



硕士学位论文

多因变量时空地理加权回归模型研究及其应用  
Research and Application of Multivariate  
Geographically and Temporally Weighted  
Regression Model

作 者：杨旭初  
导 师：杨毅副教授  
校外导师：徐胜华老师

江苏海洋大学  
2024年5月

## 学位论文使用授权声明

本人完全了解江苏海洋大学有关保留、使用学位论文的规定，同意本人所撰写的学位论文的使用授权按照学校的管理规定处理：

作为申请学位的条件之一，学位论文著作权拥有者须授权所在学校拥有学位论文的部分使用权，即：①学校档案馆和图书馆有权保留学位论文的纸质版和电子版，可以使用影印、缩印或扫描等复制手段保存和汇编学位论文；②为教学和科研目的，学校档案馆和图书馆可以将公开的学位论文作为资料在档案馆、图书馆等场所或在校园网上供校内师生阅读、浏览。另外，根据有关法规，同意中国国家图书馆保存研究生学位论文。

(保密的学位论文在解密后适用本授权书)。

作者签字： 杨旭初

导师签字： 柳毅

2024年 05月 28 日

2024年 05月 28 日

中图分类号 P208

学校代码 11641

UDC 519.842

密 级 公开

# 江苏海洋大学

## 硕士学位论文

多因变量时空地理加权回归模型研究及其应用

Research and Application of Multivariate

Geographically and Temporally Weighted

Regression Model

作 者 杨旭初

导 师 杨毅 徐胜华

申请学位 工程硕士

培养单位 海洋技术与测绘学院

学科专业 土木水利

研究方向 时空数据挖掘分析

答辩委员会主席 魏冠军

评 阅 人 学院盲审

二〇二四年五月

## 论文审阅认定书

研究生 杨旭初 在规定的年限内，按照研究生培养方案的要求，完成了研究生课程的学习，成绩合格；在我的指导下完成本学位论文，经审阅，论文中的观点、数据、表述和结构为我所认同，论文撰写格式符合学校的相关规定，同意将本论文作为学位申请论文送专家评审。

导师签字： 杨毅

2024 年 05 月 28 日

## 致 谢

几载寒暑荏苒，感念光阴如梭，忆昨日初入校园仿若昨日。韶华之逝，光阴蹉跎，惟恨读书岁月少。回首在江苏海洋大学 2021 年至 2024 年匆忙而充实的时光，是满载的收获；回顾特殊时期隔着屏幕在线学习的岁月，是守望和互助；回望往来于海洋技术与测绘学院这步履匆匆的最后一年，是辛酸与不舍……

首先，感谢我一生的恩师，我的导师，杨毅老师。2021 年 4 月 25 日，一个电话与杨老师建立不断的联系。初次见面，老师站于学校测绘楼门口，翘首相望，笑面相迎，记忆犹新。此后，老师始终以身作则，从老师身上迸发出来严谨的求知精神、一丝不苟的工作态度、开放多元的国际视野、热情澎湃的运动气息、如履薄冰的育人的心态……春夜喜雨般润物细无声的引导我们师门砥砺前行。感谢江苏海洋大学的哈麦（Hamed Karimian）老师和张廷领老师两位外教带我领略异国风情。感谢武汉大学李仲玢老师对我的升学指导。感谢中国测绘科学研究院徐胜华老师、罗安老师、赵阳阳老师、亢晓琛老师、仇阿根老师对我学业的帮助与支持。再次感谢杨毅老师让我结识这些的可爱的老师们！愿老师们身体健康，桃李芬芳！

此外，短短几年的时光，我也收获了小伙伴们们的友情与陪伴。韩加宽师兄、柴正媛师姐、贺鹏师兄、杨肖月师姐、袁阿兰师姐、王康帅师弟、金凌波师妹、刘旺旺师弟、王皓辰学弟、徐延昌师弟、张宇杰师弟、王飞学弟、范晋晖学弟，以及同届管一卫、杨刚、晏雪峰、胡英屹、张顺顺、汪鑫建、王子毅，本科的好伙伴陶洋，来自尼日利亚国际留学生 Terhide、Pascal。感恩和你们的相遇相知；也感谢同学们在学习与生活上的帮助和陪伴。

最后，感谢我的家人给予我的支持，你们，我的铠甲，是我做出任何决定的勇气和底气；也是我的港湾，有你们在的地方才是我的家，你们的爱与关怀造就了现在的这个我。希望有一天，我能拥有力量成为你们的铠甲，你们的港湾。

几年时光匆匆而过，这漫长又短暂的时光里所有的经历和收获都将化为我人生的财富和宝藏，令我的生命更加充实而丰富！



# 摘要

时空地理加权回归（Geographically and Temporally Weighted Regression, GTWR）模型是一种被广泛应用于多个学科领域的时空分析工具，用于探索地理数据的时空非平稳性。GTWR 模型通过为每个观测点分配不同的权重，能够捕捉数据的空间和时间非平稳性。然而，时空地理加权回归模型在处理多个相互关联的结果变量（因变量）时存在一定的局限性，其通常只针对单一结果变量及其对应的解释变量（自变量）进行分析，无法提供多结果的数据洞察。本文引入多因变量多元线性回归（Multivariate Multiple Liner Regression , MMLR）模型和多因变量地理加权回归（Multivariate Geographically Weighted Regression, MGWR）模型的思想，并融入时间维，提出了多因变量时空地理加权回归（Multivariate Geographically and Temporally Weighted Regression, MGTWR）模型。本文收集了中国城市空气污染物浓度实测数据，实证分析了该模型能够同时处理多个相互关联的结果变量，并剖析出复合空气污染物的时空特征。本文主要研究内容及创新点如下：

## （1）提出多因变量时空地理加权回归算法

在探究具有时空变化系数的多结果局部模式时，传统多因变量时空地理加权回归未考虑时空非平稳的影响，多因变量地理加权回归模型顾及了空间效应但忽略时间效应，显得相对有限。为弥补这一方法上的不足，本研究提出了多因变量时空地理加权回归框架，该框架将多因变量多元回归整合到局部时空地理加权回归建模中。多变量的时空地理加权回归模型不仅提供了一个灵活的工具来适应多个响应变量，还能够有效探索时空非平稳性问题，对推动多元空间多因变量建模领域的发展具有现实意义。

## （2）实现多因变量时空地理加权回归模型在中国城市群空气污染的应用

为了验证多因变量时空地理加权回归模型的可行性和实用性，本文采用多因变量时空地理加权回归方法对复合空气污染物进行实证分析，并通过地理可视化技术绘制了多因变量模型产生的丰富结果。具体而言，本研究使用多因变量时空地理加权回归方法分析了气象条件、环境因素以及人类活动对两种高相关性空气污染物（PM<sub>2.5</sub> 和 PM<sub>10</sub>）浓度的影响。通过将多因变量时空地理加权回归方法与传统的非空间多因变量多元线性回归方法和基于空间距离加权的多因变量地理加权回归方法进行比较，研究发现融合时空因素的多因变量时空地理加权回归模型的拟合优度方面表现出色。多因变量时空地理加权回归模型将多种污染物作为一个整体而非独立且不相关的结果变量进行考虑，实验结果表明，城市工业粉尘排放量仅对 PM<sub>2.5</sub> 或 PM<sub>10</sub> 中的单一结果变量进行考虑，忽略了对这

两种污染物的共同影响。这些关系不能用单因变量技术恰当地反映出来，具有区分给定协变量对多因变量的影响的能力是多因变量时空地理加权回归模型的主要优势，并突出了本研究的价值所在。

该论文有图 20 幅，表 5 个，参考文献 77 篇。

**关键词：** 多因变量时空地理加权回归；MGTWR；复合空气污染物；时空非平稳

# Abstract

Geographically and Temporally Weighted Regression (GTWR) model is a popular tool widely used in various disciplines to explore the spatiotemporal non-stationarity of geographic data. GTWR model captures the heterogeneity of data in space and time by assigning different weights to each observation point. However, it has limitations when dealing with multiple correlated response variables, as it typically analyzes only a single response variable and its corresponding explanatory variables, providing limited insights and modeling each response variable separately. This thesis introduces the ideas of Multivariate Multiple Liner Regression (MMLR) model and Multivariate Geographically Weighted Regression (MGWR) model, and incorporates the temporal dimension to propose the Multivariate Geographically and Temporally Weighted Regression (MGTWR) model. The study collects air pollution concentration data in Chinese cities and empirically analyzes how the model can handle multiple correlated response variables and uncover the spatio-temporal characteristics of composite air pollutants. The main contributions and innovations of this thesis are as follows:

1. Proposing the algorithm of Multivariate Geographically and Temporally Weighted Regression.

Traditional multivariate spatial modeling methods are relatively limited when exploring local patterns with spatiotemporal varying coefficients, especially when dealing with geographic data involving two or more correlated response variables. To address this limitation, this study proposes the framework of Multivariate Geographically and Temporally Weighted Regression, which ingeniously integrates multivariate multiple liner regression into local spatiotemporal geographically weighted regression modeling. MGTWR not only provides a flexible tool to accommodate multiple response variables but also effectively explores spatio-temporal non-stationarity, which is of practical significance for advancing the field of multivariate spatial modeling with multiple response variables.

2. Implement the application of Multivariate Geographically and Temporally Weighted Regression model to air pollution in Chinese urban agglomerations.

To validate the feasibility and practicality of the Multivariate Geographically and Temporally Weighted Regression model, this thesis applies the MGTWR method to empirically analyze composite air pollutants and visually presents the rich results generated by the multivariate model using geographic visualization techniques. Specifically, this study uses the MGTWR method to analyze the effects of meteorological conditions, environmental factors, and human activities on the concentrations of two highly correlated air pollutants ( $PM_{2.5}$  and  $PM_{10}$ ). By comparing the MGTWR method with traditional non-spatial MMLR method and spatially

weighted MGWR method, the study finds that the MGTWR model incorporating spatio-temporal factors performs well in terms of model fit. The MGTWR model considers multiple pollutants as a whole rather than independent and unrelated outcomes. For instance, solely considering urban industrial dust emissions in relation to a single outcome variable of either PM2.5 or PM10 overlooks the shared influences on both pollutants. These relationships cannot be adequately reflected using univariate techniques. The ability of the MGTWR model to distinguish the impact of given covariates on multiple dependent variables is its primary strength, emphasizing the significance of this study.

The thesis includes 20 figures, 5 tables, and 77 references.

**Keywords:** Multivariate Geographically and Temporally Weighted Regression; MGTWR; Composite air pollutants; Spatiotemporal non-stationarity

# 目 录

摘要 .....	I
目录 .....	III
图清单 .....	VII
表清单 .....	IX
变量注释表 .....	X
1 绪论 .....	1
1.1 研究背景与意义 .....	1
1.2 国内外研究现状 .....	2
1.3 主要研究内容及方法 .....	4
1.4 论文组织与章节安排 .....	8
2 基础理论与研究方法 .....	10
2.1 多因变量线性回归模型理论 .....	10
2.2 时空地理加权回归模型理论 .....	13
2.3 多因变量地理加权回归算法模型 .....	16
2.4 模型的评价指标 .....	18
2.4 本章小结 .....	19
3 多因变量时空地理加权回归算法实现 .....	21
3.1 多因变量时空地理加权回归算法模型 .....	21
3.2 模型回归估计方法 .....	22
3.3 多因变量时空带宽的选择 .....	23
3.4 多因变量时空非平稳性检验 .....	24
3.5 本章小结 .....	27
4 多因变量时空地理加权回归模型实证分析 .....	28
4.1 空气复合污染物相关性分析 .....	28
4.2 空气复合污染物驱动因素研究应用分析 .....	39
4.3 本章小结 .....	49
5 总结与展望 .....	50
5.1 研究总结 .....	50

5.2 研究展望 .....	50
参考文献.....	52
作者简历.....	57
学位论文原创性声明.....	58
学位论文数据集.....	59

# Contents

<b>Abstract .....</b>	<b>I</b>
<b>Contents.....</b>	<b>III</b>
<b>List of Figures .....</b>	<b>VII</b>
<b>List of Tables.....</b>	<b>IX</b>
<b>Table of Variables.....</b>	<b>X</b>
<b>1 Introduction .....</b>	<b>1</b>
1.1 Research Background and Significance .....	1
1.2 Research Status at Home and Abroad .....	2
1.3 Main Research Content and Methods.....	4
1.4 Thesis Organization and Chapter Arrangement .....	8
<b>2 Basic Theory and Research Methods .....</b>	<b>10</b>
2.1 Theory Of Multivariate Multiple Linear Regression Model.....	10
2.2 Theory of Geographically and Temporally Weighted Regression .....	13
2.3 Multivariate Geographically Weighted Regression Model .....	16
2.4 Model Evaluation Metrics .....	18
2.4 Chapter Summary .....	19
<b>3 Multivariate Geographically and Temporally Weighted Regression Model....</b>	<b>21</b>
3.1 Multivariate Geographically and Temporally Weighted Regression Model.....	21
3.2 Model Regression Estimation Methods.....	22
3.3 Bandwidth Selection for MGTWR .....	23
3.4 Test of Spatiotemporal Non-stationarity .....	24
3.5 Chapter Summary .....	27
<b>4 Empirical Analysis of MGTW Model.....</b>	<b>28</b>
4.1 Correlation Analysis of Air Composite Pollutants .....	28
4.2 Study and Application Analysis .....	39
4.3 Chapter Summary .....	49
<b>5 Summary and Outlook.....</b>	<b>50</b>
5.1 Research Summary.....	50
5.2 Research Outlook .....	50

<b>References .....</b>	<b>52</b>
<b>CV .....</b>	<b>57</b>
<b>Claim of Originality for the Degree Thesis .....</b>	<b>58</b>
<b>Dataset for the Degree Thesis .....</b>	<b>59</b>

# 图清单

图序号	图名称	页码
图 1-1	总技术路线图	6
Figure 1-1	General technical pathway	6
图 4-1	研究区域-中国	29
Figure 4-1	Study area - China	29
图 4-2	空气污染物的年均趋势变化	30
Figure 4-2	Annual trend changes in air pollutants	30
图 4-3	空气污染物的月均趋势变化	30
Figure 4-3	Monthly trend changes in air pollutants	30
图 4-4	2015 - 2021 年中国 PM <sub>2.5</sub> 浓度年/季节平均变化	32
Figure 4-4	Annual and seasonal average changes of PM <sub>2.5</sub> concentration in China from 2015 to 2021	32
图 4-5	2015 - 2021 年中国 PM <sub>10</sub> 浓度年/季节平均变化	33
Figure 4-5	Annual and seasonal average changes of PM <sub>10</sub> concentration in China from 2015 to 2021	33
图 4-6	2015 - 2021 年中国 SO <sub>2</sub> 浓度年/季节平均变化	34
Figure 4-6	Annual and seasonal average changes of SO <sub>2</sub> concentration in China from 2015 to 2021	34
图 4-7	2015 - 2021 年中国 NO <sub>2</sub> 浓度年/季节平均变化	34
Figure 4-7	Annual and seasonal average changes of NO <sub>2</sub> concentration in China from 2015 to 2021	34
图 4-8	2015 - 2021 年中国 CO 浓度年/季节平均变化	35
Figure 4-8	Annual and seasonal average changes of CO concentration in China from 2015 to 2021	35
图 4-9	2015 - 2021 年中国 O <sub>3</sub> 浓度年/季节平均变化	35
Figure 4-9	Annual and Seasonal Average Changes of O <sub>3</sub> Concentration in China from 2015 to 2021	35
图 4-10	城市大气污染物的相关性	37
Figure 4-10	Correlation of urban air pollutants	37
图 4-112	时空系数范围图	38
Figure 4-112	Range of spatiotemporal coefficients	38
图 4-12	1 月(冬季)和 7 月(夏季)时空系数空间映射	38
Figure 4-12	Spatial mapping of spatiotemporal coefficients for January (Winter) and July (Summer)	38
图 4-13	研究区域-华北平原	39
Figure 4-13	Study area - North China Plain	39
图 4-14	2015-2021 年国内生产总值 (GDP) 回归系数图	43
Figure 4-14	Regression coefficients of Gross Domestic Product (GDP) from 2015 to 2021	43

图 4-15	2015-2021 年粉尘排放量回归系数图	44
Figure 4-15	Regression coefficients of dust emissions from 2015 to 2021	44
图 4-16	2015-2021 年人均绿化面积回归系数图	45
Figure 4-16	Regression coefficients of per capita green area from 2015 to 2021	45
图 4-17	2015-2021 年人口密度回归系数图	46
Figure 4-17	Regression coefficients of population density from 2015 to 2021	46
图 4-18	2015-2021 年降水回归系数图	47
Figure 4-18	Regression coefficients of precipitation from 2015 to 2021	47
图 4-19	2015-2021 年气温回归系数图	48
Figure 4-19	Regression coefficients of temperature from 2015 to 2021	48

## 表清单

表序号	表名称	页码
表 4-1	2015 年至 2021 年年度/季节性全局 Moran's I	32
Figure 4-1	Annual/Seasonal Global Moran's I from 2015 to 2021	32
表 4-2	变量范围、方差诊断结果	41
Figure 4-2	Variable range, variance diagnostic results	41
表 4-3	MMLR、MGWR、MGTWR 模型的性能表现	42
Figure 4-3	Performance of MMLR, MGWR, and MGTWR Models	42
表 4-4	MGTWR 系数范围	42
Figure 4-4	MGTWR coefficient range	42
表 4-5	时空系数非平稳性诊断结果	42
Figure 4-5	Spatiotemporal coefficient non-stationarity diagnostic results	42

## 变量注释表

$X$	自变量矩阵
$X^T$	自变量矩阵的转置
$Y$	多因变量矩阵
$\hat{y}$	因变量的拟合值
$S$	帽子矩阵
$W_i$	权重矩阵
$\beta_k$	回归系数
$Cov$	协方差矩阵
$Q$	最小目标函数
$X_{tp}$	回归点 $t$ 的自变量矩阵
$Y_{tm}$	回归点 $t$ 的因变量 $m$ 矩阵
$\beta_{pm}$	回归点 $t$ 的多因变量回归系数
$\hat{\sigma}^2$	无偏估计项
$e$	残差向量
CV	交叉验证最小值
AICc	赤池信息准则法校正值
$d_{ij}^s$	空间距离
$d_{ij}^{st}$	时空距离
$h$	模型带宽
RSS	残差平方和
$R^2$	拟合优度

# 1 绪论

## 1 Introduction

### 1.1 研究背景与意义(Research Background and Significance)

随着科技的发展，人们正逐渐步入一个时空数据所驱动的时代。从 GPS 定位、社交媒体互动，到智能交通和万物互联。时空数据成为人类认识和改变世界的重要工具，空间分析算法已经难以满足当代的实时需求。将时间与空间分析的观点融入许多领域的研究中，这一过程取得了快速发展，催生了各种解决地理问题的时空分析工具，如时空数据可视化、时空数据挖掘等。特别是，时空建模已成为一种重要的方法，使研究人员能够评估感兴趣的变量之间的关联性，并了解不同地理尺度上的结构特征<sup>[1,2]</sup>。

在时空建模中，当将线性回归拟合时，需要考虑两个效应，一是观察值之间的空间自相关性<sup>[2,3]</sup>，另一个是关于因变量和自变量之间关系的空间非平稳性<sup>[4-6]</sup>。这两个特征都与地理学第一定律有关<sup>[7]</sup>。空间自相关性关注相邻空间数据和残差的相似性，而空间非平稳性则关注空间建模估计在空间上的可比性水平。文献中提出了许多方法来处理空间自相关性的影响，包括同时自回归或条件自回归模型<sup>[8,9]</sup>、空间滞后和误差模型<sup>[3]</sup>和空间滤波技术<sup>[10]</sup>。空间非平稳性指的是回归系数在空间上对模型的不稳定性，而不是对数据的不稳定性<sup>[5,11-13]</sup>。对于建模此类非平稳性关系的常见方法是贝叶斯分层空间变系数建模<sup>[14]</sup>和地理加权回归<sup>[4,15]</sup>。随着地理数据从空间数据结构向时空数据结构的转变，地理加权回归算法已经逐步发展成熟的时空地理加权回归算法<sup>[16,17]</sup>。虽然这些方法提高了空间自相关性和时空非平稳性的能力，能够更深入地理解和分析时空数据，但其是针对单个结果变量进行分析的。单个因变量的回归分析更多的是研究多个解释变量对单个被解释变量的影响和相互作用，然而在经济学、医学、环境科学、市场营销和教育学等领域<sup>[46]</sup>，经常需要同时考虑多个解释变量对多个被解释变量的依赖关系。由于多个被解释变量之间存在某种相关关系，将其化为多个多元回归问题来解决会丢失一部分相互联系的信息。多对多回归分析的研究存在一定的复杂性，效应的比较更是涉及多个层面，例如同一解释变量在不同被解释变量上的影响效应差异等。因此，当需要处理多个相关的结果变量时，单变量建模可能是不足够的。

在时空数据研究中，常常涉及到多个结果变量。这些变量通常高度相关，因为其评估的是同一时空位置的相关概念和条件。研究人员有时有兴趣在控制其他空间效应的同

时评估这些多个空间结果之间的相互作用。例如大气复合污染，为多种污染物同时存在，浓度较高并发生复杂的相互作用。这种污染现象表现为大气氧化性增强、能见度明显下降以及污染蔓延等<sup>[18]</sup>。在污染本质上，其表现为各种污染物相互交错并产生耦合作用<sup>[19]</sup>。气态及固态前体物，如二氧化硫、氮氧化物、有机碳等通过均相和非均相反应都可转化为颗粒物，使得颗粒物的成分变得更加复杂。各种大气污染物的气态前体物和转化后的颗粒物共存于大气中，形成了复杂的大气污染物化学体系<sup>[20,21]</sup>。

为满足现实世界中复杂的多结果需求，多因变量多元回归是一种重要的工具，通过构建一个“统一”的框架来估计协变量对多个相关结果变量的影响<sup>[22]</sup>。科学家已经在不同的应用中使用了这个工具或其扩展或变体。与本研究相关的是，研究者已经做出了许多努力来探索面板数据的多元空间模型，通常采用贝叶斯框架作为自然的推断方法。例如，Babcock 等人建议使用带有空间随机效应的多元贝叶斯空间回归模型来适应地理位置内和之间结果变量的依赖关系<sup>[23]</sup>。刘等人采用类似的贝叶斯空间建模方法，用于联合建模多元小区域计数数据<sup>[24]</sup>。虽然这些研究对多元空间分析方法做出了贡献，但对于时间和空间非平稳性（或时空异质性）的局部时空模型却缺乏深入的探讨。陈等人提出了在分析多个相互关联的响应变量时，地理加权回归（Geographically Weighted Regression, GWR）不能提供足够的数据信息，因为其只允许对每个响应变量单独建模<sup>[25]</sup>。通过引入多因变量地理加权回归(Multivariate Geographically Weighted Regression, MGWR)技术来解决这一差距，该技术不仅探索了空间非平稳性，而且还融合了多个响应变量之间的相关性，这种模型探索了预测变量和多元结果变量之间回归参数（变化系数）的不稳定性。然而，MGWR 所揭示的数据规律和特征只涉及时空数据的空间部分，忽视了数据纵向的时间变化规律。

## 1.2 国内外研究现状(Research Status at Home and Abroad)

### 1.2.1 时空地理加权回归研究现状(Current Status of Geographically and Temporally Weighted Regression)

时空地理加权回归（Geographically and Temporally Weighted Regression, GTWR）是一种顾及时间维和空间维加权的回归分析方法。在时空数据挖掘领域，空间分析方法从最初的普通线性回归模型逐渐演变为地理加权回归（Geographically Weighted Regression, GWR）模型，并最终发展成为时空地理加权回归模型，旨在揭示空间和时间上的非平稳性效应，并提供更准确的预测和解释结果<sup>[26,27]</sup>。

时空地理加权回归模型融合了地理加权回归和时空分析技术。其基于地理加权回归模型框架，通过引入时间维度，将时间因素纳入到地理加权回归模型中。时空地理加权回归通过在空间和时间上设置权重，对不同地理位置和时间段的样本进行加权，从而更准确地反映地理现象的变化。时空地理加权回归模型的数学表达式通常采用线性回归模型，但权重随空间和时间变化而变化，以反映非平稳性效应。时空地理加权回归模型的推断和参数估计方法是研究的重要内容。研究者通过不同的统计方法和算法，推导时空地理加权回归模型的参数估计方法。一种方法为最小二乘法，用于数据拟合，通过最小化误差平方和估算参数，常用于线性回归分析。值得注意的是其假设误差独立且服从正态分布。相比之下，另外一种贝叶斯方法结合先验和观测数据，提供更精确和不确定性的参数估计，但计算复杂且理解有难度<sup>[28]</sup>。

