



分类号.....

UDC.....

密 级 公 开.....

黑龙江大学

硕士研究生学位论文

基于相关性分析的 PM_{2.5} 影响因素研究

申 请 人:张凯丽

学 号: 2170096

培养单位: 数学科学学院

学科专业: 应用统计

研究方向: 应用统计

指导教师:马维军

完成日期: 2019年3月10日

中文摘要

空气质量关乎着人们的身体健康,PM2.5 是影响空气质量的重要指标,并且其浓度变化受到多种因素的影响。哈尔滨是黑龙江省的省会城市,东北地区中心城市之一,空气质量问题备受政府和市民关心。为此,对哈尔滨 PM2.5 浓度的变化规律、影响因素进行研究显得尤为重要。

首先,研究了哈尔滨 PM2.5 浓度变化的时间规律,发现其浓度具有中午、晚间高,下午低的分布特征。然后分析了天气、风向、气温等气象因素对 PM2.5 浓度的影响,结果表明,不同程度的降雨可以有效地稀释和清除空气中的 PM2.5,西南风向下的 PM2.5 浓度较高,并且当风力达到 8 级时,PM2.5 的浓度最高,PM2.5 浓度变化与气温、相对湿度具有负相关关系,与气压具有正相关关系。接着基于气温、气压、相对湿度、风力建立了 PM2.5 的多元线性回归模型,通过交叉验证发现该模型的预测效果是比较好的。最后,基于不同季节影响,利用 Spearman 秩相关系数分析了 PM2.5 浓度与气象因素之间的相关性,研究表明,季节不同,气象因素对其浓度的影响也有所不同,并且与春、夏、秋三个季节相比较,冬季的 PM2.5 浓度最高。

关键词: PM2.5; 相关分析; 偏相关分析; 多元线性回归模型

Abstract

Air quality has an important impact on people's health and PM_{2.5} is an important indicator of air quality which is affected by many factors. Harbin is the capital of Heilongjiang Province, it is one of the central cities in Northeast China, Harbin's air quality issues are concerned by the government and the public.

First, the time law of PM_{2.5} concentration in Harbin is studied. Its concentration is found to be higher at noon and evening, lower in the afternoon. Then the effects of meteorological factors such as weather, wind direction and temperature on the concentration of PM_{2.5} are analyzed. The results show that different degrees of rainfall have significant dilution and scavenging effects on PM_{2.5}, the concentration of PM_{2.5} in the southwest wind is higher. When the wind reaches 8, the concentration of PM_{2.5} is the highest. The concentration change has a negative correlation with temperature and relative humidity, and it has a positive correlation with air pressure. Based on temperature, pressure, relative humidity and wind, a linear regression model is established about PM_{2.5}. The result of cross-validation shows that the prediction effect of the model is better. Finally, using the Spearman rank correlation coefficient, the correlation between PM_{2.5} concentration and meteorological factors is analyzed. The effects of meteorological factors on PM_{2.5} concentration are different in different seasons, and compared to the other three seasons, the PM_{2.5} concentration of winter is the highest.

Keywords: PM_{2.5}; Correlation analysis; Partial correlation coefficient; Multivariable linear regression model

目录

中文摘要	I
Abstract	II
目录	III
第1章 绪论	1
1.1 本文研究背景与意义	1
1.2 国内外研究现状	2
1.3 本文的主要内容及结构安排	3
1.3.1 本文的主要内容	3
1.3.2 本文结构安排	4
第2章 相关知识简介	5
2.1 相关分析	5
2.1.1 Spearman 相关系数	5
2.1.2 偏相关分析	6
2.1.3 相关性的显著性检验	7
2.2 多元线性回归分析	8
2.2.1 多元线性回归模型	8
2.2.2 多元线性回归模型的基本假定	9
2.2.3 拟合优度	9
2.3 逐步回归	10
2.3.1 逐步回归的基本思路	10
2.3.2 逐步回归的主要过程	10
2.4 多重共线性	12
2.4.1 多重共线性的后果	12
2.4.2 多重共线性的检测	13
2.5 本章小结	13

第3章 哈尔滨 PM2.5 浓度变化的影响因素研究	14
3.1 哈尔滨市数据来源	14
3.2 PM2.5 浓度变化的时间规律	14
3.2.1 PM2.5 浓度的日均值变化	14
3.2.2 PM2.5 浓度的昼夜变化	15
3.2.3 PM2.5 浓度的一周变化	16
3.3 PM2.5 与气象因素之间的关系	17
3.3.1 与天气、风向以及风力的关系	17
3.3.2 与气温、气压以及相对湿度的关系	21
3.3.3 与气温、气压、相对湿度以及风力的偏相关分析	25
3.4 基于主要气象因素的多元线性回归分析	27
3.4.1 模型建立	27
3.4.2 模型检验	29
3.4.3 模型分析	31
3.5 本章小结	31
第4章 气象因素对 PM2.5 浓度影响的季节性差异	32
4.1 PM2.5 的时间分布特征	32
4.1.1 PM2.5 随季节的变化规律	32
4.1.2 PM2.5 随月份的变化规律	33
4.2 天气、风向对 PM2.5 浓度的影响	34
4.3 气温、气压、相对湿度以及风力对 PM2.5 浓度的影响	37
4.4 本章小结	40
结论与建议	41
参考文献	43
致谢	49
攻读学位期间发表的学术论文	50
独创性声明	51

第1章 绪论

1.1 本文研究背景与意义

空气质量对人体健康有着重要的影响，然而随着经济的快速发展和工业的不断进步，当前的空气质量不容乐观。同时大气中的颗粒物 PM2.5 对空气质量产生的重要影响也不容忽视，由于城市人口增长，机动车数量增多，使得 PM2.5 大量增加。为此，对 PM2.5 的影响因素进行研究分析刻不容缓，以便采取相应的措施改善空气质量，满足人民群众的基本要求。

研究发现，影响 PM2.5 浓度的因素有很多，其中气象因素就不容忽视，比如气温、气压、相对湿度以及风向和风力都会影响 PM2.5 的扩散，而且不同的天气状况也会对 PM2.5 产生直接或间接的影响，也就是受不同天气状况的影响，PM2.5 的扩散程度也会不同。因此，通过对 PM2.5 与气象因素之间的相关关系进行分析就显得非常重要。同时，空气质量的好坏也与地域有关，虽然已有学者对主要地区和城市进行了研究分析，但是从统计学的角度，定量分析气象因素如何影响空气质量的研究相对较少。为此，本文基于相关性分析，对影响哈尔滨市 PM2.5 浓度的因素进行研究，这将为哈尔滨市空气质量监控辨别重要影响因素提供理论依据。

哈尔滨是黑龙江省省会，东北地区中心城市之一，省会城市作为全省的龙头和重心，对全省各地区的经济文化发展具有重要的引领作用，因此哈尔滨在发展经济文化的同时还要着重强调对环境的保护。环境是制约一个城市经济文化发展的重要因素，哈尔滨作为省会城市必须时刻保持自身的优良形象，其各方面的发展都具有重要的示范作用，并且良好的发展对促进全省各地区经济快速平稳健康发展具有重要意义。此外，保护环境，改善空气质量也是全面建设小康社会和推进现代化进程中的一项重大任务。国家现在从各个方面支持对空气质量的研究和治理，分析哈尔滨 PM2.5 浓度变化特征，这对如何改善空气质量有着重要的意义。

因为空气中的颗粒物浓度会影响大气的能见度，进而影响着交通出行安全，尤其是 PM2.5，对人们的呼吸系统也有一定的影响，这关乎着人们的身体健康。

近几年来,伴随着日益严重的空气污染问题,国内越来越重视对 PM_{2.5} 的研究,并且取得了一定的成果,但是这些研究大部分集中京津冀、上海和珠三角地区,对哈尔滨市 PM_{2.5} 的研究还比较少,而且目前的研究大多局限于宏观区域上,缺乏对具体问题和具体区域的深入研究。

为此,本文分析了哈尔滨 PM_{2.5} 浓度的变化规律及其与气象因素之间的相关关系,然后结合实际情况分析原因,这将为改善空气质量提供理论依据。同时,“让数据说话”,这不仅可以让人们对空气污染采取有效防范措施,维护身体健康,还可以为政府有关部门科学治理大气污染提供定量依据,具有重要的借鉴意义和应用价值。

1.2 国内外研究现状

空气质量关乎着城市经济文化的发展,也与人们的身体健康紧密相连,为此对 PM_{2.5} 的研究受到不少国内外学者的青睐,研究内容主要包括颗粒物 PM_{2.5} 的组成、来源、对人体和空气的影响^[1-3]等几大方面。

国外很早就开始对颗粒物 PM_{2.5} 进行研究,尤其美国最先开始研究 PM_{2.5},研究内容也较为深入。美国国家环保局于 1998 年开始监测颗粒物的浓度变化,并逐步完善颗粒物监测系统^[4]。Sawant 等^[5]研究了美国加利福尼亚州南部颗粒物 PM_{2.5} 的化学组成,分析表明 PM_{2.5} 的组成成分中有机物占比达到了 41%,其次是硫占到了 37%。Huy 分析出美国阿拉斯加州费尔班克斯的 PM_{2.5} 浓度在静风、低温、相对湿度比较大的条件下,逆温层越厚越稳定,PM_{2.5} 浓度就越高^[6]。Mejia J F 和 Wraith D^[7]对澳大利亚布里斯班亚热带地区颗粒物浓度的变化趋势进行了分析,提出空气中污染物所形成的二次颗粒物是 PM_{2.5} 的主要成分,二次污染过程导致 PM_{2.5} 浓度升高,从而加重了空气污染。Greene J S 和 Kalkstein L S^[8]研究了美国 4 个城市大气污染和气象因素之间的相关关系,研究表明污染物浓度在很大程度上受气象因素(如风速、相对湿度)的影响。Flocas H 和 Kelessis A^[9]分析出地中海地区的空气污染与当地天气有关。Johnson 等人^[10]研究表明冬季交通对 PM_{2.5} 浓度影响较大,其贡献率高达 30%。

在国内,汪永英等^[11]分析了2014年哈尔滨PM_{2.5}浓度与气象因素之间的相关性,结果表明PM_{2.5}浓度与日均气温、日均最大能见度存在正相关关系,与日均气压、日均风速之间的相关关系不太显著;王薇等研究了2015年哈尔滨供暖期颗粒物浓度与气象因子的关系,分析得出供暖期燃煤数量增加导致颗粒物浓度上升,其浓度变化与能见度、相对湿度存在显著的相关关系^[12];周一敏等^[13]对近年来北京PM_{2.5}浓度的变化特征与边界层气象要素之间的相关关系进行了分析,提出风向、边界层高度、降水以及相对湿度对PM_{2.5}有显著影响;杜宇静等^[14]分析了2016年吉林市冬季和春季的PM_{2.5}浓度与气象因素之间的关系,研究表明PM_{2.5}浓度与日均温度、风速和相对湿度存在相关关系,最后又对这两种不同季节的PM_{2.5}浓度变化规律进行了对比分析;江鸿等^[15]分析了武汉市PM_{2.5}浓度与降水、气温、风速等气象因素之间的相关关系,结果表明PM_{2.5}浓度与月平均气温、月降水量呈负相关,与月平均气压呈正相关,而最大风速和平均相对湿度对PM_{2.5}的影响不稳定。徐敬、董雪玲等^[16,17]研究了气象条件对PM_{2.5}浓度的影响,分析发现PM_{2.5}浓度与气温的相关关系不明显,与风速、气压以及相对湿度存在显著的相关关系。

此外,对PM_{2.5}浓度的影响因素研究^[18-25]以及多元线性回归模型在PM_{2.5}浓度变化方面的应用还有很多^[26-29]。

1.3 本文的主要内容及结构安排

1.3.1 本文的主要内容

本文主要基于相关性分析对哈尔滨市PM_{2.5}浓度变化特征及其影响因素进行研究,研究内容主要包括两大方面:一方面是研究PM_{2.5}浓度变化的影响因素,另一方面是研究PM_{2.5}浓度变化的季节性差异。PM_{2.5}浓度变化的影响因素研究主要包括PM_{2.5}浓度的变化规律、原因分析以及基于气象因素所进行的多元线性回归分析,季节性差异研究涉及到不同季节PM_{2.5}浓度的变化规律以及气象因素对其影响的季节性差异。具体内容如下:

(1) 首先运用统计分析软件R语言对数据进行处理与分析,研究了哈尔

滨 PM2.5 浓度变化的时间规律，主要有日均值变化、昼夜变化以及周一至周日的变化特征。

(2) 从气象因素出发，计算 PM2.5 浓度与气象因素（气温、气压、相对湿度、天气、风向和风力）之间的 Spearman 秩相关系数，分析不同气象因素对 PM2.5 浓度的影响。

(3) 由于 PM2.5 浓度的变化受多个气象因素的影响，而且不同气象因素之间也存在相关关系，故为了全面分析气象因素与 PM2.5 之间的相关性，采用偏相关分析方法研究 PM2.5 与气象因素之间的关系。

(4) 基于上述研究内容，运用回归分析建立了 PM2.5 与主要气象因素的多元线性回归模型，并采用逐步回归方法进行模型选择，然后进行检验、优化。

(5) 因为不同季节的气象因素对 PM2.5 的影响有所不同，为此研究了 PM2.5 浓度的月度、季节变化规律并运用箱线图分析不同季节 PM2.5 浓度与天气、风向和风力之间的相关性。

(6) 根据研究结果，针对哈尔滨大气污染治理提出合理性建议。

1.3.2 本文结构安排

第 1 章介绍了 PM2.5 浓度研究的背景和意义，然后对国内外研究现状进行了阐述，最后说明了研究的主要内容以及本文的结构安排。

第 2 章对课题研究所涉及的一些理论知识进行了说明，主要包括相关分析、偏相关分析和回归分析。

第 3 章基于相关性分析，研究 PM2.5 浓度的变化规律及其与气象因素之间的相关关系并建立了多元线性回归模型。

第 4 章研究了 PM2.5 的月度、季节变化特征以及受气象因素影响的季节性差异。

最后基于论文研究结果，对哈尔滨相关部门科学治理空气污染提出合理性建议。

第2章 相关知识简介

2.1 相关分析

2.1.1 Spearman 相关系数

Spearman 相关系数又称为秩相关系数，Spearman 秩相关系数与分布无关，主要用来衡量两个变量联系的强弱，对应的相关分析称为非参数相关分析^[30]。Spearman 秩相关系数适用于定序变量或者是不符合正态分布假设的数据，它可以不用考虑两个变量的变化情况，能够克服 Pearson 相关系数仅用于两个变量需呈线性相关的缺点，适用范围更广泛^[31]。同时，在变量分布不明确，信息缺乏条件下，Spearman 秩相关系数得出的结论更准确^[32]。

如果两个变量的 Spearman 秩相关系数为 +1 或者 -1，表明这两个变量是完全 Spearman 相关。如果对原始数据 x_i, y_i 进行从大到小排列，并将 x_i, y_i 排列后所处位置的名次记为 R_i, S_i ，得到 n 对秩 $(R_1, S_1), (R_2, S_2), \dots, (R_n, S_n)$ ，这 n 对秩的情况见表 2-1。

