编号：B460

# 基于地理加权回归模型的京津冀地区碳排放相关指标体系构建与影响因素分析

论文题目：基于地理加权回归模型的京津冀地区碳排放相关指标体系构建与影响因素分析

参赛学校：中央民族大学

参赛成员（作者）：田玲玲、蔡奕杉、梁桂宇

指导老师：苏宇楠

# 目录

# 表目录



# 图目录

# 摘要

2019 年，我国的碳排放量仍居于全球榜的首位，减排的国际压力仍然巨大，如何进一步落实节能减排十分关键。京津冀协同发展是党中央在新的历史条件下做出的重大决策，是习总书记亲自指导的重大国家战略，因此本文选取京津冀地区为研究对象，探讨区域碳排放的空间效应，寻找碳排放影响因素，贯彻落实节能减排。

本文构建京津冀三地协同发展统计监测指标体系，该指标体系由四个层次构成，通过国家数据收集了京津冀三地 2008～2017 年的碳排放相关面板数据，选取 5 个指标作为区域排碳因素进行实证分析；考虑到碳排放溢出效应，本文还将山东、河南、江苏等其他周边 10 个省份与京津冀区域碳排放情况进行对比分析，发现这 13 个省市中，只有京津冀地区碳排放量呈下降趋势。

由于京津冀地区存在空间上的联系与差异，存在城市间排碳的空间规律，因此采用本文使用 Moran’s I 指数进行空间自相关检验，根据 LISA 显著性聚集图探究京津冀及周边地区的排碳溢出效应；同时采用地理加权模型探究京津冀地区的人均 GDP、人均民用汽车、城市化水平、总人口、第二产业占比探究碳排放量的影响；并选取高斯核函数为权函数，利用交叉验证法确定窗宽，还讨论了地理加权模型中的估计推导及检验问题。

实证分析得出的主要结论有：天津与河北作为能源消耗型城市，人均民用汽车量、第二产业占比与城市化水平均对排碳有明显的促进作用；北京作为科技人才聚集型城市，人均 GDP 与城市化水平反而成为保护环境的因素；而总人口对京津冀地区排碳的影响并不大，最终根据结论提出五个方面的建议。



**关键字：**指标体系；京津冀；碳排放；面板数据；地理加权模型

# Abstract

In 2019, China's carbon emissions still ranked first in the world and international pressure on China to cut its carbon emissions remains intense. So it’s really important to figure out how to further implement energy conservation and emission reduction. The coordinated development of Beijing-Tianjin-Hebei region is a major decision made by the Party Central Committee under the new historical conditions, and it’s also a major national strategy which is planned, and promoted by General Secretary Xi. Therefore, this paper selects the Beijing-Tianjin-Hebei region as the research object to explore the spatial effect of regional carbon emissions and find the influencing factors of carbon emissions and also implement energy conservation and emission reduction.

In this paper, we construct a statistical monitoring index system for the coordinated development of Beijing-Tianjin – Hebei region, which is composed of four levels. Based on the national data, we collected the panel data of carbon emissions in Beijing-Tianjin – Hebei region from 2008 to 2017, and selected five indicators as the regional carbon emission factors for empirical analysis. Considering the spillover effect of carbon emissions, we also compared the carbon emissions of 10 neighboring provinces, including Shandong, Henan and Jiangsu, with the Beijing-Tianjin-Hebei region, and found that among the 13 provinces and cities, only the Beijing-Tianjin-Hebei region showed a downward trend in carbon emissions.

Since there are spatial connections and differences in Beijing-Tianjin-Hebei region, as well as spatial rules of carbon emission between cities, Moran's I index was used in this paper to conduct spatial autocorrelation test. According to LISA significance cluster map, we explored the carbon spillover effect both in

Beijing-Tianjin-Hebei region and in its surrounding areas. Next, a geographically weighted model was used to explore the impact of per capita GDP, per capita civil vehicles, urbanization level, total population, and the proportion of the secondary industry in the Beijing-Tianjin-Hebei region. The Gaussian kernel function is selected as the weight function, and the window width is determined by cross-validation method in this model. The derivation and verification of the estimation in the geographically weighted model are also discussed.

According to the empirical analysis, we get the following conclusions. First, Tianjin and Hebei are energy-consuming cities, so in these two areas, the per capita number of civil vehicles, the proportion of the secondary industry and the level of urbanization all have significant promoting effects on carbon emission. Secondly, Beijing, as a city of scientific and technological talents, its per capita GDP and urbanization level have become environmental protection factors. However, the total population has no significant impact on the carbon emission in the Beijing-Tianjin-Hebei region. Finally, five suggestions are put forward based on the conclusions in this paper.

**Key words:** Indicator System; Beijing-Tianjin-Hebei; Carbon Emission; Panel data; Geographically Weighted Model

# 前言

2019 年，中央全面深化改革委员会第十一次通过《关于构建更加完善的要素市场化配置体制机制的意见》，次年，中共中央、国务院明确将数据作为一种新型生产要素纳入《意见》，此举充分肯定了数据这一新型要素对于完善社会主义市场经济体制起着重要作用，以及数据对劳动力要素、技术要素等其他要素活力的促进作用。

改革开放进入了全新阶段，在其他要素持续发挥作用的条件下，加快培育发展数据要素市场，使数据要素在新时代展现焕发社会创造力与市场活力的作用，成为刺激经济高质量发展的新动能。从不同角度、以不同方法利用数据促进经济高速、高质量发展，深入探讨数据新动能的统计测度，更好地挖掘数据要素潜在规律，体现数据资源价值。

近几年，在习总书记的亲自推动下，京津冀协同发展战略取得了显著的成果，为该区域及其周边人民带来了重大利好。构建具有代表性的区域协同发展指标体系，量化协同发展为区域经济、社会和环境带来的利好，利用数据要素反映京津冀地区的多方面发展，并深究影响区域发展的重点因素，针对不同影响因素给出多方面的改进建议显得尤为关键。

## 一、问题的提出

### （一）研究背景

在人类的经济活动中，以化石能源为主排放的二氧化碳是导致温室效应、全球变暖的主要原因。2007 年我国的碳排放量排名世界第一，中国早已成为名副其实的“ 排碳 ”大国。区域经济发展是否协同会直接影响到国家经济发展的质量，城市发展过程中工业化推进导致的碳排放一直是国民关注的热点问题，而京津冀三地是北方的经济中心，在我国经济发展队伍中占据引领地位。

京津冀协同发展，是党中央在新的历史条件下做出的重要决策，是习总书记亲自谋划、亲自决策、亲自推动的重大国家战略。在京津冀协同发展上升为国家战略的这些年来，京津冀“ 一张图 ”规划、“ 一盘棋 ”建设、“ 一体化 ”发展，三地优势互补、良性互动、共赢发展，协同发展成果日渐显现。

长久以来，以京津冀地区为典型的空气污染一直困扰着当地及周边区域的居民，当地也出台了许多政策来改善空气质量。对于发展中国家来说，城市化建设带来的能源消费量与碳排放量庞大是长久的特征。我国的交通运输业、建筑业与工业是碳排放的最主要行业，同时也是我国关键减排行业。寻找碳排放影响因素，贯彻落实节能减排，同时充分保障经济快速稳步发展，进一步掌握可持续发展要领是研究的重点。

### （二）研究意义

2019 年，我国的碳排放量仍居于全球榜的首位，中国碳减排的国际压力[1]仍然巨大，如何进一步落实节能减排十分重要。京津冀协同发展战略正处于关键时期，以交通一体化为代表可持续发展决策取得了可喜成果，不仅给三地经济发展与居民出行带来重大利好，更为节能减排做出了重要贡献。

在环境污染亟需改善的背景下，城市交通造成的污染是不容忽视的，交通运输业已经成为我国减排降碳的重点行业[2]。为深入研究影响京津冀地区碳排放的

多种影响因素，本文结合已有的研究成果，将交通规模纳入碳排放的影响因素中，深入分析京津冀交通一体化对三地碳排放影响程度，可以为地区经济与节能减排协调发展[3]提出更具体的建议。

本文收集了京津冀三地 2008～2017 年的碳排放等相关的面板数据，结合地理加权模型，分析探讨京津冀地区碳排放的影响因素及其空间效应。通常情况下认为，地理位置相邻近的城市，其经济发展的相关性越显著。本文选取与经济建设密切相关的碳排放量面板数据，分析京津冀三地碳排放量在空间上的关联程度，探讨如何形成三地协同降低碳排放，为京津冀地区交通一体化健康发展提出建议。

为了缓解交通压力，更为了节能降碳，京津冀三地坚持落实限号出行政策，为保护环境做出了很大贡献。在京津冀联系越来越紧密的情况下，本文结合人口规模、产业结构、交通规模等因素，挖掘每个变量背后隐藏的可能对降碳有益的信息，希望以京津冀地区为代表形成连锁减排效应，为其他地区节能减排做出表率作用，进而缓解我国降碳压力，为改善全球温室效应做出贡献。

## （三）文献综述

碳排放及其影响因素一直是国内外的研究重点，多数城市的碳排放量会随着经济增长而增加，产业结构与城市的碳排放量也有正向的相关关系。二氧化碳排量的影响因素众多，诸多学者做出了不同因素作用于排碳量的说明，Ehrlich [4]、Holden 和 Commoner [5] 等学者提出了 IPAT 模型，在该模型中，变量 I 是指人类活动对环境的影响，变量 P 指人口规模，变量 A 指富裕程度，变量 T 指技术水平。也有学者指出，土地利用类型、产业结构和城市绿化等方面[6,7,8]影响大城市碳排放量。另外，科技因素[9]与交通因素[10]等也会影响到城市的碳排放。京津冀三地在地理位置非常邻近，在经济关系上联系非常紧密，苑清敏、张宝荣和李健[11]通过实证研究，得出京津冀三地工业碳排放强度与多个因素存在长期均衡关系的结论。为节能减排制定政策建议时，考虑影响排碳量的因素应当越充分越好，

但在模型构建方面，纳入过多的自变量可能会因多重共线性导致本该显著的变量表现地不显著，进而无法科学地解释诸多自变量如何对排碳量产生影响。

地理学家指出空间上存在相互作用机制，即某地区的碳排放会被邻域人口规模、产业类型、贸易结构和自然环境等因素影响[12]。杨世杰[13]通过实证得出，在我国省域背景下，能源消耗产生的二氧化碳排放在空间上表现出较强的正相关性。SU Wensong[14]等学者的研究结果表明，随着时间的推移，中国城市级别碳排放量的区域差异有所减少，这与明显的空间溢出效应和高排放量的空间集聚相吻合。李力、洪雪飞[15]指出，由于实证表明我国省市的碳排放量、技术等诸多现象存在空间溢出效应，所以需要将碳排放量的空间效应，以及影响因素的空间效应纳入政策制定与实施过程中。因此，结合地理加权模型分析京津冀三地碳排放及其影响因素，有助于了解城市群排碳的空间规律，进而能为集聚城市联合减排、协同发展制定政策方案提供理论基础。

对于我国碳排放的问题，许多学者将研究范围放到国家层面上，利用传统的空间计量经济模型进行实证分析，但这并未充分反映出不同城市碳排放在空间位置上与经济活动中的相互作用。陈志建、王铮[16]指出，我国的经济地理有很大差异，具有变量之间关系不因空间变化而发生改变假定的空间计量经济模型，已经不再适用具有地理空间非稳态性的数据类型，而地理加权回归模型更能反映出变量之间变动的关系。王亚政[17]利用中国省域数据，通过实证分析得出地理加权回归模型相比一般回归模型更优的结论。探索性空间数据分析包括全局自相关、局部自相关、Moran’s I 和 LISA 集聚分析等，由此来检验我国不同省份的能源消耗碳排放及其影响因素具有可解释的空间相关关系。

空间数据携带地理位置的信息，不同的地理位置之间是有差异性的。本文研究京津冀地区碳排放量情况。Holtz-Eakin 和 Selden（1995）在环境库兹涅茨曲线研究中得出二氧化碳排放量随着收入的增加而增加。Shafik（1994）在研究环

境质量指标与人均 GDP 之间的关系中指出：二氧化碳的排放量与人均 GDP 之间呈线性关系。而不同地区的发展情况是不相同的，因此不同的指标对各地区的影响程度不相同，普通的线性回归不能体现出这种异质性。

### （四）研究过程

本文构建了京津冀三地协同发展指标体系，收集京津冀及周边共十三个省的碳排放相关数据，根据已有的碳排放及其影响因素相关文献构造指标，结合地理加权模型进行实证分析，详细过程如下图。