时空地理加权回归模型在多个领域得到了广泛验证，证明了其实际的应用能力。在城市规划领域，研究者利用时空地理加权回归方法研究城市发展的空间和时间变化规律，揭示城市各区域的发展趋势和影响因素，为城市规划和决策提供科学依据<sup>[29-31]</sup>。在环境科学领域，时空地理加权回归方法可以用于研究环境污染的空间分布和时间变化，解释环境污染的影响因素，为环境管理和保护提供支持<sup>[32-34]</sup>。研究者致力于改进和扩展时空地理加权回归方法，以提高模型的预测能力和解释能力。一方面，研究者尝试引入其他因素和变量，如社会经济因素、气候因素等，来丰富模型的解释变量，提高模型的拟合度。另一方面，研究者探索时空地理加权回归方法与其他时空分析方法的结合，如时空插值方法<sup>[35]</sup>、时空聚类方法<sup>[36]</sup>等，以增强对时空数据的理解和分析能力。

### 1.2.2 多因变量回归模型研究现状(Current Status of Multivariate Regression Model)

经典的多因变量回归模型，主要包括多因变量线性回归模型与多因变量地理加权回归模型，是回归分析领域中重要的研究工具。多因变量线性回归模型作为一种回归分析方法，用于研究多个因变量与单个或者多个自变量之间的关系<sup>[37-40]</sup>。其是在传统的多元线性回归模型的基础上扩展而来<sup>[41]</sup>，能够同时考虑多个因变量的影响因素，提供更全面和综合的分析结果。

多因变量线性回归模型的方法理论基于多元统计学和线性回归分析。其使用多元向量表示多个因变量，并构建线性回归模型来描述因变量与自变量之间的关系。由 Vivian 等人提出的多因变量地理加权回归模型是对多因变量线性模型的重要拓展，旨在解决传

统地理加权回归在分析多个相关响应变量时存在的局限性<sup>[25]</sup>。传统地理加权回归模型主要聚焦于单独建模每个响应变量，然而在面对多个相关响应变量时，其无法提供充分的数据信息和深入的空间关系洞察。多因变地理加权回归模型顾及了空间非平稳因素，不仅能探索空间非平稳性，还能解释多变量响应之间的相关性<sup>[25]</sup>，有助于更好地理解和解释地理数据的复杂性和多样性。

多因变量线性回归模型的推断和参数估计是研究的重要内容。研究者通过不同的统计方法和算法，推导多因变量线性回归模型的参数估计方法。其中，常用的方法包括普通最小二乘法（Ordinary Least Square, OLS）、广义最小二乘法（Generalized Least Squares, GLS）和迭代加权最小二乘法（Iterative Weighted Least Square, IWLS）<sup>[42,43]</sup>。这些方法可以用于估计模型的回归系数、方差协方差矩阵和模型的解释变量。此外，还可以进行假设检验来评估模型的显著性和可靠性<sup>[22]</sup>。

对多因变量线性回归模型的评价和选择是研究的重要环节。研究者通常使用多元统计学中的指标来评估模型的拟合度和预测能力，如决定系数、调整型绝对系数、残差平方和、标准误差等。此外，模型诊断包括检验模型的线性、正态性<sup>[44]</sup>、异方差性和多重共线性<sup>[45]</sup>等假设。通过综合评估这些指标和方法，研究者可以选择出最合适的多因变量线性回归模型。

多因变量线性回归模型的研究不断扩展和应用于不同领域。例如，在金融领域，研究者使用多因变量线性回归模型研究股票收益率之间的关系，考虑多个因素如市场指数、利率和公司基本面对股票收益的综合影响<sup>[46]</sup>。研究者还通过引入非线性、非参数和机器学习等方法，进一步拓展多因变量线性回归模型的能力和应用范围<sup>[47-49]</sup>。例如，可以将多项式回归<sup>[50]</sup>、岭回归<sup>[51,52]</sup>和支持向量回归等方法与多因变量线性回归模型相结合，以捕捉非线性关系和提高模型的预测性能<sup>[53]</sup>。

总体而言，多因变量回归模型是一种重要的统计分析方法，在多个学科领域具有广泛的应用。其能够同时考虑多个因变量的影响因素，提供更全面和综合的分析结果。研究者通过不断改进和扩展方法，提高模型的预测能力和解释能力，并将其应用于不同领域的实际问题中。随着数据科学迅速发展，多因变量线性回归模型的研究将继续深入，并与其他方法相结合，为解决复杂实际问题提供更有效的分析工具。

## 1.3 主要研究内容及方法(Main Research Content and Methods)

### 1.3.1 研究内容(Research Content)

在地理大数据时代，有效挖掘海量时空数据中有价值的信息是地理数据分析中的重要问题。探索有效的大数据量地理数据分析和建模方法，对于解释社会情景和分析自然变化过程具有重要意义。

本文通过研究比较相关文献理论，发现多因变量地理加权回归模型在处理空间非平稳性方面表现出色，但忽略时间效应可能导致模型在解释某些现象时存在局限性。考虑到时间维度对于理解地理现象的动态变化至关重要，尤其是在环境科学、社会科学和经济学等领域，时间的引入可以显著提高模型的预测精度和解释力。针对这一问题，本文探索将时间因素整合到多因变量时空地理加权回归模型中。这种扩展模型，本文称为多因变量时空地理加权回归模型（Multivariate Geographically and Temporally Weighted Regression, MGTWR），旨在捕捉地理空间和时间上同时存在的变化。

故本文具体内容及创新点为如下几个方面：

### (1) 多因变量时空地理回归方法研究

在时空地理数据分析中，研究者常常需要综合考虑多个因变量之间的关联性，并充分顾及空间和时间维度的影响。然而，传统的时空地理回归方法往往难以有效融合多个因变量的分析，大多限制于单一的因变量，使得同时建模和分析多因变量成为一项具有挑战性的任务。为了克服这一难题，本文结合了多因变量多元性回归的优势，其能够探测多因变量的特征，同时借鉴了经典的时空地理加权回归模型在探测时空非平稳性方面的特长。在 MGWR 先验知识的基础上，将二维地理空间转换为三维时空环境，并将最小二乘估算方法和最优带宽选择方法拓展于提出的多因变量时空地理回归方法，为多因变量下的时空非平稳性建模提供了有效的方法。

### (2) 将多因变量时空地理加权回归模型应用于复合空气污染研究

空气污染是当今社会面临的重要环境问题之一，而现实中的空气污染往往是由多个因素共同造成的复合污染。因此，研究者对于多因变量方法在解析空气复合污染中的应用进行了研究。多因变量方法可以帮助研究者同时考虑多个污染物的影响因素，揭示污染物之间的相互作用和共同影响机制。例如，研究者可以使用多因变量线性回归模型来分析不同污染物之间的关系，并探索其与气象条件、交通流量和人口密度等因素之间的交互作用。这样的分析有助于深入理解复合污染的形成机制和影响因素，为制定针对性的环境治理措施提供科学依据。

## 1.3.2 技术路线(Technical Pathway)

围绕大数据环境下多结果地理时空非平稳关系如何精准建模的核心问题，本文以多因变量多元线性回归理论和时空地理加权回归模型为基础。本文采用了模型推导与算法实现相结合的研究思路，将理论研究与应用研究相结合，形成了如下技术路线（见图 1-1）。本文以大数据多结果时空非平稳性关系如何精准建模问题为引导，重点展开了多变量的时空地理加权回归模型和模型实证分析的研究。

在新型方法实现方面，本文构建了包含基础理论研究、核心模型实现和诊断分析方法等内容。在基础理论研究层面，本文提炼了经典空间统计学、回归分析理论以及诊断分析相关理论，形成了多因变量时空地理加权回归方法的整体研究脉络。在核心模型与技术实现层面，本文重点研究了多因变量时空带宽选择，并深入探索了选取最优带宽方法算法，进一步构建了系列时空非平稳性诊断分析方法。



图 1-1 总技术路线图  
Figure 1-1 General technical pathway

在模型实证分析层面，以中国范围城市复合空气污染物为例，本文验证了多结果变量之间的相关性，并以华北平原为例分析了驱动因素与多结果的地理时空关系建模。本文从地理空间关系和复杂时空关系范畴给出了模型定义与设计实现，对方法进行了测试、

验证与应用，完成了模型的性能评价与对比工作，并揭示了驱动因素和多复合空气污染物发生与作用的机理，为中国环境空气监测与预测提供了技术支撑。

## 1.4 论文组织与章节安排(Thesis Organization and Chapter Arrangement)

本文组织结构如下：

### 第一章 绪论

本章阐述了本文的研究背景与意义，对目前时空非平稳建模方法进行了深入剖析，明确了在处理多因变量回归时，现有方法在时空非平稳建模方面存在的不足，并进一步强调了开展多因变量研究的迫切性与重要性。重点陈述了本文研究中核心内容和创新点。罗列了本文的整体组织结构，包括各章节的内容安排。

### 第二章 基础理论与研究方法

本章详细阐述了研究方法依托的理论知识，包括多因变量线性回归、时空地理加权回归及建模方法以及多因变量地理加权回归算法。多因变量线性回归揭示了多因变量对多自变量之间的影响关系。在时空地理加权回归中，主要讨论了模型估算方法——最小二乘法，以及最优带宽选择和探测时空非平稳性的优势。多因变量地理加权回归算法的基础理论为第三章提出的多因变量时空地理加权算法提供了先验知识，后续章节中的算法提出和应用奠定了基础。

### 第三章 多因变量时空地理加权回归算法实现

本章详细阐述了本文提出的多因变量时空地理加权回归方法的实现流程，包括模型回归估计方法、多因变量时空带宽的选择，以及多因变量时空非平稳性检验。首先介绍了该方法的基本原理，解释了多因变量在时空地理加权回归中的作用和影响关系。接着，详细讨论了模型回归估计方法，包括最小二乘法等常用方法，并探讨了其在多因变量时空地理加权回归中的适用性。此外，还介绍了多因变量时空带宽的选择方法，以确保回归模型的准确性和稳定性。最后，推导了多因变量时空非平稳性检验方法，以验证模型在多结果下时空维度上的实用性和可靠性。

### 第四章 多因变量时空地理加权回归模型实证分析

本章通过实证分析，对多因变量时空地理加权回归模型的有效性进行了验证和评估。首先验证城市范围内空气污染的存在性及相关性，继而采用多因变量时空地理加权回归算法进行实证分析。构建复合污染物的多因变量时空地理加权回归模型，引入时空权重函数，准确估计影响因子对空气质量的影响程度，并揭示复合污染中的相互作用机制。最后对比 MMLR, MGWR, MGTWR 模型的性能对比实验。通过对实证结果的深入解读，给出模型分析论证结论。

## 第五章 总结与展望

总结了研究成果，指出了研究的不足与局限性。对未来研究方向和展望进行了讨论，  
望为后续研究提供参考。

## 2 基础理论与研究方法

### 2 Basic Theory and Research Methods

#### 2.1 多因变量线性回归模型理论(Theory Of Multivariate Multiple Linear Regression Model)

回归分析方法在区域研究中被广泛采用，并被证明为一种有效的分析预测工具。然而，在区域研究领域中，往往需要研究多个变量之间的关系。这些变量之间可能存在复杂的相关性，不仅存在于自变量与因变量之间，还可能存在于自变量自身以及各个因变量之间。因此，不仅需要深入探究单个因变量与一组自变量之间的依赖关系，更需要全面地研究一组多个因变量与一组自变量之间的相互制约关系。为了实现这一目标，研究者在传统多元线性回归方法的基础上进行扩展。具体而言，通过对每个因变量与自变量之间的回归方程参数进行估计，并联立求解，最终可以得到一组回归方程的参数估计值。这种方法的基本求解方法仍然是基于最小二乘法。通过这种方法，可以综合地求解多个自变量与多个因变量之间的线性相关关系，这种回归方法被称为多因变量多元线性回归模型。这种方法为研究者提供了一种有效的工具，可以更全面地揭示变量之间的复杂关系，从而为区域研究提供更准确、更深入的见解。

##### 2.1.1 多因变量线性回归模型定义(Definition of Multivariate Linear Regression Model)

假设有  $p$  个自变量  $x_1, x_2, x_3, \dots, x_p$ ,  $m$  个因变量,  $Y_1, Y_2, Y_3, \dots, Y_m$ , 以及  $n$  个样本总数, 令  $Y_t = (y_{t1}, y_{t2}, \dots, y_{tm})$ , 其对应关系则有:

$$\begin{cases} y_{t1} = \beta_{01} + \beta_{11}X_{t1} + \beta_{21}X_{t2} + \dots + \beta_{p1}X_{tp} + \varepsilon_1 \\ y_{t2} = \beta_{02} + \beta_{12}X_{t1} + \beta_{22}X_{t2} + \dots + \beta_{p2}X_{tp} + \varepsilon_2 \\ \vdots \\ y_{tm} = \beta_{0m} + \beta_{1m}X_{t1} + \beta_{2m}X_{t2} + \dots + \beta_{pm}X_{tp} + \varepsilon_m \end{cases} \quad (2-1)$$

其中,  $(\beta_{01}, \beta_{11}, \beta_{21}, \dots, \beta_{p1}; \beta_{02}, \beta_{12}, \beta_{22}, \dots, \beta_{p2}; \beta_{0m}, \beta_{1m}, \beta_{2m}, \dots, \beta_{pm})$ ,  $(p+1) \times m$  个回归系数, 即  $(x_{t1}, x_{t2}, \dots, x_{tp}; y_{t1}, y_{t2}, \dots, y_{tm})$ ,  $t = 1, 2, \dots, n$ , 为一组观测值。

矩阵表达形式如下:

$$X_{(n \times p)} = \begin{bmatrix} x_{11} & x_{12} \dots x_{1p} \\ x_{21} & x_{22} \dots x_{2p} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \dots x_{np} \end{bmatrix} = (X_1, X_2, \dots, X_p) \quad (2-2)$$

$$Y_{(n \times m)} = \begin{bmatrix} y_{11} & y_{12} \dots y_{1p} \\ y_{21} & y_{22} \dots y_{2p} \\ \vdots & \vdots \\ y_{n1} & y_{n2} \dots y_{np} \end{bmatrix} = (Y_1, Y_2, \dots, Y_m) \quad (2-3)$$

其中：

$$X_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{bmatrix}, \quad i = 1, 2, \dots, p \quad (2-4)$$

$$Y_j = \begin{bmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{nj} \end{bmatrix}, \quad j = 1, 2, \dots, m \quad (2-5)$$

多变量对多变量线性回归的基本前提是，因变量  $Y_1, Y_2, \dots, Y_m$  与自变量  $X_1, X_2, \dots, X_p$  之间存在线性关系。这意味着在数据中，因变量和自变量之间存在可以通过线性回归模型来描述和预测这种关系。在多变量线性回归中，需要考虑多个自变量对因变量的影响，并确定之间的权重和交互作用。可表达为：

矩阵表示为：

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} = \begin{bmatrix} \beta_{01} & \beta_{01} & \beta_{01} & \dots & \beta_{01} \\ \beta_{02} & \beta_{01} & \beta_{01} & \dots & \beta_{01} \\ \vdots & \vdots & \vdots & & \vdots \\ \beta_{0m} & \beta_{01} & \beta_{01} & \dots & \beta_{01} \end{bmatrix} * \begin{bmatrix} 1 \\ X_1 \\ \vdots \\ X_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix} \quad (2-6)$$

简化为：

$$Y = (1|X)\beta + \varepsilon \quad (2-7)$$

其中， $\beta$  回归系数估计矩阵，相应的随机残差矩阵为  $\varepsilon$ 。

$$\varepsilon = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \dots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{12} & \dots & \varepsilon_{2m} \\ \vdots & \vdots & & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \dots & \varepsilon_{nm} \end{bmatrix} = \begin{bmatrix} \varepsilon'_{(1)} \\ \varepsilon'_{(2)} \\ \vdots \\ \varepsilon'_{(n)} \end{bmatrix} \quad (2-8)$$

其中，分量  $\varepsilon'_{(a)} = (\varepsilon_{a1}, \varepsilon_{a2}, \dots, \varepsilon_{am})$ ,  $a=1,2,3,4\dots,n$  相互之间独立。每个向量  $\varepsilon'_{(a)}$  的均值向量为 0，协方差矩阵设为  $Cov$ 。即  $\varepsilon$  为随机数，则相应的  $Y$  的均值向量也同样为随机数。故多因变量线性回归模型表示为：

$$\begin{cases} E(Y) = (1|X)\beta \\ \{Y'_{(1)}, Y'_{(2)}, \dots, Y'_{(n)}\} \text{ 互不相关, 协方差矩阵 } Cov > 0 \end{cases} \quad (2-9)$$

### 2.1.2 多因变量回归模型理论假设(Theoretical Assumptions of Multivariate Regression Model)

多因变量回归模型通常建立在以下基本假设的基础上：

**线性关系假设：**假设自变量与因变量之间存在线性关系，这意味着自变量的变化与因变量的变化之间可以用线性函数来描述。

**独立观测假设：**假设样本观测是相互独立的，即每个观测值是独立地从总体中获取的。这意味着样本之间的观测结果不会相互影响。

**多重共线性假设：**假设自变量之间不存在高度相关性。多重共线性指的是自变量之间存在较高的线性相关性，可能导致参数估计不准确或不可靠。

**常态性假设：**假设误差项（残差）服从独立同分布的正态分布。常态性假设使得最小二乘法成为参数估计的最优方法，同时也允许进行统计推断和假设检验。

**零均值误差假设：**假设误差项的均值为零，即误差的期望值为零。这个假设是为了确保回归模型中的截距项是可解释的，并且不引入额外的偏差。

**同方差性假设：**假设误差项的方差在所有自变量值上是恒定的，即误差项的方差不随自变量的变化而变化。

**模型可加性假设：**假设自变量的影响是可加的，即因变量的期望值可以通过自变量的线性组合来表示。

这些假设为多因变量回归模型提供了基本的理论基础和可解释性。在实际应用中，需要对这些假设进行检验和验证，以确保模型的适用性和准确性。如果某些假设被违反，可能需要采取相应的方法进行模型修正或转换。

### 2.1.3 模型回归估计方法(Model Regression Estimation Methods)

多因变量多元回归模型是一种用于分析多个自变量与多个因变量之间关系的统计模型。回归估计指的是通过对该模型进行拟合，从而估计自变量与因变量之间的关系。模型回归估计方法目标在于找到最佳的回归系数，以最精确地描述自变量与因变量之间的关联。

本文的回归估计采用最小二乘法，即通过最小化观测值与预测值之间的残差平方和

来确定最优回归系数。最小二乘法的目标是最小化观测值 $Y$ 与预测值 $\hat{Y}$ 之间的残差平方和，即最小化以下目标函数为 $Q$ :

$$Q = \sum_{i=0}^n \sum_{j=1}^p \varepsilon_{ij}^2 = (Y - \hat{Y})'(Y - \hat{Y}) \quad (2-10)$$

则有，最小目标函数的矩阵表现形式:

$$\varepsilon^T \varepsilon = [Y - (1|X)\hat{\beta}]^T [Y - (1|X)\hat{\beta}] \quad (2-11)$$

求得函数解 $\hat{\beta}$ 值为:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2-12)$$

其中，矩阵 $X^T X$ 约束条件为存在逆矩阵 $(X^T X)^{-1}$ ，则有:

$$\hat{Y} = X\hat{\beta} = (X^T X)^{-1} X^T Y = SY \quad (2-13)$$

故 $S$ 表示预测值 $\hat{y}$ 和因变量 $y$ 建立的关系矩阵，被成为帽子矩阵，即:

$$S = X(X^T X)^{-1} X^T \quad (2-14)$$

帽子矩阵的作用是将观测值映射到预测值，并计算模型的残差。给定输入矩阵 $X$ 和对应的响应变量 $y$ ，则其残差平方和可以表示为:

$$RSS = \sum_{i=1}^n \varepsilon_i^2 = Y^T (I - S)^T (I - S) Y \quad (2-15)$$

由 $RSS$ 将中机误差项方差的无偏估计记为:

$$\hat{\sigma}^2 = \frac{RSS}{n-p-1} \quad (2-16)$$

其中， $(n-p-1)$ 为无偏估计的 $RSS_{MMLR}$ 的自由度。

## 2.2 时空地理加权回归模型理论 (Theory of Geographically and Temporally Weighted Regression)

### 2.2.1 时空相关性以及非平稳性(Spatiotemporal Correlation and Non-stationarity)

地理实体之间普遍呈现出时空依赖性，并非相互独立的，常通过空间自相关衡量。通常情况下有正相关和负相关两种，正空间自相关表示相邻属性的值相似，负空间自相关表示相邻属性不相似。空间位置的相似性采用空间权重矩阵来度量，而空间属性的相似性以计算属性值之间的空间统计量来衡量。

地理空间相关性是指地理上相邻区域之间存在的相关性。相邻区域之间的相关性可能是由地理特征引起的，例如相似的地理特征、相似的人口构成和相似的经济条件。地理空间相关性能够借助空间自相关分析来量化。空间相关性分析通常分为两种：全局和局部，全局空间相关性通常采取使用莫兰指数进行定量分析，而局部空间相关性通常采

取使用局部莫兰指数进行定量分析<sup>[54]</sup>。这些指标能够帮助确定地理空间数据是否相关，以及相关程度如何。在地理信息科学中，地理空间关联的研究能够帮助更好地了解地理现象的空间分布，从而在城市规划、资源管理和环境保护方面做出更好的决策。

时空非平稳性是地理空间关联研究的核心问题，是指地球表面上的不同地点和不同时间之间存在差异和变化的现象，包括空间非平稳性和时间非平稳性两个方面<sup>[55,56]</sup>。空间非平稳性表现为地球表面上不同地点的地理特征、自然环境、人文活动等的差异和多样性，如地形、气候、土壤类型、植被分布和水文体系。这些差异形成了地球表面的空间非平稳性。时间非平稳性则指地球上的自然和人文过程在不同时间段内呈现出不同的特征和趋势，包括气候变化、人口迁移、城市发展和土地利用变化等。时空数据挖掘通过对时空数据进行处理和分析，从时空大数据中发现新的规律，挖掘深层次的知识。时空数据挖掘的关键和核心是提高时空分析和建模能力。其中，时空地理加权回归 (GTWR) 是一种时空数据挖掘的分析方法，可以通过局部加权回归系数来反映时间和空间上的非平稳性，从而提高对现实世界的建模能力。

## 2.2.2 时空地理加权回归模型定义(Definition of Geographically and Temporally Weighted Regression Model)

传统的全局回归模型在捕捉时空系数的空间非平稳性和识别解释变量与响应变量之间的局部特征方面面临挑战。空间环境的多样性特征，如不同城市之间工业结构、人口规模和排放源的变化，存在着显著的空间非平稳性。为了解决这些问题，GWR 模型引入了空间权重，实现了对研究区域内的局部回归，并有效地捕捉了不同区域的非平稳信息，具有很高的实际价值<sup>[57]</sup>。

地理加权回归模型扩展了基于假设空间的平稳的多元线性回归模型，将局部的地理位置引入全局回归模型来观测回归系数的空间特征，公式如下：

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \quad i = 1, 2, 3 \dots n \quad (2-17)$$

其中  $y_i$  代表在空间位置  $i$  的因变量， $x_{ik}$  代表在空间位置  $i$  的第  $p$  个因变量， $\beta_{i0}$  是空间变化截距。 $\beta_k(u_i, v_i)$  表示时空位置坐标  $(u_i, v_i)$  到第  $i$  个自变量回归系数， $\varepsilon_i$  是误差项， $n$  表示观测点总数。

此外，时空地理加权回归模型扩展了基于空间加权的地理加权回归模型，通过引入除空间信息的时间维，更好地考虑地理现象中的时间和空间复杂性，解决存在的时空非

平稳性问题。其空间坐标由二维坐标 $(u_i, v_i)$ 向三维坐标 $(u_i, v_i, t_i)$ 转变，即公式如下：

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^p \beta_k(u_i, v_i, t_i) x_{ik} + \varepsilon_i, \quad i = 1, 2, 3 \dots n \quad (2-18)$$

其中 $y_i$ 代表在时空位置 $i$ 的自变量， $x_{ik}$ 代表在时空位置 $i$ 的第 $p$ 个因变量， $\beta_{i0}$ 是时空变化截距。 $\beta_k(u_i, v_i, t_i)$ 表示时空位置坐标 $(u_i, v_i, t_i)$ 到第 $i$ 个自变量回归系数， $\varepsilon_i$ 是误差项， $n$ 表示观测点总数。

### 2.2.3 模型回归估计方法(Model Regression Estimation Methods)

在加权回归算法中，每个样本点的权重与其地理位置和时间距离有关。这些权重用于加权最小二乘法的求解过程中。加权最小二乘法通过对每个样本点进行加权，使得离群值和误差较大的点对模型拟合的影响减小，通过最小化以下目标函数实现：

$$Q(\beta) = \sum_{i=0}^n (y_i - x_{i1}\beta_{x1} - x_{i2}\beta_{x2} - \dots - x_{ip}\beta_{xp}) \quad (2-19)$$