表 2-1 x 和 y 的 n 对秩情况Table 2-1 n pairs of ranks for x and y

完全相同的评秩		完全相反的评秩		不完全相同的评秩	
x 的秩 R	y 的秩 S	x 的秩 R	y 的秩 S	x 的秩 R	y 的秩 S
1	1	1	n	1	3
2	2	2	$n-1$	2	5
3	3	3	$n-2$	3	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$n-1$	$n-1$	$n-1$	2	$n-1$	n
n	n	n	1	n	$n-1$

记 d_i 为 x_i 与 y_i 的秩次之差, 如果没有相同的秩次, 那么 Spearman 秩相关系数为:

$$\rho_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (2-1)$$

如果两个变量之间的秩次相同, 那么可以计算秩次之间的 Pearson 相关系数:

$$\rho_s = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2-2)$$

Spearman 秩相关系数的符号反应了变量 X 和 Y 相关关系的方向。如果随着变量 X 的增加, 变量 Y 也在增加, 那么变量 X 和 Y 的 Spearman 秩相关系数符号为正, 相反, 在变量 X 增加的时候, 变量 Y 在减少, 则 Spearman 秩相关系数符号为负。当 Spearman 秩相关系数等于 0 时, 表示随着变量 X 的增加, 变量 Y 并没有随之增加或减少。当变量 X 和 Y 严格单调递增时, Spearman 秩相关系数为 1, 而当变量 X 和 Y 严格单调递减时, Spearman 秩相关系数为 -1。变量 X 和 Y 之间越具有严格单调的函数关系, Spearman 相关系数就越大。

2.1.2 偏相关分析

简单相关分析一般是通过计算变量之间的相关系数来衡量相关程度, 但是在大多数情况下, 变量之间的相关关系往往受很多因素的影响, 有时会因为第三个变量的存在导致相关系数不能正确反映两个变量间的相关关系, 偏相关分析就是在控制其他变量的情况下, 研究这两个变量之间的相关关系^[4]。偏相关分析是在简单相关分析的基础上, 考虑了其他变量对这两个变量相关关系的影响, 也就是在控制其他变量之后, 重新计算这两个变量的相关系数, 即偏相关系数。

在三个变量中, 任意两个变量之间的偏相关系数是在控制剩下的一个变量的作用后计算得到的, 称为一阶偏相关系数, 公式如下:

$$r_{ij \cdot h} = \frac{r_{ij} - r_{ih}r_{jh}}{\sqrt{(1 - r_{ih}^2)(1 - r_{jh}^2)}} \quad (2-3)$$

在公式 (2-3) 中, r_{ij} 为变量 x_i 和 x_j 的简单相关系数, r_{ih} 为变量 x_i 和 x_h 的简单相关系数, r_{jh} 为变量 x_j 和 x_h 的简单相关系数。

在四个变量中，任意两个变量之间的偏相关系数是在控制其余两个变量的影响后计算得到的，称为二阶偏相关系数，公式如下：

$$r_{ij \cdot hm} = \frac{r_{ij \cdot h} - r_{im \cdot h} r_{jm \cdot h}}{\sqrt{(1 - r_{im \cdot h}^2)(1 - r_{jm \cdot h}^2)}} \quad (2-4)$$

显然，由公式 (2-4) 知，二阶偏相关系数是由一阶偏相关系数求得的。

2.1.3 相关性的显著性检验

一、Spearman 秩相关检验步骤如下：

(1) 建立假设

$H_0 : X \text{ 和 } Y \text{ 不相关} \leftrightarrow H_1 : X \text{ 和 } Y \text{ 相关}$

或 $H_0 : X \text{ 和 } Y \text{ 不相关} \leftrightarrow H_1 : X \text{ 和 } Y \text{ 正相关}$

或 $H_0 : X \text{ 和 } Y \text{ 不相关} \leftrightarrow H_1 : X \text{ 和 } Y \text{ 负相关}$

(2) 计算检验统计量

Spearman 秩相关系数是衡量两个样本相关性强弱的重要统计指标，由公式 (2-1) 知，Spearman 秩相关系数的取值范围在 -1 和 1 之间， $\rho_s > 0$ 为正相关， $\rho_s < 0$ 为负相关， $\rho_s = 1$ 为完全正相关， $\rho_s = -1$ 为完全负相关。 $|\rho_s|$ 越接近 1，表示两个样本之间相关性越强。一般情况下，认为 $|\rho_s| > 0.8$ 的样本相关性较好。

(3) 做出决策

当 $\rho_s > \rho_s^\alpha$ 时，拒绝原假设 H_0 ；

当 $\rho_s < \rho_s^\alpha$ 时，不能拒绝原假设 H_0 。

其中， ρ_s^α 是临界值，它是根据样本个数 n 、备择假设以及给定的显著性水平查表得到的。

二、偏相关系数检验过程如下：

(1) 建立原假设 H_0 ：两个变量之间的偏相关系数等于 0。

(2) 使用 t 检验方法，检验统计量如下：

$$t = \frac{\sqrt{n-k-2} \cdot r}{\sqrt{1-r^2}} \quad (2-5)$$

在公式 (2-5) 中, r 表示偏相关系数, n 是样本观测数, k 是可控制的变量个数, $n - k - 2$ 是自由度。

(3) 当 $t > t_{0.05}(n - k - 2)$ 或 $p < 0.05$ 时, 应拒绝原假设。

2.2 多元线性回归分析

回归分析一般可以用来解决以下问题:

(1) 测定变量间的关联程度, 关联程度大的话, 就可以建立对应的回归方程。

(2) 依据一个变量的信息, 可以预测其他变量的取值, 然后使用回归方程验证其准确性。

(3) 可以进行因素分析, 从数据的主从关系来进行分析。

回归分析的基本思想是有时自变量与因变量存在不明显的函数关系, 但可以通过数学建模来找出与之对应的回归方程, 多元线性回归主要是用来分析多个变量之间的关系^[33]。

2.2.1 多元线性回归模型

设因变量 y 与自变量 x_1, x_2, \dots, x_p 的线性回归模型为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (2-6)$$

其中, $\beta_0, \beta_1, \dots, \beta_p$ 是 $p + 1$ 个未知参数, β_0 是回归常数, β_1, \dots, β_p 是回归系数, ε 是随机误差, 当 $p \geq 2$ 时, 称该模型为多元线性回归模型^[34]。

假设有 n 组观测数据 $(x_{i1}, x_{i2}, \dots, x_{ip}; y_i)(i = 1, 2, \dots, n)$, 那么多元线性回归模型可以表示为

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} + \varepsilon_2 \\ \dots\dots\dots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + \varepsilon_n \end{cases}$$

将其写成矩阵形式为

$$y = X\beta + \varepsilon \quad (2-7)$$

其中

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

2.2.2 多元线性回归模型的基本假定

(1) 自变量 x_1, x_2, \dots, x_p 是确定性变量，不是随机变量，并且满足 $\text{rank}(X) = p+1 < n$ ，也就是矩阵 X 的列之间不相关，样本量的个数要大于自变量的个数。

(2) 随机误差项满足零均值和等方差的假定，即

$$\begin{cases} E(\varepsilon) = 0 \\ \text{Var}(\varepsilon) = \sigma^2 \end{cases}$$

随机误差项的均值为 0，意思是假设观测值没有系统误差，在正态假定下，随机误差项不存在序列相关，具有相同的精度。

(3) 正态分布的假定条件为

$$\begin{cases} \varepsilon_i \sim N(0, \sigma^2), i=1, 2, \dots, n \\ \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \text{ 相互独立} \end{cases}$$

一般情况下，对含有 p 个自变量的多元线性回归方程来讲，每个回归系数 β_i 表示在其他自变量保持不变的情况下，自变量 x_i 每增加一个单位，因变量 y 的平均增加（减少）程度。

2.2.3 拟合优度

拟合优度一般用来检验回归模型的拟合程度，在进行多元线性回归时，样本决定系数为

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (2-8)$$

其中, 由公式 (2-8) 知, $SST = \sum_{i=1}^n (y_i - \bar{y})^2$, $SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ 。样本决定系数 R^2 可以更直观地反映回归方程的拟合效果, 取值在 0 到 1 之间, R^2 越接近于 1, 回归方程的拟合效果就越好; R^2 越接近于 0, 回归方程的拟合效果就越差。

2.3 逐步回归

2.3.1 逐步回归的基本思路

逐步回归就是通过每次添加或者是删除一个变量, 当模型满足某个判停规则时就停止。逐步回归又分为向前逐步回归、向后逐步回归以及向前向后逐步回归, 向前逐步回归是指模型每次添加一个自变量, 直到再添加变量不会让模型有所改进为止; 向后逐步回归是从模型中的全部自变量中, 每次删除一个自变量, 直到模型的质量降低为止; 向前向后逐步回归是在向前逐步回归和向后逐步回归的基础上, 每次进入一个变量, 并要对变量重新进行评价, 从而可以删除对模型影响不大的变量, 直到获得最优模型^[35]。

实际问题中, 我们都是先选出与因变量 y 有相关关系的变量作为自变量, 然后通过多元线性回归分析建立最优模型, 最后再进行预测。逐步回归中, 引入一个变量或者删除一个变量, 模型都要进行 F 检验, 这样会在引入新变量前或者删除变量后, 保证回归方程中的自变量对因变量 y 有显著影响, 而不具有显著关系的自变量已经被剔除^[33]。

逐步回归实质上就是在最小二乘理论基础上进行的线性回归, 其基本思想就是删除一个变量或者增加一个变量的过程, 重复进行这一过程, 直至所有与因变量 y 之间呈显著关系的自变量均被引入到回归模型中, 而与因变量 y 没有显著关系的自变量均被剔除, 此时逐步回归就完成了^[36]。

2.3.2 逐步回归的主要过程

假设回归方程中已经有 $b-1$ 个变量, 得到矩阵 A , 将新变量 X_k 放进方程

中，计算偏 F 统计量

$$F = \frac{SSR(A) - SSR(A, X_k)}{\frac{SSE(A, X_k)}{n-b-1}} = \frac{SSR(X_k|A)}{MSE(A, X_k)} \quad (2-9)$$

依据显著性水平，给出用于 F 检验的两个判断值 F_1 和 F_2 ， F_1 主要用来选择自变量， F_2 用来剔除自变量，一般情况下都是令 F_1 和 F_2 相等，接下来再按如下过程进行计算^[36]。

(1) 拟合仅含一个自变量 $X_k (k = 1, 2, \dots, n)$ 的一元线性回归方程

$$Y = \beta_0 + \beta_k X_k + \varepsilon \quad (2-10)$$

计算每个自变量对应的偏 F 统计量

$$F_k^{(1)} = \frac{SSR(X_k|A)}{MSE(A, X_k)}, \quad k = 1, 2, \dots, m \quad (2-11)$$

当 X_k 进入方程后，残差误差平方和变小了，设

$$F_{k_1}^{(1)} = \max_{1 \leq k \leq m} F_k^{(1)} \quad (2-12)$$

并且当 $F_{k_1}^{(1)} > F_1$ 时，将含有自变量 X_{k_1} 的回归模型作为新模型，即

$$Y = \beta_0 + \beta_{k_1} X_{k_1} + \varepsilon \quad (2-13)$$

相反，对于所有的自变量，都有 $F \leq F_{k_1}^{(1)}$ ，那么就不需要引入其他自变量了，回归分析就结束了。

(2) 基于过程 (1)，将剩下的 $m - 1$ 个自变量依次引入到回归方程中，计算

$$F_k^{(2)} = \frac{SSR(X_k|X_{k_1})}{MSE(X_{k_1}, X_k)}, \quad k \neq k_1 \quad (2-14)$$

设

$$F_{k_2}^{(2)} = \max_{k \neq k_1} F_k^{(2)} \quad (2-15)$$

如果 $F_{k_2}^{(2)} > F_1$ ，那么将变量 X_{k_2} 放进过程 (1) 建立的模型中，即

$$Y = \beta_0 + \beta_{k_1} X_{k_1} + \beta_{k_2} X_{k_2} + \varepsilon \quad (2-16)$$

如果 $F_{k_2}^{(2)} \leq F_1$ ，那么意味着回归结束，上一过程所建的回归模型为最优回归模型。当 X_{k_2} 进入方程后，计算

$$F_{k_1}^{(2)} = \frac{SSR(X_{k_1}|X_{k_2})}{MSE(X_{k_1}, X_{k_2})} \quad (2-17)$$

如果 $F_{k_1}^{(2)} \leq F_2$ ，那么应该去除变量 X_{k_1} ，此时回归模型只包含 X_{k_2} 。

(3) 基于过程 (2) 建立的模型，将自变量逐个放入到模型中，并计算相应模型的偏 F 统计量，再与 F_1 相比较，确定变量是否能进入到回归方程中。当该自变量能够引入到模型中时，再验证是否要删除原模型中的自变量。直到模型不再引入新变量，或者也不需要删除自变量时，此时得到的回归模型就是最优模型。

2.4 多重共线性

多重共线性表示回归模型中的两个或多个自变量之间呈现出高度相关性，从而使模型结果失真^[36]。一般实际经济问题中常会出现多重共线性，这是因为经济变量之间常存在内在联系、滞后变量的引入以及样本资料的限制都可能使自变量之间存在多重共线性。

2.4.1 多重共线性的后果

1、完全共线性下，参数估计量是不存在的

回归模型 $Y = X\beta + \varepsilon$ 的普通最小二乘估计量为 $\hat{\beta} = (X'X)^{-1}X'Y$ ，但是在完全共线性下， $(X'X)^{-1}$ 是不存在的，也就是无法得到参数的估计量。

2、近似共线性时，参数的普通最小二乘估计量是非有效的

近似共线性下，参数估计量的方差为 $Cov(\hat{\beta}) = \sigma^2(X'X)^{-1}$ ，由于 $|X'X|$ 约等于 0，导致 $(X'X)^{-1}$ 主对角线上的元素较大，这会增大参数估计值的方差，得到的普通最小二乘参数估计量就不是有效估计量。

3、变量的显著性检验失去意义

因为当存在多重共线性时，参数估计值的方差会变大，这会使样本计算出来的 t 统计量值小于临界值，从而做出错误的推断，很有可能排除重要的自变量。

4、模型的预测失效

因为多重共线性会使模型参数估计值的方差变大，预测区间就容易变大，从而导致模型预测不准确。

2.4.2 多重共线性的检测

多重共线性可用方差膨胀因子 (VIF) 来进行检测， VIF 是指变量回归参数的置信区间能膨胀为与模型无关的自变量的程度，一般情况下，当 $0 < VIF < 10$ 时，认为不存在多重共线性；当 $10 < VIF < 100$ 时，表示自变量之间存在较强的多重共线性； $VIF \geq 100$ 时，表明自变量间存在严重的多重共线性^[37]。

2.5 本章小结

本章主要对论文所涉及的相关知识进行了概述，主要包括相关分析和回归分析，相关分析有 Spearman 秩相关、偏相关以及相关性的显著性检验，回归分析包括多元线性回归、逐步回归和多重共线性的一些理论知识。

