

Research on the Spatial and Temporal Distribution Characteristics and Dynamic Trend of Air Pollution in China

Rongli Li

School of Mathematics, Yunnan Normal University, Kunming, Yunnan, 650500, China

Abstract

With the advancement of industrialization and urbanization, the pollution situation in China is becoming more and more serious, and shows obvious regional characteristics. In this paper, $PM_{2.5}$ is selected as the index to measure the pollution situation. We study the monthly $PM_{2.5}$ concentration data of 282 cities. The Fourier basis function is used to fit the data into a functional curve. Then the functional principal component analysis is carried out in 282 cities according to seven geographical regions in China. The results show that the 282 cities can be divided into two principal components. According to the $PM_{2.5}$ pollution situation in the seven major geographical regions, it can be divided into three categories: low pollution in the South China region, Southwest region, and Northeast region; mild pollution in the North China region and Northwest region; and high pollution in the East China region and Central China region. The research results can provide important reference information for pollution prevention and control in various regions.

Keywords

$PM_{2.5}$ concentration; basis function; functional data; functional principal component analysis

全国空气污染状况的时空分布特征及动态趋势研究

李荣丽

云南师范大学数学学院, 中国·云南昆明 650500

摘要

随着工业化和城镇化的发展, 中国污染情况日趋严重且呈现出区域特征。论文选取 $PM_{2.5}$ 浓度衡量污染情况, 基于全国282个城市的 $PM_{2.5}$ 浓度月度数据, 用傅里叶基函数将上述数据拟合为函数曲线, 并按七大地理分区进行函数型主成分分析。研究表明全国282个城市可划为2个主成分, 且 $PM_{2.5}$ 的污染情况按七大地理分区可分为三类, 第一类是低污染地区, 包括华南地区、西南地区和东北地区; 第二类是轻度污染地区, 包括华北地区、西北地区, 第三类是高污染地区, 包括华东地区、华中地区。研究结果将为各地区污染防治提供重要参考信息。

关键词

$PM_{2.5}$ 浓度; 基函数; 函数型数据; 函数型主成分分析

1 引言

大气污染事件频发已经引起全球的高度重视, 根据生态环境部消息, 北京、长三角区域、汾渭平原等众多地区的首要污染物均为 $PM_{2.5}$ 。探究全国多个城市的 $PM_{2.5}$ 浓度动态变化、对 $PM_{2.5}$ 浓度数据进行函数型主成分分析, 将更好地了解各地区首要污染物浓度的分布情况, 为大气污染治理工作建言献策。目前国内关于大气污染问题的研究内容主要分为以下三个方面: 对大气污染的时空特征进行研究^[1,2], 对大气污染的未来趋势进行预测^[3]。近几十年来, 由于统计学的不断发展, 函数型数据分析方法逐渐受到关注。目前函数型数据分析方法主要运用于金融股票研究方

面和社会经济研究方面, 但也有部分学者运用函数型数据分析方法对大气污染进行分析^[4,5]。目前还没有学者利用函数型数据主成分分析方法对中国七大地理分区的 $PM_{2.5}$ 浓度进行研究。基于此, 论文将基于 $PM_{2.5}$ 浓度的数据, 运用函数型数据分析方法, 将282个城市的 $PM_{2.5}$ 浓度共108个月的数据拟合为函数曲线, 分析 $PM_{2.5}$ 浓度动态变化特征, 并对 $PM_{2.5}$ 浓度曲线族运用函数型数据进行主成分分析方法对七大地区 $PM_{2.5}$ 大气污染特征进行分类。

2 数据说明与方法

2.1 数据说明及基函数

论文选取全国282个地级市的 $PM_{2.5}$ 浓度监测数据作为研究对象, 整理了282个城市从2015年1月到2023年12月的 $PM_{2.5}$ 浓度月度数据, 数据来自中国空气质量在线监测分析平台。

【作者简介】李荣丽(1982-), 女, 中国云南丽江人, 博士, 副教授, 从事时间序列分析、纵向数据分析研究。

论文采用傅里叶基函数方法拟合观测数据，因为该数据具有周期性。傅里叶基函数拟合方法如下：选择一组基函数的线性组合，估计 $x_i(t)$ 的值：

$$x_i(t) = \sum_{k=1}^K b_{ik} \phi_k(t) \quad (1)$$

其中， $x_i(t)$ 是第 i 个样本拟合的曲线； $\phi_k(t)$ 是第 k 个基函数； b_{ik} 是与之相对应的系数。通过最小二乘法得到的估计值，从而拟合曲线 $x_i(t)$ 。