寻找 $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n)$ 的估计值，使得残差平方和最小：

$$Q(\hat{\beta}) = \min(Q(\beta)) \quad (2-20)$$

当探测 GWR 时 $\beta_i$ 表示 $\beta_{i0}(u_i, v_i)$ ， $\sum_{k=1}^p \beta_k(u_i, v_i)$ ；当探测 GTWR 时空系数时， $\beta_i$ 可表示 $\beta_{i0}(u_i, v_i, t_i)$ ， $\sum_{k=1}^p \beta_k(u_i, v_i, t_i)$ 。 $w_{ij}$ 为当前位置点 $i$ 与其他受影响位置点 $j$ 之间时空（空间）距离的时空（空间）权重。即， $\beta_i = [\beta_{i0}, \beta_{i1}, \beta_{i2}, \dots, \beta_{ip}]^T$ ， $W_i = diag(w_{i1}, w_{i2}, w_{i3}, \dots, w_{in})$ ，第 $i$ 个回归系数 $\hat{\beta}_i$ 估算如下：

$$\hat{\beta}_i = (X^T W_i X)^{-1} X^T W_i y \quad (2-21)$$

其中， $X$ 是自变量矩阵的表示形式，维度为： $n \times (1+p)$ 。 $W_i$ 是根据指定的核函数和带宽计算的。空间分析中最广泛使用的核函数是 Bi-squared 和 Gaussian 函数，其他函数如逆距离加权(Inverse Distance Weighting, IDW)也是常用的方法。Bi-squared 和 Gaussian 核函数时空变形的方程如下：

$$w_{ij} = \exp\left[-\left(\frac{d_{ij}^{st}}{h}\right)^2\right] \quad (2-22)$$

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}^{st}}{h}\right)^2\right]^2 & d_{ij} \leq h \\ 0 & d_{ij} > h \end{cases} \quad (2-23)$$

其中，时空距离 $d_{ij}^{st}$ 扩展了 GWR 中的空间距离 $d_{ij}^s$ ，其考虑了观测点和受影响的时间

距离。在现实生活中，人们感知地球为一个平面，但获取的是经纬度坐标。因此，为了更符合实际需求，需要将这些坐标转换为高斯坐标系，以便进行欧几里得距离的计算。

#### 2.2.4 时空带宽选择(Spatiotemporal Bandwidth Selection)

带宽参数  $h$  在 GWR/GTWR 中起着至关重要的作用。其表示模型中权值与距离之间的非负衰减关系，选择合适的带宽对于保证模型估计的准确性至关重要。当带宽太大时，会包含对估计结果影响不大的点。这可能导致精度的损失，并可能模糊变量之间的关系。另一方面，如果带宽过小，估计结果可能会出现过拟合，即模型与数据拟合过于紧密，不能很好地推广到新的观测值。为了选择合适的带宽，有多种方法，包括交叉验证法(Cross Validation, CV)和赤池信息准则法校正值(Akaike Information Criterion corrected, AICc)。CV 方法在整个数据集上进行模型拟合和评估，并使用不同的带宽值评估模型的性能，在每个数据点上都会进行模型的训练和评估产生最佳预测性能的带宽。AICc 方法旨在通过惩罚具有较大带宽的模型来找到平衡模型拟合和复杂性的带宽。通常选择使 AICc 值最小的带宽。CV 和 AICc 方法都为两种模型估计选择最优带宽提供了系统的方法，同时考虑了模型复杂性和预测性能之间的权衡。

由空间距离  $(d_{ij}^s)^2 = (u_i - u_j)^2 + (v_i - v_j)^2$  推算得到时空距离  $d_{ij}^{st}$ ，GWR 的广泛运用验证了空间距离其合理性，进一步推论出的时空距离公式为：

$$(d_{ij}^{st})^2 = \lambda [(u_i - u_j)^2 + (v_i - v_j)^2] \divideontimes \mu (t_i - t_j)^2 \quad (2-24)$$

其中， $\lambda$ ， $\mu$  分别为时间和空间距离的比例因子， $\divideontimes$  作为一个操作符来整合两者之间的关系。在这里，黄波等人提出两者结合可以有效地提高模型的精度<sup>[58]</sup>，则将实验所持假设的公式定义为：

$$(d_{ij}^{st})^2 = [(u_i - u_j)^2 + (v_i - v_j)^2] + \tau (t_i - t_j)^2 \quad (2-25)$$

其中  $\tau$  代替  $\lambda$  和  $\mu$ ，简化了计算，同时达到了同样的训练效果。

### 2.3 多因变量地理加权回归算法模型(Multivariate Geographically Weighted Regression Model)

多因变量地理加权回归算法模型结合了地理加权回归和多因变量回归的思想，在考虑地理位置的同时，能够同时处理多个因变量。该模型首先确定多因变量之间的显著的相关性。然后，使用地理加权回归的方法为每个因变量建立一个局部回归模型。与传统

的多因变量全局回归模型不同，多因变量地理加权回归模型考虑了地理位置的影响，使得每个因变量的回归模型能够根据地理位置的不同进行调整。

### 2.3.1 多因变量地理加权回归模型定义(Definition of Multivariable Geographically and Temporally Weighted Regression Model)

MGWR 模型是在 MMLR 模型的拓展，其在计算回归系数时，考虑了由于空间位置的差异化而引起的空间关系变化。MGWR 公式具体表现如下：

$$y_{ij} = \beta_{0j}(u_i, v_i) + \sum_{k=1}^p \beta_{kj}(u_i, v_i) x_{ik} + \varepsilon_{ij}, \quad i=1, 2, \dots, n; j=1, 2, \dots, m \quad (2-26)$$

其中  $n$  个为样本位置的数量，每个地点包含  $m$  个响应和  $p$  个预测变量。设  $y_{ij}$  为位置  $i$  的第  $j$  个响应变量。其中，截距项为  $\beta_{0j}(u_i, v_i)$ ，空间变化系数为  $\beta_{kj}(u_i, v_i)$ ， $(u_i, v_i)$  是第  $i$  个位置的空间坐标，第  $j$  个响应变量的截距和第  $k$  个预测变量的回归系数； $\varepsilon_{ij}$  是第  $i$  个位置的第  $j$  个响应变量的随机误差。与 MMLR 类似，假设第  $i$  个位置的所有响应变量的误差行向量， $\varepsilon_i = [\varepsilon_{i1}, \dots, \varepsilon_{ij}]^\top$ ，具有均值为 0 的联合分布，其  $m \times m$  协方差矩阵为  $\Sigma$ 。如果响应变量之间存在强相关性，则  $\Sigma$  将不是对角矩阵。

为了便于说明，可以用矩阵形式表示 MGWR：

$$Y_i = X_i B(u_i, v_i) + \varepsilon_i \quad (2-27)$$

其中， $Y_i$  是位置  $i$  的响应变量的  $1 \times m$  行向量， $X_i$  表示相应的行向量( $1 \times (p+1)$ )包括截距在内的所有解释变量， $\varepsilon_i$  的定义与之前描述的一样。 $B(u_i, v_i)$  代表一个  $(p+1) \times m$  的参数矩阵，其对应的行  $B(u_i, v_i)$  作为每个预测变量的系数向量：

$$B(u_i, v_i) = \begin{bmatrix} \beta_{01}(u_i, v_i) & \beta_{02}(u_i, v_i) & \dots & \beta_{0m}(u_i, v_i) \\ \beta_{11}(u_i, v_i) & \beta_{12}(u_i, v_i) & \dots & \beta_{1m}(u_i, v_i) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p1}(u_i, v_i) & \beta_{p2}(u_i, v_i) & \dots & \beta_{pm}(u_i, v_i) \end{bmatrix} = \begin{bmatrix} \beta_0(u_i, v_i) \\ \beta_1(u_i, v_i) \\ \vdots \\ \beta_p(u_i, v_i) \end{bmatrix} \quad (2-28)$$

### 2.3.2 模型回归估计方法(Model Regression Estimation Methods)

从公式 (2-26) 中可以看出，MGWR 可以看作是传统（全局）MMLR 模型的局部变体。在 GWR 和 MMLR 技术的基础上，MGWR 可以通过局部回归方法进行校准，其中参数采用基于核的方法在逐点进行地理加权最小二乘估计，即最小化目标函数  $Q$ ：

$$Q(\beta) = \sum_{i=0}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=0}^n (Y_i - X_1 \beta(u_i, v_i) - X_2 \beta(u_i, v_i) - \dots - X_p \beta(u_i, v_i)) \quad (2-29)$$

寻找  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n)$  的估计值，使得残差平方和最小：

$$Q(\hat{\beta}) = \min(Q(\beta)) \quad (2-30)$$

然后，可以通过以下公式获得位置( $u_i, v_i$ )处第 $j$ 个因变量系数 $\beta$ 的估计值：

$$\hat{\beta}_j(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \quad (2-31)$$

其中， $W(u_i, v_i)$ 是一个 $n \times n$ 的对角权重矩阵，其对角线元素 $w_{ij}$ 是分配给在位置 $i$ 估计系数的第 $j$ 个观测值( $s=1, \dots, n$ )的地理权重。类似的， $W(u_i, v_i)$ 的值由核函数(公式 2-22, 2-23)决定，该函数在样本点接近 $(u_i, v_i)$ 的观测值上分配更多的权重，而核函数的计算取决于带宽的选择。

### 2.3.3 带宽选择(Bandwidth Selection)

同样的，带宽 $h$ 的选择在多因变量地理加权回归中也扮演着关键的角色。其直接影响模型的复杂性和最终估计的准确性。为了找到最佳带宽，需要谨慎权衡，以确保模型在保持精确性的同时不至于过于复杂。研究中通常采用不同的方法，例如第二章提及的交叉验证和 AIC 最小化，来优化带宽的选择。这些决策直接影响了模型的性能和可靠性。Harini 和 Purhadi 提出了一种留一交叉验证(CV)技术来最小化预测误差<sup>[59]</sup>，具体表达如下：

$$CV(h) = \sum_{i=1}^n \|Y_i - \hat{Y}_{\neq i}(h)\|^2 \quad (2-32)$$

其中， $\|\cdot\|$ 为 Frobenius 范数，即矩阵各项元素的绝对值平方的总和，求得最后的一维向量 CV 值，确定最小预测误差。

尽管 GTWR 理论中，CV 方法很有用，但 MGTWR 采用 AICc 最小化方法来选择带宽。AICc 通常被认为比 CV 更符合简洁性原则<sup>[60]</sup>。此外，研究表明，CV 分数可能会受到影响观测值的影响<sup>[61]</sup>。因此，本文建议在误差服从多元正态分布的假设下使用 AICc：

$$AICc = n \log |\hat{\Sigma}| + nm \log(2\pi) + \frac{n(n+tr(H))m}{n-tr(H)-m-1} \quad (2-33)$$

其中， $|\cdot|$ 表示矩阵的行列式， $\hat{\Sigma}$ 是误差协方差矩阵 $\Sigma$ 的最大似然估计值， $tr(S)$ 是帽子矩阵 $S$ 的迹，该矩阵的每一行都是对应观测值的估计系数向量 $r_i$ ：

$$r_i = X_i (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) Y \quad (2-34)$$

公式(3-7)中的 AICc 用作模型选择的手段，并且  $m=1$  时，其类似于单因变量的 GWR 的 AICc<sup>[16, 17]</sup>。

## 2.4 模型的评价指标(Model Evaluation Metrics)

相关性是一种衡量两个变量之间的关联程度的统计指标。相关性通过计算变量之间

的协方差或相关系数来评估它们之间的线性关系的强度和方向。相关性可以告诉数据变量之间是否存在一种趋势或模式，但不能确定因果关系或预测一个变量的值基于另一个变量。

回归分析是一种统计方法，用于建立自变量和因变量之间的数学模型。其目的是通过拟合一个数学方程来描述自变量对因变量的影响，并通过模型的参数来量化这种影响。回归分析可用于预测因变量的值，也可以用于解释因变量的变化。通过回归分析，可以推断自变量和因变量之间的因果关系，并利用模型进行预测和推断。

因此，相关性和回归分析是紧密相关的，但又有区别。相关性指示变量之间的关联程度，而回归分析则通过建立模型来揭示自变量对因变量的影响。相关性分析可以为回归分析提供初步的线索，而回归分析则可以进一步探究变量之间的关系并提供更具体的预测和解释。综合使用相关性和回归分析可以帮助研究者更好地理解和解释变量之间的关系。

在环境空气污染相关性研究中，传统的相关算法通常只评估城市主要空气污染物之间的相关性，而不考虑地理空间位置和时间变化的相互作用。为了解决这一局限性，本文采用了经典而有效的局部时空回归分析方法，即 GWR/GTWR 局部模型。该模型允许在特定时空位置测量城市空气污染物(如 PM<sub>10</sub> 对 PM<sub>2.5</sub>、PM<sub>10</sub> 对 O<sub>3</sub>、PM<sub>10</sub> 对 NO<sub>2</sub>)对彼此的影响。在这种情况下，模型可以计算统计 R<sup>2</sup> 指数来表示这些污染物之间的相关性。R<sup>2</sup> 指数衡量的是模型中因变量(如 PM<sub>10</sub>)可以解释的自变量(如 PM<sub>2.5</sub>)的拟合优度，公式(2-15)RSS 同样适用于 GWR 和 GTWR 的残差分析。两种模型中的空间、时空系数作为局部贡献程度的评价指标，其表明变量之间的关系在不同的地点和时间段是正负变化关系，以及变化的程度关系。在多因变量回归模型中，R<sup>2</sup> 亦可以作为复合指标来探测模型的拟合优度<sup>[22]</sup>。R<sup>2</sup> 指标可以用下式表示：

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (2-35)$$

## 2.5 本章小结(Chapter Summary)

本章介绍了时空地理加权回归算法通过考虑时空非平稳性，利用模型理论和最小二乘估计方法，通过时空带宽选择来优化模型性能。通过引入时空权重，有效地适应不同地理位置和时间点之间的差异性，从而更好地捕捉数据的时空特征。最小二乘估计方法被应用于拟合回归模型，通过最小化观测值与模型预测值之间的差异，确定最优参数，以提高模型的拟合效果。空间和时间带宽选择是一个关键步骤，其确定了权重的范围和

衰减程度。通过合理选择时空带宽，可以平衡模型的拟合精度和时空非平稳性的考虑，从而提高预测能力。多因变量多元回归算法能够同时考虑多个自变量对多个因变量的影响，提供更全面的分析。时空地理加权回归算法则能够探测时空非平稳性，在考虑时间和空间变化的情况下对数据进行建模和分析。多因变量地理加权回归模型分别是在多因变量多元线性回归模型和地理加权回归模型的方法融合和拓展，是对多因变量下空间非平稳关系建模的成功尝试。基于以上理论方法为构建考虑多个结果、顾及时空非平稳的多因变量的时空加权回归算法奠定了基础理论。

### 3 多因变量时空地理加权回归算法实现

## 3 Multivariate Geographically and Temporally Weighted Regression Model

由多元线性回归模型发展到地理加权回归模型，在发展到时空地理加权回归模型。类似地，多因变量多元线性回归模型<sup>[62]</sup>发展到多因变地理加权回归模型<sup>[25]</sup>，然而再进一步发展中，多因变量的时空地理加权回归模型仍存在研究的空白。

GTWR 模型基于空间和时间距离加权的思想，通过为每个观测值赋予不同的权重来考虑地理和时间上的非平稳性。在模型中，权重取决于每个观测值与其他观测值之间的空间距离和时间间隔。通过这种方式，GTWR 模型能够在建模过程中更准确地反映不同地理位置和时间点之间的关系。GTWR 模型目前主要用于单因变量的回归分析，但其仍然具有广泛的应用前景。研究人员可以利用 GTWR 模型来探索不同地理位置和时间点对单个因变量的影响差异，并在空间和时间维度上提供更精细的分析结果。

MMLR 和 GTWR 模型的框架为多因变量的时空地理加权回归算法提供了灵感。MGWR 模型的提出为本文 MGTWR 的引入提供了先验知识，确保了实施的可行性。将 GTWR 模型扩展到多因变量的情境，有助于更全面地理解多个因变量在时空上的非平稳性和相互关系。

### 3.1 多因变量时空地理加权回归算法模型(Multivariate Geographically and Temporally Weighted Regression Model)

同理，GWR 扩展到 GTWR，其空间位置到时空位置是由二维到三维的转变，所以由 MGWR 的概念扩展到 MGTWR，空间坐标变换为时空系坐标系( $u_i, v_i, t_i$ )， $\beta_{0j}(u_i, v_i, t_i)$ 和 $\beta_{kj}(u_i, v_i, t_i)$ 分别是第  $i$  个位置截距项和时空系数。则 MGTWR 模型可以定义为：

$$y_{ij} = \beta_{0j}(u_i, v_i, t_i) + \sum_{k=1}^p \beta_{kj}(u_i, v_i, t_i) x_{ik} + \varepsilon_{ij}, \quad i=1, 2, \dots, n; j=1, 2, \dots, m \quad (3-1)$$

其中，其中  $n$  个为样本数量，每个地点包含  $m$  个响应变量和  $p$  个预测变量。设  $y_{ij}$  为位置  $i$  的第  $j$  个响应变量。 $\varepsilon_{ij}$  是第  $i$  个位置的第  $j$  个响应变量的随机误差。与 MGWR 类似，假设第  $i$  个位置的所有响应变量的误差行向量， $\varepsilon_i = [\varepsilon_{i1}, \dots, \varepsilon_{im}]^T$ ，具有均值为 0 的联合分布，其  $m \times m$  协方差矩阵为  $\Sigma$ 。如果响应变量之间存在强相关性，则  $\Sigma$  将不是对角矩阵。

当  $m=1$  时，本文提出的 MGTWR 模型可以简化为经典单因变量的 GTWR 模型。

为了便于说明，可以用矩阵形式表示 MGTWR：

$$Y_i = X_i B + \varepsilon_i \quad (3-2)$$

其中， $Y_i$  是位置  $i$  的响应变量的  $1 \times m$  行向量， $X_i$  表示相应的行向量( $1 \times (p+1)$ )包括截距在内的所有解释变量， $\varepsilon_i$  的定义与之前描述的一样。 $B$  代表一个  $(p+1) \times m$  的参数矩阵，其对应的行  $\beta_k(u_i, v_i, t_i)$  作为每个预测变量的系数向量：

$$B = \begin{bmatrix} \beta_{01}(u_i, v_i, t_i) & \beta_{02}(u_i, v_i, t_i) & \dots & \beta_{0m}(u_i, v_i, t_i) \\ \beta_{11}(u_i, v_i, t_i) & \beta_{12}(u_i, v_i, t_i) & \dots & \beta_{1m}(u_i, v_i, t_i) \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{p1}(u_i, v_i, t_i) & \beta_{p2}(u_i, v_i, t_i) & \dots & \beta_{pm}(u_i, v_i, t_i) \end{bmatrix} = \begin{bmatrix} \beta_0(u_i, v_i, t_i) \\ \beta_1(u_i, v_i, t_i) \\ \vdots \\ \beta_p(u_i, v_i, t_i) \end{bmatrix} \quad (3-3)$$

### 3.2 模型回归估计方法(Model Regression Estimation Methods)

类似的，根据 MGWR 的加权回归最小二乘估计，即最小化目标函数  $Q$ ：

$$\begin{aligned} Q(\beta) &= \sum_{i=0}^n (Y_i - \hat{Y}_i) \\ &= \sum_{i=0}^n (Y_i - X_1\beta(u_i, v_i, t_i) - X_2\beta(u_i, v_i, t_i) - \dots - X_p\beta(u_i, v_i, t_i)) \end{aligned} \quad (3-4)$$

寻找  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n)$  的估计值，使得残差平方和最小：

$$Q(\hat{\beta}) = \min(Q(\beta)) \quad (3-5)$$

然后，可以通过以下公式获得位置  $(u_i, v_i, t_i)$  处系数  $\beta$  的估计值：

$$\hat{\beta}(u_i, v_i, t_i) = (X^T W(u_i, v_i, t_i) X)^{-1} X^T W(u_i, v_i, t_i) Y \quad (3-6)$$

其中， $W(u_i, v_i, t_i)$  是一个  $(n \times n)$  的对角权重矩阵，其对角线元素  $w_{ij}$  是分配给在位置  $i$  估计系数的第  $j$  个观测值的地理权重。 $X$  表示自变量矩阵， $Y$  表示多因变量矩阵，表达如下：

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \dots x_{1p} \\ 1 & x_{21} & x_{22} \dots x_{2p} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \dots x_{np} \end{bmatrix}, \quad Y = \begin{bmatrix} y_{11} & y_{12} \dots y_{1p} \\ y_{21} & y_{22} \dots y_{2p} \\ \vdots & \vdots \\ y_{n1} & y_{n2} \dots y_{np} \end{bmatrix} \quad (3-7)$$

第  $i$  个观测点  $Y$  的拟合值  $\hat{Y}_i$ ，可计算表达如下：

$$\hat{Y} = X_i B_i = X (X^T W(u_i, v_i, t_i) X)^{-1} X^T W(u_i, v_i, t_i) Y \quad (3-8)$$

其中， $X_i$  是自变量  $X$  的第  $i$  行向量， $B_i$  是多因变量对应的系数矩阵，MGTWR 的帽子矩阵  $S$  公式为：

$$S = X (X^T W(u_i, v_i, t_i) X)^{-1} X^T W(u_i, v_i, t_i) \quad (3-9)$$

将公式(3-8)与(3-9)结合，则有：

$$\hat{Y} = SY \quad (3-10)$$

故多因变量矩阵  $Y$  与多因变量拟合向量  $\hat{Y}$  之差即为残差向量  $E$ ：

$$E = Y - \hat{Y} = (I - S)Y \quad (3-11)$$

MGTWR 模型的残差平方和用  $RSS_{MGTWR}$  表示，计算公式为：

$$RSS_{MGTWR} = \sum_{i=1}^n \varepsilon_i^2 = E^T E = Y^T (I - S)^T (I - S)Y \quad (3-12)$$

其中， $tr(S)$  为帽子矩阵  $S$  的迹， $n - 2tr(S) + tr(S^T S)$  为 MGTWR 方程的有效自由度， $tr(S^T S)$  与  $tr(S)$  的结果十分相近，模型中随机误差项方差的无偏估计记为  $\hat{\sigma}^2$ ：

$$\hat{\sigma}^2 = \frac{RSS}{n - tr(S)} \quad (3-13)$$

### 3.3 多因变量时空带宽的选择(Bandwidth Selection for Multivariate Geographically and Temporally Weighted Regression)

在 MGWR 模型中，尽管可以通过矩阵范数求得 CV 值以及适用于多因变量的 AICc，从而确定一个特定的带宽，但这种方式却未能充分揭示带宽对于多个因变量的加权影响。而在 MGTWR 模型中，不仅需要根据空间距离进行加权，还需要融合时空因子进行加权，这进一步加剧了时空权重影响的遮蔽问题。若仅仅依赖传统的 CV 或 AICc 进行校正值，往往会遇到多个带宽并存的情况。这导致每个因变量都会得到一个独立的带宽值，即从单一的带宽转变为一个带宽集合  $H$ ，假设有  $p$  个因变量，则带宽集包含了  $h_1, h_2, \dots, h_p$  等多个带宽值。为了克服这一挑战，本文大胆假设了的  $fun()$  函数，旨在综合考虑时空因素，从而为模型提供一个更为合理且统一的带宽选择。则有 Bi-squared 和 Gaussian 核函数的方程表示如下：

$$w_{ij} = \exp \left[ - \left( \frac{d_{ij}^{st}}{fun(H)} \right)^2 \right] \quad (3-14)$$

$$w_{ij} = \begin{cases} \left[ 1 - \left( \frac{d_{ij}^{st}}{fun(H)} \right)^2 \right]^2 & d_{ij} \leq h \\ 0 & d_{ij} > h \end{cases} \quad (3-15)$$

其中时空距离定义如下：

$$(d_{ij}^{st})^2 = [(u_i - u_j)^2 + (v_i - v_j)^2] + \tau(t_i - t_j)^2 \quad (3-16)$$

平均数值反应数据的集中趋势常用的方法，带宽集合取平均数，以估计多因变量的

时空权重的影响，故 $fun()$ 函数具体设定为平均函数 $mean()$ ，得到对应的公式如下：

$$w_{ij} = \exp\left[-\left(\frac{d_{ij}^{st}}{mean(H)}\right)^2\right] \quad (3-17)$$

$$w_{ij} = \begin{cases} \left[1 - \left(\frac{d_{ij}^{st}}{mean(H)}\right)^2\right]^2 & d_{ij} \leq h \\ 0 & d_{ij} > h \end{cases} \quad (3-18)$$