第3章 哈尔滨 PM_{2.5} 浓度变化的影响因素研究

3.1 哈尔滨市数据来源

通过使用软件 Python 自带的爬虫软件库 BeautifulSoup 解析中国天气网网页 (<http://www.weather.com.cn/>) 以及 windows7 定时计划任务来抓取哈尔滨五个区域 (道里、南岗、道外、平房、松北) 的实时天气数据, 并将其存贮在 SQL Sever 数据库中。数据抓取的时间维度为 2018 年 3 月 1 日至 2019 年 2 月 28 日, 其中, 每隔两个小时产生 5 条观测, 数据采集截止时间为 2019 年 2 月 28 日 22:00, 一共采集到 19819 条观测数据。

3.2 PM_{2.5} 浓度变化的时间规律

3.2.1 PM_{2.5} 浓度的日均值变化

基于哈尔滨市 2018 年 3 月 1 日至 2019 年 2 月 28 日的 PM_{2.5} 浓度数据, 绘制出日均 PM_{2.5} 浓度变化曲线, 如图 3-1 所示。

从图中可以看出, PM_{2.5} 浓度日均值呈锯齿型变化, 变化范围为 5 ~ 460 $\mu\text{g}/\text{m}^3$, 日均浓度为 42 $\mu\text{g}/\text{m}^3$ 。根据 2012 年中华人民共和国环境保护部修订的《环境空气质量标准》, 我国 PM_{2.5} 日均值标准为 35 $\mu\text{g}/\text{m}^3$ [4]。依据此标准来衡量哈尔滨市 PM_{2.5} 浓度达标情况, 其中日均浓度超标率为 36.7%, 整体来看, 哈尔滨市空气污染情况还是比较可观的。

从 PM_{2.5} 浓度日均值变化曲线可以看出, 比较大的 PM_{2.5} 日均值浓度集中在 2018 年的 3 月 ~ 4 月、11 月 ~ 12 月以及 2019 年的 1 月 ~ 2 月, 这段时间总共统计了 178 天 PM_{2.5} 浓度日均值, 其中有 117 天的 PM_{2.5} 日均浓度超过日均值标准, 超标率为 66%, 其中, 2019 年 2 月份有一天的 PM_{2.5} 浓度日均值高达 460 $\mu\text{g}/\text{m}^3$, 是日均值标准的 13 倍, 空气严重污染。分析可知, PM_{2.5} 浓度较大的这几个月, 哈尔滨正处于供暖期, 燃煤数量增加, 大量燃

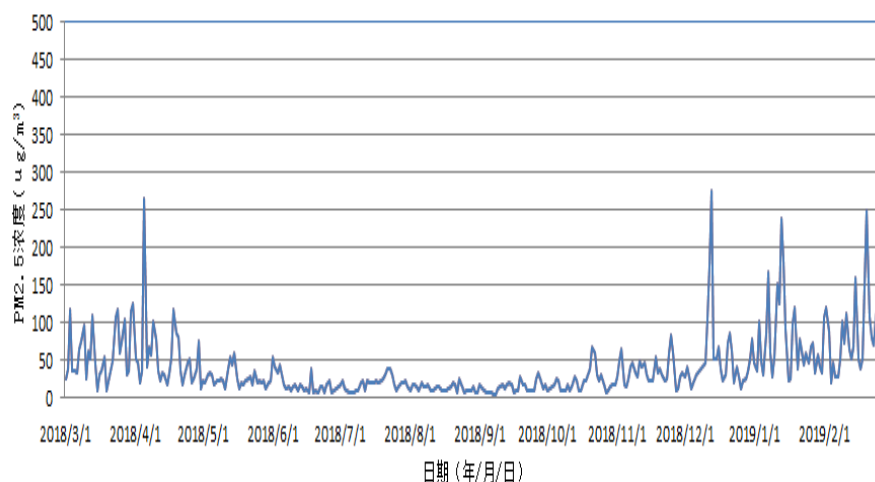


图 3-1 PM2.5 日均浓度变化曲线

Figure 3-1 Daily average concentration curve of PM2.5

煤产生的有害物质排放到空气中，导致颗粒物浓度增加，空气污染比较严重。然而，2018 年 5 月到 10 月期间的 PM2.5 日均浓度变化较为平稳，尤其是在夏季（6 月 ~ 8 月），PM2.5 日均浓度基本在 6 到 $54 \mu\text{g}/\text{m}^3$ 之间变化，仅有 6 天的 PM2.5 浓度日均值超过了日均值标准 $35 \mu\text{g}/\text{m}^3$ ，究其原因，这很有可能是夏季温度高，降水比较频繁，使 PM2.5 浓度有所下降。

3.2.2 PM2.5 浓度的昼夜变化

统计哈尔滨全天各时间点 PM2.5 的平均浓度值，来研究其昼夜变化规律，其折线图如图 3-2 所示。由图 3-2 可知，哈尔滨全天的 PM2.5 浓度呈双峰变化，分别在中午 12:00 和夜间 2:00 达到峰值。从图中可以看出，PM2.5 浓度从早上 8:00 开始上升，到中午 12:00 达到最高，然后下降，在下午 18:00 左右出现了一天中的最低浓度值，即谷值，随之，PM2.5 浓度又开始上升。

经分析这种变化类型是由于早上 8:00 到晚上 22:00，人们需要进行大量户外活动，早上人们要乘车或者驾车去上班，汽车尾气、扬尘以及车辆产生的颗粒物排放量增加，Wehner B^[38] 和 M O Nkk O Nen P^[39] 表明城市中的颗粒物粒径较小，不易扩散，大量的颗粒物排放使 PM2.5 浓度上升。中午 12:00 以后，随着气温升高，太阳辐射增强，相对湿度逐渐降低，风速不断加大，这样的气象条件

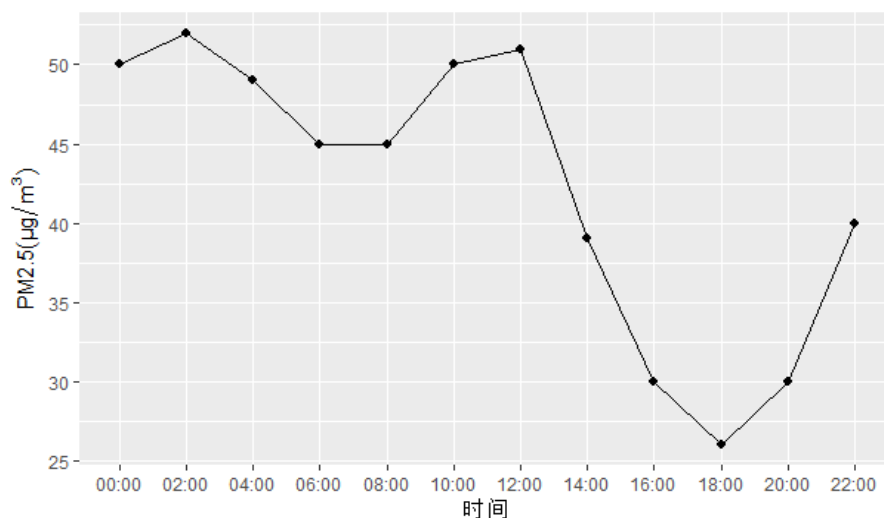


图 3-2 PM2.5 浓度昼夜变化曲线

Figure 3-2 The curve of PM2.5 concentration change in day and night

有利于污染物扩散，所以 PM2.5 浓度开始下降，一般在下午 18:00 左右下降到最低浓度值。18 时之后，气温下降，相对湿度上升，风速一般也会减小，加之在 18:00 到 20:00 期间，人们下班要乘车回家，同时厨房油烟的排放导致 PM2.5 浓度开始上升。夜间以后，大气活动趋于稳定，这样会使白天产生的颗粒物聚集在大气中，不利于污染物扩散^[40]，尤其是在晴天，夜间的地表温度可能会低于大气温度，从而形成逆温^[41]，这样更不利于扩散，从而导致凌晨 2:00 左右的 PM2.5 浓度达到了峰值。

3.2.3 PM2.5 浓度的一周变化

哈尔滨周一至周日 PM2.5 平均浓度值变化，如图 3-3 所示。

由图 3-3 可知，周一至周日 PM2.5 浓度呈单峰型分布，在周四达到峰值，其周均值为 $42 \mu\text{g}/\text{m}^3$ 。周二至周四 PM2.5 浓度逐渐上升，在周四达到最高浓度，但在周末，PM2.5 浓度快速下降。分析可知，这种变化规律可能是因为周一至周五属于工作日，上班乘车或者驾车排除的汽车尾气以及工厂排放的颗粒物都会使 PM2.5 浓度升高，很有可能在大气中积累，使周四 PM2.5 浓度出现了峰值，并且大气中的颗粒物一部分来源于交通排放，还有一部分是来源于固定源^[42]。而周末属于双休时间，车流量减少，同时，大部分工厂停止了生产运营，污染气

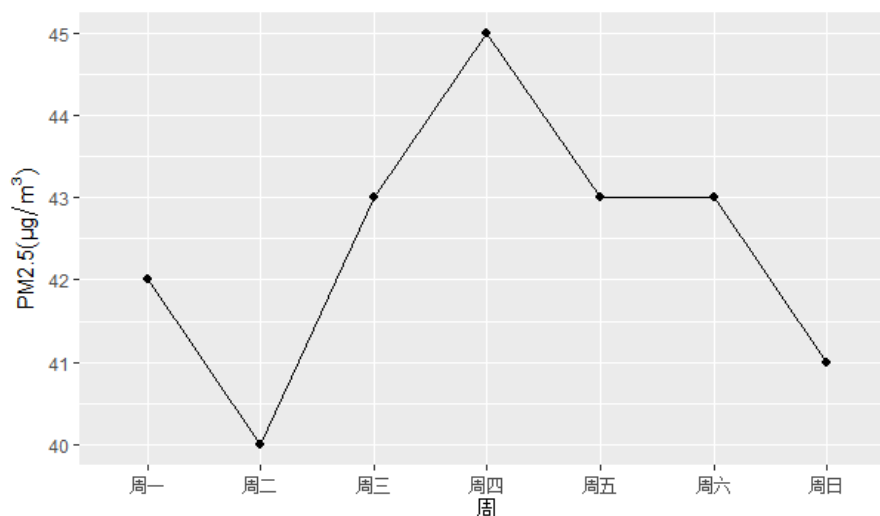


图 3-3 PM2.5 浓度的周变化曲线

Figure 3-3 The curve of PM2.5 concentration change for week

体的排放也会减少,有利于减轻大气对 PM2.5 稀释作用的压力,导致 PM2.5 浓度下降。另外,哈尔滨本地机动车周末不限号,外地汽车周末限号,机动车限号会在一定程度上减少一部分由车辆产生的颗粒物排放量。

将周一至周日 PM2.5 的平均浓度分成三个不同时间段来进行分析,具体分为仅工作日(周一~周五)、周末(周六~周日)、工作日和周末(周一~周日)。对不同时间段的 PM2.5 平均浓度进行比较,发现仅工作日的 PM2.5 平均浓度比仅周末的浓度高。这可能是由于仅工作日期间人口多、工业生产、交通运输集中、燃煤排放物大幅增加导致 PM2.5 浓度较高。

3.3 PM2.5 与气象因素之间的关系

3.3.1 与天气、风向以及风力的关系

大气中 PM2.5 浓度与气象因素之间有着密切的联系,气象条件影响着空气质量,空气质量水平同时影响着人们的身体健康^[43],并且空气质量对气象状况同样也存在影响。当大气中的污染物相对稳定时,污染物被稀释和扩散的能力主要受气象因素的影响^[44]。为了研究天气与 PM2.5 浓度之间的关系,对不同天气情况

下的 PM2.5 浓度进行整理，见表 3-1。图 3-4 为箱线图，更直观地表示了不同天气状况下 PM2.5 浓度的变化。

表 3-1 不同天气下的 PM2.5 平均浓度

Table 3-1 Average concentration of PM2.5 in different weather

编码	天气	PM2.5 平均浓度($\mu\text{g}/\text{m}^3$)	编码	天气	PM2.5 平均浓度($\mu\text{g}/\text{m}^3$)
1	大雨	15	9	雨夹雪	52
2	冻雨	92	10	阵雪	134
3	多云	38	11	中雪	51
4	雷阵雨	16	12	中雨	14
5	晴	40	13	阵雨	21
6	小雪	50	14	霾	252
7	小雨	29	15	雾	347
8	阴	50			

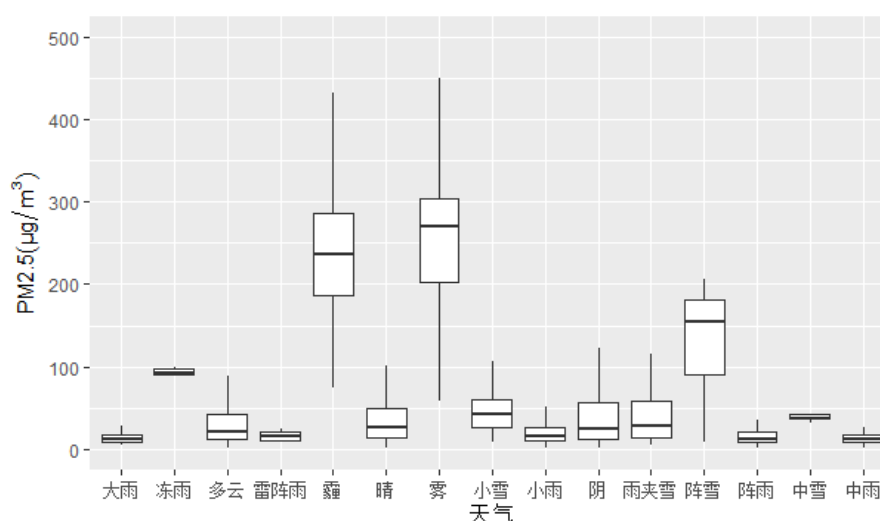


图 3-4 不同天气下的PM2.5浓度变化

Figure 3-4 Changes in PM2.5 concentration under different weather

从表 3-1 可以看出，出现霾、雾和阵雪天气时的 PM2.5 平均浓度比较高，分别为 $252 \mu\text{g}/\text{m}^3$ 、 $347 \mu\text{g}/\text{m}^3$ 、 $134 \mu\text{g}/\text{m}^3$ ，中雨、大雨、雷阵雨时的 PM2.5 平均浓度较低，分别为 $14 \mu\text{g}/\text{m}^3$ 、 $15 \mu\text{g}/\text{m}^3$ 、 $16 \mu\text{g}/\text{m}^3$ 。为此，可以得到不同降雨类型对应的 PM2.5 浓度值大小，即冻雨>小雨>阵雨>雷阵雨>大雨>中雨；不同

降雪类型对 PM_{2.5} 浓度的影响，即阵雪>雨夹雪>中雪>小雪。

分析天气情况可知，雾、霾天气下，空气受到不同程度的污染，严重污染时，PM_{2.5} 浓度高达 1335 $\mu\text{g}/\text{m}^3$ ，并且雾、霾天气主要集中在 2019 年的 1 月份、2 月份这两个月，此时哈尔滨处于供暖期，污染物排放较多，加上冬天气温较低，容易出现逆温现象。冬季天气干燥，空气流动缓慢，污染物主要集中在浅层大气中，容易形成霾，导致空气污染。此外，冬季夜间漫长，晴天风小出现的次数较多，不利于污染物的扩散，导致 PM_{2.5} 浓度升高。降雨对空气中的 PM_{2.5} 浓度具有一定的稀释作用，这是因为降雨可以直接清除空气中的一部分颗粒物，潮湿的地面也会吸附一定浓度的颗粒物^[45]。雨水对空气中的污染物有一定的清除作用，从而降低了污染物的浓度，雨水较大的情况下，空气污染物更能得到有效清除^[46]。