2.2 函数型主成分分析模型

假设得到拟合曲线 $x_i(t) (t \in T, i = 1, 2, \dots, n)$ ，每条函数曲线的主成分得分为：

$$\tau_i = \int \alpha(t) x_i(t) dt, i = 1, 2, \dots, n$$

其中， $\alpha(t)$ 是权重函数，因此求解函数曲线族的第一主成分就变成了求解带有如下所示约束条件的问题：

$$\begin{cases} \max \frac{1}{n} \sum_{i=1}^n \tau_i^2 \max \frac{1}{n} \sum_{i=1}^n \left(\int \alpha(t) x_i(t) dt \right)^2 \\ \text{s.t. } \|\alpha\|^2 = \int \alpha(t) \alpha(t) dt = 1 \end{cases} \quad (2)$$

通过约束条件的优化问题，就可以得到第一主成分 $\alpha_1(t)$ 。同理，可求得第 k 个主成分，在前 $k-1$ 个主成分权重函数互相垂直的条件下，求解上述优化问题，即：

$$\begin{cases} \max \frac{1}{n} \sum_{i=1}^n \tau_i^2 \max \frac{1}{n} \sum_{i=1}^n \left(\int \alpha(t) x_i(t) dt \right)^2 \\ \text{s.t. } \|\alpha\|^2 = \int \alpha(t) \alpha(t) dt = 1 \\ \int \alpha(t) \alpha_l(t) dt = 0, l = 1, 2, \dots, k-1 \end{cases} \quad (3)$$

我们可以通过拟合曲线的协方差函数矩阵，求解函数型主成分的权重函数 $\alpha(t)$ 。记协方差函数为：

$$v(s, t) = \frac{1}{n-1} \sum_{i=1}^n (x_i(s) - \bar{x}(s))(x_i(t) - \bar{x}(t))$$

则权重矩阵 $\alpha(t)$ 满足特征方程：变换得：

$$V\alpha(t) = \int v(s, t) \alpha(t) dt$$

其中 V 为协方差算子，且有 $V\alpha(t) = \lambda\alpha(t)$ 。设函数 $x_i(t)$ 的基函数展开式为：

$$x_i(t) = \sum_{k=1}^K b_{ik} \phi_k(t)$$

令函数向量 $X(t) = (x_1(t), x_2(t), \dots, x_n(t))'$ ， $\Phi(t) = (\phi_1(t), \phi_2(t), \dots, \phi_n(t))'$ ，则所有曲线的基函数展开式为 $x=A\Phi$ ，协方差函数的矩阵形式为：

$$v(s, t) = \frac{1}{n-1} X'X = \frac{1}{n-1} \Phi(s)'A'A\Phi(t)$$

现设函数 $\alpha(t)$ 的基函数展开式为：

$$\alpha(t) = \sum_{k=1}^K a_k \phi_k(t)$$

其中： $a = (a_1, a_2, \dots, a_k)'$ ，则上式可写为矩阵形式 $\alpha(t) = \Phi(s)'a$ ，从而得：

$$\int v(s, t) \alpha(t) dt = \int \frac{1}{n-1} \Phi(s)'A'A\Phi(t)\Phi(s)'adt = \frac{1}{n-1} \Phi(s)'A'AZa$$

其中 Z 为 K 阶对称矩阵，则有：

$$\frac{1}{n-1} \Phi(t)'A'AWa = \lambda\Phi(t)'a$$

对于 $\forall t \in T$ ，式 (4) 均成立，故 $\frac{1}{n-1}A'AWa = \lambda a$ ，即可求解出权重函数 $\alpha(t)$ 。

3 实证分析

3.1 全国 282 个城市浓度的实证分析

绘制全国 282 个城市的折线图，通过曲线拟合方法，基于 R 绘制出函数化的全国 282 个城市 $PM_{2.5}$ 浓度 2015 年 1 月至 2023 年 12 月的变化曲线如图 1 所示。