这种方法能够在 MGTWR 中有效地平衡多个因变量的带宽选择，实现统一的带宽设定。这确保了模型能够全面考虑各因变量的时空非平稳性，并综合其影响。

### 3.4 多因变量时空非平稳性检验(Test of Spatiotemporal Non-stationarity)

由 Huang 和 Fotheringham 表明在进行 GTWR 模型分析时，有必要检验其全局时空非平稳性以及调查特定一组回归系数是否真的随时空位置的变化而变化<sup>[16, 17]</sup>。与经典的 GTWR 模型类似，MGTWR 模型在进行非平稳性检验时，也会基于一系列的假设来检验模型的适用性。

#### 3.4.1 全局时空非平稳性假设检验(Global Spatiotemporal Nonstationary Hypothesis Testing)

全局时空非平稳性检验确实是用于判断整个模型是否同时随空间和时间表现出非平稳性的重要手段。在进行全局时空非平稳性检验之前，通常需要首先检测模型是否存在空间非平稳性，因为空间非平稳性是时空非平稳性的一个组成部分。关于空间非平稳性的假设检验，原假设是所有回归系数在空间上恒定不变。备择假设则至少有一个系数随空间变化，反映数据空间非平稳性，全局回归模型即多 MMLR 可能不适用，用使用估计空间非平稳性的 MGWR。假设如下：

原假设( $H_0$ ):

$$y_{ij} = \beta_{0j} + \sum_{k=1}^p \beta_{kj} x_k + \varepsilon_{ij}, \quad i=1, 2, \dots, n; j=1, 2, \dots, m \quad (3-19)$$

备择假设( $H_1$ ):

$$y_{ij} = \beta_{0j}(u_i, v_i) + \sum_{k=1}^p \beta_{kj}(u_i, v_i) x_{ik} + \varepsilon_{ij}, \quad i=1, 2, \dots, n; j=1, 2, \dots, m \quad (3-20)$$

检验方法通常涉及构建统计量，比如基于残差或拟合优度的统计量，然后通过比较统计量的值与临界值来决定是接受原假设还是备择假设。如果统计量的值超过临界值，则拒绝原假设，认为模型存在全局时空非平稳性，适合使用 MGWR 模型。

相对于单变量的 GWR, 假设具有  $m$  个因变量的 MGWR 模型。由 3.3 节最小化方差可知残差平方根  $RSS = [rss_1, rss_2, \dots, rss_m]$  的公式可以变换为:

$$rss_j = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_{ij} - \hat{y}_{ij})^2, j = 1, 2, \dots, m \quad (3-21)$$

其矩阵形式则为:

$$RSS = Y^T(I - S)^T(I - S)Y$$

其中, 帽子矩阵  $S$  表示如下:

$$S = \begin{bmatrix} X_1(X^T W(u_1, v_1) X)^{-1} X^T W(u_1, v_1, t_1) \\ X_2(X^T W(u_2, v_2) X)^{-1} X^T W(u_2, v_2, t_2) \\ \vdots \\ X_n(X^T W(u_n, v_n, t_n) X)^{-1} X^T W(u_n, v_n, t_n) \end{bmatrix} \quad (3-22)$$

当权重为  $W$  为单位矩阵时, 则退化为 MMLR。构建  $F$  统计量, 则有:

$$F_1 = \frac{\left[ (Y^T RSS_{MMLR} Y - Y^T RSS_{MGWR} Y) / \frac{Y^T RSS_{MGWR} Y}{tr(RSS_{MGWR})} \right]}{Y^T RSS_{MGWR} Y / tr(RSS_{MGWR})} \quad (3-23)$$

其中,  $F_1$  服从自由度  $((tr^2(RSS_{MMLR} - RSS_{MGWR}) / tr[RSS_{MMLR} - RSS_{MGWR}]^2), (tr^2(RSS_{MMLR}) / tr[RSS_{MMLR}]^2))$ , 检验  $p$ -值为:

$$p_1 = P_{H_0}(F_1) \quad (3-24)$$

如果  $p_1$  值不满足上述空间非平稳性的条件 (即  $p_1$  值大于  $\alpha$ ), 那么不能简单地认为数据集存在空间非平稳性, 即接受  $H_0$  假设, 即数据集回归系数不随空间位置变化而变化, 但是拒绝了随空间变化的假设, 即不存空间非平稳性。

同样的, 关于时空非平稳性的假设检验, 原假设是所有回归系数在时空上恒定不变。备择假设则至少有一个时空系数随时间变化, 反映数据时空非平稳性, 回归模型即多 MGWR 可能不适用, 需使用顾及时空非平稳性的 MGTWR。

基于以下两个假设:

原假设( $H_0$ ):

$$y_{ij} = \beta_{0j}(u_i, v_i) + \sum_{k=1}^p \beta_{kj}(u_i, v_i) x_{ik} + \varepsilon_{ij}, \quad i=1, 2, \dots, n; j=1, 2, \dots, m \quad (3-25)$$

备择假设( $H_1$ ):

$$y_{ij} = \beta_{0j}(u_i, v_i, t_i) + \sum_{k=1}^p \beta_{kj}(u_i, v_i, t_i) x_{ik} + \varepsilon_{ij}, \quad i=1, 2, \dots, n; j=1, 2, \dots, m \quad (3-26)$$

检验方法通常涉及构建统计量, 比如基于残差或拟合优度的统计量, 然后通过比较统计量的值与临界值来决定是接受原假设还是备择假设。如果统计量的值超过临界值,

则拒绝原假设，认为模型存在全局时空非平稳性，适合使用 MGTWR 模型。

相对于单变量的 GTWR, 假设具有  $m$  个因变量的 MGWTR 模型。由 3.3 节最小化方差可知残差平方根  $RSS = [rss_1, rss_2, \dots, rss_m]$  的公式可以变换为：

$$rss_j = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_{ij} - \hat{y}_{ij})^2, j = 1, 2, \dots, m \quad (3-27)$$

其矩阵形式则为：

$$RSS = Y^T(I - S)^T(I - S)Y \quad (3-28)$$

其中，帽子矩阵  $S$  表示如下：

$$S = \begin{bmatrix} X_1(X^T W(u_1, v_1, t_1)X)^{-1} X^T W(u_1, v_1, t_1) \\ X_2(X^T W(u_2, v_2, t_2)X)^{-1} X^T W(u_2, v_2, t_2) \\ \vdots \\ X_n(X^T W(u_n, v_n, t_n)X)^{-1} X^T W(u_n, v_n, t_n) \end{bmatrix} \quad (3-29)$$

构建  $F$  统计量，则有：

$$F_2 = \frac{\left[ (Y^T RSS_{MGWR} Y - Y^T RSS_{MGTWR} Y) / \frac{Y^T RSS_{MGTWR} Y}{tr(RSS_{MGTWR})} \right]}{Y^T RSS_{MGTWR} Y / tr(RSS_{MGTWR})}$$

其中， $F_2$  服从自由度  $((tr^2(RSS_{MGWR} - RSS_{MGTWR}) / tr[RSS_{MGWR} - RSS_{MGTWR}]^2), (tr^2(RSS_{MGWR}) / tr[RSS_{MGWR}]^2))$ ，检验  $p$ -值为：

$$p_2 = P_{H_0}(F_2) \quad (3-30)$$

如果  $p_2$  值不满足上述空间非平稳性的条件（即  $p_2$  值大于  $\alpha$ ），那么不能简单地认为数据集存在时空非平稳性，即接受  $H_0$  假设，即数据集回归系数随空间位置变化而变化，但是拒绝了随时间变化的假设，即不存在时间的非平稳性。

### 3.4.2 回归系数时空非平稳性假设检验(Hypothesis Testing for Spatio-Temporal Nonstationarity of Regression Coefficients)

RSS 检验可以帮助了解回归模型的整体时空非平稳性，但要深入了解每个自变量的回归系数的时空变化，还需要进行更详细的统计分析和检验<sup>[63]</sup>。单变量回归系数时空非平稳性检验更为具体，其关注模型中每个自变量的回归系数是否分别具有时空非平稳性。

在多因变量回归模型中，多个因变量对应多个自变量的关系，则有对第  $m$  个因变量的第  $k$  个自变量的回归系数  $\beta_{mk}(u_i, v_i, t_i) = [\beta_{1k}, \beta_{2k}, \dots, \beta_{nk}] = B_{ik}$  进行检验，判断其在  $i$  点上是否存在显著的时空变化。同样基于一系列假设：

原假设( $H_0$ ):  $B_{1k} = B_{2k} = \dots = B_{nk}$

备择假设( $H_1$ ): 至少存在  $i \neq j$ ,  $B_{ik} \neq B_{jk}$

构建第  $k$  个自变量的统计量  $Z_k^2$  为:

$$Z_k^2 = \frac{1}{n} \sum_{i=1}^n \left( \hat{B}_{ik} - \frac{1}{n} \sum_{i=1}^n \hat{B}_{ik} \right)^2 \quad (3-31)$$

令  $\hat{B}_k = [\hat{B}_{1k}, \hat{B}_{2k}, \dots, \hat{B}_{nk}]^T$ , 且  $L$  表示为  $n \times n \times m$  全为 1 的矩阵, 那么可以将  $Z_k^2$  表示为:

$$Z_k^2 = \frac{1}{n} \hat{B}_k^T \left( I - \frac{1}{n} L \right) \hat{B}_k \quad (3-32)$$

基于零假设, 令  $I - \frac{1}{n} L = O$ , 得到  $Z_k^2$  的进一步表示:

$$Z_k^2 = \frac{1}{n} \left( \hat{B}_k - E(\hat{B}_k) \right)^T \left( I - \frac{1}{n} L \right) \left( \hat{B}_k - E(\hat{B}_k) \right) \quad (3-33)$$

令  $e_k$  表示第  $m$  行第  $(k+1)$  个值为 1, 其他值为 1 的  $(p+1) \times m$  矩阵,  $m$  表示因变量个数,  $p$  表示自变量个数。

则  $m$  行向量第  $k$  个回归系数  $B_k$  的  $F$  统计量为:

$$F_3 = \frac{Z_k^2 / \text{tr}\left(\frac{1}{n} B^T \left(I - \frac{1}{n} L\right) B\right)}{\hat{\delta}^2} \quad (3-34)$$

其中,  $\hat{\delta}^2$  假定为无偏估计, 并参考 Leung 等人的基础理论<sup>[63]</sup>, 则有:

$$p_3 = P_{H_0}(F_3) \quad (3-35)$$

同样给定一个显著性水平  $a$ , 如果  $p_3 > a$ , 则拒绝  $H_0$  假设, 即存在回归系数的非平稳性, 否则接受  $H_0$  假设, 不存在回归系数的时空非平稳性。

### 3.5 本章小结(Chapter Summary)

本章所提出的多因变量时空地理加权回归算法模型, 相较于经典的时空地理加权回归模型, 其突破了单因变量的限制, 引入多因变量的概念, 弥补了单变量时空分析的空白。相较于传统的多因变量多元线性回归模型, 有效克服了传统时空分析在处理非平稳性问题时的局限性。在模型适用性和有效性评估上, 本文研究了全局时空非平稳性假设检验方法, 来判断多因变量时空地理加权回归模型在能够在解决非平稳性问题上更为有效, 揭示因变量与自变量间的复杂时空关系。同时, 还研究了回归系数时空非平稳性检验, 判断模型的回归系数是否具有显著的时空非平稳性, 即不同时空位置下解释变量对被解释变量的影响程度不同。

综上所述, 本章提出的多因变量时空地理加权回归算法模型不仅拓宽了时空数据分析的视野, 也为深入探究复杂时空动态提供了新的视角。

## 4 多因变量时空地理加权回归模型实证分析

### 4 Empirical Analysis of Multivariate Geographically and Temporally Weighted Regression Model

本章旨在将多因变量时空地理加权回归算法应用于中国城市空气污染数据集，通过对实验结果和评价指标的比较，验证该方法的性能和适用性。在进行多因变量时空地理加权回归算法的实证分析之前，需要首先验证城市范围内空气污染的存在性及其相关性。为此，本章收集了全国范围内的空气质量监测数据，并运用统计分析方法揭示不同污染物之间的相关性以及它们的时空变化规律。这些结果为后续的实证分析提供了重要的数据支持和理论依据。

实验研究在确认了城市空气污染的存在性及其相关性后，本章进一步采用多因变量时空地理加权回归算法进行实证分析。本文构建了一个复合污染物的多因变量时空地理加权回归模型，引入时空权重函数，以更准确地估计各影响因子对空气质量的影响程度，并揭示复合污染中各污染物之间的相互作用机制。

#### 4.1 空气复合污染物相关性分析(Correlation Analysis of Air Composite Pollutants)

##### 4.1.1 实验区域及数据概(Overview of Experimental Area and Data)

实验区位于亚洲东部的太平洋西海岸的中国，北纬  $3^{\circ}86' - 53^{\circ}55'$ ，东经  $73^{\circ}66' - 135^{\circ}05'$ 。中国幅员辽阔，陆地面积约 960 万平方公里，是继俄罗斯和加拿大之后的世界第三大国。近年来，中国经济经历了快速增长，以满足其人口日益增长的物质和文化需求。然而，这种增长是有代价的，导致了各种挑战。这些挑战包括对不可再生能源的持续依赖、优化产业结构的必要性、加快城市交通发展的迫切需要，以及在预防和控制空气污染方面相对滞后努力。

本研究利用了 2015 年至 2021 年从中国国家环境监测中心（CNEMC，<http://www.cnemc.cn>）收集的来自中国各地 1,497 个监测点的开源数据。这些地点在图 4-1 中以点单元方式表示，以线方式刻画了中国特定的地理边界。中国根据省级行政区划将其领土划分为七个主要地理区域，旨在促进地理、人类历史、气候、经济和行政治理等各方面的研究和管理。中国区域以黑河-腾冲线划分了人口密度，以秦岭-淮河线在地理上将中国北部和南部地区与秦岭山脉和淮河相隔开来。数据包括直径小于等于 2.5 微米细颗粒

粒物 PM<sub>2.5</sub>(Particulate Matter 2.5, PM<sub>2.5</sub>)和直径小于等于 10 微米的细微颗粒物(Particulate Matter 10)、二氧化氮(Nitrogen Dioxide, NO<sub>2</sub>)、二氧化硫(Sulfur Dioxide, SO<sub>2</sub>)、一氧化碳(carbon monoxide, CO)和臭氧(Ozone, O<sub>3</sub>)的观测值。为了使数据标准化，城市内所有监测站的污染物 24 小时平均监测浓度被用来代表城市污染物，O<sub>3</sub> 污染物的监测浓度被替换为 8 小时平均值。最后，通过计算月平均值以确保时间连续性。为了解决城市污染浓度数据缺失的问题，采用普通克里格插值法<sup>[64]</sup>。这种插值技术填补了部分数据的缺失，从而获得了 2015 年-2021 年中国 366 个城市的月度数据，本文同时验证了数据符合正态分布。

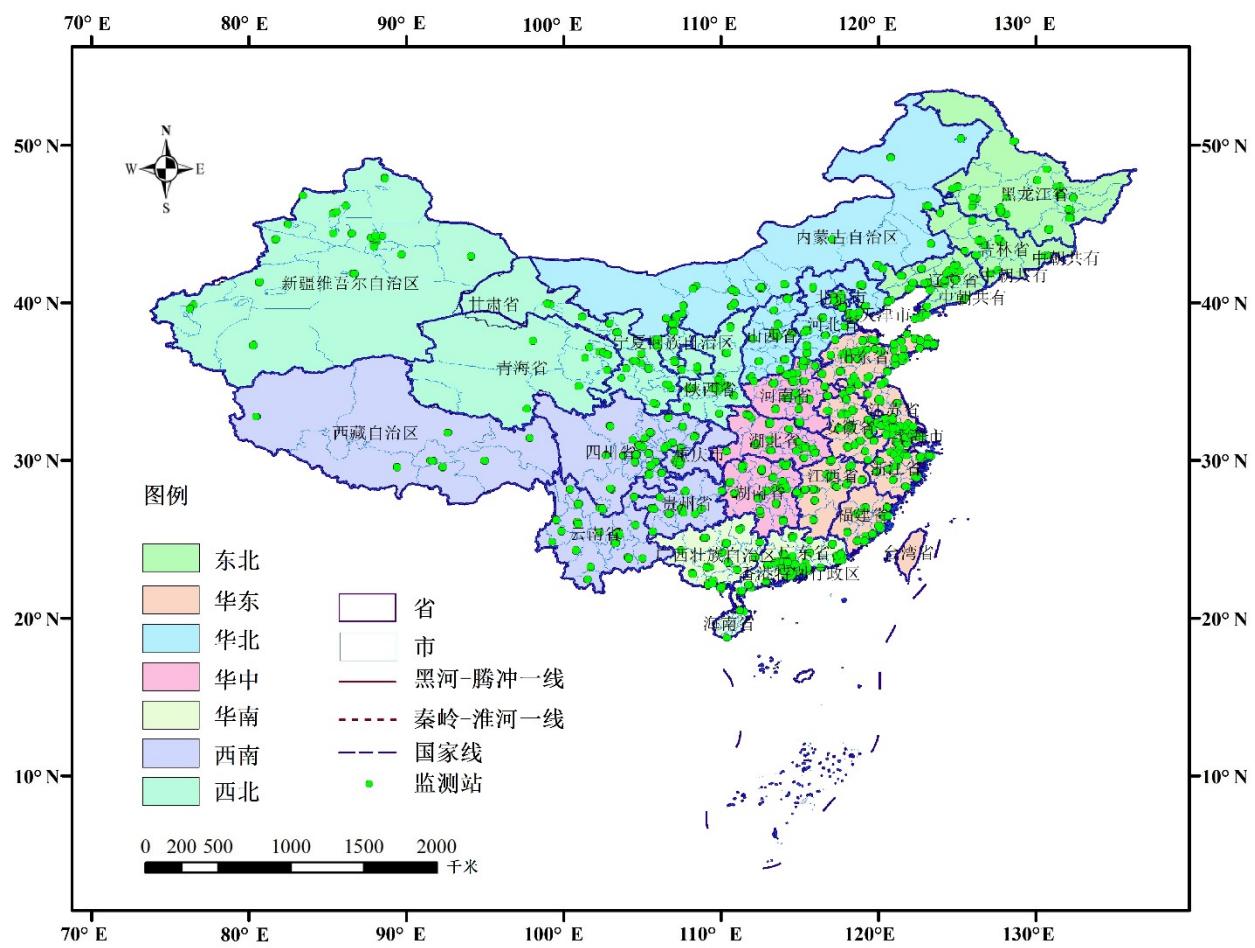


图 4-1 研究区域-中国

Figure 4-1 Study area - China

为了比较不同污染物的年度和月度变化，本文使用箱线图分析了浓度范围和平均值的变化。本研究发现表明，包括 PM<sub>2.5</sub>、PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub> 和 CO 在内的各种污染物都表现出政策治理下的下降趋势（图 4-2）。在 2022 年世界环境日国家活动中，生态环境部黄部长指出，中国地级及以上城市 2021 年 PM<sub>2.5</sub> 平均浓度比 2015 年下降 34.8%。本文所呈现的研究数据与此声明基本一致，显示中国地级及以上城市 PM<sub>2.5</sub> 年度平均浓度下降了 37.9%。同样，其他污染物的浓度也有所下降：PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub> 和 CO 分别下降了 28.0%、

64.4%、19.0%和37.4%。然而， $O_3$ 的浓度呈现波动状态，增加了9%。研究表明，颗粒物的增加可能是抑制 $O_3$ 浓度的上升原因，而从2015年到2021年颗粒物污染的显著减少可能是导致 $O_3$ 浓度增加的一个因素。各种污染物的月度趋势变化呈现出非线性模式， $PM_{2.5}$ 、 $PM_{10}$ 、 $SO_2$ 、 $CO$ 和 $NO_2$ 与 $O_3$ 呈负相关，前者在冬季污染期间更为严重，后者在夏季表现出更高的浓度（图4-3）。

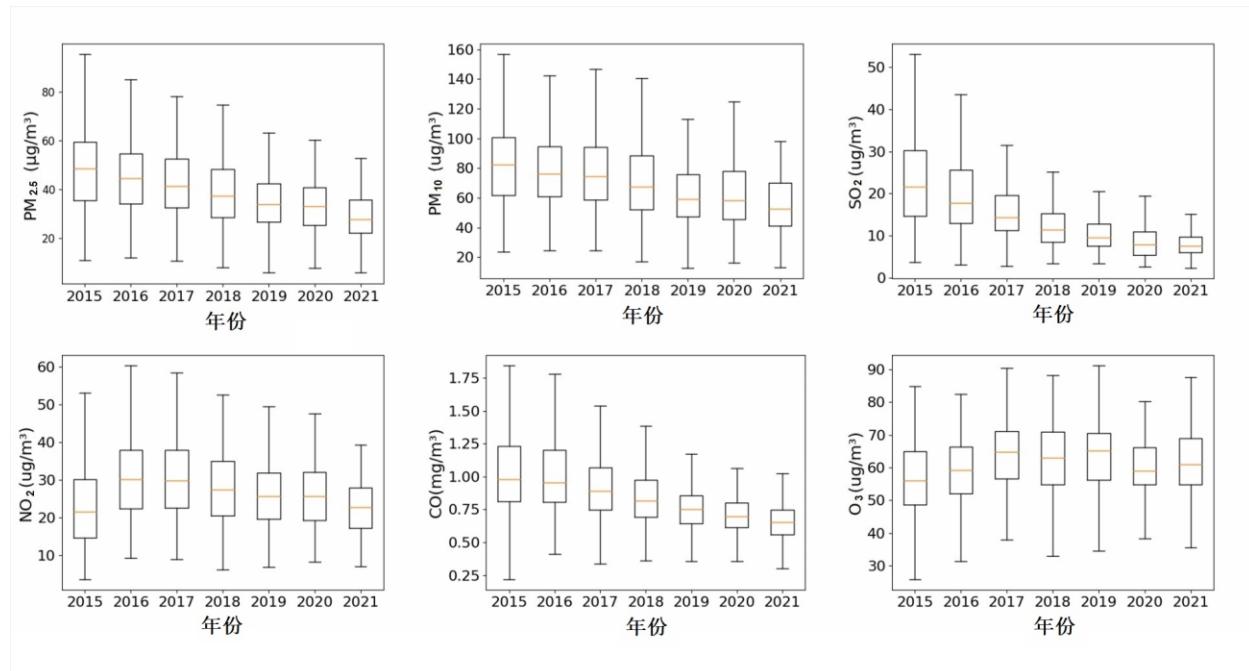


图4-2 空气污染物的年均趋势变化  
Figure 4-2 Annual trend changes in air pollutants

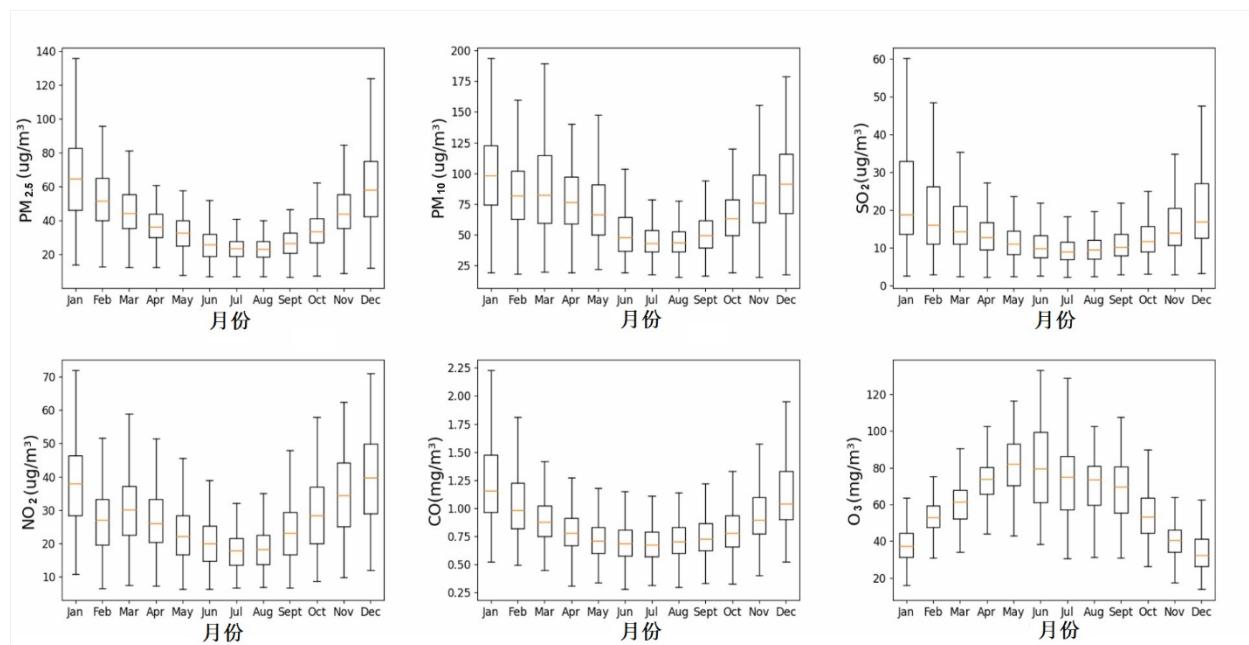


图4-3 空气污染物的月均趋势变化  
Figure 4-3 Monthly trend changes in air pollutants

### 4.1.2 实验设计与结果分析(Experimental Design and Results Analysis)

本文以中国地级及以上城市空气质量为研究数据，并利用莫兰指数来检测城市空气污染物的空间聚类。这种分析有助于更好地了解城市空气污染物的分布特征，包括聚类的程度和范围，这对于制定环境规划和决策至关重要。从 2015 年到 2021 年的各种污染物全球莫兰指数图表明，中国区域城市之间的大气环境污染存在显著的聚类现象（表 4-1a）。从图表中可以观察到，从 2015 年到 2019 年，莫兰指数保持相对稳定，但从 2020 年到 2021 年，出现了显著的波动。这种波动可能是由于防疫期间采取的封锁措施导致的城市间活动不一致和系数值下降。颗粒物聚类在 2020 年显著减少，而臭氧的聚类呈现上升趋势，需要进一步解释。此外，莫兰指数统计的应用揭示了城市空气污染物季节性变化之间的显著相关性（表 4-1b）。具体来说，PM<sub>2.5</sub> 作为主要的城市空气污染物，其浓度表现出较高的聚集程度，其值超过 0.5，但在春季除外。相反，臭氧的浓度在夏季表现出较高的聚集水平，而在冬季其聚集程度最低。此外，其他污染物的聚集程度在夏季较低，在冬季较高。