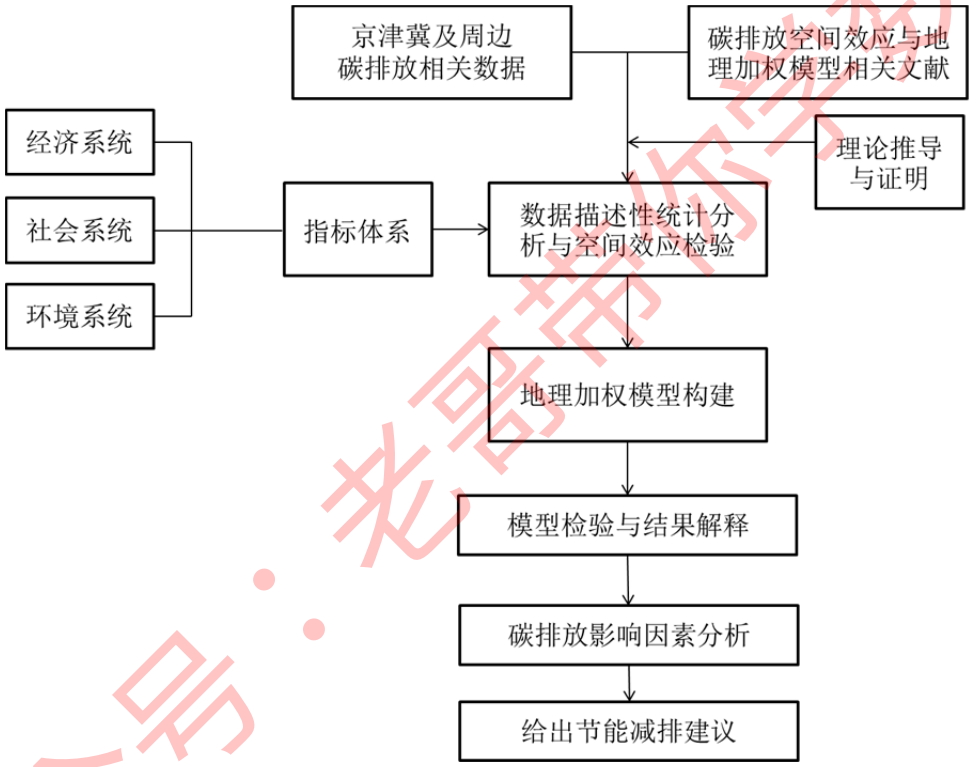


图 1　研究过程图

本文首先依照经济、社会、环境等相关知识构建京津冀碳排放指标体系，然后根据地理加权模型，构造参数估计及其置信区间，最后根据收集的数据计算各碳排放指标，并进行模型求解，最后得出节能减排结论。

## 二、研究方法

### （一）变系数模型

在实际应用中，常需要研究两个变量之间的关系，如相关关系、回归关系等。比起单独考虑两个变量之间的相关关系，我们感兴趣的往往是回归关系，因为回归关系能提供更多信息，在建立回归模型的过程中也能加入我们想要研究的更多变量与前提条件。

我们常把回归模型记做以下形式：

$$E(Y|X=x)=f(x) \qquad \text{公式（1）}$$

在回归模型中，最简单的是线性模型，如下：

$$\mathrm{Y}=X^T\beta+\varepsilon \qquad \text{公式（2）}$$

其中，$\mathrm{X}=(X_1,X_2,\dots,X_p)^T$是$p$维随机变量，$\beta=(\beta_1,\beta_2,\dots,\beta_p)^T$是$p$维参数，$\varepsilon$是均值为 0，方差为$\sigma^2$的随机误差。

在上述模型中，可看到回归函数的形式是已知的，只存在一些未知的参数，也就是说，模型中因变量 y 与自变量 x 之间的关系是静态的，对此，常利用最小二乘法对模型进行求解来获得最优的参数估计。但在具体的实际应用与研究中，回归函数往往是未知的，我们需要考虑利用非参数回归模型进行研究。

在非参数回归模型中，本文主要运用变系数模型，介绍如下。

1.模型形式

变系数模型是由 Hastie 和 Tibshirani 提出，一般形式如下：

$$y=x_1\beta_1(u_1)+x_2\beta_2(u_2)+\cdots+x_p\beta_p(u_p)+\varepsilon \qquad \text{公式（3）}$$

其中，$y$是响应变量，$\mathrm{X}=(x_1,x_2,\dots,x_p)^T$和$\mathrm{U}=(u_1,u_2,\dots,u_p)^T$是$p$维回归变量，$\varepsilon$是均值为 0，方差为$\sigma^2$的随机误差，$x_i\beta_i(u_i)$是未知函数$(i=1,2,\dots,p)$，$u_1,u_2,\dots,u_p$通过未知函数$\beta_i(u_i)$来确定$x_1,x_2,\dots,x_p$的系数，因此$\beta_i(u_i)$体现了$u_i$和$x_i$的关系。

可看到，变系数模型依然保留着线性结构，且一般的线性回归模型属于变系数模型的特殊形式，但由于变系数模型的回归参数是非参数函数形式，因此变系数模型通常比一般线性模型有更强的普适性。

2.模型类型

根据参数形式，变系数模型有多种分类形式，如：

（1）若上式中$\beta_i(u_i)=\beta_i$是常数时，即所有系数都是常数，该变系数模型即为一般线性回归模型；

（2）若在第$i$项中，$x_i\equiv 1$，即为第$i$项$\beta_i(u_i)$时，则该模型所对应的模型为可加模型；

（3）若前 p-1 个变量 x 对应的$\beta_i(u_i)=\beta_i(i=1,2,\ldots,p-1)$，第 P 个项$x_p\equiv 1,\beta_p(u_p)=f(u)$，即该模型为$Y=X^T\beta+f(u)+\varepsilon$，是部分线性可加模型；

（4）若在部分线性可加模型中$X^T\beta$ 部分外部加上一个函数，即模型为：$Y=g(X^T\beta)+f(u)Z^T+\varepsilon$，作为该模型是部分变系数单指标模型，是部分线性可加模型的推广；

（5）若将部分变系数模型进行合并，即模型形式为：$Y=g(X^T\beta)Z+\varepsilon$，则该模型为单指标变系数模型。

由于变系数模型的模型形式有较大的变化空间，因此针对不同类型的数据，研究学者们推出了各式各样的模型类型，包括本文所采用的地理加权模型也是变系数模型的一个推广，在介绍地理加权模型之前，本文先介绍一些变系数模型的相关内容，下面学习一下变系数模型进行参数估计的方法，为地理加权模型的利用与延伸打下基础。

3.参数估计方法

参数估计是回归模型中的关键步骤之一，如何估计变系数模型中的参数一直是学者们研究的重点，为了减少估计的偏差。假设模型为：

$$Y=f(X)+\sigma(X)\varepsilon \qquad \text{公式（4）}$$

根据变系数模型中形式和参数都未知的特点，目前有样条逼近，如 B 样条估计方法，和局部光滑逼近，如局部估计方法这两大类方法。

（1） B 样条估计

B 样条是数值分析学科中样条曲线一种特殊的表示形式。该算法是用一小段一小段的曲线连接或逼近整个曲线，可用于逼近的 B 样条曲线主要有均匀 B 样条曲线、准均匀 B 样条曲线、分段 Bezier 曲线、非均匀 B 样条曲线等。由于本文不采用方法进行参数估计，因此不再进行详细介绍。

（2）局部估计方法

由于非参数模型中既包含未知参数，又包含未知函数形式，因此常利用局部估计方法对其进行估计，主要收集了一下三种方法：

在介绍这三种局部估计方法之前，先介绍一下核函数的概念，核函数记做 K，它代表了局部领域的大小，并且核函数对估计结果的影响较大、较为敏感。一般有如下写法：

$$K_h(.) = \frac{1}{h}K(\frac{\cdot}{h}) \qquad \text{公式（5）}$$

其中，h 称为窗宽（bandwidth）或光滑参数（smoothing parameter），核函数的具体形式与选取在地理加权回归模型部分会有详细介绍。

1）Nadaraya-Watson 估计

Nadaraya-Watson 方法考虑了距离因素，即若要估计在 $x_0$ 处的函数值时，可认为距离 $x_0$ 近的样本点的影响较大，距离 $x_0$ 远的样本点的影响较小，针对此想法，考虑取一个权函数，每一个样本点都对应这一个权函数，使得距离 $x_0$ 越近的样本对应的权函数越大。

记该权函数为 K，也称为核函数，是一个实值函数。Nadaraya-Watson 估计方法根据加权平均求解函数值，即

$$\hat{f}_h(x_0) = \frac{\sum_{i=1}^{n} K_h(X_i - x_0)Y_i}{\sum_{i=1}^{n} K_h(X_i - x_0)} \qquad \text{公式（6）}$$

其中，h 称为窗宽或光滑参数。

2）Gasser-Muller 估计

Gasser-Muller 估计方法是 Gasser 和 Muller 提出的方法，避免了上述 Nadaraya-Watson 估计中分母具有随机性的问题，随机分母会使得在讨论其渐进性质时不方便。Gasser-Muller 估计方法如下：

令$x_0 = -\infty, x_{n+1} = +\infty$;假设$x_0 < x_1 \le x_2 \le \cdots \le x_n < x_{n+1}$;

记$m_i = \frac{x_i + x_{i+1}}{2}$，在第$i$个样本点$x_i$处的权重取为$\int_{m_{i-1}}^{m_i} K_h(u - x_0)du$.

易知，$\sum_{i=1}^{n} \int_{m_{i-1}}^{m_i} K_h(u - x_0)du = 1$,此时 Nadaraya-Watson 估计变成如下结果：

$$\widehat{f_h}(x_0) = \sum_{i=1}^{n} \int_{m_{i-1}}^{m_i} K_h(u - x_0)du\, Y_i \qquad \text{公式（7）}$$

这一结果成为 Gasser-Muller 估计。

3）局部多项式估计

由于变系数模型存在函数形式未知的情况，通常我们会考虑利用泰勒展式表示未知函数，这就是局部多项式估计的思路来源，此为线性估计中的最优估计，在这里介绍局部线性估计的方法如下。

假设$f(x)$在$x_0$附近有二阶导数，则在$x_0$的某一邻域有

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)$$

在上式中，我们将$f(x_0)$看做是$\beta_0$，$f'(x_0)$看做是$\beta_1$.即在估计时，$\widehat{\beta_0} = \hat{f}(x_0), \widehat{\beta_1} = \widehat{f'}(x_0)$,结合最小二乘法,极小化下式,则可得到$\beta_0$与$\beta_1$的估计：

$$\sum_{i=1}^{n} [Y_i - \beta_0 - \beta_1(x_i - x_0)]^2 K_h(x_i - x_0) \qquad \text{公式（8）}$$

$$K_h(.) = \frac{1}{h} K(\frac{\cdot}{h}) \qquad \text{公式（9）}$$

$K$核函数，且有紧支撑，根据最小二乘法得到：

$$\hat{f}(x_0) = \frac{\sum_{i=1}^{n} W_i Y_i}{\sum_{i=1}^{n} W_i} \qquad \text{公式（10）}$$

$$W_i = K_h(x_i - x_0) * \{m_{n,2} - (x_i - x_0)m_{n,1}\}, m_{n,j} = \sum_{i=1}^{n} K_h(x_i - x_0)\,(x_i - x_0)^j$$

此时，$\hat{f}(x_0)$称为$f(x_0)$的局部线性估计。

在上述三种局部估计方法中, 最小偏差与方差是局部线性估计的优势, 因此, 局部线性估计由于 Nadaraya-Watson 估计和 Gasser-Muller 估计。

将一元的局部线性估计方法推广如下：

假设$f(x_0)$在$x_0$附近有$p+1$阶导数，则在$x_0$的某一邻域有：

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{f^{(p)}(x_0)}{p!}(x - x_0)^p \qquad \text{公式（11）}$$

将$f(x)$看做是$\beta_0$，第$m$阶导数$f'(x_0)$看做是$\beta_m, m = 1,2,\ldots p$, 求下式即可：

$$min \sum_{i=1}^{n}[Y_i - \sum_{j=0}^{p}\beta_j(x_i - x_0)^j]^2 K_h(x_i - x_0) \qquad \text{公式（12）}$$

若$\tilde{X} = ((x_i - x_0)^j)_{1\le i\le n, 0\le j\le p}, y = (Y_i)_{i=1}^{n}, \hat{\beta} = (\hat{\beta}_i)_{i=0}^{p}, W = diag\{K_h(x_i - x_0)\}$,则：

$$\hat{\beta} = (\tilde{X}^T W \tilde{X})^{-1} \tilde{X}^T W y \qquad \text{公式（13）}$$

在利用局部多项式估计变系数模型参数时，往往采用两步法，进行两阶段估计。如在部分变系数模型中，第一步利用该方法估计函数形式未知部分；第二步再第一步的条件下，估计线性参数部分。

局部多项式估计是常用的变系数估计方法, 不仅像上述所说的它有相对小的偏差、方差；该方法还没有边界效应，因为权函数的存在，使得不同区域或部分的变化是缓慢的, 即在边界点的估计情况与在内部的估计情况或偏差阶数差别不大，不必减少边界效应。

## （二）地理加权回归模型

地理加权回归模型是变系数模型的推广模型之一。

在空间分析中，地理位置的变化往往影响着变量间的关系与结构，这往往造成空间非平稳性。在之前学习的线性回归分析、非线性回归分析等模型中，这些全局性模型往往把各变量视为具有同质性, 忽视了不同时间或空间的变量的局部特性。如在研究房屋结构时，对于北方地区而言，有无暖气对房屋销售价格或销售量有着很大影响，但对于南方地区的居民来说，有无暖气对房屋销售价格或销

售量影响不大。因此，空间的非平稳性往往造成回归参数随地理位置变化而变化，这就需要更加先进、合适的模型来刻画空间地理数据的特点[18]。

考虑到变量的局部特征，容易想到的模型有局域回归分析，如分区回归模型、移动窗口回归分析等方法。其中，分区回归模型将区域划分成多个小块，在每个小块上进行回归分析，但由于每个小块上的样本点数不一样，则会导致块与块之间存在较大的采样误差，使得块与块交界处的参数系数值发生跳变；移动窗口回归分析模型将每个样本点为中心，设定相同或不同大小的窗口，以该窗口内的样本点数进行回归分析，解决了分区回归模型中采样误差的问题，但每个窗口边界的参数系数值还是会发生跳变，使得模型参数估计的曲面仍是不连续光滑的，并且根据常识我们也容易理解：在不同区域边界处的参数变化总是缓慢变化的，不会产生突变，因此常用的两类局域回归分析方法仍然存在缺陷。