冻雨，也称“滴水成冰”，逆温层是其产生的重要条件之一，逆温层现象也就是空气上热下冷，这样就不会发生空气对流，导致空气中的水汽和颗粒物越积越多，不利于污染物扩散，从而使 PM_{2.5} 浓度升高。小雨使得空气湿度较大，有利于颗粒物的吸湿增长，PM_{2.5} 又是吸湿性气溶胶粒子，是大气中重要的凝结核，具有“吸湿增长”的化学特性^[47]，这会使大气中颗粒物的浓度增加，不利于颗粒物的扩散。但连续降雨时，PM_{2.5} 浓度也不会一直递减，而是下降后处于一个较低水平。

降雪对 PM_{2.5} 的清除作用没有降雨好，表明降雨比降雪更能有效降低 PM_{2.5} 浓度。另外可以看出降雪时的 PM_{2.5} 浓度普遍比其它天气情况下的浓度高，这可能与此时哈尔滨还处于供暖期，大量燃煤导致 PM_{2.5} 浓度较高，增加了其被稀释的压力。在晴天、多云天气条件下，PM_{2.5} 浓度相对于降水普遍较高，因为没有降水，地面更易扬尘，造成 PM_{2.5} 浓度升高。阴天条件下，PM_{2.5} 浓度相对来说比较低，这是因为阴天大部分出现在夏季（6 月~8 月），哈尔滨在夏季盛行东南风，空气比较洁净，大气的稀释作用也比较强，此外，夏季温度高，光照强，PM_{2.5} 浓度容易扩散，各种因素综合作用使其浓度普遍较低。也就是说，除了污染源排放到空气中的颗粒物对空气质量有影响外，气象条件也是影响 PM_{2.5} 浓度的重要因素^[48]。

此外，还统计了不同风向下的 PM_{2.5} 平均浓度，如表 3-2 所示。风影响着污染物的扩散和稀释，风向决定污染物的扩散方向，风力的大小决定污染物被稀

释的速度。在风的输送下，污染物一般会顺着下风方向扩散，风力越大，污染物扩散速度也越快，在移动的过程中，不断与周围空气发生碰撞，使得污染物不断被稀释，从而降低了 PM2.5 浓度。图 3-5 为不同风向下 PM2.5 浓度变化的箱线图。

表 3-2 不同风向下 PM2.5 浓度

Table 3-2 The concentration of PM2.5 in different wind direction

编码	风向	PM2.5 平均浓度($\mu\text{g}/\text{m}^3$)	编码	风向	PM2.5 平均浓度($\mu\text{g}/\text{m}^3$)
1	北风	34	5	南风	42
2	东北风	39	6	西北风	36
3	东风	45	7	西风	44
4	东南风	37	8	西南风	48

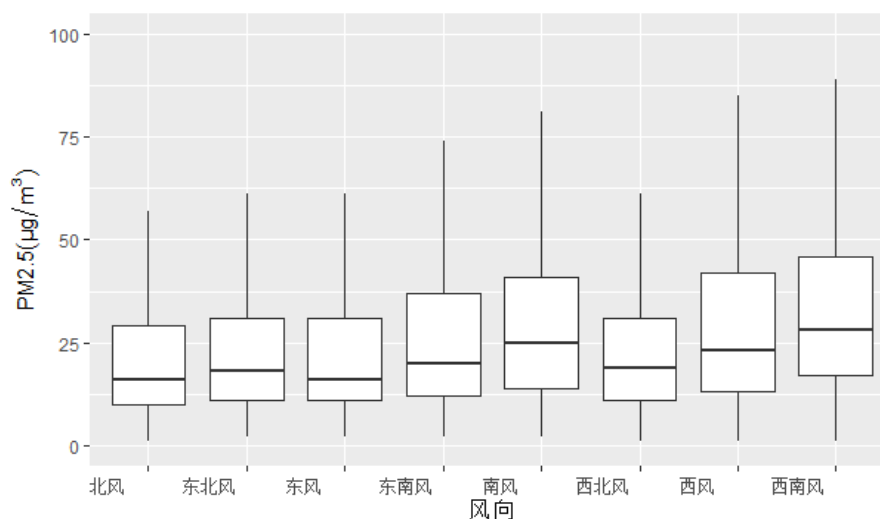


图 3-5 不同风向下PM2.5的浓度变化图

Figure 3-5 Concentration change of PM2.5 in different wind direction

由表 3-2 可知，西南风、东风下的 PM2.5 浓度较高，分别为 $48 \mu\text{g}/\text{m}^3$ 、 $45 \mu\text{g}/\text{m}^3$ ，北风时的 PM2.5 浓度相对来说比较低，所以可以推断北风方向的气流易于 PM2.5 的扩散，西南风、东风容易造成 PM2.5 的积累，不利于扩散。风向对 PM2.5 的影响是因为高污染浓度通常聚集在污染源的下风方向，哈尔滨的主导风向是西南风，工业区位于东北和西南方向，工厂排放出来的污染物会在风的作用下进入市

区，导致市区空气污染严重，这也与西南风方向下，PM2.5 浓度较高相吻合。

风力与 PM2.5 浓度也具有一定的关系，研究发现，当风力 ≤ 3 级时，PM2.5 浓度被稀释的作用不明显，当风力在 3 ~ 5 级时，PM2.5 浓度开始下降。PM2.5 粒径小，更容易在风的输送下运到其他地方，但是风力对 PM2.5 浓度的影响具有两面性，当风力过大时，地面就更容易扬尘，PM2.5 浓度明显升高，随着风力增大到 8 级，PM2.5 浓度会达到最高值，之后在高位震荡。还有一个方面就是当输送的空气比较干净时，风力的增大有利于污染物的扩散，PM2.5 浓度会有所降低，当输送的空气不干净时，地面上的尘土会加重污染，风力增强会使 PM2.5 浓度升高^[49]。

3.3.2 与气温、气压以及相对湿度的关系

气温在一定程度上会影响大气污染物的扩散和稀释。空气中悬浮粒子的布朗运动与气温有着紧密的联系，气温不同，粒子布朗运动的能力也不同，这样就会影响空气中悬浮粒子的浓度，低温时，近地面的悬浮粒子浓度较高，高空中的粒子浓度较低；相反，高温时，近地面的悬浮粒子浓度较低，高空中的粒子浓度会有所升高^[50]。

表 3-3 PM2.5 浓度与气象因素之间的相关系数

Table 3-3 Correlation coefficient between PM2.5 concentration and meteorological factors

气象因素	相关系数
日均气温	-0.40
日均气压	0.28
日均相对湿度	-0.19

由表 3-3 知，哈尔滨 PM2.5 日均浓度与气温具有负相关关系，并且通过了相关性检验。图 3-6 为气温随时间变化的关系，图 3-7 为 PM2.5 浓度与气温的相关性曲线。从图中可以看出，随着气温的升高，PM2.5 浓度在下降，尤其是在夏季（6 月~8 月），这种变化更明显，9 月至 10 月期间，哈尔滨气温在下降，对应的 PM2.5 浓度有所上升。

因为哈尔滨地理位置特殊，位于中国的最北端，是纬度最高、气温最低的大

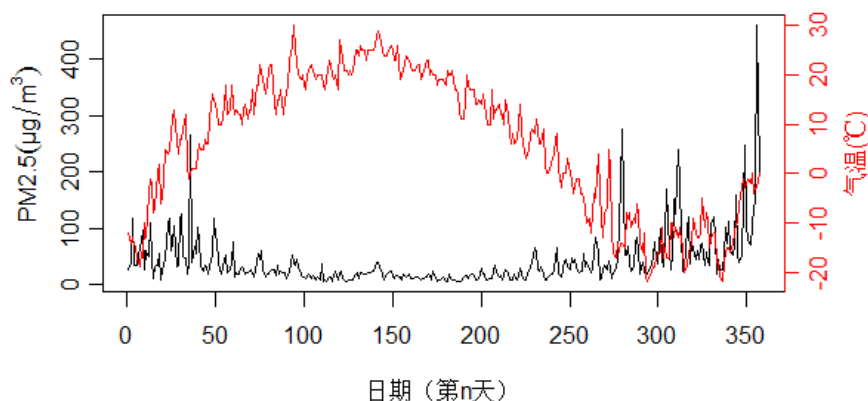


图 3-6 PM2.5 浓度与气温的关系

Figure 3-6 Relationship between PM2.5 concentration and temperature

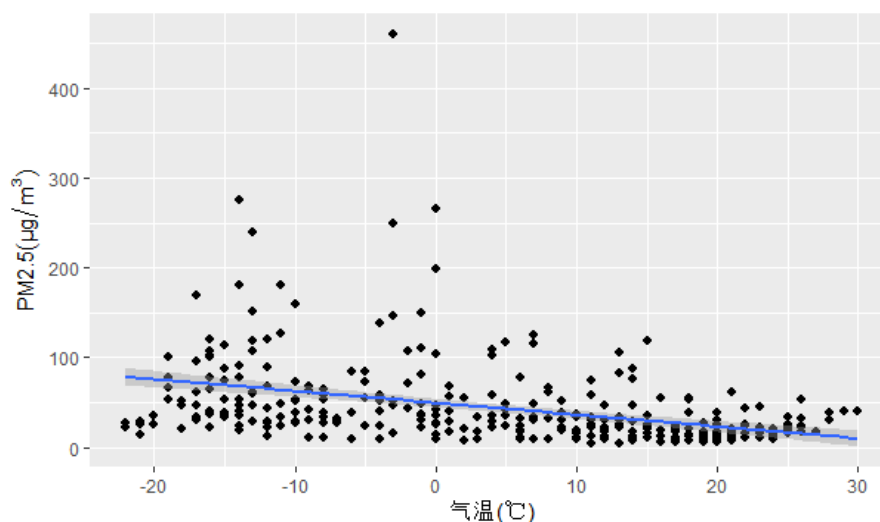


图 3-7 PM2.5浓度与气温相关性曲线

Figure 3-7 Correlation curve between PM2.5 concentration and temperature

城市，冬季漫长寒冷，从图中可以看出，2018 年 3 月以及 2019 年 1 月到 2 月，哈尔滨的 PM2.5 日均浓度较高，此时气温一般都在 10° 以下。因此 PM2.5 日均浓度与气温呈负相关关系的原因可能是当气温升高时会使得近地面大气变得活跃，悬浮粒子的布朗运动能力变强，加快了 PM2.5 的扩散，导致 PM2.5 浓度下降；还有一个方面就是温度升高，混合层上升，大气在竖直方向上运动，混合层越高，污染物越容易被大气稀释，同样会降低 PM2.5 浓度^[51]。相反冬季气温低，燃煤

供暖增加了 PM2.5 的来源，并且这个季节的大气层相对来说比较稳定，静稳状态的大气不利于 PM2.5 的扩散和稀释，在静风或微风条件下，容易造成 PM2.5 在城区和近郊区积聚，从而导致其浓度上升。

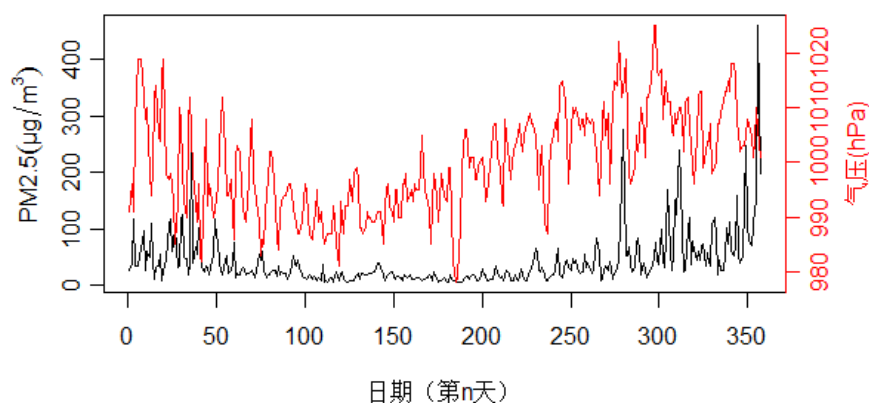


图 3-8 PM2.5 浓度与气压的关系

Figure 3-8 Relationship between PM2.5 concentration and atmospheric pressure

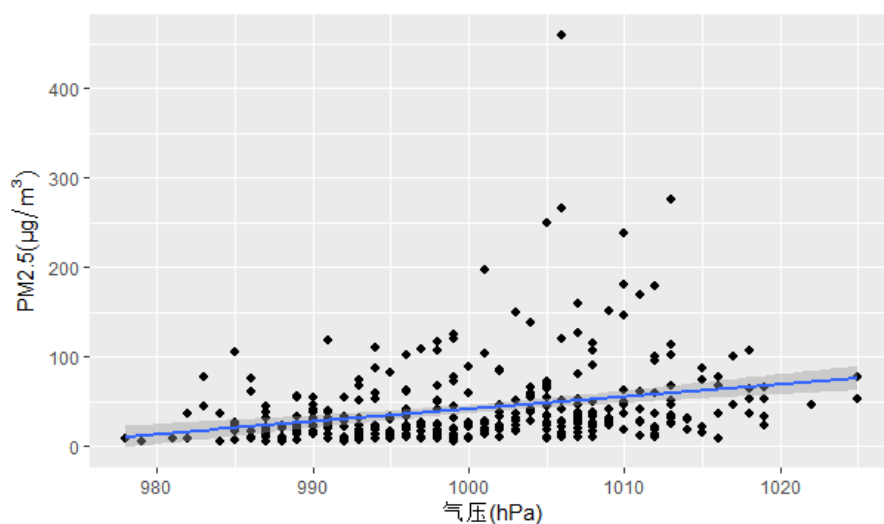


图 3-9 PM2.5浓度与气压相关性曲线

Figure 3-9 Correlation curve between PM2.5 concentration and atmospheric pressure

气压受地面特征与其他气象因素影响，与气温、云团的关系比较密切。由表 3-3 知，PM2.5 日均浓度与气压呈正相关，相关系数为 0.28，并且通过了相关

性检验。图 3-8 为气压随时间变化的关系，图 3-9 为 PM2.5 浓度与气压的相关性曲线。从图 3-8 可以看出，哈尔滨在 2018 年 3 月、12 月以及 2019 年 1 月、2 月的气压偏高，受高压系统控制，气团比较稳定，晴朗天气较多，高气压导致空气下沉，容易形成逆温层，空气中的 PM2.5 难以扩散和被稀释就会被阻滞在低空或者近地面，导致其浓度升高^[52]。5 月到 8 月哈尔滨气压比较低，此时哈尔滨受低压系统控制，大气一般不太稳定，在垂直方向上频繁运动，降雨天气较多，空气湿度大，PM2.5 得到了很好的扩散和稀释。正如图 3-9 描述的 PM2.5 浓度与气压之间的关系，随着气压的升高，PM2.5 浓度也在升高，两者呈明显的正相关关系。