进一步的调查显示，城市空气污染物存在地理集中现象。详细的图表展示了 2015 年至 2021 年间中国城市空气污染物浓度的年度和季节性平均变化，以及局部空间关联指标（Local Indicators of Spatial Association, LISA），如图 6 至图 11 所示。年度平均变化的特征表明，PM<sub>2.5</sub> 和 PM<sub>10</sub> 年度浓度最高的站点主要位于华北地区，其次是东北和华中地区。LISA 特征图显示了一种高-高的聚类状态，特别是在华北和华中东部地区（图 4-4 和图 4-5）。该地区属于华北平原，地形平坦，河流湖泊众多，交通便利，经济发展，导致人类活动相对密集。新疆西部是另一个 PM<sub>2.5</sub> 和 PM<sub>10</sub> 浓度较高的地区。该地区存在严重的荒漠化、植被覆盖率低、水土流失严重、粗颗粒污染物含量高<sup>[65]</sup>。相反，PM<sub>2.5</sub> 和 PM<sub>10</sub> 年度浓度最低的站点通常位于青藏高原、华南和西南地区。特别是华南地区的沿海地区，该地区强海风对污染物具有一定的吸附作用。珠江三角洲地区经济发展迅速，表明采取了有效的控制和预防措施。

表 4-1 2015 年至 2021 年年度/季节性全局 Moran's I  
Table 4-1 Annual/Seasonal Global Moran's I from 2015 to 2021

时间	PM <sub>2.5</sub>	PM <sub>10</sub>	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub>
(a) 年份						
2015	0.62	0.61	0.60	0.60	0.51	0.49
2016	0.62	0.61	0.62	0.54	0.52	0.51
2017	0.61	0.59	0.55	0.57	0.49	0.55
2018	0.64	0.62	0.53	0.59	0.50	0.58
2019	0.59	0.58	0.50	0.57	0.47	0.55
2020	0.52	0.61	0.52	0.50	0.50	0.63
2021	0.57	0.44	0.41	0.50	0.38	0.43
(b) 季节						
春	0.45	0.45	0.53	0.53	0.45	0.46
夏	0.54	0.42	0.40	0.47	0.45	0.58
秋	0.56	0.49	0.51	0.57	0.43	0.53
冬	0.57	0.54	0.63	0.56	0.45	0.39

就季节性空间变化而言，新疆西部的 PM<sub>2.5</sub> 和 PM<sub>10</sub> 的季节性变化影响较小（图 4-4 和图 4-5）。该地区的气候特征是春季变化快且不稳定，秋季季节短而快，以晴天为主，阳光充足，空气干燥，沙尘暴频繁。就空间浓度而言，PM<sub>2.5</sub> 在夏季、秋季和冬季显著下降。另一方面，PM<sub>10</sub> 始终呈现出高-高聚集状态，周围地区受到这种高-高聚集状态的影响，呈现出低-高聚集状态。

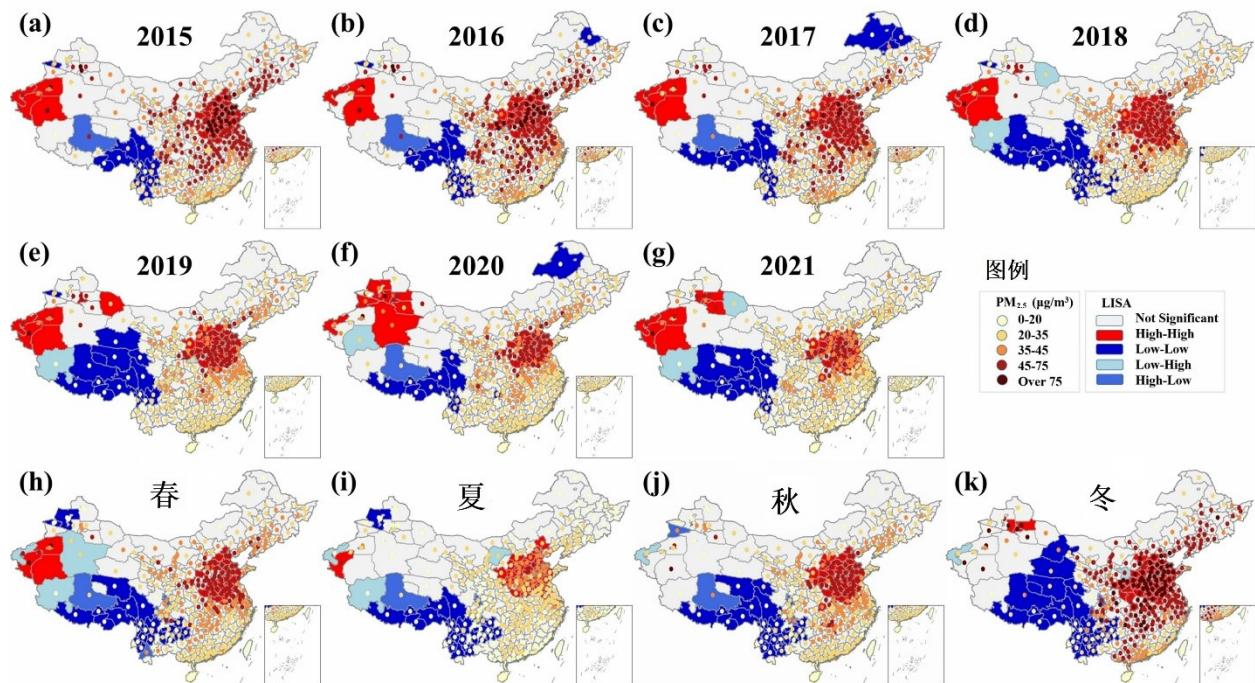


图 4-4 2015 - 2021 年中国 PM<sub>2.5</sub> 浓度年/季节平均变化

Figure 4-4 Annual and seasonal average changes of PM<sub>2.5</sub> concentration in China from 2015 to 2021

北京、天津和河北省及其周边地区的二氧化硫和二氧化氮浓度一直最高，这些地区构成了重化工产业集群。该区域呈现高-高浓度模式（图 4-6 和图 4-7）。虽然颗粒物污染得到了有效控制，但 CO 浓度也得到了很好的控制。自 2015 年以来，污染范围不断缩小，空气质量显著改善，特别是自 2019 年以来（图 4-8）。O<sub>3</sub> 浓度的年度峰值出现在华北、华东和华中地区，尤其是京津冀城市群。从 2017 年到 2018 年，O<sub>3</sub> 浓度呈上升趋势（图 4-9）。在 2020-2021 年的疫情期间，工业活动的减少导致城市空气污染物减少，能见度提高。这反过来又在一定程度上促进了光化学反应，促进了 O<sub>3</sub> 的形成。O<sub>3</sub> 的 LISA 特征图揭示了一个大而宽的高-高聚集区，从人口稠密的京津冀地区延伸到人口稀少的西北地区。城市空气污染物的浓度，包括 PM<sub>2.5</sub>、PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub> 和 CO，都受到人类活动的显著影响。前人对人类活动对城市的影响进行了具体调查<sup>[66]</sup>。由于空气污染具有非线性性质，有效管理人类活动对于降低污染物浓度至关重要。薛等人<sup>[67]</sup>研究了 2016-2019 年政策控制下华北平原城市大气污染物（O<sub>3</sub> 除外）的减少情况。这项研究还发现，在新冠肺炎大流行期间，O<sub>3</sub> 的浓度并没有随着人类活动的减少而降低<sup>[19]</sup>。

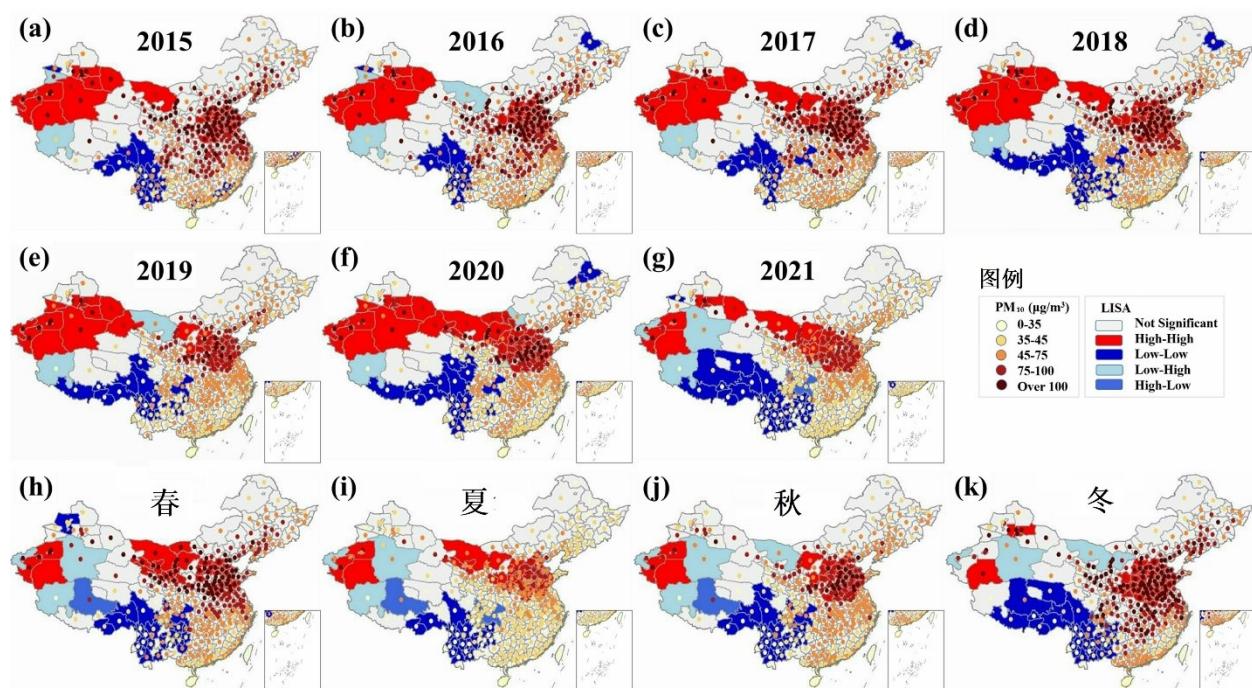


图 4-5 2015 - 2021 年中国 PM10 浓度年/季节平均变化

Figure 4-5 Annual and seasonal average changes of PM10 concentration in China from 2015 to 2021

在冬季，PM<sub>2.5</sub>、PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub> 和 CO 表现出高浓度，尤其是在华北平原地区（图 4-4k 至 10k）。最严重的污染物是 PM<sub>2.5</sub> 和 SO<sub>2</sub>，这两种颗粒物从北京-天津-河北（BTH）地区向外扩散（图 6 和 8）。在宁夏平原、华北平原和东北平原等平原地区，SO<sub>2</sub> 在冬季表现出较高的污染水平（图 4-6）。在这些平原地区，冬季 SO<sub>2</sub> 的污染水平较高。

西南气流将各种污染物从河北省输送到全国平原地区<sup>[68]</sup>。西北地区全年持续的污染归因于新疆沙漠地区对颗粒物污染浓度的影响<sup>[69]</sup>，以及二次层、生物质和废物燃烧、车辆排放、地壳矿物、工业污染和煤炭燃烧<sup>[70]</sup>。总体而言，污染物浓度的变化并没有导致污染物空间分布的转变。

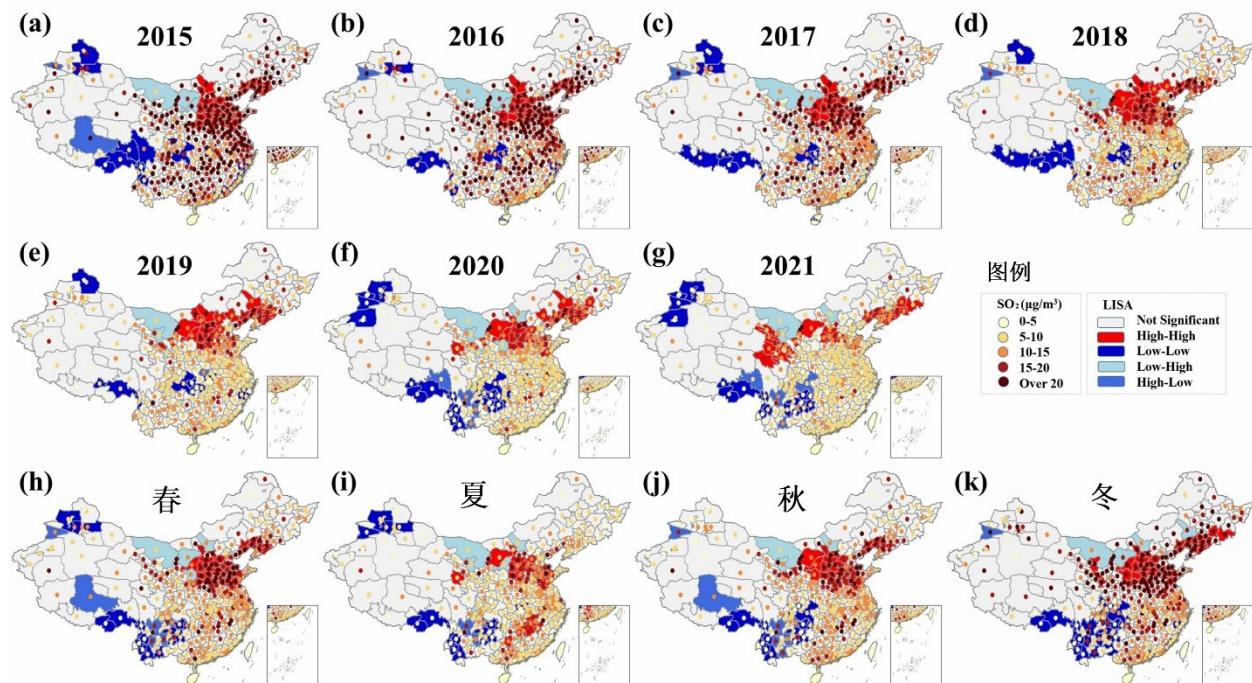


图 4-6 2015 - 2021 年中国 SO<sub>2</sub> 浓度年/季节平均变化

Figure 4-6 Annual and seasonal average changes of SO<sub>2</sub> concentration in China from 2015 to 2021

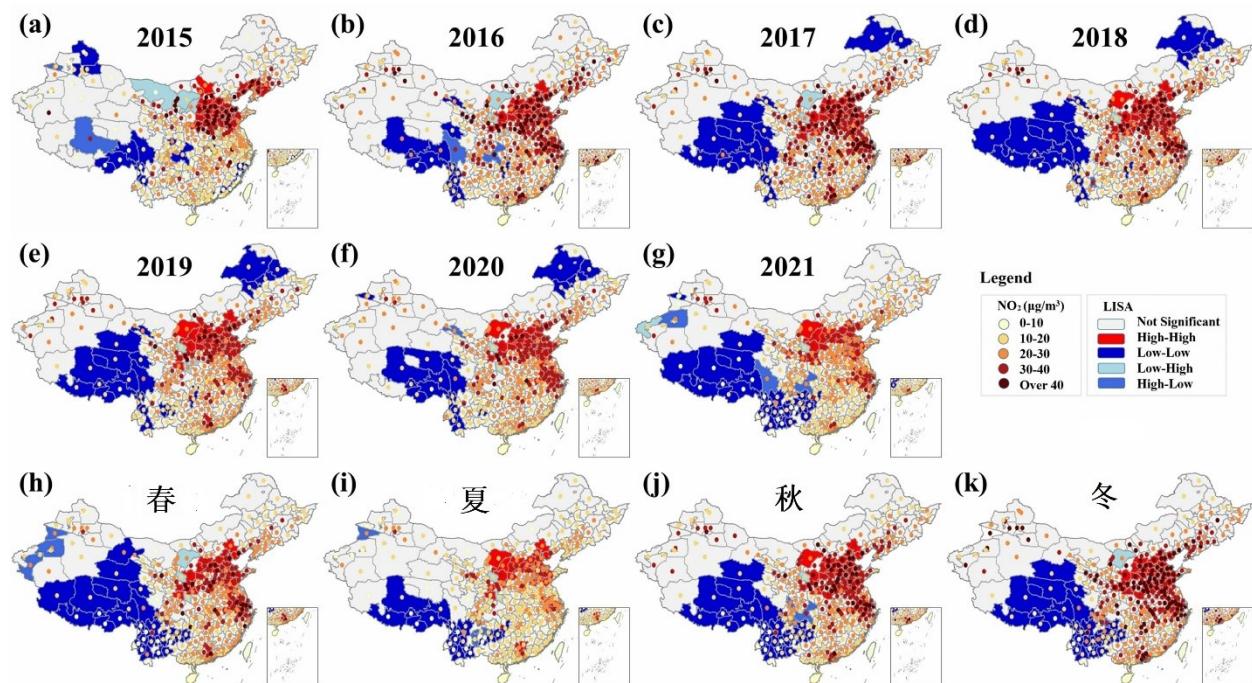


图 4-7 2015 - 2021 年中国 NO<sub>2</sub> 浓度年/季节平均变化

Figure 4-7 Annual and seasonal average changes of NO<sub>2</sub> concentration in China from 2015 to 2021

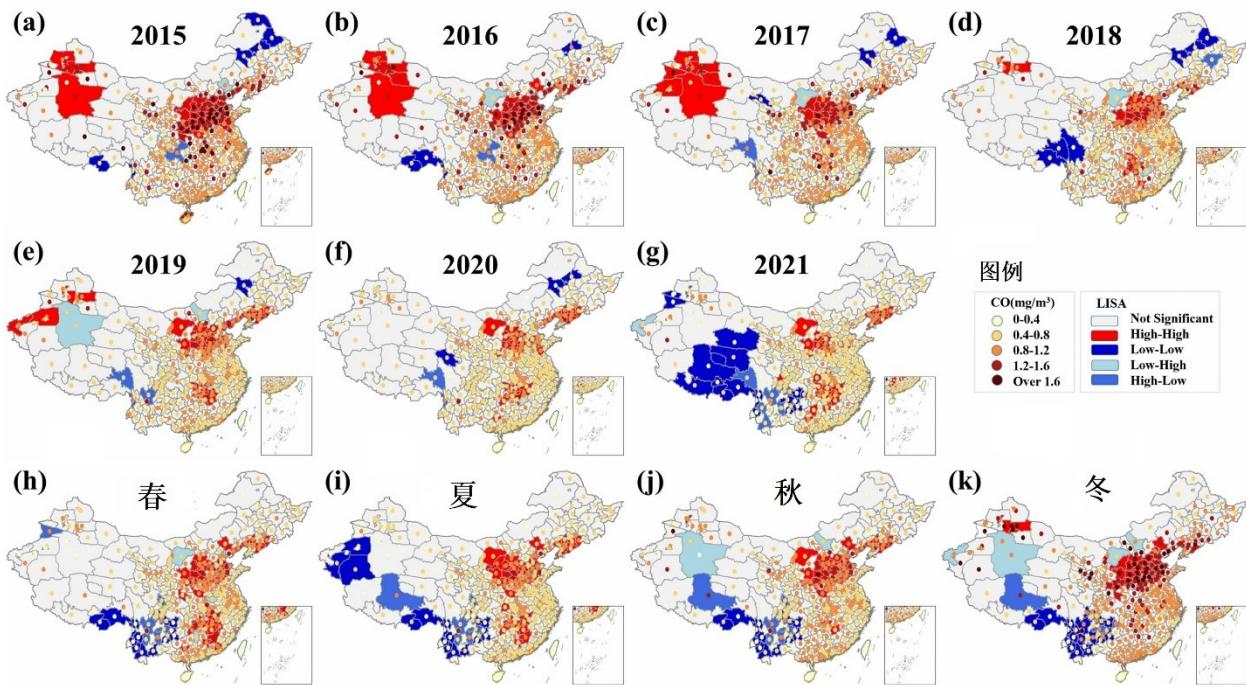


图 4-8 2015 - 2021 年中国 CO 浓度年/季节平均变化

Figure 4-8 Annual and seasonal average changes of Co concentration in China from 2015 to 2021

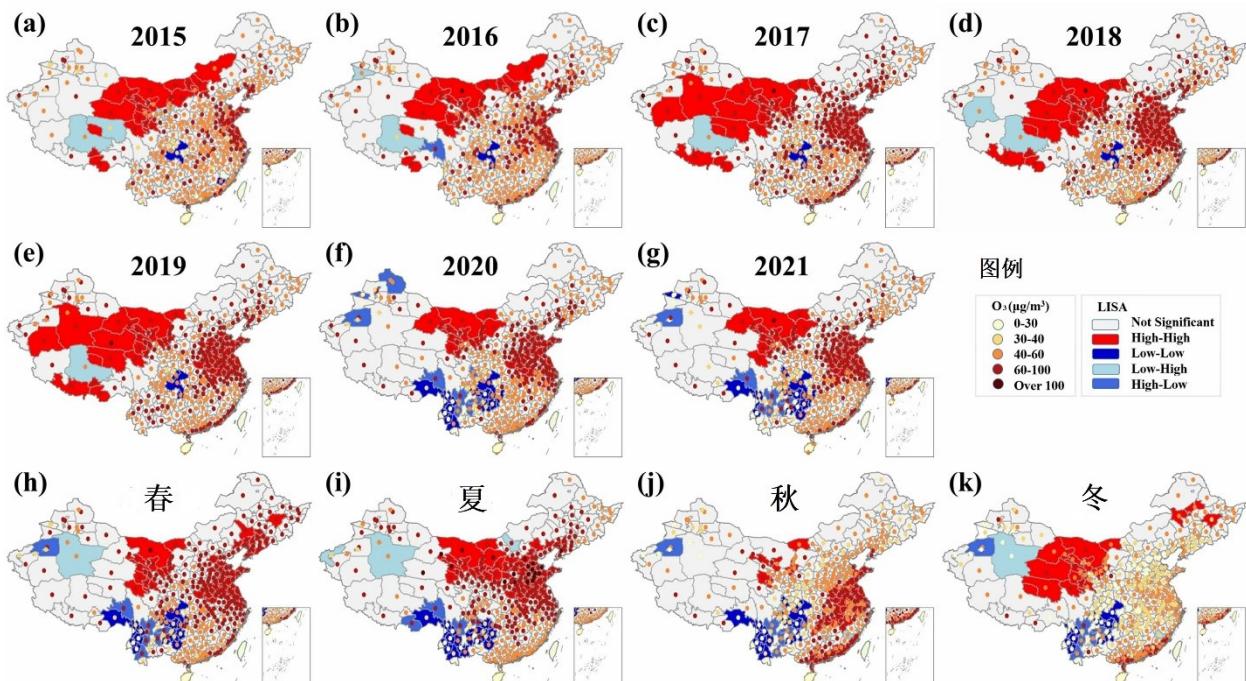


图 4-9 2015 - 2021 年中国 O3 浓度年/季节平均变化

Figure 4-9 Annual and Seasonal Average Changes of O3 Concentration in China from 2015 to 2021