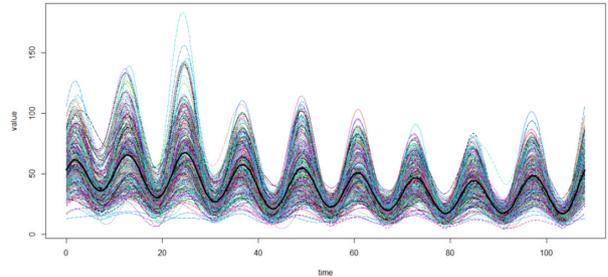


图 1 全国 282 个城市 $PM_{2.5}$ 浓度的曲线拟合图

黑色粗线条为整体均值线，从图中可看出 282 个城市的 $PM_{2.5}$ 浓度在 2015 至 2017 呈现上升趋势、2017 年至 2023 年波动范围整体相对稳定无较大变动，说明全国 $PM_{2.5}$ 浓度污染情况在 2017 年最为显著。在 2015 年至 2023 年大部分城市的 $PM_{2.5}$ 浓度差距在每年的冬春时期最大、夏秋时期较小，这说明大部分城市的 $PM_{2.5}$ 浓度呈现季节性波动， $PM_{2.5}$ 浓度变化与季节变化高度相关；其次每个城市的 $PM_{2.5}$ 浓度月变化在 9 年间形成波峰、波谷，大致形成 9 个波峰、9 个波谷，且绝大多数城市在年初出现波峰、年中出现波谷。就全国 282 个城市的整体波动趋势而言，冬春期间全国绝大多数城市大气污染最为严重、空气质量最差，在夏秋期间全国绝大多数城市大气污染最轻、空气质量最好。

通过函数型主成分分析看出全国 282 个城市 $PM_{2.5}$ 浓度曲线族的前 2 个主成分的贡献率达（超过），可对 282 个城市进行较全面的解释，因此选用两个主成分来分析全国 282 个城市 $PM_{2.5}$ 浓度的变化情况。

函数型主成分分析贡献率和累计贡献率见表 1。

表 1 函数型主成分分析贡献率和累计贡献率

函数型主成分	FPC1	FPC2
贡献率 (%)	82.8	3.9
累计贡献率 (%)	82.8	86.7

第一个函数型主成分表明 $PM_{2.5}$ 整体污染最严重，整体无明显增减趋势，在 2015 年至 2023 年期间呈季节性规律

波动,波峰出现在每年的冬春时期;第二个函数型主成分 $PM_{2.5}$ 在 2015 年至 2017 年污染较为严重,2017 年至 2023 年污染较轻;且整体浓度在 2015 年至 2023 年间有递减趋势,波峰正好与第一主成分相反,出现在夏秋时期。说明在 2017 年至 2023 年期间,全国一部分城市 $PM_{2.5}$ 浓度在每年的冬春高、夏秋低,一部分城市是冬春低、夏秋高。由于第一主成分和第二主成分的贡献率分别为 82.8%、3.9%,共同描述了数据中的主要变化方向,所以选用二维平面直观展示不同城市在两个方向上的得分情况,绘制函数型主成分分析的第一主成分和第二主成分得分图如图 2 所示。

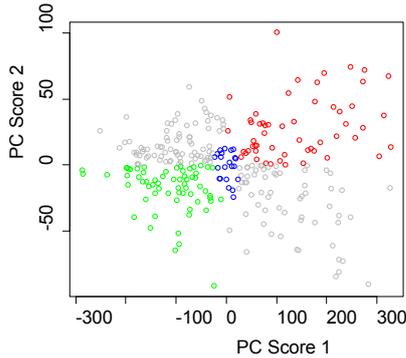


图 2 全国第一和第二函数型主成分得分散点图

根据如上所示的得分图,可以将全国 282 个城市分为四类。

①第一类红色部分:北京市等 55 个城市两个主成分得分均为正,说明这些城市 $PM_{2.5}$ 浓度高于全国的平均水平,大气质量较差,尤其是河北石家庄、山东聊城等七个城市的两个主成分得分之和均超过 300,在 282 个城市中排名前七,说明它们的 $PM_{2.5}$ 浓度在 2015 年至 2023 年期间都很高,是大气质量最差的城市。

②第二类蓝色部分:兰州市等 22 个城市的第一主成分和第二主成分均在零点附近,两个主成分得分与零点的差值均不大于 25,这说明这些城市 $PM_{2.5}$ 浓度接近全国的平均水平,大气质量居中。