相对湿度是大气中水汽压与饱和水汽压的比值，用百分数表示，反映了实际空气距离饱和空气的程度。以往研究发现，灰霾天气与相对湿度与也具有一定的关系，可以依据相对湿度的大小来确定灰霾程度^[53]。并且相对湿度与气温和降水具有密切的关系。由表 3-3 知，PM2.5 日均浓度与相对湿度呈负相关，其相关系数为 -0.19，并且通过了显著性水平为 0.01 的相关性检验。与气温、气压相比，PM2.5 日均浓度与相对湿度的相关关系不是很强。图 3-10 为相对湿度随时间变化的关系，图 3-11 为 PM2.5 浓度与相对湿度的相关性曲线。

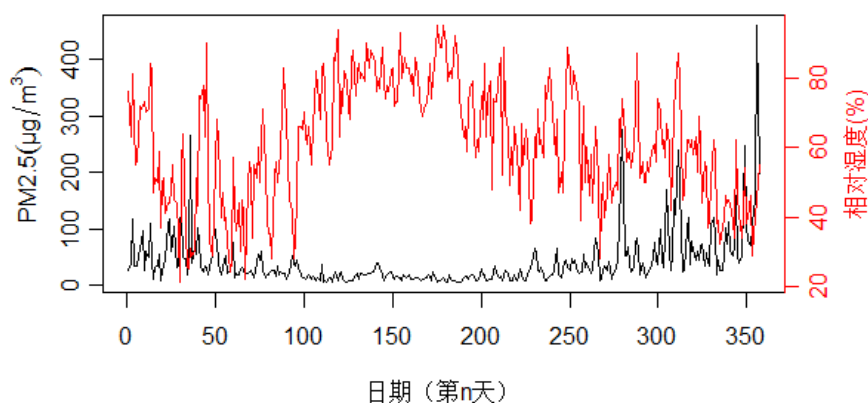


图 3-10 PM2.5 浓度与相对湿度的关系

Figure 3-10 Relationship between PM2.5 concentration and humidity

由图 3-10 知，当相对湿度大于 80% 时，PM2.5 日均浓度大幅下降，表明在

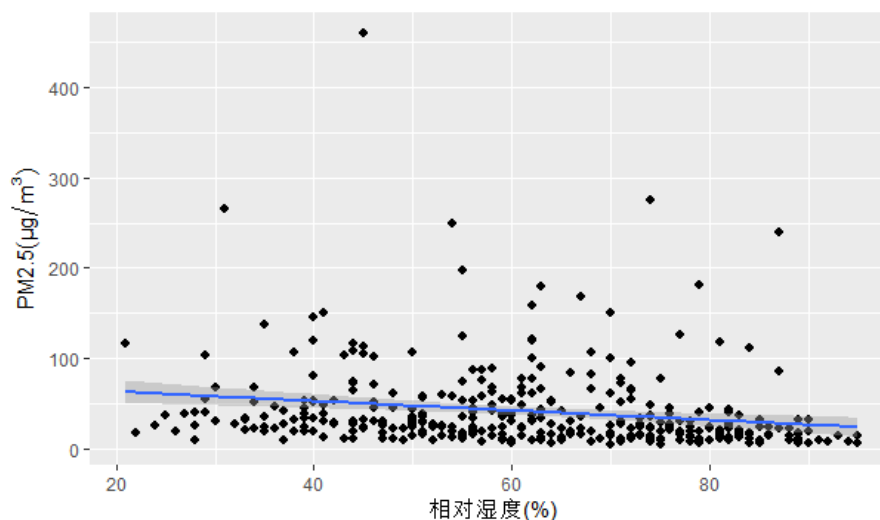


图 3-11 PM2.5浓度与相对湿度相关性曲线

Figure 3-11 Correlation curve between PM2.5 concentration and humidity

相对湿度较大的条件下, PM2.5 得到了明显的清除, 导致其浓度下降。其中, 相对湿度较大的时间集中在夏季 (6 月~8 月), 因为哈尔滨的气候属中温带大陆性季风气候, 夏季气候温热湿润多雨, 降水比较多, 占全年降水量的 60%, 而冬季的湿度最小, 冬季天气寒冷干燥, 对应的 PM2.5 浓度较高。分析原因可能是因为湿度较大时, 大气中水汽含量就会增多, PM2.5 容易被液滴吸附, 再沉降到地面, 导致其浓度下降, 另外, 当相对湿度大于 80% 时, 容易形成降雨, 不同程度的降雨对 PM2.5 具有清除作用。

3.3.3 与气温、气压、相对湿度以及风力的偏相关分析

基于前面研究结果可知, PM2.5 浓度在一定程度上受气温、气压、相对湿度以及风力的影响, 对其之间的相关性进行可视化, 如图 3-12 所示。其中, 对角线显示的是气压、相对湿度、气温、PM2.5 与风力的自身分布, 可以看出, PM2.5 浓度呈右偏分布; 下三角形 (对角线的左下方), 给出了两个变量的散点图, 比如第四行第三列的散点图显示 PM2.5 与气温之间存在比较明显的负相关关系; 上三角形 (对角线的右上方) 中的数字指的是两个变量间的相关系数, 型号表示显著程度 (星星越多表明越显著)。由图 3-12 可知, 各气象因素之间也相互影响, 存在比较明显的相关关系, 比如气压与气温、相对湿度之间的相关系

数为 -0.75、-0.33，相对湿度与风力之间的相关系数为 -0.29。

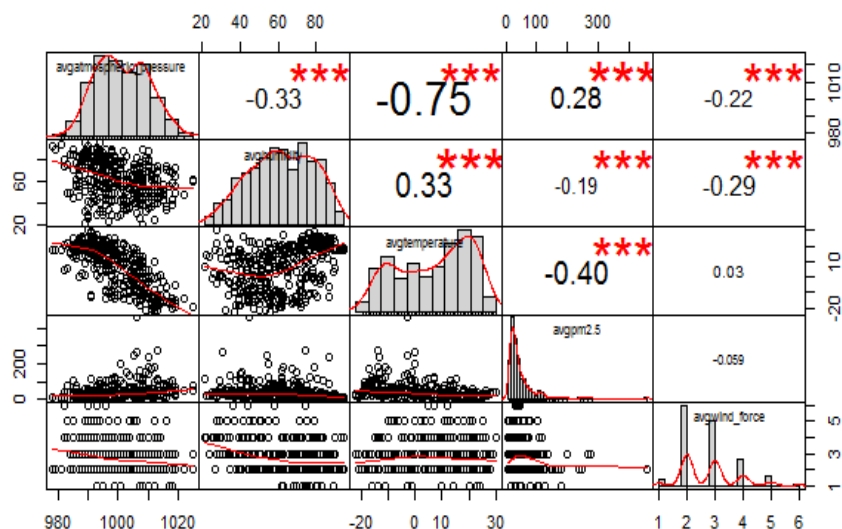


图 3-12 PM2.5 与气象因素相关图

Figure 3-12 Correlation curve between PM2.5 and meteorological factors

以往的研究主要集中在单一气象因素对 PM2.5 浓度的影响，但是 PM2.5 浓度的变化是受多个气象因素的综合影响。因此研究单一气象因素对 PM2.5 浓度的影响无法全面反映 PM2.5 与气象因素之间的关系。故本文采用偏相关分析来研究 PM2.5 与气象因素之间的关系，计算 PM2.5 与气温、气压、风力以及相对湿度之间的偏相关系数，结果见表 3-4。

表 3-4 PM2.5 与气象因素偏相关性分析

Table 3-4 Analysis of partial correlation between PM2.5 and meteorological factors

影响因素	偏相关系数	相关系数	控制变量
相对湿度	-0.13	-0.19	气温、气压、风力
气温	-0.31	-0.40	气压、风力、相对湿度
气压	-0.08	0.28	气温、风力、相对湿度
风力	-0.10	-0.06	气温、气压、相对湿度

由表 3-4 可知，控制气温、气压和风力这三个变量后，PM2.5 与相对湿度的偏相关系数为 -0.13，与控制变量前的相关系数 -0.19 相比，变化并不明显，说明

相对湿度与 PM2.5 的关系不太受气温、气压和风力的影响；同理，在控制气压、风力、相对湿度后，PM2.5 与气温的偏相关系数为 -0.31，与控制变量前的相关系数 -0.40 相比，变化也并不太明显，说明 PM2.5 与气温之间的关系并不受气压、风力、相对湿度的影响，同理，风力在控制变量前与控制变量后变化也不大。但是，在没有控制气温、风力、相对湿度前，PM2.5 与气压的相关系数为 0.28，控制这三个变量后，PM2.5 与气压的偏相关系数为 -0.08，说明气温、风力、相对湿度影响了气压与 PM2.5 浓度之间的关系。总而言之，PM2.5 与气温、气压、风力、相对湿度的关系不是简单地用一对一分析就能说明的，其浓度的变化受多个气象因素的共同影响。

3.4 基于主要气象因素的多元线性回归分析

3.4.1 模型建立

由前面研究可知，PM2.5 日均浓度受气温、气压、相对湿度以及风力的综合作用影响，为此，基于气温、气压、相对湿度以及风力这些气象因素，以 PM2.5 日均浓度为因变量，建立 PM2.5 和气象因素的多元线性回归模型。在建立模型前，首先对因变量进行正态性检验，如图 3-13 所示。

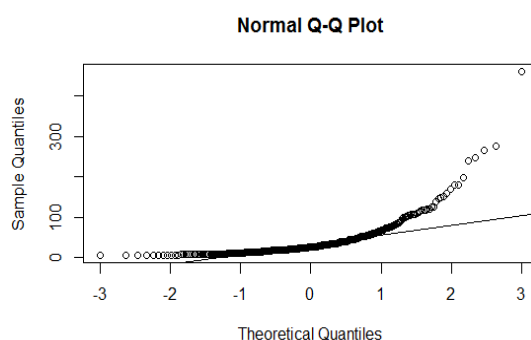


图 3-13 PM2.5 的正态 Q-Q 图

Figure 3-13 Normal Q-Q diagram of PM2.5

图 3-13 为 PM2.5 的正态 Q-Q 图，其中，纵坐标表示分布分位数，横坐标表示样本分位数，要判断 PM2.5 是否近似于正态分布，可以观察 Q-Q 图上的点，

看这些点是否近似地分布在一条直线上。由图 3-13 知，Q-Q 图上的点明显偏离直线，表明 PM2.5 不服从正态分布。为此，可以对 PM2.5 取自然对数，再对取完对数后的 PM2.5 进行检验。由图 3-14 知，取过对数后的 PM2.5 近似服从正态分布。

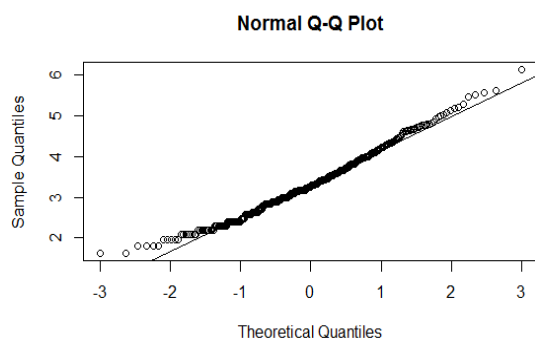


图 3-14 $\log(\text{PM}_{2.5})$ 的正态 Q-Q 图

Figure 3-14 Normal Q-Q diagram of $\log(\text{PM}_{2.5})$

表 3-5 PM2.5 浓度线性回归模型系数

Table 3-5 Linear regression model coefficients of PM2.5 concentration

参数	估计值	标准差	t-统计量	p 值
常数项	13.251437	6.671798	1.986	0.047788 *
气温	-0.031905	0.004045	-7.887	3.93e-14 ***
风力	-0.027780	0.043107	-0.644	0.519709
气压	-0.009081	0.006569	-1.382	0.167739
相对湿度	-0.008952	0.002505	-3.573	0.000402 ***

表 3-6 PM2.5 浓度的最优线性回归模型系数

Table 3-6 Optimal linear regression model coefficients for PM2.5 concentration

参数	估计值	标准差	t-统计量	p 值
常数项	4.009885	0.141694	28.300	<2e-16 ***
气温	-0.027914	0.002762	-10.107	<2e-16 ***
相对湿度	-0.007982	0.002299	-3.471	0.000582 ***

表 3-7 自变量的 VIF 值

Table 3-7 The independent variable's VIF value

自变量	气温	相对湿度
VIF	1.123847	1.123847

基于 2018 年 3 月至 2019 年 2 月的 PM2.5 浓度、气象因素数据，用统计软件 R 语言将其处理为日均值，采用多元线性回归方法，建立哈尔滨 PM2.5 日均浓度的对数线性回归模型，参数估计结果见表 3-5。为了进一步说明各气象因素对 PM2.5 浓度的影响，采用向后逐步回归方法，分析气温、风力、气压、相对湿度与 PM2.5 之间的关系。开始时模型包含 4 个（全部）自变量，模型对应的 AIC(AIC 越小越好)为 -255.72，逐步回归过程中，变量气压以及风力被删除，删除后模型的 AIC 从 -255.72 降低到 -256.19，然后再删除变量，模型的 AIC 会增加，因此终止选择过程。通过逐步回归分析，获得了最优多元线性回归模型，模型的参数估计结果见表 3-6。

整体来看，模型的回归系数是显著的，并且在显著性水平 0.01 下，PM2.5 与气温、相对湿度有显著的线性关系，即回归方程是显著的。用 R 语言进行多重共线性检验，检验结果见表 3-7。一般认为， $\sqrt{VIF} > 2$ 就表示存在多重共线性问题，由表 3-7 知，自变量是不存在多重共线性问题的。

3.4.2 模型检验

对模型进行检验，检验实质上是基于多元线性回归模型的假设进行的。

假设 1：自变量与因变量之间呈线性关系。

假设 2：残差基本呈正态分布。

假设 3：残差方差具有同方差性。

图 3-15 主要用来检验假设 1，整体来看，残差和拟合值没有明显的系统关系，说明模型中的自变量能够解释大部分 PM2.5 浓度的变化。图 3-16 用来检验假设 2，从图中可以看出，残差的分布尾部概率较小，但没有明显偏离正态性，说明残差正态性良好。