O<sub>3</sub>浓度在春季和夏季在北方地区最高，在秋季和冬季在南方地区最高（图 4-9）。具体来说，夏季O<sub>3</sub>污染严重，主要分布在宁夏、内蒙古和青海。内蒙古的温差特征是春季气温突然升高，夏季风大、降水集中、夏季短而炎热，秋季气温骤降、霜冻早。青海位

于大气稀薄、气压低的高原上，氧分子光解率增加，促进  $O_3$  形成。虽然污染物浓度在冬季普遍增加，但随着清洁供暖措施的实施和冬季供暖影响的减少，污染物逐渐减少<sup>[71]</sup>。

刘等人分析了 2015 年至 2020 年夏季华北平原的盛行南风和东南风，发现风向加剧了北方地区的  $O_3$  污染<sup>[72]</sup>。同样，冬季盛行的北风和西北风加剧了南方地区的  $O_3$  污染，这与本研究的结果一致。夏季，宁夏、内蒙古和青海的  $O_3$  浓度较高，其中宁夏和陕西的煤炭排放、内蒙古和宁夏的车辆和生物质排放以及陕西和河南的粉尘排放是主要的污染源。 $O_3$  的高聚类位于内蒙古和甘肃省，浓度范围呈现不规则的增加和减少。夏季  $O_3$  浓度增加的原因应综合考虑大气垂直运动、化学反应和气象条件<sup>[70,73]</sup>。

从 2015 年到 2021 年，污染物浓度的空间格局没有明显变化。因此，研究京津冀地区及其周边城市以及内蒙古自治区的呼和浩特和包头等城市等高浓度地区的重点防治措施至关重要。由于城市空气污染物浓度和空间格局随季节而变化，因此加强管理措施以减少高污染季节和高浓度地区的污染源排放非常重要。在低浓度地区也应采取预防措施，以改善整体空气质量。基于上述分析，城市空气污染物表现出强烈的时空相关性。为了定量和比较相关程度，需要采用额外的时空分析方法来表征污染物之间的关系。

传统的相关算法在不考虑时间和空间位置影响的情况下，并没有充分揭示城市中各种污染物之间的相互关系<sup>[74,75]</sup>。为了解决这一局限性，本文采用了 OLS 回归和 GWR 技术来研究空间相关性。此外，本文利用 GTWR 来研究时空变化的影响，从而建立了污染物之间的时空关系。通过比较每月和每年的污染物趋势图（图 4-2 和图 4-3），观察到每月变化呈 U 形模式，表明与年度数据中观察到的近似线性趋势相比，随着时间的推移，污染物排放量呈非平稳行为。相比年度变化，污染物在月份变化较为显著。

与 OLS 和 GWR 模型相比，GTWR 模型考虑了时间和空间变化，提供了对污染物相关性的更全面理解（图 4-10）。总体而言，GTWR 模型展现出更高的解释能力。尽管 GTWR 模型和皮尔逊相关系数都显示了污染物之间存在强烈的相关性，但是皮尔逊相关系数未考虑时空因素，且不足以显示局部的污染物相关程度。与之相比，GTWR 模型揭示了时空变化对污染物相关性的影响，还能够进一步分析污染物在时空层面上的相关程度。

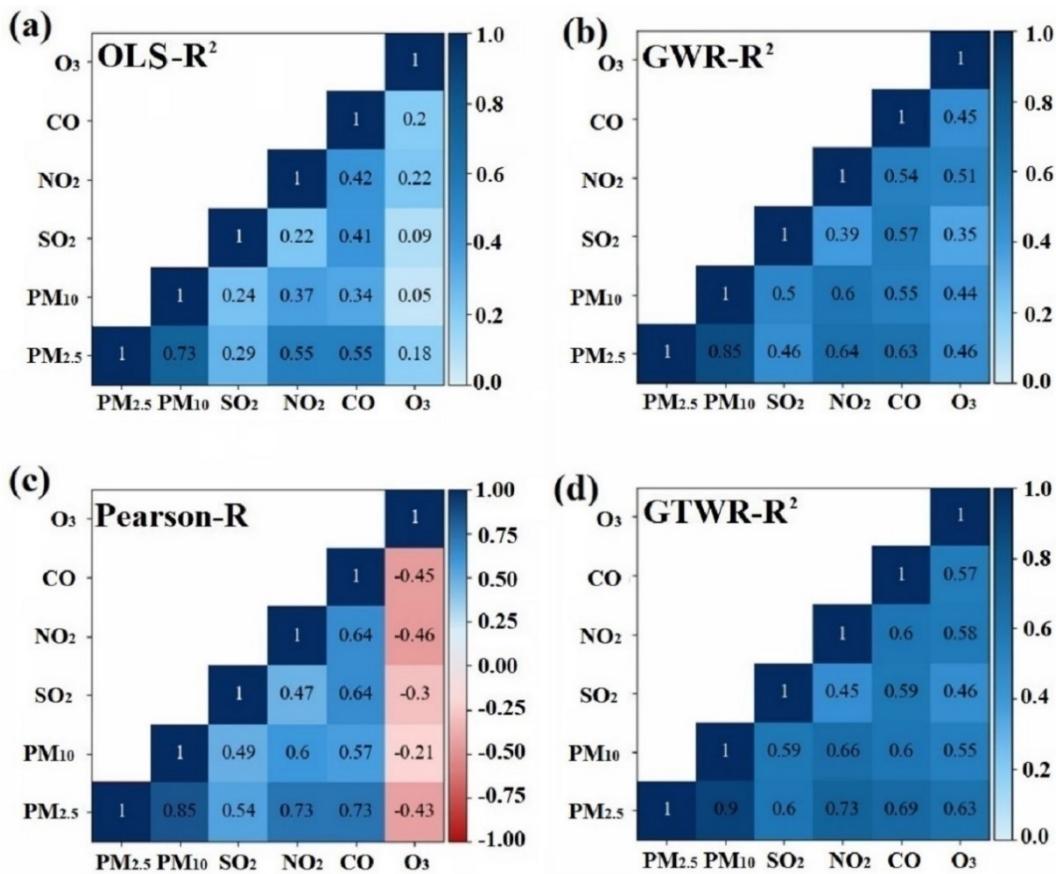


图 4-10 城市大气污染物的相关性。 (a) 表示 OLS-R<sup>2</sup>。 (b) 表示 GWR-R<sup>2</sup>。 (c) 代表 Pearson-R。 (d) 表示 GTWR-R<sup>2</sup>

Figure 4-10 Correlation of urban air pollutants

在本文的进一步研究中，使用 GTWR 模型分析了时空系数变化。具体的统计结果如图 4-11 所示，展示了六种污染物的系数变化。观察到的系数范围表明随着时间的推移不断变化。值得注意的是，研究观察到夏季和冬季的系数范围存在显著差异。为了探索这些季节之间的空间差异，本文对 1 月和 7 月进行了时空系数的空间映射（图 4-12）。结果表明，冬季和夏季的系数空间分布存在明显的差异。具体来说，黑河腾冲线上的污染物之间的相关性强度各不相同，其中 SO<sub>2</sub>、NO<sub>2</sub> 和颗粒物显示出更强的相关性。CO 与其他污染物之间的相关性在秦淮河线上的强度不同，季节变化不太显著。O<sub>3</sub> 与其他污染物之间的相关性随季节变化而显著变化。

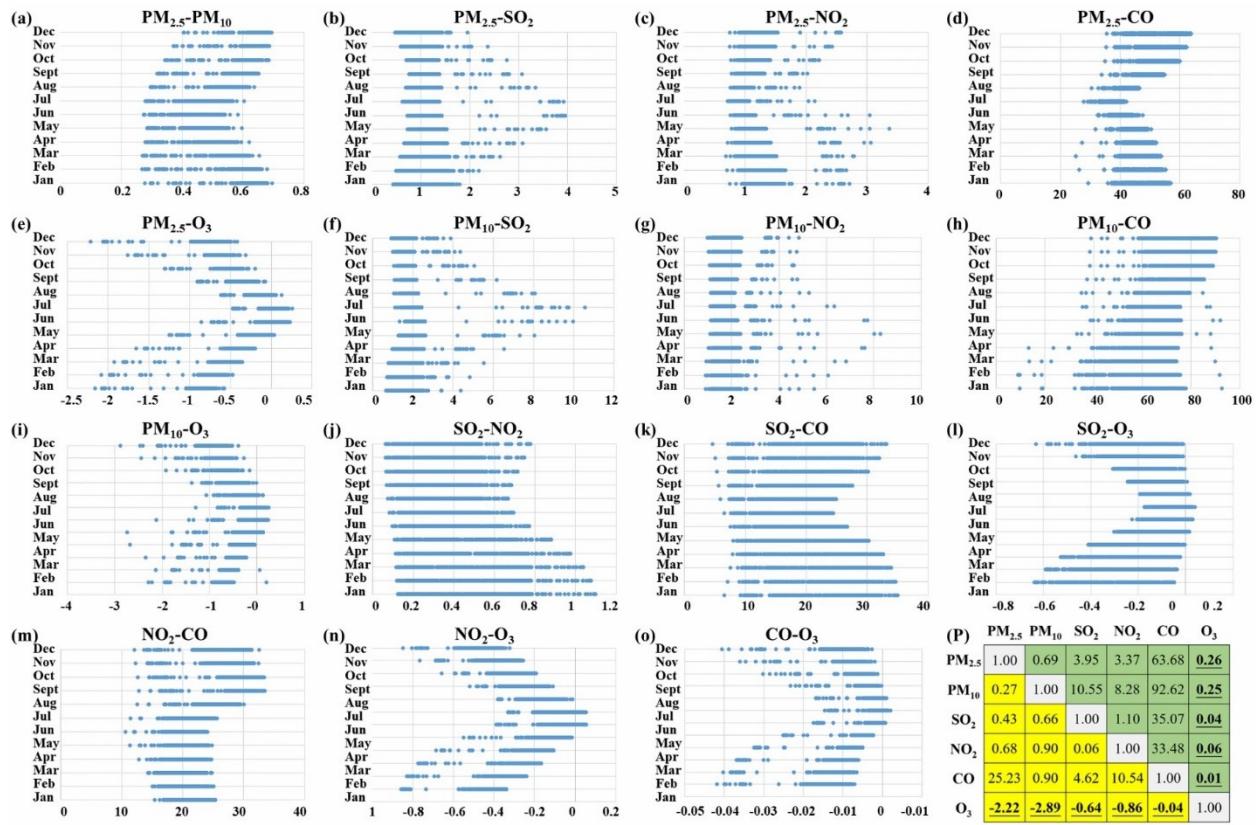


图 4-11 时空系数范围(P)，其中黄色背景表示最小系数值，绿色背景表示最大系数值。散点图(a)至(o)表示污染物水平的月时空系数值

Figure 4-11 Range of spatiotemporal coefficients

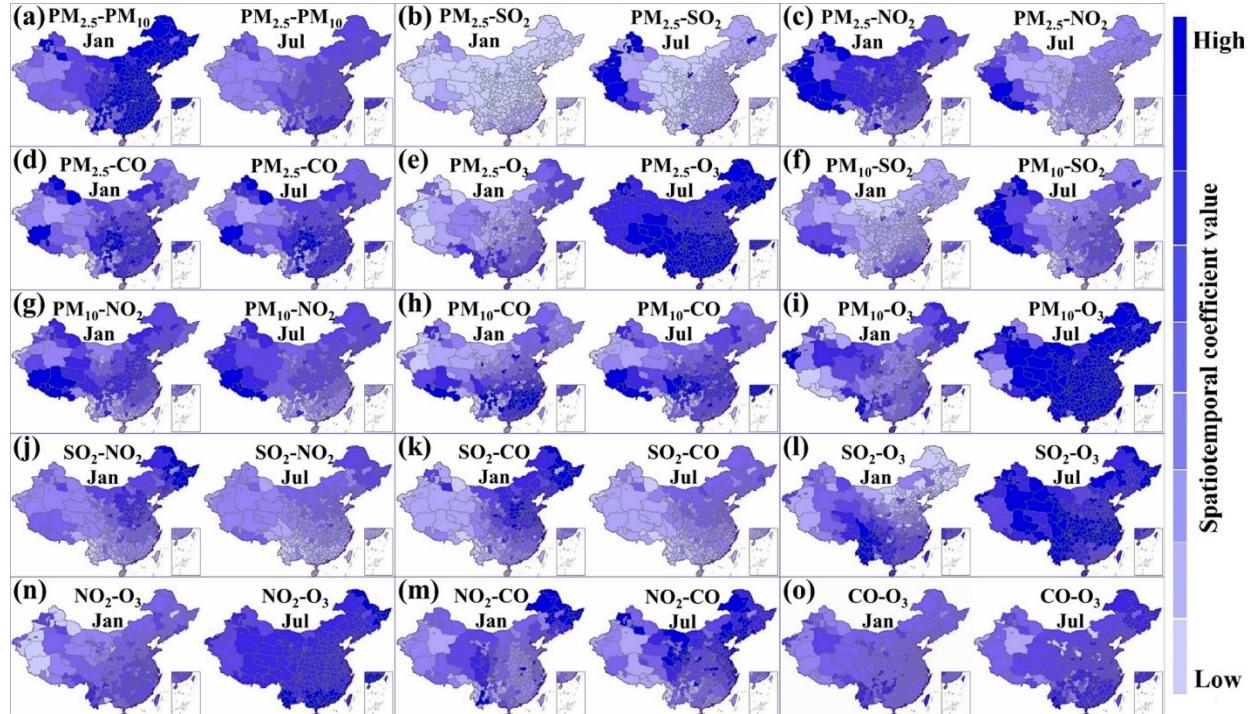


图 4-12 1月(冬季)和7月(夏季)时空系数空间映射

Figure 4-12 Spatial mapping of spatiotemporal coefficients for January (Winter) and July (Summer)

这些研究结果为中国城市空气污染问题提供了深刻的见解。研究成果表明，空气污

染是中国城市面临的一个重大环境问题，特别是 PM<sub>2.5</sub>、PM<sub>10</sub>颗粒物和 O<sub>3</sub>（图 4-4、图 4-5 和图 4-9）。这突显了在快速城市化和工业化过程中对空气质量管理的迫切需求。

空气质量管理应针对季节和地理位置对空气污染进行协同控制。研究表明，不同季节和地理位置的城市空气污染水平存在差异。冬季和夏季是污染最严重的季节，不同地区的污染水平也有所不同（图 4-4 至图 4-9）。因此，在制定空气污染控制策略时，必须考虑季节和地理因素的影响，并采取有针对性的措施。

协同管理是制定空气污染策略的重要方法，综合考虑各种因素，确保在控制一种污染物的同时，不会引发其他污染物的增加。研究表明，各种污染物的浓度普遍下降，特别是 PM<sub>2.5</sub>。然而，O<sub>3</sub>浓度有所增加<sup>[76]</sup>，可能是由于颗粒物污染减少导致 O<sub>3</sub>产生途径发生变化。这表明在解决空气污染问题时需要综合考虑不同污染物之间的相互关系（图 4-11 和图 4-12）。

PM<sub>2.5</sub>、PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub> 和 CO 表现出相似的时空变化格局，其中以 PM 颗粒物具有显著的时空相关性。与其他污染物相比，O<sub>3</sub>表现出明显的时空特征，但其时空变化与其他污染物相似。这些时空系数的空间分布格局解释了污染物相关性的时空非平稳性。然而，值得注意的是仅使用单一因变量进行回归建模，极大是忽略了污染物之间的相互作用。故应将这相关性高的污染物作为一个整体来分析，同时考虑它们的互联性和时空差异<sup>[25]</sup>。

## 4.2 空气复合污染物驱动因素研究应用分析(Study and Application Analysis of Driving Factors of Air Composite Pollutants)

由上述章节，PM<sub>2.5</sub> 和 PM<sub>10</sub> 在空气污染研究中通常具有较高的相关性，利用多因变量时空地理加权回归算法来验证两者是否存在相同的驱动成因，揭示其与其他自变量（例如环境因素、社会经济、人口密度等）之间的关系。这种模型可以通过引入时空权重函数来考虑不同区域和时间点的权重差异，从而更准确地估计不同因素对于 PM<sub>2.5</sub> 和 PM<sub>10</sub> 浓度的影响程度。

### 4.2.1 实验区域及数据概况(Overview of Experimental Area and Data)

4.1 章研究发现空气污染高聚集区域位于华北平原，华北平原位于中国东部，占据重要的地理位置，其纬度范围为北纬 32°至 40°，经度范围为东经 114°至 121°（图 4-13）。其北界延伸至燕山南麓，南界至大别山北侧，西侧倚靠太行山和伏牛山，东侧则直面渤海和黄海。华北平原横跨北京、天津、河北、山东、河南、安徽和江苏七个省和直辖市，

不仅是中国三大平原之一，更是中国人口最为稠密的平原地区。

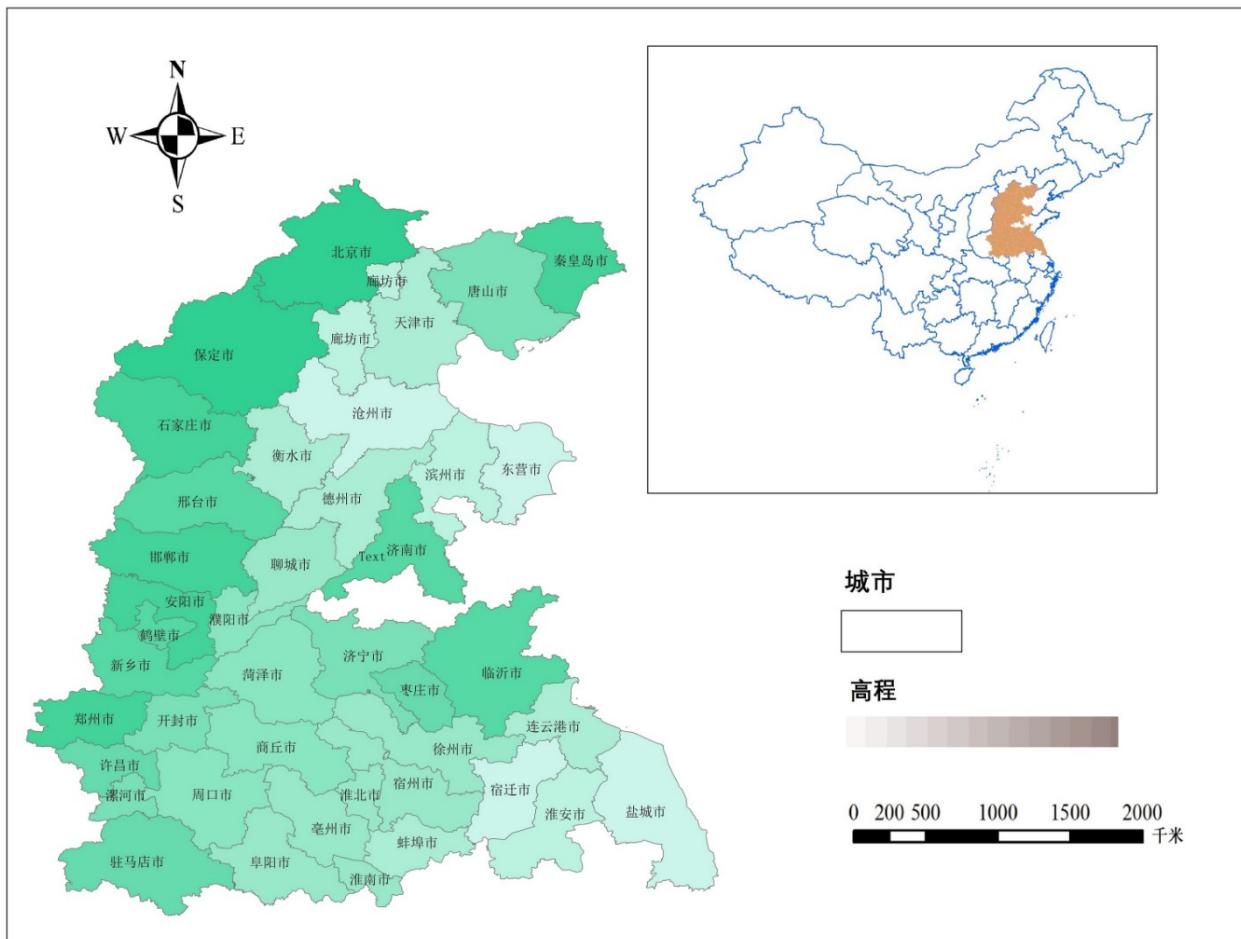


图 4-13 研究区域-华北平原  
Figure 4-13 Study area - North China Plain

#### 4.2.2 实验设计与结果分析(Experimental Design and Results Analysis)

本研究系统地搜集和整理了与华北平原空气污染成因密切相关的数据，并对这些数据进行了共线性诊断<sup>[45]</sup>。其中，符合研究要求的数据方差膨胀因子（VIF<10）<sup>[55]</sup>，主要包括人口社会经济发展数据（表 4-2），如人口密度（单位：万人/平方公里）、粉尘排放量（单位：万吨）、人均绿化面积（单位：万人/平方公里）以及年 GDP 等，这些数据均来源于中国统计年鉴。此外，还搜集了气象数据，包括气温（单位：°C）、降水量（单位：毫米）等，这些数据来源于国家气象科学数据中心。

本研究采用全局多因多元线性回归、多因变量地理加权回归和多因变量时空地理加权回归模型三种方法，对华北平原不同阶段的复合污染物 PM<sub>2.5</sub> 和 PM<sub>10</sub> 影响程度进行了定量分析。结果如表 4-2 所示，MGTWR 模型在不同阶段的 RSS 均小于其他模型，复合 R<sup>2</sup> 均大于 MMLR 和 MGWR 模型，表明 MGTWR 模型在拟合复合污染 PM 影响因素方面

的拟合度优于其他两种模型（表 4-3），进一步验证了 MGTWR 模型在处理时空非平稳性数据时的优越性和有效性。

表 4-2 变量范围、方差诊断结果  
Table 4-2 Variable range, variance diagnostic results

模型	最小值	最大值	平均值	VIF
降水量	387.68	1419.68	722.33	1.28
气温	11.148	21.77	15.28	1.51
人口密度	0.10	1089.28	736.74	1.15
GDP	0.07	4.02	0.41	1.19
人均绿化	1.60	133.54	48.61	1.22
粉尘排放	0.10	185.98	3.18	1.03

综合表 MGTWR 的系数范围，发现降雨与 PM 颗粒物主要呈负相关，大部分情况下降雨能有效降低 PM 浓度。气温与 PM 颗粒物关系则较为复杂，低温可能促进颗粒物沉降或扩散，而高温则可能利于颗粒物积聚。人口密度对 PM 颗粒物影响并不显著，表明其他因素可能更为重要。人均绿化面积与 PM 颗粒物关系呈 U 型，适量绿化有益于减少颗粒物，但过度绿化可能适得其反。粉尘排放量与 PM 颗粒物关系复杂，低排放有利于改善空气质量，而高排放则是 PM 浓度上升的重要因素。

通过对表 4-5 中各变量获取的  $F$  值和  $p$  值检验，研究发现截距项、降水量、气温、人均绿化和粉尘排放量在 PM2.5 和 PM10 的时空系数中表现出非平稳性，因为其  $p$  值均小于显著性水平 0.05，拒绝了平稳性假设。相反，人口密度与 GDP 的时空系数在该显著性水平下未被拒绝，可能呈现平稳性，表明该变量的时空关系相对稳定，不受观测时间和地理位置的显著影响或者在进行分析时未能捕捉到系统性的时空非平稳性，这需要进一步的对其研究。

表 4-3 MMLR、MGWR、MGTWR 模型的性能表现  
Table 4-3 Performance of MMLR, MGWR, and MGTWR Models

模型	RSS	R <sup>2</sup>	B	AICc	F <sub>1</sub>	p <sub>1</sub>	F <sub>2</sub>	p <sub>2</sub>
MMLR	42992.54	0.33/0.23	-	2315.98	-	-	-	-
MGWR	18422.40	0.71/0.69	0.1149	2251.20	0.18/0.19	0.00	-	-
MGTWR	5356.89	0.91/0.90	0.1134	2158.40	-	-	0.11/0.13	0.00
MGWR/ MMLR	58%	57%/66%	-	-	-	-	-	-
MGTWR/ MGWR	71%	21%/23%	-	-	-	-	-	-

RSS: 残差平方和; B: 带宽; R<sup>2</sup>: 多因变量复合 R<sup>2</sup>。

表 4-4 MGTWR 系数范围  
Table 4-4 MGTWR coefficient range

变量	最小值		下四分位		中位数		上四分位		最大值	
	PM <sub>2.5</sub>	PM <sub>10</sub>								
截距	-138.81	-269.1	51.50	112.26	85.52	161.76	108.64	219.27	257.58	550.88
降水量	-0.11	-0.20	-0.02	-0.05	-0.01	-0.03	0.00	-0.01	0.08	0.06
气温	-5.79	-13.60	-0.87	-3.90	0.10	-1.22	1.62	1.02	10.58	21.56
人口密度	-0.07	-0.13	-0.01	-0.01	0.00	0.00	0.01	0.02	0.05	0.11
GDP	-49.76	-131.5	-5.06	-11.10	-0.29	0.45	5.03	11.02	34.05	79.89
人均绿化	-0.63	-0.95	-0.09	-0.17	-0.04	-0.06	0.02	0.07	0.36	0.82
粉尘排放量	-3.55	-15.25	-0.14	-0.01	0.41	1.24	1.24	2.98	7.44	11.33