③第三类绿色部分:盘锦市等 64 个城市第一主成分与第二主成分得分均为负,说明这些城市 $PM_{2.5}$ 浓度低于全国的平均水平,大气质量较好,尤其是海南三亚、云南丽江、福建龙岩等七个城市的两个主成分得分之和均小于 -200,在 282 个城市中排名倒数七位,这说明这七个城市的 $PM_{2.5}$ 浓度在 2015 年至 2023 年期间都很低,是全国大气质量最好的城市。

④第四类灰色部分:六安市等 141 个城市第一主成分与第二主成分均为一正一负,第一主成分与第二主成分差值的绝对值均大于 25,说明这 141 个城市的 $PM_{2.5}$ 浓度在 2015 年至 2023 年期间变化较大。其中西藏拉萨、云南保山等 5 个城市的差值均小于 -200,说明这 5 个城市大气质量

较好、变化大、不稳定;陕西咸阳等四个城市的差值均大于 300,说明大气质量较差、变化大、不稳定。

3.2 中国七大地理分区 $PM_{2.5}$ 浓度的实证分析

为进一步探究全国 $PM_{2.5}$ 浓度的变化特征,根据中国七大地理分区各区的 $PM_{2.5}$ 浓度进行函数型主成分分析。分别绘制东北地区 36 个城市、华北地区 28 个城市、华东地区 76 个城市、华南地区 37 个城市、华中地区 42 个城市、西北地区 30 个城市、西南地区 33 个城市的折线图,再对各区的 $PM_{2.5}$ 浓度曲线进行拟合,并对七大区的曲线族做函数型主成分分析,得到各区的主成分函数图像和主成分得分图,以下以华中地区的 42 个城市为例,其他地区都是类似进行分析。

华中地区 42 个城市 $PM_{2.5}$ 浓度的曲线拟合图见图 3。

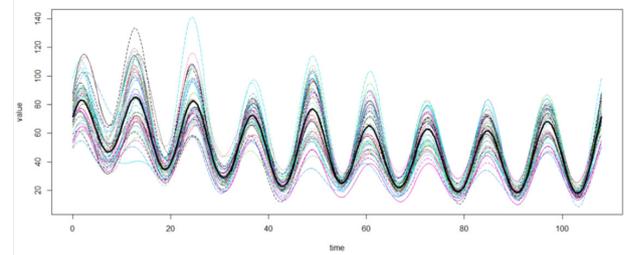


图 3 华中地区 42 个城市 $PM_{2.5}$ 浓度的曲线拟合图

图 4 为华中地区的函数型主成分分析的主成分得分图,根据每个地区的得分图可将各个地区城市 $PM_{2.5}$ 污染分为三类,红色代表该区内 $PM_{2.5}$ 高污染城市,绿色代表该区内 $PM_{2.5}$ 低污染城市,灰色代表该区内的 $PM_{2.5}$ 污染变化大的城市。

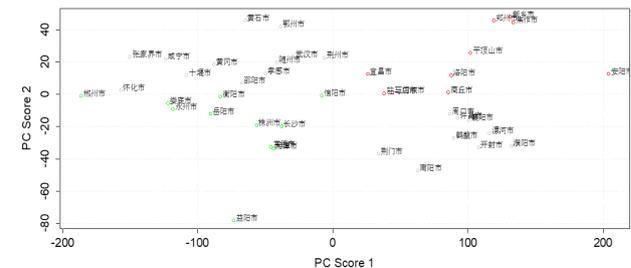


图 4 华中地区 42 个城市第一和第二函数型主成分得分图

4 结论与建议

4.1 结论

论文运用函数型主成分分析方法,对全国 282 个城市的 $PM_{2.5}$ 浓度数据进行分析,结果表明全国 282 个城市可化成 2 个主成分,第一个函数型主成分 $PM_{2.5}$ 整体污染最严重、无明显增减趋势;第一主成分在 2015 年至 2023 年期间呈季节性规律波动,波峰出现在每年的冬春时期;同时第一个函数主成分整体较均值变化较大,在 2016 年至 2017 年间 $PM_{2.5}$ 浓度偏离均值程度最大,2019 年至 2020 年间 $PM_{2.5}$ 浓

