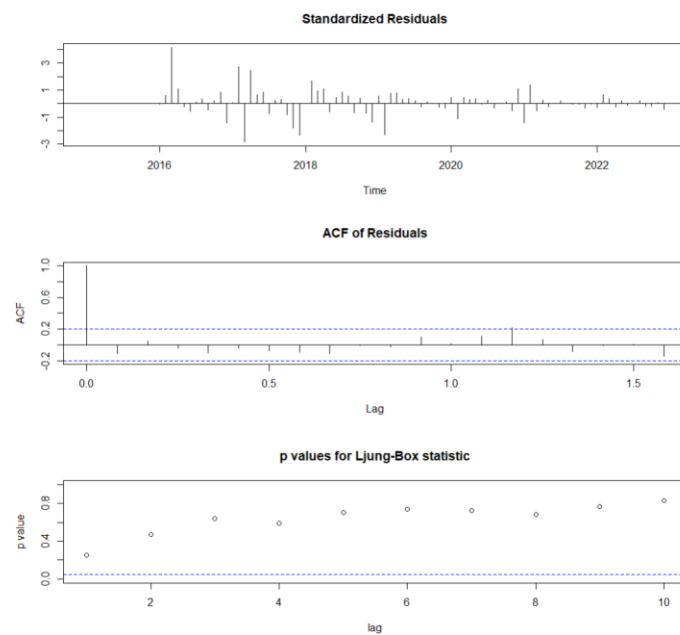


图 11 ARIMA 季节乘法模型残差检验

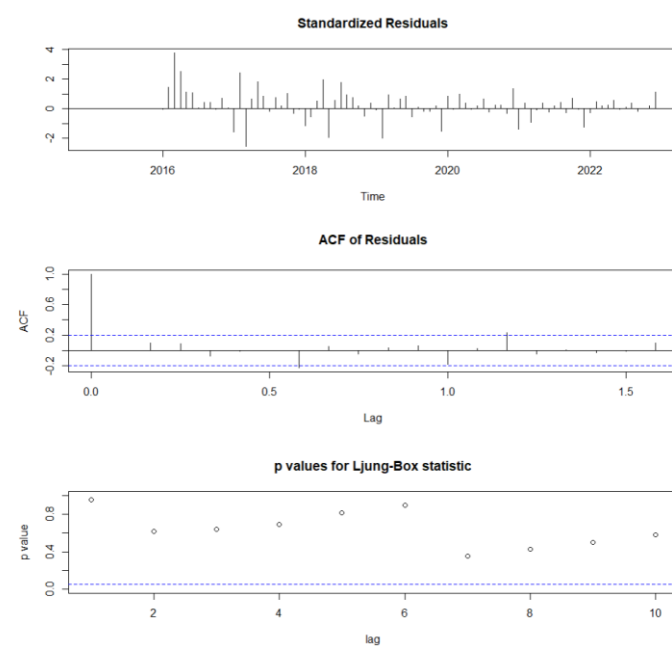


图 12 ARIMA 多元协整模型残差检验