表 4-5 时空系数非平稳性诊断结果  
Table 4-5 Spatiotemporal coefficient non-stationarity diagnostic results

变量	PM <sub>2.5</sub>		PM <sub>10</sub>	
	F <sub>3</sub>	p <sub>3</sub>	F <sub>3</sub>	p <sub>3</sub>
截距	24.65	0.00	26.38	0.00
降水量	-3.26	0.00	-7.54	0.00
气温	3.58	0.00	-3.74	0.00
人口密度	0.13	<b>0.90</b>	-0.17	<b>0.87</b>
GDP	-0.40	<b>0.69</b>	-0.07	<b>0.94</b>
人均绿化	-4.64	0.00	-3.40	0.00
粉尘排放量	6.06	0.00	5.88	0.00

本研究进一步进行可视化分析空间上和时间位置的相关影响, 图 4-14 对比 PM<sub>2.5</sub> 和 PM<sub>10</sub> 与 GDP 相关性的强弱程度。从 2015 年到 2018 年, PM 和 GDP 相关性系数都呈现了先下降后上升的趋势。然而, 2015-2018 年相关性呈现逐年下降趋势, 至 2019 年趋于稳定状态。实验结果表明, 华北平原在城市发展的同时, 对环境保护方面也做了进一步发

展。空间上，高相关性主要集中在华北平原西南部，具体为安徽与河南城市群。

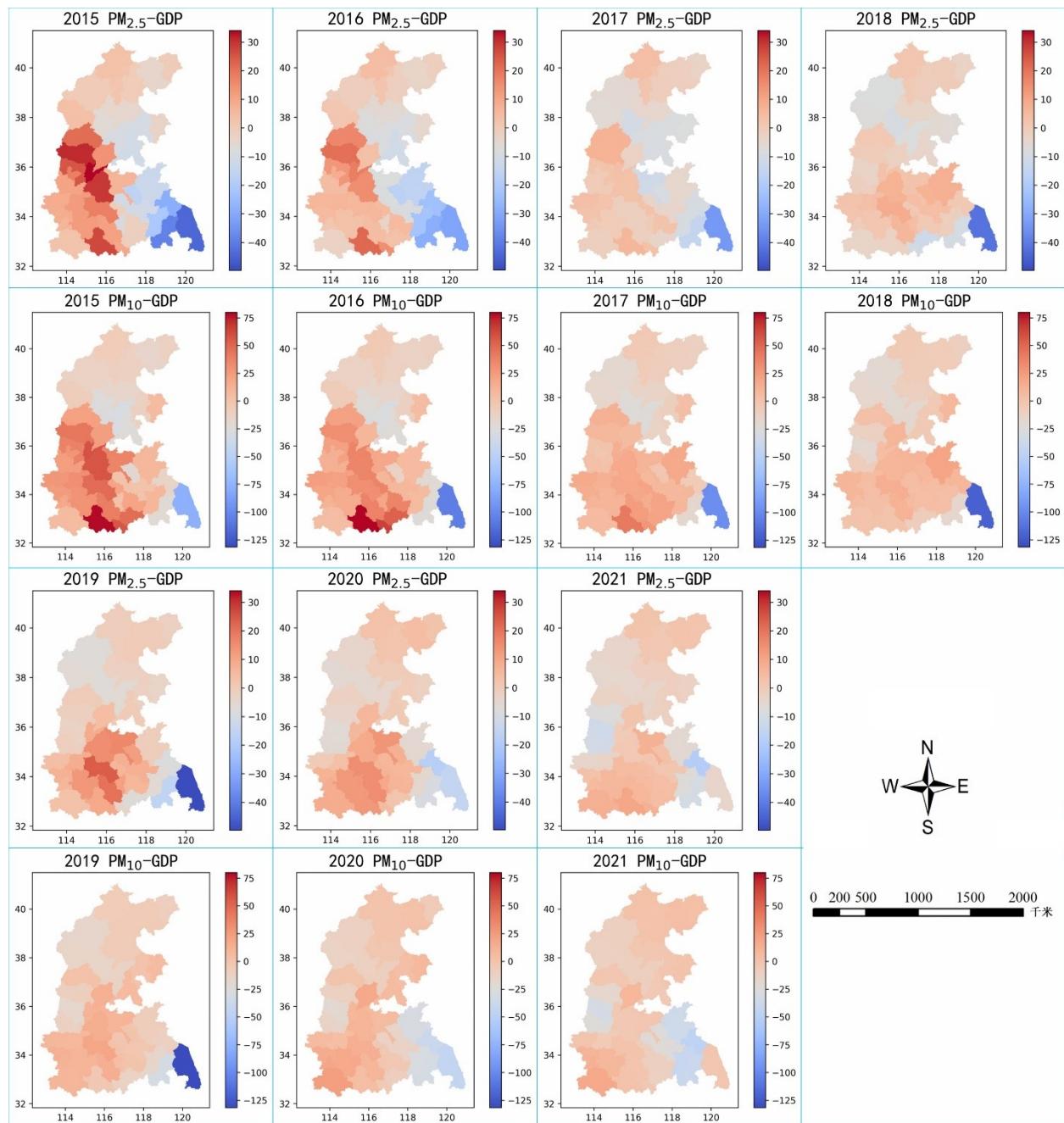


图 4-14 2015-2021 年国内生产总值 (GDP) 回归系数图

Figure 4-14 Regression coefficients of Gross Domestic Product (GDP) from 2015 to 2021

在 2015 年至 2021 年期间，粉尘排放量的回归系数呈现出一种先下降后上升的趋势（图 4-15）。这一变化趋势反映了在这一时期内，环境保护和经济发展之间的复杂关系。初始，随着经济的快速增长，颗粒物排放量也随之增加。然而，随着政府对环保问题的重视和一系列政策的出台，颗粒物排放量的回归系数开始逐渐下降。这表明，宏观策略在控制环境污染方面取得了一定的成效。高系数主要集中在安徽省一带，这与该地区的工业结构和能源使用有关。随着时间的推移，这些高系数区域由安徽省的西部逐步向东部

迁移。实验结果表明，随着经济的发展和产业结构的调整，污染源也在逐步转移。尽管颗粒物排放量的回归系数在早期呈现出下降趋势，但在后期却出现上升。

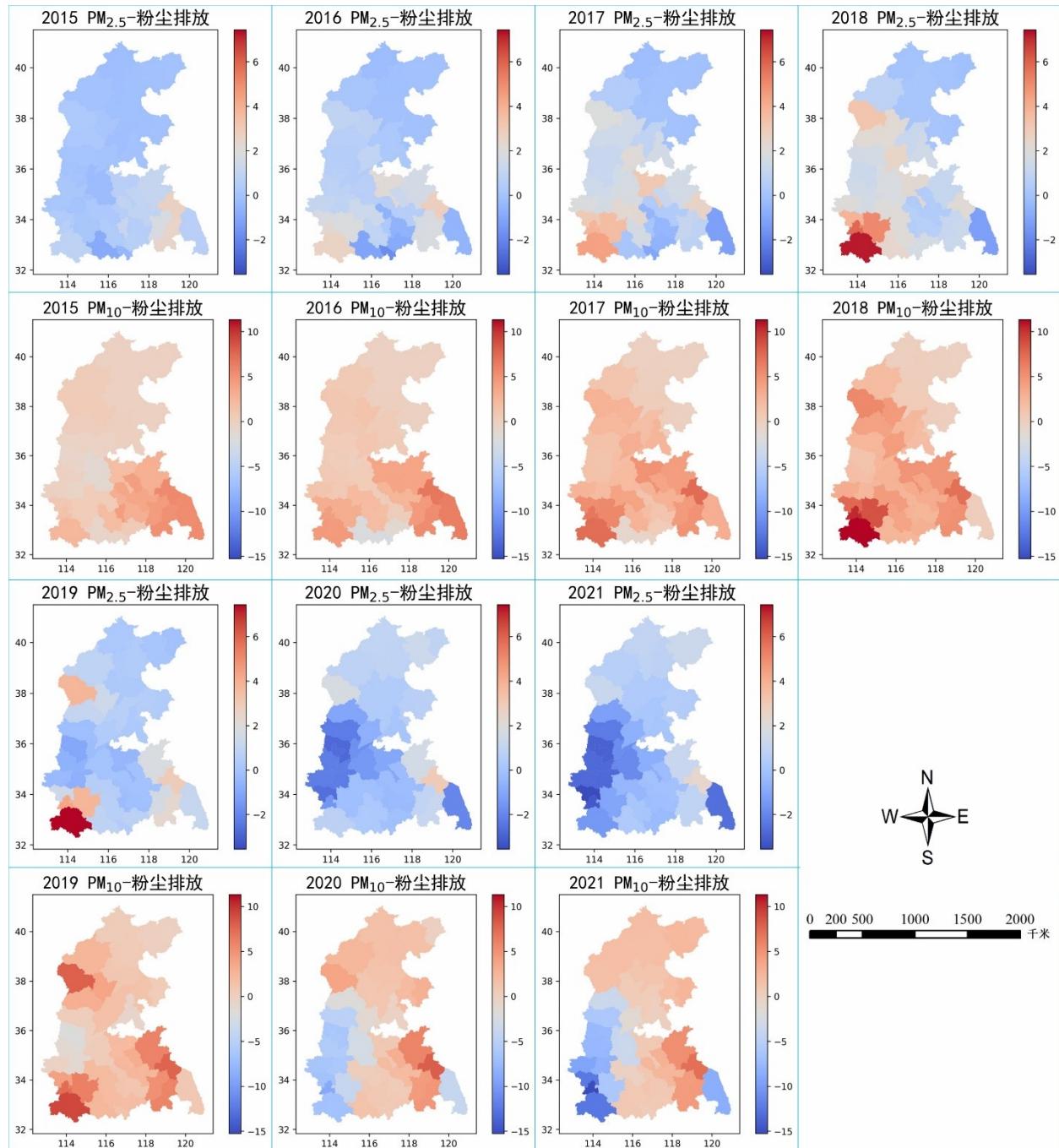


图 4-15 2015-2021 年粉尘排放量回归系数图

Figure 4-15 Regression coefficients of dust emissions from 2015 to 2021

人均绿化面积的系数在 2015 年到 2021 年期间呈现了先下降后上升的趋势（图 4-15）。具体来说，在 2015 年到 2016 年期间，人均绿化面积的系数有所下降，而在 2017 年到 2021 年期间，该系数又呈现上升趋势。人均绿化在空间上与 PM 呈现大部分的负相关关系，然而 2020 年和 2021 年河北省城市群呈现正相关关系。

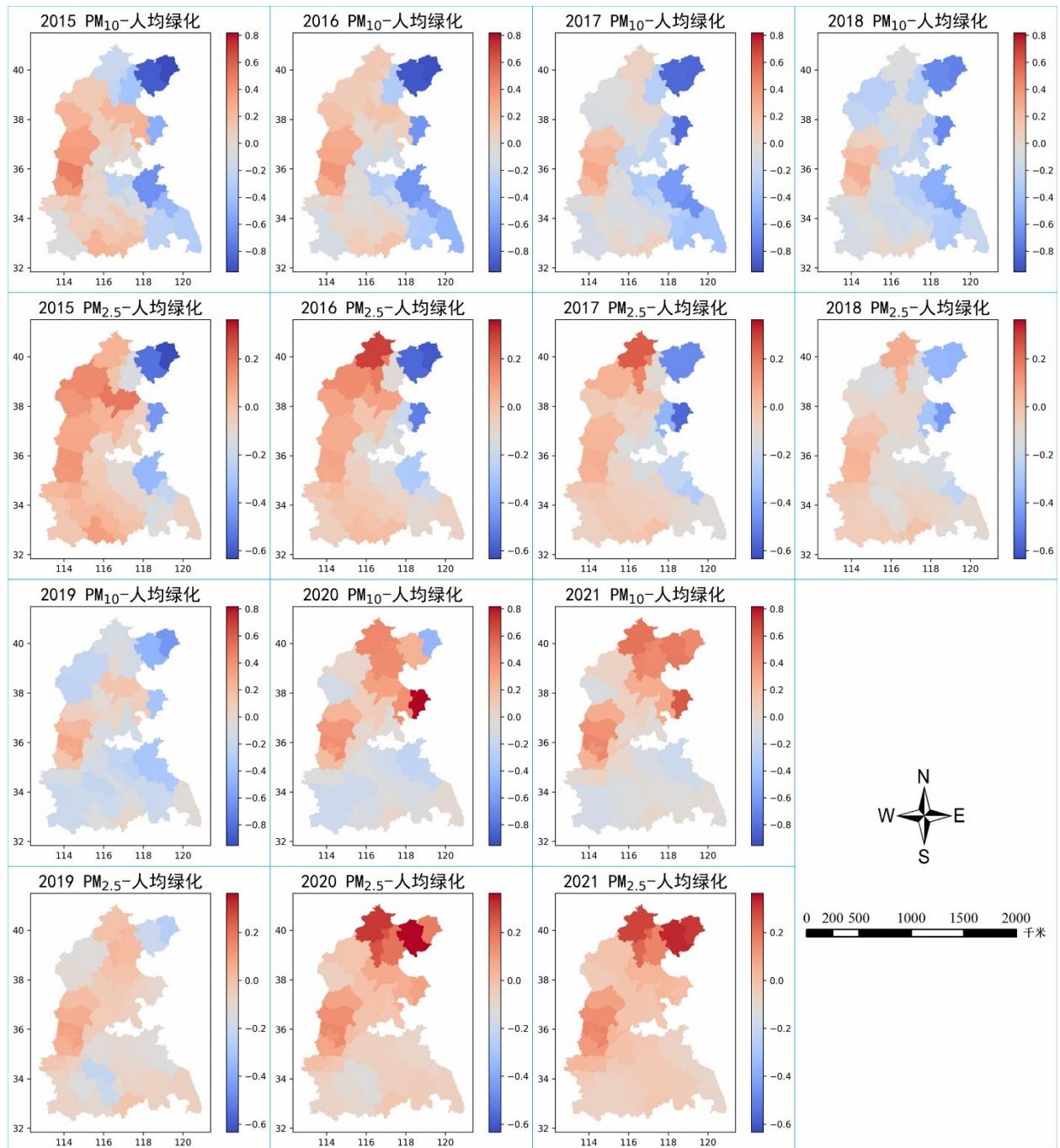


图 4-16 2015-2021 年人均绿化面积回归系数图

Figure 4-16 Regression coefficients of per capita green area from 2015 to 2021

从 2015 年至 2021 年的七年间，人口密集系数区域的变化呈现出一种先上升后下降的趋势（图 4-17）。具体来说，在 2015 年至 2018 年的这段时间内，河北省西北部的城市群显示出一种负相关性。然而，从 2019 年开始，这种关系逐渐转变为正相关。值得注意的是，尽管绿化通常被认为有助于改善空气质量，但在河北省的某些地区，绿化的增加却伴随着空气质量的下降。这引发了对于绿化效果的深入思考。为了更好地理解这一现象，需要进一步研究绿化的植物种类和配置方式是否科学合理。同时，也不能忽视城市中可

能存在的其他严重污染源，如工业排放和交通尾气等。这些污染源的存在可能会抵消绿化带来的正面效果，从而导致空气质量的下降。

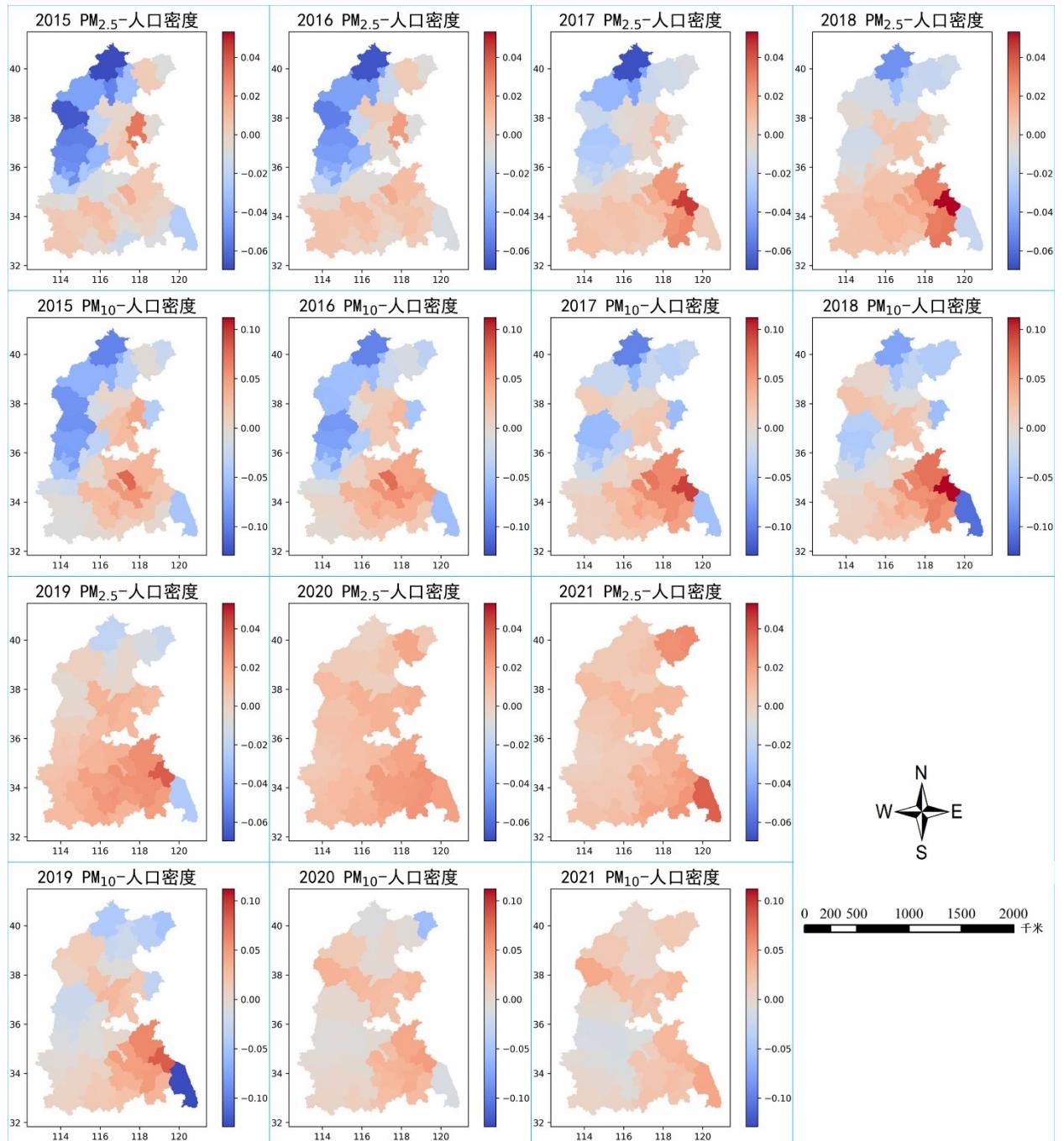


图 4-17 2015-2021 年人口密度回归系数图

Figure 4-17 Regression coefficients of population density from 2015 to 2021

在 2015 年到 2016 年期间，降水量与  $\text{PM}_{2.5}$ 、 $\text{PM}_{10}$  的相关性呈现出正相关，即降水量增加时， $\text{PM}_{2.5}$  和  $\text{PM}_{10}$  的浓度也相应增加（图 4-14）。这可能是因为降水量增加了空气中的湿度，有利于颗粒物的悬浮和扩散。然而，到了 2017 年，这种正相关关系开始逐渐减弱。随着时间的推移，到了 2019 年到 2021 年期间，降水量与  $\text{PM}_{2.5}$ 、 $\text{PM}_{10}$  的相关性逐渐

转为负相关，即降水量增加时，PM<sub>2.5</sub>和PM<sub>10</sub>的浓度反而有所下降。这可能是因为随着降水量的增加有助于清洗空气中的颗粒物，进一步降低PM<sub>2.5</sub>和PM<sub>10</sub>的浓度。在空间上，相关性也呈现出一定的变化。降水对PM<sub>2.5</sub>和PM<sub>10</sub>在整体上存在一定的差异以2016年-2018年较为显著。

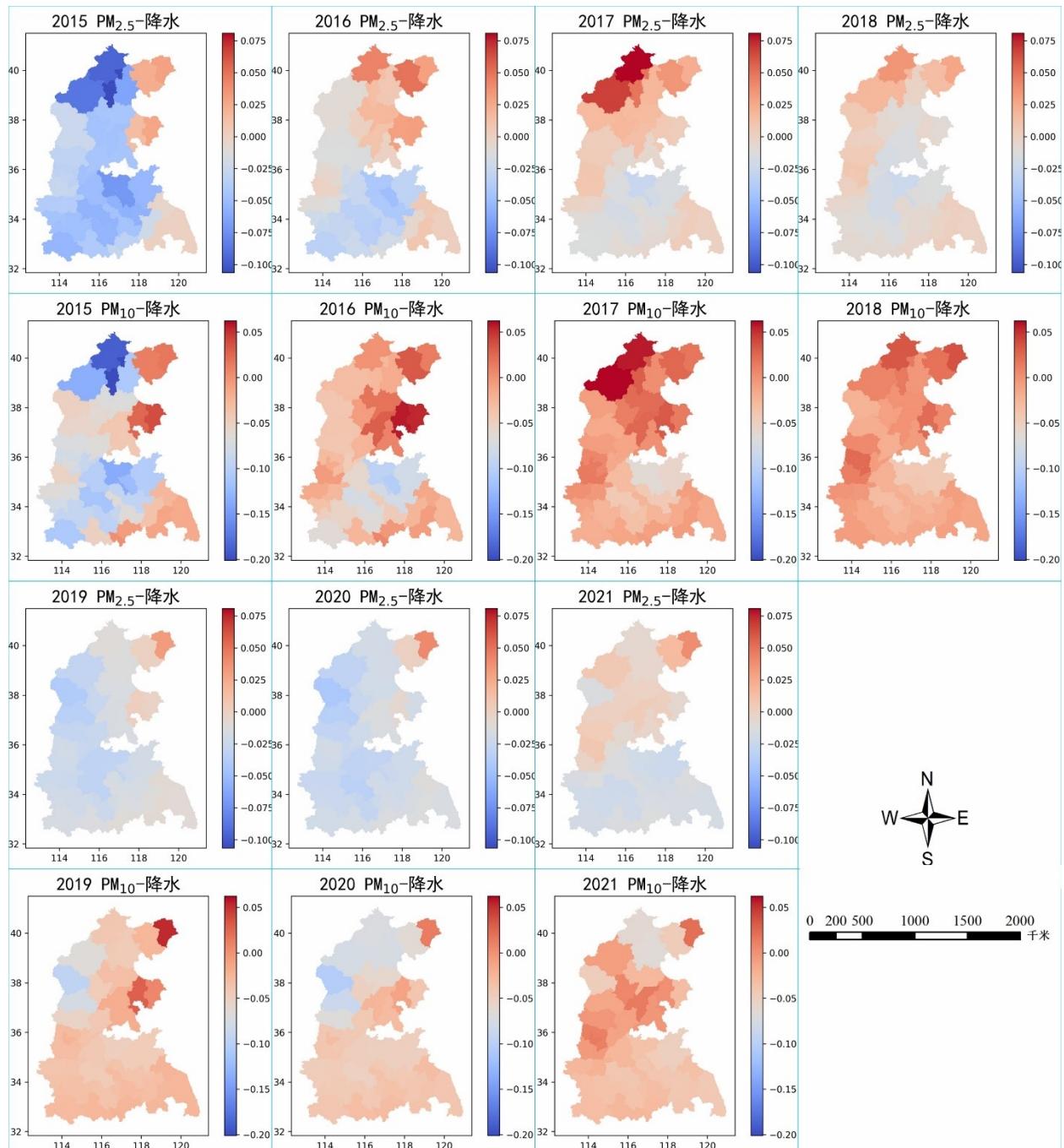


图 4-18 2015-2021 年降水回归系数图

Figure 4-18 Regression coefficients of precipitation from 2015 to 2021

从时间上看，随着时间的推移，气温与PM<sub>2.5</sub>、PM<sub>10</sub>的相关性呈现先增强后减弱的趋势（图4-19）。具体来说，在2015年至2018年期间，气温与PM<sub>2.5</sub>、PM<sub>10</sub>的相关性逐渐