图 3-17 主要用来检验假设 3，可以看出，残差基本沿着预测值均匀分布，没

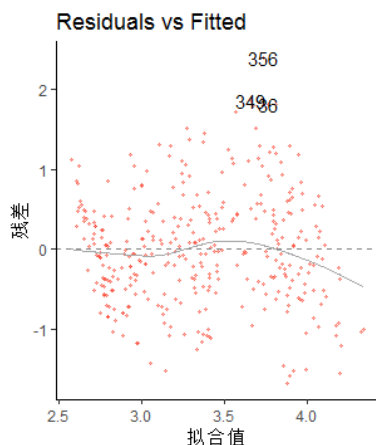


图 3-15 残差vs 拟合值

Figure 3-15 Residual vs fit value

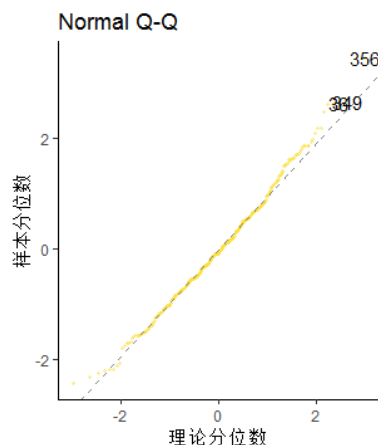


图 3-16 残差 Q-Q 图

Figure 3-16 Residual Q-Q diagram

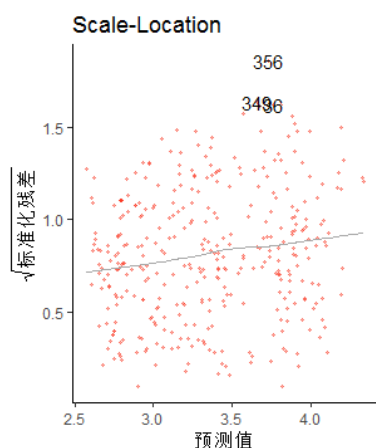


图 3-17位置尺度图

Figure 3-17 Location scale diagram

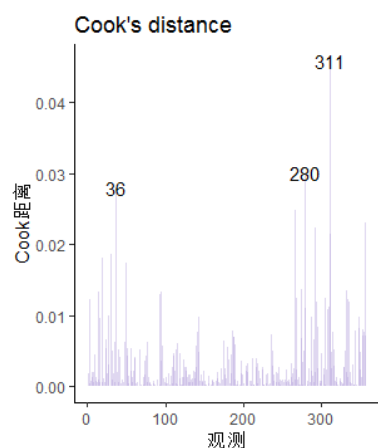


图 3-18 Cook距离图

Figure 3-18 Cook distance diagram

有明显的异方差性。图 3-18 主要用来检验异常值，异常值与三个概念有关：

- (1) 离群点，模型预测效果不佳的点。
- (2) 高杠杆值点，表示自变量值出现异常的点。
- (3) 强影响点，影响模型参数估计值的点。

从图 3-18 中可以看出，有 3 个样本像强影响点，在其他检验图中也有出现，查看原始数据发现，与相邻日期的 PM2.5 浓度相比，这 3 个样本中的 PM2.5 浓度出现了比较大的波动，其浓度值要么过大，要么过小。将其剔除后，对模型影响不大，所以可保留。

3.4.3 模型分析

实际应用中,有时收集到的样本量不充足,为了客观评价模型的预测效果,需要用到交叉验证法,该方法指的是先挑选出来一定比例的样本作为训练集,其余样本作为测试集,用训练集进行回归分析,然后在测试集上做预测。因为测试集不用选择模型参数,所以能够获得更精确的估计^[54]。

K 折交叉验证是对交叉验证法的推广,该方法是轮流将 $k-1$ 个子样本组合作为训练集,另外 1 个样本作为保留集。最后会得到 k 个回归方程,对这 k 个保留样本求平均值^[35]。对于小样本数据,该方法的检验结果比简单地将原始样本分为训练集和测试集得到的结果更为可靠^[55]。

交叉验证可以验证回归模型效果的好坏,对回归模型的拟合优度 R^2 进行 k 重交叉验证,利用 R 语言 bootstrap 包中的 crossval 函数来实现,取 $k=10$ 。结果发现,原始的 $R^2=0.304$,交叉验证后的 $R^2=0.288$, R^2 减少了 0.016。对于交叉验证, R^2 减少的越少,预测越精确。从结果来看, R^2 减少的较少,所以可认为该回归模型较好。

3.5 本章小结

本章主要对哈尔滨 PM2.5 浓度变化的影响因素进行了研究,首先分析了 PM2.5 浓度变化的时间规律,发现其浓度呈现下午低,晚间高的规律,工作日的浓度比周末时的浓度高,然后分析了 PM2.5 与气象因素之间的关系,结果表明,不同的天气、风向、风力对 PM2.5 浓度变化都有一定的影响,最后基于这些因素建立了 PM2.5 的多元线性回归模型。

第4章 气象因素对 PM2.5 浓度影响的季节性差异

4.1 PM2.5 的时间分布特征

4.1.1 PM2.5 随季节的变化规律

夏季哈尔滨天气凉爽，冬季漫长寒冷，春季和秋季，气温变化幅度较大，此时属于过渡季节，持续时间比较短。为此，受不同季节的影响，PM2.5 浓度与气象因素变化存在较大的差异性，将时间划分为春（2018 年 3 月~2018 年 5 月）、夏（2018 年 6 月~2018 年 8 月）、秋（2018 年 9 月~2018 年 11 月）、冬（2018 年 12 月~次年 2 月）四个季节，然后对不同季节 PM2.5 和气象因素之间的关系进行研究，其中，PM2.5 浓度随季节变化的差异见表 4-1。

表 4-1 PM2.5 浓度变化的季节性差异

Table 4-1 Seasonal differences of PM2.5 concentration change

季节	平均浓度($\mu\text{g}/\text{m}^3$)	浓度波动范围($\mu\text{g}/\text{m}^3$)
春季	47	9~266
夏季	17	6~54
秋季	25	1~163
冬季	81	2~1335

由表 4-1 知，不同季节 PM2.5 平均浓度大小关系是：冬季 > 春季 > 秋季 > 夏季。并且冬季、春季的 PM2.5 浓度分别为 $81 \mu\text{g}/\text{m}^3$ 和 $47 \mu\text{g}/\text{m}^3$ ，均超过了季节标准限值（ $41.94 \mu\text{g}/\text{m}^3$ ），夏季浓度最低，平均浓度为 $17 \mu\text{g}/\text{m}^3$ ，秋季次之。PM2.5 浓度随季节变化情况如图 4-1 所示。

从图中可以看出，冬季的 PM2.5 平均浓度要比夏季高很多，这可能是因为进入冬季以来，气温较低，风速较小，伴随着居民供暖，燃煤数量增加，污染物排放到空气中，不易于扩散，以及多发的雾、霾天气导致 PM2.5 浓度升高^[56]。春季气温变化较快，升温或降温一次可达 10°C 左右，并且时常出现大风天气，空气中的沙尘较多，这样 PM2.5 就不容易被稀释，导致其浓度升高。夏季 PM2.5 浓

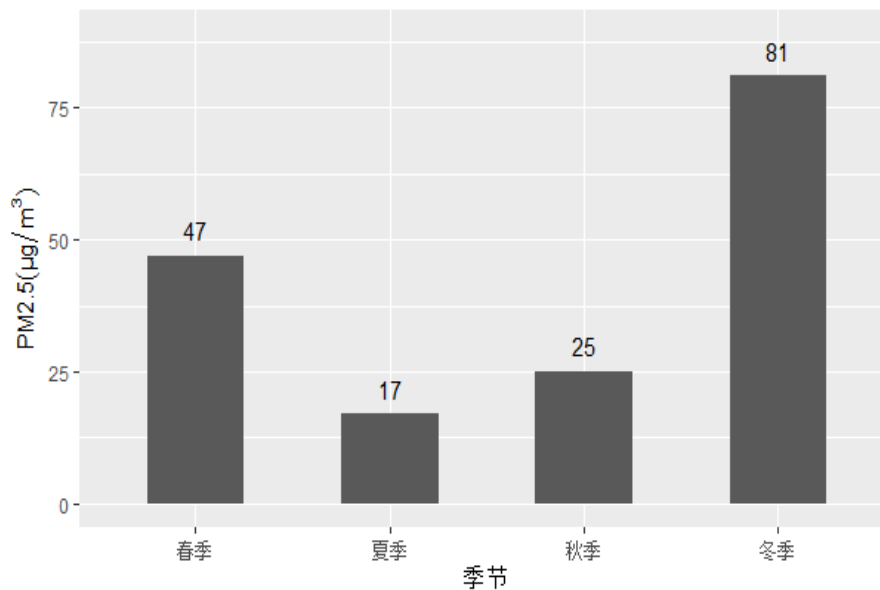


图 4-1 PM2.5 浓度的季节变化趋势

Figure 4-1 Seasonal trends of PM2.5 concentration

度相对较低，这是因为夏季温度较高，大气活动旺盛，不易形成逆温层，并且夏季降水较频繁，一定程度的降雨能很好地清除空气中的颗粒物^[57]，所以夏季空气比较湿润，易于 PM2.5 的吸收和扩散，同时没有燃煤排放的污染物，致使夏季的 PM2.5 浓度最低。

4.1.2 PM2.5 随月份的变化规律

为了进一步研究 PM2.5 的变化规律，对其月均值进行了整理，见表 4-2。

由表 4-2 知，2019 年 2 月 PM2.5 浓度最大，达到了 $108 \mu\text{g}/\text{m}^3$ ，2018 年 8 月的 PM2.5 浓度最低，最低浓度为 $13 \mu\text{g}/\text{m}^3$ 。这是因为 2018 年 3 月~5 月份随着气温升高，逆温时间变短，PM2.5 浓度开始下降；6~8 月份降水次数增多，降雨量也比较大，降水可以有效清除空气中的 PM2.5，所以在此期间，PM2.5 浓度普遍较低；9 月份天气比较凉爽，少雨，PM2.5 浓度略有升高；10 月份到次年 2 月份随着气温下降、逆温时间变长，加上哈尔滨市一般在 10 月中下旬开始供暖，导致 PM2.5 浓度迅速升高，在 2019 年 2 月其浓度值达到最高。

表 4-2 哈尔滨市各月份 PM2.5 浓度的变化情况

Table 4-2 Changes of Harbin's PM2.5 concentration in each month

月份	最小值	最大值	平均值
2018年3月	9	125	60
2018年4月	13	266	55
2018年5月	12	61	26
2018年6月	6	54	19
2018年7月	7	40	20
2018年8月	7	25	13
2018年9月	5	34	14
2018年10月	8	67	21
2018年11月	1	163	36
2018年12月	10	193	58
2019年1月	16	120	77
2019年2月	34	397	108

4.2 天气、风向对 PM2.5 浓度的影响

研究不同季节天气状况对哈尔滨 PM2.5 浓度的影响,如图 4-2、4-3、4-4、4-5 所示。可以看出,春季雨夹雪、冻雨会导致 PM2.5 浓度升高,冻雨天气出现在 2018 年 3 月 3 号,并且持续了两个小时,冻雨天气的出现使 PM2.5 浓度迅速升高,相对于两个小时前的 PM2.5 平均浓度为 $58 \mu\text{g}/\text{m}^3$ 而言,冻雨出现后其浓度升高到了 $92 \mu\text{g}/\text{m}^3$ 。而小雨天气下的 PM2.5 浓度最低,这可能是因为春季哈尔滨气温还是比较低的,大气运动缓慢,小雨天气会使空气湿度增大,使 PM2.5 易于被稀释,导致其浓度比较低。夏季,哈尔滨气候温热湿润多雨,平均降水量占全年的 60% - 70%,受雨水的冲刷作用,PM2.5 能够被有效清除,所以,夏季不同程度的降水对应的 PM2.5 浓度整体都比较低。

秋季天气凉爽,因为哈尔滨特殊的地理位置原因,10 月底开始就出现了不同程度的降雪,但是此时居民供暖增加了额外的污染物排放,所以,相对于春、冬两个季节,秋季 PM2.5 浓度比较低,但比夏季高。从 2018 年 10 月哈尔滨进入秋

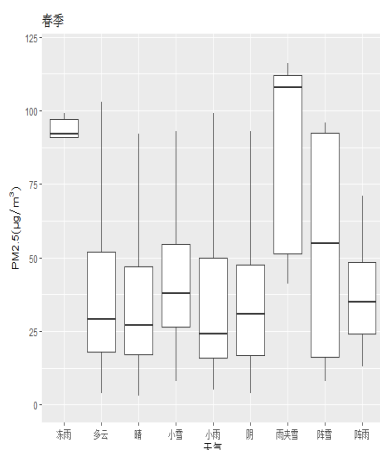


图 4-2 天气对 PM2.5 的影响

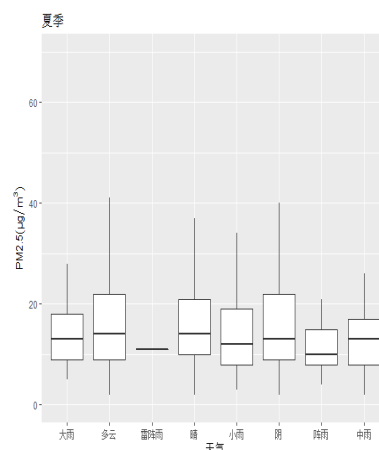


图 4-3 天气对 PM2.5 的影响

Figure 4-2 The effect of weather on PM2.5 Figure 4-3 The effect of weather on PM2.5

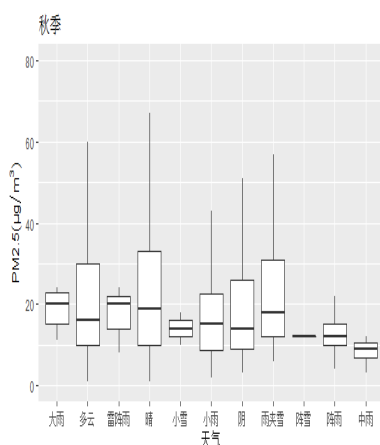


图 4-4 天气对 PM2.5 的影响

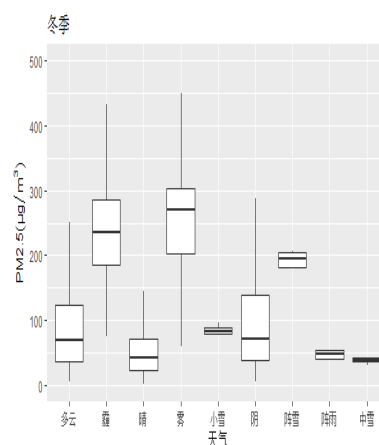


图 4-5 天气对 PM2.5 的影响

Figure 4-4 The effect of weather on PM2.5 Figure 4-5 The effect of weather on PM2.5

冬交替季节后，出现了持续两个多月的好天气，空气质量整体上也一直维持优、良水平。但是在 12 月中旬期间，连续几天哈尔滨出现了严重的雾霾天气，平均空气质量指数超过 400，空气重度污染以及严重污染，分析空气污染原因，发现是燃烧秸秆所致，严重影响了空气质量。并且自哈尔滨进入冬季以来，雪天天气很少，空气寒冷干燥，容易发生逆温天气，这样的气象条件不利于颗粒物的扩散。尤其在 2019 年 1 月到 2 月期间，雾霾天气频繁发生，PM2.5 不易被稀释，导致其浓度升高。从图 4-5 可以看出，雾霾天气下的 PM2.5 要比其他天气条件下的浓度高很多，平均浓度高达 $285 \mu\text{g}/\text{m}^3$ 。

图 4-6、4-7、4-8、4-9 表示不同季节风向对 PM2.5 浓度的影响，哈尔滨春、夏、秋主导风向为西南风，冬季主导风向为西（东）北风。从图中可以看出，春、夏、秋季西南风方向下的 PM2.5 浓度普遍较高，冬季东北风方向的 PM2.5 浓度较高，而四个季节北（西北）风下的 PM2.5 浓度普遍较低。