Fortheringham et al（1996）整理了局域回归分析与变参数研究，基于局部光滑思想，提出了地理加权回归（Geographically Weighted Regression--GWR）模型，该模型采用了数据的空间位置。地理加权模型的提出解决了以往利用空间数据进行建模时存在的弊端与缺陷[18]。

1.模型基本形式

地理加权回归方法是对普通线性回归分析方法的拓展，回归参数包含了空间位置信息，即：

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_0(u_i, v_i)x_{ij} + \varepsilon_i, \quad i = 1,2,\dots,n \qquad \text{公式（14）}$$

基本假设为随机误差项服从零均值，方差$\sigma^2$的正态分布，且互不相关，即

$$\varepsilon_i \sim N(0, \sigma^2), Cov(\varepsilon_i, \varepsilon_j) = 0 (i \neq j)$$

可简写为：

$$y_i = \beta_{i0} + \sum_{k=1}^{p} \beta_{ik}x_{ik} + \varepsilon_i \quad i = 1,2,\dots,n \qquad \text{公式（15）}$$

$$y = (X \otimes \beta')I + \varepsilon \qquad \text{公式（16）}$$

表示矩阵的逻辑乘运算，$X$与$\beta$是$n\times(p+1)$维矩阵，设有$n$个样本点和$p$个自变量，$I$为$(p+1)\times 1$的单位向量，则$\beta$形式如下：

$$\beta=\begin{bmatrix}\beta_{10} & \cdots & \beta_{k0} & \cdots & \beta_{n0}\\ \beta_{11} & \cdots & \beta_{k1} & \cdots & \beta_{n1}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ \beta_{1p} & \cdots & \beta_{kp} & \cdots & \beta_{np}\end{bmatrix} \qquad \text{公式（17）}$$

2.加权最小二乘法

由于地理加权回归模型的参数与空间位置有关，参数个数往往大于样本量，直接采用一般参数回归估计往往是不可行的。充分考虑地理加权模型的特点后，采用加权最小二乘法（Weighted Least Squares--WLS）对模型参数进行估计，即不同样本点的重要性不同，距离样本点 i 越近其他样本点重要性越大，距离越远的样本点的其他样本点的重要性越小，因此模型参数可通过下式来求解：

$$\min\sum_{j=1}^{n}\omega_{ij}(y_j-\beta_{i0}-\sum_{m=1}^{p}\beta_{ik}x_{ik})^2$$

$\omega_{ij}$为样本点$i$与其他样本点j之间的距离的单调递减函数

若令$\beta_i=[\beta_{i0}\ \beta_{i1}\ \ldots\ \beta_{ip}]^T$，$W_i=\mathrm{diag}(\omega_{i1},\omega_{i2},\ldots,\omega_{in})$，则i点上的回归参数估计为：

$$\widehat{\beta_i}=(X^TW_iX)^{-1}X^TW_iy \qquad \text{公式（18）}$$

3.预测

拟合值为：

$$\widehat{y_i}=X_i\widehat{\beta_i}=X_i(X^TW_iX)^{-1}X^TW_iy \qquad \text{公式（19）}$$

拟合值得数学期望：

$$E(\widehat{y_i})=X_i\widehat{\beta_i}=X_i(X^TW_iX)^{-1}X^TW_iE(y) \qquad \text{公式（20）}$$

由于$X^TW_iE(y)$可写作：

$$X^TW_iE(y)=(X_1^T,X_2^T,\ldots,X_n^T)\begin{bmatrix}\omega_{i1} & 0 & \ldots & 0\\ 0 & \omega_{i2} & \ldots & 0\\ \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & \ldots & \omega_{in}\end{bmatrix}\begin{bmatrix}X_1\widehat{\beta_1}\\ X_2\widehat{\beta_2}\\ \ldots\\ X_n\widehat{\beta_n}\end{bmatrix} \qquad \text{公式（21）}$$

所以期望可变为：

$$E(\hat{y_i}) = \sum_{k=1}^{n} \omega_{ik} X_i (X^T W_i X)^{-1} X_k^T X_k \widehat{\beta_k} \qquad \text{公式（22）}$$

又根据基本假设知，$\varepsilon_i \sim N(0, \sigma^2), Cov(\varepsilon_i, \varepsilon_j) = 0 (i \neq j)$，所以$Var(y) = \sigma^2 I_n$，有：

$$Var(\hat{y_i}) = Var(X_i (X^T W_i X)^{-1} X^T W_i y) = X_i (X^T W_i X)^{-1} X^T W_i Var(y) W_i X (X^T W_i X)^{-1} X_i^T$$
$$= \sigma^2 X_i (X^T W_i X)^{-1} X^T W_i^2 X (X^T W_i X)^{-1} X_i^T$$

残差是回归分析中预测值与真实值之差，在统计推断问题中残差有着重要意义与作用。

设观测值为$y$，预测值或回归值为$\hat{y}$，则可计算残差为$e_i = y_i - \hat{y}_i = y_i - S_i y$

其中，$S$为帽子矩阵，记做$S = X_i (X^T W_i X)^{-1} X^T W_i$

写作矩阵形式如下：

$$e = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} - \begin{bmatrix} S_1 \\ S_2 \\ \dots \\ S_n \end{bmatrix} y = (I - S) y \qquad \text{公式（23）}$$

4.参数估计与预测值的置信区间

在统计学中，统计推断是主要研究问题，统计推断包括参数估计和假设检验，二者相互区别又互有联系。在参数估计中，需要依赖随机变量的分布函数，利用样本对未知参数进行估计刻画；在假设检验中，对已知参数进行假设，通过样本构建检验统计量来判断假设是否正确。

（1）参数的估计与置信区间

在上面的介绍中，本文已解释了地理加权回归模型的参数估计，即利用加权最小二乘法对模型中个指标参数进行估计。这里不再赘述。

根据随机误差分布与估计参数的期望，可设计如下统计量：

由$\widehat{\beta_i} = (X^T W_i X)^{-1} X^T W_i y$知，$E(\widehat{\beta_i}) = (X^T W_i X)^{-1} X^T W_i E(y) = \beta_i$，因此地理加权回归估计是无偏估计，令$M_i = (X^T W_i X)^{-1} X^T W_i, D_i = M_i M_i^T$，则有$Var(\widehat{\beta_i}) = D_i \sigma^2$，所以在假设条件下，存在统计量：

$$T = \frac{\hat{\beta}_{ik} - \beta_{ik}}{\hat{\sigma}\sqrt{d_{kk}^{i}}} \sim t(r_s) \quad i = 1,2,\dots,n,\ k = 0,1,\dots,p \qquad \text{公式（24）}$$

其中，$d_{kk}^{i}$表示矩阵$D_i$的对角线元素，$r_s = \frac{(tr(R_s))^2}{tr(R_s)^2}$

因此在给定置信水平为$\alpha$的条件下，参数估计值$\hat{\beta}_{ik}$的置信区间为：

$$\hat{\beta}_{ik} \pm \hat{\sigma}\sqrt{d_{kk}^{i}} * t_{1-\frac{\alpha}{2}}(r_s) \qquad \text{公式（25）}$$

（2）预测值的置信区间

在上述分析与介绍中，本文也介绍地理加权模型的预测值求解，但需要注意的是，每一个地理位置都有一组预测值，因此在给出一组新的数据时，要根据新观测点确定权重矩阵，再根据上述步骤，带入模型拟合得到相应预测值结果。

因为$\hat{y}_0$与$y_0$相互独立，根据上面讨论的预测值估计的方差可推出：

$$\begin{aligned} Var(\hat{y}_0 - y_0) &= Var(\hat{y}_0) + Var(y_0) = \sigma^2 X_0 (X^T W_0 X)^{-1} X^T W_0^2 X (X^T W_0 X)^{-1} X_0^T + \sigma^2 \\ &= (1 + S_0^2)\sigma^2 \end{aligned}$$

因此，当满足$\varepsilon_i \sim N(0, \sigma^2 I_n), Cov(\varepsilon_i, \varepsilon_j) = 0 (i \neq j)$的假定时，有

$$T(X_0) = \frac{y_0 - \hat{y}_0}{\hat{\sigma}\sqrt{1+S_0^2}} \sim t(r_s) \qquad \text{公式（26）}$$

当给定置信度为$1-\alpha$时，因变量y在$(u_0, v_0)$处的真实值的置信区间如下：

$$\hat{y}_0 \pm \hat{\sigma}\sqrt{1 + S_0^2} \cdot t_{\alpha/2}(r_s) \qquad \text{公式（27）}$$

5.空间权重矩阵与带宽的选择

由于地理加权模型研究的就是数据在空间中的特点，因此空间权重矩阵是地理加权模型研究的核心。空间权重矩阵的正确选择与否直接影响到研究者对该数据空间关系的分析与认识。

（1）空间权重矩阵的选择

1）距离阈值法

在衡量空间关系值，自然想到的一个衡量尺度就是距离。距离阈值法是最基本、最简单的空间权重矩阵的构造方法，该方法的核心在于取一个合适的阈值D，若样本点i与样本点j之间的距离大于阈值D，则权重取为0，否则取1，即：

$$\omega_{ij}=\begin{cases}1 & d_{ij}\le D\\ 0 & d_{ij}>D\end{cases}$$ 公式（28）

该方法与上述提到的局域回归方法中的移动窗口回归分析法类似，通过判断距离是否超出设定的“窗口”来确定权重。

该方法的优点是计算简单，空间权重矩阵容易构造；但缺点也很明显，由于该权重不是 0 就是 1，因此往往会造成权重的突变，并且在实际应用中，随着空间信息的变化，参数估计会因为某个样本点的移动发生突变，不宜将该方法使用在地理加权模型中。

2）距离反比法

为了避免距离阈值法中权重发生突变的问题，人们又提出了另一个用距离衡量空间关系的方法——距离反比法，构造如下：

$$\omega_{ij}=\frac{1}{d_{ij}^{\alpha}}$$ 公式（29）

该方法的优点是纳入了距离变量，使得不同空间上的权重不会产生突变，该方法简洁，直接考虑了空间远近；但缺点是对于距离很近的样本点来说，权值趋于无穷大，会导致参数估计过程中会剔除一些对研究对象有着重要影响作用的变量，也不适用于地理加权模型中。

3）Gauss 核函数法

为了避免权值趋于无穷大给建模过程中带来的麻烦，学者提出了 Gauss 核函数法，该方法通过一个连续单调递减函数来衡量权值与距离之间的关系，其函数形式如下：

$$w_{ij}=e^{-(\frac{d_{ij}}{h})^2}$$ 公式（30）

其中，$h$用来描述权重、距离间函数关系，即带宽（Bandwidth）。由图可知，若带宽一定，距离越近，权值越大；若同一距离，带宽越大，权值越大，即带宽越大，权重随着距离的增加而减小得越慢[18]。

4）bi-square 核函数法

除了 Guass 核函数外，还可利用 bi-square 核函数来提高计算效率：

$$w_{ij}=\begin{cases}[1-\left(\frac{d_{ij}}{h}\right)^2]^2 & d_{ij}\le D\\ 0 & d_{ij}>D\end{cases} \qquad \text{公式（31）}$$

bi-square 核函数法引入了阈值，在该阈值范围（即带宽$h$）内，通过 bi-square 函数来计算权重；在该阈值范围（即带宽$h$）外，权重为 0，该法使得在阈值附近的点的区中不会发生突变，使得对地理加权模型建模时影响不大。

（2）核函数带宽的确定

在常用的 Guass 核函数与 bi-square 核函数中，当带宽过大，参数估计偏差也会过大，带宽过小却会引起参数估计的方差过大。因此，如何选择一个合适的带宽很重要[18]。

1）交叉验证法

交叉验证法是一种划分样本集或数据集的方法，该方法往往把数据集划分成两个部分，一部分用于模型建立，另一部分用于模型预测或检验。

在核函数的带宽选择问题中，我们利用交叉验证法获得一个差异函数，表达式如下：

$$\mathrm{CV}=\frac{1}{n}\sum_{i=1}^{n}[y_i-\hat{y}_{\neq i}(h)]^2 \qquad \text{公式（32）}$$

将第$i$（$i=1,2,\dots n$）个样本点剔除后进行模型拟合，即只根据某些样本点数据进行回归计算。把不同带宽$h$与其CV值绘制成曲线，取所有CV值中最小的CV值对应的带宽$h$。

为简便计算，提出了广义交叉验证方法，具体计算公式如下：

$$\mathrm{GCV}=\frac{1}{n}\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i(h))^2}{(1-tr\left(\frac{S(h)}{n}\right))^2}=\frac{n\sum_{i=1}^{n}(y_i-\hat{y}_i(h))^2}{(n-tr(S(h))^2} \qquad \text{公式（33）}$$