增强，而在 2019 年至 2021 年期间，相关性则逐渐减弱。这可能与华北平原地区的气候变化和污染物排放情况有关。从空间上看，华北平原不同地区的气温与 PM<sub>2.5</sub>、PM<sub>10</sub> 的相关性也存在差异。

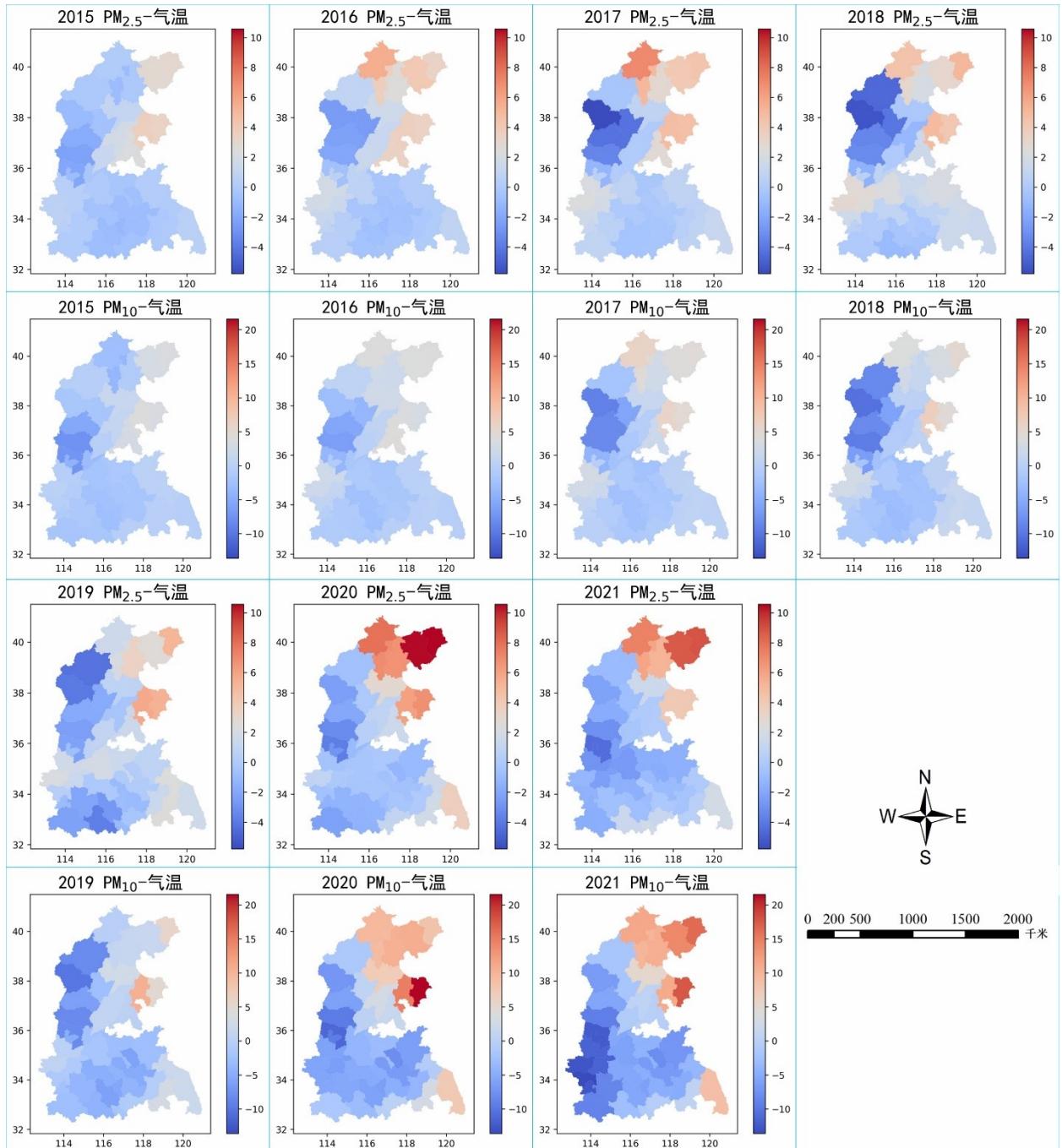


图 4-19 2015-2021 年气温回归系数图

Figure 4-19 Regression coefficients of temperature from 2015 to 2021

最后，MGTWR 模型通过对多个因变量的联合建模和分析，揭示了 PM<sub>2.5</sub> 和 PM<sub>10</sub> 之间的共同影响因子。研究结果显示，气温、人口密度、GDP、人均绿化以及粉尘排放量对颗粒物 PM 的共同影响在时空格局中表现得尤为显著。这些影响因素在不同的时间和空

间尺度上相互作用，共同塑造着  $PM_{2.5}$  和  $PM_{10}$  的浓度变化。值得注意的是，降水量对颗粒物 PM 的影响表现出一定的差异性。在不同的地区和时间点，降水量的影响程度和方式可能有所不同，这可能需要更多的影响因素来探索降水量对颗粒物 PM 影响的时空非平稳性。在整体范围上， $PM_{10}$  的浓度变化受驱动因子的影响大于  $PM_{2.5}$ 。

### 4.3 本章小结(Chapter Summary)

本章首先运用时空地理加权回归模型，深入剖析了中国城市空气污染中六种污染物间的相关性，特别是  $PM_{2.5}$  与  $PM_{10}$  之间的显著相关性，这充分满足了多因变量回归建模的分析条件。这种复合关系需要从整体视角来考量，而非单一变量的独立分析。因此，在理论基础上了，探测这种空气复合污染物驱动因素适用于本文提出的多银边时空地理加权回归模型。通过对比多因变量多元线性回归、多因变量地理加权回归和多因变量时空地理加权回归三种模型的实验结果，实验发现多因变量地理加权回归模型因考虑了空间非平稳性，即空间变化因素，其拟合效果决定系数较多因变量多元线性回归模型更为优越。而多因变量时空地理加权回归模型则进一步纳入了时空非平稳性，从而使其拟合效果决定系数超越了多因变量地理加权回归模型。经过时空非平稳性检验，确认了多因变量时空地理加权回归模型的应用必要性和有效性。这些实验结果不仅验证了模型的可靠性，更重要的是揭示了细微颗粒物之间共同驱动因子在局部上作用程度。这一发现对于深入解析城市空气污染的成因，以及构建高效治理措施具有启示价值。

## 5 总结与展望

### 5 Conclusion and Outlook

#### 5.1 研究总结(Research Summary)

多因变量时空地理加权回归是一种回归分析方法，可以同时考虑多个因变量的时空变化，并通过对时空数据进行加权处理来揭示因变量之间的相互关系和影响机制。下面是对该算法及其应用在复合空气污染物研究方面的总结：

(1) 算法提出：多因变量时空地理加权回归算法是在多因变量的多元线性回归算法和单因变量的时空地理加权回归方法的基础上发展而来。多因变量时空地理加权回归算法通过引入多个因变量，克服了传统方法只能处理单一因变量的局限性。其利用加权函数对空间和时间上的数据进行加权，以考虑空间邻近性和时间相关性，从而更准确地建模和分析多个因变量的时空关系。

(2) 应用在复合空气污染物研究中：多因变量时空地理加权回归方法在复合空气污染物研究中的应用取得了显著成果。通过将多因变量时空地理加权回归应用于分析复合污染物（如  $PM_{2.5}$  和  $PM_{10}$ ）的时空变化，可以揭示复合污染物之间的关联性、共同的驱动因素以及复合污染的形成机制。研究首先使用了时空地理加权回归方法对污染物浓度数据进行时空相关性分析，揭示中国城市群中不同污染物之间时空位置下的关联性。例如，研究可以发现  $PM_{2.5}$  和  $PM_{10}$  之间存在显著的时空相关性，并通过时空地理加权回归模型对复合污染物的影响因素进行解析。这些影响因素包括气象条件、环境因素和人类活动等共性因素。通过对这些因素进行加权回归分析，揭示了其对复合污染物时空变化的贡献程度和相互作用关系。

总而言之，多因变量时空地理加权回归算法在复合空气污染物研究中展现出了重要的应用价值。其通过同时考虑多个因变量的时空变化，揭示了复合污染物之间的关联性和影响机制，有望为改善空气质量、保护人类健康做出实质性贡献。

#### 5.2 研究展望(Research Outlook)

随着统计和计算技术的发展，未来的研究可以进一步优化多因变量时空地理加权回归模型，提高其预测精度和计算效率。同时，也可以探索其他多因变量时空分析方法<sup>[49,77]</sup>，如多因变量时空随机森林、多因变量时空神经网络等，以提供更多样化的视角和研究工

具。

尽管多因变量时空地理加权回归模型已经揭示了  $PM_{2.5}$  和  $PM_{10}$  的在时空位置上某些共同驱动因素，但未来的研究可以进一步精细化这些因素的识别。例如，可以进一步细分粉尘工业排放和交通排放的来源，以更准确地识别出对  $PM_{2.5}$  和  $PM_{10}$  影响最大的排放源。未来的研究环境空气质量领域具有广阔的前景和众多的挑战。随着环境问题的日益严重和公众对空气质量的关注度不断提高，对  $PM_{2.5}$  和  $PM_{10}$  以及其他空气污染物的研究将更加深入和全面。

在建立多因变量时空地理加权回归模型时，本文深入研究了变量之间的关系，发现人口密度和 GDP 在时空位置下的表现呈现出的非平稳性并不显著，这一发现对模型的准确性构成了挑战。具体来说，随时空的变化并不显著。为了应对这一挑战，需要在多因变量时空地理加权回归模型中进一步融合平稳性参数和非平稳参数，以提升模型的准确性。

综上所述，地理时空关系建模领域仍存在许多有待深入研究的问题。通过进一步的理论探索、数据质量改进、方法和技术创新以及跨学科合作，不断提升地理时空关系建模的能力和应用水平。

## 参考文献

- [1] Goodchild M F, Anselin L, Appelbaum R P, et al. Toward spatially integrated social science[J]. International Regional Science Review, 2000, 23(2): 139-159.
- [2] Anselin L. Spatial econometrics: methods and models[M]. 4. Springer Science & Business Media, 2013.
- [3] Lesage J, Pace R K. Introduction to spatial econometrics[M]. Chapman and Hall/CRC, 2009.
- [4] Fotheringham A S, Brunsdon C, Charlton M. Geographically weighted regression: the analysis of spatially varying relationships[M]. John Wiley & Sons, 2003.
- [5] Harris P. A simulation study on specifying a regression model for spatial data: Choosing between autocorrelation and heterogeneity effects[J]. Geographical Analysis, 2019, 51(2): 151-181.
- [6] Mcmillen D P. Issues in spatial data analysis[J]. Journal of Regional Science, 2010, 50(1): 119-141.
- [7] Tobler W R. A computer movie simulating urban growth in the Detroit region[J]. Economic geography, 1970, 46(1): 234-240.
- [8] Lee L F. Asymptotic distributions of quasi - maximum likelihood estimators for spatial autoregressive models[J]. Econometrica, 2004, 72(6): 1899-1925.
- [9] Zhu J, Huang H-C, Reyes P E. On selection of spatial linear models for lattice data[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2010, 72(3): 389-402.
- [10] Griffith D A. Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR)[J]. Environment and Planning A, 2008, 40(11): 2751-2769.
- [11] Myers D E. To be or not to be... stationary? That is the question[J]. Mathematical Geology, 1989, 21: 347-362.
- [12] Atkinson P M. Geographical information science: geocomputation and nonstationarity[J]. Progress in Physical Geography, 2001, 25(1): 111-122.
- [13] Lloyd C D. Local models for spatial analysis[M]. CRC press, 2010.
- [14] Gelfand A E, Kim H-J, Sirmans C, et al. Spatial modeling with spatially varying coefficient processes[J]. Journal of the American Statistical Association, 2003, 98(462): 387-396.
- [15] Brunsdon C, Fotheringham A S, Charlton M E. Geographically weighted regression: a method for exploring spatial nonstationarity[J]. Geographical analysis, 1996, 28(4): 281-298.
- [16] Huang B, Wu B, Barry M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices[J]. International Journal of Geographical Information Science, 2010, 24(3): 383-401.
- [17] Fotheringham A S, Crespo R, Yao J. Geographical and Temporal Weighted Regression (GTWR)[J]. Geographical Analysis, 2015, 47(4): 431-452.
- [18] 乔林, 张旖禾, 郑明辉, et al. 大气中高风险新污染物的筛查进展及展望[J]. 科学通报, 2024, 69(06): 659-671.

- [19] 任珂. 面向空气污染的时空多变量数据可视分析[D]. 东北师范大学, 2021.
- [20] 胡秋灵, 郭帅. 宁夏沿黄城市群大气复合污染物关联规则挖掘——以污染季 PM<sub>2.5</sub> 质量浓度为例[J]. 宁夏大学学报(自然科学版), 2021, 42(02): 219-225.
- [21] 邓思欣, 司徒淑娉, 周雪玲, et al. 2017 年佛山市大气复合污染特征及来源分析[J]. 环境监控与预警, 2021, 13(06): 43-50.
- [22] 高惠璇. 应用多元统计分析. 北京大学出版社, 2005.
- [23] Babcock C, Matney J, Finley A O, et al. Multivariate spatial regression models for predicting individual tree structure variables using LiDAR data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2012, 6(1): 6-14.
- [24] Liu H, Zhu X. Joint modeling of multiple crimes: A Bayesian spatial approach[J]. ISPRS International Journal of Geo-Information, 2017, 6(1): 16.
- [25] Chen V Y-J, Yang T-C, Jian H-L. Geographically Weighted Regression Modeling for Multiple Outcomes[J]. Annals of the American Association of Geographers, 2022, 112(5): 1278-1295.
- [26] Jiang F, Chen B, Li P, et al. Spatio-temporal evolution and influencing factors of synergizing the reduction of pollution and carbon emissions - Utilizing multi-source remote sensing data and GTWR model[J]. Environ Res, 2023, 229: 115775.
- [27] Xu S, Wang X, Zhu R, et al. Spatio-temporal effects of regional resilience construction on carbon emissions: Evidence from 30 Chinese provinces[J]. Sci Total Environ, 2023, 887: 164109.
- [28] 吴世朋, 张辉国, 胡锡健. 贝叶斯地理加权回归模型的稳健性研究[J]. 统计与决策, 2019, 35(18): 25-30.
- [29] 冯一凡, 李翅, 冯君明. 基于 GTWR 模型的济南都市圈生态系统服务价值对城市扩张时空响应[J]. 北京林业大学学报, 2024, 46(01): 104-118.
- [30] 尹庆民, 林茵茵. 基于 GTWR 模型的数字经济对经济高质量发展的时空效应研究[J]. 资源与产业, , 2023, 26(1): 1-14.
- [31] 黄依婷, 崔秦毓, 刘畅. 基于 GTWR 模型的广东省乡村聚落时空动态特征及驱动机理研究[C]. 2022/2023 中国城市规划年会, 2023: 10.
- [32] 孙莹莹. 中国中部六省省会城市碳排放和大气污染的协同治理可行性研究[D]. 湖北大学, 2023.
- [33] 刘佳明, 王洁. GTWR 模型联合地理探测的长三角地区 PM<sub>2.5</sub> 驱动因素研究[J]. 长江流域资源与环境, 2024, 33(01): 139-149.
- [34] 徐泽龙. 基于时空地理加权回归模型的长江经济带近地面 NO<sub>2</sub> 浓度估算与分析[D]. 江西师范大学, 2023.
- [35] 李佳琪. 时空克里格插值系统设计与实现[D]. 华中农业大学, 2022.
- [36] 顾昱骅. 地理时空大数据高效聚类方法研究[D]. 浙江大学, 2018.
- [37] 董霖. 浅谈多因变量的多元回归算法的研究[J]. 科学与财富, 2018, 001(024).
- [38] 岳珠. 多因变量线性回归中的联合影响分析[J]. 山西师大学报(自然科学版), 1993(02): 6-9.

- [39] Fattahi S, Ravandi S a H, Taheri S M. Two-way prediction of cotton yarn properties and fiber properties using multivariate multiple regression[J]. Journal of the Textile Institute, 2011, 102(10): 849-856.
- [40] Zhu R, Zou G, Zhang X. Model averaging for multivariate multiple regression models[J]. Statistics, 2017, 52(1): 205-227.
- [41] 王惠文, 孟洁. 多元线性回归的预测建模方法[J]. 北京航空航天大学学报, 2007(04): 500-504.
- [42] 韩汉鹏. 偏最小二乘法在回归设计多因变量建模中的应用及其优化[J]. 数理统计与管理, 2007, 26(2): 303-307.
- [43] 刘钦圣. 最小二乘问题计算方法[M]. 北京: 北京工业大学出版社, 1989.
- [44] 周洪伟. 正态性检验的几种常用的方法[J]. 南京晓庄学院学报, 2012, 28(03): 13-18.
- [45] 张凤莲. 多元线性回归中多重共线性问题的解决办法探讨[D]. 华南理工大学, 2010.
- [46] 黄婷婷. 解释变量组合对企业多目标影响效应比较[D]. 南京大学, 2018.
- [47] 杨自强. 用多因变量逐步回归作多类逐步判别[J]. 计算数学, 1979(03): 25-36.
- [48] Du Z, Wang Z, Wu S, et al. Geographically neural network weighted regression for the accurate estimation of spatial non-stationarity[J]. International Journal of Geographical Information Science, 2020, 34(7): 1353-1377.
- [49] Wu S, Wang Z, Du Z, et al. Geographically and temporally neural network weighted regression for modeling spatiotemporal non-stationary relationships[J]. International Journal of Geographical Information Science, 2021, 35(3): 582-608.
- [50] 赵阳阳. 顾及全局和局部时空非平稳差异的地理加权回归方法研究[D]. 辽宁工程技术大学, 2017.
- [51] 何秀丽. 多元线性模型与岭回归分析[D]. 华中科技大学, 2005.
- [52] 高延超. 基于多元回归分析——支持向量机的财务危机预警模型[J]. 商业经济, 2008(9S): 3.
- [53] 徐伟, 孙涛, 刘竹林. 关于多因变量综合线性回归的几点注记[J]. 统计与信息论坛, 2018, 33(05): 13-18.
- [54] Moran P A. The interpretation of statistical maps[J]. Journal of the Royal Statistical Society. Series B (Methodological), 1948, 10(2): 243-251.
- [55] 柴正媛. 时空卷积神经网络加权回归方法研究[D]. 江苏海洋大学, 2022.
- [56] 杨毅. 顾及时空非平稳性的地理加权回归方法研究[D]. 武汉大学, 2016.
- [57] Comber A, Brunsdon C, Charlton M, et al. A Route Map for Successful Applications of Geographically Weighted Regression[J]. Geographical Analysis, 2023, 55(1): 155-178.
- [58] Huang B, Wu B, Barry M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices[J]. International Journal of Geographical Information Science, 2010, 24(3): 383-401.
- [59] Harini S. Parameter estimation of Multivariate Geographically Weighted Regression model using matrix laboratory[C]. 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), 2012: 1-4.

- [60] Nakaya T, Fotheringham A S, Brunsdon C, et al. Geographically weighted Poisson regression for disease association mapping[J]. *Statistics in medicine*, 2005, 24(17): 2695-2717.
- [61] Farber S, Páez A. A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations[J]. *Journal of Geographical Systems*, 2007, 9: 371-396.
- [62] Johnson R A, Wichern D W. *Applied multivariate statistical analysis*. Springer, 2002.
- [63] Leung Y, Mei C-L, Zhang W-X. Statistical Tests for Spatial Nonstationarity Based on the Geographically Weighted Regression Model[J]. *Environment and Planning A: Economy and Space*, 2000, 32(1): 9-32.
- [64] 李杰, 翟亮, 桑会勇, et al. PM<sub>2.5</sub>浓度插值中不同空间插值方法对比[J]. *测绘科学*, 2016, 41(04): 50-54+101.
- [65] Rupakheti D, Yin X, Rupakheti M, et al. Spatio-temporal characteristics of air pollutants over Xinjiang, northwestern China[J]. *Environ Pollut*, 2021, 268(Pt A): 115907.
- [66] Liu L, Wang Q. Is the effect of human activity on air pollution linear or nonlinear? Evidence from Wuhan, China, under the COVID-19 lockdown[J]. *Cities*, 2022, 127.
- [67] Xue Y G, Wang L X, Zhang Y M, et al. Air pollution: A culprit of lung cancer[J]. *Journal of Hazardous Materials*, 2022, 434.
- [68] Zhao H, Gui K, Ma Y, et al. Effects of Different Aerosols on the Air Pollution and Their Relationship With Meteorological Parameters in North China Plain[J]. *Frontiers in Environmental Science*, 2022, 10.
- [69] Turap Y, Talifu D, Wang X, et al. Temporal distribution and source apportionment of PM2.5 chemical composition in Xinjiang, NW-China[J]. *Atmospheric Research*, 2019, 218: 257-268.
- [70] Chen J, Lu J, Ning J, et al. Characteristics and sources analysis of PM2.5 in a major industrial city of northern Xinjiang, China[J]. *SN Applied Sciences*, 2020, 2(4).
- [71] Wang J, Wang S, Xu X, et al. The diminishing effects of winter heating on air quality in northern China[J]. *J Environ Manage*, 2023, 325(Pt B): 116536.
- [72] Liu S, Zhao W, Li L, et al. Meteorological mechanisms of regional PM2.5 and O<sub>3</sub> transport in the North China Plain driven by the East Asian monsoon[J]. *Atmospheric Pollution Research*, 2023, 14(1).
- [73] Zhou M, Hu T, Zhang W, et al. COVID-19 pandemic: impacts on air quality and economy before, during and after lockdown in China in 2020[J]. *Environ Technol*, 2022, 44(20): 1-11.
- [74] Kuerban M, Waili Y, Fan F, et al. Spatio-temporal patterns of air pollution in China from 2015 to 2018 and implications for health risks[J]. *Environ Pollut*, 2020, 258: 113659.
- [75] Zhao C, Sun Y, Zhong Y, et al. Spatio-temporal analysis of urban air pollutants throughout China during 2014-2019[J]. *Air Qual Atmos Health*, 2021, 14(10): 1619-1632.
- [76] Wei J, Li Z, Li K, et al. Full-coverage mapping and spatiotemporal variations of ground-level ozone (O<sub>3</sub>) pollution from 2013 to 2020 across China[J]. *Remote Sensing of Environment*, 2022, 270.
- [77] Dai Z, Wu S, Wang Y, et al. Geographically convolutional neural network weighted regression: a method

for modeling spatially non-stationary relationships based on a global spatial proximity grid[J].  
International Journal of Geographical Information Science, 2022, 36(11): 1-22.

# 作者简历

## 一、基本情况

姓名：杨旭初；性别：男；民族：汉；出生年月：1994-07-20；籍贯：江苏省淮安市

## 二、学术论文

1\*\*. Yang, X., Yang, Y., Xu, S., Karimian, H., Zhao, Y., Jin, L., Xu, Y., Qi, Y., 2024. Unveiling the air pollution tapestry in China: A comprehensive assessment of spatiotemporal variations through geographically and temporally weighted regression. Atmospheric Pollution Research 15.

## 三、获奖情况

1\*\*. 出血性脑卒中临床智能诊疗建模. 中国研究生创新实践系列大赛“华为杯”第二十届中国研究生数学建模竞赛，二等奖

## 学位论文原创性声明

本人郑重声明：所呈交的学位论文《多因变量时空地理加权回归模型研究及其应用》，是本人在导师指导下，在江苏海洋大学攻读学位期间进行的研究工作所取得的成果。据我所知，除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签字：

2024年 05月 29日

# 学位论文数据集

关键词*	密级*	中图分类号*	UDC	论文资助
MGTWR; 复合空 气污染物; 时空非平 稳	公开	P208	519.842	
学位授予单位名称*	学位授予单位代码*	学位类别*		学位级别*
江苏海洋大学	11641	工程		硕士
论文题名*	并列题名*			论文语种*
多因变量时空地理加权回归模型 研究及其应用	Research and Application of Multivariate Geographically and Temporally Weighted Regression Model			中文
作者姓名*	杨旭初	学号*		2021220444
培养单位名称*	培养单位代码*	培养单位地址		邮编
江苏海洋大学	11641	江苏省连云港市		222005
学科专业*	研究方向*	学制*		学位授予年*
土木水利	时空数据挖掘分析	3 年		2024 年
论文提交日期*		2024 年 5 月		
导师姓名*	杨毅 徐胜华	职称*		副教授 研究员
评阅人	答辩委员会主席*	答辩委员会成员		
电子版论文提交格式 文本(√) 图像( ) 视频( ) 音频( ) 多媒体( ) 其他( )				
推荐格式: application/msword; application/pdf				
电子版论文出版(发布)者	电子版论文出版(发布)地	权限声明		
论文总页数*	77			
注: 共 33 项, 其中带*为必填数据, 共 22 项。				