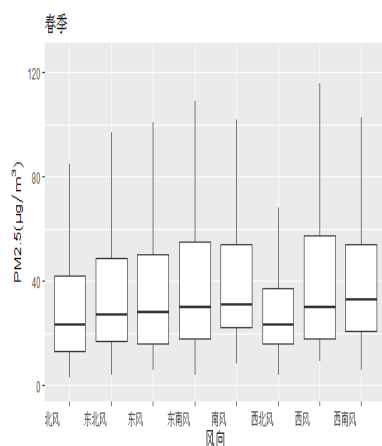


图 4-6 风向对 PM2.5 的影响

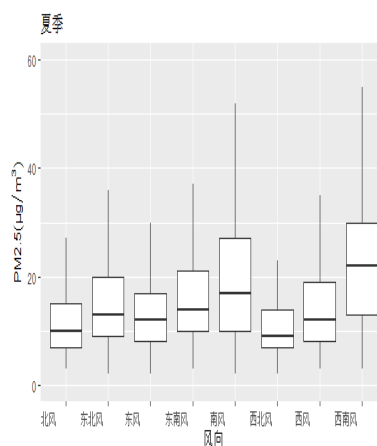


图 4-7 风向对 PM2.5 的影响

Figure 4-6 Effect of wind direction on PM2.5Figure 4-7 Effect of wind direction on PM2.5

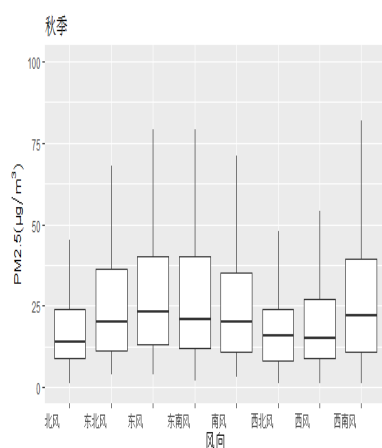


图 4-8 风向对 PM2.5 的影响

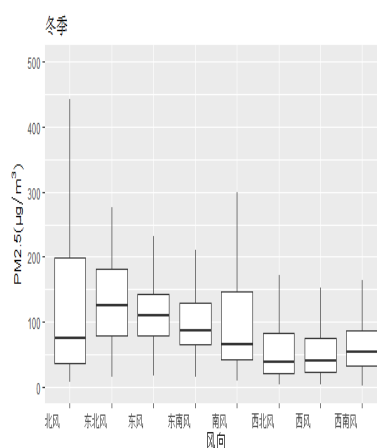


图 4-9 风向对 PM2.5 的影响

Figure 4-8 Effect of wind direction on PM2.5Figure 4-9 Effect of wind direction on PM2.5

4.3 气温、气压、相对湿度以及风力对 PM2.5 浓度的影响

基于哈尔滨 2018 年 3 月到 2019 年 2 月统计的气象数据（气温、气压、相对湿度、风力）的日平均值和 PM2.5 平均浓度，对其进行 Spearman 秩相关分析，结果见表 4-3。

表 4-3 气象因素与 PM2.5 浓度相关性

Table 4-3 Correlation between meteorological factors and PM2.5 concentration

季节	气温	气压	相对湿度	风力
春季	-0.27	-0.13	0.14	0.16
夏季	0.65	-0.26	-0.06	0.18
秋季	-0.44	0.51	-0.15	-0.10
冬季	0.30	-0.03	0.30	-0.39

由表 4-3 知，不同季节气象因素对 PM2.5 浓度的影响也有所不同。春季和夏季 PM2.5 浓度与风力都呈正相关关系，秋季和冬季 PM2.5 浓度与风力都呈负相关关系，这是因为春季天气比较干燥，风力较大易出现扬尘，导致空气中的颗粒物浓度增大；夏季空气比较湿润，风力相对较小，PM2.5 浓度较低。而在秋冬季节，伴随着不同程度的降雪，适当大的风力会加快 PM2.5 被稀释的速度，会使空气中的颗粒物浓度减小。

春季 PM2.5 浓度与气温呈负弱相关，与其它气象因素相关性都不太大；夏季 PM2.5 浓度与气温呈显著正相关，与气压呈负弱相关；秋季 PM2.5 浓度与气温存在显著的负相关关系，与气压之间呈显著正相关；冬季 PM2.5 浓度与气温呈正相关，与风力呈负相关，相比其他季节来讲，冬季风力对 PM2.5 浓度的影响更为显著。各月份气温、气压以及相对湿度与 PM2.5 浓度的关系如图 4-10、4-11、4-12 所示。

从图 4-10 中可以看出，春季随着温度的升高，PM2.5 浓度在下降，夏季 PM2.5 浓度与气温有两段变化趋势，6 月~7 月平均气温有小范围的上升，PM2.5 浓度也在上升，这可能是因为在晴朗、风力又比较小的天气下，空气比较稳定，不利于污染物扩散，造成 PM2.5 浓度有所升高；7 月~8 月平均气温在下降的同时，PM2.5 浓度也在下降，因为夏季降水较频繁，雨水的冲刷使空气

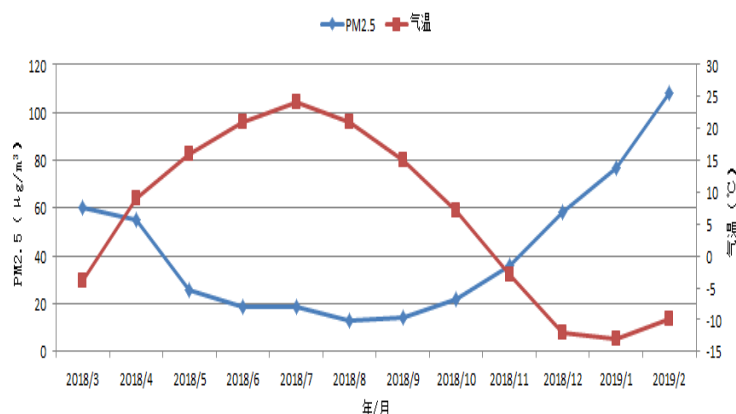


图 4-10 各月份气温与PM2.5 浓度变化

Figure 4-10 Changes of temperature and PM2.5 concentration in each month

中的 PM2.5 得到了有效清除，纵使气温有所下降，平均气温也在 19° 左右，温度高会使逆温层较弱，有利于污染物扩散，导致 PM2.5 平均浓度较低。秋季随着温度的下降，PM2.5 浓度在上升，两者呈负相关关系。10 月中旬期间，哈尔滨陆续进入供暖期，此时气温在逐渐下降，风力比较小，一般在 2 级左右，气象条件不利于污染物的扩散，导致 PM2.5 浓度升高。

进入冬季以来，气温一直在 0° 以下，2018 年 12 月至次年 1 月，气温在下降的同时，PM2.5 浓度在升高，这可能是因为供暖期居民取暖主要还是以燃煤为主，污染物排放到空气中，增加了 PM2.5 被稀释的压力；另外，因为室外温度较低，人们常驾车或者乘公共汽车出行，车辆排放的污染物也会导致其浓度升高。另外，在城市中，尤其是大城市的机动车，在上下班高峰期排放量明显增加^[58]，而且机动车尾气中的氮氧化物比燃煤还要高^[59]。而在 2019 年 1 月至 2 月期间，气温有所上升，PM2.5 浓度仍然在上升，一部分原因可能是春节过后，建筑工地开始施工，施工扬尘以及运输扬尘在风力的作用下被输送到空气中，加上降雪天气又比较少，导致 PM2.5 浓度升高。

从图 4-11 中可以看出，春夏季气压与 PM2.5 浓度之间的关系不是太显著，春季气压与 PM2.5 浓度的相关系数为 -0.13，夏季气压与 PM2.5 浓度的相关系数为 -0.26，都是呈弱相关关系，这可能是因为春夏季同一天的天气过程交替不明显，一天内的天气系统可以维持较长时间^[60]。秋季气压与 PM2.5 浓度呈显著正相关，相关系数为 0.51，相对于夏季，秋季气压较高，高压控制下的天气一般比

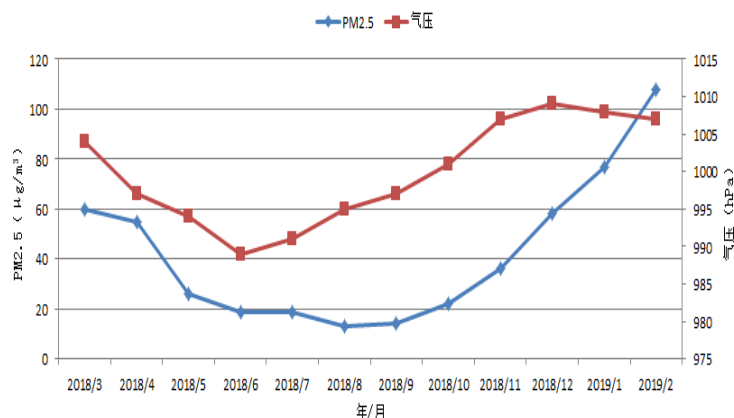


图 4-11 各月份气压与PM2.5 浓度变化

Figure 4-11 Changes of atmospheric pressure and PM2.5 concentration in each month

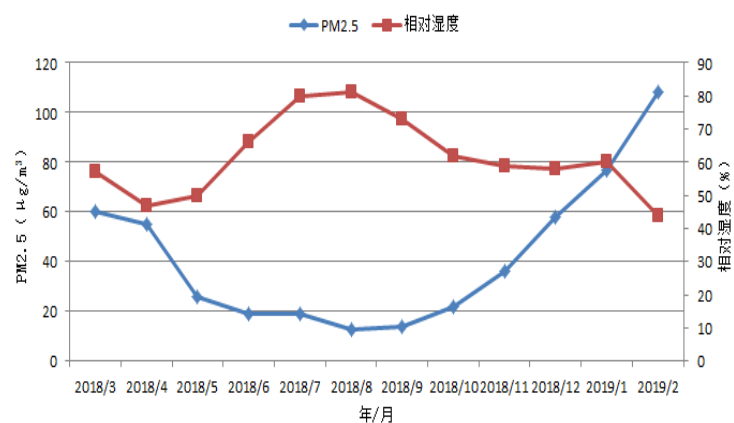


图 4-12 各月份相对湿度与PM2.5 浓度变化

Figure 4-12 Changes of humidity and PM2.5 concentration in each month

较晴朗，气流下沉不利于污染物扩散，较小的颗粒物集中在大气中，并且颗粒物质量在 PM2.5 总质量中的占比较高，从而加重空气污染^[61-63]。冬季 PM2.5 浓度与气压存在负相关关系，但并不显著。

从图 4-12 中可以看出，春季相对湿度与 PM2.5 浓度呈弱正相关关系；夏季相对湿度与 PM2.5 浓度呈弱负相关关系，因为夏季湿润，相对湿度较高，导致 PM2.5 浓度较低。秋季 PM2.5 浓度与相对湿度之间的相关关系不太显著，冬季相对湿度相较于秋季，有小范围的上升，但是冬季时常发生逆温现象，持续时间长，污染物不易扩散，容易积累，引起 PM2.5 浓度升高^[64]。

所以, 季节不同, 气象因素对 PM_{2.5} 浓度的影响也有所不同, 同时, 一些学者在对 PM_{2.5} 来源的研究中发现, 车辆排放的尾气、燃煤、扬尘都是其较大的贡献者^[65-69], 这些因素其实都不容忽视。

4.4 本章小结

本章主要对不同季节气象因素与 PM_{2.5} 浓度之间的相关关系进行了分析, 首先分析了 PM_{2.5} 浓度变化的时间分布特征, 然后又研究了不同季节天气状况、风向对 PM_{2.5} 浓度的影响, 计算出气温、气压、相对湿度、风力与 PM_{2.5} 浓度之间的 Spearman 秩相关系数, 发现季节不同, 气象因素对 PM_{2.5} 浓度的影响也存在差异。

结论与建议

随着经济的发展, 空气质量不断恶化, PM2.5 又是影响空气质量的重要因素。为此, 本文统计了哈尔滨市 2018 年 3 月至 2019 年 2 月 PM2.5 浓度和气象因素数据, 分析了哈尔滨 PM2.5 浓度变化的时间规律, 以及与气象因素(天气、风向、气温、气压、相对湿度、风力)之间的相关关系, 并基于气温、气压、相对湿度以及风力建立了多元线性回归模型, 最后分析了气象因素对 PM2.5 浓度影响的季节性差异。主要结论如下:

(1) 2018 年 3 月到 4 月哈尔滨 PM2.5 日均浓度普遍较高, 日均超标率达到了 64%, 5 月到 10 月期间 PM2.5 日均浓度变化较为缓慢, 日均浓度基本在 $50 \mu\text{g}/\text{m}^3$ 以下, 全天 PM2.5 浓度呈双峰型变化, 分别在夜间 2:00 和中午 12:00 达到峰值, 在下午 18:00 左右其浓度降到了最低, 按周一至周日来看 PM2.5 浓度变化, 发现工作日的浓度比周末时的浓度高。

(2) 研究 PM2.5 浓度与各气象因素之间的关系, 发现降雨对 PM2.5 具有一定的稀释作用, 降雪对其浓度的清除效果没有降雨好, 在风力达到一定程度的时候, PM2.5 浓度会有所下降, 但风力过大易出现扬尘, 导致空气中的污染物浓度升高, 并且相较于其他风向, 西南风方向下的 PM2.5 浓度最高。整体来看, PM2.5 浓度与气温呈负相关, 相关系数为 -0.40, 与气压呈正相关, 相关系数为 0.28, 与相对湿度呈弱负相关, 相关系数为 -0.19。

(3) 对 PM2.5 浓度与气温、气压、相对湿度、风力进行偏相关分析, 发现 PM2.5 浓度与气温的偏相关系数为 -0.31, 与气压的偏相关系数为 -0.08, 与相对湿度的偏相关系数为 -0.13, 与风力的偏相关系数为 -0.10, 说明 PM2.5 与气象因素的关系并不能简单地进行一对一分析, 其浓度变化受到多种气象因素的综合影响。

(4) 基于气温、气压、相对湿度、风力这些气象因素对 PM2.5 浓度建立对数线性回归模型, 并对 R^2 进行交叉验证, 从结果来看, 相对于原始的 $R^2 = 0.304$, 交叉验证后的 $R^2 = 0.288$, R^2 减少的比较少, 说明该回归模型的预测效果还是比较好的。

(5) 研究发现, 冬春季的 PM2.5 浓度明显比夏季高, 夏季 PM2.5 浓度与气

温呈显著正相关，与其它气象因素之间的关系不显著，秋季 PM_{2.5} 浓度与气压呈显著正相关，而在冬季，风力对 PM_{2.5} 浓度的影响较大。

针对论文研究成果，在改善哈尔滨空气质量方面，建议有关部门加大对不合格煤的检查力度，从源头控制空气污染；加强对机动车辆的环保检验，改进汽车尾气排放装置；通过制定一些补贴政策鼓励市民乘坐公共交通出行；加大监测秸秆燃烧力度，控制秸秆燃烧造成的大气污染。

综上所述，本文的研究工作即将结束，但是对 PM_{2.5} 浓度方面的研究还有待进一步完善，本文对 PM_{2.5} 浓度影响因素的分析相对较浅，考虑的气象因素比较少，希望在今后的研究中能够进一步深入对这方面的研究，同时希望在将来的研究中加入 PM_{2.5} 对人体健康和气候的影响。

参考文献

- [1] Mingxing Wang, Renjian Zhang. Frontier of atmospheric aerosols researches[J]. Climatic and Environmental Research, 2001, 6(1):119-124.
- [2] 吴海龙, 余新晓, 师忱, 等. PM_{2.5}特征及森林植被对其调控研究进展[J]. 中国水土保持科学, 2012, 10(6):116-122.
- [3] 张怡文, 胡静宜, 王冉. 基于神经网络的PM_{2.5}预测模型[J]. 江苏大学学报:自然科学版, 2015, 33(1):63-65.
- [4] 张振华. PM_{2.5}浓度时空变化特性、影响因素及来源解析研究[D]. 浙江大学, 2014.
- [5] Sawant A A, Na K, Zhu X, et al. Chemical characterization of outdoor PM_{2.5} and gas-phase compounds in Mira Loma, California[J]. Atmospheric Environment, 2004, 38(33):5517-5528.
- [6] Tran H N Q, Molders N. Investigations on meteorological conditions for elevated PM_{2.5} in Fairbanks, Alaska[J]. Atmospheric Research, 2011, 99(1):39-49.
- [7] Mejia J F, Wraith D, Mengersen K. Trends in size classified particle number concentration in subtropical Brisbane, Australia, based on a 5 year study[J]. Atmospheric Environment, 2006, 40:1064-1079.
- [8] Greene J S, Kalkstein L S, Ye H, et al. Relationships between synoptic climatology and atmospheric pollution at 4 US cities[J]. Theoretical & Applied Climatology, 1999, 62:163-174.
- [9] Flocas H, Kelessis A, Helmis C, et al. Synoptic and local scale atmospheric circulation associated with air pollution episodes in an urban Mediterranean area[J]. Theoretical & Applied Climatology, 2009, 95:265 - 277.
- [10] Johnson,R., Marsik,T., Lee,M., Cahill,C.F. Helping Fairbanks meet new air quality requirements: developing ambient PM_{2.5} management strategies[J].