度偏离均值程度最小。第二个函数型主成分 $PM_{2.5}$ 整体污染在 2015 年至 2017 年污染较为严重且在 2015 年至 2023 年间有递减趋势,波峰正好与第一主成分相反,出现在夏秋时期;且第二个函数主成分整体较均值变化较小,除均值波峰、波谷处差距较大外其余部分与均值基本重合,2017 年至 2023 年间在每个均值波峰处往负向偏离。

北京市等 55 个城市两个主成分得分都为正,说明这些城市 $PM_{2.5}$ 浓度高于全国的平均水平,大气 $PM_{2.5}$ 污染较为严重;尤其是河北保定、河北邢台、河北衡水、河北石家庄、山东聊城、山东德州、河北邯郸七个城市的两个主成分得分之和超过 300,在 282 个城市中排名前七,这说明七个城市的 $PM_{2.5}$ 浓度在 2015 年至 2023 年期间都很高,是全国大气 $PM_{2.5}$ 污染最严重的城市。盘锦市等 64 个城市第一主成分与第二主成分得分均为负,说明这些城市 $PM_{2.5}$ 浓度低于全国的平均水平,大气 $PM_{2.5}$ 污染较低;尤其是海南三亚、云南丽江、海南海口、黑龙江伊春、云南玉溪、云南普洱、福建龙岩七个城市的两个主成分得分之和小于 -200,在 282 个城市中排名倒数七位,这说明这七个城市的 $PM_{2.5}$ 浓度在 2015 年至 2023 年期间都很低,是全国大气 $PM_{2.5}$ 污染最轻的城市。

从全国七大地理分区看,华南地区、西南地区、东北地区大部分城市的 $PM_{2.5}$ 浓度在 $75\mu\text{g}/\text{m}^3$ 范围内波动,少数城市在 $35\mu\text{g}/\text{m}^3$ 范围内波动,说明这些地区的大部分城市的 $PM_{2.5}$ 污染程度在轻度污染以内,少数城市浓度达到了空气质量优级的要求。华北地区和西北地区一部分城市的 $PM_{2.5}$ 浓度在 $75\mu\text{g}/\text{m}^3$ 范围外波动,说明华北地区和西北地区有部分城市 $PM_{2.5}$ 污染程度达到轻度污染。华东地区和华中地区大多数城市 $PM_{2.5}$ 浓度的波动范围超过 $75\mu\text{g}/\text{m}^3$,说明华东地区和华中地区 $PM_{2.5}$ 的污染等级在轻度污染以上的城市占

大多数。根据整体 $PM_{2.5}$ 污染情况可分为三类,第一类是低污染:华南地区、西南地区和东北地区,第二类是轻度污染:华北地区、西北地区,第三类是高污染:华东地区、华中地区。

4.2 建议

针对河北保定、河北邢台、河北衡水等 $PM_{2.5}$ 浓度排名靠前的城市的高污染城市,应加大治理力度,实施更为严格的环保措施,推广清洁能源、提高能源利用效率等。对于华东地区、华中地区高污染地区,可建立区域性高强度联控机制,对于华北地区、西北地区,轻度污染地区可以建立区域性低强度联防联控机制,对于华南地区、西南地区、东北地区低污染地区,可以建立区域性联防联控机制,区域统一规划、区域协同治理,共同应对 $PM_{2.5}$ 污染问题。同时加强季节性污染防治:针对绝大多数城市冬春季节 $PM_{2.5}$ 污染加重的情况,各地应提前制定应急预案,加强空气质量的监测与预警,可以采取季节性临时性减排措施。夏秋时期虽然污染较轻,但也不能放松警惕,应继续加强污染源的监控与管理,防止污染反弹。

参考文献

- [1] 张红梅.时空变化特征的城市空气污染趋势研究[J].环境科学与管理,2023,48(4):34-38+44.
- [2] Puustinen A, Hämeri K, Pekkanen J, et al. Spatial variation of particle number and mass over four European cities [J]. *Atmospheric Environment*,2007,41(31):6622-6636.
- [3] 余晓美,沈永昌.中国环境保护重点城市空气质量的动态特征分析[J].统计与决策,2019,35(11):91-94.
- [4] 黄艳玲. $PM_{2.5}$ 时空特征解析与函数型数据分析研究[D].南昌:江西财经大学,2023.
- [5] 武祺然,周力凯,孙金金,等.浙江省空气质量变化特征研究——基于函数型数据分析[J].山东大学学报(理学版),2021,56(7):53-64.