同样，把不同带宽$h$与其GCV值绘制成曲线，取所有GCV值中最小的GCV值对应的带宽$h$。

2）AIC 准则

在时间序列回归模型中，我们常根据 AIC 准则选择最优的模型。AIC 准则衡量了一个统计模型优良性，该方法来参考了极大似然估计，是极大似然估计的一种推广。

一般地，假设模型的似然函数为$L(\beta,x)$,β的维数是p，对应于回归模型中说明有p个变量，则 AIC 定义为：

$$AIC = -2\ln L(\hat{\vartheta}_L, x) + 2q \qquad \text{公式（34）}$$

其中q为未知参数个数。根据似然函数的定义，我们易知似然函数越大估计量越好，但由于有时候参数的数量会导致似然函数也会增加，因此 AIC 准则中还考虑了惩罚因子2q，综上，我们在挑选模型时，往往认为 AIC 值最小对应的模型为最佳。

根据文献可知，对于一般的回归模型来说，假设随机误差项：

$$\varepsilon_i \sim N(0,\sigma^2 I_n), Cov(\varepsilon_i,\varepsilon_j) = 0(i \neq j), 则 y \sim N(X\beta,\sigma^2 I_n),$$

则似然函数为：

$$L = \frac{1}{(2\pi)^{\frac{n}{2}}}\frac{1}{(\sigma^2)^{\frac{n}{2}}}e^{-\frac{(y-X\beta)^T(y-X\beta)}{2\sigma^2}} \qquad \text{公式（35）}$$

其中的未知参数是β和$\sigma^2$，未知参数个数是$p+2$,则极大似然函数为：

$$\ln L_{max} = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\hat{\sigma}_L^2) - \frac{(y-X\beta)^T(y-X\beta)}{2\hat{\sigma}_L^2} = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\frac{RSS}{n}\right) - \frac{n}{2} \qquad \text{公式（36）}$$

整理可得回归模型的 AIC 公式为：

$$AIC = n\ln(RSS) + 2q \qquad \text{公式（37）}$$

AIC 准则用途颇多，在本文中将考虑利用 AIC 准则，在地理加权模型中进行核函数的带宽选择，即：

$$AIC = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n\left[\frac{n+tr(S)}{n-2-tr(S)}\right] \qquad \text{公式（38）}$$

其中，带宽$h$的函数以tr(S)表示，随机误差方差的极大似然估计为$\hat{\sigma}$。因此，在所有带宽的地理加权回归模型中，AIC 最小的模型所对应的带宽即为最优带宽。

3）AICc 准则

在样本量较大的情况下，AIC准则会存在误差缺乏解释性，因此对 AIC 准则进行了改进，得到 AICc 准则，如下：

$$\mathrm{AICc} = -2\mathrm{Loglikelihood} + 2\mathrm{k} + \frac{2\mathrm{k}(\mathrm{k}+1)}{\mathrm{n}-\mathrm{k}-1} \qquad \text{公式（39）}$$

其中，k是模型中估计参数的个数，n是模型中的样本量。

## （三）本文研究方法相关理论推导

本文将根据邻近标准确定空间权重矩阵，并且推导出加权最小二乘法进行地理加权回归模型的参数估计，具体推导过程如下：

1.空间权重矩阵的确定

与因变量的空间自回归过程相联系的矩阵称作空间权值矩阵。在研究过程中，矩阵的选取是外生的，因为$n*n$维的矩阵$W$隐含着不同地区$i$与$j$相关空间的外生信息，该矩阵通过权值计算就能得到。

空间矩阵$W$中对角线元素$w_{ij}$设定为 0，而$w_{ij}$表示区域$i$与$j$在空间上相联系的原因，为了减少区域间的外在影响，权值矩阵会被标准化（$w_{ij}^{*} = w_{ij} / \sum_{i=1}^{n} w_{ij}$）为行元素之和为 1 的矩阵。

衡量地理联系方法有两种：邻近指标与距离指标。据此确定的空间矩阵为二进制的邻近空间权值矩阵，矩阵元素$w_{ij}$的表示采用邻近标准或距离标准，定义空间对象的相互邻近关系，方便将地理位置信息的有关特征放在空间上进行比较。

空间矩阵$W$可以用以下矩阵来表示：

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} \end{bmatrix} \qquad \text{公式（40）}$$

当选择相邻标准，$w_{ij}$表示如下：

$$w_{ij}=\begin{cases}1，\ i、j相邻；\\0，\ i、j不相邻；\end{cases}$$ 公式（41）

其中，$i=1,2,\ldots,n$，$j=1,2,\ldots,m$，基于邻近概念的空间权值矩阵有一阶邻近矩阵和高阶邻近矩阵两种形式。

一阶邻近矩阵假定两个地区要具备共同边界，有 Rook 邻近与 Queen 邻近两种算法。Rook 邻近只用共同边界来定义邻近，Queen 邻近除去公共边界外还包括共同顶点。所以说，Queen 邻近定义的空间矩阵往往表示周围单元与该单元具备更紧密的关联。另外，区域公共边界的长度不同时其空间作用强度也不同，也可以将公共边界长度纳入考虑范围。

提出高阶邻近矩阵在于去除矩阵创建时出现的循环。二阶邻近矩阵表示了一种空间滞后的邻近矩阵，即表达了邻近单元的相邻单元其空间信息。当分析时空数据并假设随着时间推移产生空间溢出效应时，这种类型的空间权值矩阵将很有用，所以邻近空间权值矩阵运用广泛。

提出 K 值最邻近空间矩阵是因为使用简单空间矩阵可能会导致不平衡的邻近矩阵结构。试想一下，当地区面积差异很大时，会出现面积较小的地区具有更多邻近单元，而较大的地区的邻近单元可能有很少甚至没有，此时用到 K 值最邻近空间矩阵。通常在给出空间单元周围选取最邻近的若干单元来计算 K 值最邻近空间矩阵权值的大小。

基于距离的空间权值矩阵$w_{ij}$表示如下：

$$w_{ij}(d)=\begin{cases}1，\ i、j在距离\ \mathrm{d}\ 之内（相邻）\\0，\ i、j在距离\ \mathrm{d}\ 之外（不相邻）\end{cases}$$ 公式（42）

基于距离的空间权值矩阵方法，假定空间相互作用的强度取决于两地质心距离或区域行政中心间的距离，这种空间权值矩阵较为常见。时空数据中两地的距离通过经纬度计算得出。欧氏距离、弧式距离可用来计算两点的距离，需要用到

具有地理位置信息的变量。另外，还有一种较复杂的权重矩阵，包括但不限于经济、社会等因素来界定距离，这里不再赘述。

2.参数估计

（1）加权最小二乘法

Brunsdon 等（1996）首次提出了地理加权回归模型，设定如下：

$$Y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)X_{i1} + \beta_2(u_i, v_i)X_{i2} + \ldots + \beta_p(u_i, v_i)X_{ip} + \varepsilon_i \quad \text{公式（43）}$$

其中，$\beta_j(u, v)$（j=0,1,⋯,p）是反应为空间地理位置的函数，$u_i, v_i$通常是某地的经度和纬度。上述模型将地理位置考虑进去，能够更好地反应异质性。

Tobler 地理学第一定律（Tobler,1970）认为：万物都是空间相关的，距离越近的事物其相关性越大，反之其相关性越小。基于此梅长林，王宁[19]等人通过在$(u_0, v_0)$处采用核函数构造一组权重，并用加权最小二乘法对参数进行了估计：

$$\omega_i(u_0, v_0) = K(\frac{d_{0i}}{h}), i=1,2,\ldots,n \quad \text{公式（44）}$$

其中，$d_{0i}$指第$i$个地区与$(u_0, v_0)$之间的距离，h 为光滑函数，$K(\cdot)$是一个核函数，通常取 Gaussian 核函数：

$$\mathrm{K(x)} = \sqrt{2\pi}^{-1}\exp(-\mathrm{x}^2/2) \quad \text{公式（45）}$$

通过极小化

$$\sum_{i=1}^{n}\{Y_i - \beta_0(u_0, v_0) - \sum_{j=0}^{p}\beta_j(u_0, v_0)X_{ij}\}^2\omega_i(u_0, v_0) \quad \text{公式（46）}$$

求得$\beta_j(u_0, v_0)$的估计值，地理加权回归的矩阵形式如下：

$$Y = X\beta(u_0, v_0) + \varepsilon \quad \text{公式（47）}$$

其中，$X = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{pmatrix}$，$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$，$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$，

所以有：

$$\beta(u_0, v_0) = (\beta_0(u_0, v_0), \beta_1(u_0, v_0), \ldots, \beta_p(u_0, v_0))^T ; \quad \text{公式（48）}$$

$$W(u_0, v_0) = Diag(\omega_1(u_0, v_0), \omega_2(u_0, v_0), \ldots, \omega_n(u_0, v_0)). \quad \text{公式（49）}$$

加权最小二乘结果为：

$$\hat{\beta}(u_0,v_0)=\left(\hat{\beta}_0(u_0,v_0),\hat{\beta}_1(u_0,v_0),\ldots,\hat{\beta}_p(u_0,v_0)\right)^T$$
$$=(X^TW(u_0,v_0)X)^{-1}X^TW(u_0,v_0)Y \quad \text{公式（50）}$$

特别地，当$(u_0,v_0)$分别取$(u_i,v_i)$时，就可以得到不同地区的估计值：

$$\hat{\beta}(u_i,v_i)=\left(\hat{\beta}_0(u_i,v_i),\hat{\beta}_1(u_i,v_i),\ldots,\hat{\beta}_p(u_i,v_i)\right)^T$$
$$=(X^TW(u_i,v_i)X)^{-1}X^TW(u_i,v_i)Y \quad \text{公式（51）}$$

其中，$i=1,2,\ldots n$.

（2）局部线性估计法

地理加权回归实质上是变系数回归模型，张日权，卢一强等[20]在变系数模型研究中采用局部线性的思想对系数进行估计。但是他们考虑的系数函数$\beta(\cdot)$是一维情况，而地理位置是一个二维变量，下面将介绍空间变系数的局部线性估计。

对于公式：

$$Y_i=\beta_0(u_i,v_i)+\beta_1(u_i,v_i)X_{i1}+\beta_2(u_i,v_i)X_{i2}+\ldots+\beta_p(u_i,v_i)X_{ip}+\varepsilon_i \quad \text{公式（52）}$$

可以写成矩阵形式：

$$Y_i=\sum_{j=0}^{p}\beta_j(u_i,v_i)X_{ij}+\varepsilon_i,i=1,2,\ldots,n, \quad \text{公式（53）}$$

当$X_{i1}\equiv 1$时，该空间变系数模型有截距项。

若$\beta_j(u_i,v_i)$（$j=0,1,2,\ldots,p$）有连续的二阶偏导数，那么在$(u_0,v_0)$附近有：

$$\beta_j(u_i,v_i)\approx\beta_j(u_0,v_0)+\beta_j'(u)(u_0,v_0)(u-u_0)+\beta_j'(v)(u_0,v_0)(v-v_0),$$

公式（54）

其中，j = 0,1,2,…,p，$\beta_j'(u)(u_0,v_0)$和$\beta_j'(v)(u_0,v_0)$分别表示$\beta_j(u_i,v_i)$关于$u_i$和$v_i$得偏导在$(u_0,v_0)$上的值，我们采用局部线性估计的思想，极小化下式：

$$min_{\alpha_j,\beta_j,j=1,2,\ldots p}\sum_{i=1}^{n}\{Y_i-$$

$$\sum_{j=0}^{p}(\beta_j(u_0,v_0)+\beta_j'(u)(u_0,v_0)(u_i-u_0)+$$

$$\beta_j'(v)(u_0,v_0)(v_i-v_0))X_{ij}\}^2K_h[(u_i,v_i)-(u_0,v_0)] \quad \text{公式（55）}$$

其中，$K_h(\cdot)=h^{-1}K(\frac{\cdot}{h})$，$K(\cdot)$是一个核函数，通常取 Gaussian 核函数：

$$K(x)=\sqrt{2\pi}^{-1}exp(-x^2/2) \quad \text{公式（56）}$$

对$\beta_j(u_0,v_0)$、$\beta_j'(u)(u_0,v_0)$、$\beta_j'(v)(u_0,v_0)$求偏导得：

$$\sum_{i=1}^{n}\{Y_i-\sum_{j=0}^{p}(\beta_j(u_0,v_0)+\beta_j'(u)(u_0,v_0)(u_i-u_0)+\beta_j'(v)(u_0,v_0)(v_i-v_0))X_{ij}\}K_h[(u_i,v_i)-(u_0,v_0)]\begin{pmatrix}X_j^T\\(u_i-u_0)X_j^T\\(v_i-v_0)X_j^T\end{pmatrix}^T=0 \quad \text{公式（57）}$$

所以有：

$$\left(\beta_j(u_0,v_0),\beta_j'(u)(u_0,v_0),\beta_j'(v)(u_0,v_0)\right)^T=$$

$$\sum_{i=1}^{n}(\tilde{X}^T\tilde{L}\tilde{X})^{-1}(X_j^T,(u_i-u_0)X_j^T,)^T K_h[(u_i,v_i)-(u_0,v_0)]Y_i \quad \text{公式（58）}$$

进而，对所有$j=1,\ldots,p$有：

$$\hat{\beta}_j(u_0,v_0)=\sum_{i=1}^{n}e_{j,2p}^T(\tilde{X}^T\tilde{L}\tilde{X})^{-1}(X_j^T,(u_i-u_0)X_j^T,(v_i-v_0)X_j^T)^T K_h[(u_i,v_i)-(u_0,v_0)]Y_i \quad \text{公式（59）}$$

其中，$e_{j,2p}$是一个$2p\times 1$的向量，其第$j$个元素为 1，其余为 0，$\tilde{\mathrm{X}}$是一个的矩阵，其第$i$行为：

$$(X_j^T,(u_i-u_0)X_j^T,(v_i-v_0)X_j^T),$$

$$\tilde{L}=Diag(K_{h1}[(u_1,v_1)-(u_0,v_0)],K_{h2}[(u_2,v_2)-(u_0,v_0)],\ldots,K_{hn}[(u_n,v_n)-(u_0,v_0)]). \quad \text{公式（60）}$$

综上，就是本文具体研究方法的理论推导，在具体实证分析中，本文还将采用 Moran’s I、Geary ’s C、Getis 指数等进行空间自相关检验，具体分析见模型建立部分。

## 三、统计监测指标体系的建立

为了方便准确地展示京津冀三地协同发展的本质特征，需要寻找与协同发展相关联的影响指标，对其进行提炼与归类形成指标体系，实现定性到定量的转化，便于对区域协同发展进行测度。在构建指标体系过程中，既要保证数据来源真实，还要充分考虑能反映出京津冀三地协同发展的特征，从不同角度描述三地之间协同的关联性与独特性。

区域协同发展的指标体系不仅要反映经济表现，更要考虑到社会状况与环境状况两方面。因此，本文构建京津冀三地协同发展统计监测指标体系，该指标体系首先由四个层次构成，分别是目标层、子系统、一级综合指标与二级综合指标构成，具体构成及解释如下所示。

表 1　京津冀协同发展综合指标体系

| 目标层 | 子系统 | 一级综合指标 | 二级综合指标（三级指标见下文） |
|---|---|---|---|
| 协同发展 | 经济系统 | 对内综合指标 | 经济发展（7 个子指标） |
| | | | 经济结构（2 个子指标） |
| | | 对外综合指标 | 对外贸易（2 个子指标） |
| | 社会系统 | 人文综合指标 | 人口水平（3 个子指标） |
| | | | 科教水平（4 个子指标） |
| | | “行&用”综合指标 | 交通水平（4 个子指标） |
| | | | 公用设施（2 个子指标） |
| | 环境系统 | 环境综合指标 | 环境污染（3 个子指标） |
| | | | 环境治理（3 个子指标） |

## （一）子系统之经济系统

京津冀三地区域协同发展子系统之经济系统，包括对内综合指标与对外综合指标，体现当地经济发展状况、经济结构状况与对外贸易规模。经济发展状况包含地区生产总值、人均 GDP、地区固定资产投资等，经济结构子指标包括地区农林牧渔业产值占比与地区第二产业产值，对外贸易规模通过外商投资企业出口总额与外商投资企业进口总额体现。

## （二）子系统之社会系统

京津冀三地区域协同发展子系统之社会系统，包括人文综合指标与“行&用”综合指标，反映了地区人口水平、科教水平、交通水平和公用设施状况。人口水平通过地区总人口、地区城镇人口与城市化水平指标来体现；科教水平通过地区普通高等学校数、地区教育经费和每十万人口高等学校平均在校生数指标来体现；交通水平通过地区公共交通车辆运营数、轨道交通配属车辆数、民用汽车拥有量等来体现；公用设施状况通过城市人均绿地公园面积和地区人均日生活用水量来体现。

## （三）子系统之环境系统

京津冀三地区域协同发展子系统之环境系统，由环境综合指标构成，反映了环境污染现状与环境治理现状。环境污染现状通过八种主要能源排碳、地区烟(粉)尘排放量和地区生活垃圾清运量来体现；环境治理现状通过生活垃圾无害化处理率、地区造林总面积与林业投资来体现。

对二级指标细分得到三级指标，三级指标层总共由 30 个指标构成，经济系统指标内容、社会系统指标内容、环境系统指标内容分别如下表所示。

表 2　京津冀协同发展指标体系——经济系统指标内容

| 二级指标 | 三级指标 | 单位 | 变量 |
|---|---|---|---|
| 经济发展 | 地区生产总值 | (亿元) | $I_1$ |
| | 人均 GDP | (元/人) | $I_2$ |
| | 地区固定资产投资 | (亿元) | $I_3$ |
| | 地区房地产开发投资 | (亿元) | $I_4$ |
| | 地方财政税收收入 | (亿元) | $I_5$ |
| | 地区社会消费品零售总额 | (亿元) | $I_6$ |
| | 地区城镇单位就业人员 | (万人) | $I_7$ |
| 经济结构 | 地区农林牧渔业产值占比 | % | $I_8$ |
| | 地区第二产业产值 | (亿元) | $I_9$ |
| 对外贸易 | 外商投资企业出口总额 | (千美元) | $I_{10}$ |
| | 外商投资企业进口总额 | (千美元) | $I_{11}$ |

表 3　京津冀协同发展指标体系——社会系统指标内容

| 二级指标 | 三级指标 | 单位 | 变量 |
|---|---|---|---|
| 人口水平 | 地区总人口 | (万人) | $S_1$ |
| | 地区城镇人口 | (万人) | $S_2$ |
| | 城市化水平 | % | $S_3$ |
| 科教水平 | 地区普通高等学校数 | (所) | $S_4$ |
| | 地区普通高等学校教职工总数 | (万人) | $S_5$ |
| | 地区教育经费 | (万元) | $S_6$ |
| | 每十万人口高等学校平均在校生数 | (人) | $S_7$ |
| 交通水平 | 地区公共交通车辆运营数 | (万辆) | $S_8$ |
| | 轨道交通配属车辆数 | (万辆) | $S_9$ |
| | 民用汽车拥有量 | (万辆) | $S_{10}$ |
| | 私人汽车拥有量 | (万辆) | $S_{11}$ |
| 公用设施 | 城市人均绿地公园面积 | (平方米/人) | $S_{12}$ |
| | 地区人均日生活用水量 | (升) | $S_{13}$ |

表 4　京津冀协同发展指标体系——环境系统指标内容

| 二级指标 | 三级指标 | 单位 | 变量 |
|---|---|---|---|
| 环境污染 | 八种主要能源排碳 | (万吨) | $E_1$ |
| | 地区烟(粉)尘排放量 | (万吨) | $E_2$ |
| | 地区生活垃圾清运量 | (万吨) | $E_3$ |
| 环境治理 | 地区造林总面积 | (千公顷) | $E_4$ |
| | 地区林业投资 | (万元) | $E_5$ |
| | 生活垃圾无害化处理率 | % | $E_6$ |

### （四）模型使用指标构造选取原则

对于具体的模型建立与结论梳理过程来说，京津冀协同发展指标体系较为庞大且不易解释，不同系统间、相同系统内具体指标单位均有差异，所以需要在综合指标与具体指标之间达到平衡，消除量纲影响，尤其考虑因人口数量差异引起的多数指标的差距，由此得到能够反映京津冀区域协同发展过程的具体指标。后文将分别从区域经济角度、社会角度与环境角度，利用三级指标内容构建具体适用的指标，详细解释指标的构造含义与使用。

## 四、数据来源与分析

### （一）数据来源

为检验近些年来京津冀区域碳排放的溢出效应，要根据区域的真实情况选择合适的指标，对选择的指标进行基本的统计分析，然后结合数学模型来解决以下问题：

（1）京津冀三地的碳排放溢出指标是否具有空间自相关性；

（2）产业结构、交通规模、城市化水平等因素是否对排碳放溢出有影响；

（3）2008 年至 2017 年京津冀三地碳排放的基本变化情况与原因。

京津冀区域经济规划的提出在 2006 年，本文使用 2008 年至 2017 年的面板数据进行分析，希望探究区域经济计划施行后，京津冀三地碳排放溢情况出有何变化。通过研究京津冀三地及其周边地区碳排放面板数据，探讨京津冀内部的影响作用如何，以及周边地区对于京津冀的影响。面板数据的时间跨度为十年，过长时期的面板数据有可能导致模型效果不理想，因此后文会将重点研究时期主要投放在较为邻近的年份。

### （二）变量选取

影响碳排放量的因素有很多，众多文献已做出碳排放量与经济规模等因素相关的说明，并且既有直接影响的因素也有间接影响的因素。本文根据京津冀协同发展指标体系中的三级指标内容，构建多个变量作为影响碳排放量的关键因素，模型涉及变量如下表所示。

表 5　变量说明

| 变量 | 定义 | 单位 |
|---|---|---|
| 二氧化碳排放量 | 八种主要能源消耗产生的二氧化碳排放量 | 万吨 |
| 人均 GDP | 人均实际 GDP | 美元 |
| 城市化水平 | 城镇人口占总人口比例 | % |
| 总人口 | 地区总人口 | 万人 |
| 交通规模 | 每万人民用汽车拥有量 | 万辆/万人 |
| 第二产业占比 | 第二产业生产总值占比 | % |

考虑到过多的自变量可能导致多重共线性，因此选择以上 5 个变量作为碳排放量的影响因素，理由如下：第一，人均 GDP 可以充分反映出当地人民的富裕程度；第二，城市化水平可以反映城市化发展进度；第三，地区总人口对二氧化碳排放量有直接影响；第四，交通一体化是京津冀区域联系的关键领域，本文选择影响碳排放的民用汽车拥有量作为自变量之一，探讨交通规模与碳排放量的关系；第五，工业是最大的能源消耗业，第二产业占比能反映产业结构，是评价城市工业化是否发展充分的重要指标。

## （三）变量说明

本文使用的自变量与因变量数据来自国家数据发布的统计公报，碳排放系数来源于《省级温室气体清单编制指南》（发改办气候[2011]1041 号），所有数据来源真实。

1.碳排放溢出效应指标

为了构造碳排放溢出指标，本文选取京津冀三地及其周边相邻省地，包括内蒙古、黑龙江、辽宁省、陕西省、山西省、安徽省、山东省、湖北省、河南省和江苏省，每万人碳排放量作为反映碳排放溢出效应的指标，构造形式如下：

$$CARB_{it} = \frac{E_{it}}{S_{1it}}$$

其中，$E_{it}$表示该年份第$i$个地区的八种主要能源产生的碳排量总和，$S_{1it}$表示该年份第$i$个地区的年末人口总量。

为了计算各省市碳排放量，我们选取了北京市、天津市、河北省等十三个省市 2008 年～2017 年的八种主要能源消费量作为城市碳排放量的构成部分，包括煤炭、焦炭、原油、天然气等能源的消费量，利用以上数据可以计算得出不同地区在不同年份的二氧化碳排放量。

北京市、天津市、河北省 2008 年～2017 年的间接二氧化碳排放量作为因变量，排碳数据计算内容包括八种主要能源的消费量与二氧化碳排放系数两部分，计算公式如下：

$$E_{it} = \sum_{j=1}^{8} e_{itj} = \sum_{j=1}^{8} c_{itj} \times coef_j$$

其中，$E_{it}$为第$i$个地区在第$t$年间八种主要能源消耗产生的二氧化碳排放量；$e_{itj}$为第$i$个地区在第$t$年的第$j$种主要能源的间接二氧化碳排放量；$c_{itj}$为第$i$个地区在第$t$年的第$j$种主要能源的消费量；$coef_j$为第$j$种主要能源的二氧化碳排放系数。公式中的$i = 1,2,3$，分别对应的地区为北京市、天津市、河北省；$t = 2008,2009,\cdots,2017$，即$t$为年份；$j = 1,2,\cdots,8$，分别对应的主要能源为煤炭、焦炭、原油等。

碳排放系数指每单位该种能源所产生的二氧化碳排放量。通常情况下，根据 IPCC 的假定，能源的碳排放系数是固定的，具体系数值见下表。

表 6　八种能源折标准碳排放参考系数

| 能源类别（单位不同） | CO2 排放系数（单位不同） |
|---|---|
| 煤炭（kg） | 1.9003kg-$CO_2$/kg |
| 焦炭（kg） | 2.8604kg-$CO_2$/kg |
| 原油（kg） | 3.0202kg-$CO_2$/kg |
| 汽油（kg） | 2.9251kg-$CO_2$/kg |
| 煤油（kg） | 3.0179kg-$CO_2$/kg |
| 柴油（kg） | 3.0959kg-$CO_2$/kg |
| 燃料油（kg） | 3.1705kg-$CO_2$/kg |
| 天然气（$m^3$） | 2.1622kg-$CO_2$/$m^3$ |

2.影响因素

模型的因变量为当地居民活动对环境的影响，以二氧化碳排放量（万吨/万人）表示，对于模型的自变量，本文选取了京津冀地区 2008 年～2017 年的人均 GDP(元)反映财富规模、地区城镇人口对年末总人口占比（%）反映城市化水平、人口总数(万人)反映人口规模、地区民用汽车拥有量（万辆）反映交通规模、第二产业产值对地区生产总值占比（%）反映产业结构，为了行文的规范与方便，分别用不同字母代表上述变量，如下表所示。

表 7　变量表示

| 变量表示 | 变量说明 | 单位 |
| --- | --- | --- |
| $E_{it}$ | 人类活动对环境影响 | 万吨/万人 |
| $W_{it}$ | 当地居民富裕程度 | 元/人 |
| $U_{it}$ | 地区城市化水平 | % |
| $S_{it}$ | 人口规模 | 万人 |
| $V_{it}$ | 交通规模 | 万辆/万人 |
| $R_{it}$ | 产业结构 | % |

对于上表中的当地居民富裕程度指标，其构造如下：

$$W_{it} = \frac{I_{1it}}{S_{1it}}$$

其中，$I_{1it}$是来自经济系统的三级指标，表示该年份第$i$个地区的生产总值，$S_{1it}$是来自社会系统的三级指标，表示该年份第$i$个地区的年末人口总量。

地区城市化水平指标，其构造如下：

$$U_{it} = \frac{S_{2it}}{S_{1it}}$$

其中，$S_{2it}$是来自社会系统的三级指标，表示该年份第$i$个地区的城镇人口数，人口规模指标直接使用三级指标$S_{1it}$表示。

交通规模指标，其构造如下：

$$V_{it} = \frac{S_{10it}}{S_{1it}}$$

其中，$S_{10it}$是来自社会系统的三级指标，表示该年份第$i$个地区的民用汽车拥有量。

产业结构指标，其构造如下：

$$R_{it} = \frac{I_{9it}}{I_{1it}}$$

其中，$I_{9it}$是来自经济系统的三级指标，表示该年份第$i$个地区的第二产业产值。

## （四）描述性统计分析

1.京津冀及周边地区碳排放量对比分析

在建立模型前，通过直接数据了解京津冀地区与周边其他省区在二氧化碳排放量的异同，由于数据量较大，在此选择了 2008 年、2013 年和 2017 年三年的二氧化碳排放量进行对比分析。

表 8　京津冀及周边地区碳排放量数据

| 区域 | 省份 | 二氧化碳排放量（单位：万吨/年） | | |
|---|---|---|---|---|
| | | 2008 | 2013 | 2017 |
| 京津冀 | 北京 | 13314.5984 | 11911.65603 | 11113.72163 |
| | 天津 | 14037.9694 | 20965.07508 | 18712.69533 |
| | 河北 | 70553.28893 | 92948.48304 | 85745.27132 |
| 周边区域 | 内蒙古 | 50399.87052 | 76566.00638 | 83154.21636 |
| | 黑龙江 | 30267.21721 | 36070.73254 | 37363.28217 |
| | 辽宁 | 59340.89911 | 71207.8032 | 71759.35557 |
| | 陕西 | 26626.59113 | 46136.62162 | 50537.27919 |
| | 山西 | 63044.72019 | 79026.28857 | 91413.39514 |
| | 安徽 | 26911.26519 | 37970.26195 | 40550.68328 |
| | 山东 | 94377.14119 | 116828.2957 | 142134.8514 |
| | 湖北 | 29321.80606 | 35417.914633 | 36400.96762 |
| | 河南 | 55007.267722 | 61734.50616 | 58877.11088 |
| | 江苏 | 57560.060353 | 80884.89805 | 85876.6198 |

由上表，除北京以外，其他各省的二氧化碳排放量在这 3 年基本呈上升趋势，这可能是由于北京作为首都，在环境治理、车辆限行等方面有着严格的秩序要求，使得碳排放量呈下降趋势。在京津冀地区，河北省的碳排放量一直是三者中最高者，由于河北省是资源型地区，采矿业、重工加工业等发达，使得河北省的碳排放量较大这体现了京津冀地区的重要差别。纵向对比各省二氧化碳排放量，京津冀地区在 2013 年到 2017 年的碳排放量是呈下降趋势的，其他省区都呈上升趋势，由于京津冀地区受到历史、政策或经济方面等利益因素的影响，使得这一地区又呈现出一定的联系。

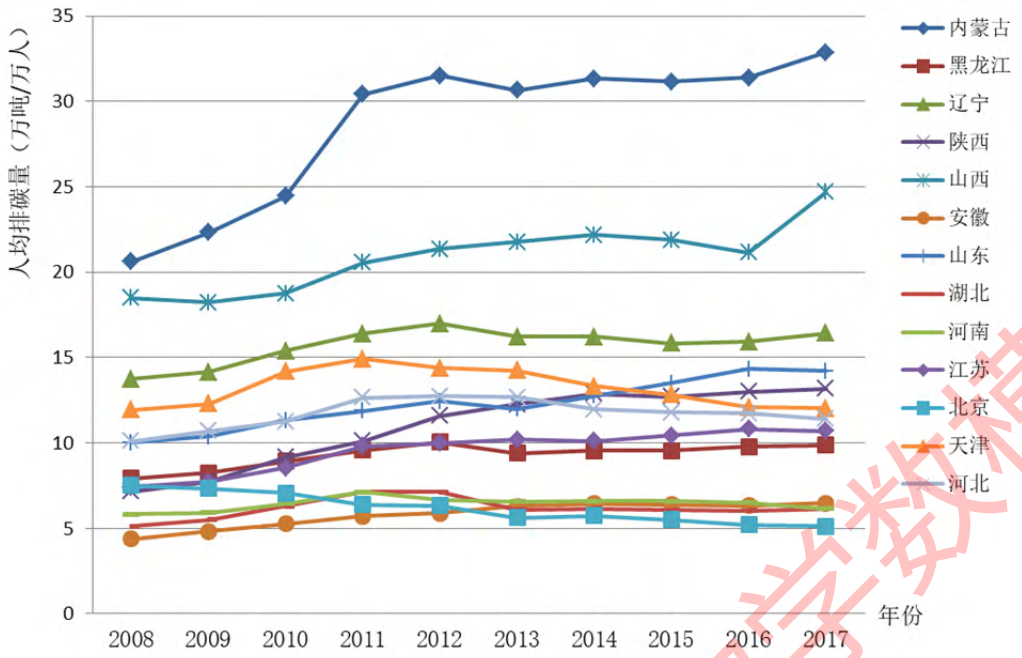


图 2　2008～2017 年京津冀及周边地区人均碳排放

由上图可看到，随着时间推移，京津冀（北京、天津、河北）地区的人均碳排放量会有下降趋势，由于这三个地区存在一定的联系，相应政策或环境保护措施会相互影响，使得碳排放量会得到控制，呈现减少的情况。明显可看到内蒙古的人均碳排放量明显高于其他各省，这可能是因为内蒙古地广人稀所导致的结果；其次是山西，山西本省就是一个煤矿大省，煤炭资源丰富，相应的煤炭加工厂等相对较多，导致其人均碳排放量较高。

2.京津冀地区人均 GDP 对比分析

GDP 是国内生产总值，代表着一个地区经济状况和发展水平。了解或研究京津冀地区的人均 GDP 是其二氧化碳排放量的一个重要影响因素。

表 9　京津冀地区人均 GDP 数据

| 区域 | 省份 | 人均 GDP（单位：元/人） | | |
|---|---|---|---|---|
| | | 2008 | 2013 | 2017 |
| 京津冀 | 北京 | 68541 | 101023 | 137596 |
| | 天津 | 45242 | 68937 | 79837 |
| | 河北 | 20385 | 33187 | 40833 |

上表中，抽取了京津冀地区（北京、天津、河北）2008 年、2013 年和 2017 年这三年的 GDP 数据，可看到作为首都的北京人均 GDP 是三者中的最高者，且涨幅也最高，其次是天津、河北。由于北京是一个人才聚集的知识型地区，文化产业、高新技术产业等是主要产业，科技和技术的进步促进了人民生活水平的发展，因此人均 GDP 较高。

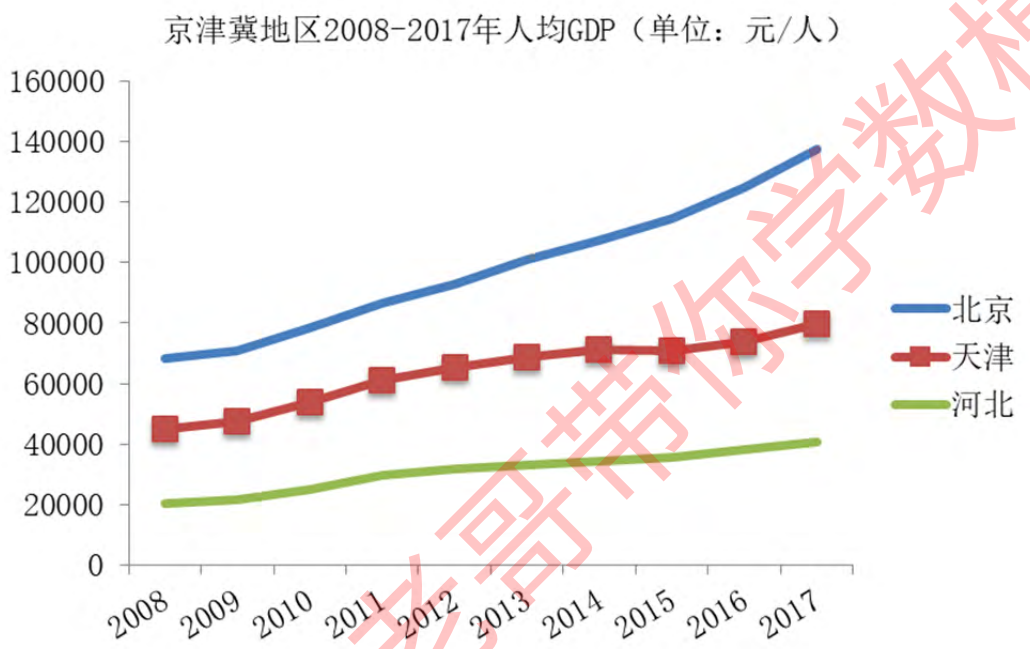


图 3　京津冀 2008 年～2017 年人均 GDP 数据对比

由上图可知，天津地区与河北地区的人均 GDP 呈缓慢上升状态，而北京地区的人均 GDP 增幅明显大于天津与河北地区，且从 2014 年起，北京的人均 GDP 增加速度明显提高。还可看到天津人均 GDP 明显高于河北人均 GDP，由于天津是加工工业占优势的加工型区域，其技术与产业相对于资源型的河北来说较为发达，因此人均 GDP 比河北高。

3.京津冀 2008 年～2017 年城市化水平对比分析

城市化水平，也称为“城镇化率”。城市化水平体现了一个地区在教育、交通、基础设施、工业等建设方面的水平，也是一个地区二氧化碳排放量的潜在因素之一，接下来对 2008 年至 2017 年京津冀地区城市化水平进行描述性分析如下。

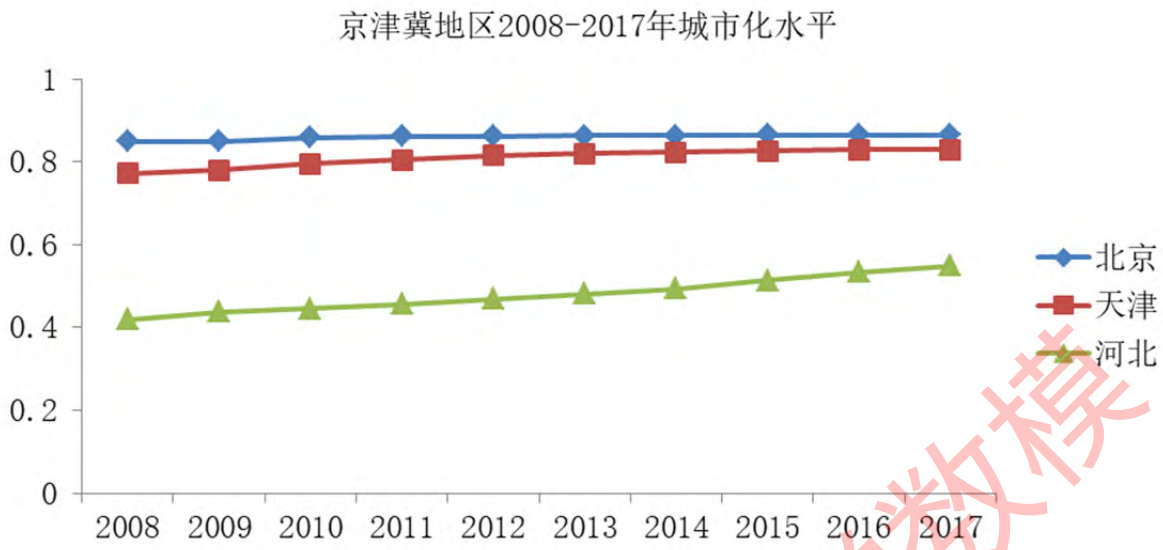


图 4　京津冀 2008 年～2017 年城市化水平对比

由上图可看到，河北与北京、天津的城市化水平存在较大差距，这可能是因为北京与天津同是直辖市，而河北是一个省级地区，区域跨度较大，省内部发展情况差距较大导致了这样的局面。还可看到北京与天津的城市化水平较为相似，并且由于北京与天津存在科技与工业技术上的密切交流与扩散，使得天津的城市化水平随着时间发展越来越接近于北京的城市化水平。

## 五、模型建立与参数估计

### （一）空间自相关检验

在建立模型前，检验空间相关是否明显。当空间效应的影响明显，就需要将空间效应考虑其中；当空间效应并未充分表现出来时，就可以采用普通的估计方法来估计模型参数，通常采用最小二乘估计。常见的空间自相关数包括 Moran’s I、Geary ’ s C、Getis 指数，本文使用 Moran’s I 指数来检验空间效应是否存在，其定义如下：

$$Moran's\ I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\,(Y_i - \bar{Y})(Y_j - \bar{Y})}{S^2 \sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}}$$

其中，$S^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})$，$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$，$Y_i$表示第$i$个地区的观测值，$n$为地区总数，$W_{ij}$为邻近空间权值矩阵，矩阵的元素采用邻近标准或距离标准，意义在于给出空间上的相对远近关系。莫兰指数的检验统计量为$Z$，当其大于临界值，则表示区域经济行为在空间分布上具有明显的正向相关关系，即邻近地区的相似特征值有集群现象。

对于 2008～2017 年京津冀及其周边共十三个省，使用其对应的每万人碳排放量指标进行空间自相关检验，以下分别是 2008 年、2009 年基于 Rook 权重矩阵全局 Moran’s I 散点图。

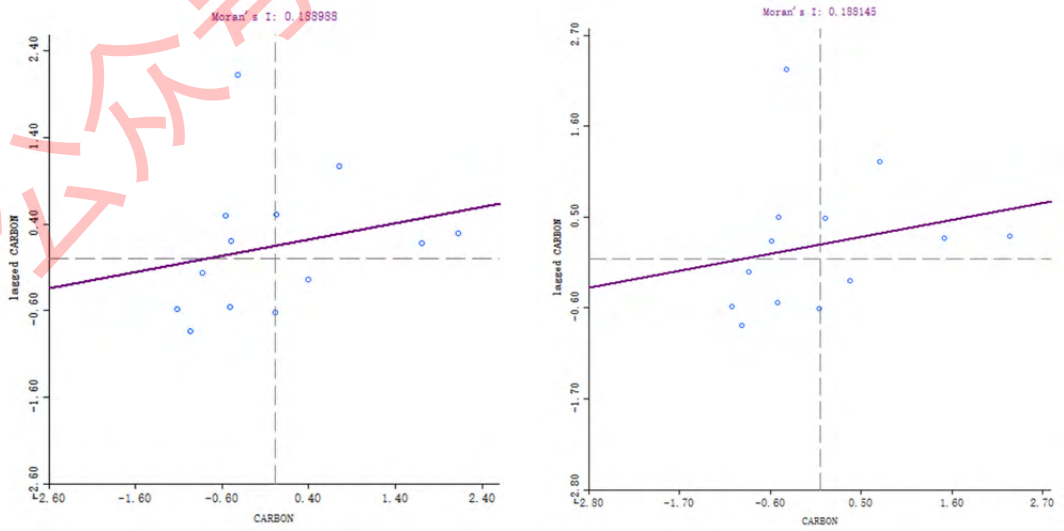


图 5　基于 Rook 权重矩阵全局 Moran’s I 散点图（2008、2009 年）

由上图可知，2008、2009 年京津冀及其周边共十三个省，基于 Rook 权重矩阵全局 Moran's I 指数分别为 0.188988、0.188145，可以认为包括京津冀在内的十三个省的排碳溢出可能具有正向空间相关性。以 2008 年数为例，全局 Moran's I 散点图从 H-H、H-L、L-H 以及 L-L 四个角度分析了指标$CARB_{it}$在空间上的相关性分布，结果如下表所示：

表 10　2008 年排碳溢出指标地区分布情况

| 象限 | 基于邻接矩阵的地区分布 |
|---|---|
| 第一象限<br>(High-High) | 北京、黑龙江、陕西 |
| 第二象限<br>(Low-High) | 河北、辽宁、山西、内蒙古 |
| 第三象限<br>(Low-Low) | 河南、安徽、湖北、江苏 |
| 第四象限<br>(High-Low) | 天津、山东 |

对于邻近年份的每万人碳排放量指标进行空间自相关检验，以下分别是 2016 年、2017 年基于 Rook 权重矩阵全局 Moran's I 散点图。

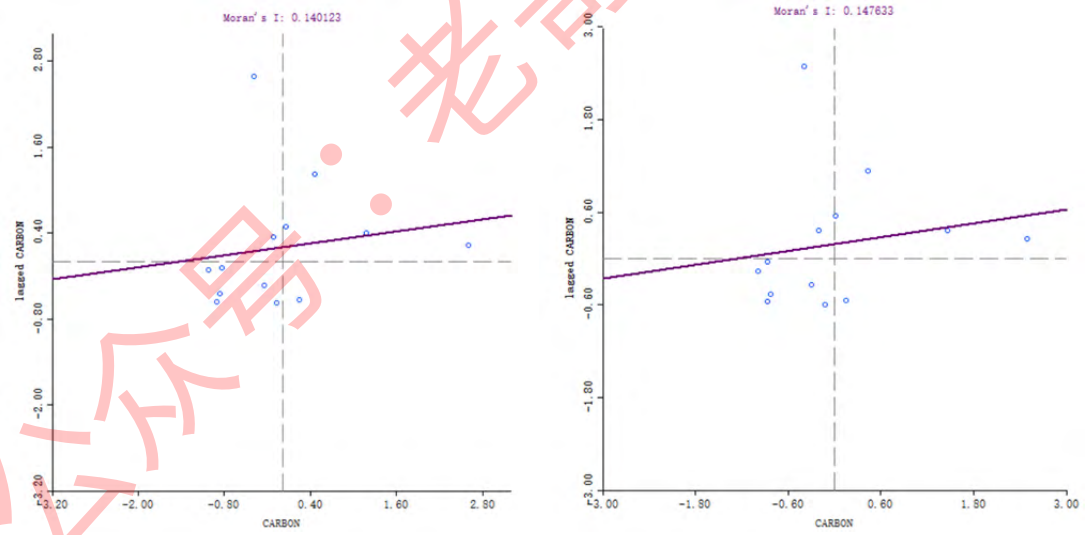


图 6　基于 Rook 权重矩阵全局 Moran's I 散点图（2016、2017 年）

由上图可知，2016、2017 年京津冀及其周边共十三个省，基于 Rook 权重矩阵全局 Moran's I 指数分别为 0.140123、0.147633，可以认为包括京津冀在内的十三个省的排碳溢出可能具有正向空间相关性。以 2017 年数为例，指标$CARB_{it}$在空间上的相关性分布如下表所示：

表 11　2017 年排碳溢出指标地区分布情况

| 象限 | 基于邻接矩阵的地区分布 |
|---|---|
| 第一象限（High-High） | 黑龙江、河北 |
| 第二象限（Low-High） | 陕西、辽宁、山西、内蒙古 |
| 第三象限（Low-Low） | 河南、北京、安徽、湖北、江苏、天津 |
| 第四象限（High-Low） | 山东 |

2008 年～2019 年京津冀及其周边共十三个省，基于 Rook 权重矩阵全局 Moran’s I 指数如下表所示。

表 12　2008 年～2019 年十三省全局 Moran’s I 指数

| 年份 | 全局 Moran’s I　指数 |
|---|---|
| 2008 | 0.188988 |
| 2009 | 0.188145 |
| 2010 | 0.162575 |
| 2011 | 0.127526 |
| 2012 | 0.159124 |
| 2013 | 0.147271 |
| 2014 | 0.13481 |
| 2015 | 0.148198 |
| 2016 | 0.140123 |
| 2017 | 0.147633 |

利用局部 Moran’s I 指数，绘出 LISA 显著性集聚图，如下所示：

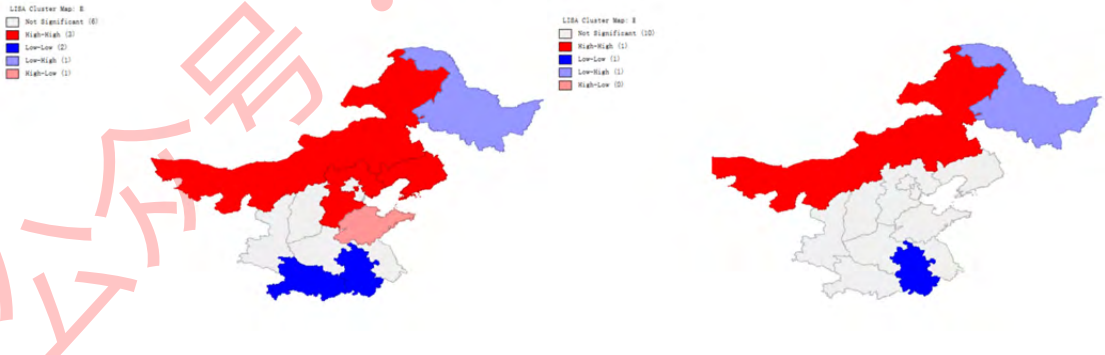

图 7　2008 年、2017 年 LISA 显著性集聚图

由 2008 年的 LISA 图可知，空间相关性较显著的 7 个省份中，有 3 个是“高-高集聚”，分别是东部的内蒙古、河北和辽宁。“低-低集聚”的包括 2 个省份，分别是湖北和安徽。到了 2017 年，排碳溢出效应不及 2008 年明显。

## （二）地理加权回归系数解读

通过 CV 准则得到最优窗宽为 h=0.128，用地理加权回归得到京津冀地区的模型系数如下表：

表 13　模型系数

| 地区 | 截距项 | 人均 GDP | 人均民用汽车 | 城市化水平 | 地区总人口 | 第二产业占比 |
|---|---|---|---|---|---|---|
| 北京 | 36.47 | -0.00002 | 15.53 | -24 | -0.0048 | -5.671 |
| 天津 | -168.2 | -0.000065 | 25.07 | 278.6 | -0.036 | 16.07 |
| 河北 | -45.08 | 0.000555 | -154.2 | 106.5 | -0.00088 | 25.3 |

所以，我们得到京津冀三个地区的地理加权回归方程如下：

$$\begin{cases} CARB_{1t} = 36.47 - 0.00002W_{1t} + 15.53V_{1t} - 24U_{1t} - 0.0048S_{11t} - 5.671R_{1t} \\ CARB_{2t} = -168.2 - 0.000065W_{2t} + 25.07V_{2t} + 278.6U_{2t} - 0.036S_{12t} + 16.07R_{2t} \\ CARB_{3t} = -45.08 + 0.000555W_{3t} - 154.2V_{3t} + 106.5U_{3t} - 0.00088S_{13t} + 25.3R_{3t} \end{cases}$$

其中，$CARB_{1t}$，$CARB_{2t}$，$CARB_{3t}$北京、天津、河北的碳排放量。

由此可得，各指标对京津冀三地的人均碳排放量存在显著差异。北京和天津人均碳排放量与人均 GDP 呈负相关，河北人均碳排放与人均 GDP 呈正相关。北京和天津的人均民用汽车量对人均碳排放量有显著正影响，河北的人均民用汽车量对人均碳排放量有显著负影响。天津和河北的城市化水平和第二产业占比对人均碳排放量有显著正影响。

## 六、结论与建议

### （一）结论

北京和天津人均碳排放量与人均 GDP 呈负相关，而河北人均碳排放与人均 GDP 呈正相关，人均 GDP 每增加 1 元，人均碳排放量增加 0.000555 万吨。这是因为河北属于重点能源消耗型城市，其经济相对于北京和天津来说发展较缓慢，因此河北在经济建设时对环境的影响程度更大。

北京和天津的人均碳排放随着人均民用汽车量的增加而增加，且天津的增速大于北京，人均民用汽车量每增加 1 万辆，天津人均碳排放量就会增加 25 万吨，河北人均碳排放量就会增加 15.53 万吨。这是因为北京和天津城市化水平较高，科技和经济发展较快，更多的高收入人群聚集于此，人均民用汽车拥有量较高，因此对人均碳排放量影响较大。另外，由于北京出行限号等相关政策对汽车出行量限制更强，所以北京人均民用汽车对人均碳排放量的增幅相对天津来说更小。

天津与河北的城市化水平对人均碳排放量的影响最大，且呈正相关，而北京的城市化水平与人均碳排放量呈负相关，城市化水平每提高 1%，北京的人均碳排放量会下降 24 万吨，天津和河北人均碳排放量分别增加 278.6 万吨和 106.5 万吨。这是因为天津与河北城市化相对北京来说较落后，城市化仍有较大发展空间，此时两地城市化进程对环境的影响更大，而北京的城市化水平原本较高，当地居民保护环境的意识较强，且有更优的科技手段来保护环境，综合来说对环境的影响较小。另外作为首都，北京市会更加注重保护环境，投入更多的财力和物力保护与改善环境。

三个地区的总人口数对人均碳排放量的影响都不大。天津与河北的第二产业占比对人均碳排放量有很大的影响，第二产业占比越大，人均碳排放量越高。河北与天津注重工业发展，所以第二产业占比较高，导致能源的消耗很高。而北京

重点在于科技与创新，第二产业能源消耗相对较小，另外将科技应用在环境保护中大大提高了节能减排效率。

## （二）建议

基于京津冀三地人均碳排放量的空间差异性，本文针对不同地区提出了以下相关建议：

1．鼓励天津河北，创新产业结构

当今世界能源紧缺，环境问题显著，我国的产业结构复杂，河北和天津注重第二产业发展，但这也带来了很多环境问题。低能耗、低污染、高产出是文化产业的优势，它对我国产业结构的创新有积极作用。北京文化底蕴丰厚、资源丰富，而京津冀三地地缘相接，人缘相亲，这有助于文化产业的协同发展。从国家层面应当给予大力的鼓励和政策扶持，鼓励其文化产业的发展。

2.针对北京和天津，优化车辆出行政策

尽管北京有限号出行政策，但是在高峰期堵车现象依然严峻。堵车的同时汽车排放的气体对环境的污染非常严重，因此相关部门应当优化交通路线，减少车辆重复路段。另外鼓励市民使用新能源汽车，加强对汽车出行的管控，尾气排放不合格的汽车严格控制出行。

3.鼓励碳的资源化发展

过量排放二氧化碳会导致许多环境问题，但是它是光合作用必不可少的一环，可以用于农业、消防、人工降雨等诸多领域。因此将碳“存储”起来，用在有需求的地方是一个可行手段。国家应当积极鼓励对二氧化碳资源化深层探索，大力支持相关环保科技科研机构的研发，为碳资源化提供“智力保障”，同时鼓励国家相关领域人才在碳资源化方向的研究。

4.加强发展碳市场，征收碳排放税

碳市场不仅可以帮助解决高排碳行业的碳排放压力，同时可以增加碳排量低的地区的经济收入。另外还可以刺激碳咨询、碳金融以及新能源汽车行业的发展，间接带动碳资源化等节能技术的普及，国家可以通过征收碳排放税等手段协调碳排放和经济之间的平衡，绿色发展经济。

5.采取清洁能源，提高能源的利用率

除了保护环境，增加植被覆盖率之外，减排才能从根本上解决问题。优化能源消费结构，使用含碳量低的能源。如用天然气代替煤炭，尽可能使用风能、太阳能等可再生能源，绿色发展经济，减少对环境的损害。另外，倡导全社会节能减排，呼吁每个人改掉不良习惯，为实现低碳生活贡献一份力量。

6．不足与改进

在实际应用中，并不是所有参数都随着空间位置的变化而变化，某些参数在空间上可能是不变的，此时的模型中只有部分参数是变化的，因此模型既包含变参数又包含常参数，因此在未来，对于空间上碳排放的研究还可利用混合地理加权模型进行建模分析。

# 参考文献

[1]林伯强,刘希颖.中国城市化阶段的碳排放:影响因素和减排策略[J].经济研究,2010,45(08):66-78.

[2]王毅萌.城市交通温室气体核算与减排潜力研究[D].河北工程大学,2020.

[3]佟新华,周红岩,陈武,段志远,徐梦鸿,段海燕.工业化不同发展阶段碳排放影响因素驱动效应测度[J].中国人口·资源与环境,2020,30(05):26-35.

[4]EHRLICH P R, HOLDREN J P. Impact of population growth[J].Science, 1971(171).

[5]COMMONER B. Economic growth and ecology — A biologist's view[J].Monthly Labor Review, 1971(94).

[6]赵荣钦,黄贤金,钟太洋,揣小伟.区域土地利用结构的碳效应评估及低碳优化[J].农业工程学报,2013,29(17):220-229.

[7]周迪,罗东权.绿色税收视角下产业结构变迁对中国碳排放的影响[J].资源科学,2021,43(04):693-709.

[8]杨夏星.武汉城市空间形态与碳排放的关系研究[J].当代经济,2020(03):81-83.

[9]闫树熙,陈璐.交通碳排放影响因素分析：以西安市为例[J].统计与决策,2020,36(04):62-66.

[10]赵小曼,张帅,袁长伟.中国交通运输碳排放环境库兹涅茨曲线的空间计量检验[J].统计与决策,2021,37(04):23-26.

[11]苑清敏,张宝荣,李健.京津冀地区工业碳排放影响因素的门限效应分析[J].环境科学与技术,2019,42(11):213-221.

[12]候勃,岳文泽,王腾飞.中国大都市区碳排放时空异质性探测与影响因素——以上海市为例[J].经济地理,2020,40(09):82-90.

[13]杨世杰.中国省域能源消耗碳排放的空间效应研究：基于不同空间权重矩阵视角[J].环境科学与技术,2019,42(S2):180-185.

[14]Wensong Su,Yanyan Liu,Shaojian Wang,Yabo Zhao,Yongxian Su,Shijie Li. Regional inequality, spatial spillover effects, and the factors influencing city-level energy-related carbon emissions in China[J]. Journal of Geographical Sciences,2018,28(4).

[15] Li, L. , X. Hong , and S. School . "Spatial Effects of Energy-Related Carbon Emissions and Environmental Pollution——STIRPAT Durbin Model Based on Energy Intensity and Technology Progress." Journal of Industrial Technological & Economics(2017).

[16]陈志建,王铮.中国地方政府碳减排压力驱动因素的省际差异——基于STIRPAT模型[J].资源科学,2012,34(04):718-724.

[17]王亚政. 基于 ESDA-GWR 的中国区域碳排放空间差异研究[D].天津大学,2017.

[18] 覃文忠. 地理加权回归基本理论与应用研究[D].同济大学,2007.

[19]梅长林,王宁.近代回归分析方法[M]:北京：科学出版社,2012.

[20]张日权, 卢一强.变系数模型[M]北京：科学出版社, 2004.

[21]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.

[22]张连发. 基于流数据的地理加权回归建模方法的研究[D].武汉大学,2019.

[23]冯三营. 一类变系数模型的统计方法与理论研究[D].北京工业大学,2015.

[24]谢琍. 空间自回归模型的统计推断理论、方法与应用[D].北京工业大学,2019.

[25]Shen Si Lian and Cui Jian Ling and Wu Xin Qian. A simple test for spatial heteroscedasticity in spatially varying coefficient models[J]. Journal of Statistical Computation and Simulation, 2021, 91(8) : 1580-1592.

# 附录

附录一：其余结果

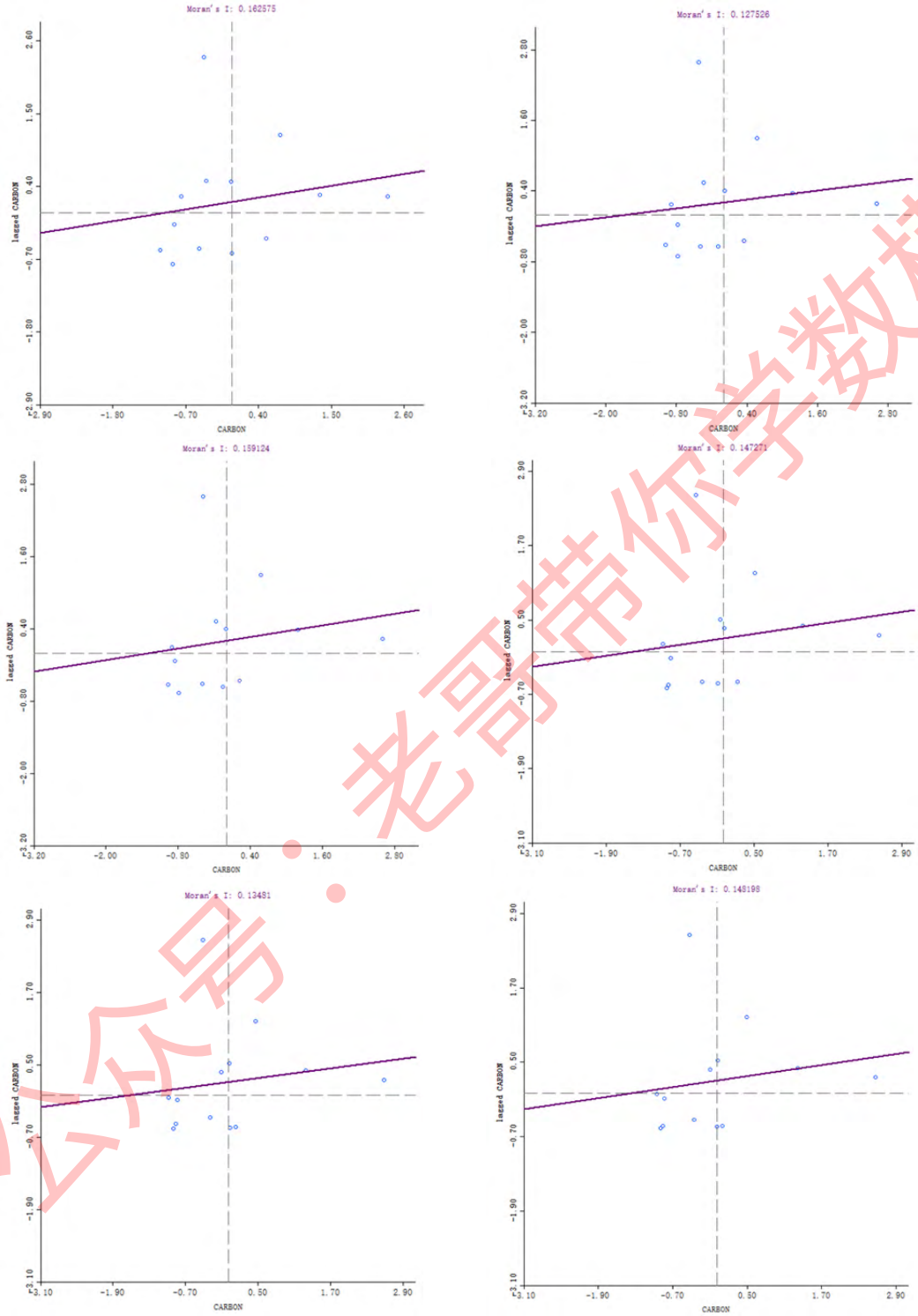


图 8　基于 Rook 权重矩阵全局 Moran’s I 散点图（2010 年～2015 年）

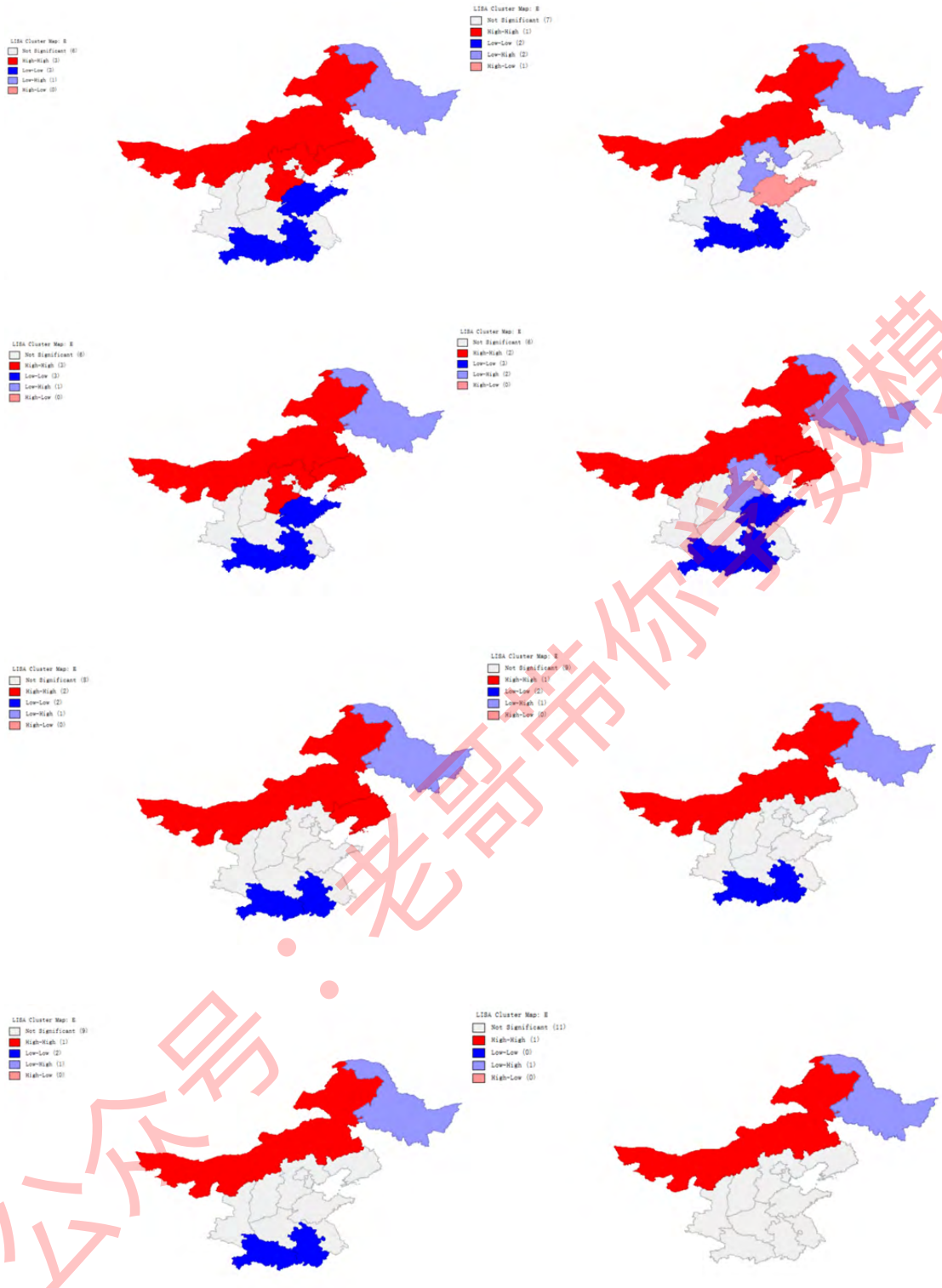

图 9　2009～2016 年 LISA 显著性集聚图

## 致谢

首先感谢举办方组织这次竞赛，让我们有了学习与展示的机会；然后感谢学校为我们提供了舒适的竞赛环境；从确定选题到撰写报告，离不开老师的指导，在此十分感谢我们的老师；感谢队友之间相互鼓励，我们一起克服了很多困难；并且也很感谢相关研究方向的学者，在完成论文的过程中，我们团队从他们的研究成果中得到很多启发。

本次统计建模竞赛接近尾声，在奋力完成此次比赛的过程中，带给我们的是满满的收获：我们的文献阅读与快速学习能力有所提高；数据收集与使用考虑更加全面；理论学习与推导能力有所加强；更加理解了论文质量与美观的重要性。这些收获都将继续激励我们奋发图强、迎难而上。