- Transportation safety, security, and innovation in cold regions, 2009, 16(6):158-161.
- [11] 汪永英, 张雪梅, 韩冬荟, 等. 哈尔滨供暖期间PM2.5污染状况及其与气象因子的相关性分析[J]. 安徽农业科学, 2015,(19):185-189.
- [12] 王薇, 黄丽坤, 田甜. 哈尔滨市采暖期颗粒物数浓度变化特征[J]. 黑龙江大学学报, 2016,(2):46-51.
- [13] 周一敏, 赵昕奕. 北京地区PM2.5浓度与气象要素的相关分析[J]. 北京大学学报(自然科学版), 2017,(1):111-124.
- [14] 杜宇静, 李佩媛. PM2.5与气象要素之间相关关系的回归分析——以吉林市为例[J]. 科教导刊-电子版(中旬), 2017,(10):294-296.
- [15] 江鸿, 柳戊弼, 张翠荣, 等. 2015年武汉市PM2.5变化特征及其与气象要素的关系[J]. 安徽农业科学, 2016,(20):190-191,236.
- [16] 徐敬, 丁国安, 颜鹏, 等. 北京地区PM2.5的成分特征及来源分析[J]. 应用气象学报, 2007, 18(5):645-654.
- [17] 董雪玲, 刘大锰, 袁杨森, 等. 北京市2005年夏季大气颗粒物污染特征及影响因素[J]. 环境工程学报, 2007, 1(9):100-104.
- [18] Chu N, Kadane J B, Davidson C I. Identifying likely PM2.5 sources on days of elevated concentration: a simple statistical approach[J]. Environment Science & Technology, 2009, 43(7):2407-2411.
- [19] 王涛. 乌鲁木齐城市群大气质量及其与气象因素条件的相关性分析[D]. 新疆大学, 2015.
- [20] Perez P, Gramsch E. Forecasting hourly PM2.5 in Santiago de Chile with emphasis on night episodes[J]. Atmospheric Environment, 2016, 124:22-27.
- [21] 刘慧君. 武汉市PM2.5污染的演变预测及成因分析和仿真[D]. 湖南大学, 2014.
- [22] 毛恒青. 气象条件对大气污染物扩散的影响[J]. 山东气象, 1991(4):9-13.

- [23] Dawson J P, Adams P J, Pandis S N. Sensitivity of PM_{2.5} to climate in the Eastern U.S.:a modeling case study[J]. Atmospheric Chemistry&Physics, 2007, 7(16):4295-4309.
- [24] Donato A, Contini D, Belosi F. Real time measurements of PM_{2.5} concentrations and vertical turbulent fluxes using an optical detector[J]. Atmospheric Environment, 2006, 40(7):1346-1360.
- [25] 王艳秋, 杨晓丽. 哈尔滨市降水形势对大气污染物浓度稀释的影响[J]. 自然灾害学报, 2007, 16(5):65-68.
- [26] 姜新华, 刘霞, 薛河儒, 张存厚. 基于逐步回归的空气质量影响因素分析——以呼和浩特市为例[J]. 内蒙古农业大学学报(自然科学版), 2015, 36(02):123-126.
- [27] 彭金龙, 关欣, 杨思培, 王广, 李巧云. 气象因素对长沙颗粒物浓度影响的季节性差异分析[J]. 四川环境, 2017, 36(04):103-108.
- [28] 胡玉筱, 段显明. 基于高斯烟羽和多元线性回归模型的 $PM_{2.5}$ 扩散和预测研究[J]. 干旱区资源与环境, 2015, 29(06):86-92.
- [29] 谢心庆, 郑薇, 开璇, 等. 基于时间序列和多元方法的乌鲁木齐 $PM_{2.5}$ 浓度分析[J]. 云南大学学报:自然科学版, 2016, 38(4):595-601.
- [30] 张利田, 卜庆杰, 杨桂华, 等. 环境科学领域学术论文中常用数理统计方法的正确使用问题[J]. 环境科学学报, 2007, 27(1):171-173.
- [31] 陶澍. 应用数理统计方法[M]. 中国环境科学出版社, 1994.
- [32] 吴喜之, 王兆军. 非参数统计方法[M]. 高等教育出版社, 1996.
- [33] 滕浩宇. 基于常规大气监测数据对 $PM_{2.5}$ 的评估和预测[D]. 东北林业大学, 2015.
- [34] 何晓群, 刘文卿. 应用回归分析-第2版[M]. 中国人民大学出版社, 2007.
- [35] Kabacoff R, 卡巴科弗, Kabacoff, et al. R语言实战[M]. 人民邮电出版社, 2013.

- [36] 张亚吉. 基于主成分分析和神经网络的 $PM_{2.5}$ 浓度预测研究[D]. 燕山大学, 2017.
- [37] 樊聪聪. 跨国制造企业内部知识转移影响因素研究[D]. 北京交通大学, 2018.
- [38] Wehner B, Birmili W, Gnauk T, et al. Particle number size distributions in a street canyon and their transformation into the urban-air background: measurements and a simple model study[J]. Atmospheric Environment, 2002, 36(13):2215-2223.
- [39] M O Nkk O Nen P, Koponen I K, Lehtinen K, et al. Measurements in a highly polluted Asian mega city: observations of aerosol number size distribution, modal parameters and nucleation events[J]. Atmospheric Chemistry and Physics, 2005, 5(1):57-66.
- [40] 覃国荣. 保定市 PM_{10} 和 $PM_{2.5}$ 污染特征研究[D]. 河北大学, 2014.
- [41] 李小平, 司瑶冰, 杨武杰. 呼和浩特市冬季气象条件对空气质量的影响[J]. 干旱区资源与环境, 2009, 23(4):136-139.
- [42] Shi J P, Evans D E, Khan A A, et al. Sources and concentration of nanoparticles(< 10nm diameter) in the urban atmosphere[J]. Atmospheric Environment, 2001, 35(7):1193-1202.
- [43] Mnece W. Healthy risk of particulate matter from long range transboundary air pollution(preliminary assessment)[R]. Geneva: World Health Organization, 1999.
- [44] 李宗恺. 空气污染气象学原理及应用[M]. 气象出版社, 1985.
- [45] 席云. 大气中 PM_{10} 浓度与气象因素的关系[J]. 保山学院学报, 2004, 23(5):25-28.
- [46] 吴燊先, 董惠青, 黄香杏. 南宁市夏季影响大气污染的气象条件分析[J]. 广西气象, 2001, 22(4):34-36.

- [47] 江鸿, 柳戊弼, 张翠荣, 等. 2015年武汉市PM_{2.5}变化特征及其与气象要素的关系[J]. 安徽农业科学, 2016, 44(20):190-191.
- [48] 刘辉, 贺克斌, 马永亮, 等. 2008年奥运前后北京城、郊 PM_{2.5} 及其水溶性离子变化特征[J]. 环境科学学报, 2011, 31(1):177-185.
- [49] 蒋雷敏, 李佶. 天气因数对PM_{2.5}浓度的影响[J]. 科技视界, 2014(29):125-125.
- [50] 黄善斌. 关于PM_{2.5}与气象条件关系及其预报研究[D]. 齐鲁工业大学, 2017.
- [51] 黄金星, 张林, 陈欢林, 等. 杭州市区空气中PM_{2.5}细微粒监测及污染状况分析[J]. 环境科学与技术, 2006, 29(9):49-51.
- [52] 魏玉香, 童尧青, 银燕, 等. 南京 SO₂、NO₂和 PM₁₀ 变化特征及其与气象条件的关系[J]. 大气科学学报, 2009, 32(3):451-457.
- [53] 吴兑. 关于霾与雾的区别和灰霾天气预警的讨论[J]. 气象, 2005, 31(4):3-7.
- [54] 杨思琪. 随机森林模型在太原市细颗粒物浓度预测中的应用[D]. 太原理工大学, 2017.
- [55] 周启清, 李毓. 分类树集成算法在县域金融贷款风险分类评估中的应用[J]. 经济问题, 2009(12):94-97.
- [56] 李旭芳, 初钧晗, 于令达, 等. 北京市城区单监测点PM_{2.5}质量浓度变化特征[J]. 北京师范大学学报(自然科学版), 2011, 47(3):258-261.
- [57] 张静. 北京降水化学组成特征及其对颗粒物的去除作用[D]. 北京大学, 2005.
- [58] 陈丹青, 师建中, 肖亮洪, 等. 粤东3市PM_{2.5}和PM₁₀质量浓度分布特征[J]. 上海环境科学, 2013, 32(2):72-78.
- [59] Yao X, Chan C K, Fang M, et al. The water-soluble ionic composition of PM_{2.5} in Shanghai and Beijing, China[J]. Atmospheric Environment, 2002, 36(26):4223-4234.
- [60] 万欣, 王磊, 张亚楠, 等. 气象因素对南京市南郊大气PM_{2.5}质量浓度的影响[J]. 江苏林业科技, 2015, 42(6):1-4.

- [61] Pitz M, W.G.Kreyling, Holseher B, et al. Change of the ambient particle size distribution in East Germany between 1993 and 1999[J]. Atmospheric Environment, 2001, 35:4357-4366.
- [62] Clarle A.G., Azadi-Boogar G.A., Andrews G.E. Particle size and chemical composition of urban aerosols[J]. The science of the Total Environment, 1999, 235:15-24.
- [63] Tuch T., Brand P., Wichmann H.E., et al. Variation of Particle number and mass concentration in various size ranges of ambient aerosols in Eastern Germany[J]. Atmospheric Environment, 1997, 31:4193-4197.
- [64] Tao Chen, Jun He, Xiaowei Lu, et al. Spatial and Temporal Variations of P-M2.5 and Its Relation to Meteorological Factors in the Urban Area of Nanjing, China[J]. Environmenal Research and Public Health, 2016, 13(9):921.
- [65] Jurgen Schnelle-Kreis, Martin Sklocz, Jurgen Orasche, et al. Semi Volatile Organic Compounds in Ambient PM2.5 Seasonal Trends and Daily Resolved Source Contributions[J]. Environ. Sci. Technol, 2007, 41(11):3821-3828.
- [66] Robert A. Kotchenruther. Source apportionment of PM2.5 at Northwest U.S. sites: Assessing regional winter wood smoke impacts from residential wood combustion[J]. Atmospheric Environment, 2016, 142:210-219.
- [67] John G. Watson, Judith C. Chow, Douglas H. Lowenthal, et al. PM2.5 source apportionment with organic markers in the Southeastern Aerosol Research and Characterization(SEARCH) study[J]. Journal of the Air & Waste Management Association, 2016, 65(9):1104-1118.
- [68] Pathak R K, Wu W S, Wang T. Summertime PM2.5 ionic species in four major cities of China: nitrate formation in an ammonia-deficient atmosphere[J]. Atmospheric Chemistry & Physics, 2008, 9(5):1711-1722.
- [69] Lonati G, Giugliano M, Butelli P, et al. Major chemical components of PM2.5 in Milan(Italy)[J]. Atmospheric Environment, 2005, 39(10):1925-1934.

致谢

时光荏苒，两年的研究生生活就要结束，在黑龙江大学的点点滴滴都还历历在目，记得入学报到时对周围事物的那种好奇心与新鲜感，记得第一次与导师见面时的那种激动与紧张……。然而，我们终究要毕业，毕业意味着离别，新的征程即将开始，同时也要面对更多的挑战。回首这两年的学习生活，我不仅学到了更多专业方面的知识，还明白了许多为人处事的道理，很幸运遇到许多良师和益友，正是他们的帮助和鼓励，让我不断进步和成长。

首先，感谢我的导师马维军老师，马老师渊博的学术知识和严谨的治学态度给我留下了深刻的印象。在毕业论文的撰写过程中，尤其在承担繁重的教学、工作任务情况下，无论从方法研究还是文章内容，马老师都给予我很大帮助。从老师身上，不仅学到了许多专业知识，还明白了为人处世的道理。所以在此向我的导师表示诚挚的感谢！

同时，特别感谢王楠老师在学习、论文方面给予我的指导和帮助，王老师诲人不倦、平易近人的人格魅力对我影响很深。从论文选题到论文初稿再到定稿，每当我有不懂的问题时，老师总是耐心地讲解，细心地指导，并且提出宝贵的修改意见，每每想到这，感激之情油然而生。正是在老师耐心地指导和帮助下，我的论文才得以顺利完成。在此我再次向王楠老师表示由衷的感谢！

然后，我也要感谢研究生学习中教导过我的各位老师，钦佩各位老师的丰富学识，让我学到很多专业知识，这为我以后工作打下了坚实的基础。同时，还要感谢我的父母和朋友们的支持和鼓励，让我能够全身心地投入到学业中，顺利完成毕业论文。

最后，还要向百忙中抽出宝贵时间审阅和莅临答辩评审的各位老师和专家表示衷心的感谢！

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知,除了文中特别加以标注和致谢的地方外,论文中不包含其他人已经发表或撰写过的研究成果,也不包含为获得黑龙江大学或其他教育机构的学位或证书而使用过的材料。

学位论文作者签名:张凯丽

签字日期:2019年6月10日

学位论文授权使用授权书

本人完全了解黑龙江大学有关保留、使用学位论文的规定,同意学校保留并向国家有关部门或机构送交论文的复印件和电子版,允许论文被查阅和借阅。本人授权黑龙江大学可以将学位论文的全部或部分内容编入有关数据库进行检索,可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

学位论文作者签名:张凯丽

导师签名:



签字日期:2019年6月10日

签字日期:2019年6月10日

学位论文作者毕业后去向:

工作单位:

电话:

通讯地址:

邮编: